YACHAY-ETL: Sistema de Extracción Transformación y Carga para la Bodega de Datos del Registro Poblacional de Cáncer del Municipio de Pasto



MANUAL DE REFERENCIA

Universidad Nacional Abierta y a Distancia

Universidad de Nariño

INSTITUCIÓN UNIVERSITARIA COLEGIO MAYOR DEL CAUCA

El conocimiento es de todos

Minciencias

Contenido

1. Introducción 3
2. Objetivos del Proyecto 3
Objetivo General 3
3. Metodología de Trabajo 3
4. Resultados Obtenidos 3
5. Estructura y Arquitectura del Sistema 3
Desarrollo backend 4
Desarrollo frontend 8
6. Diseño de la Bodega de Datos 14
ANEXOS 59

1. Introducción

Durante el año 2025 se desarrolló el Sistema ETL para la Bodega de Datos del Registro Poblacional de Cáncer del municipio de Pasto, una aplicación web orientada a la visualización, gestión y ejecución de procesos ETL ya existentes, así como al acceso y descarga de tableros analíticos construidos en Power BI.

2. Objetivos del Proyecto

Objetivo General:
Desarrollar una aplicación web que permita visualizar y ejecutar scripts ETL existentes, así como facilitar la consulta y descarga de tableros analíticos.

3. Metodología de Trabajo

El trabajo se desarrolló de forma incremental, priorizando el diseño de la interfaz, la integración de scripts existentes, el desarrollo de funcionalidades de visualización y la validación del sistema mediante pruebas funcionales.

4. Resultados Obtenidos

- Aplicación web funcional para la gestión de procesos ETL existentes.
- Visualización estructurada de scripts SQL.
- Acceso centralizado a tableros analíticos.
- Manual de usuario para facilitar la adopción del sistema.

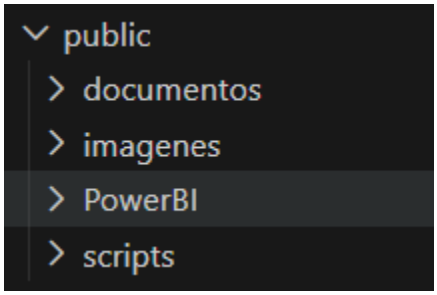
5. Estructura y Arquitectura del Sistema.

Para el desarrollo de la aplicación se realizó una sección backend la cual se desarrolló con lenguaje node.Js.

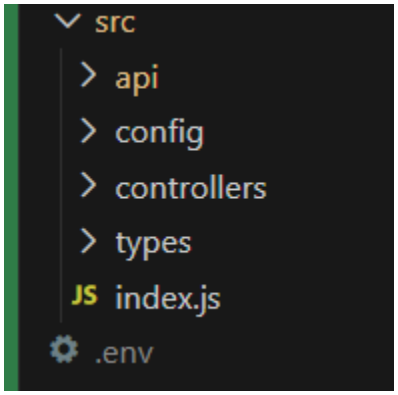
Una segunda sección frontend desarrollada con framework react. A continuación, realizaremos una descripción de cada sección.

Desarrollo backend

- Inicialmente se tiene una carpeta publica en la cual se alojan los documentos o recursos de interés común, como imágenes, reportes o documentos



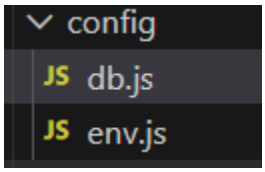
- Seguido se tiene la carpeta src donde se encuentra el código fuente y la configuración de nuestra aplicación back



- El archivo env contiene las variables de entorno, en nuestro caso únicamente las variables para la conexión a base de datos y el puerto del servidor

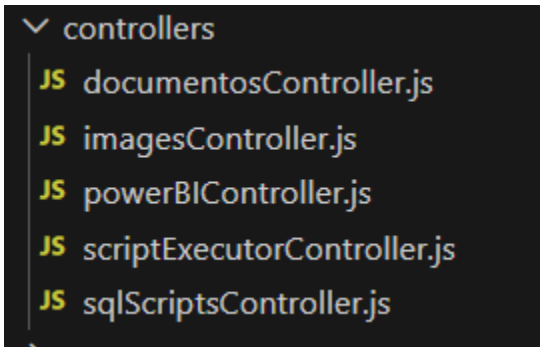
```
.env
Import to Postman
1  DB_USER=postgres
2  DB_HOST=10.10.10.117
3  DB_NAME=yachaydesarrollo
4  DB_PASSWORD=wyR4SbJFFHGaMnRZ1mqGOKc5
5  DB_PORT=5432
6  PORT = 3002
7
```

- La carpeta config, contiene la configuración completa para la base de datos.



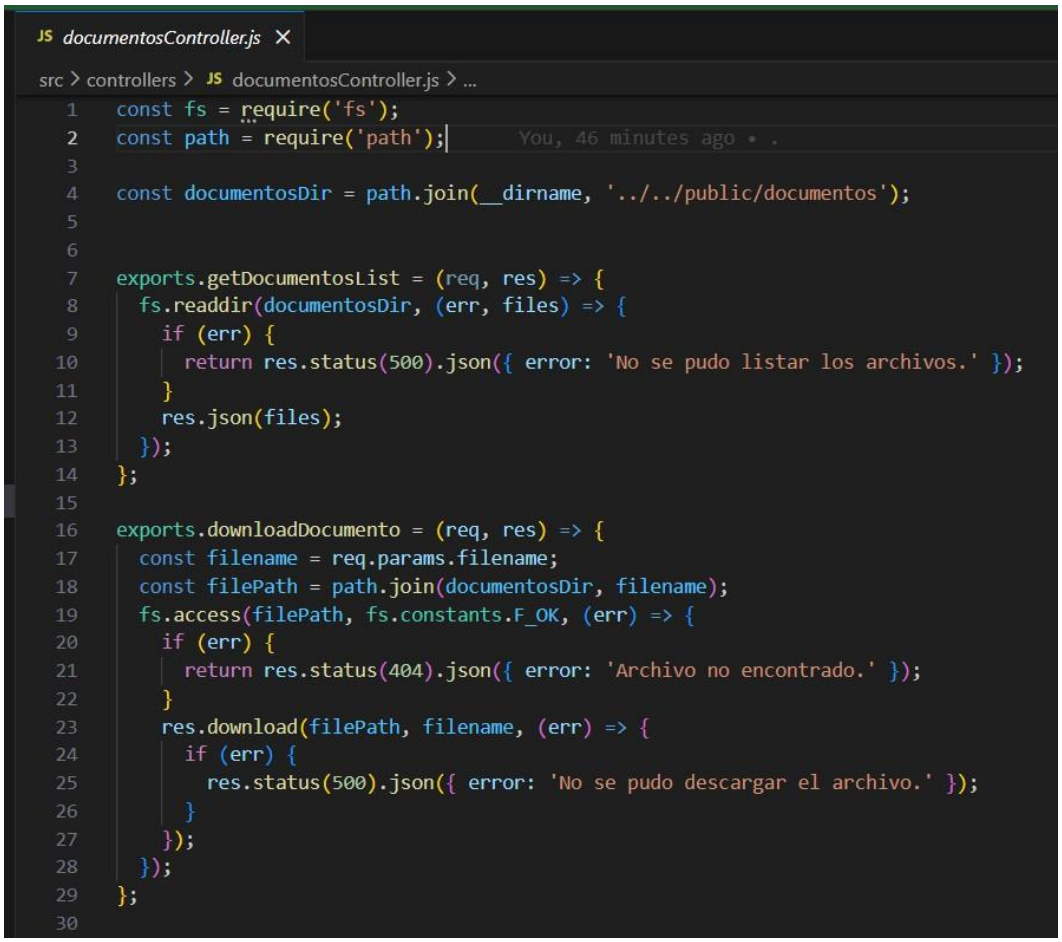

```
db.js
src > config > db.js > ...
You, last month | 1 author (You)
1  const { Sequelize } = require('sequelize');      You, last month • . …
2  require('dotenv').config();
3
4  // Crear la instancia de Sequelize usando las variables de entorno
5  const sequelize = new Sequelize(
6    process.env.DB_NAME,    // Nombre de la base de datos
7    process.env.DB_USER,    // Usuario de la base de datos
8    process.env.DB_PASSWORD, // Contraseña de la base de datos
9    {
10     host: process.env.DB_HOST,    // Host de la base de datos
11     port: process.env.DB_PORT,    // Puerto de la base de datos
12     dialect: 'postgres',
13     logging: false        //// Desactiva los logs de las consultas SQL
14   }
15 );
16
17 // Probar la conexión
18 sequelize.authenticate()
19   .then(() => {
20     console.log('Conexión Bd establecida');
21   })
22   .catch(err => {
23     console.error('No se logró establecer conexión a la Bd', err);
24   });
25
26 module.exports = sequelize;
27
```

- La carpeta controllers contiene todos los controladores necesarios para el funcionamiento de la aplicación.

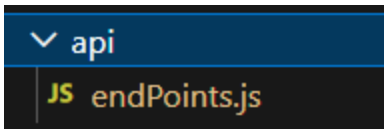


Cada controlador con la configuración necesario para el manejo de cada componente.

A continuación, se muestra uno de los controladores



- La carpeta api, contiene el archivo para los endpoint, los cuales son los enlaces para la aplicación front



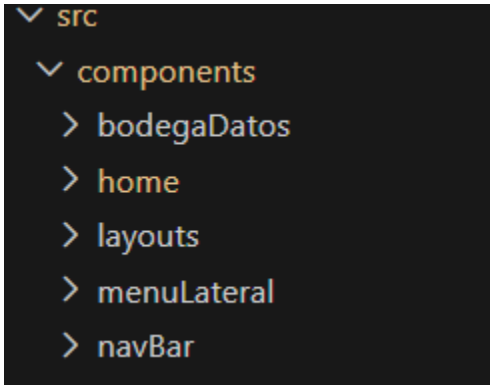
endPoints.js M

src > api > endPoints.js > ...

```
1   const express = require('express')
2   const router = express.Router()
3
4
5   const sqlScriptsController = require('../controllers/sqlScriptsController')
6   const powerBIController = require('../controllers/powerBIController')
7   const imagesController = require('../controllers/imagesController')
8   const scriptExecutorController = require('../controllers/scriptExecutorController')
9   const documentosController = require('../controllers/documentosController')
10
11  // Rutas para scripts SQL       You, last month • .
12  router.get('/sql-scripts', sqlScriptsController.getScriptsList)
13  router.get('/sql-scripts/:filename', sqlScriptsController.getScriptContent)
14
15  // Rutas para archivos Power BI
16  router.get('/powerbi-files', powerBIController.getPowerBIFilesList)
17  router.get('/powerbi-files/:filename/download', powerBIController.downloadPowerBIFile)
18
19  // Rutas para imágenes
20  router.get('/images', imagesController.getImagesList)
21  router.get('/images/:filename', imagesController.getImage)
22
23  // Ruta para documentos
24  router.get('/documentos', documentosController.getDocumentosList)
25  router.get('/documentos/:filename/download', documentosController.downloadDocumento)
26
27  // Rutas para ejecutar scripts SQL
28  router.post('/execute-script/:filename', scriptExecutorController.executeScript)
29  router.post('/execute-all-scripts', scriptExecutorController.executeAllScripts)
30
31
32  module.exports = router
33
```

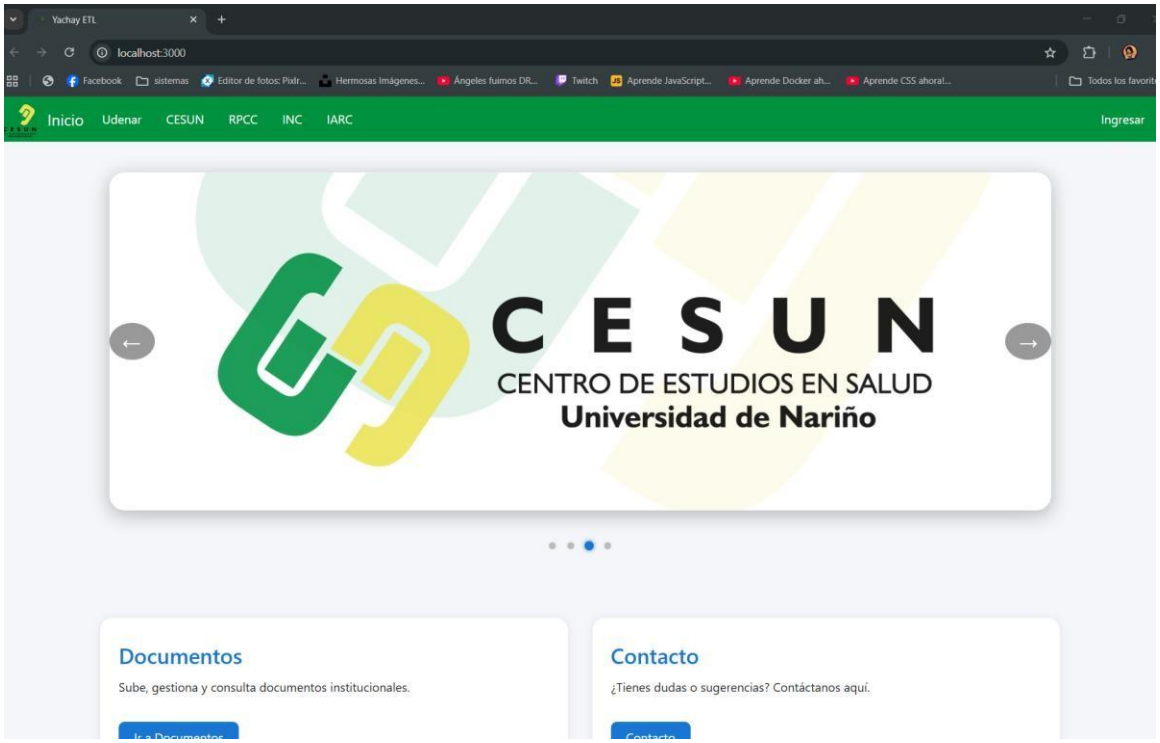
Desarrollo frontend

- A nivel de frontend tenemos una arquitectura por componentes donde cada sección de la aplicación corresponde a un componente diferente, de esta forma asegurando la mantenibilidad y facilidad de modificación en el código.



- A continuación, se presenta los componentes en forma de interfaz para su visualización.

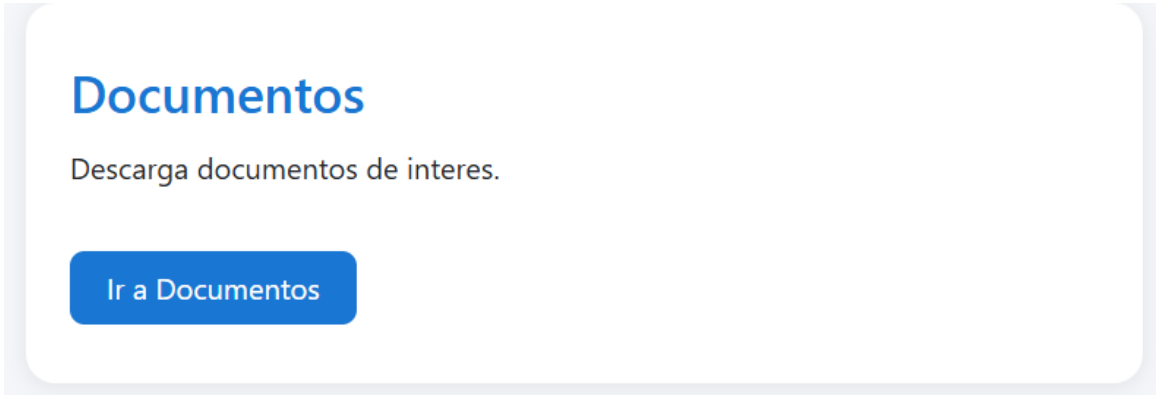
- Se tiene una interfaz inicial, la cual tiene una barra de navegación, un panel informativo con imágenes y tarjetas interactivas.



- La barra de navegación contiene enlaces o botones para interactuar con las diferentes páginas de interés que se pueda tener, también tiene una opción para ingresar a la aplicación yachay etl

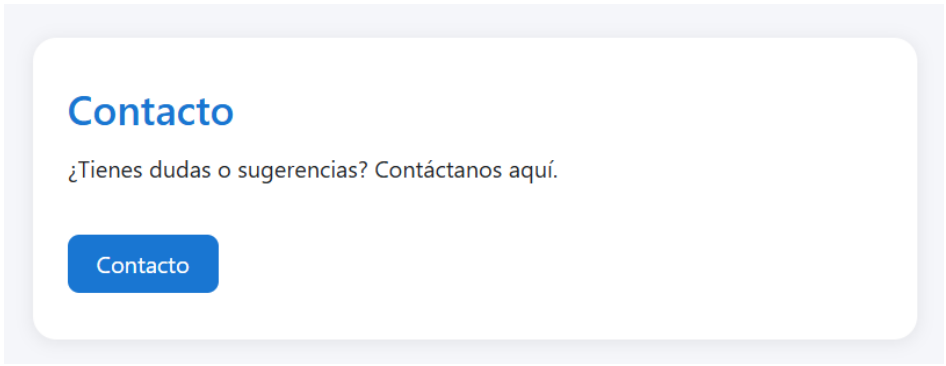


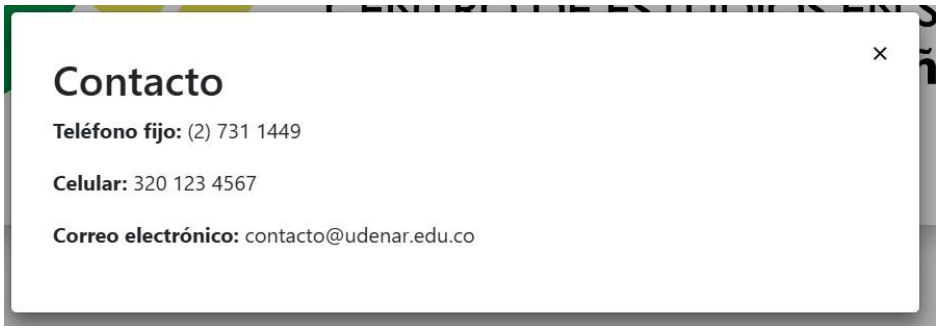
- La tarjeta de documentos, permitirá visualizar un listado de documentos disponibles y habilitará la opción para descargar estos documentos al momento de seleccionarlos.



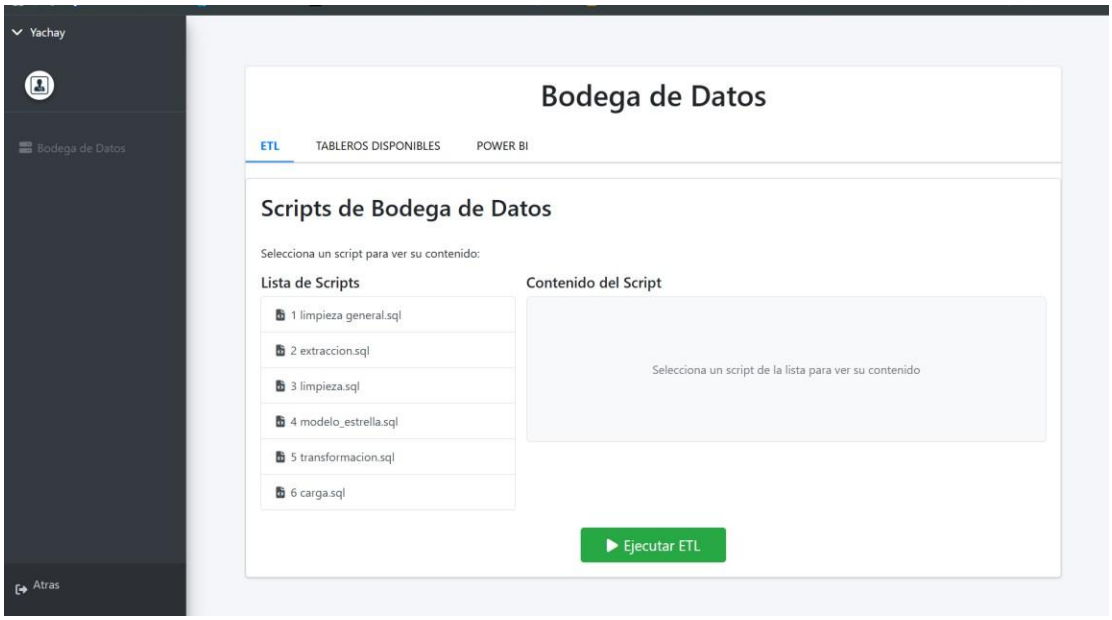


- La tarjeta de contacto abrirá una ventana con información para contactar, como correo, teléfono

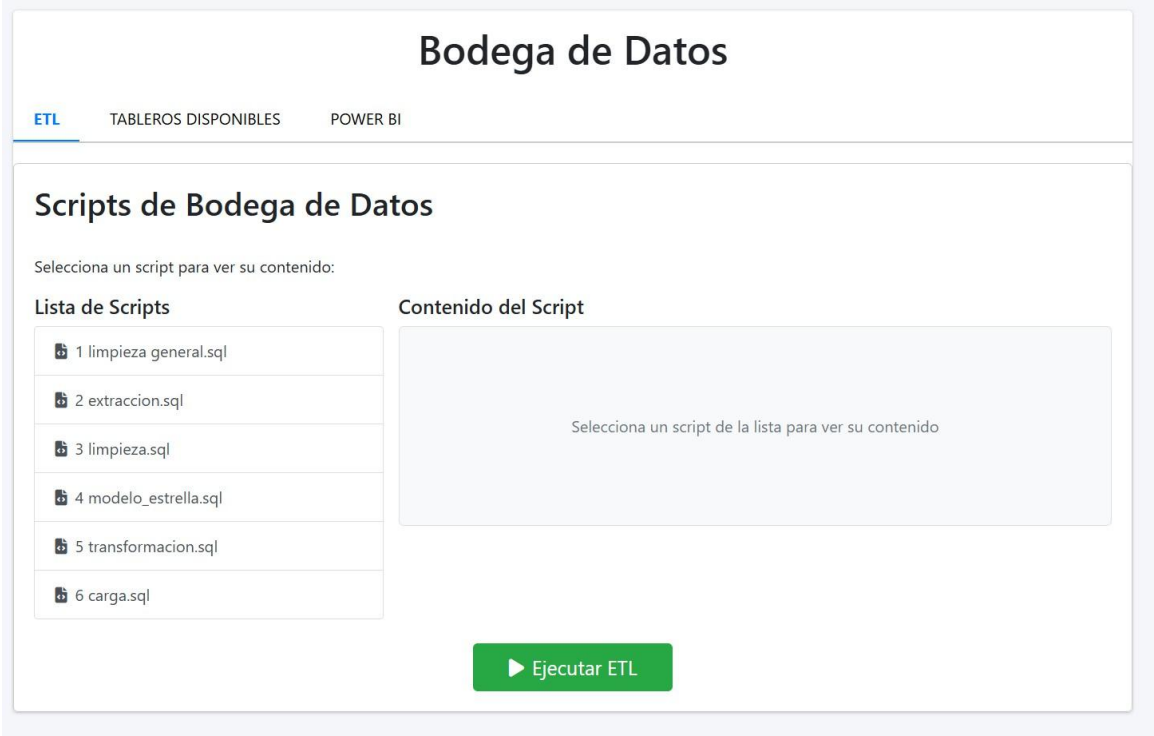




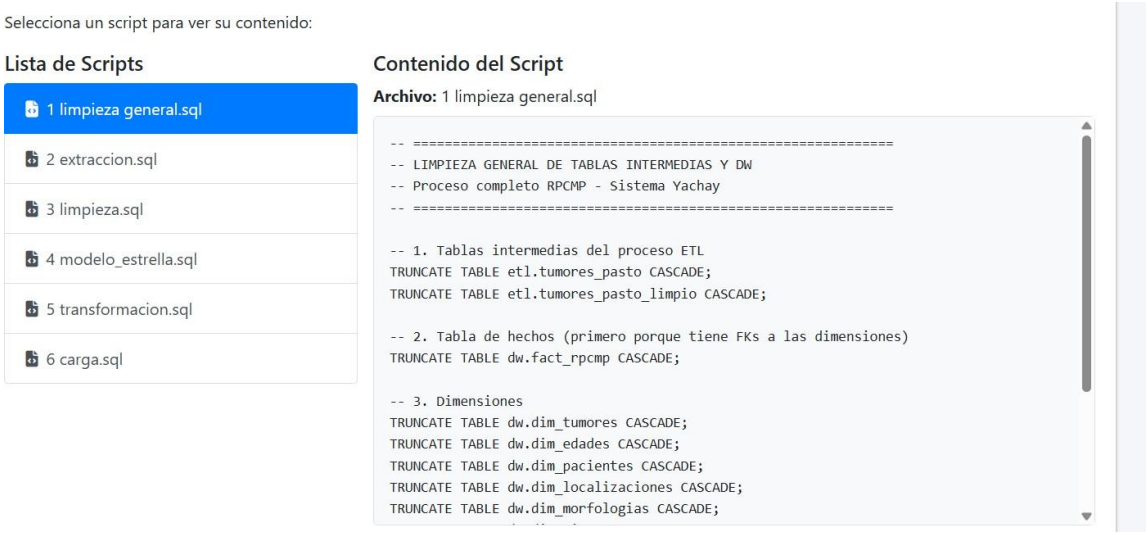
- Al ingresar a la aplicación se tiene la siguiente interfaz, la cual se compone por un panel lateral y un panel principal en el que se visualiza el contenido.



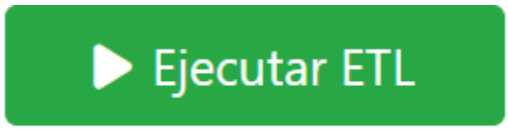
- En el panel principal se mira una ventana con 3 pestañas, la primera de ellas llamada ETL, nos muestra los scripts de la bodega de datos



- En la lista de la izquierda podemos seleccionar un script y visualizar su contenido en el panel de la derecha.

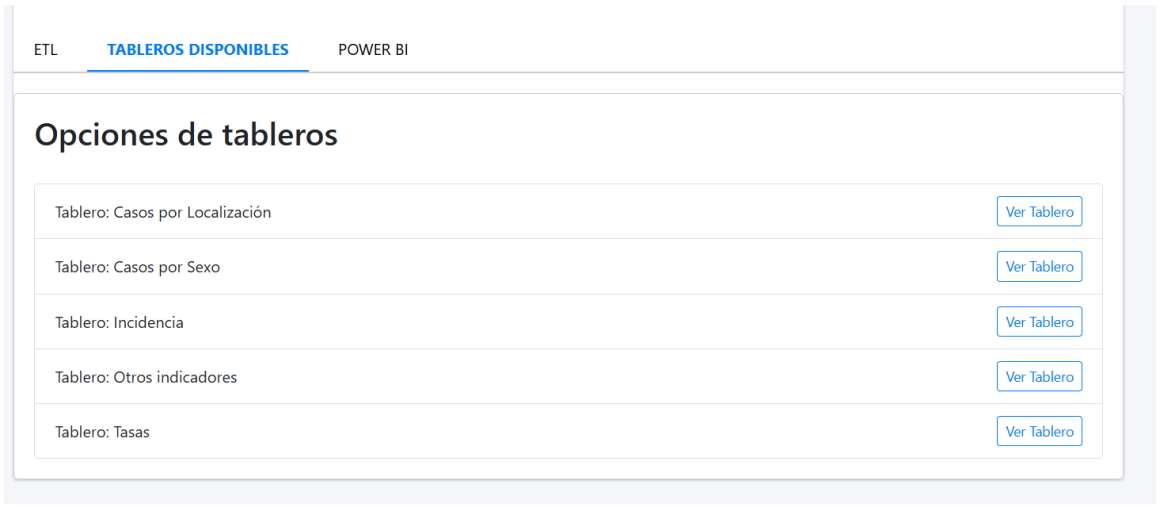


- Y se tiene una opción para ejecutar los scripts en la base de datos.

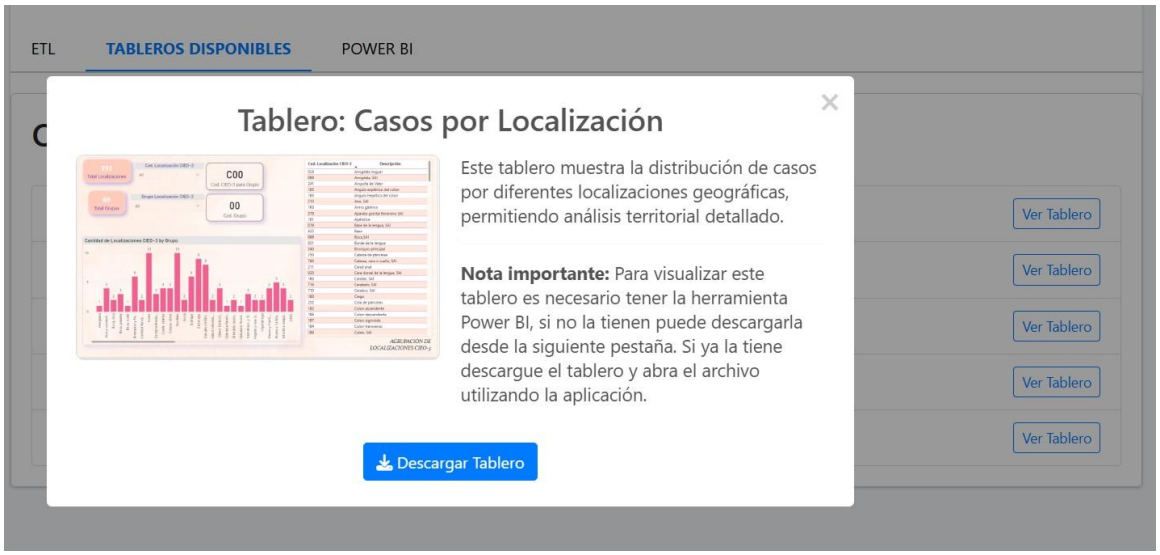




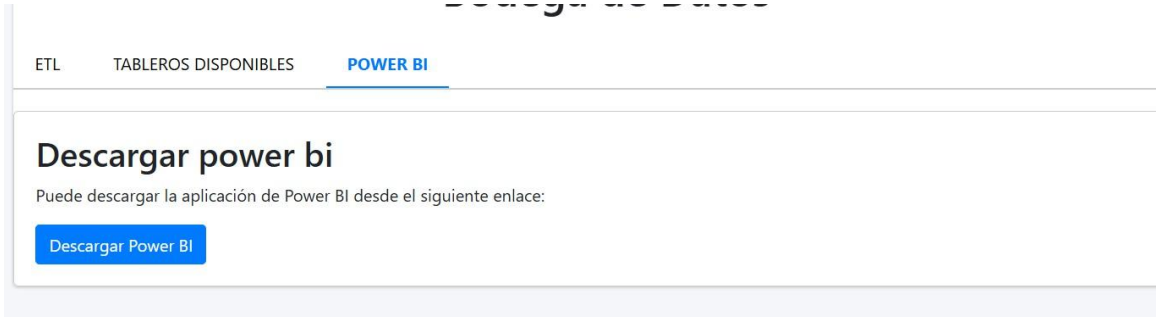
- La segunda pestaña muestra una lista de las opciones de tableros que tiene la aplicación.



- Al seleccionar una y dar clic en el botón ver tablero correspondiente se abre una ventana emergente con información del tablero y la posibilidad de descargarlo.



- La tercera pestaña tiene un botón con un enlace a la pagina oficial de power bi para poder descargar he instalar la herramienta necesaria para la visualización de los tableros.



6. Diseño de la Bodega de Datos.

Con base en los requerimientos tanto del proyecto YACHAY, como del Registro Poblacional de Cáncer del Municipio de Pasto, y en la literatura consultada se plantea el diseño de la Bodega de Datos a construir de la mano de la metodología específica a emplear en esta fase, que es HEFESTO.

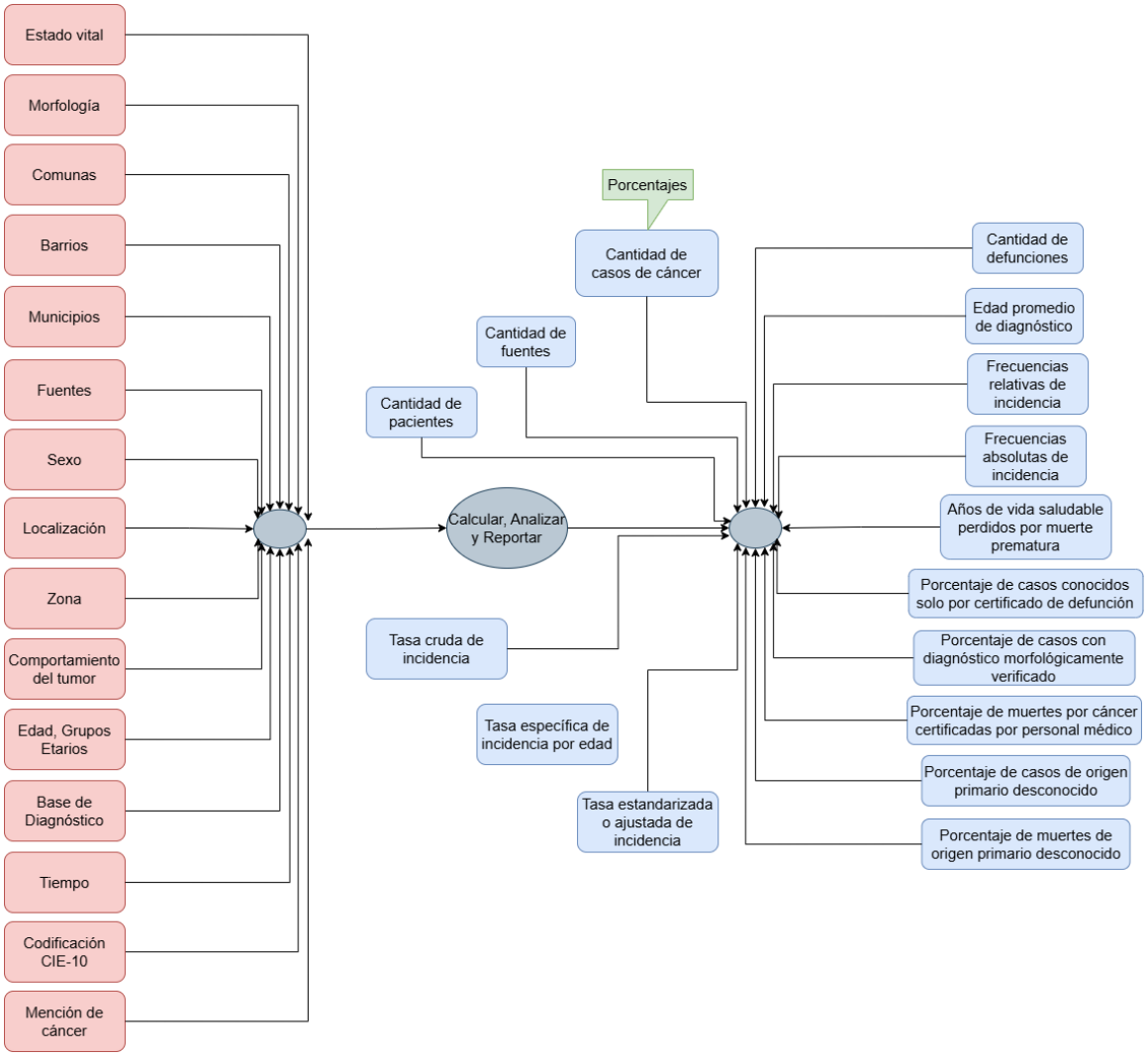
HEFESTO es una metodología que tiene inmerso la metodología Kimball para su desarrollo y que permite la construcción de Data Warehouse de forma sencilla, ordenada e intuitiva. Está fundamentada en una extensa investigación, comparación de metodologías existentes y el aporte de experiencias propias de sus implementadores. Se acopla muy bien a cualquier ciclo de vida de desarrollo de software, sus objetivos y resultados son fáciles de distinguir y comprender, utiliza modelos conceptuales y lógicos sencillos de interpretar y analizar, es secuencial puesto que el resultado de una fase se constituye en la entrada de la siguiente y sobre todo se adapta con facilidad y rapidez a los cambios y a los requerimientos de los usuarios, que es fundamental dentro del entorno cambiante de operación del RPCMP y por ende a la hora de diseñar e implementar la Bodega de Datos y su sistema ETL, planteados en este trabajo.

Se analizan las fuentes de datos para determinar cómo construir los Indicadores, señalando el mapeo correspondiente y seleccionando los campos de estudio de cada Perspectiva.

Las fuentes de información son indispensables para contar con las variables de análisis y poder obtener finalmente los Indicadores que reflejan el comportamiento de los datos desde diferentes Perspectivas.

En esta etapa de la metodología se busca analizar las fuentes de datos para determinar ¿cómo se van a calcular los Indicadores? y para hacer un mapeo entre el Modelo Conceptual creado y los datos o fuentes de datos del RPCMP. Para esto es necesario definir dos conceptos:

- Hecho o hechos: son los factores que componen la fórmula de cálculo del Indicador.
- Función de agregación: es la operación común en el análisis de datos mediante la que se agrupan los resultados para resumir datos y obtener resultados coherentes al cálculo que se está realizando del Indicador. Ejemplo: suma (SUM), promedio (AVG), conteo (COUNT), etc.



En la figura anterior se muestra un diagrama conceptual, a la izquierda perspectivas y a la derecha indicadores relacionados mediante el proceso calcular, analizar y reportar.

A continuación, se definen uno a uno los Indicadores y Perspectivas con su respectiva fórmula identificando los Hechos y la Función de agregación.

- Indicador: Cantidad de casos de cáncer (CCC)

$$count(casos) \quad (1)$$

  Hechos: casos de cáncer o tumores
  Función de agregación: conteo (count)
  Este indicador representa el conteo de los casos de cáncer o tumores que se han registrado en el RPCMP en un tiempo determinado bajo las características que se desee analizar.

- Tasa cruda de incidencia (TCI)

$$\frac{count(casos) \times 100.000}{Proyeccion\ de\ Poblacion} \quad (2)$$

Hechos: casos de cáncer o tumores, proyección de población de Pasto y una constante de 100.000.
Función de agregación: conteo (count)
Este indicador representa la tasa de incidencia presentada para una o varias localizaciones específicas y que se han registrado en el RPCMP en un tiempo determinado bajo las características que se desee analizar.

- Tasa específica de incidencia por edad ($TEEI$)

$$\frac{count(casos) \times Población\ Mundial}{Proyeccion\ de\ Poblacion} \quad (3)$$

Hechos: casos de cáncer o tumores, población mundial estándar y proyección de población de Pasto
Función de agregación: conteo (count)
Este indicador representa la tasa específica de incidencia por edad presentada para una o varias localizaciones específicas y que se han registrado en el RPCMP en un tiempo determinado bajo las características que se desee analizar.

- Tasa estandarizada o ajustada de incidencia ($TAEI$)

$$sum(TEEI) \quad (4)$$

Hechos: tasa específica de incidencia por edad
Función de agregación: suma (sum)
Este indicador representa la tasa ajustada de incidencia por edad presentada para una o varias localizaciones específicas y que se han registrado en el RPCMP en un tiempo determinado bajo las características que se desee analizar.

- Frecuencias absolutas de incidencias de casos de cáncer (f_{AI})

$$count(casos) \quad (5)$$

Hechos: casos de cáncer o tumores
Función de agregación: conteo (count)
Este indicador representa la frecuencia absoluta de incidencia para una o varias localizaciones específicas y que se han registrado en el RPCMP en un tiempo determinado. En otras palabras, establece el número de veces que un determinado caso de incidencia de cáncer, que cumple algunas condiciones, se encuentra registrado en el RPCMP. Este indicador se considera igual al CCC pero se describen por aparte porque los reportes así lo mencionan.

- Frecuencias relativas de incidencias de casos de cáncer (f_{RI})

$$\frac{f_{AI}}{Numero\ total\ de\ casos} \times 100 \quad (6)$$

Hechos: frecuencia absoluta de incidencia y número total de casos de cáncer o tumores

Función de agregación: conteo (count) proveniente de f_{AI}
Este indicador representa la frecuencia relativa de incidencia para una o varias localizaciones específicas y que se han registrado en el RPCMP en un tiempo determinado. En otras palabras, establece la proporción en que un determinado caso de incidencia de cáncer, que cumple algunas condiciones, se encuentra registrado en el RPCMP.

- Cantidad de defunciones (CD)

$$count(defunciones) \tag{7}$$

Hechos: defunciones
Función de agregación: conteo (count)
Este indicador representa el conteo de las defunciones que se han registrado en el RPCMP en un tiempo determinado bajo las características que se desee analizar.

- Cantidad de pacientes (CP)

$$count(pacientes) \tag{8}$$

Hechos: pacientes
Función de agregación: conteo (count)
Este indicador representa el conteo de los pacientes que se han registrado en el RPCMP en un tiempo determinado bajo las características que se desee analizar.

- Cantidad de fuentes (CF)

$$count(fuentes) \tag{9}$$

Hechos: fuentes
Función de agregación: conteo (count)
Este indicador representa el conteo de las fuentes con las que cuenta el RPCMP y que por ende le reportan información de pacientes con casos oncológicos en un tiempo determinado bajo las características que se desee analizar.

- Edad promedio de diagnóstico (EPD)

$$avg(edad\ de\ diagnostico) \tag{10}$$

Hechos: edades de diagnóstico
Función de agregación: promedio (avg)
Este indicador representa la edad promedio de diagnóstico para un subconjunto de casos de incidencia de cáncer o tumores que se han registrado en el RPCMP en un tiempo determinado bajo las características que se desee analizar.

- Porcentaje de casos con un diagnóstico morfológicamente verificado ($CDMV(\%)$)

$$\frac{count(casos) * 100}{Numero\ total\ de\ casos} \tag{11}$$

Hechos: casos de cáncer o tumores y número total de casos
Función de agregación: conteo (count)

Este indicador representa el porcentaje de aquellos casos que tienen un diagnóstico morfológicamente verificado, es decir que cuentan con verificación microscópica: citología - hematología e histología del tumor primario, aspirado de médula ósea y citometría de flujo y que se han registrado en el RPCMP en un tiempo determinado bajo las características que se desee analizar.

- Porcentaje de casos conocidos sólo por certificado de defunción ($CSCD$(%))

$$\frac{count(casos) * 100}{Numero\ total\ de\ casos} \quad (12)$$

Hechos: casos de cáncer o tumores y número total de casos
Función de agregación: conteo (count)
Este indicador representa el porcentaje de aquellos casos que solo se han registrado en el RPCMP mediante el certificado de defunción en un tiempo determinado y bajo las características que se desee analizar.

- Porcentaje de muertes por cáncer que fueron certificadas por personal médico ($MCCPM$(%))

$$\frac{count(defunciones) * 100}{Numero\ total\ de\ defunciones} \quad (13)$$

Hechos: defunciones y número total de defunciones
Función de agregación: conteo (count)
Este indicador representa el porcentaje de muertes por cáncer certificadas por personal médico y que se han registrado en el RPCMP en un tiempo determinado bajo las características que se desee analizar.

- Porcentaje de casos de cáncer de origen primario desconocido - localización C809 ($CCOPD$(%))

$$\frac{count(casos) * 100}{Numero\ total\ de\ casos} \quad (14)$$

Hechos: casos y número total de casos o tumores
Función de agregación: conteo (count)
Este indicador representa el porcentaje de casos de cáncer cuya localización es desconocida y que se han registrado en el RPCMP en un tiempo determinado bajo las características que se desee analizar.

- Porcentaje de muertes de cáncer de origen primario desconocido - localización C809 ($MCOPM$(%))

$$\frac{count(defunciones) * 100}{Numero\ total\ de\ defunciones} \quad (15)$$

Hechos: defunciones y número total de defunciones
Función de agregación: conteo (count)
Este indicador representa el porcentaje de muertes cuya localización es desconocida y que se han registrado en el RPCMP en un tiempo determinado bajo las características que se desee analizar.

- Años de vida saludable perdidos por muerte prematura ($AVMP$)

$$sum\ (Esperanza\ de\ vida - Edad\ de\ muerte) \quad (16)$$

Hechos: esperanza de vida y edad de muerte
Función de agregación: suma (sum)
Este indicador representa una aproximación de los años que personas que padecen la enfermedad pierden debido a que probablemente mueren antes de lo que se esperaría o de lo que se proyectaría y que sus casos se han registrado en el RPCMP en un tiempo determinado bajo las características que se desee analizar.
A continuación, es necesario validar las fuentes de datos y verificar que se posible extraer de ellas todos los datos requeridos para el cálculo de los Indicadores definidos anteriormente. Por esta razón se realiza un mapeo que indica cómo serán obtenidos los elementos definidos en el Modelo Conceptual estableciendo una relación directa entre este y las Fuentes de Datos.
En este sentido y complementando lo expuesto en la sección anterior de Materiales, se consideran fuentes de datos las siguientes:

- Entidades de salud (hospitales, laboratorios, etc.)
- Base de datos *yachay* del sistema transaccional que se está construyendo para el RPCMP en el marco del macroproyecto YACHAY.
- La base de datos *yachay* se alimenta de los casos de cáncer registrados hasta la fecha dentro del sistema CanReg empleado para tal fin por el RPCMP, así como también de exámenes de laboratorios entregados por las entidades que reportan al RPCMP y por la información de egresos también compartida por diferentes entidades de salud.
- DANE con sus proyecciones de población para el municipio de Pasto.
- Tablas adicionales como la tabla de esperanza de vida, de agrupaciones de codificación CIE-10 y la tabla de población mundial estándar.

El mapeo entre el Modelo Conceptual y las fuentes de datos es:

- La Perspectiva barrios se relaciona con la tabla tab_direcciontumor (diccionario: tab_barrios).
- La Perspectiva comunas se relaciona con la tabla tab_barrios.
- La Perspectiva municipios se relaciona con la tabla tab_direcciontumor (diccionario: tab_municipios).
- La Perspectiva sexo se relaciona con el campo cod_sexo (diccionario: tab_sexo) de la tabla tab_pacientes.
- La Perspectiva edad, grupos etarios (quinquenales de edad) se relaciona con el campo edad_paciente de la tabla tab_fichas.
- La Perspectiva localización se relaciona con la tabla tab_fichas (diccionario: tab_localizacion).
- La Perspectiva morfología se relaciona con la tabla tab_fichas (diccionario: tab_morfologia).
- La Perspectiva tiempo se relaciona con la fecha de diagnóstico (fecha de incidencia) por lo tanto con el campo fe_diagnostico de la tabla tab_fichas.

- La Perspectiva comportamiento del tumor (tipo de tumor) se relaciona con el campo cod_comportamiento (diccionario: tab_comportamiento) de la tabla tab_fichas.
- La Perspectiva base de diagnóstico (método de diagnóstico) se relaciona con el campo cod_metododiagnostico (diccionario: tab_metododiagnostico) de la tabla tab_fichas.
- La Perspectiva mención de cáncer o mención de cáncer defunción se relaciona con el campo cod_mensioncancer (diccionario: tab_mencioncancerdefuncion) de la tabla tab_certidefuncion.
- La Perspectiva estado vital se relaciona con el campo cod_estadovital (diccionario: tab_estadovital) de la tabla tab_pacientes.
- La Perspectiva fuentes se relaciona con la tabla tab_historias (diccionario: tab_fuentes).
- La Perspectiva codificación CIE-10 se relaciona con la tabla tab_fichas (diccionario: tab_cie10).
- La Perspectiva zona de residencia se relaciona con el campo cod_zona de la tabla tab_barrios (diccionario: tab_zona).

En cuanto a los Indicadores se menciona lo siguiente:

- Los Indicadores que incluyen el *count(casos)* se relacionan con el campo id_ficha de la tabla tab_fichas que identifica un tumor o caso específico, quedando la fórmula de cálculo, así:

$$count(id\_ficha) \tag{17}$$

- El indicador TCI se relaciona con el campo id_ficha de la tabla tab_fichas y el valor de proyección de población de Pasto (vr_pobpasto) que depende de la zona de residencia, si es zona urbana, la proyección de población se relaciona con el campo vr_cab_muni de la tabla tab_dane_pobpasto_base mientras que si la zona es rural, la proyección de población se relaciona con el campo vr_cent_pob_rural de la tabla tab_dane_pobpasto_base y si se calcula el indicador para el total de la población independientemente de la zona de residencia, entonces la proyección de población se relaciona con el campo vr_total de la tabla tab_dane_pobpasto_base, quedando la fórmula de cálculo, así:

$$\frac{count(id\_ficha) \times 100.000}{vr\_pobpasto} \tag{18}$$

- Para la TEEI además de tener en cuenta lo anterior también hay relación con el campo poblacion de la tabla tab_pobmundialestandar, así:

$$\frac{count(id\_ficha) \times poblacion}{vr\_pobpasto} \tag{19}$$

- El Indicador TAEI se relaciona con el Indicador TEEI.
- El Indicador f_{RI} se relaciona con el Indicador f_{AI}.
- El Indicador CD se relaciona con los campos id_ficha y cod_estadovital de la tabla tab_pacientes (diccionario: tab_estadovital) y el campo

cod_mencioncancer de la tabla tab_certidefuncion (diccionario: tab_mencioncancerdefuncion), así:

$$count(id\_ficha\ donde\ cod\_estadovital = 2\ y\ cod\_mencioncancer = 1) \quad (20)$$

- El Indicador CP se relaciona con el campo id_paciente de la tabla tab_pacientes, así:

$$count(id\_paciente) \quad (21)$$

- El Indicador CF se relaciona con el campo fuente de la tabla tab_historias, así:

$$count(fuente) \quad (22)$$

- El Indicador EPD se relaciona con el campo edad_paciente de la tabla tab_fichas, así:

$$avg(edad\_paciente) \quad (23)$$

- El Indicador CDMV se relaciona con los campos id_ficha y cod_metododiagnostico de la tabla tab_fichas, quedando la fórmula así:

$$\frac{count(id\_ficha\ donde\ cod\_metododiagnostico\ = 5\ o\ 7) * 100}{count(id\_ficha)} \quad (24)$$

- El Indicador CSCD se relaciona con los campos id_ficha y cod_metododiagnostico de la tabla tab_fichas, quedando la fórmula así:

$$\frac{count(id\_ficha\ donde\ cod\_metododiagnostico\ = 0) * 100}{count(id\_ficha)} \quad (25)$$

- El Indicador MCCPM se relaciona con el campo id_ficha de la tabla tab_fichas, el campo cod_estadovital de la tabla tab_pacientes (diccionario: tab_estadovital) y los campos cod_mencioncancer y cod_profesionexpedidor de la tabla tab_certidefuncion (diccionarios: tab_mencioncancerdefuncion y tab_profesionexpedidor), así:

$$\frac{count(id\_ficha\ donde\ cod\_estadovital = 2, cod\_mencioncancer = 1\ y\ c}{count(id\_ficha\ donde\ cod\_estadovital = 2\ y\ cod\_} \quad (26)$$

- El Indicador CCOPD se relaciona con el campo id_ficha de la tabla tab_pacientes y el campo cod_localizacion de la tabla tab_fichas (diccionario: tab_localizacion), quedando la fórmula así:

$$\frac{count(id\_ficha\ donde\ cod\_localizacion\ = 809) * 100}{count(id\_ficha)} \quad (27)$$

- El Indicador MCOPD se relaciona con el campo id_ficha de la tabla tab_fichas, los campos cod_estadovital y cod_localizacion de la tabla tab_pacientes (diccionarios: tab_estadovital y tab_localizacion) y el campo cod_mencioncancer de la tabla tab_certidefuncion (diccionarios: tab_mencioncancerdefuncion), así:

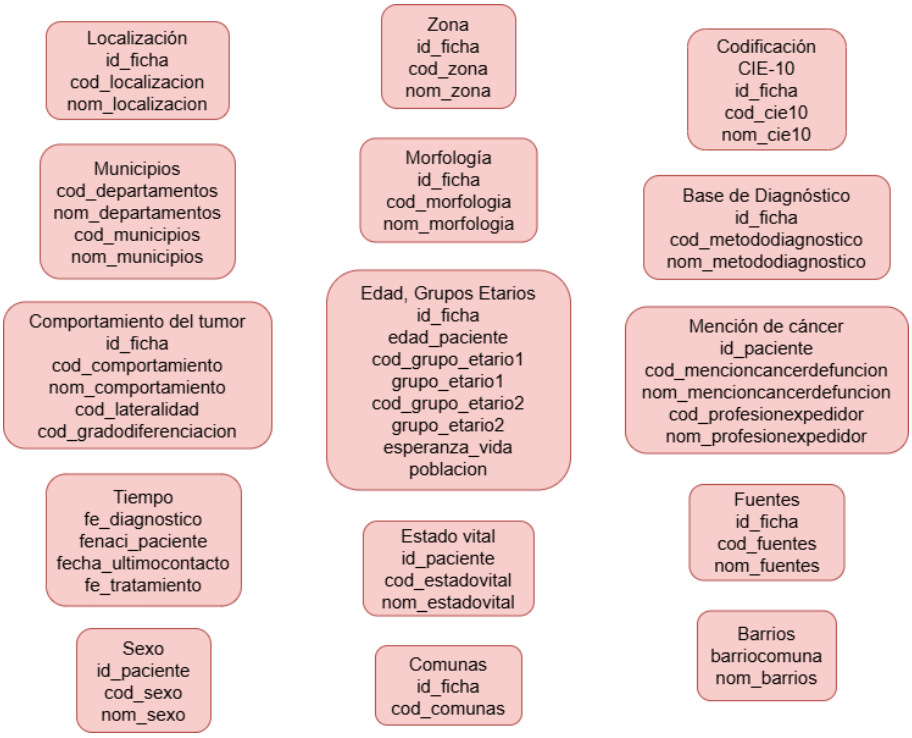
$$\frac{count(id\_ficha\ donde\ cod\_estadovital = 2, cod\_mencioncancer = 1\ y\ c}{count(id\_ficha\ donde\ cod\_estadovital = 2, cod\_menci} \quad (2$$

8
)

- El indicador AVMP se relaciona con el campo esperanza_vida de la tabla tab_aux_esperanzadevida, el campo fenaci_paciente de la tabla tab_pacientes, el campo edad_paciente de la tabla tab_fichas y el campo fe_muerte de la tabla tab_certidefuncion, sin embargo, el valor de edad de muerte (edad_muerte) depende de las fechas anteriores, quedando la fórmula así:

$$sum(esperanza\_vida + edad\_paciente - edad\_muerte) \quad (29)$$

A continuación, es necesario definir la Granularidad en la Bodega de Datos seleccionando los campos de las fuentes de datos que integran cada Perspectiva puesto que a través de ellos se analizan los Indicadores planteados. La granularidad es indispensable para conocer minuciosamente la documentación de las fuentes de datos. Los campos identificados para cada Perspectiva se relacionan en la siguiente figura.



a) Modelo lógico del Data Warehouse

Se construye el *Modelo Lógico del Data Warehouse* bajo el esquema en Estrella definido en el presente trabajo puesto que es una de las arquitecturas más utilizadas en la construcción de Bodegas de Datos porque ofrece entre otras cosas

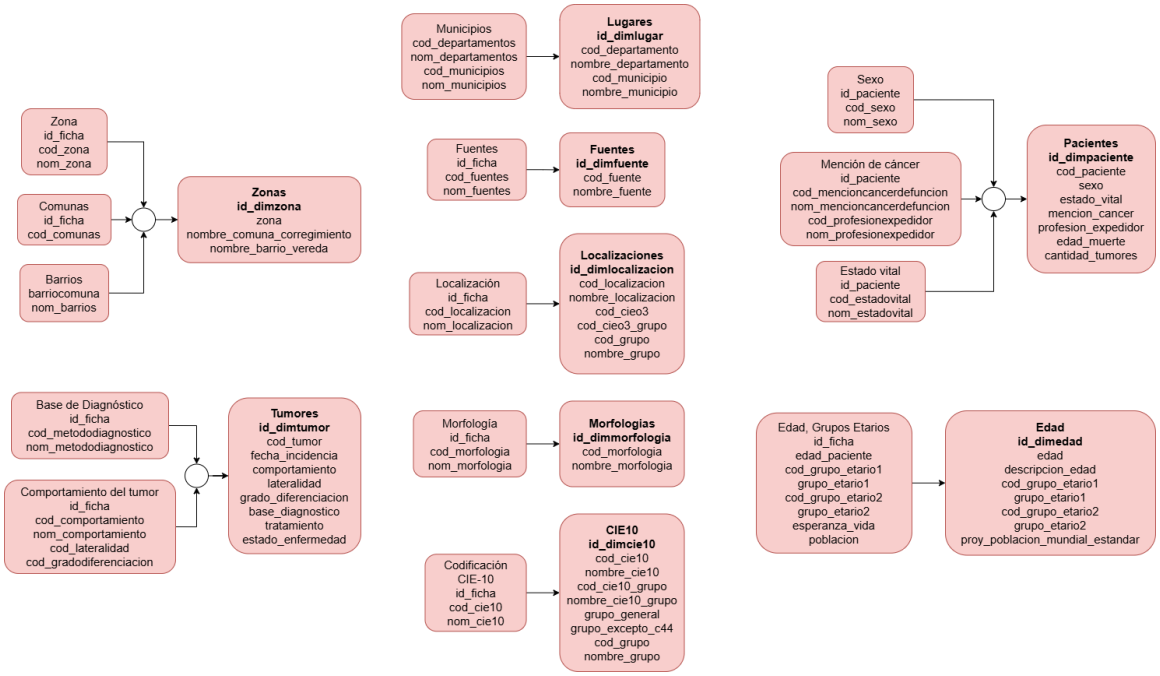
rendimiento, simplicidad y facilidad de consulta [35] acoplándose al objeto de este trabajo. Seguidamente, se confeccionan las tablas de *Dimensiones* y la tabla de *Hechos*, para luego efectuar sus respectivas uniones.

Un Modelo Lógico es la representación de una estructura de datos que puede procesarse y almacenarse en un SGBD. En cuanto a la tipología, para este trabajo se ha seleccionado el Modelo Estrella, que es el modelo base para la implementación de otros modelos, es el más simple de interpretar, mantener y actualizar y su principal ventaja radica en que posee los mejores tiempos de respuesta y facilita la formulación de las sentencias SQL, debido a la desnormalización requerida en un modelo multidimensional, para evitar uniones (JOIN) entre las tablas cuando se realizan consultas.

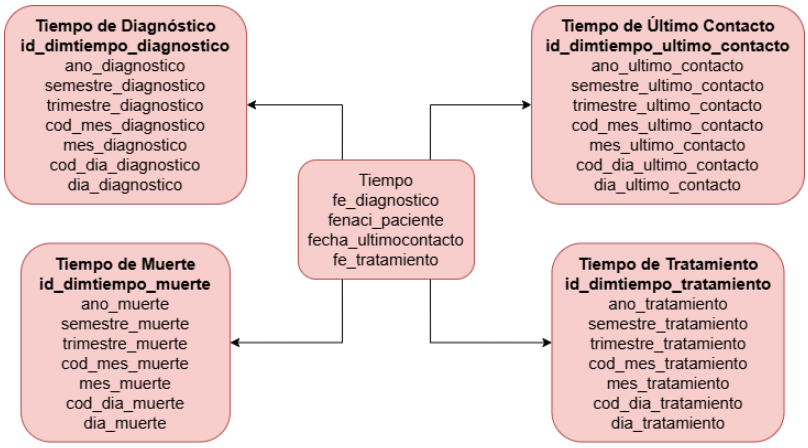
Posteriormente y partiendo del resultado del paso anterior (Modelo Conceptual), se reajustan los campos de las Perspectivas definidas para construir las Tablas de Dimensiones de la Bodega de Datos puesto que, en el Modelo Estrella, como en otros modelos, cada Perspectiva se constituye en una tabla de Dimensión de la Bodega de Datos. Para esto, se realiza el siguiente proceso:

- Se elige un nombre que identifique la tabla de Dimensión.
- Se adiciona un campo que represente su clave principal.
- Se redefinen, si es necesario, los campos y sus nombres para que sean más intuitivos.

Siguiendo el proceso se han construido las tablas de Dimensiones con su respectivo nombre y su clave principal como se evidencia en la siguiente figura:



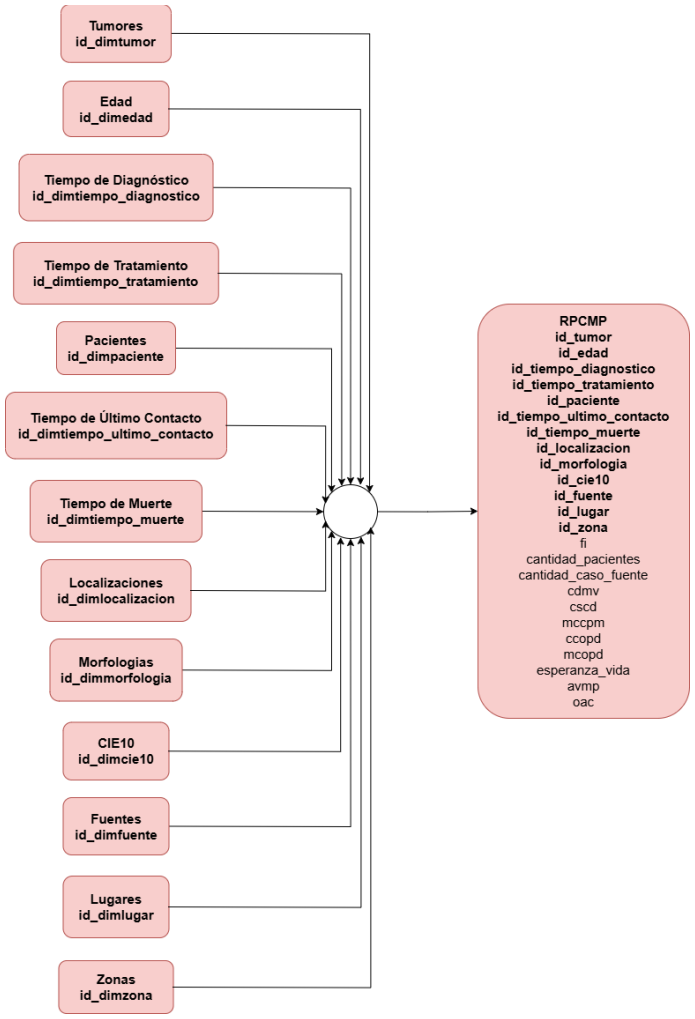
En cuanto a la Perspectiva Tiempo, luego del análisis se encuentra que los Indicadores pueden ser analizados desde una Perspectiva de Tiempo de Diagnóstico, pero también desde una Perspectiva de Tiempo de Muerte que estaría asociada al paciente y no al tumor, además desde una Perspectiva de Tratamiento y de Último Contacto para aquellos casos en los que se tiene una fecha registrada en esta información, por lo que se encontró necesario especificar la Perspectiva Tiempo en cuatro Dimensiones específicas como se aprecia en la siguiente figura:



Enseguida se construye la tabla de Hechos del Modelo en Estrella, para esto se realiza el siguiente proceso:

- Se asigna un nombre a la tabla de Hechos que represente la información contenida en ella.
- Se define su clave primaria, que se compone de la combinación de las claves primarias de cada tabla de Dimensión relacionada.
- Se crean tantos campos de Hechos como Indicadores se hayan definido en el Modelo Conceptual y se les asigna un nombre que los identifique.

En ese orden de ideas, la tabla de Hechos construida se evidencia en la figura siguiente:



b) Integración de datos (Proceso ETL)

Finalmente, utilizando técnicas de limpieza y calidad de datos, se implementa un proceso ETL y se define la política o estrategia para la Carga Inicial del Data Warehouse y su respectiva Actualización.

Para la construcción del ETL se empleó la herramienta pgAdmin4 y se construyeron las diferentes sentencias SQL que permiten la extracción, transformación y carga de los datos a la Bodega de Datos. Este proceso consiste específicamente en extraer de las fuentes de datos todos los campos que hacen parte del Modelo Lógico definido en la fase anterior y consolidarlos en lo que se conoce como un área de stage o staging area que es un área de almacenamiento temporal que puede estar en diversos formatos, como bases de datos relacionales, XML, NoSQL y archivos planos , en este caso se encuentra en la misma base de datos *yachay* en un esquema independiente llamado *etl* y una vez

extraídos, acorde al análisis se limpian y transforman en la misma área para posteriormente ser cargados a la Bodega de Datos.

❖ *Extracción y análisis de datos.*

La Extracción es el primer paso del proceso ETL, en este caso se elaboró una consulta SQL (ver anexo 1) que relaciona las tablas requeridas de la base de datos *yachay* y las consolida en el staging area con el nombre de *etl.tumores_pasto*. Posteriormente, con el fin de obtener un conjunto de datos consistente, organizado y listo para usar en Power BI como herramienta de visualización, se realizó un análisis de la calidad de los datos correspondientes a cada uno de los campos que hacen parte del Modelo Lógico definido en la fase anterior. Esto con el fin de conocer entre otras cosas, su comportamiento y por cada atributo el número de valores distintos, el número de valores nulos y no nulos y lo que representa el atributo. Para esto se empleó una tarea automática construida con la librería pandas de Python (ver anexo 2) que arroja los resultados expuestos.

❖ *Limpieza de datos.*

El segundo paso del ETL es la Limpieza de los datos donde partiendo del análisis anterior se ejecutan diferentes procedimientos por medio de sentencias SQL con el propósito especialmente de reemplazar valores nulos y corregir valores erróneos. A continuación, se nombran algunos de estos procedimientos:

- Creación de la tabla *etl.tumores_pasto_limpio* a partir de la tabla *etl.tumores_pasto*.
- Se eliminan espacios en blanco o de relleno de aquellos atributos que los contienen como los atributos *sexo*, *cod_localización*, *cod_cie10*, entre otros.
- Se estandariza algunos atributos a mayúsculas, especialmente aquellos cuyos valores son de extensión corta en caracteres, por ejemplo, los atributos *sexo*, *estado_vital*, *lateralidad* y *comportamiento*.
- Con el fin de corregir valores erróneos o anómalos y eliminar valores nulos, se actualizan algunos valores de algunos atributos como se evidencia en la **¡Error! No se encuentra el origen de la referencia.**.
- Se encontró que el atributo *cod_cie10* tiene algunos valores a 3 caracteres y debido a la extensión solicitada por CanReg, el RPCMP optó por complementar los 4 caracteres con una “X” al final del código por lo que en este proceso de Limpieza se elimina la “X” al final de los códigos, en los casos en que aplique. Ejemplo: *cod_cie10 = C73X*, el código CIE-10 en realidad es C73 por lo que se elimina la “X” del final para conservar el código acorde al valor real.
- En cuanto a los atributos extraídos para la codificación CIE-10 se encuentra el atributo *cod_cie10_grupo* (contiene los códigos CIE-10 para clasificación por la agrupación CI5 XII) donde debido a las agrupaciones indicadas por el personal del RPCMP y a que para el RPCMP es necesario excluir algunos casos (neoplasias malignas de la piel) de conteos y reportes, se brinda la posibilidad de por medio de este atributo filtrar los conteos en los reportes de casos. Para esto se ha asignado un nuevo valor para el atributo, “EXC” para

todos aquellos casos asociados a *cod_cie10 = C44* siempre y cuando los códigos de morfología (atributo *cod_morfologia*) estén entre 8070 y 8078 y 8090 y 8098, de lo contrario el valor C44 de *cod_cie10_grupo* se conserva. De igual forma y en concordancia con lo anterior, los atributos *cie10_grupo_general* y *grupo_cie10* actualizan su valor a "*Otras neoplasias malignas de la piel que agrupan los carcinomas de células basales y escamosas de la piel*" y el atributo *cod_grupo_cie10* adiciona un nuevo *id* de grupo cuando el atributo *cod_cie10_grupo* es igual a "EXC".

- Los atributos *cod_lateralidad* y *lateralidad* sufren un cambio teniendo en cuenta que la lateralidad no aplica para todos los casos de cáncer puesto que algunas localizaciones no tienen lateralidad, simplemente no corresponde. Por esto y con base en la sección Laterality expuesta en [37], como la indicación dada por el RPCMP a la hora de identificar las codificaciones CIE-10 a las que se les debería reconocer una lateralidad asignada, se reemplazan los valores nulos de los atributos *cod_lateralidad* y *lateralidad* por *4* y *No corresponde*, respectivamente, teniendo en cuenta los códigos CIE-10 para los que aplique y para aquellos códigos CIE-10 que se deba reconocer una lateralidad, se asigna los valores *9* y *Desconocido*, respectivamente indicando que aunque la lateralidad aplica, no se conoce específicamente la que corresponde a cada uno de los tumores en cuestión.
- Para las variables mención de cáncer (atributos *cod_mencion_cancer* y *mencion_cancer*) y profesión de quien expide el certificado de defunción, se crea un nuevo código con el valor de "0" y la descripción "NO APLICA", cuando el atributo *estado_vital* del tumor sea "VIVO". Además, cuando el atributo *estado_vital* es "MUERTO", se actualiza el valor nulo de las variables a "MUERE POR CAUSAS DESCONOCIDAS" y "PROFESIÓN DEL EXPEDIDOR DESCONOCIDA", respectivamente, puesto que se sabe que está muerto, pero se desconoce las causas y la profesión de quien expide el certificado de defunción, y si el *estado_vital* es "DESCONOCIDO" o "PERDIDO EN EL SEGUIMIENTO", los atributos *cod_mencion_cancer* y *mencion_cancer* de la variable mención de cáncer toman los valores de *9* y *DESCONOCIDO*, respectivamente y los atributos *cod_profesion_expedidor* y *profesion_expedidor*, de igual forma.

❖ *Transformación de datos.*

A continuación, el tercer y último proceso del ETL es la Transformación de los datos cuyo objetivo es que la staging area esté limpia y transformada con datos de calidad como la materia prima de carga para la Bodega de Datos. Algunos de los procedimientos ejecutados en este proceso son:

- Creación de la tabla *etl.tumores_pasto_transformado* a partir de la tabla *etl.tumores_pasto_limpio*.
- Se actualiza el tipo de dato a *char(1)* para algunos atributos que son códigos de un solo caracter como es el caso de *cod_sexo*, *cod_estado_vital* y *cod_base_diagnostico*, entre otros.
- Se actualiza el tipo de datos a *text* para algunos atributos que son descriptivos como es el caso de *sexo*, *estado_vital* y *base_diagnostico*, entre otros.

- El atributo *edad_paciente* se discretiza y asocia a un grupo etario en las dos definiciones, grupo_etario1 y grupo_etario2, teniendo en cuenta lo expuesto líneas arriba sobre grupos etarios. Por esto se crean los atributos *cod_grupo_etario1* y *cod_grupo_etario2* de tipo *smallint* y *grupo_etario1* y *grupo_etario2* de tipo *text*.
- Se crean nuevos atributos (*cod_dia_muerte, dia_muerte, cod_mes_muerte, mes_muerte, ano_muerte*) para diferenciar la fecha de muerte y se actualizan con la separación respectiva en día, mes y año de los valores del atributo *fecha_ultimo_contacto* en los casos donde el estado vital es “MUERTO”.
- Se modifica el tipo de dato a *text* para el atributo *fecha_diagnostico* (fecha de incidencia) con el fin de separar la fecha en campos independientes para el día, el mes y el año de diagnóstico, por eso se crean los atributos *dia_diagnostico*, *mes_diagnostico* y *ano_diagnostico* de tipo *smallint*. Estos atributos pueden contener el valor “DESCONOCIDO” cuando el día sea igual a 99 y/o el mes igual a 99, respectivamente. Por lo anterior estos atributos son usados para los nombres respectivos en los ejes de gráficos de visualización, pero para su orden tanto numéricamente como el orden correcto de los meses (ENERO, FEBRERO, MARZO, etc.) se crean los atributos *cod_dia_diagnostico* y *cod_mes_diagnostico*, donde el valor “DESCONOCIDO” corresponde con el código 32 y 13, respectivamente, para que aparezca siempre al final de la lista. Un ejemplo de esto se observa en la **¡Error! No se encuentra el origen de la referencia.**. A su vez, el atributo *mes_diagnostico* se discretiza en dos nuevos atributos, *trimestre_diagnostico* y *semestre_diagnostico*, El procedimiento descrito en este punto aplica también para las fechas de último contacto, de muerte y de tratamiento, aunque en la fecha de muerte se tiene en cuenta los valores de 0 para los atributos *cod_dia_muerte* y *cod_mes_muerte* y “NO APLICA” para los atributos *dia_muerte*, *mes_muerte* y *ano_muerte* cuando el estado vital es “VIVO” y el valor 33 para el atributo *cod_dia_muerte*, el valor 14 para el atributo *cod_mes_muerte* y el valor “SIN INFORMACIÓN” para los atributos *dia_muerte*, *mes_muerte* y *ano_muerte* para otros estados vitales (“PERDIDO EN EL SEGUIMIENTO” y “DESCONOCIDO”).
- Los valores del atributo *localizacion* inician con el código de localización en CIEO-3 por lo que se omite su parte inicial puesto que esta información se encuentra en el atributo *cod_localizacion* para que el atributo *localizacion* contenga solo la descripción del respectivo *cod_localizacion*.
- Se crean los atributos *barrio_vereda* y *comuna_corregimiento* de tipo *text*, con el fin de incluir los valores del barrio o la vereda y de la comuna o del corregimiento, dependiendo de la zona, “URBANA” o “RURAL”, respectivamente, a partir de los atributos *barrio_comuna* y *cod_comuna_corregimiento*. Posteriormente se actualizan los valores del atributo *barrio_vereda* de “Rural desconocido” a “Vereda desconocida” y de “Urbano desconocido” a “Barrio desconocido”. También se actualizan los valores del atributo *comuna_corregimiento* de “99” a “Comuna desconocida”, de “Desconocido” donde la zona es “RURAL” a “Corregimiento desconocido”, de “Desconocido” donde la zona es “DESCONOCIDA” a “Comuna o

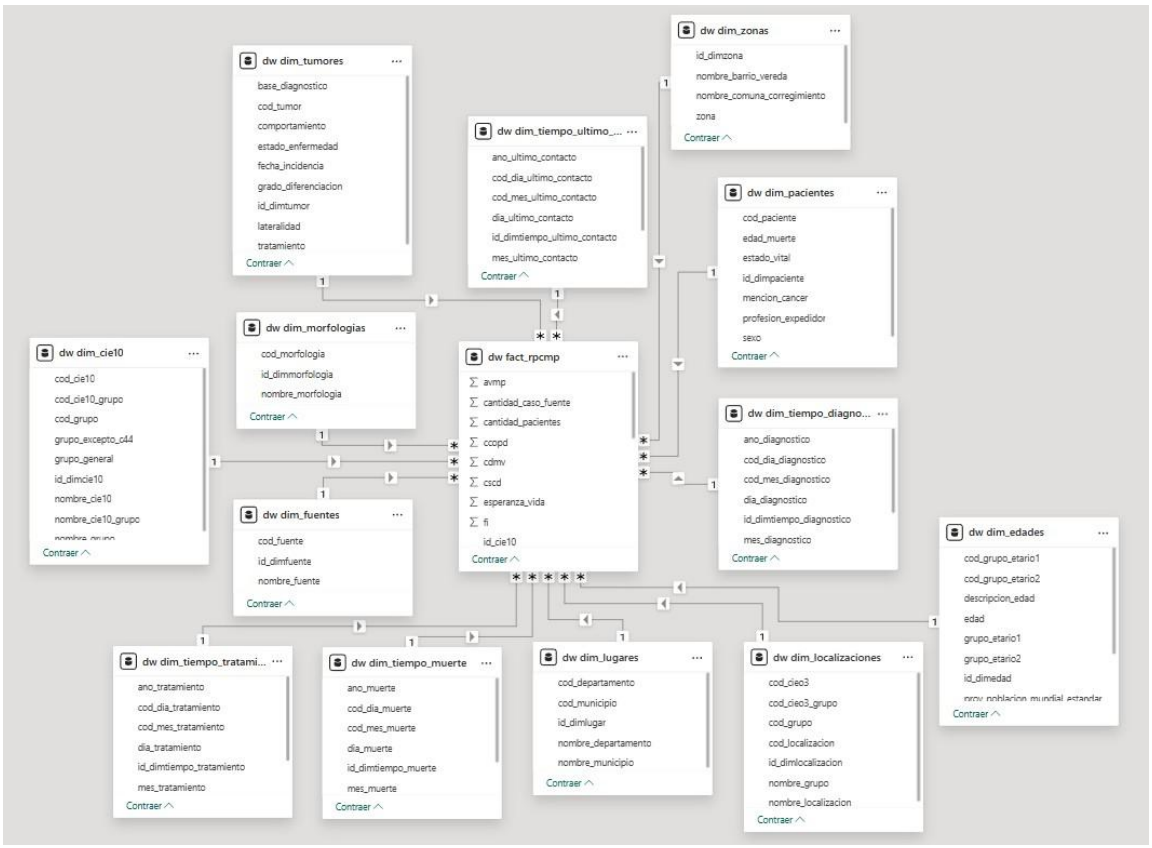
Corregimiento desconocido", los valores con números de comuna, se les concatena la palabra "Comuna" antes del número de la comuna y los valores con nombres de corregimientos se les concatena la palabra "Corregimiento" antes del nombre del corregimiento.

- *Carga de datos.*

Finalmente, el proceso de Carga de los datos consiste en implementar el Modelo en Estrella creado en la fase anterior y seleccionar los atributos que alimentan cada una de las Dimensiones y la tabla de Hechos para cargarlos a las tablas correspondientes en el esquema de la Bodega de Datos implementado. Para esto en primer lugar se ejecuta el script que se evidencia en el 3 donde se crean las 13 Dimensiones y la tabla de Hechos en un esquema independiente de nombre *dw* y posteriormente se ejecuta el script que se evidencia en el 4 que distingue de la staging area, los valores de los atributos que conforman cada una de las Dimensiones y los inserta como nuevos registros en cada una de las tablas, luego para la Carga de la tabla de Hechos, se alimenta de los *id* de cada una de las Dimensiones para construir su llave primaria y también distingue de la staging area, los valores de los atributos que la conforman.

Teniendo en cuenta todo el proceso descrito en esta fase, HEFESTO está fundamentada en una extensa investigación, comparación de metodologías existentes y el aporte de experiencias propias de sus implementadores. Además, se acopla muy bien a cualquier ciclo de vida de desarrollo de software, sus objetivos y resultados son fáciles de distinguir y comprender, utiliza modelos conceptuales y lógicos sencillos de interpretar y analizar, es secuencial puesto que el resultado de una fase se constituye en la entrada de la siguiente y sobre todo se adapta con facilidad y rapidez a los cambios y a los requerimientos de los usuarios, que es fundamental dentro del entorno cambiante de operación del RPCMP y por ende a la hora de diseñar e implementar la Bodega de Datos.

El diseño multidimensional de la bodega de datos con topologia en Estrella se muestra en la siguiente figura:



c) Fase de visualización

En esta fase se pretende mostrar a grandes rasgos el conocimiento que se puede generar a partir del análisis de la información contenida en la Bodega de Datos y por medio de la integración con la herramienta de Inteligencia de Negocios Power BI, en su suite Desktop, desarrollada por Microsoft puesto que la implementación de Power BI no solo mejora la capacidad de análisis y la toma de decisiones, en este caso frente a los datos de cáncer de este trabajo, sino que también puede proporcionar una cantidad de resultados en poco tiempo debido a la facilidad de uso de sus herramientas tanto para consultar la información como para crear reportes y visualizar el comportamiento de los datos.

En esta fase se construyeron diferentes informes con temas específicos y se hizo uso de las herramientas que ofrece Power BI para que los gráficos sean comprensibles y permitan obtener información de calidad rápidamente. En este sentido, entre otras cosas, es importante resaltar el uso del lenguaje DAX (Expresiones de Análisis de Datos) de Power BI que permite el uso de diferentes fórmulas para trabajar con datos tabulares y hacer consultas sobre esos datos estructurados.

Cabe resaltar que en la aplicación YACHAY-ETL, todos los dashboards que se implementaron en Power Bi hacen referencia a los casos de incidencia de cáncer

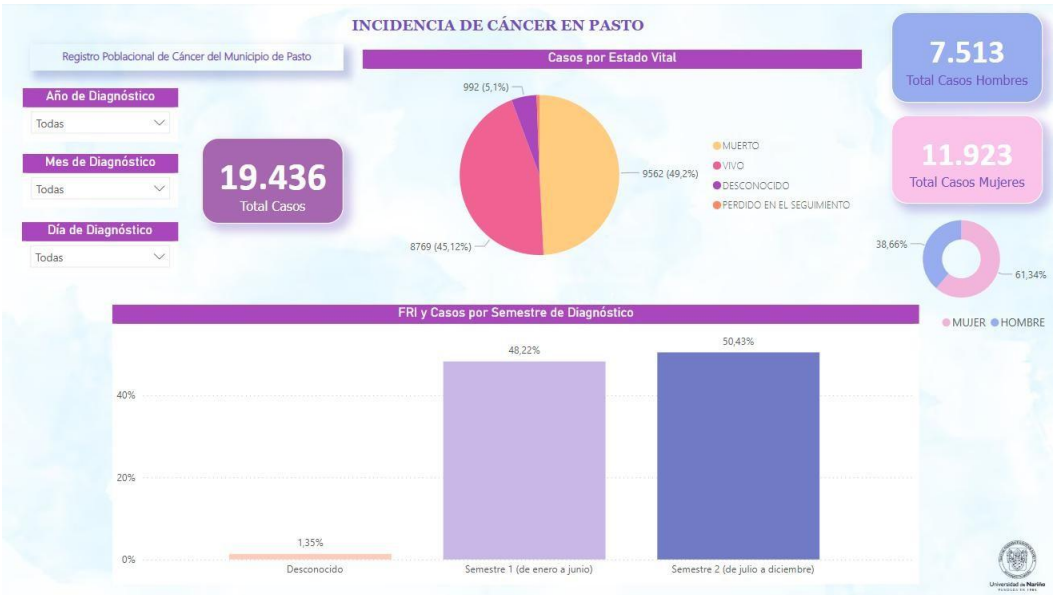
registrados con fechas de diagnóstico a partir del año 1998, un total de *19.436 casos* del municipio de Pasto.

A continuación, se muestran los dashboards disponibles en la aplicación YACHAY-ETL:

A. INCIDENCIA

1) Incidencia - Tiempo

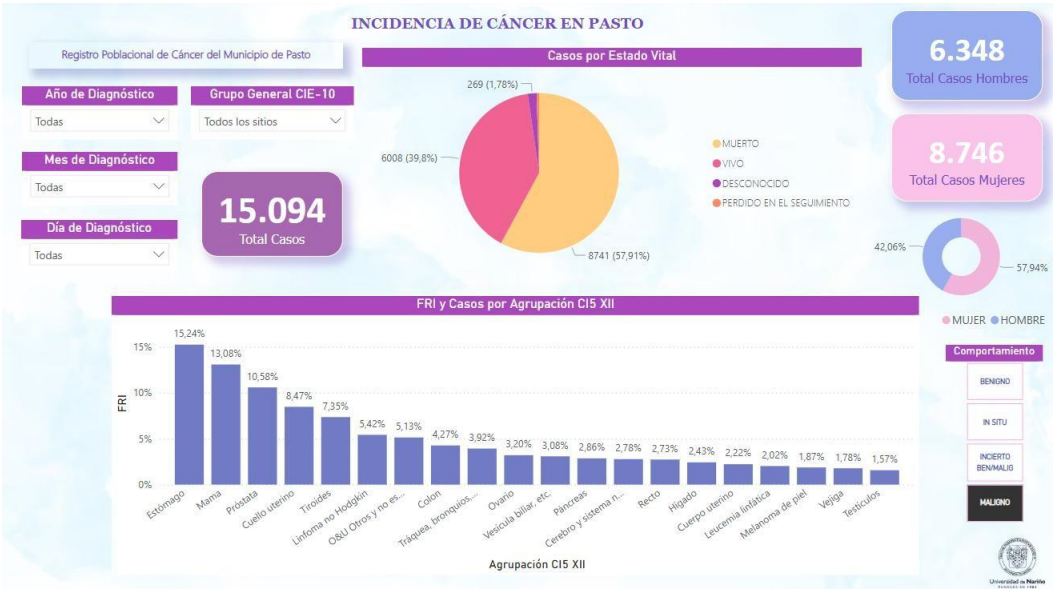
Permite visualizar la evolución del número total de casos de incidencia de cáncer en el municipio de Pasto y su respectiva frecuencia relativa de incidencia (FRI) detallados por estado vital y por fecha de diagnóstico. En cuanto a la fecha de diagnóstico, la información se organiza en diferentes niveles de jerarquía temporal (semestre, trimestre, mes y día), proporcionando una visión detallada de la progresión de la enfermedad.



2) Incidencia - Localización

Permite visualizar las localizaciones (agrupadas con base en la agrupación CI5 XII) más frecuentes del tumor a partir del número total de casos de incidencia de cáncer en el municipio de Pasto y su respectiva frecuencia relativa de incidencia (FRI) detallados por comportamiento. La información también es presentada en niveles de jerarquía de localización que incluyen la discriminación por codificación CIE-10, agrupación CI5 XII (con base en la codificación CIE-10), por CIEO-3 y por grupo de localización CIEO-3, con la posibilidad de excluir, a través del filtro “Grupo General CIE-10”, los casos asociados a los códigos de morfología entre 8070 y 8078 y entre 8090 y 8098 correspondientes a descripciones de neoplasias malignas de la piel que no se suelen incluir en conteos y reportes. Lo anterior con

el fin de especificar con mayor detalle las localizaciones más frecuentes de los tumores registrados.



3) Incidencia - Año

Permite visualizar la evolución del número total de casos de incidencia de cáncer en el municipio de Pasto y su respectiva frecuencia relativa de incidencia (FRI) detallados por estado vital y por fecha de diagnóstico y comportamiento. En cuanto a la fecha de diagnóstico, la información se organiza en diferentes niveles de jerarquía temporal incluyendo el año (año, semestre, trimestre, mes y día), proporcionando una visión detallada en un gráfico lineal de la progresión de la enfermedad.

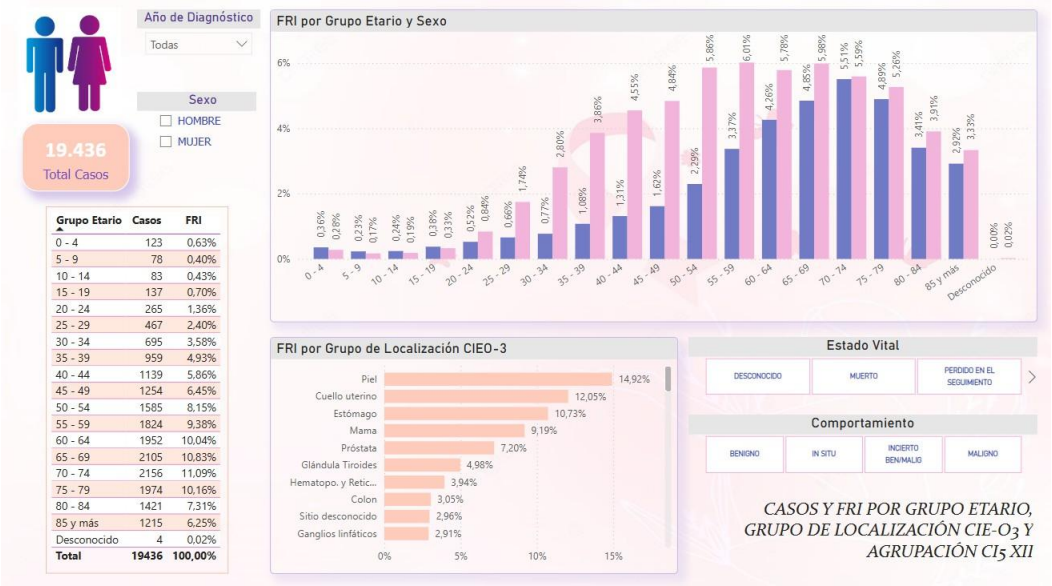


B. CASOS POR SEXO

Busca contabilizar los casos de cáncer del RPCMP enfocado especialmente en la perspectiva del Sexo.

4) Casos por Sexo

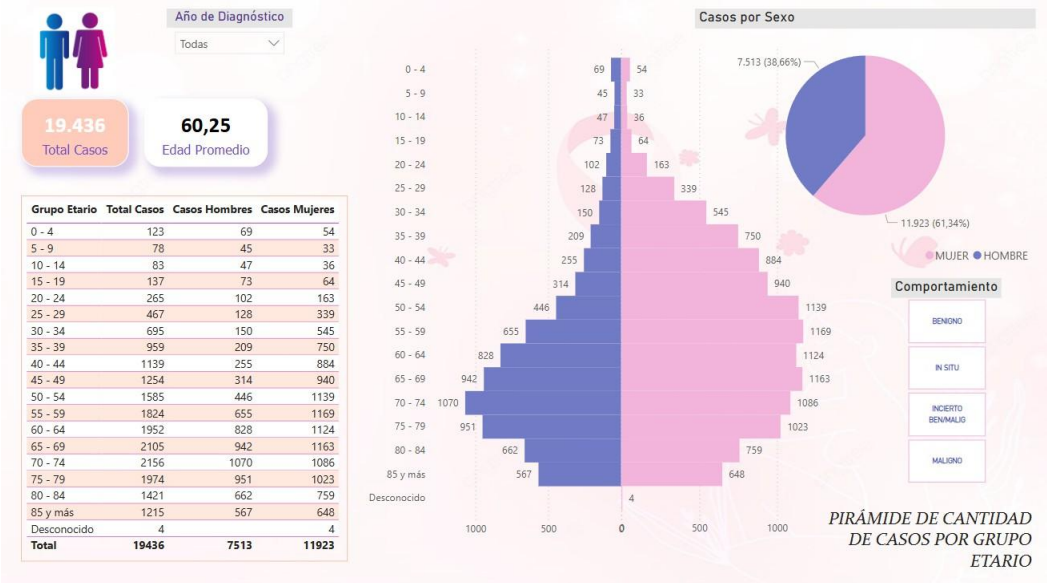
Este informe permite visualizar de forma consolidada, la frecuencia relativa de incidencia (FRI) discriminada por grupo etario, sexo, grupo de localización en CIEO-3 y agrupación CI5 XII (codificación CIE-10), filtrando por año de diagnóstico, sexo, estado vital y por comportamiento. Lo anterior con el fin de diferenciar el comportamiento del cáncer en diferentes grupos de edad tanto en hombres como en mujeres.



Grupo Etario	Casos	FRI
0 - 4	123	0,63%
5 - 9	78	0,40%
10 - 14	83	0,43%
15 - 19	137	0,70%
20 - 24	265	1,36%
25 - 29	467	2,40%
30 - 34	695	3,58%
35 - 39	959	4,93%
40 - 44	1139	5,86%
45 - 49	1254	6,45%
50 - 54	1585	8,15%
55 - 59	1824	9,38%
60 - 64	1952	10,04%
65 - 69	2105	10,83%
70 - 74	2156	11,09%
75 - 79	1974	10,16%
80 - 84	1421	7,31%
85 y más	1215	6,25%
Desconocido	4	0,02%
Total	**19436**	**100,00%**

5) Pirámide de Casos

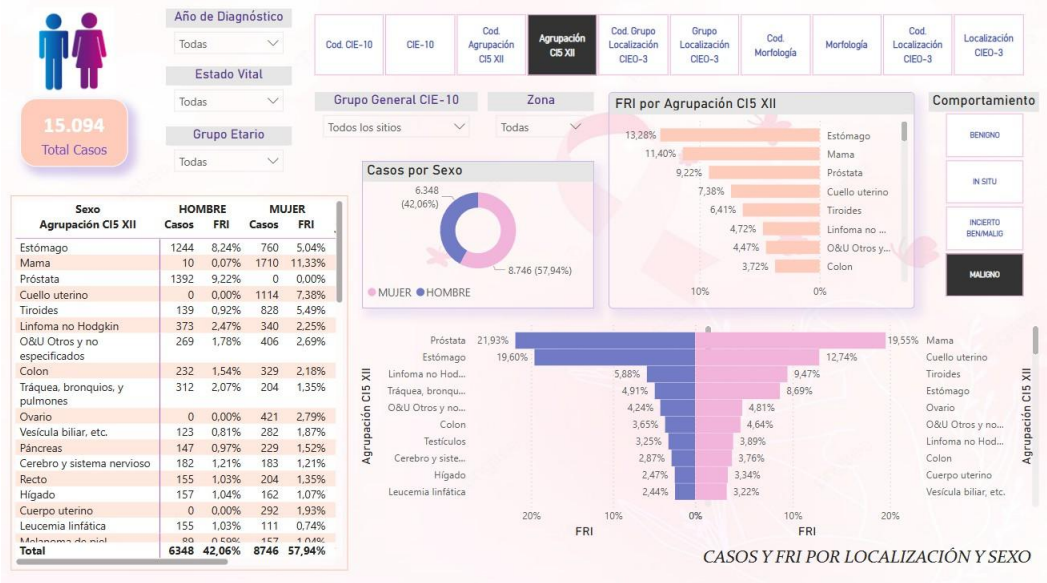
Permite visualizar el número total de casos de incidencia de cáncer del municipio de Pasto discriminados por sexo y grupo etario en un gráfico de pirámide con la posibilidad de segmentar la muestra por año de diagnóstico y comportamiento, además de reportar el cálculo de la edad promedio al momento del diagnóstico. Los análisis que resulten de la interacción con este informe proporcionan un contexto demográfico por edad detallado para los casos de incidencia de cáncer en el municipio y permiten comparar fácilmente los rangos de edad en el comportamiento de la incidencia.



Grupo Etario	Total Casos	Casos Hombres	Casos Mujeres
0 - 4	123	69	54
5 - 9	78	45	33
10 - 14	83	47	36
15 - 19	137	73	64
20 - 24	265	102	163
25 - 29	467	128	339
30 - 34	695	150	545
35 - 39	959	209	750
40 - 44	1139	255	884
45 - 49	1254	314	940
50 - 54	1585	446	1139
55 - 59	1824	655	1169
60 - 64	1952	828	1124
65 - 69	2105	942	1163
70 - 74	2156	1070	1086
75 - 79	1974	951	1023
80 - 84	1421	662	759
85 y más	1215	567	648
Desconocido	4		4
Total	**19436**	**7513**	**11923**

6) FRI por Sexo y Localización

Este informe permite visualizar la frecuencia relativa de incidencia (FRI) discriminada por sexo y por localización del tumor que incluye la discriminación por codificación CIE-10, agrupación CI5 XII (codificación CIE-10), por CIEO-3, por grupo de localización CIEO-3 y por morfología con la posibilidad de segmentar el informe por año de diagnóstico, estado vital, grupo etario, comportamiento y por zona, con la posibilidad de excluir, a través del filtro “Grupo General CIE-10”, los casos asociados a los códigos de morfología entre 8070 y 8078 y entre 8090 y 8098 correspondientes a descripciones de neoplasias malignas de la piel que no se suelen incluir en estos conteos y reportes.
Complementario al informe anterior, este refleja la proporción de los casos en un gráfico piramidal en orden de las localizaciones del tumor más frecuentes discriminadas por sexo para evidenciar el predominio de la incidencia del cáncer tanto en hombres como en mujeres, así como para el total de la población.

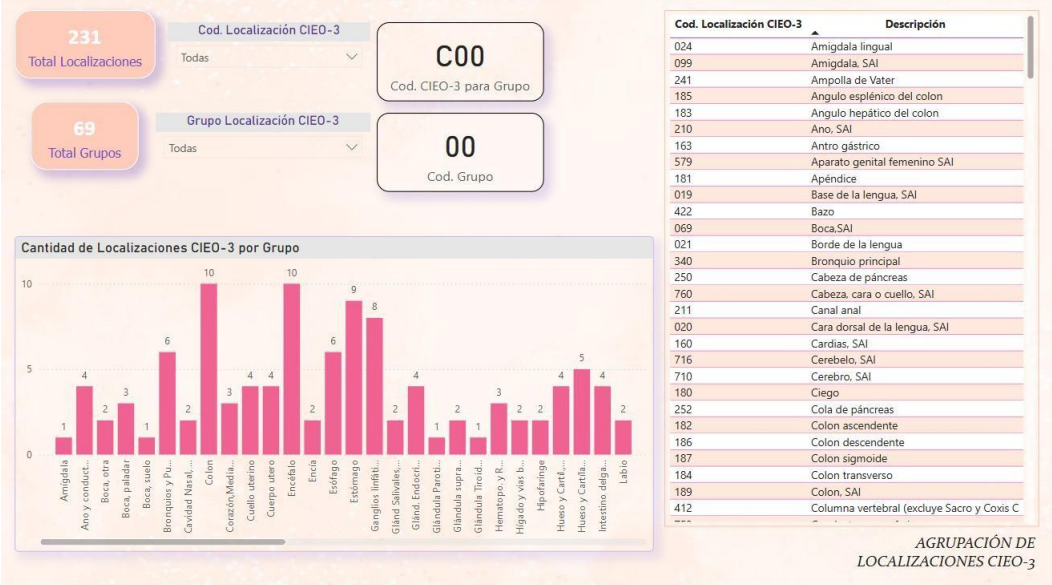


C. CASOS POR LOCALIZACIÓN

En estos informes se busca contabilizar los casos de cáncer del RPCMP y su respectiva frecuencia relativa de incidencia (FRI) enfocado especialmente en la perspectiva de Localización que puede incluir una o varias de las siguientes variables: codificación CIE-10, agrupación CI5 XII (con base en la codificación CIE-10), localización en CIEO-3, grupo de localización CIEO-3 y morfología en CIEO-3.

1) Agrupación de Localizaciones CIEO-3

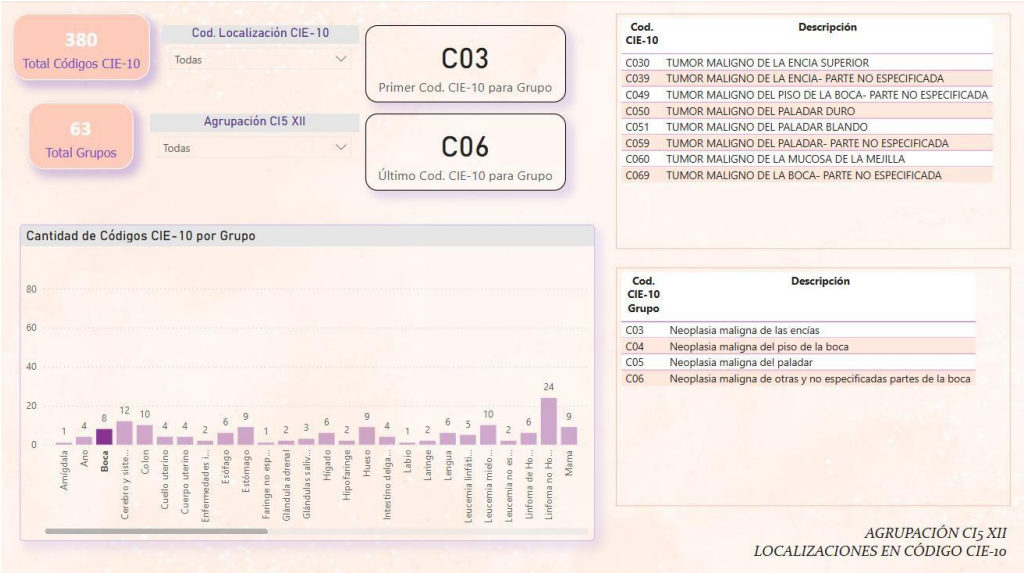
El informe se crea con el fin de especificar las localizaciones del tumor en codificación CIEO-3 que hacen parte de un mismo grupo de localización CIEO-3. Su utilidad radica en la importancia de ser consciente de las sub localizaciones del tumor que se estarían incluyendo en un informe cuando se reportan los casos por grupo de localización CIEO-3. De esta forma el dashboard permite visualizar el número total de localizaciones CIEO-3 por grupo, filtrando por código de localización CIEO-3 o por nombre del grupo y obteniendo la descripción de las localizaciones que hacen parte de un mismo grupo.



AGRUPACIÓN DE LOCALIZACIONES CIEO-3

2) Agrupación CI5 XII (CIE-10)

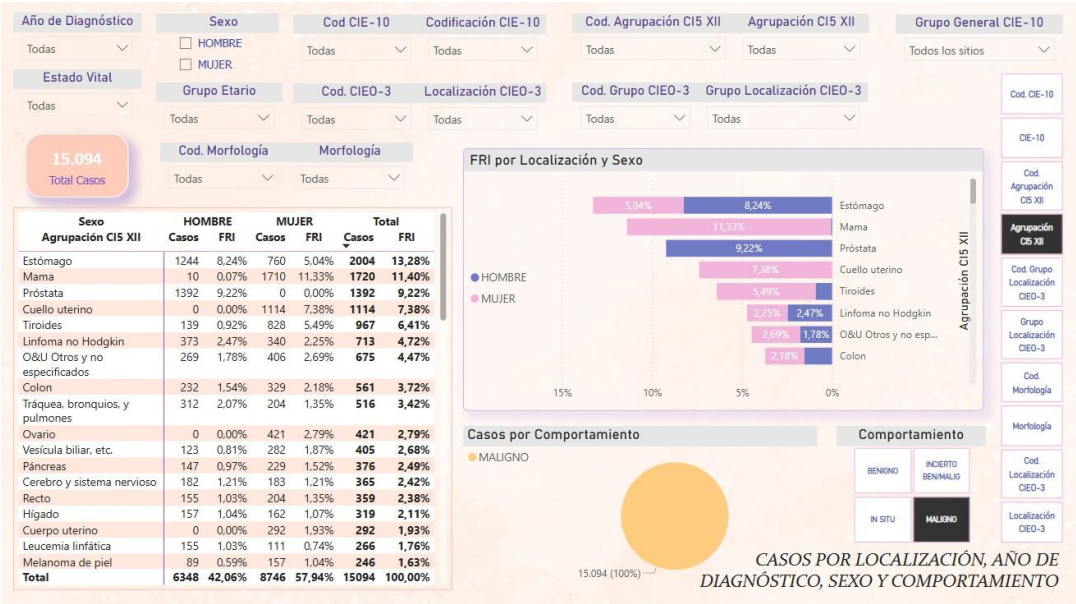
De igual forma este informe permite especificar las localizaciones del tumor en codificación CIE-10 que hacen parte de una misma agrupación CI5 XII según la clasificación consolidada en la Bodega de Datos a partir de la información dada en la Incidencia de Cáncer en los 5 Continentes [28]. De esta forma se visualiza el número total de códigos CIE-10 por grupo, filtrando por código CIE-10 o por nombre del grupo CI5 XII y obteniendo la descripción de los códigos CIE-10 que hacen parte de un mismo grupo.



AGRUPACIÓN CI5 XII LOCALIZACIONES EN CÓDIGO CIE-10

3) Casos por Localización

Este informe se enfoca en visualizar el número total de casos de incidencia de cáncer del municipio de Pasto y su respectiva frecuencia relativa de incidencia (FRI) discriminados por comportamiento y localización. Brinda la posibilidad de analizar este comportamiento por año de diagnóstico, por sexo, por grupo etario, por estado vital y por comportamiento del tumor, con la posibilidad de excluir, a través del filtro “Grupo General CIE-10”, los casos asociados a los códigos de morfología entre 8070 y 8078 y entre 8090 y 8098 correspondientes a descripciones de neoplasias malignas de la piel que no se suelen incluir en estos conteos y reportes. Lo anterior brinda una perspectiva amplia y específica de la localización del tumor en los casos registrados por el RPCMP.

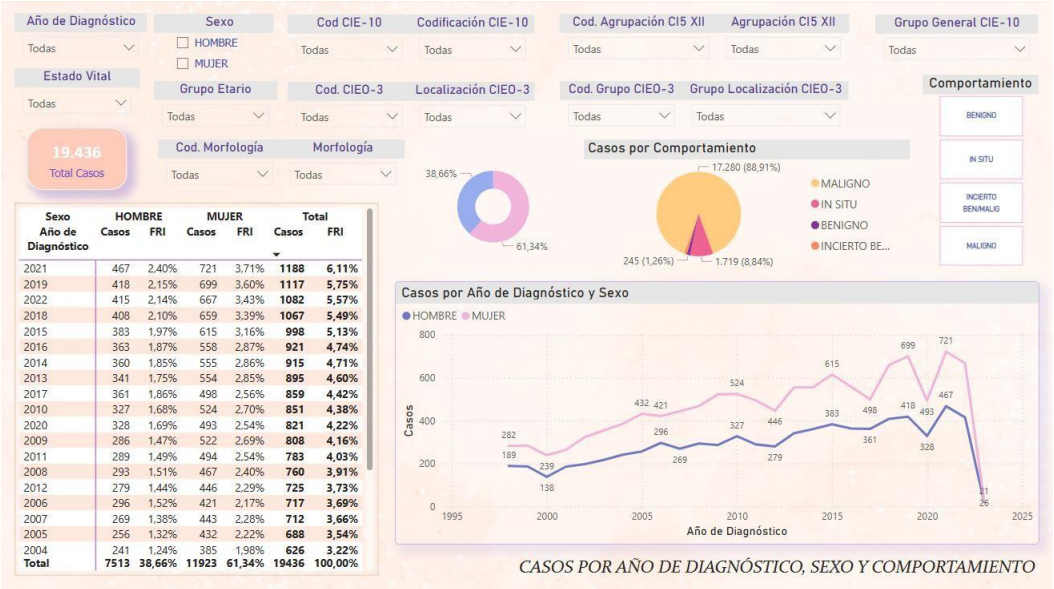


Sexo / Agrupación CI5 XII	HOMBRE Casos	HOMBRE FRI	MUJER Casos	MUJER FRI	Total Casos	Total FRI
Estómago	1244	8,24%	760	5,04%	**2004**	**13,28%**
Mama	10	0,07%	1710	11,33%	**1720**	**11,40%**
Próstata	1392	9,22%	0	0,00%	**1392**	**9,22%**
Cuello uterino	0	0,00%	1114	7,38%	**1114**	**7,38%**
Tiroides	139	0,92%	828	5,49%	**967**	**6,41%**
Linfoma no Hodgkin	373	2,47%	340	2,25%	**713**	**4,72%**
O&U Otros y no especificados	269	1,78%	406	2,69%	**675**	**4,47%**
Colon	232	1,54%	329	2,18%	**561**	**3,72%**
Tráquea, bronquios, y pulmones	312	2,07%	204	1,35%	**516**	**3,42%**
Ovario	0	0,00%	421	2,79%	**421**	**2,79%**
Vesícula biliar, etc.	123	0,81%	282	1,87%	**405**	**2,68%**
Páncreas	147	0,97%	229	1,52%	**376**	**2,49%**
Cerebro y sistema nervioso	182	1,21%	183	1,21%	**365**	**2,42%**
Recto	155	1,03%	204	1,35%	**359**	**2,38%**
Hígado	157	1,04%	162	1,07%	**319**	**2,11%**
Cuerpo uterino	0	0,00%	292	1,93%	**292**	**1,93%**
Leucemia linfática	155	1,03%	111	0,74%	**266**	**1,76%**
Melanoma de piel	89	0,59%	157	1,04%	**246**	**1,63%**
Total	**6348**	**42,06%**	**8746**	**57,94%**	**15094**	**100,00%**

CASOS POR LOCALIZACIÓN, AÑO DE DIAGNÓSTICO, SEXO Y COMPORTAMIENTO

4) Casos en el Tiempo – Comportamiento

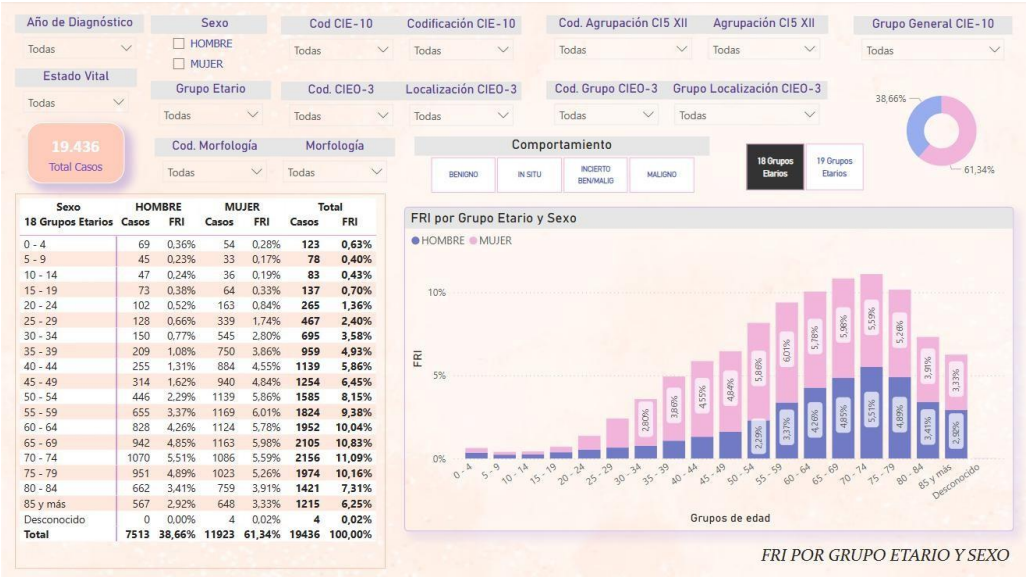
Este informe se enfoca en la evolución del número total de casos de incidencia de cáncer en el municipio de Pasto detallados por comportamiento y en el tiempo, por sexo y por año de diagnóstico. Es posible configurar un escenario de análisis que dependa del año de diagnóstico, del sexo, grupo etario, estado vital, comportamiento y/o localización para evidenciar el comportamiento del cáncer en el municipio y determinar las localizaciones más frecuentes o menos frecuentes y el comportamiento a lo largo del tiempo.



Sexo / Año de Diagnóstico	HOMBRE Casos	HOMBRE FRI	MUJER Casos	MUJER FRI	Total Casos	Total FRI
2021	467	2,40%	721	3,71%	**1188**	**6,11%**
2019	418	2,15%	699	3,60%	**1117**	**5,75%**
2022	415	2,14%	667	3,43%	**1082**	**5,57%**
2018	408	2,10%	659	3,39%	**1067**	**5,49%**
2015	383	1,97%	615	3,16%	**998**	**5,13%**
2016	363	1,87%	558	2,87%	**921**	**4,74%**
2014	360	1,85%	555	2,86%	**915**	**4,71%**
2013	341	1,75%	554	2,85%	**895**	**4,60%**
2017	361	1,86%	498	2,56%	**859**	**4,42%**
2010	327	1,68%	524	2,70%	**851**	**4,38%**
2020	328	1,69%	493	2,54%	**821**	**4,22%**
2009	286	1,47%	522	2,69%	**808**	**4,16%**
2011	289	1,49%	494	2,54%	**783**	**4,03%**
2008	293	1,51%	467	2,40%	**760**	**3,91%**
2012	279	1,44%	446	2,29%	**725**	**3,73%**
2006	296	1,52%	421	2,17%	**717**	**3,69%**
2007	269	1,38%	443	2,28%	**712**	**3,66%**
2005	256	1,32%	432	2,22%	**688**	**3,54%**
2004	241	1,24%	385	1,98%	**626**	**3,22%**
Total	**7513**	**38,66%**	**11923**	**61,34%**	**19436**	**100,00%**

5) FRI por Grupo Etario – Comportamiento

A diferencia del informe anterior, este dashboard reporta la frecuencia relativa de incidencia (FRI) por grupo etario y por sexo con la posibilidad de segmentar la muestra por el año de diagnóstico, sexo, grupo etario, estado vital, comportamiento y por localización en los diferentes códigos y grupos. Un análisis de este tipo brinda información sobre el grupo quinquenal de edad más impactado por el cáncer en el municipio tanto en hombres como en mujeres y especialmente es relevante cuando se analiza una localización del tumor específica.



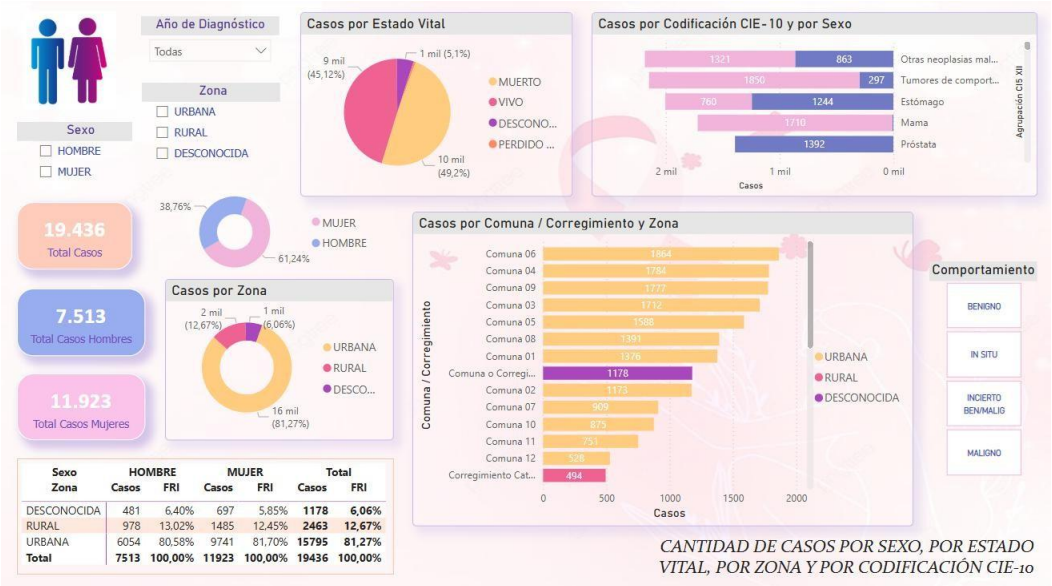
Sexo / 18 Grupos Etarios	HOMBRE Casos	HOMBRE FRI	MUJER Casos	MUJER FRI	Total Casos	Total FRI
0 - 4	69	0,36%	54	0,28%	**123**	**0,63%**
5 - 9	45	0,23%	33	0,17%	**78**	**0,40%**
10 - 14	47	0,24%	36	0,19%	**83**	**0,43%**
15 - 19	73	0,38%	64	0,33%	**137**	**0,70%**
20 - 24	102	0,52%	163	0,84%	**265**	**1,36%**
25 - 29	128	0,66%	339	1,74%	**467**	**2,40%**
30 - 34	150	0,77%	545	2,80%	**695**	**3,58%**
35 - 39	209	1,08%	750	3,86%	**959**	**4,93%**
40 - 44	255	1,31%	884	4,55%	**1139**	**5,86%**
45 - 49	314	1,62%	940	4,84%	**1254**	**6,45%**
50 - 54	446	2,29%	1139	5,86%	**1585**	**8,15%**
55 - 59	655	3,37%	1169	6,01%	**1824**	**9,38%**
60 - 64	828	4,26%	1124	5,78%	**1952**	**10,04%**
65 - 69	942	4,85%	1163	5,98%	**2105**	**10,83%**
70 - 74	1070	5,51%	1086	5,59%	**2156**	**11,09%**
75 - 79	951	4,89%	1023	5,26%	**1974**	**10,16%**
80 - 84	662	3,41%	759	3,91%	**1421**	**7,31%**
85 y más	567	2,92%	648	3,33%	**1215**	**6,25%**
Desconocido	0	0,00%	4	0,02%	**4**	**0,02%**
Total	**7513**	**38,66%**	**11923**	**61,34%**	**19436**	**100,00%**

D. CASOS POR ZONA Y CODIFICACIÓN CIE-10

Teniendo en cuenta que generalmente se reporta la localización del tumor en codificación CIE-10, este informe busca contabilizar los casos de cáncer del RPCMP enfocado especialmente en las perspectivas Zona y Codificación CIE-10.

1) Casos por Zona y Codificación CIE-10

Este dashboard permite de forma generalizada, visualizar el número total de casos de incidencia de cáncer del municipio de Pasto discriminados por estado vital del paciente, por sexo y por codificación CIE-10 que incluye la discriminación por código CIE-10 y la agrupación CI5 XII (con base en la codificación CIE-10), además por zona, por comuna / corregimiento y por barrio / vereda, segmentando el escenario de análisis por año de diagnóstico, por sexo, por zona y por comportamiento, esto con el fin de inspeccionar el comportamiento de la incidencia del cáncer en diferentes regiones del municipio y evidenciar por ejemplo, si existen sectores del municipio que sean recurrentes en distintos periodos de tiempo.



Sexo / Zona	HOMBRE Casos	HOMBRE FRI	MUJER Casos	MUJER FRI	Total Casos	Total FRI
DESCONOCIDA	481	6,40%	697	5,85%	**1178**	**6,06%**
RURAL	978	13,02%	1485	12,45%	**2463**	**12,67%**
URBANA	6054	80,58%	9741	81,70%	**15795**	**81,27%**
Total	**7513**	**100,00%**	**11923**	**100,00%**	**19436**	**100,00%**

CANTIDAD DE CASOS POR SEXO, POR ESTADO VITAL, POR ZONA Y POR CODIFICACIÓN CIE-10

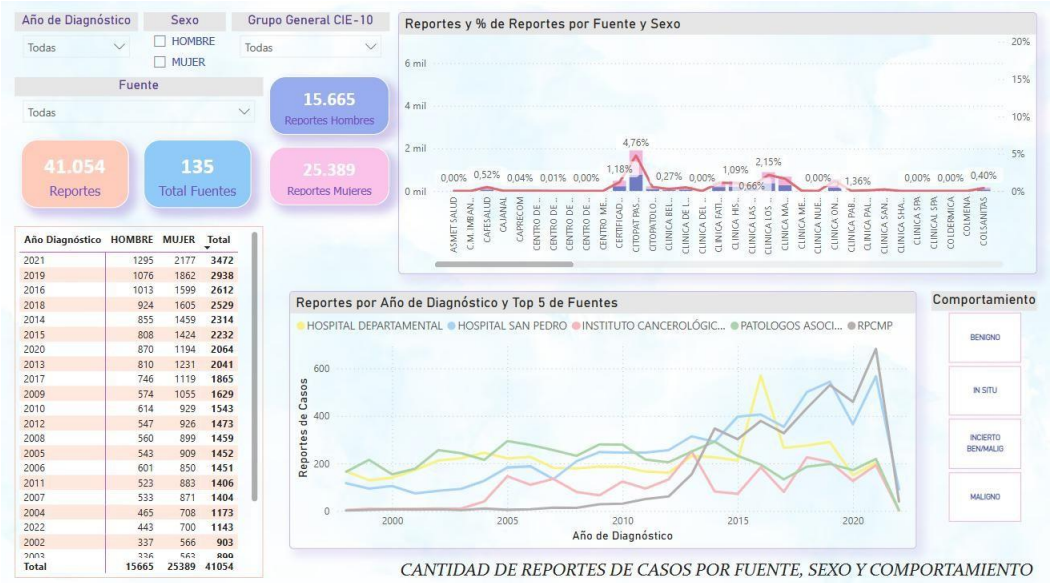
E. CASOS POR FUENTE

Busca contabilizar los reportes de casos de cáncer del RPCMP enfocados especialmente en la perspectiva Fuente con el fin de generar gráficos y reportes para una fuente específica cuando esto es solicitado por la fuente al RPCMP.

1) Casos por Fuente

Este dashboard muestra el panorama de las fuentes de información que contribuyen al RPCMP y especialmente aquellas que aportan en mayor porcentaje

con la información de los casos, es decir también representa las fuentes de información, entidades de salud, hospitales, clínicas o laboratorios a los que más acuden los pacientes con cáncer para su atención en diferentes etapas de la enfermedad. En este sentido, permite visualizar el número total de reportes de casos de incidencia de cáncer del municipio de Pasto y su respectivo porcentaje discriminados por año de diagnóstico y por fuente. Es posible adecuar el escenario de reporte para una fuente específica filtrando el año de diagnóstico, el sexo, la fuente y el comportamiento del tumor, con la posibilidad de excluir, a través del filtro “Grupo General CIE-10”, los reportes de casos asociados a los códigos de morfología entre 8070 y 8078 y entre 8090 y 8098 correspondientes a descripciones de neoplasias malignas de la piel que no se suelen incluir en estos conteos y reportes. Cabe resaltar que un reporte de un caso de cáncer hace referencia a cualquier reporte hecho por alguna fuente sobre un caso de cáncer determinado, por lo que un mismo caso de cáncer puede tener varios reportes de diferentes fuentes.

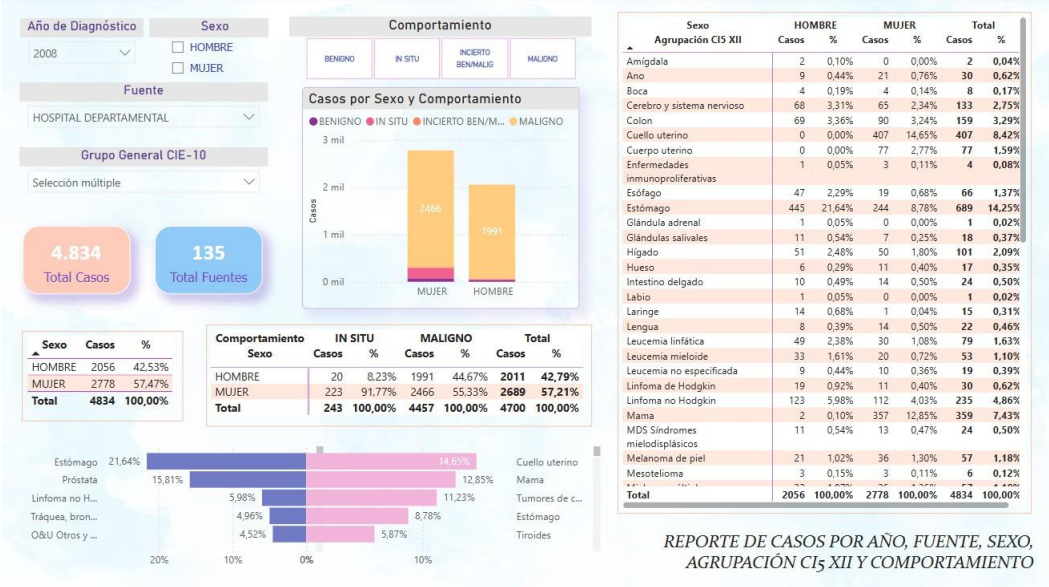


Año Diagnóstico	HOMBRE	MUJER	Total
2021	1295	2177	3472
2019	1076	1862	2938
2016	1013	1599	2612
2018	924	1605	2529
2014	855	1459	2314
2015	808	1424	2232
2020	870	1194	2064
2013	810	1231	2041
2017	746	1119	1865
2009	574	1055	1629
2010	614	929	1543
2012	547	926	1473
2008	560	899	1459
2005	543	909	1452
2006	601	850	1451
2011	523	883	1406
2007	533	871	1404
2004	465	708	1173
2022	443	700	1143
2002	337	566	903
Total	15665	25389	41054

CANTIDAD DE REPORTES DE CASOS POR FUENTE, SEXO Y COMPORTAMIENTO

2) Casos – Fuente

Este informe es más detallado para presentar los reportes de casos hechos especialmente por una misma fuente en un periodo de tiempo determinado. A través de los reportes de incidencia, se puede conocer el número total de casos de incidencia reportados por la fuente y su respectivo porcentaje discriminados por sexo, por comportamiento y por agrupación CI5 XII (con base en la codificación CIE-10). Es posible segmentar el informe por año de diagnóstico, fuente, sexo y por comportamiento y para garantizar un análisis más preciso, existe la opción de excluir aquellos casos asociados a códigos de morfología entre 8070 y 8078 y entre 8090 y 8098, que corresponden a ciertos tipos de neoplasias malignas de la piel, comúnmente no considerados en este tipo de estudios y reportes.



Sexo	Casos	%
HOMBRE	2056	42,53%
MUJER	2778	57,47%
Total	**4834**	**100,00%**

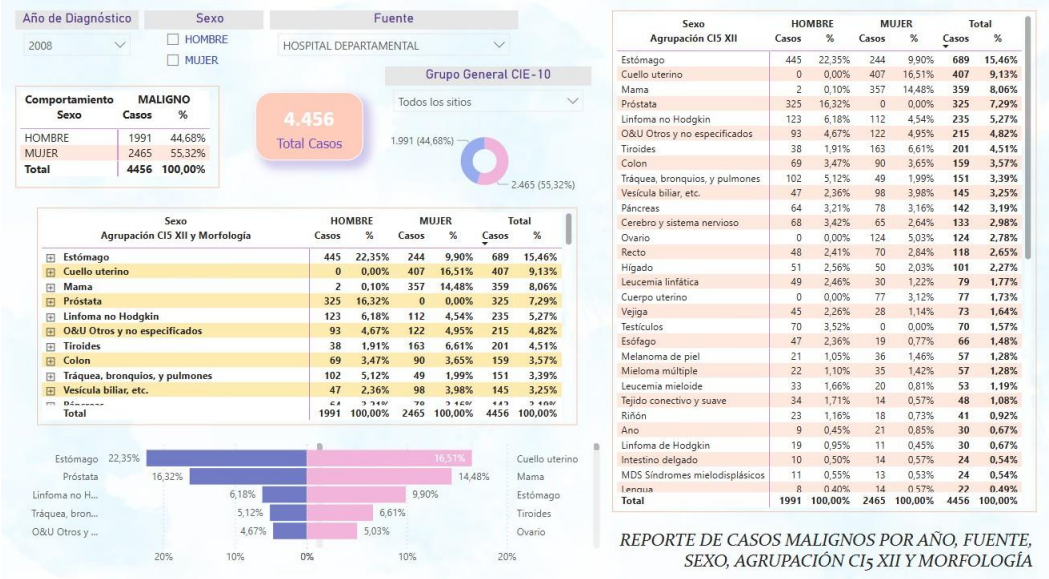
Comportamiento	IN SITU		MALIGNO		Total	
Sexo	Casos	%	Casos	%	Casos	%
HOMBRE	20	8,23%	1991	44,67%	**2011**	**42,79%**
MUJER	223	91,77%	2466	55,33%	**2689**	**57,21%**
Total	**243**	**100,00%**	**4457**	**100,00%**	**4700**	**100,00%**

Sexo	HOMBRE		MUJER		Total	
Agrupación CI5 XII	Casos	%	Casos	%	Casos	%
Amígdala	2	0,10%	0	0,00%	**2**	**0,04%**
Ano	9	0,44%	21	0,76%	**30**	**0,62%**
Boca	4	0,19%	4	0,14%	**8**	**0,17%**
Cerebro y sistema nervioso	68	3,31%	65	2,34%	**133**	**2,75%**
Colon	69	3,36%	90	3,24%	**159**	**3,29%**
Cuello uterino	0	0,00%	407	14,65%	**407**	**8,42%**
Cuerpo uterino	0	0,00%	77	2,77%	**77**	**1,59%**
Enfermedades inmunoproliferativas	1	0,05%	3	0,11%	**4**	**0,08%**
Esófago	47	2,29%	19	0,68%	**66**	**1,37%**
Estómago	445	21,64%	244	8,78%	**689**	**14,25%**
Glándula adrenal	1	0,05%	0	0,00%	**1**	**0,02%**
Glándulas salivales	11	0,54%	7	0,25%	**18**	**0,37%**
Hígado	51	2,48%	50	1,80%	**101**	**2,09%**
Hueso	6	0,29%	11	0,40%	**17**	**0,35%**
Intestino delgado	10	0,49%	14	0,50%	**24**	**0,50%**
Labio	1	0,05%	0	0,00%	**1**	**0,02%**
Laringe	14	0,68%	1	0,04%	**15**	**0,31%**
Lengua	8	0,39%	14	0,50%	**22**	**0,46%**
Leucemia linfática	49	2,38%	30	1,08%	**79**	**1,63%**
Leucemia mieloide	33	1,61%	20	0,72%	**53**	**1,10%**
Leucemia no especificada	9	0,44%	10	0,36%	**19**	**0,39%**
Linfoma de Hodgkin	19	0,92%	11	0,40%	**30**	**0,62%**
Linfoma no Hodgkin	123	5,98%	112	4,03%	**235**	**4,86%**
Mama	2	0,10%	357	12,85%	**359**	**7,43%**
MDS Síndromes mielodisplásicos	11	0,54%	13	0,47%	**24**	**0,50%**
Melanoma de piel	21	1,02%	36	1,30%	**57**	**1,18%**
Mesotelioma	3	0,15%	3	0,11%	**6**	**0,12%**
Total	**2056**	**100,00%**	**2778**	**100,00%**	**4834**	**100,00%**

REPORTE DE CASOS POR AÑO, FUENTE, SEXO, AGRUPACIÓN CI5 XII Y COMPORTAMIENTO

3) Casos Malignos – Fuente

Este informe especifica aún más el dashboard anterior, haciendo énfasis en los tumores de comportamiento maligno puesto que su análisis permite visualizar el impacto de esta enfermedad en la población. Permite visualizar el número total de reportes de casos de incidencia de cáncer del municipio de Pasto de comportamiento maligno y su respectivo porcentaje discriminados por agrupación CI5 XII. El escenario de análisis puede especificar el año de diagnóstico, la fuente, discriminar los reportes por sexo y por agrupación CI5 XII. De igual forma permite excluir los casos de ciertas neoplasias malignas de la piel, comúnmente no considerados en este tipo de estudios y reportes.



Comportamiento	MALIGNO		
Sexo	Casos	%	
HOMBRE	1991	44,68%	
MUJER	2465	55,32%	
Total	**4456**	**100,00%**	

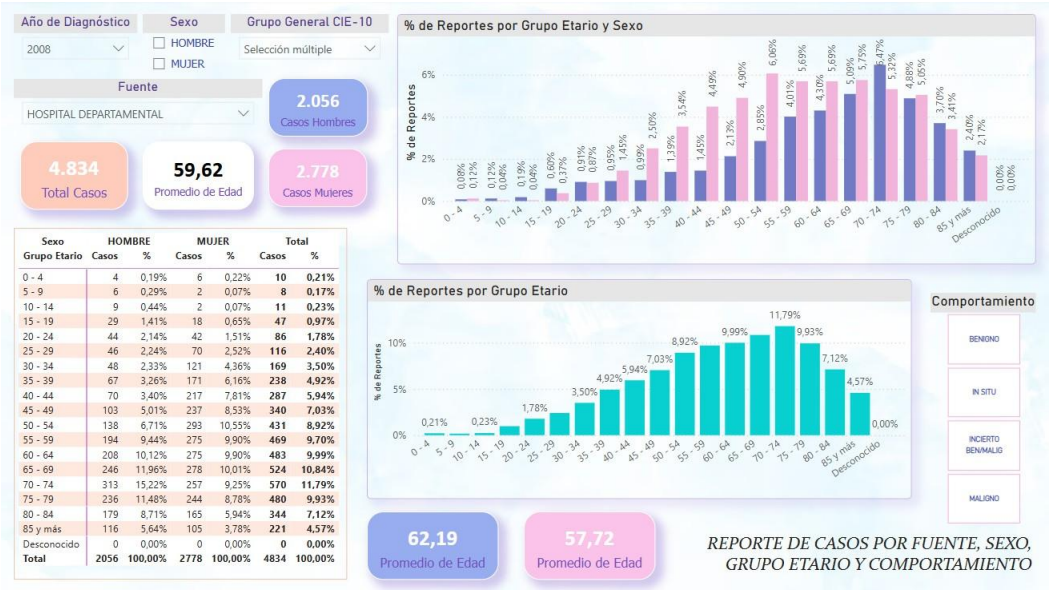
Sexo	HOMBRE		MUJER		Total	
Agrupación CI5 XII y Morfología	Casos	%	Casos	%	Casos	%
⊞ Estómago	445	22,35%	244	9,90%	689	15,46%
⊞ Cuello uterino	0	0,00%	407	16,51%	407	9,13%
⊞ Mama	2	0,10%	357	14,48%	359	8,06%
⊞ Próstata	325	16,32%	0	0,00%	325	7,29%
⊞ Linfoma no Hodgkin	123	6,18%	112	4,54%	235	5,27%
⊞ O&U Otros y no especificados	93	4,67%	122	4,95%	215	4,82%
⊞ Tiroides	38	1,91%	163	6,61%	201	4,51%
⊞ Colon	69	3,47%	90	3,65%	159	3,57%
⊞ Tráquea, bronquios, y pulmones	102	5,12%	49	1,99%	151	3,39%
⊞ Vesícula biliar, etc.	47	2,36%	98	3,98%	145	3,25%
Total	**1991**	**100,00%**	**2465**	**100,00%**	**4456**	**100,00%**

Sexo	HOMBRE		MUJER		Total	
Agrupación CI5 XII	Casos	%	Casos	%	Casos	%
Estómago	445	22,35%	244	9,90%	**689**	**15,46%**
Cuello uterino	0	0,00%	407	16,51%	**407**	**9,13%**
Mama	2	0,10%	357	14,48%	**359**	**8,06%**
Próstata	325	16,32%	0	0,00%	**325**	**7,29%**
Linfoma no Hodgkin	123	6,18%	112	4,54%	**235**	**5,27%**
O&U Otros y no especificados	93	4,67%	122	4,95%	**215**	**4,82%**
Tiroides	38	1,91%	163	6,61%	**201**	**4,51%**
Colon	69	3,47%	90	3,65%	**159**	**3,57%**
Tráquea, bronquios, y pulmones	102	5,12%	49	1,99%	**151**	**3,39%**
Vesícula biliar, etc.	47	2,36%	98	3,98%	**145**	**3,25%**
Páncreas	64	3,21%	78	3,16%	**142**	**3,19%**
Cerebro y sistema nervioso	68	3,42%	65	2,64%	**133**	**2,98%**
Ovario	0	0,00%	124	5,03%	**124**	**2,78%**
Recto	48	2,41%	70	2,84%	**118**	**2,65%**
Hígado	51	2,56%	50	2,03%	**101**	**2,27%**
Leucemia linfática	49	2,46%	30	1,22%	**79**	**1,77%**
Cuerpo uterino	0	0,00%	77	3,12%	**77**	**1,73%**
Vejiga	45	2,26%	28	1,14%	**73**	**1,64%**
Testículos	70	3,52%	0	0,00%	**70**	**1,57%**
Esófago	47	2,36%	19	0,77%	**66**	**1,48%**
Melanoma de piel	21	1,05%	36	1,46%	**57**	**1,28%**
Mieloma múltiple	22	1,10%	35	1,42%	**57**	**1,28%**
Leucemia mieloide	33	1,66%	20	0,81%	**53**	**1,19%**
Tejido conectivo y suave	34	1,71%	14	0,57%	**48**	**1,08%**
Riñón	23	1,16%	18	0,73%	**41**	**0,92%**
Ano	9	0,45%	21	0,85%	**30**	**0,67%**
Linfoma de Hodgkin	19	0,95%	11	0,45%	**30**	**0,67%**
Intestino delgado	10	0,50%	14	0,57%	**24**	**0,54%**
MDS Síndromes mielodisplásicos	11	0,55%	13	0,53%	**24**	**0,54%**
Lengua	8	0,40%	14	0,57%	**22**	**0,49%**
Total	**1991**	**100,00%**	**2465**	**100,00%**	**4456**	**100,00%**

REPORTE DE CASOS MALIGNOS POR AÑO, FUENTE, SEXO, AGRUPACIÓN CI5 XII Y MORFOLOGÍA

4) Casos por Grupo Etario - Fuente

En este punto también suele ser importante conocer la proporción de casos que se reportan por grupos de edad desde una fuente determinada puesto que puede ayudar a identificar la recurrencia de ciertos tipos de población (juvenil, adulto mayor, etc.) a entidades de salud determinadas. Este informe permite visualizar el porcentaje de reportes de casos de incidencia de cáncer del municipio de Pasto discriminado por grupo etario y por sexo. Un análisis específico filtra el año de diagnóstico, la fuente, el sexo y el comportamiento. Además, se reporta la edad promedio por grupo etario, general y general por sexo de los casos reportados que cumplen los filtros planteados.



Sexo / Grupo Etario	HOMBRE Casos	HOMBRE %	MUJER Casos	MUJER %	Total Casos	Total %
0 - 4	4	0,19%	6	0,22%	10	0,21%
5 - 9	6	0,29%	2	0,07%	8	0,17%
10 - 14	9	0,44%	2	0,07%	11	0,23%
15 - 19	29	1,41%	18	0,65%	47	0,97%
20 - 24	44	2,14%	42	1,51%	86	1,78%
25 - 29	46	2,24%	70	2,52%	116	2,40%
30 - 34	48	2,33%	121	4,36%	169	3,50%
35 - 39	67	3,26%	171	6,16%	238	4,92%
40 - 44	70	3,40%	217	7,81%	287	5,94%
45 - 49	103	5,01%	237	8,53%	340	7,03%
50 - 54	138	6,71%	293	10,55%	431	8,92%
55 - 59	194	9,44%	275	9,90%	469	9,70%
60 - 64	208	10,12%	275	9,90%	483	9,99%
65 - 69	246	11,96%	278	10,01%	524	10,84%
70 - 74	313	15,22%	257	9,25%	570	11,79%
75 - 79	236	11,48%	244	8,78%	480	9,93%
80 - 84	179	8,71%	165	5,94%	344	7,12%
85 y más	116	5,64%	105	3,78%	221	4,57%
Desconocido	0	0,00%	0	0,00%	0	0,00%
Total	**2056**	**100,00%**	**2778**	**100,00%**	**4834**	**100,00%**

F. TASAS

El objetivo fundamental de los registros poblacionales es el cálculo de las tasas de incidencia de cáncer y su utilización para estudiar el riesgo de padecer determinados tipos de cáncer en el área cubierta por el registro en comparación con cualquier otra área, o comparar diferentes subgrupos de la población dentro de la propia área cubierta por el registro.

La incidencia expresa el número de casos nuevos de cáncer que aparecen en una población definida de individuos libres de la enfermedad, siendo la tasa de incidencia el número de tales acontecimientos durante un periodo de tiempo específico.

Por lo anterior, los siguientes dashboard son los más importantes tanto para el RPCMP como para las publicaciones que se hacen en torno a la incidencia del cáncer en el municipio de Pasto puesto que el cálculo de las diferentes Tasas refleja la evolución de la enfermedad en el tiempo, su impacto y el de los diversos tipos de cáncer en grupos de edad, por sexo y en una población específica. Los

informes buscan calcular las diferentes Tasas expuestas en este trabajo a partir de la cantidad de casos de cáncer del RPCMP, las Proyecciones de Población del municipio de Pasto y la Población Mundial Estándar, con el fin de generar tableros dinámicos para filtros específicos por diferentes perspectivas entre las que se encuentra la localización del tumor que incluye, para todos los informes, la discriminación por codificación CIE-10, agrupación CI5 XII (con base en la codificación CIE-10), la localización en CIEO-3, grupo de localización CIEO-3 y morfología, con la posibilidad de excluir, para un análisis más preciso, a través del filtro “Grupo General CIE-10”, los casos asociados a los códigos de morfología entre 8070 y 8078 y entre 8090 y 8098 de ciertas neoplasias malignas de la piel que generalmente no se incluyen en estos estudios. Adicionalmente, en algunos dashboards se brinda la opción de discriminar la respectiva Tasa en 18 grupos quinquenales de edad o en 19 grupos de edad, donde el grupo quinquenal de 0 a 4 años de edad, se subdivide en dos grupos, uno para edad de 0 años conocidos como Lactantes y otro grupo para los casos con edad de diagnóstico entre 1 y 4 años. Estos informes también incluyen la posibilidad de visualizar en algunos gráficos de forma independiente o en la misma gráfica el cálculo por sexo de la respectiva Tasa x 100.000 hombres y/o mujeres año. En cuanto a las tablas incluidas en los informes, cabe anotar que cuando el nombre del informe incluye la expresión “(General)” presenta una tabla del cálculo de Tasas por Grupo Etario totalizada para el Sexo y cuando se refiere la palabra “(Extendida)” el reporte tabular también discrimina el cálculo de las Tasas por Sexo.

1) TCI (General) por Año – Sexo

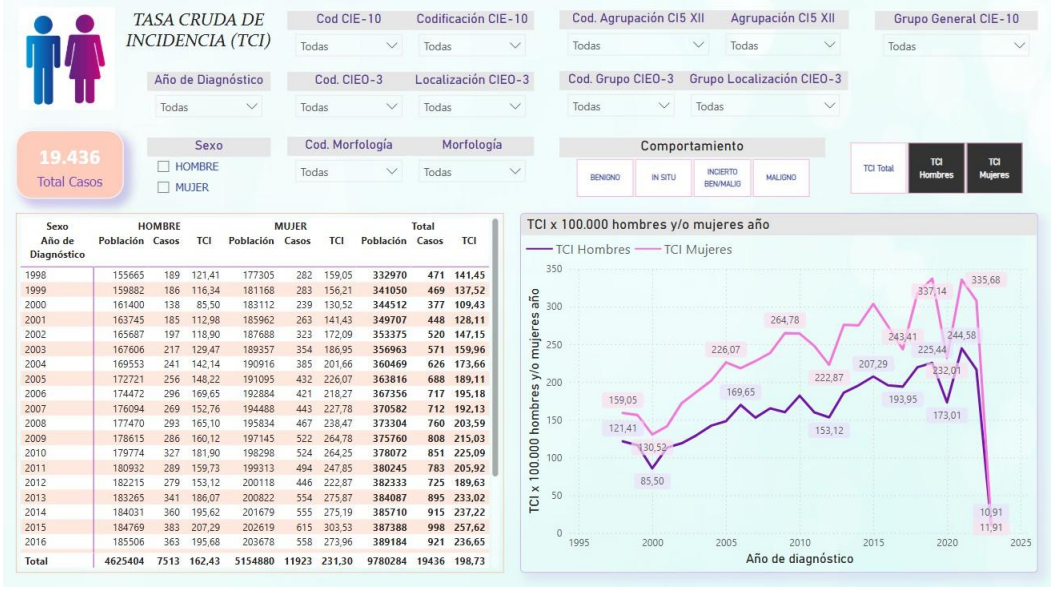
El cálculo de la Tasa Cruda de Incidencia (TCI) general (x 100.000 hombres y/o mujeres año) permite observar cómo afecta el cáncer a distintos grupos de edad en el municipio de Pasto. Esta medida proporciona una estimación directa de la probabilidad o riesgo de enfermar y es fundamental en los estudios epidemiológicos. Este análisis muestra la distribución de casos según el año de diagnóstico, el sexo y estado vital de los pacientes, el tipo de comportamiento tumoral y las características de localización y morfología del tumor basadas en las codificaciones CIEO-3 y CIE-10. Este enfoque ayuda a generar datos más precisos y adaptados a los estándares de investigación.



Año de Diagnóstico	Población	Casos	TCI
1998	332970	471	141,45
1999	341050	469	137,52
2000	344512	377	109,43
2001	349707	448	128,11
2002	353375	520	147,15
2003	356963	571	159,96
2004	360469	626	173,66
2005	363816	688	189,11
2006	367356	717	195,18
2007	370582	712	192,13
2008	373304	760	203,59
2009	375760	808	215,03
2010	378072	851	225,09
2011	380245	783	205,92
2012	382333	725	189,63
2013	384087	895	233,02
2014	385710	915	237,22
2015	387388	998	257,62
Total	9780284	19436	198,73

2) *TCI (Extendida) por Año – Sexo*

Fue necesario recrear un segundo informe con el cálculo de la Tasa Cruda de Incidencia (TCI) con el fin de presentar en un mismo dashboard la gráfica y los datos tabulados para el cálculo de este indicador por grupo etario tanto totalizado como segregado por sexo.



Sexo / Año de Diagnóstico	HOMBRE Población	HOMBRE Casos	HOMBRE TCI	MUJER Población	MUJER Casos	MUJER TCI	Total Población	Total Casos	Total TCI
1998	155665	189	121,41	177305	282	159,05	332970	471	141,45
1999	159882	186	116,34	181168	283	156,21	341050	469	137,52
2000	161400	138	85,50	183112	239	130,52	344512	377	109,43
2001	163745	185	112,98	185962	263	141,43	349707	448	128,11
2002	165687	197	118,90	187688	323	172,09	353375	520	147,15
2003	167606	217	129,47	189357	354	186,95	356963	571	159,96
2004	169553	241	142,14	190916	385	201,66	360469	626	173,66
2005	172721	256	148,22	191095	432	226,07	363816	688	189,11
2006	174472	296	169,65	192884	421	218,27	367356	717	195,18
2007	176094	269	152,76	194488	443	227,78	370582	712	192,13
2008	177470	293	165,10	195834	467	238,47	373304	760	203,59
2009	178615	286	160,12	197145	522	264,78	375760	808	215,03
2010	179774	327	181,90	198298	524	264,25	378072	851	225,09
2011	180932	289	159,73	199313	494	247,85	380245	783	205,92
2012	182215	279	153,12	200118	446	222,87	382333	725	189,63
2013	183265	341	186,07	200822	554	275,87	384087	895	233,02
2014	184031	360	195,62	201679	555	275,19	385710	915	237,22
2015	184769	383	207,29	202619	615	303,53	387388	998	257,62
2016	185506	363	195,68	203678	558	273,96	389184	921	236,65
Total	4625404	7513	162,43	5154880	11923	231,30	9780284	19436	198,73

3) *TEEI (General) por Grupo Etario*

Con frecuencia, se obtiene el número de casos por grupos de edad quinquenales y la respectiva población en riesgo para calcular tasas de incidencia, proviene de

las estadísticas de proyección de población publicadas por el DANE. Por este motivo se calcula otro indicador relevante para el RPCMP que es la Tasa Específica por Edad de Incidencia (TEEI) que permite calcular la tasa de incidencia teniendo en cuenta la variación de la estructura de edad de la población. La TEEI es la tasa teórica que se produciría si las tasas específicas por edad observadas se aplicasen a una población de referencia, que normalmente se conoce como Población Estándar donde frecuentemente se emplea la Población Mundial Estándar. Se conoce al número de personas en cada grupo de edad de la Población Estándar como los pesos a usar en el proceso de estandarización de las tasas de incidencia. Su amplio uso facilita la comparación de los niveles de cáncer en diferentes áreas.

Con base en lo anterior, este dashboard presenta el cálculo general de la Tasa Específica por Edad de Incidencia (TEEI) x 100.000 hombres y mujeres año por grupos de edad. Se brinda la opción de discriminar el indicador en 18 grupos quinquenales de edad o en 19 grupos de edad, donde el grupo quinquenal de 0 a 4 años de edad, se subdivide en dos grupos, uno para edad de 0 años conocidos como Lactantes y otro grupo para los casos con edad de diagnóstico entre 1 y 4 años. Adicionalmente, se presenta la información tabulada por grupo etario y el gráfico evidencia tanto el número total de casos de incidencia de cáncer como la TEEI por grupo de edad para los respectivos filtros aplicados de año de diagnóstico, sexo de los pacientes, comportamiento tumoral y las características de localización y morfología basadas en las codificaciones CIEO-3 y CIE-10.



4) TEEI (General) por Grupo Etario – Sexo

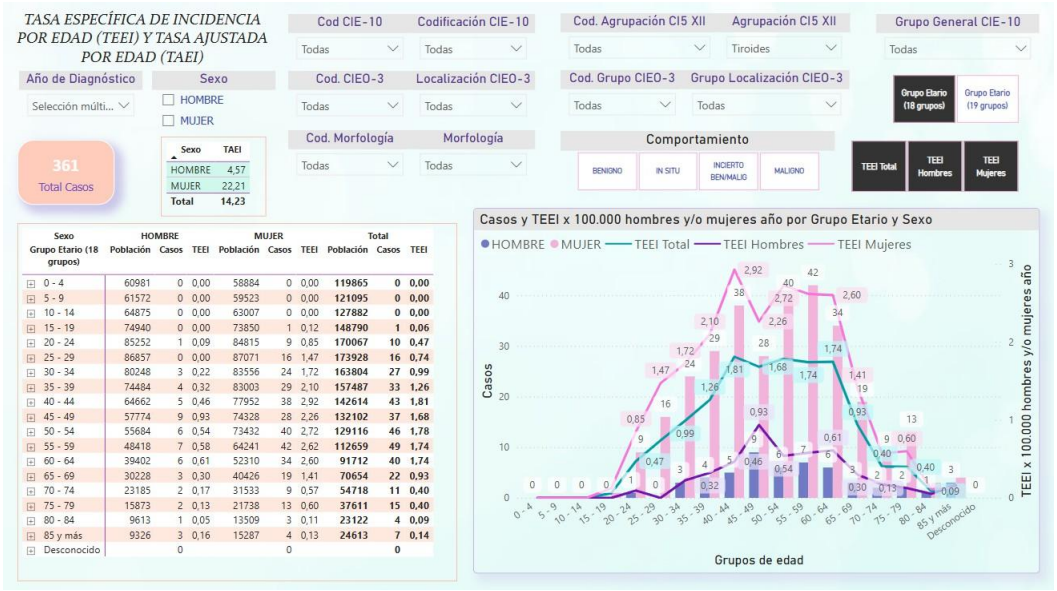
Este segundo informe del cálculo de la Tasa Específica por Edad de Incidencia (TEEI) se crea con el fin de discriminar por sexo el cálculo del indicador en los diferentes grupos etarios. Aunque la tabla presenta el cálculo general, la gráfica

puede segmentarse por sexo y presentar el cálculo totalizado de la TEEI para los filtros que se apliquen por año de diagnóstico, sexo del paciente, comportamiento tumoral y por las características de localización y morfología basadas en las codificaciones CIEO-3 y CIE-10. De esta forma, en un solo gráfico se puede representar por sexo tanto el número total de casos de incidencia de cáncer como el comportamiento de la TEEI en los diferentes grupos etarios mostrando las edades con mayor incidencia del cáncer ajustada a la población de riesgo y la Población Mundial Estándar.



5) TEEI (Extendida) por Grupo Etario – Sexo y TAEI

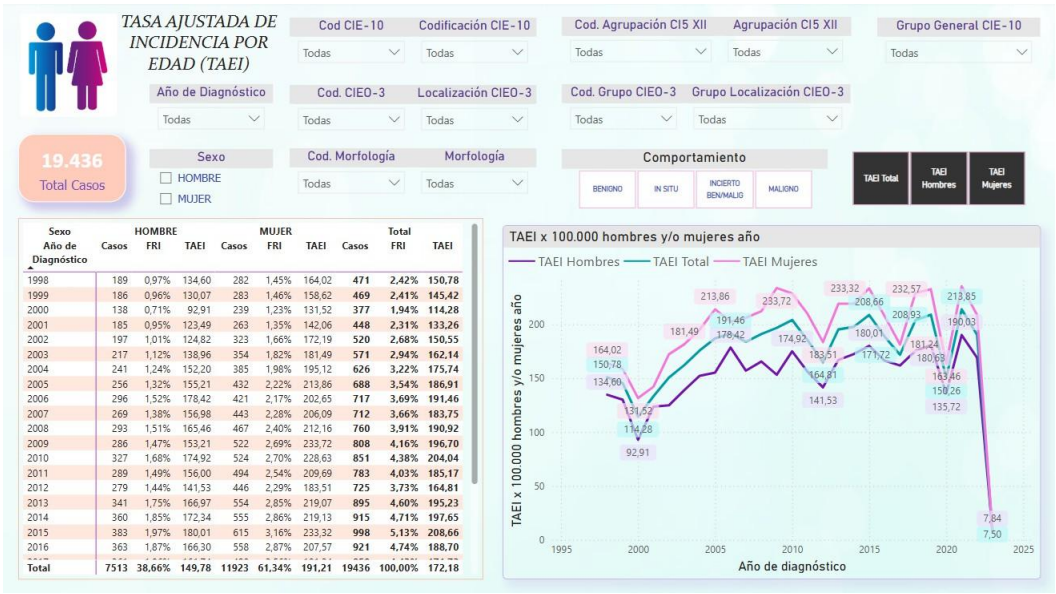
Este informe complementario al anterior, presenta en un mismo dashboard la gráfica y los datos tabulados para el cálculo de la TEEI por grupo etario tanto totalizado como segregado por sexo (de forma extendida). Además, reporta el valor totalizado y por sexo de la Tasa Ajustada por Edad de Incidencia (TAEI), a la que se dedican los siguientes informes, dependiendo de los filtros aplicados.



Sexo / Grupo Etario (18 grupos)	HOMBRE Población	HOMBRE Casos	HOMBRE TEEI	MUJER Población	MUJER Casos	MUJER TEEI	Total Población	Total Casos	Total TEEI
0 - 4	60981	0	0,00	58884	0	0,00	**119865**	**0**	**0,00**
5 - 9	61572	0	0,00	59523	0	0,00	**121095**	**0**	**0,00**
10 - 14	64875	0	0,00	63007	0	0,00	**127882**	**0**	**0,00**
15 - 19	74940	0	0,00	73850	1	0,12	**148790**	**1**	**0,06**
20 - 24	85252	1	0,09	84815	9	0,85	**170067**	**10**	**0,47**
25 - 29	86857	0	0,00	87071	16	1,47	**173928**	**16**	**0,74**
30 - 34	80248	3	0,22	83556	24	1,72	**163804**	**27**	**0,99**
35 - 39	74484	4	0,32	83003	29	2,10	**157487**	**33**	**1,26**
40 - 44	64662	5	0,46	77952	38	2,92	**142614**	**43**	**1,81**
45 - 49	57774	9	0,93	74328	28	2,26	**132102**	**37**	**1,68**
50 - 54	55684	6	0,54	73432	40	2,72	**129116**	**46**	**1,78**
55 - 59	48418	7	0,58	64241	42	2,62	**112659**	**49**	**1,74**
60 - 64	39402	6	0,61	52310	34	2,60	**91712**	**40**	**1,74**
65 - 69	30228	3	0,30	40426	19	1,41	**70654**	**22**	**0,93**
70 - 74	23185	2	0,17	31533	9	0,57	**54718**	**11**	**0,40**
75 - 79	15873	2	0,13	21738	13	0,60	**37611**	**15**	**0,40**
80 - 84	9613	1	0,05	13509	3	0,11	**23122**	**4**	**0,09**
85 y más	9326	3	0,16	15287	4	0,13	**24613**	**7**	**0,14**
Desconocido		0			0			**0**	

6) TAEI por Año – Sexo

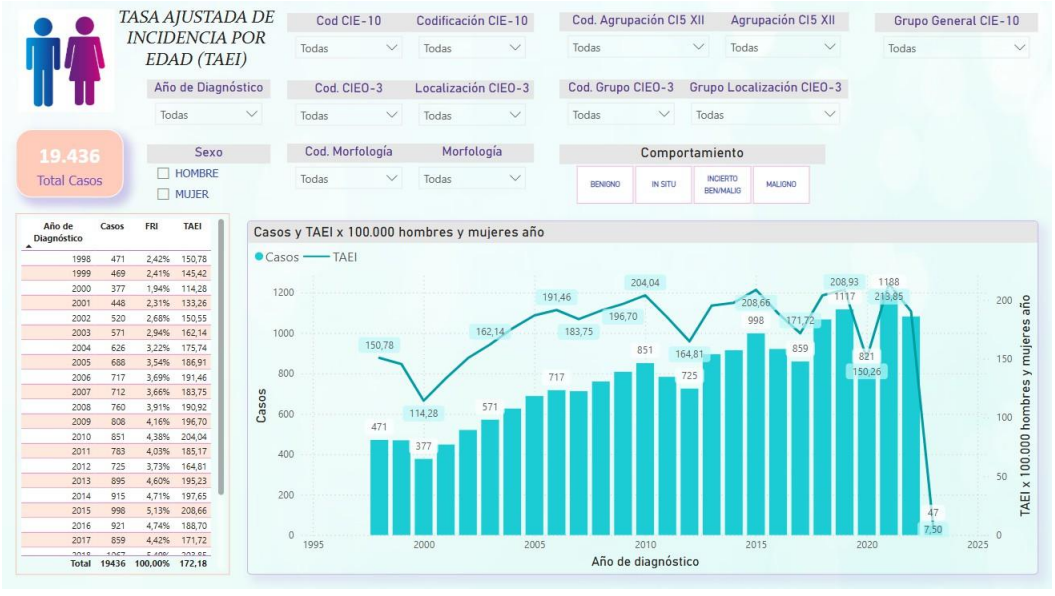
La Tasa Ajustada o Acumulada por Edad de Incidencia (TAEI) es clave en el análisis epidemiológico del cáncer. Se usa especialmente para comparar la frecuencia de casos entre diferentes poblaciones o periodos de tiempo, eliminando el efecto de las diferencias en la estructura por edades. Este indicador permite hacer comparaciones más precisas teniendo en cuenta que el riesgo de padecer cáncer varía significativamente con la edad, por lo que principalmente se emplea para que los resultados no sean distorsionados por el envejecimiento de la población. De esta forma, este dashboard presenta el cálculo de la TAEI x 100.000 hombres y/o mujeres año en el tiempo, de forma totalizada o discriminada por sexo del paciente. Además, la información se organiza junto con la frecuencia relativa de incidencia (FRI) en una tabla por año de diagnóstico y por sexo del paciente.



Sexo / Año de Diagnóstico	HOMBRE Casos	HOMBRE FRI	HOMBRE TAEI	MUJER Casos	MUJER FRI	MUJER TAEI	Total Casos	Total FRI	Total TAEI
1998	189	0,97%	134,60	282	1,45%	164,02	471	2,42%	150,78
1999	186	0,96%	130,07	283	1,46%	158,62	469	2,41%	145,42
2000	138	0,71%	92,91	239	1,23%	131,52	377	1,94%	114,28
2001	185	0,95%	123,49	263	1,35%	142,06	448	2,31%	133,26
2002	197	1,01%	124,82	323	1,66%	172,19	520	2,68%	150,55
2003	217	1,12%	138,96	354	1,82%	181,49	571	2,94%	162,14
2004	241	1,24%	152,20	385	1,98%	195,12	626	3,22%	175,74
2005	256	1,32%	155,21	432	2,22%	213,86	688	3,54%	186,91
2006	296	1,52%	178,42	421	2,17%	202,65	717	3,69%	191,46
2007	269	1,38%	156,98	443	2,28%	206,09	712	3,66%	183,75
2008	293	1,51%	165,46	467	2,40%	212,16	760	3,91%	190,92
2009	286	1,47%	153,21	522	2,69%	233,72	808	4,16%	196,70
2010	327	1,68%	174,92	524	2,70%	228,63	851	4,38%	204,04
2011	289	1,49%	156,00	494	2,54%	209,69	783	4,03%	185,17
2012	279	1,44%	141,53	446	2,29%	183,51	725	3,73%	164,81
2013	341	1,75%	166,97	554	2,85%	219,07	895	4,60%	195,23
2014	360	1,85%	172,34	555	2,86%	219,13	915	4,71%	197,65
2015	383	1,97%	180,01	615	3,16%	233,32	998	5,13%	208,66
2016	363	1,87%	166,30	558	2,87%	207,57	921	4,74%	188,70
Total	7513	38,66%	149,78	11923	61,34%	191,21	19436	100,00%	172,18

7) TAEI por Año – FAI

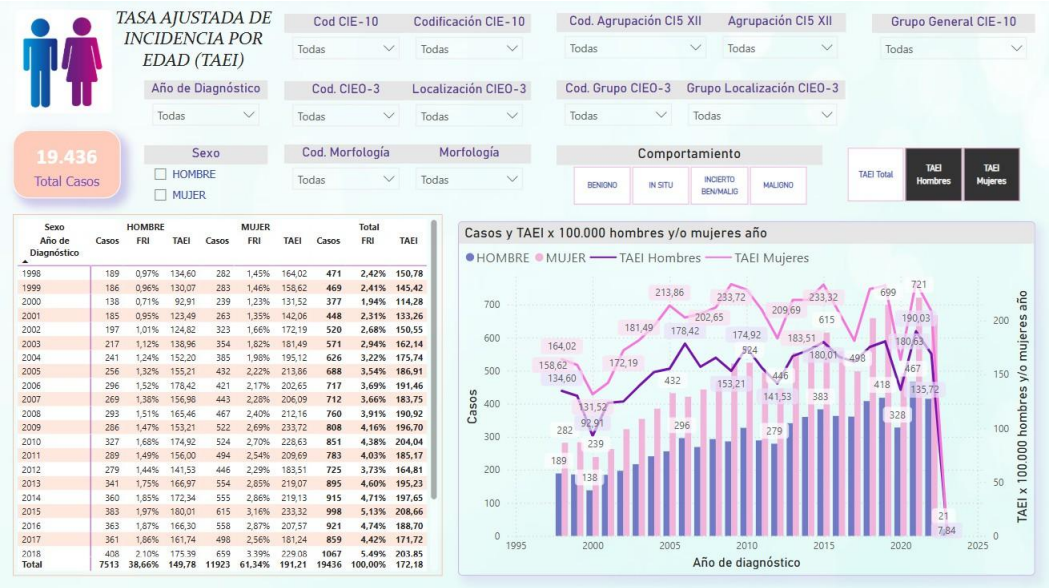
Este dashboard presenta la información tabulada por año de diagnóstico y el gráfico que evidencia en el tiempo, tanto el número total de casos de incidencia de cáncer como el cálculo general de la Tasa Ajustada por Edad de Incidencia (TAEI) x 100.000 hombres y mujeres año para los respectivos filtros aplicados de año de diagnóstico, sexo de los pacientes, comportamiento tumoral y las características de localización y morfología basadas en las codificaciones CIEO-3 y CIE-10.



Año de Diagnóstico	Casos	FRI	TAEI
1998	471	2,42%	150,78
1999	469	2,41%	145,42
2000	377	1,94%	114,28
2001	448	2,31%	133,26
2002	520	2,68%	150,55
2003	571	2,94%	162,14
2004	626	3,22%	175,74
2005	688	3,54%	186,91
2006	717	3,69%	191,46
2007	712	3,66%	183,75
2008	760	3,91%	190,92
2009	808	4,16%	196,70
2010	851	4,38%	204,04
2011	783	4,03%	185,17
2012	725	3,73%	164,81
2013	895	4,60%	195,23
2014	915	4,71%	197,65
2015	998	5,13%	208,66
2016	921	4,74%	188,70
2017	859	4,42%	171,72
Total	19436	100,00%	172,18

8) TAEI por Año – Sexo y FAI

Este informe complementario al anterior, presenta en un mismo dashboard la gráfica y los datos tabulados para el cálculo de la TAEI en el tiempo tanto totalizado como segregado por sexo. Además, brinda la posibilidad de representar en la misma gráfica tanto el número total de casos de incidencia de cáncer como el comportamiento en el tiempo de la TAEI por sexo.

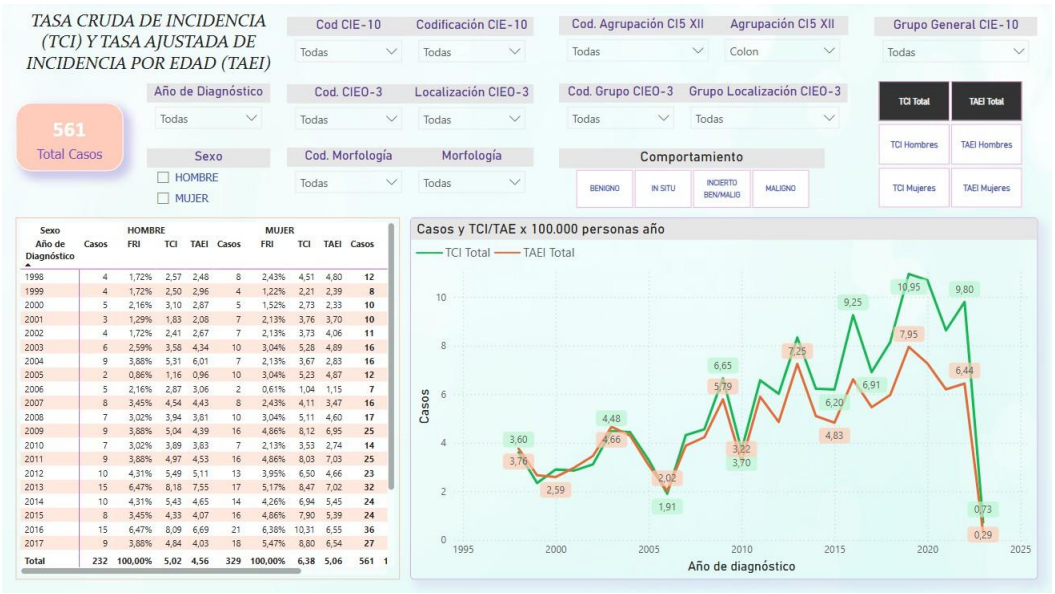


Sexo / Año de Diagnóstico	HOMBRE Casos	HOMBRE FRI	HOMBRE TAEI	MUJER Casos	MUJER FRI	MUJER TAEI	Total Casos	Total FRI	Total TAEI
1998	189	0,97%	134,60	282	1,45%	164,02	471	2,42%	150,78
1999	186	0,96%	130,07	283	1,46%	158,62	469	2,41%	145,42
2000	138	0,71%	92,91	239	1,23%	131,52	377	1,94%	114,28
2001	185	0,95%	123,49	263	1,35%	142,06	448	2,31%	133,26
2002	197	1,01%	124,82	323	1,66%	172,19	520	2,68%	150,55
2003	217	1,12%	138,96	354	1,82%	181,49	571	2,94%	162,14
2004	241	1,24%	152,20	385	1,98%	195,12	626	3,22%	175,74
2005	256	1,32%	155,21	432	2,22%	213,86	688	3,54%	186,91
2006	296	1,52%	178,42	421	2,17%	202,65	717	3,69%	191,46
2007	269	1,38%	156,98	443	2,28%	206,09	712	3,66%	183,75
2008	293	1,51%	165,46	467	2,40%	212,16	760	3,91%	190,92
2009	286	1,47%	153,21	522	2,69%	233,72	808	4,16%	196,70
2010	327	1,68%	174,92	524	2,70%	228,63	851	4,38%	204,04
2011	289	1,49%	156,00	494	2,54%	209,69	783	4,03%	185,17
2012	279	1,44%	141,53	446	2,29%	183,51	725	3,73%	164,81
2013	341	1,75%	166,97	554	2,85%	219,07	895	4,60%	195,23
2014	360	1,85%	172,34	555	2,86%	219,13	915	4,71%	197,65
2015	383	1,97%	180,01	615	3,16%	233,32	998	5,13%	208,66
2016	363	1,87%	166,30	558	2,87%	207,57	921	4,74%	188,70
2017	361	1,86%	161,74	498	2,56%	181,24	859	4,42%	171,72
2018	408	2,10%	175,39	659	3,39%	229,08	1067	5,49%	203,85
Total	7513	38,66%	149,78	11923	61,34%	191,21	19436	100,00%	172,18

9) TCI y TAEI por Año

Finalmente, este informe permite visualizar de forma compilada la comparación en el tiempo entre la TCI y la TAEI x 100.000 hombres y/o mujeres año, puede ser totalizada o discriminada por sexo y presenta la información tabulada por año de diagnóstico y por sexo, para un escenario específico construido aplicando los filtros disponibles por año de diagnóstico, sexo de los pacientes, comportamiento tumoral y por las características de localización y morfología basadas en las codificaciones CIEO-3 y CIE-10. Esta comparación de los dos indicadores busca comprender el impacto de la estructura de edad de la población en los datos de incidencia puesto que si, por ejemplo, la TCI es significativamente mayor que la TAEI podría indicar que la población de estudio tiene una alta proporción de personas mayores, en quienes el cáncer es más frecuente.

Es importante resaltar que la comparación de estas Tasas entre diferentes periodos de tiempo o diferentes poblaciones permite evaluar tendencias de cáncer sin que los resultados se vean sesgados por diferencias demográficas y también contribuye a diseñar estrategias de prevención y control adaptadas a la realidad demográfica de cada población.



Sexo	HOMBRE				MUJER				
Año de Diagnóstico	Casos	FRI	TCI	TAEI	Casos	FRI	TCI	TAEI	Casos
1998	4	1,72%	2,57	2,48	8	2,43%	4,51	4,80	12
1999	4	1,72%	2,50	2,96	4	1,22%	2,21	2,39	8
2000	5	2,16%	3,10	2,87	5	1,52%	2,73	2,33	10
2001	3	1,29%	1,83	2,08	7	2,13%	3,76	3,70	10
2002	4	1,72%	2,41	2,67	7	2,13%	3,73	4,06	11
2003	6	2,59%	3,58	4,34	10	3,04%	5,28	4,89	16
2004	9	3,88%	5,31	6,01	7	2,13%	3,67	2,83	16
2005	2	0,86%	1,16	0,96	10	3,04%	5,23	4,87	12
2006	5	2,16%	2,87	3,06	2	0,61%	1,04	1,15	7
2007	8	3,45%	4,54	4,43	8	2,43%	4,11	3,47	16
2008	7	3,02%	3,94	3,81	10	3,04%	5,11	4,60	17
2009	9	3,88%	5,04	4,39	16	4,86%	8,12	6,95	25
2010	7	3,02%	3,89	3,83	7	2,13%	3,53	2,74	14
2011	9	3,88%	4,97	4,53	16	4,86%	8,03	7,03	25
2012	10	4,31%	5,49	5,11	13	3,95%	6,50	4,66	23
2013	15	6,47%	8,18	7,55	17	5,17%	8,47	7,02	32
2014	10	4,31%	5,43	4,65	14	4,26%	6,94	5,45	24
2015	8	3,45%	4,33	4,07	16	4,86%	7,90	5,39	24
2016	15	6,47%	8,09	6,69	21	6,38%	10,31	6,55	36
2017	9	3,88%	4,84	4,03	18	5,47%	8,80	6,54	27
Total	232	100,00%	5,02	4,56	329	100,00%	6,38	5,06	561

G. POBLACIÓN DANE

El siguiente informe busca representar en un gráfico de pirámide poblacional la proyección de población del municipio de Pasto publicada por el DANE que se emplea en los cálculos de Tasas presentados anteriormente.

1) Población por Grupo Etario y Sexo

El cálculo de las Tasas está asociado a una población de riesgo que en este caso es la población del municipio de Pasto. Los registros de cáncer normalmente tienen que calcular Tasas de incidencia para poblaciones enteras por lo que lo habitual es tomar la población a mediados del año (o la media de las poblaciones a comienzos y a finales del año en cuestión) obtenida de las oficinas del censo. Para este caso y según las indicaciones del RPCMP la población de riesgo se considera como el valor de población proyectado por el DANE para efectos de cálculos de Tasas. En este sentido, este dashboard permite visualizar en un gráfico de pirámide poblacional el porcentaje de la proyección de población del municipio de Pasto (fuente DANE) discriminado por sexo y por grupo etario para un escenario específico en el que se defina el o los años de diagnóstico y el sexo a estudiar.



Año de Diagnóstico	HOMBRE	MUJER	Total
1998	155665	177305	332970
1999	159882	181168	341050
2000	161400	183112	344512
2001	163745	185962	349707
2002	165687	187688	353375
2003	167606	189357	356963
2004	169553	190916	360469
2005	172721	191095	363816
2006	174472	192884	367356
2007	176094	194488	370582
2008	177470	195834	373304
2009	178615	197145	375760
2010	179774	198298	378072
2011	180932	199313	380245
2012	182215	200118	382333
2013	183265	200822	384087
2014	184031	201679	385710
2015	184769	202619	387388
2016	185506	203678	389184
2017	186134	204593	390727
2018	185707	207223	392930
2019	185416	207332	392748
2020	189579	212494	402073
2021	190942	214788	405730
2022	191730	216628	408358
2023	192494	218341	410835
Total	4625404	5154880	9780284

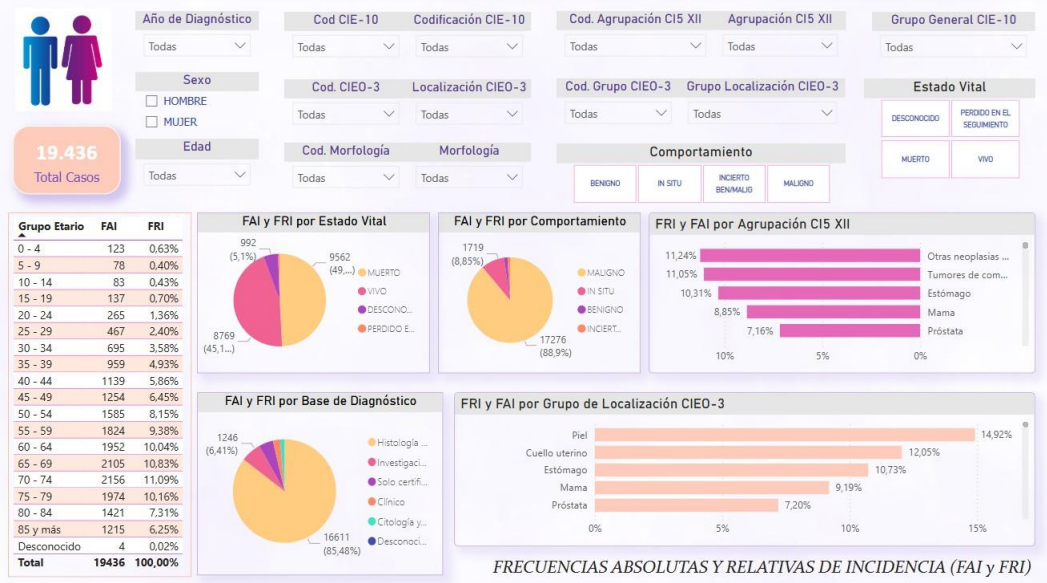
Grupo Etario (18 grupos)	HOMBRE	MUJER	Total
0 - 4	355337	351079	706416
5 - 9	383192	378490	761682
10 - 14	409231	406658	815889
15 - 19	436257	440384	876641
20 - 24	431846	442713	874559
25 - 29	406087	428666	834753
30 - 34	365202	406749	771951
35 - 39	331853	390962	722815
40 - 44	293203	360175	653378
45 - 49	260155	328605	588760
50 - 54	236921	299588	536509
55 - 59	200166	251743	451909
60 - 64	163517	205052	368569
65 - 69	123469	156171	279640
70 - 74	92618	119519	212137
75 - 79	64002	83936	147938
80 - 84	39576	53094	92670
85 y más	32772	51296	84068
Total	4625404	5154880	9780284

H. FRECUENCIAS

El siguiente informe busca calcular las Frecuencias, Absoluta y Relativa expuestas en este trabajo a partir de la cantidad de casos de cáncer del RPCMP, con el fin de generar gráficos y reportes para filtros específicos por diferentes perspectivas.

1) Frecuencia Absoluta y Relativa de Incidencia

El cálculo de las Frecuencias Absoluta (FAI) y Relativa (FRI) de Incidencia permite interpretar la magnitud y distribución de los casos en una población de estudio. La FAI es útil para conocer la carga de la enfermedad en términos numéricos sin considerar el tamaño de la población y la FRI se expresa como una proporción de los casos para hacer comparaciones de la incidencia entre diferentes poblaciones o periodos de tiempo. Así, este dashboard presente una visión compilada del número total de casos de incidencia de cáncer del municipio de Pasto (FAI) y su respectivo porcentaje (FRI) discriminados por estado vital de los pacientes, por la base de diagnóstico de los casos registrados, por el comportamiento tumoral y por localización del tumor en agrupación CI5 XII y grupo de localización CIEO-3. Es posible crear un escenario de estudio filtrando el informe por el año de diagnóstico, sexo, edad y estado vital de los pacientes y por las características de comportamiento y localización del tumor.



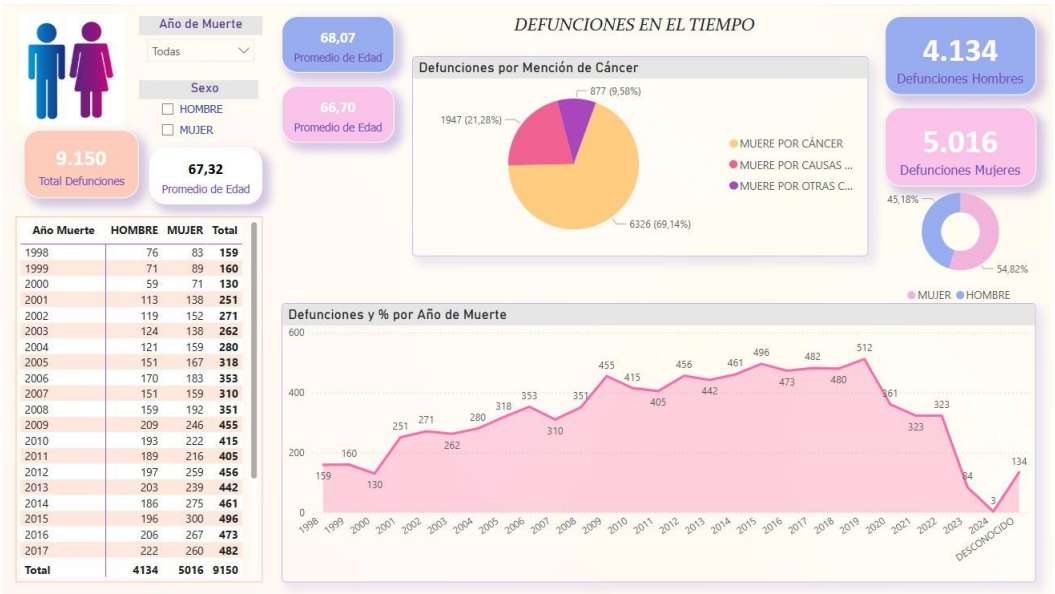
FRECUENCIAS ABSOLUTAS Y RELATIVAS DE INCIDENCIA (FAI y FRI)

I. DEFUNCIONES

En este ítem, se busca contabilizar las defunciones del RPCMP desde diferentes perspectivas teniendo en cuenta que la defunción está asociada al paciente y no al tumor.

1) Defunciones – Año

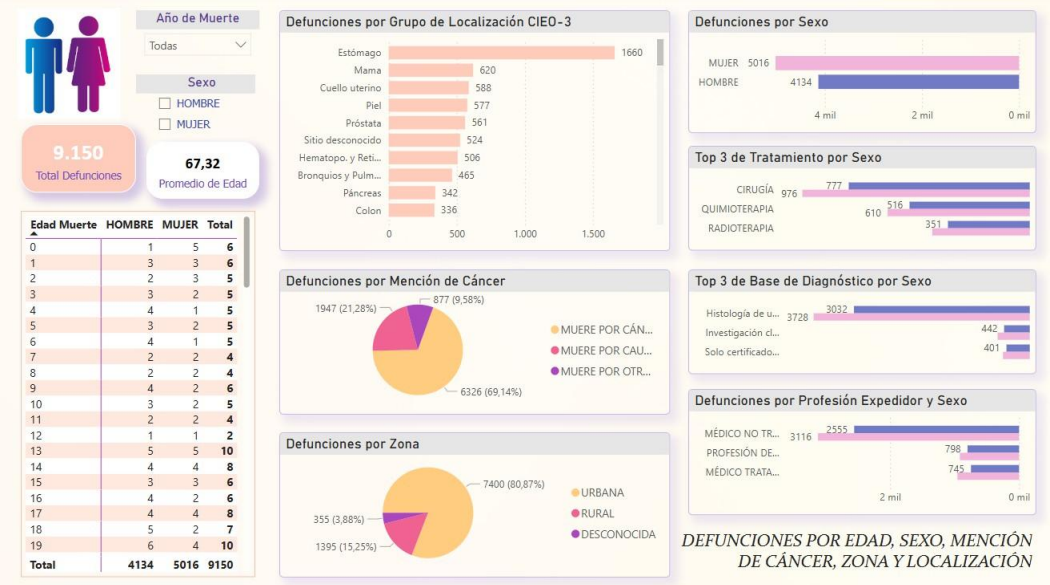
Permite visualizar la evolución del número total de defunciones en el tiempo por fecha de muerte y detalladas por sexo y por la variable mención de cáncer que indica si un paciente murió por cáncer o por otras causas. En cuanto a la fecha de muerte, la información se organiza en diferentes niveles de jerarquía temporal (año, semestre, trimestre, mes y día), proporcionando una visión detallada de la progresión de las defunciones en el municipio de Pasto, de pacientes registrados por el RPCMP.



Año Muerte	HOMBRE	MUJER	Total
1998	76	83	159
1999	71	89	160
2000	59	71	130
2001	113	138	251
2002	119	152	271
2003	124	138	262
2004	121	159	280
2005	151	167	318
2006	170	183	353
2007	151	159	310
2008	159	192	351
2009	209	246	455
2010	193	222	415
2011	189	216	405
2012	197	259	456
2013	203	239	442
2014	186	275	461
2015	196	300	496
2016	206	267	473
2017	222	260	482
Total	4134	5016	9150

2) Defunciones

Aunque el RPCMP realiza un análisis exhaustivo de la mortalidad de pacientes que padecen cáncer en el municipio de Pasto, incluyendo las bases de datos de mortalidad del DANE, el estudio de las defunciones registradas en el RPCMP proporciona una primera visión generalizada sobre el impacto del cáncer en la población del municipio de Pasto. Por esto, el informe presenta de forma compilada el número total de defunciones discriminadas por edad, sexo y zona de residencia del paciente, por la variable mención de cáncer, por grupo de localización CIEO-3 y por agrupación CI5 XII asociadas al tumor que padecen los pacientes que fallecen. También se incluyen datos interesantes sobre los tres tratamientos más frecuentes aplicados a los pacientes que fallecen, las tres bases de diagnóstico más utilizadas en los casos relacionados con esos pacientes y la profesión del expedidor del certificado de defunción. El informe se puede segmentar por año de muerte y por sexo del paciente que fallece. Este enfoque contribuye a generar un panorama amplio y a detalle sobre la incidencia del cáncer en la mortalidad de los pacientes que padecen la enfermedad.



Edad Muerte	HOMBRE	MUJER	Total
0	1	5	6
1	3	3	6
2	2	3	5
3	3	2	5
4	4	1	5
5	3	2	5
6	4	1	5
7	2	2	4
8	2	2	4
9	4	2	6
10	3	2	5
11	2	2	4
12	1	1	2
13	5	5	10
14	4	4	8
15	3	3	6
16	4	2	6
17	4	4	8
18	5	2	7
19	6	4	10
Total	4134	5016	9150

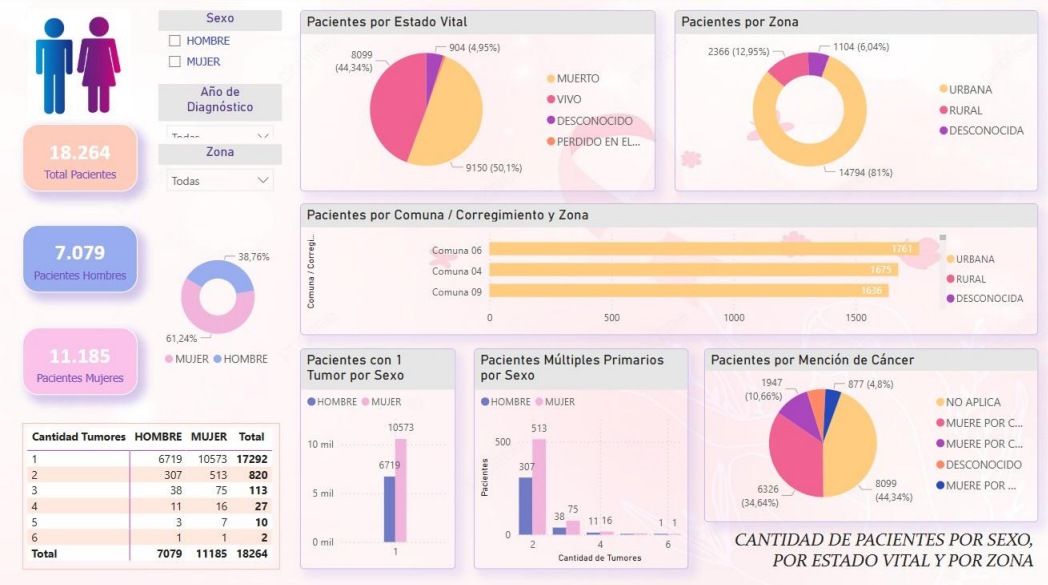
DEFUNCIONES POR EDAD, SEXO, MENCIÓN DE CÁNCER, ZONA Y LOCALIZACIÓN

J. PACIENTES

Aunque la unidad de estudio del RPCMP es el tumor y no el paciente, en ocasiones es necesario conocer el panorama de los pacientes que padecen la enfermedad, por ello se busca contabilizar los pacientes asociados a los casos de cáncer del RPCMP desde diferentes perspectivas.

1) Pacientes

El siguiente dashboard brinda una visión del panorama de los pacientes que padecen cáncer en el municipio de Pasto. Permite visualizar el número total de pacientes y su respectivo porcentaje discriminados por sexo, estado vital, zona de residencia, comuna / corregimiento, barrio / vereda según sea el caso, por la variable mención de cáncer y por la cantidad de tumores que tienen los pacientes, siendo múltiples primarios cuando hay más de un tumor para el mismo paciente. El análisis puede efectuarse sobre el total de la población o sobre una muestra de ella filtrando por sexo de los pacientes, año de diagnóstico de los casos que padecen los pacientes y por la zona de residencia del paciente en el momento del diagnóstico. El paciente también es una unidad importante en el análisis de la incidencia del cáncer puesto que, si se analizan otras variables que impacten el cáncer que padece el paciente, como hábitos de consumo y de vida, estas variables están asociadas directamente con el paciente.



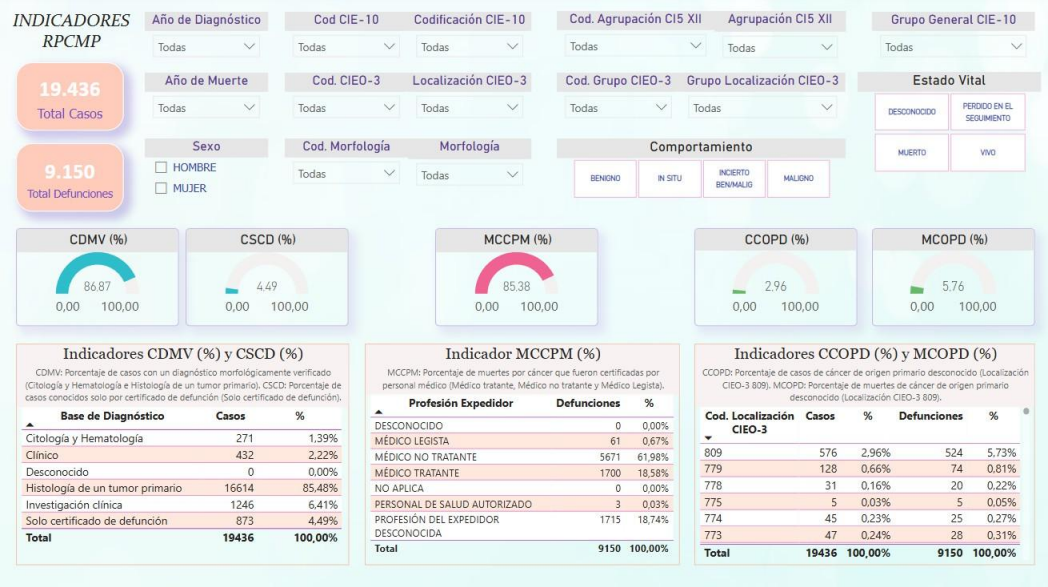
Cantidad Tumores	HOMBRE	MUJER	Total
1	6719	10573	**17292**
2	307	513	**820**
3	38	75	**113**
4	11	16	**27**
5	3	7	**10**
6	1	1	**2**
Total	**7079**	**11185**	**18264**

K. OTROS INDICADORES

El RPCMP cuenta con diferentes indicadores para el análisis de cáncer en el municipio, entre ellos los que se han definido en este trabajo, por ello, los siguientes informes están destinados a calcular estos Indicadores a partir de la cantidad de casos de cáncer del RPCMP con el fin de generar gráficos y reportes para filtros específicos por diferentes perspectivas.

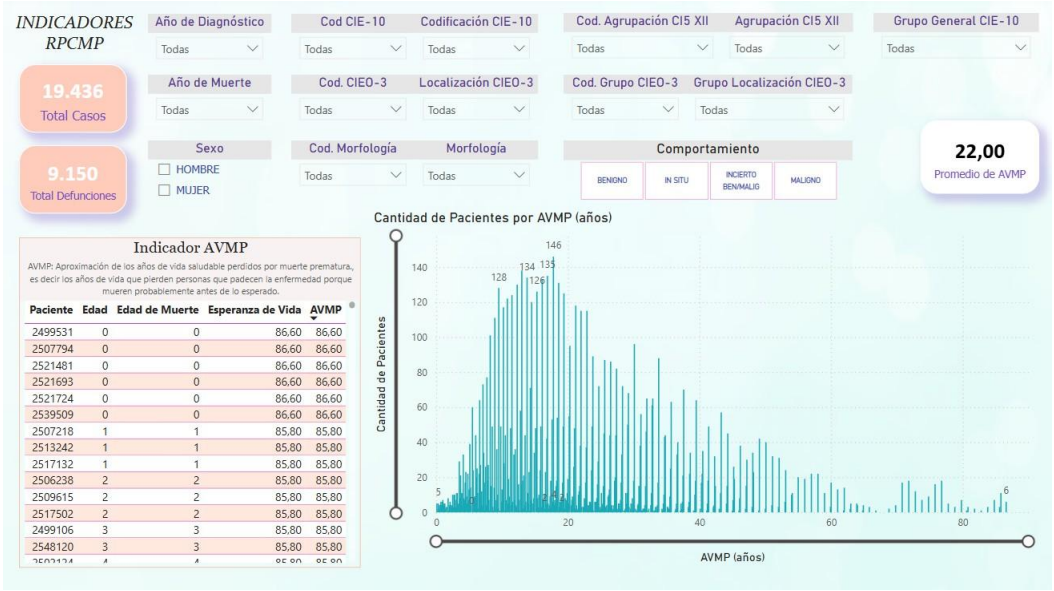
1) Otros Indicadores RPCMP

Este dashboard está destinado a presentar el cálculo de los indicadores de validez del RPCMP definidos en este trabajo. Estos indicadores permiten evaluar la credibilidad, eficacia y confiabilidad en la labor y misión del RPCMP. Por lo anterior, este informe permite visualizar el valor de los indicadores porcentaje de casos con un diagnóstico morfológicamente verificado (CDMV), porcentaje de casos conocidos sólo por certificado de defunción (CSCD), porcentaje de muertes por cáncer que fueron certificadas por personal médico (MCCPM), porcentaje de casos de cáncer de origen primario desconocido (Localización CIEO-3 809, CCOPD) y el porcentaje de muertes de cáncer de origen primario desconocido (Localización CIEO-3 809, MCOPD) en un escenario de estudio construido a partir del año de diagnóstico, año de muerte, sexo y estado vital de los pacientes y por las características de comportamiento, localización y morfología basadas en las codificaciones CIEO-3 y CIE-10.



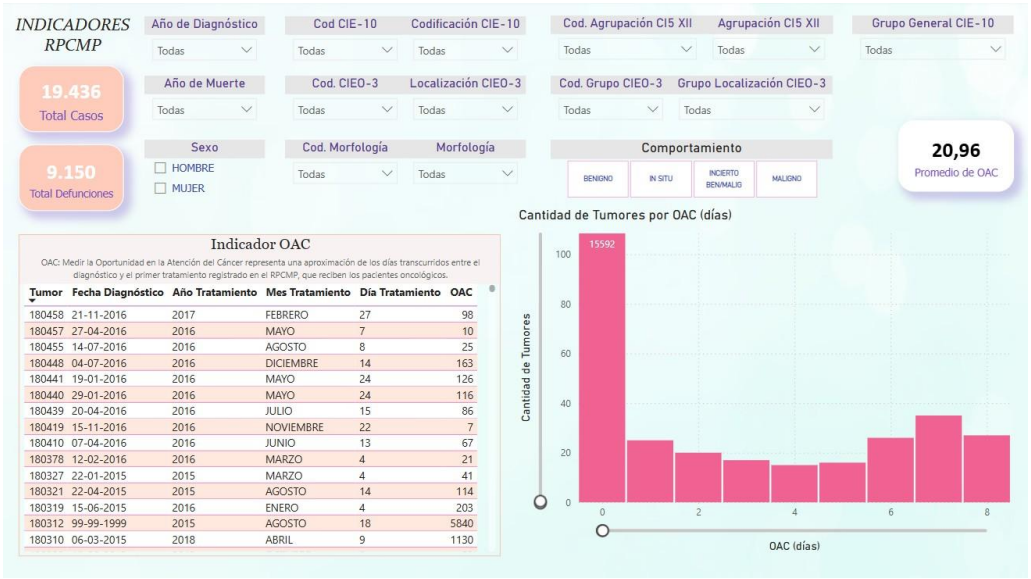
2) Otros Indicadores RPCMP – AVMP

Teniendo en cuenta que para la fecha de culminación de este trabajo, no se ha profundizado en la revisión y cálculo del Indicador de años de vida saludable perdidos por muerte prematura (AVMP), este informe permite visualizar una primera aproximación del valor y el promedio del indicador de AVMP calculado para el estado vital “MUERTO” discriminado por cantidad de pacientes en un escenario de estudio distinguido por el año de diagnóstico de los casos, el sexo y año de muerte de los pacientes asociados y por las características de comportamiento, localización y morfología del tumor basadas en las codificaciones CIEO-3 y CIE-10. Este es un Indicador relevante puesto que la vida de los pacientes cambia drásticamente desde el momento del diagnóstico y hasta la muerte.



3) Otros Indicadores RPCMP – OAC

De igual forma que el dashboard anterior, este permite visualizar una primera aproximación del valor y el promedio del indicador de oportunidad en la atención del cáncer (OAC) calculado para los casos que tienen una fecha de primer tratamiento conocida discriminado por cantidad de tumores en una muestra de estudio identificada por los mismos filtros del informe anterior. Se considera que este Indicador es relevante porque su profundización podría permitir identificar tratamientos efectos e ineficaces aplicados en los diferentes tipos de cáncer y además determinar los tiempos en que se tardan los sistemas y servicios de salud en atender y tratar a una persona que padece la enfermedad desde el momento del diagnóstico.



ANEXOS

Anexo 1. Script empleado en la Extracción de datos como primer paso del proceso ETL.

A continuación, se presenta la sentencia SQL que permite relacionar las tablas de la base de datos *yachay* y consolidar el staging area como base de la implementación del proceso ETL.

```
SELECT tab_fichas.id_paciente AS paciente,
        tab_pacientes.cod_sexo AS cod_sexo,
        tab_sexo.nom_sexo AS sexo,
        tab_pacientes.fenaci_paciente AS fecha_nacimiento,
        tab_fichas.edad_paciente AS edad_paciente,
        tab_pacientes.cod_estadovital AS cod_estado_vital,
        tab_estadovital.nom_estadovital AS estado_vital,
        tab_fichas.id_ficha AS tumor,
        tab_fichas.fe_diagnostico AS fecha_diagnostico,
        tab_fichas.cod_metododiagnostico AS cod_base_diagnostico,
        tab_metododiagnostico.nom_metododiagnostico AS base_diagnostico,
        tab_fichas.cod_localizacion AS cod_localizacion,
        tab_localizacion.nom_localizacion AS localizacion,
        SUBSTRING(tab_localizacion.nom_localizacion FROM 1 FOR POSITION(' ' IN
tab_localizacion.nom_localizacion) - 1) AS cod_cieo3,
        SUBSTRING(tab_localizacion.nom_localizacion FROM 1 FOR POSITION('.' IN
tab_localizacion.nom_localizacion) - 1)::CHAR(4) AS cod_cieo3_grupo,
        REGEXP_REPLACE(SUBSTRING(tab_localizacion.nom_localizacion FROM 1 FOR
POSITION('.' IN tab_localizacion.nom_localizacion) - 1),'[^\d]','','g')::CHAR(4) AS no_cieo3_grupo,
        tab_localizacion_grupo.nom_localizacion AS cieo3_grupo,
        tab_fichas.cod_morfologia AS cod_morfologia,
        tab_morfologias.nom_morfologias AS morfologia,
        tab_fichas.cod_cie10 AS cod_cie10,
        tab_cie10.nom_cie10 AS cie10,
        etl.tab_aux_grupocie10.cod_cie10 AS cod_cie10_grupo,
        etl.tab_aux_grupocie10.nom_cie10_es AS cie10_grupo,
        etl.tab_aux_grupocie10.nom_todos_sitios AS cie10_grupo_general,
        etl.tab_aux_grupocie10.nom_excepto_C44 AS cie10_grupo_excepto_c44,
        etl.tab_aux_grupocie10.cod_grupo_cie10 AS cod_grupo_cie10,
        etl.tab_aux_grupocie10.nom_grupo_cie10_es AS grupo_cie10,
        tab_fichas.cod_lateralidad AS cod_lateralidad,
        tab_lateralidad.nom_lateralidad AS lateralidad,
        tab_fichas.cod_comportamiento AS cod_comportamiento,
        tab_comportamiento.nom_comportamiento AS comportamiento,
        tab_fichas.cod_gradodiferenciacion AS cod_grado_diferenciacion,
        tab_gradodiferenciacion.nom_gradodiferenciacion AS grado_diferenciacion,
        tab_certidefuncion.cod_mencioncancer AS cod_mencion_cancer,
        tab_mencioncancerdefuncion.nom_mencioncancerdefuncion AS mencion_cancer,
        tab_certidefuncion.cod_profesionexpedidor AS cod_profesion_expedidor,
        tab_profesionexpedidor.nom_profesionexpedidor AS profesion_expedidor,
        tab_pacientes.fecha_ultimocontacto AS fecha_ultimo_contacto,
        subtab_historias.cod_fuentes AS cod_fuente,
        tab_fuentes.nom_fuentes AS fuente,
        subtab_tratamientosficha.cod_tratamientos AS cod_tratamiento,
        tab_tratamientos.nom_tratamientos AS tratamiento,
```

```
        subtab_tratamientosficha.cod_estadoenfermedad AS cod_estado_enfermedad,
        tab_estadoenfermedad.nom_estadoenfermedad AS estado_enfermedad,
        subtab_tratamientosficha.fe_tratamiento AS fecha_tratamiento,
        subtab_tratamientosficha.orden_tratamiento AS orden_tratamiento,
        tab_municipios.cod_departamentos AS cod_departamento,
        tab_departamentos.nom_departamentos AS departamento,
        tab_municipios.cod_municipios AS cod_municipio,
        tab_municipios.nom_municipios AS municipio,
        tab_barrios.cod_zona AS cod_zona,
        tab_zona.nom_zona AS zona,
        subtab_direccionTumor.barriocomuna AS cod_barrio_comuna,
        tab_barrios.nom_barrios AS barrio_comuna,
        tab_barrios.cod_comunas AS cod_comuna_corregimiento,
        tab_barrios.cod_barriovereda AS cod_barrio_vereda
INTO etl.tumores_pasto
FROM tab_fichas
INNER JOIN tab_pacientes ON tab_fichas.id_paciente = tab_pacientes.id_paciente
LEFT JOIN tab_sexo USING (cod_sexo)
LEFT JOIN tab_estadovital USING (cod_estadovital)
LEFT JOIN tab_metododiagnostico USING (cod_metododiagnostico)
LEFT JOIN tab_localizacion USING (cod_localizacion)
LEFT JOIN tab_localizacion tab_localizacion_grupo ON
(REGEXP_REPLACE(SUBSTRING(tab_localizacion.nom_localizacion FROM 1 FOR POSITION('.' IN
tab_localizacion.nom_localizacion) - 1),'[^\d]','','g')::CHAR(4)) = tab_localizacion_grupo.cod_localizacion
LEFT JOIN tab_morfologias ON tab_fichas.cod_morfologia = tab_morfologias.cod_morfologias
LEFT JOIN tab_cie10 USING (cod_cie10)
LEFT JOIN etl.tab_aux_grupocie10 ON
SUBSTRING(tab_fichas.cod_cie10 FROM 1 FOR 3) = etl.tab_aux_grupocie10.cod_cie10
LEFT JOIN tab_lateralidad USING (cod_lateralidad)
LEFT JOIN tab_comportamiento USING (cod_comportamiento)
LEFT JOIN tab_gradodiferenciacion USING (cod_gradodiferenciacion)
LEFT JOIN tab_certidefuncion ON tab_fichas.id_paciente = tab_certidefuncion.id_paciente
LEFT JOIN tab_mencioncancerdefuncion ON tab_certidefuncion.cod_mencioncancer =
tab_mencioncancerdefuncion.cod_mencioncancerdefuncion
LEFT JOIN tab_profesionexpedidor USING (cod_profesionexpedidor)
LEFT JOIN
(SELECT DISTINCT id_ficha,cod_fuentes FROM tab_historias ORDER BY id_ficha) subtab_historias
USING (id_ficha)
LEFT JOIN tab_fuentes USING (cod_fuentes)
LEFT JOIN
(SELECT DISTINCT ON (id_ficha) *
FROM tab_tratamientosficha
ORDER BY id_ficha, orden_tratamiento ASC) subtab_tratamientosficha
USING (id_ficha)
LEFT JOIN tab_tratamientos USING (cod_tratamientos)
LEFT JOIN tab_estadoenfermedad USING (cod_estadoenfermedad)
LEFT JOIN
(SELECT * FROM tab_direccionTumor
WHERE id_ficha BETWEEN 1 AND
(SELECT MAX(id_ficha) FROM tab_fichas) AND orden_registro='1') subtab_direccionTumor
USING (id_ficha)
LEFT JOIN tab_barrios ON subtab_direccionTumor.barriocomuna = tab_barrios.cod_barrios
LEFT JOIN tab_municipios USING (cod_municipios)
LEFT JOIN tab_departamentos USING (cod_departamentos)
LEFT JOIN tab_zona USING (cod_zona)
WHERE tab_fichas.id_ficha BETWEEN 1 AND (SELECT MAX(id_ficha) FROM tab_fichas)
```

```
AND (tab_municipios.cod_departamentos = '52' AND tab_barrios.cod_municipios = '52001')
OR subtab_direcciontumor.barriocomuna = '999999'
ORDER BY tumor,paciente;
```

Anexo 2. Script empleado en el análisis de calidad de los datos dentro del proceso ETL.

A continuación, se presenta el código del ejecutable de Python (.py) que permite analizar la calidad de los datos mediante la impresión por cada campo, de la cantidad de registros nulos, no nulos y valores diferentes con el fin de conocer los atributos a trabajar, su comportamiento y los requerimientos a tener en cuenta para el proceso de Limpieza de los datos.

```
#!/usr/bin/env python
# coding: utf-8

# In[1]:
import pandas as pd
from sqlalchemy import create_engine

# In[2]:
def connect_with_db():
    #Conexión con la base de datos
    # Parámetros de conexión
    user = 'postgres'
    password = 'postgresLore'
    host = 'localhost'
    port = '5445'
    db_name = 'yachay_local2'
    connection = f'postgresql://{user}:{password}@{host}:{port}/{db_name}'
    engine = create_engine(connection)
    return engine

# In[3]:
def read_table(relation,key):
    engine = connect_with_db()
    query = f"""SELECT * FROM {relation} ORDER BY {key};"""
    df = pd.read_sql(query,engine)
    return df

# In[4]:
def data_quality_analysis(df,relation):
    data_quality = []
    engine = connect_with_db()
    for field in df.columns:
        #query = f"""SELECT COUNT(*) AS records,COUNT({field}) AS non_null,COUNT(*) -
COUNT({field}) AS null,COUNT(DISTINCT{field}) AS different_values
        #FROM tab_aux_extraccion;"""
        #result = pd.read_sql(query,engine)

#data_quality.append((field,result["records"][0],result["null"][0],result["non_null"][0],result["different_val
ues"][0]))
        query = f"""SELECT COUNT(*) AS records FROM {relation};"""
        total_records = pd.read_sql(query,engine)
        query = f"""SELECT COUNT(*) AS null FROM {relation} WHERE {field} IS NULL;"""
```

```
        total_null_records = pd.read_sql(query,engine)
        query = f"""SELECT COUNT(*) AS non_null FROM {relation} WHERE {field} IS NOT NULL;"""
        total_non_null_records = pd.read_sql(query,engine)
        query = f"""SELECT COUNT(DISTINCT {field}) AS values FROM {relation};"""
        different_values = pd.read_sql(query,engine)
        query = f"""SELECT {field} AS values FROM {relation} GROUP BY 1 ORDER BY 1 LIMIT 10;"""
        values = pd.read_sql(query,engine)
        values = values["values"].tolist()
        values = ', '.join(map(str,values))

data_quality.append((field,total_records["records"][0],total_null_records["null"][0],total_non_null_records
["non_null"][0],different_values["values"][0],values))
    results = pd.DataFrame(data_quality,columns=["Campo","Total Registros","Registros Nulos","Registros
No Nulos","Total Valores Diferentes","Valores Diferentes"])
    return results

# In[5]:
#Leer una tabla de la base de datos (table) y ordenarla por un campo en específico (key)
#relation = "public.tab_aux_extraccion_pasto"
relation = "etl.tumores_pasto"
key = "paciente"
tumores_pasto = read_table(relation,key);
tumores_pasto.head()

# In[6]:
tumores_pasto.shape

# In[7]:
data_quality = data_quality_analysis(tumores_pasto,relation)
data_quality.head()

# In[8]:
data_quality.tail()

# In[9]:
data_quality.to_excel('data_quality.xlsx',index=False)

# In[ ]:
```

Anexo 3. Script empleado en la Limpieza de datos dentro del proceso ETL.

A continuación, se presenta el código SQL que permite hacer la Limpieza de los datos mediante la actualización de algunos valores, el reemplazo de nulos, la eliminación de espacios en blanco de relleno y la estandarización de algunos atributos para preparar la staging area para el proceso de Transformación de los datos.

```
--BODEGA DE DATOS SISTEMA YACHAY - RPCMP
--REGISTRO POBLACIONAL DE CANCER DEL MUNICIPIO DE PASTO
--
========================================================================
--LIMPIEZA DE DATOS A PARTIR DE LA TABLA DE EXTRACCION DE LOS TUMORES-FUENTE
DEL MUNICIPIO DE PASTO
```

```
--Crear la tabla tumores_pasto_limpio a partir de la tabla tumores_pasto.
BEGIN;
CREATE TABLE etl.tumores_pasto_limpio AS SELECT * FROM etl.tumores_pasto;
COMMIT;
--SELECT 41345
--Query returned successfully in 247 msec.
--COMMIT
--Query returned successfully in 671 msec.
-------------------------------------------------------------------------------------------------------------------------------------------
--1. Campo sexo
--Reemplazar los valores Masculino por HOMBRE y Femenino por MUJER segun nomenclatura del RPCMP
BEGIN;
UPDATE etl.tumores_pasto_limpio SET sexo = 'HOMBRE' WHERE sexo = 'Masculino';
COMMIT;
--UPDATE 15755
--Query returned successfully in 486 msec.
--COMMIT
--Query returned successfully in 154 msec.
BEGIN;
UPDATE etl.tumores_pasto_limpio SET sexo = 'MUJER' WHERE sexo = 'Femenino';
COMMIT;
--UPDATE 25684
--Query returned successfully in 493 msec.
--COMMIT
--Query returned successfully in 156 msec.
--Eliminar espacios en blanco de relleno
BEGIN;
UPDATE etl.tumores_pasto_limpio SET sexo = TRIM(BOTH FROM sexo);
COMMIT;
--UPDATE 41345
--Query returned successfully in 901 msec.
--COMMIT
--Query returned successfully in 731 msec.
--Estandarizar el atributo a mayusculas
BEGIN;
UPDATE etl.tumores_pasto_limpio SET sexo = UPPER(sexo);
COMMIT;
--UPDATE 41345
--Query returned successfully in 542 msec.
--COMMIT
--Query returned successfully in 768 msec.
-------------------------------------------------------------------------------------------------------------------------------------------
--2. Campo fecha_nacimiento
--Reemplazar las fechas nulas por fechas desconocidas bajo el formato 99-99-9999
BEGIN;
UPDATE etl.tumores_pasto_limpio SET fecha_nacimiento = '99-99-9999' WHERE fecha_nacimiento IS NULL;
COMMIT;
--UPDATE 161
--Query returned successfully in 91 msec.
--COMMIT
--Query returned successfully in 133 msec.
-------------------------------------------------------------------------------------------------------------------------------------------
```

```
--3. Campo cod_estado_vital
--Reemplazar los codigos nulos por codigo 9 asignado a Desconocido
BEGIN;
UPDATE etl.tumores_pasto_limpio SET cod_estado_vital = '9' WHERE cod_estado_vital IS NULL;
COMMIT;
--UPDATE 39
--Query returned successfully in 173 msec.
--COMMIT
--Query returned successfully in 117 msec.
-----------------------------------------------------------------------------------------------------------------------------------------------------------
--4. Campo estado_vital
--Reemplazar los estados vital nulos en DESCONOCIDO
BEGIN;
UPDATE etl.tumores_pasto_limpio SET estado_vital = 'DESCONOCIDO' WHERE estado_vital IS NULL;
COMMIT;
--UPDATE 39
--Query returned successfully in 161 msec.
--COMMIT
--Query returned successfully in 116 msec.
--Estandarizar el atributo a mayusculas
BEGIN;
UPDATE etl.tumores_pasto_limpio SET estado_vital = UPPER(estado_vital);
COMMIT;
--UPDATE 41345
--Query returned successfully in 384 msec.
--COMMIT
--Query returned successfully in 716 msec.
-----------------------------------------------------------------------------------------------------------------------------------------------------------
--5. Campo cod_localizacion
--Eliminar espacios en blanco de relleno
BEGIN;
UPDATE etl.tumores_pasto_limpio SET cod_localizacion = TRIM(BOTH FROM cod_localizacion);
COMMIT;
--UPDATE 41345
--Query returned successfully in 1 secs 23 msec.
--COMMIT
--Query returned successfully in 758 msec.
-----------------------------------------------------------------------------------------------------------------------------------------------------------
--6. Campo cod_cieo3_grupo
--Eliminar espacios en blanco de relleno
BEGIN;
UPDATE etl.tumores_pasto_limpio SET cod_cieo3_grupo = TRIM(BOTH FROM cod_cieo3_grupo);
COMMIT;
--UPDATE 41345
--Query returned successfully in 764 msec.
--COMMIT
--Query returned successfully in 882 msec.
-----------------------------------------------------------------------------------------------------------------------------------------------------------
--7. Campo no_cieo3_grupo
--Eliminar espacios en blanco de relleno
BEGIN;
UPDATE etl.tumores_pasto_limpio SET no_cieo3_grupo = TRIM(BOTH FROM no_cieo3_grupo);
COMMIT;
--UPDATE 41345
--Query returned successfully in 872 msec.
--COMMIT
```

```
--Query returned successfully in 670 msec.
-----------------------------------------------------------------------------------------------------------------------------------------------------------
--8. Campo cod_cie10
--Eliminar X al final de codigos que en realidad son de 3 caracteres y eliminar espacios en blanco de relleno
BEGIN;
UPDATE etl.tumores_pasto_limpio SET cod_cie10 = CASE
WHEN RIGHT(TRIM(BOTH FROM cod_cie10),1) = 'X' THEN LEFT(cod_cie10,LENGTH(cod_cie10) -
1)
ELSE TRIM(BOTH FROM cod_cie10)
END;
COMMIT;
--UPDATE 41345
--Query returned successfully in 306 msec.
--COMMIT
--Query returned successfully in 771 msec.
-----------------------------------------------------------------------------------------------------------------------------------------------------------
--9. Campo cod_cie10_grupo
--Reemplazar los codigos nulos por codigo 000 asignado a 'Tumores de comportamiento Benigno, Incierto e
In situ'
BEGIN;
UPDATE etl.tumores_pasto_limpio SET cod_cie10_grupo = '000' WHERE cod_cie10_grupo IS NULL;
COMMIT;
--UPDATE 4254
--Query returned successfully in 453 msec.
--COMMIT
--Query returned successfully in 299 msec.
--Reemplazar los codigos C44 por el codigo EXC asignado a los casos con codigos de morfologia 8070 a
8078 y 8090 a 8098
--de cancer de piel excluidos de reportes y conteos
BEGIN;
UPDATE etl.tumores_pasto_limpio SET cod_cie10_grupo = 'EXC' WHERE cod_cie10_grupo = 'C44' AND
(cod_morfologia BETWEEN '8070' AND '8078' OR cod_morfologia BETWEEN '8090' AND '8098');
COMMIT;
--UPDATE 4250
--Query returned successfully in 858 msec.
--COMMIT
--Query returned successfully in 389 msec.
--Eliminar espacios en blanco de relleno
BEGIN;
UPDATE etl.tumores_pasto_limpio SET cod_cie10_grupo = TRIM(BOTH FROM cod_cie10_grupo);
COMMIT;
--UPDATE 41345
--Query returned successfully in 939 msec.
--COMMIT
--Query returned successfully in 862 msec.
-----------------------------------------------------------------------------------------------------------------------------------------------------------
--10. Campo cie10_grupo
--Reemplazar los valores nulos por 'Tumores de comportamiento Benigno, Incierto e In situ' asociado al
codigo 000 de
--cod_cie10_grupo
BEGIN;
UPDATE etl.tumores_pasto_limpio SET cie10_grupo = 'Tumores de comportamiento Benigno, Incierto e In
situ'
WHERE cie10_grupo IS NULL;
COMMIT;
--UPDATE 4254
```

--Query returned successfully in 328 msec.
--COMMIT
--Query returned successfully in 322 msec.
--Reemplazar los valores 'Otras neoplasias malignas de la piel' por 'Otras neoplasias malignas de la piel que agrupan
--los carcinomas de células basales y escamosas de la piel' asociado al codigo EXC de cod_cie10_grupo
BEGIN;
UPDATE etl.tumores_pasto_limpio SET
cie10_grupo = 'Otras neoplasias malignas de la piel que agrupan los carcinomas de células basales y escamosas de la piel'
WHERE cod_cie10_grupo = 'EXC';
COMMIT;
--UPDATE 4250
--Query returned successfully in 460 msec.
--COMMIT
--Query returned successfully in 481 msec.

--11. Campo cie10_grupo_general
--Reemplazar los valores nulos por 'Tumores de comportamiento Benigno, Incierto e In situ' asociado al codigo 000 de
--cod_cie10_grupo
BEGIN;
UPDATE etl.tumores_pasto_limpio SET cie10_grupo_general = 'Tumores de comportamiento Benigno, Incierto e In situ'
WHERE cie10_grupo_general IS NULL;
COMMIT;
--UPDATE 4254
--Query returned successfully in 410 msec.
--COMMIT
--Query returned successfully in 300 msec.
--Reemplazar los valores 'Todos los sitios' por 'Otras neoplasias malignas de la piel que agrupan los carcinomas de
--células basales y escamosas de la piel' asociado al codigo EXC de cod_cie10_grupo
BEGIN;
UPDATE etl.tumores_pasto_limpio SET cie10_grupo_general = 'Otras neoplasias malignas de la piel que agrupan los carcinomas de células basales y escamosas de la piel'
WHERE cod_cie10_grupo = 'EXC';
COMMIT;
--UPDATE 4250
--Query returned successfully in 403 msec.
--COMMIT
--Query returned successfully in 403 msec.

--12. Campo cie10_grupo_excepto_c44
--Reemplazar los valores nulos por 'Tumores de comportamiento Benigno, Incierto e In situ' asociado al codigo 000 de
--cod_cie10_grupo
BEGIN;
UPDATE etl.tumores_pasto_limpio SET cie10_grupo_excepto_c44 = 'Tumores de comportamiento Benigno, Incierto e In situ'
WHERE cie10_grupo_excepto_c44 IS NULL;
COMMIT;
--UPDATE 4254
--Query returned successfully in 350 msec.
--COMMIT
--Query returned successfully in 285 msec.

```
--------------------------------------------------------------------------------------------------------------------------------------------------------
--13. Campo cod_grupo_cie10
--Reemplazar los valores nulos por 0 asociado al codigo 000 de cod_cie10_grupo
BEGIN;
UPDATE etl.tumores_pasto_limpio SET cod_grupo_cie10 = 0 WHERE cod_grupo_cie10 IS NULL;
COMMIT;
--UPDATE 4254
--Query returned successfully in 567 msec.
--COMMIT
--Query returned successfully in 547 msec.
--Reemplazar los valores 26 por el maximo de este codigo + 1 creando un nuevo grupo para las morfologias excluidas de
--cancer de piel asociados al codigo EXC de cod_cie10_grupo
BEGIN;
UPDATE etl.tumores_pasto_limpio SET cod_grupo_cie10 = (
SELECT MAX(cod_grupo_cie10) + 1 FROM etl.tumores_pasto_limpio
) WHERE cod_cie10_grupo = 'EXC';
COMMIT;
--UPDATE 4250
--Query returned successfully in 750 msec.
--COMMIT
--Query returned successfully in 964 msec.
--------------------------------------------------------------------------------------------------------------------------------------------------------
--14. Campo grupo_cie10
--Reemplazar los valores nulos por 'Tumores de comportamiento Benigno, Incierto e In situ' asociado  al codigo 000 de
--cod_cie10_grupo
BEGIN;
UPDATE etl.tumores_pasto_limpio SET grupo_cie10 = 'Tumores de comportamiento Benigno, Incierto e In situ'
WHERE grupo_cie10 IS NULL;
COMMIT;
--UPDATE 4254
--Query returned successfully in 442 msec.
--COMMIT
--Query returned successfully in 670 msec.
--Reemplazar los valores 'Otras neoplasias de piel' por 'Otras neoplasias malignas de la piel que agrupan los carcinomas de células
--basales y escamosas de la piel' asociado al codigo EXC de cod_cie10_grupo
BEGIN;
UPDATE etl.tumores_pasto_limpio
SET grupo_cie10 = 'Otras neoplasias malignas de la piel que agrupan los carcinomas de células basales y escamosas de la piel'
WHERE cod_cie10_grupo = 'EXC';
COMMIT;
--UPDATE 4250
--Query returned successfully in 387 msec.
--COMMIT
--Query returned successfully in 372 msec.
--------------------------------------------------------------------------------------------------------------------------------------------------------
--15. Campo cod_lateralidad
--A partir de la tabla de lateralidad compartida por el RPCMP que identifica los codigos de cancer que deben tener
--registrado una lateralidad, se reemplaza los codigos de lateralidad nulos con el valor 4 que es No corresponde
--para aquellos registros con codigos cie10 que no aparecen en la tabla
```

```
BEGIN;
UPDATE etl.tumores_pasto_limpio SET cod_lateralidad = 4 WHERE cod_lateralidad IS NULL
AND cod_cie10 NOT IN (
SELECT DISTINCT(cod_cie10) FROM etl.tumores_pasto
WHERE cod_cie10 BETWEEN 'C079' AND 'C081'
OR cod_cie10 BETWEEN 'C090' AND 'C091'
OR cod_cie10 BETWEEN 'C098' AND 'C099'
OR cod_cie10 BETWEEN 'C300' AND 'C301'
OR cod_cie10 BETWEEN 'C340' AND 'C349'
OR cod_cie10 BETWEEN 'C400' AND 'C403'
OR cod_cie10 BETWEEN 'C413' AND 'C414'
OR cod_cie10 BETWEEN 'C441' AND 'C443'
OR cod_cie10 BETWEEN 'C445' AND 'C447'
OR cod_cie10 BETWEEN 'C471' AND 'C472'
OR cod_cie10 BETWEEN 'C491' AND 'C492'
OR cod_cie10 BETWEEN 'C500' AND 'C509'
OR cod_cie10 BETWEEN 'C569' AND 'C570'
OR cod_cie10 BETWEEN 'C620' AND 'C631'
OR cod_cie10 BETWEEN 'C690' AND 'C699'
OR cod_cie10 BETWEEN 'C710' AND 'C714'
OR cod_cie10 BETWEEN 'C722' AND 'C725'
OR cod_cie10 BETWEEN 'C740' AND 'C749'
OR cod_cie10 IN
('C310','C312','C384','C649','C659','C669','C700','C754')
GROUP BY cod_cie10
);
COMMIT;
--UPDATE 10998
--Query returned successfully in 571 msec.
--COMMIT
--Query returned successfully in 1 secs 3 msec.
--Reemplazar los codigos de lateralidad nulos con el valor de 9 que es Desconocido para los codigos cie10
que
--corresponden con los listados en la tabla de lateralidad
BEGIN;
UPDATE etl.tumores_pasto_limpio SET cod_lateralidad = 9 WHERE cod_lateralidad IS NULL
AND cod_cie10 IN (
SELECT DISTINCT(cod_cie10) FROM etl.tumores_pasto
WHERE cod_cie10 BETWEEN 'C079' AND 'C081'
OR cod_cie10 BETWEEN 'C090' AND 'C091'
OR cod_cie10 BETWEEN 'C098' AND 'C099'
OR cod_cie10 BETWEEN 'C300' AND 'C301'
OR cod_cie10 BETWEEN 'C340' AND 'C349'
OR cod_cie10 BETWEEN 'C400' AND 'C403'
OR cod_cie10 BETWEEN 'C413' AND 'C414'
OR cod_cie10 BETWEEN 'C441' AND 'C443'
OR cod_cie10 BETWEEN 'C445' AND 'C447'
OR cod_cie10 BETWEEN 'C471' AND 'C472'
OR cod_cie10 BETWEEN 'C491' AND 'C492'
OR cod_cie10 BETWEEN 'C500' AND 'C509'
OR cod_cie10 BETWEEN 'C569' AND 'C570'
OR cod_cie10 BETWEEN 'C620' AND 'C631'
OR cod_cie10 BETWEEN 'C690' AND 'C699'
OR cod_cie10 BETWEEN 'C710' AND 'C714'
OR cod_cie10 BETWEEN 'C722' AND 'C725'
OR cod_cie10 BETWEEN 'C740' AND 'C749'
```

```
OR cod_cie10 IN
('C310','C312','C384','C649','C659','C669','C700','C754')
GROUP BY cod_cie10
);
COMMIT;
--UPDATE 2479
--Query returned successfully in 598 msec.
--COMMIT
--Query returned successfully in 808 msec.
-------------------------------------------------------------------------------------------------------------------------------------------
--16. Campo lateralidad
--Acorde a la actualizacion anterior de los codigos de lateralidad es necesario asociar un nombre a la lateralidad
--de los registros actualizados entonces reemplazar los registros nulos de lateralidad y con codigos cie10 que no
--estan en la tabla de lateralidad por el valor asociado al codigo de lateralidad 4 (NO CORRESPONDE)
BEGIN;
UPDATE etl.tumores_pasto_limpio SET lateralidad =(
SELECT nom_lateralidad FROM tab_lateralidad WHERE cod_lateralidad = 4
) WHERE lateralidad IS NULL
AND cod_cie10 NOT IN (
SELECT DISTINCT(cod_cie10) FROM etl.tumores_pasto
WHERE cod_cie10 BETWEEN 'C079' AND 'C081'
OR cod_cie10 BETWEEN 'C090' AND 'C091'
OR cod_cie10 BETWEEN 'C098' AND 'C099'
OR cod_cie10 BETWEEN 'C300' AND 'C301'
OR cod_cie10 BETWEEN 'C340' AND 'C349'
OR cod_cie10 BETWEEN 'C400' AND 'C403'
OR cod_cie10 BETWEEN 'C413' AND 'C414'
OR cod_cie10 BETWEEN 'C441' AND 'C443'
OR cod_cie10 BETWEEN 'C445' AND 'C447'
OR cod_cie10 BETWEEN 'C471' AND 'C472'
OR cod_cie10 BETWEEN 'C491' AND 'C492'
OR cod_cie10 BETWEEN 'C500' AND 'C509'
OR cod_cie10 BETWEEN 'C569' AND 'C570'
OR cod_cie10 BETWEEN 'C620' AND 'C631'
OR cod_cie10 BETWEEN 'C690' AND 'C699'
OR cod_cie10 BETWEEN 'C710' AND 'C714'
OR cod_cie10 BETWEEN 'C722' AND 'C725'
OR cod_cie10 BETWEEN 'C740' AND 'C749'
OR cod_cie10 IN
('C310','C312','C384','C649','C659','C669','C700','C754')
GROUP BY cod_cie10
);
COMMIT;
--UPDATE 10998
--Query returned successfully in 718 msec.
--COMMIT
--Query returned successfully in 822 msec.
--Reemplazar los registros de lateralidad nulos con el valor de Desconocido correspondiente al codigo 9 de lateralidad
--para los codigos cie10 que corresponden con los listados en la tabla de lateralidad
BEGIN;
UPDATE etl.tumores_pasto_limpio SET lateralidad = (
SELECT nom_lateralidad FROM tab_lateralidad WHERE cod_lateralidad = 9
)
```

```
WHERE lateralidad IS NULL
AND cod_cie10 IN (
SELECT DISTINCT(cod_cie10) FROM etl.tumores_pasto
WHERE cod_cie10 BETWEEN 'C079' AND 'C081'
OR cod_cie10 BETWEEN 'C090' AND 'C091'
OR cod_cie10 BETWEEN 'C098' AND 'C099'
OR cod_cie10 BETWEEN 'C300' AND 'C301'
OR cod_cie10 BETWEEN 'C340' AND 'C349'
OR cod_cie10 BETWEEN 'C400' AND 'C403'
OR cod_cie10 BETWEEN 'C413' AND 'C414'
OR cod_cie10 BETWEEN 'C441' AND 'C443'
OR cod_cie10 BETWEEN 'C445' AND 'C447'
OR cod_cie10 BETWEEN 'C471' AND 'C472'
OR cod_cie10 BETWEEN 'C491' AND 'C492'
OR cod_cie10 BETWEEN 'C500' AND 'C509'
OR cod_cie10 BETWEEN 'C569' AND 'C570'
OR cod_cie10 BETWEEN 'C620' AND 'C631'
OR cod_cie10 BETWEEN 'C690' AND 'C699'
OR cod_cie10 BETWEEN 'C710' AND 'C714'
OR cod_cie10 BETWEEN 'C722' AND 'C725'
OR cod_cie10 BETWEEN 'C740' AND 'C749'
OR cod_cie10 IN
('C310','C312','C384','C649','C659','C669','C700','C754')
GROUP BY cod_cie10
);
COMMIT;
--UPDATE 2479
--Query returned successfully in 659 msec.
--COMMIT
--Query returned successfully in 734 msec.
--Estandarizar el atributo a mayusculas
BEGIN;
UPDATE etl.tumores_pasto_limpio SET lateralidad = UPPER(lateralidad);
COMMIT;
--UPDATE 41345
--Query returned successfully in 638 msec.
--COMMIT
--Query returned successfully in 982 msec.
-------------------------------------------------------------------------------------------------------------------------------------------------
--17. Campo comportamiento
--Estandarizar el atributo a mayusculas
BEGIN;
UPDATE etl.tumores_pasto_limpio SET comportamiento = UPPER(comportamiento);
COMMIT;
--UPDATE 41345
--Query returned successfully in 362 msec.
--COMMIT
--Query returned successfully in 691 msec.
-------------------------------------------------------------------------------------------------------------------------------------------------
--18. Campo cod_mencion_cancer
--Reemplazar los codigos de mencion de cancer por un nuevo codigo con valor 0 (NO APLICA) para tumores
con estado
--vital 'VIVO'
BEGIN;
UPDATE etl.tumores_pasto_limpio SET cod_mencion_cancer = 0 WHERE estado_vital = 'VIVO';
COMMIT;
```

--UPDATE 19900
--Query returned successfully in 712 msec.
--COMMIT
--Query returned successfully in 387 msec.
--Reemplazar los codigos de mencion de cancer por el codigo 3 que corresponde a 'MUERE POR CAUSAS DESCONOCIDAS' para
--aquellos registros cuyo estado vital es MUERTO y la variable codigo mencion de cancer esta en blanco o se habia
--registrado como igual a 9 (Desconocido)
BEGIN;
UPDATE etl.tumores_pasto_limpio SET cod_mencion_cancer = 3
WHERE estado_vital = 'MUERTO' AND (cod_mencion_cancer = 9 OR cod_mencion_cancer IS NULL);
COMMIT;
--UPDATE 4061
--Query returned successfully in 546 msec.
--COMMIT
--Query returned successfully in 120 msec.
--Reemplazar los codigos de mencion de cancer por el codigo 9 que corresponde a DESCONOCIDO para aquellos registros
--donde el estado vital es DESCONOCIDO o 'PERDIDO EN EL SEGUIMIENTO'
BEGIN;
UPDATE etl.tumores_pasto_limpio SET cod_mencion_cancer = 9
WHERE estado_vital = 'DESCONOCIDO' OR estado_vital = 'PERDIDO EN EL SEGUIMIENTO';
COMMIT;
--UPDATE 2526
--Query returned successfully in 426 msec.
--COMMIT
--Query returned successfully in 125 msec.

--19. Campo mencion_cancer
--Estandarizar el atributo a mayusculas
BEGIN;
UPDATE etl.tumores_pasto_limpio SET mencion_cancer = UPPER(mencion_cancer);
COMMIT;
--UPDATE 41345
--Query returned successfully in 881 msec.
--COMMIT
--Query returned successfully in 1 secs 196 msec.
--Acorde a la actualizacion del campo cod_mencion_cancer es necesario asociar un nombre a la mencion de cancer de los
--registros actualizados entonces reemplazar el valor de mencion cancer por NO APLICA para tumores con estado vital
--'VIVO'
BEGIN;
UPDATE etl.tumores_pasto_limpio SET mencion_cancer = 'NO APLICA' WHERE estado_vital = 'VIVO';
COMMIT;
--UPDATE 19900
--Query returned successfully in 594 msec.
--COMMIT
--Query returned successfully in 485 msec.
--Reemplazar la mencion de cancer por el valor MUERE POR CAUSAS DESCONOCIDAS para aquellos registros donde el estado
--vital es MUERTO y la variable mencion de cancer esta en blanco o habia sido registrada como 'DESCONOCIDO'
BEGIN;
UPDATE etl.tumores_pasto_limpio SET mencion_cancer = 'MUERE POR CAUSAS DESCONOCIDAS'

```
WHERE estado_vital = 'MUERTO' AND (mencion_cancer = 'DESCONOCIDO' OR mencion_cancer IS
NULL);
COMMIT;
--UPDATE 4061
--Query returned successfully in 465 msec.
--COMMIT
--Query returned successfully in 157 msec.
--Reemplazar la mencion de cancer por el valor DESCONOCIDO para aquellos registros donde el estado
vital es
--DESCONOCIDO y PERDIDO EN EL SEGUIMIENTO
BEGIN;
UPDATE etl.tumores_pasto_limpio SET mencion_cancer = 'DESCONOCIDO'
WHERE estado_vital = 'DESCONOCIDO' OR estado_vital = 'PERDIDO EN EL SEGUIMIENTO';
COMMIT;
--UPDATE 2526
--Query returned successfully in 465 msec.
--COMMIT
--Query returned successfully in 164 msec.
-----------------------------------------------------------------------------------------------------------------------------------
--20. Campo cod_profesion_expedidor
--Reemplazar los valores del codigo de profesion expedidor por un nuevo codigo 0 (NO APLICA) para
tumores con estado
--vital 'VIVO'
BEGIN;
UPDATE etl.tumores_pasto_limpio SET cod_profesion_expedidor = 0 WHERE estado_vital = 'VIVO';
COMMIT;
--UPDATE 19900
--Query returned successfully in 1 secs 190 msec.
--COMMIT
--Query returned successfully in 349 msec.
--Reemplazar los valores nulos o con valor Desconocido del codigo de profesion expedidor por el codigo 5
que corresponde
--a 'PROFESIÓN DEL EXPEDIDOR DESCONOCIDA' para aquellos registros donde el estado vital es
MUERTO
BEGIN;
UPDATE etl.tumores_pasto_limpio SET cod_profesion_expedidor = 5
WHERE estado_vital = 'MUERTO' AND (cod_profesion_expedidor = 9 OR cod_profesion_expedidor IS
NULL);
COMMIT;
--UPDATE 3159
--Query returned successfully in 496 msec.
--COMMIT
--Query returned successfully in 155 msec.
--Reemplazar los valores del codigo de profesion expedidor por el codigo 9 que corresponde a
DESCONOCIDO para
--aquellos registros donde el estado vital es PERDIDO EN EL SEGUIMIENTO o DESCONOCIDO
BEGIN;
UPDATE etl.tumores_pasto_limpio SET cod_profesion_expedidor = 9
WHERE estado_vital = 'PERDIDO EN EL SEGUIMIENTO' OR estado_vital = 'DESCONOCIDO';
COMMIT;
--UPDATE 2526
--Query returned successfully in 428 msec.
--COMMIT
--Query returned successfully in 151 msec.
-----------------------------------------------------------------------------------------------------------------------------------
--21. Campo profesion_expedidor
```

```
--Estandarizar el atributo a mayusculas
BEGIN;
UPDATE etl.tumores_pasto_limpio SET profesion_expedidor = UPPER(profesion_expedidor);
COMMIT;
--UPDATE 41345
--Query returned successfully in 881 msec.
--COMMIT
--Query returned successfully in 1 secs 344 msec.
--Acorde a la actualizacion del campo cod_profesion_expedidor es necesario asociar un nombre a la profesion del
--expedidor del certificado de defuncion de los registros actualizados entonces reemplazar el valor de profesion
--expedidor por el valor NO APLICA para tumores con estado vital 'VIVO'
BEGIN;
UPDATE etl.tumores_pasto_limpio SET profesion_expedidor = 'NO APLICA' WHERE estado_vital = 'VIVO';
COMMIT;
--UPDATE 19900
--Query returned successfully in 680 msec.
--COMMIT
--Query returned successfully in 490 msec.
--Reemplazar los valores nulos y DESCONOCIDO del campo de profesion expedidor por el valor de 'PROFESIÓN DEL EXPEDIDOR
--DESCONOCIDA' para aquellos registros donde el estado vital es MUERTO
BEGIN;
UPDATE etl.tumores_pasto_limpio SET profesion_expedidor = 'PROFESIÓN DEL EXPEDIDOR DESCONOCIDA'
WHERE estado_vital = 'MUERTO' AND (profesion_expedidor = 'DESCONOCIDO' OR profesion_expedidor IS NULL);
COMMIT;
--UPDATE 3159
--Query returned successfully in 519 msec.
--COMMIT
--Query returned successfully in 144 msec.
--Reemplazar los valores del campo de profesion expedidor por el valor de DESCONOCIDO para aquellos registros
--donde el estado vital es PERDIDO EN EL SEGUIMIENTO y DESCONOCIDO
BEGIN;
UPDATE etl.tumores_pasto_limpio SET profesion_expedidor = 'DESCONOCIDO'
WHERE estado_vital = 'PERDIDO EN EL SEGUIMIENTO' OR estado_vital = 'DESCONOCIDO';
COMMIT;
--UPDATE 2526
--Query returned successfully in 498 msec.
--COMMIT
--Query returned successfully in 165 msec.
--Eliminar espacios en blanco de relleno
BEGIN;
UPDATE etl.tumores_pasto_limpio SET profesion_expedidor = TRIM(BOTH FROM profesion_expedidor)
WHERE profesion_expedidor != TRIM(BOTH FROM profesion_expedidor);
COMMIT;
--UPDATE 3694
--Query returned successfully in 499 msec.
--COMMIT
--Query returned successfully in 163 msec.
-------------------------------------------------------------------------------------------------------------------------------------------
--22. Campo fecha_ultimo_contacto
```

```
--Reemplazar las fechas nulas por fechas desconocidas bajo el formato 99-99-9999
BEGIN;
UPDATE etl.tumores_pasto_limpio SET fecha_ultimo_contacto = '99-99-9999'
WHERE fecha_ultimo_contacto IS NULL;
COMMIT;
--UPDATE 2941
--Query returned successfully in 243 msec.
--COMMIT
--Query returned successfully in 180 msec.
--Corregir el valor anomalo 11-30-2012 por el valor correcto en el formato DD-MM-AAAA (30-11-2012)
BEGIN;
UPDATE etl.tumores_pasto_limpio SET fecha_ultimo_contacto = '30-11-2012'
WHERE fecha_ultimo_contacto = '11-30-2012';
COMMIT;
--UPDATE 2
--Query returned successfully in 405 msec.
--COMMIT
--Query returned successfully in 102 msec.
-----------------------------------------------------------------------------------------------------------------------------------------
--23. Campo cod_fuente
--Reemplazar los codigos de fuente nulos por el valor de 999 que corresponde con el valor de
DESCONOCIDO
BEGIN;
UPDATE etl.tumores_pasto_limpio SET cod_fuente = 999 WHERE cod_fuente IS NULL;
COMMIT;
--UPDATE 1238
--Query returned successfully in 170 msec.
--COMMIT
--Query returned successfully in 160 msec.
--Eliminar espacios en blanco de relleno
BEGIN;
UPDATE etl.tumores_pasto_limpio SET cod_fuente = TRIM(BOTH FROM cod_fuente)
WHERE LENGTH(cod_fuente) < PG_COLUMN_SIZE(cod_fuente)-1;
COMMIT;
--UPDATE 41345
--Query returned successfully in 739 msec.
--COMMIT
--Query returned successfully in 921 msec.
-----------------------------------------------------------------------------------------------------------------------------------------
--24. Campo fuente
--Acorde a la actualizacion anterior del campo cod_fuente es necesario asociar un nombre a la fuente de los
registros
--actualizados entonces reemplazar los valores nulos de fuente por el valor de DESCONOCIDO
BEGIN;
UPDATE etl.tumores_pasto_limpio SET fuente = 'DESCONOCIDO' WHERE fuente IS NULL;
COMMIT;
--UPDATE 1238
--Query returned successfully in 172 msec.
--COMMIT
--Query returned successfully in 165 msec.
--Eliminar espacios en blanco de relleno
BEGIN;
UPDATE etl.tumores_pasto_limpio SET fuente = TRIM(BOTH FROM fuente) WHERE fuente !=
TRIM(BOTH FROM fuente);
COMMIT;
--UPDATE 18801
```

--Query returned successfully in 408 msec.
--COMMIT
--Query returned successfully in 255 msec.
--Estandarizar el atributo a mayusculas
BEGIN;
UPDATE etl.tumores_pasto_limpio SET fuente = UPPER(fuente);
COMMIT;
--UPDATE 41345
--Query returned successfully in 1 secs 300 msec.
--COMMIT
--Query returned successfully in 868 msec.

--25. Campo cod_tratamiento
--Reemplazar los codigos de tratamiento nulos por el valor de 99 que corresponde con el valor de DESCONOCIDO
BEGIN;
UPDATE etl.tumores_pasto_limpio SET cod_tratamiento = 99 WHERE cod_tratamiento IS NULL;
COMMIT;
--UPDATE 6456
--Query returned successfully in 801 msec.
--COMMIT
--Query returned successfully in 745 msec.

--26. Campo tratamiento
--Actualizar los valores incorrectos de tratamiento que corresponden con el codigo tratamiento No. 16 de NO REGISTRA
--a CRIOTERAPIA
BEGIN;
UPDATE etl.tumores_pasto_limpio SET tratamiento = 'CRIOTERAPIA' WHERE
cod_tratamiento = 16 AND tratamiento = 'NO REGISTRA';
COMMIT;
--UPDATE 6
--Query returned successfully in 672 msec.
--COMMIT
--Query returned successfully in 682 msec.
--Acorde a la actualizacion anterior del campo cod_tratamiento es necesario asociar un nombre al tratamiento de los
--registros actualizados entonces reemplazar los valores nulos de tratamiento por el valor de DESCONOCIDO
BEGIN;
UPDATE etl.tumores_pasto_limpio SET tratamiento = 'DESCONOCIDO' WHERE tratamiento IS NULL;
COMMIT;
--UPDATE 6456
--Query returned successfully in 840 msec.
--COMMIT
--Query returned successfully in 901 msec.
--Eliminar espacios en blanco de relleno
BEGIN;
UPDATE etl.tumores_pasto_limpio SET tratamiento = TRIM(BOTH FROM tratamiento)
WHERE tratamiento != TRIM(BOTH FROM tratamiento);
COMMIT;
--UPDATE 16369
--Query returned successfully in 1 secs 315 msec.

--27. Campo cod_estado_enfermedad

```
--Reemplazar los codigos de estado de enfermedad nulos por el valor de 9 que corresponde con el valor de
DESCONOCIDO
BEGIN;
UPDATE etl.tumores_pasto_limpio SET cod_estado_enfermedad = 9 WHERE cod_estado_enfermedad IS
NULL;
COMMIT;
--UPDATE 26648
--Query returned successfully in 1 secs 163 msec.
--COMMIT
--Query returned successfully in 1 secs 112 msec.
-----------------------------------------------------------------------------------------------------------------------------------------------
--28. Campo estado_enfermedad
--Acorde a la actualizacion anterior del campo cod_estado_enfermedad es necesario asociar un nombre al
estado de
--enfermedad de los registros actualizados entonces reemplazar los valores nulos de estado_enfermedad por
el valor de
--DESCONOCIDO
BEGIN;
UPDATE etl.tumores_pasto_limpio SET estado_enfermedad = 'DESCONOCIDO' WHERE
estado_enfermedad IS NULL;
COMMIT;
--UPDATE 26648
--Query returned successfully in 1 secs 842 msec.
--COMMIT
--Query returned successfully in 1 secs 925 msec.
-----------------------------------------------------------------------------------------------------------------------------------------------
--29. Campo fecha_tratamiento
--Corregir registros anomalos de fechas de tratamiento asignando el valor de fecha desconocida (99999999)
BEGIN;
UPDATE etl.tumores_pasto_limpio SET fecha_tratamiento = '99999999' WHERE fecha_tratamiento IN
('07070707','19050702','19709999','21200299','24','29/02929');
COMMIT;
--UPDATE 136
--Query returned successfully in 382 msec.
--COMMIT
--Query returned successfully in 356 msec.
--Reemplazar las fechas nulas por fechas desconocidas bajo el formato 99999999
BEGIN;
UPDATE etl.tumores_pasto_limpio SET fecha_tratamiento = '99999999' WHERE fecha_tratamiento IS
NULL;
COMMIT;
--UPDATE 26201
--Query returned successfully in 978 msec.
--COMMIT
--Query returned successfully in 992 msec.
--Corregir el valor anomalo 20143008 a fecha desconocida bajo el formato AAAAMMDD (99999999)
BEGIN;
UPDATE etl.tumores_pasto_limpio SET fecha_tratamiento = '99999999'
WHERE fecha_tratamiento = '20143008';
COMMIT;
--UPDATE 3
--Query returned successfully in 562 msec.
--COMMIT
--Query returned successfully in 603 msec.
-----------------------------------------------------------------------------------------------------------------------------------------------
--30. Campo orden_tratamiento
```

```
--Reemplazar los registros nulos de orden_tratamiento por 0 (no hay un tratamiento asociado)
BEGIN;
UPDATE etl.tumores_pasto_limpio SET orden_tratamiento = 0 WHERE orden_tratamiento IS NULL;
COMMIT;
--UPDATE 6456
--Query returned successfully in 1 secs 160 msec.
--COMMIT
--Query returned successfully in 1 secs 101 msec.
-----------------------------------------------------------------------------------------------------------------------------------------------
--31. Campo cod_departamento
--Reemplazar los valores de codigo de departamento de 99 por el valor de 52 porque luego del analisis esos registros
--de cancer corresponden al Municipio de Pasto
BEGIN;
UPDATE etl.tumores_pasto_limpio SET cod_departamento = '52' WHERE cod_departamento = '99';
COMMIT;
--UPDATE 2422
--Query returned successfully in 170 msec.
--COMMIT
--Query returned successfully in 172 msec.
--Eliminar espacios en blanco de relleno
BEGIN;
UPDATE etl.tumores_pasto_limpio SET cod_departamento = TRIM(BOTH FROM cod_departamento)
WHERE LENGTH(cod_departamento) < PG_COLUMN_SIZE(cod_departamento)-1;
COMMIT;
--UPDATE 41345
--Query returned successfully in 783 msec.
--COMMIT
--Query returned successfully in 647 msec.
-----------------------------------------------------------------------------------------------------------------------------------------------
--32. Campo departamento
--Acorde a la actualizacion anterior del campo cod_departamento es necesario asociar un nombre al departamento de
--los registros actualizados entonces reemplazar el valor de departamento por el valor NARIÑO
BEGIN;
UPDATE etl.tumores_pasto_limpio SET departamento = 'NARIÑO' WHERE departamento = 'DESCONOCIDO';
COMMIT;
--UPDATE 2422
--Query returned successfully in 94 msec.
--COMMIT
--Query returned successfully in 158 msec.
-----------------------------------------------------------------------------------------------------------------------------------------------
--33. Campo cod_municipio
--Reemplazar los valores de codigo de municipio de 99999 por el valor de 52001 porque luego del analisis esos
--registros de cancer corresponden al Municipio de Pasto
BEGIN;
UPDATE etl.tumores_pasto_limpio SET cod_municipio = '52001' WHERE cod_municipio = '99999';
COMMIT;
--UPDATE 2422
--Query returned successfully in 93 msec.
--COMMIT
--Query returned successfully in 208 msec.
--Eliminar espacios en blanco de relleno
BEGIN;
```

```
UPDATE etl.tumores_pasto_limpio SET cod_municipio = TRIM(BOTH FROM cod_municipio)
WHERE LENGTH(cod_municipio) < PG_COLUMN_SIZE(cod_municipio)-1;
COMMIT;
--UPDATE 41345
--Query returned successfully in 621 msec.
--COMMIT
--Query returned successfully in 582 msec.
-----------------------------------------------------------------------------------------------------------------------------------------------------------
--34. Campo municipio
--Acorde a la actualizacion anterior del campo cod_municipio es necesario asociar un nombre al municipio de los
--registros actualizados entonces reemplazar el valor de municipio por el valor PASTO
BEGIN;
UPDATE etl.tumores_pasto_limpio SET municipio = 'PASTO' WHERE municipio = 'DESCONOCIDO';
COMMIT;
--UPDATE 2422
--Query returned successfully in 101 msec.
--COMMIT
--Query returned successfully in 154 msec.
-----------------------------------------------------------------------------------------------------------------------------------------------------------
--35. Campo zona
--Estandarizar el atributo a mayusculas
BEGIN;
UPDATE etl.tumores_pasto_limpio SET zona = UPPER(zona);
COMMIT;
--UPDATE 41345
--Query returned successfully in 498 msec.
--COMMIT
--Query returned successfully in 638 msec.
-----------------------------------------------------------------------------------------------------------------------------------------------------------
--36. Campo cod_barrio_comuna
--Eliminar espacios en blanco de relleno
BEGIN;
UPDATE etl.tumores_pasto_limpio SET cod_barrio_comuna = TRIM(BOTH FROM cod_barrio_comuna);
COMMIT;
--UPDATE 41345
--Query returned successfully in 627 msec.
--COMMIT
--Query returned successfully in 1 secs 276 msec.
-----------------------------------------------------------------------------------------------------------------------------------------------------------
--37. Campo barrio_comuna
--Reemplazar el valor de barrio_comuna 'El Campanero' por 'El Campanero (Catambuco)'
BEGIN;
UPDATE etl.tumores_pasto_limpio SET barrio_comuna = 'El Campanero (Catambuco)'
WHERE barrio_comuna = 'El Campanero';
COMMIT;
--UPDATE 28
--Query returned successfully in 681 msec.
--COMMIT
--Query returned successfully in 476 msec.
--Reemplazar el valor de barrio_comuna 'Rural desconocido' por 'Rural desconocido (Desconocido)'
BEGIN;
UPDATE etl.tumores_pasto_limpio SET barrio_comuna = 'Rural desconocido (Desconocido)'
WHERE barrio_comuna = 'Rural desconocido';
COMMIT;
--UPDATE 93
```

```
--Query returned successfully in 622 msec.
--COMMIT
--Query returned successfully in 572 msec.
--Reemplazar el valor de barrio_comuna 'EL Socorro (No definido)' por 'El Socorro desconocido (El socorro)'
BEGIN;
UPDATE etl.tumores_pasto_limpio SET barrio_comuna = 'El Socorro desconocido (El socorro)'
WHERE barrio_comuna = 'El Socorro (No definido)';
COMMIT;
--UPDATE 19
--Query returned successfully in 205 msec.
--COMMIT
--Query returned successfully in 159 msec.
--Eliminar espacios en blanco de relleno
BEGIN;
UPDATE etl.tumores_pasto_limpio SET barrio_comuna = TRIM(BOTH FROM barrio_comuna)
WHERE LENGTH(barrio_comuna) < PG_COLUMN_SIZE(barrio_comuna)-1;
COMMIT;
--UPDATE 5741
--Query returned successfully in 318 msec.
--COMMIT
--Query returned successfully in 289 msec.
-------------------------------------------------------------------------------------------------------------------------------------------------------
--38. Campo cod_comuna_corregimiento
--Actualizar el valor de cod_comuna_corregimiento a '17' para los registros actualizados el campo barrio_comuna a
--'El Socorro Desconocido (El socorro)'
BEGIN;
UPDATE etl.tumores_pasto_limpio SET cod_comuna_corregimiento = '17'
WHERE barrio_comuna = 'El Socorro desconocido (El socorro)';
COMMIT;
--UPDATE 19
--Query returned successfully in 586 msec.
--COMMIT
--Query returned successfully in 320 msec.
--Eliminar espacios en blanco de relleno
BEGIN;
UPDATE etl.tumores_pasto_limpio SET cod_comuna_corregimiento = TRIM(BOTH FROM cod_comuna_corregimiento);
COMMIT;
--UPDATE 41345
--Query returned successfully in 1 secs 616 msec.
--COMMIT
--Query returned successfully in 1 secs 157 msec.
-------------------------------------------------------------------------------------------------------------------------------------------------------
=========================================================================
--TABLA FINAL
SELECT * FROM etl.tumores_pasto_limpio;
- - - -
```

discretización de algunos atributos, etc. con el fin de terminar de preparar la staging area para el proceso de Carga de los datos.

```
--=====================================================================
--BODEGA DE DATOS SISTEMA YACHAY - RPCMP
--REGISTRO POBLACIONAL DE CANCER DEL MUNICIPIO DE PASTO
--=====================================================================
--TRANSFORMACION DE DATOS A PARTIR DE LOS DATOS LIMPIOS DE LOS TUMORES-FUENTE DEL MUNICIPIO DE PASTO
--=====================================================================
--Crear la tabla tumores_pasto_transformado a partir de la tabla tumores_pasto_limpio.
BEGIN;
CREATE TABLE etl.tumores_pasto_transformado AS SELECT * FROM etl.tumores_pasto_limpio;
COMMIT;
--SELECT 41345
--Query returned successfully in 376 msec.
--COMMIT
--Query returned successfully in 764 msec.
-----------------------------------------------------------------------------------------------------------------------------
--1. Campo cod_sexo
--Modificar el tipo de dato a char(1)
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN cod_sexo SET DATA TYPE CHAR(1);
COMMIT;
--ALTER TABLE
--Query returned successfully in 426 msec.
--COMMIT
--Query returned successfully in 384 msec.
-----------------------------------------------------------------------------------------------------------------------------
--2. Campo sexo
--Modificar el tipo de dato a texto
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN sexo SET DATA TYPE TEXT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 388 msec.
--COMMIT
--Query returned successfully in 483 msec.
-----------------------------------------------------------------------------------------------------------------------------
--3. Campo edad_paciente
--Crear los atributos cod_grupo_etario1 y cod_grupo_etario2 de tipo SMALLINT y grupo_etario1 y grupo_etario2 de
--tipo TEXT
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ADD cod_grupo_etario1 SMALLINT;
ALTER TABLE etl.tumores_pasto_transformado ADD cod_grupo_etario2 SMALLINT;
ALTER TABLE etl.tumores_pasto_transformado ADD grupo_etario1 TEXT;
ALTER TABLE etl.tumores_pasto_transformado ADD grupo_etario2 TEXT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 1 secs 118 msec.
--COMMIT
--Query returned successfully in 120 msec.
```

```
--Actualizar los atributos cod_grupo_etario1 y grupo_etario1 discretizando el atributo edad_paciente en 18 grupos
--quinquenales, el grupo 0 para 0 años conocido como Lactantes y el grupo 19 para Desconocido, de acuerdo a la
--siguiente relacion:
--0 -> 0 (Lactantes) -> De 0 años de edad
--1 -> 1 - 4 -> De 1 a 4 años de edad
--2 -> 5 - 9 -> De 5 a 9 años de edad
--3 -> 10 - 14 -> De 10 a 14 años de edad
--4 -> 15 - 19 -> De 15 a 19 años de edad
--5 -> 20 - 24 -> De 20 a 24 años de edad
--6 -> 25 - 29 -> De 25 a 29 años de edad
--7 -> 30 - 34 -> De 30 a 34 años de edad
--8 -> 35 - 39 -> De 35 a 39 años de edad
--9 -> 40 - 44 -> De 40 a 44 años de edad
--10 -> 45 - 49 -> De 45 a 49 años de edad
--11 -> 50 - 54 -> De 50 a 54 años de edad
--12 -> 55 - 59 -> De 55 a 59 años de edad
--13 -> 60 - 64 -> De 60 a 64 años de edad
--14 -> 65 - 69 -> De 65 a 69 años de edad
--15 -> 70 - 74 -> De 70 a 74 años de edad
--16 -> 75 - 79 -> De 75 a 79 años de edad
--17 -> 80 - 84 -> De 80 a 84 años de edad
--18 -> 85 y más -> De 85 y más años de edad
--19 -> Desconocido
BEGIN;
UPDATE etl.tumores_pasto_transformado SET cod_grupo_etario1 = CASE
WHEN edad_paciente = 999 THEN 19
WHEN edad_paciente = 0 THEN 0
WHEN edad_paciente BETWEEN 1 AND 4 THEN 1
WHEN edad_paciente BETWEEN 5 AND 9 THEN 2
WHEN edad_paciente BETWEEN 10 AND 14 THEN 3
WHEN edad_paciente BETWEEN 15 AND 19 THEN 4
WHEN edad_paciente BETWEEN 20 AND 24 THEN 5
WHEN edad_paciente BETWEEN 25 AND 29 THEN 6
WHEN edad_paciente BETWEEN 30 AND 34 THEN 7
WHEN edad_paciente BETWEEN 35 AND 39 THEN 8
WHEN edad_paciente BETWEEN 40 AND 44 THEN 9
WHEN edad_paciente BETWEEN 45 AND 49 THEN 10
WHEN edad_paciente BETWEEN 50 AND 54 THEN 11
WHEN edad_paciente BETWEEN 55 AND 59 THEN 12
WHEN edad_paciente BETWEEN 60 AND 64 THEN 13
WHEN edad_paciente BETWEEN 65 AND 69 THEN 14
WHEN edad_paciente BETWEEN 70 AND 74 THEN 15
WHEN edad_paciente BETWEEN 75 AND 79 THEN 16
WHEN edad_paciente BETWEEN 80 AND 84 THEN 17
WHEN edad_paciente >= 85 THEN 18
ELSE 19 END;
COMMIT;
--UPDATE 41345
--Query returned successfully in 1 secs 207 msec.
--COMMIT
--Query returned successfully in 1 secs 101 msec.
BEGIN;
UPDATE etl.tumores_pasto_transformado SET grupo_etario1 = CASE
WHEN edad_paciente = 999 THEN 'Desconocido'
```

```
WHEN edad_paciente = 0 THEN '0 (Lactantes)'
WHEN edad_paciente BETWEEN 1 AND 4 THEN '1 - 4'
WHEN edad_paciente BETWEEN 5 AND 9 THEN '5 - 9'
WHEN edad_paciente BETWEEN 10 AND 14 THEN '10 - 14'
WHEN edad_paciente BETWEEN 15 AND 19 THEN '15 - 19'
WHEN edad_paciente BETWEEN 20 AND 24 THEN '20 - 24'
WHEN edad_paciente BETWEEN 25 AND 29 THEN '25 - 29'
WHEN edad_paciente BETWEEN 30 AND 34 THEN '30 - 34'
WHEN edad_paciente BETWEEN 35 AND 39 THEN '35 - 39'
WHEN edad_paciente BETWEEN 40 AND 44 THEN '40 - 44'
WHEN edad_paciente BETWEEN 45 AND 49 THEN '45 - 49'
WHEN edad_paciente BETWEEN 50 AND 54 THEN '50 - 54'
WHEN edad_paciente BETWEEN 55 AND 59 THEN '55 - 59'
WHEN edad_paciente BETWEEN 60 AND 64 THEN '60 - 64'
WHEN edad_paciente BETWEEN 65 AND 69 THEN '65 - 69'
WHEN edad_paciente BETWEEN 70 AND 74 THEN '70 - 74'
WHEN edad_paciente BETWEEN 75 AND 79 THEN '75 - 79'
WHEN edad_paciente BETWEEN 80 AND 84 THEN '80 - 84'
WHEN edad_paciente >= 85 THEN '85 y más'
ELSE 'Desconocido' END;
COMMIT;
--UPDATE 41345
--Query returned successfully in 1 secs 465 msec.
--COMMIT
--Query returned successfully in 1 secs 660 msec.
--Actualizar los atributos cod_grupo_etario2 y grupo_etario2 a partir de los valores de cod_grupo_etario1 y
--grupo_etario1 con la salvedad de que el grupo 0 desaparece y se incluye en el grupo 1 cambiando su
descripcion
--de '1 - 4' a '0 - 4' de acuerdo a la siguiente relacion:
--1 -> 0 - 4 -> De 0 a 4 años de edad
--2 -> 5 - 9 -> De 5 a 9 años de edad
--...
BEGIN;
UPDATE etl.tumores_pasto_transformado SET cod_grupo_etario2 = CASE
WHEN cod_grupo_etario1 = 0 THEN 1
ELSE cod_grupo_etario1 END;
COMMIT;
--UPDATE 41345
--Query returned successfully in 830 msec.
--COMMIT
--Query returned successfully in 720 msec.
BEGIN;
UPDATE etl.tumores_pasto_transformado SET grupo_etario2 = CASE
WHEN grupo_etario1 = '0 (Lactantes)' THEN '0 - 4'
WHEN grupo_etario1 = '1 - 4' THEN '0 - 4'
ELSE grupo_etario1 END;
COMMIT;
--UPDATE 41345
--Query returned successfully in 1 secs 93 msec.
--COMMIT
--Query returned successfully in 1 secs 201 msec.
--Crear el atributo descripcion_edad de tipo TEXT
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ADD descripcion_edad TEXT;
COMMIT;
--ALTER TABLE
```

--Query returned successfully in 167 msec.
--COMMIT
--Query returned successfully in 136 msec.
--Actualizar el atributo descripcion_edad con el valor de la edad
BEGIN;
UPDATE etl.tumores_pasto_transformado SET descripcion_edad = edad_paciente;
COMMIT;
--UPDATE 41345
--Query returned successfully in 1 secs 349 msec.
--COMMIT
--Query returned successfully in 1 secs 136 msec.
--Actualizar los valores 999 del atributo descripcion_edad por Desconocido
BEGIN;
UPDATE etl.tumores_pasto_transformado SET descripcion_edad = CASE
WHEN descripcion_edad = '999' THEN 'Desconocido' ELSE descripcion_edad END;
COMMIT;
--UPDATE 41345
--Query returned successfully in 908 msec.
--COMMIT
--Query returned successfully in 730 msec.

--4. Campo cod_estado_vital
--Modificar el tipo de dato a char(1)
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN cod_estado_vital SET DATA TYPE CHAR(1);
COMMIT;
--ALTER TABLE
--Query returned successfully in 916 msec.
--COMMIT
--Query returned successfully in 572 msec.

--5. Campo estado_vital
--Modificar el tipo de dato a texto
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN estado_vital SET DATA TYPE TEXT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 74 msec.
--COMMIT
--Query returned successfully in 77 msec.

--6. Campo fecha_diagnostico
--Modificar el tipo de dato a texto
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN fecha_diagnostico SET DATA TYPE TEXT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 100 msec.
--COMMIT
--Query returned successfully in 72 msec.
--Crear los atributos cod_dia_diagnostico,dia_diagnostico,cod_mes_diagnostico,mes_diagnostico,ano_diagnostico de
--tipo SMALLINT

```
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ADD cod_dia_diagnostico SMALLINT;
ALTER TABLE etl.tumores_pasto_transformado ADD dia_diagnostico SMALLINT;
ALTER TABLE etl.tumores_pasto_transformado ADD cod_mes_diagnostico SMALLINT;
ALTER TABLE etl.tumores_pasto_transformado ADD mes_diagnostico SMALLINT;
ALTER TABLE etl.tumores_pasto_transformado ADD ano_diagnostico SMALLINT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 120 msec.
--COMMIT
--Query returned successfully in 69 msec.
--Separar la fecha de diagnostico en los campos creados
BEGIN;
UPDATE etl.tumores_pasto_transformado SET
cod_dia_diagnostico = CAST(split_part(fecha_diagnostico, '-', 1) AS SMALLINT),
dia_diagnostico = CAST(split_part(fecha_diagnostico, '-', 1) AS SMALLINT),
cod_mes_diagnostico = CAST(split_part(fecha_diagnostico, '-', 2) AS SMALLINT),
mes_diagnostico = CAST(split_part(fecha_diagnostico, '-', 2) AS SMALLINT),
ano_diagnostico = CAST(split_part(fecha_diagnostico, '-', 3) AS SMALLINT);
COMMIT;
--UPDATE 41345
--Query returned successfully in 1 secs 639 msec.
--COMMIT
--Query returned successfully in 637 msec.
--Reemplazar el valor 99 del atributo cod_dia_diagnostico por el valor de 32 estableciendo un orden para el
--atributo dia_diagnostico
BEGIN;
UPDATE etl.tumores_pasto_transformado SET cod_dia_diagnostico = 32 WHERE cod_dia_diagnostico =
99;
COMMIT;
--UPDATE 1132
--Query returned successfully in 148 msec.
--COMMIT
--Query returned successfully in 85 msec.
--Reemplazar el valor 99 del atributo cod_mes_diagnostico por el valor de 13 estableciendo un orden para el
--atributo mes_diagnostico
BEGIN;
UPDATE etl.tumores_pasto_transformado SET cod_mes_diagnostico = 13 WHERE cod_mes_diagnostico
= 99;
COMMIT;
--UPDATE 592
--Query returned successfully in 400 msec.
--COMMIT
--Query returned successfully in 203 msec.
--Crear los atributos trimestre_diagnostico y semestre_diagnostico de tipo TEXT
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ADD trimestre_diagnostico TEXT;
ALTER TABLE etl.tumores_pasto_transformado ADD semestre_diagnostico TEXT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 88 msec.
--COMMIT
--Query returned successfully in 73 msec.
--Actualizar el atributo trimestre_diagnostico discretizando el atributo mes_diagnostico en 4 grupos de meses
y el
--grupo 0 para Desconocido, de acuerdo a la siguiente relacion:
```

```
--0 -> Desconocido -> mes 99
--1 -> Trimestre 1 (de enero a marzo) -> meses 1 al 3
--2 -> Trimestre 2 (de abril a junio) -> meses 4 al 6
--3 -> Trimestre 3 (de julio a septiembre) -> meses 7 al 9
--4 -> Trimestre 4 (de octubre a diciembre) -> meses 10 al 12
BEGIN;
UPDATE etl.tumores_pasto_transformado SET trimestre_diagnostico = CASE
WHEN mes_diagnostico = 99 THEN 'Desconocido'
WHEN mes_diagnostico BETWEEN 1 AND 3 THEN 'Trimestre 1 (de enero a marzo)'
WHEN mes_diagnostico BETWEEN 4 AND 6 THEN 'Trimestre 2 (de abril a junio)'
WHEN mes_diagnostico BETWEEN 7 AND 9 THEN 'Trimestre 3 (de julio a septiembre)'
WHEN mes_diagnostico BETWEEN 10 AND 12 THEN 'Trimestre 4 (de octubre a diciembre)'
ELSE 'Desconocido' END;
COMMIT;
--UPDATE 41345
--Query returned successfully in 771 msec.
--COMMIT
--Query returned successfully in 878 msec.
--Actualizar el atributo semestre_diagnostico discretizando el atributo mes_diagnostico en 2 grupos de meses y el
--grupo 0 para Desconocido, de acuerdo a la siguiente relacion:
--0 -> Desconocido -> mes 99
--1 -> Semestre 1 (de enero a junio) -> meses 1 al 6
--2 -> Semestre 2 (de julio a diciembre) -> meses 7 al 12
BEGIN;
UPDATE etl.tumores_pasto_transformado SET semestre_diagnostico = CASE
WHEN mes_diagnostico = 99 THEN 'Desconocido'
WHEN mes_diagnostico BETWEEN 1 AND 6 THEN 'Semestre 1 (de enero a junio)'
WHEN mes_diagnostico BETWEEN 7 AND 12 THEN 'Semestre 2 (de julio a diciembre)'
ELSE 'Desconocido' END;
COMMIT;
--UPDATE 41345
--Query returned successfully in 792 msec.
--COMMIT
--Query returned successfully in 592 msec.
-----------------------------------------------------------------------------------------------------------------------------------------
--7. Campo cod_base_diagnostico
--Modificar el tipo de dato a char(1)
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN cod_base_diagnostico SET DATA TYPE CHAR(1);
COMMIT;
--ALTER TABLE
--Query returned successfully in 522 msec.
--COMMIT
--Query returned successfully in 616 msec.
-----------------------------------------------------------------------------------------------------------------------------------------
--8. Campo base_diagnostico
--Modificar el tipo de dato a texto
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN base_diagnostico SET DATA TYPE TEXT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 136 msec.
--COMMIT
```

```
--Query returned successfully in 71 msec.
-----------------------------------------------------------------------------------------------------------------------------------
--9. Campo cod_localizacion
--Modificar el tipo de dato a texto
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN cod_localizacion SET DATA TYPE
TEXT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 736 msec.
--COMMIT
--Query returned successfully in 650 msec.
-----------------------------------------------------------------------------------------------------------------------------------
--10. Campo localizacion
--Modificar el tipo de dato a texto
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN localizacion SET DATA TYPE
TEXT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 155 msec.
--COMMIT
--Query returned successfully in 75 msec.
--Actualizar el campo localizacion omitiendo la parte inicial que contiene el codigo de localizacion en la
--nomenclatura CIEO-3
BEGIN;
UPDATE etl.tumores_pasto_transformado SET localizacion =
TRIM(SUBSTRING(localizacion FROM POSITION(' ' IN localizacion) + 1));
COMMIT;
--UPDATE 41345
--Query returned successfully in 735 msec.
--COMMIT
--Query returned successfully in 1 secs 94 msec.
-----------------------------------------------------------------------------------------------------------------------------------
--11. Campo cod_cieo3_grupo
--Modificar el tipo de dato a texto
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN cod_cieo3_grupo SET DATA TYPE
TEXT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 455 msec.
--COMMIT
--Query returned successfully in 588 msec.
-----------------------------------------------------------------------------------------------------------------------------------
--12. Campo no_cieo3_grupo
--Modificar el tipo de dato a texto
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN no_cieo3_grupo SET DATA TYPE
TEXT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 458 msec.
--COMMIT
--Query returned successfully in 508 msec.
-----------------------------------------------------------------------------------------------------------------------------------
```

```
--13. Campo cieo3_grupo
--Modificar el tipo de dato a texto
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN cieo3_grupo SET DATA TYPE
TEXT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 73 msec.
--COMMIT
--Query returned successfully in 72 msec.
-----------------------------------------------------------------------------------------------------------------------
--14. Campo cod_morfologia
--Modificar el tipo de dato a texto
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN cod_morfologia SET DATA TYPE
TEXT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 498 msec.
--COMMIT
--Query returned successfully in 516 msec.
-----------------------------------------------------------------------------------------------------------------------
--15. Campo morfologia
--Modificar el tipo de dato a texto
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN morfologia SET DATA TYPE TEXT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 70 msec.
--COMMIT
--Query returned successfully in 72 msec.
-----------------------------------------------------------------------------------------------------------------------
--16. Campo cod_cie10
--Modificar el tipo de dato a texto
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN cod_cie10 SET DATA TYPE TEXT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 785 msec.
--COMMIT
--Query returned successfully in 818 msec.
-----------------------------------------------------------------------------------------------------------------------
--17. Campo cie10
--Modificar el tipo de dato a texto
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN cie10 SET DATA TYPE TEXT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 72 msec.
--COMMIT
--Query returned successfully in 93 msec.
-----------------------------------------------------------------------------------------------------------------------
--18. Campo cod_cie10_grupo
--Modificar el tipo de dato a texto
BEGIN;
```

```
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN cod_cie10_grupo SET DATA TYPE
TEXT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 720 msec.
--COMMIT
--Query returned successfully in 571 msec.
-----------------------------------------------------------------------------------------------------------------------------
--19. Campo cie10_grupo_general
--Modificar el tipo de dato a texto
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN cie10_grupo_general SET DATA
TYPE TEXT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 123 msec.
--COMMIT
--Query returned successfully in 66 msec.
-----------------------------------------------------------------------------------------------------------------------------
--20. Campo cie10_grupo_excepto_c44
--Modificar el tipo de dato a texto
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN cie10_grupo_excepto_c44 SET
DATA TYPE TEXT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 130 msec.
--COMMIT
--Query returned successfully in 68 msec.
-----------------------------------------------------------------------------------------------------------------------------
--21. Campo cod_grupo_cie10
--Modificar el tipo de dato a texto
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN cod_grupo_cie10 SET DATA TYPE
TEXT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 220 msec.
--COMMIT
--Query returned successfully in 574 msec.
-----------------------------------------------------------------------------------------------------------------------------
--22. Campo grupo_cie10
--Modificar el tipo de dato a texto
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN grupo_cie10 SET DATA TYPE
TEXT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 114 msec.
--COMMIT
--Query returned successfully in 69 msec.
-----------------------------------------------------------------------------------------------------------------------------
--23. Campo cod_lateralidad
--Modificar el tipo de dato a char(1)
BEGIN;
```

```
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN cod_lateralidad SET DATA TYPE CHAR(1);
COMMIT;
--ALTER TABLE
--Query returned successfully in 590 msec.
--COMMIT
--Query returned successfully in 75 msec.
-----------------------------------------------------------------------------------------------------------------------------------
--24. Campo lateralidad
--Modificar el tipo de dato a texto
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN lateralidad SET DATA TYPE TEXT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 128 msec.
--COMMIT
--Query returned successfully in 78 msec.
-----------------------------------------------------------------------------------------------------------------------------------
--25. Campo cod_comportamiento
--Modificar el tipo de dato a char(1)
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN cod_comportamiento SET DATA TYPE CHAR(1);
COMMIT;
--ALTER TABLE
--Query returned successfully in 576 msec.
--COMMIT
--Query returned successfully in 71 msec.
-----------------------------------------------------------------------------------------------------------------------------------
--26. Campo comportamiento
--Modificar el tipo de dato a texto
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN comportamiento SET DATA TYPE TEXT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 134 msec.
--COMMIT
--Query returned successfully in 65 msec.
-----------------------------------------------------------------------------------------------------------------------------------
--27. Campo cod_grado_diferenciacion
--Modificar el tipo de dato a char(1)
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN cod_grado_diferenciacion SET DATA TYPE CHAR(1);
COMMIT;
--ALTER TABLE
--Query returned successfully in 831 msec.
--COMMIT
--Query returned successfully in 136 msec.
-----------------------------------------------------------------------------------------------------------------------------------
--28. Campo grado_diferenciacion
--Modificar el tipo de dato a texto
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN grado_diferenciacion SET DATA TYPE TEXT;
```

```
COMMIT;
--ALTER TABLE
--Query returned successfully in 72 msec.
--COMMIT
--Query returned successfully in 71 msec.
-----------------------------------------------------------------------------------------------------------------------------
--29. Campo cod_mencion_cancer
--Modificar el tipo de dato a char(1)
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN cod_mencion_cancer SET DATA TYPE CHAR(1);
COMMIT;
--ALTER TABLE
--Query returned successfully in 852 msec.
--COMMIT
--Query returned successfully in 1 secs 83 msec.
-----------------------------------------------------------------------------------------------------------------------------
--30. Campo mencion_cancer
--Modificar el tipo de dato a texto
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN mencion_cancer SET DATA TYPE TEXT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 146 msec.
--COMMIT
--Query returned successfully in 110 msec.
-----------------------------------------------------------------------------------------------------------------------------
--31. Campo cod_profesion_expedidor
--Modificar el tipo de dato a char(1)
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN cod_profesion_expedidor SET DATA TYPE CHAR(1);
COMMIT;
--ALTER TABLE
--Query returned successfully in 932 msec.
--COMMIT
--Query returned successfully in 1 secs 131 msec.
-----------------------------------------------------------------------------------------------------------------------------
--32. Campo profesion_expedidor
--Modificar el tipo de dato a texto
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN profesion_expedidor SET DATA TYPE TEXT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 116 msec.
--COMMIT
--Query returned successfully in 159 msec.
-----------------------------------------------------------------------------------------------------------------------------
--33. Campo fecha_ultimo_contacto
--Modificar el tipo de dato a texto
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN fecha_ultimo_contacto SET DATA TYPE TEXT;
COMMIT;
```

```
--ALTER TABLE
--Query returned successfully in 121 msec.
--COMMIT
--Query returned successfully in 116 msec.
--Crear los atributos de tipo SMALLINT
--cod_dia_muerte,dia_muerte,cod_mes_muerte,mes_muerte,ano_muerte,
--
cod_dia_ultimo_contacto,dia_ultimo_contacto,cod_mes_ultimo_contacto,mes_ultimo_contacto,ano_ultimo
_contacto
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ADD cod_dia_muerte SMALLINT;
ALTER TABLE etl.tumores_pasto_transformado ADD dia_muerte SMALLINT;
ALTER TABLE etl.tumores_pasto_transformado ADD cod_mes_muerte SMALLINT;
ALTER TABLE etl.tumores_pasto_transformado ADD mes_muerte SMALLINT;
ALTER TABLE etl.tumores_pasto_transformado ADD ano_muerte SMALLINT;
ALTER TABLE etl.tumores_pasto_transformado ADD cod_dia_ultimo_contacto SMALLINT;
ALTER TABLE etl.tumores_pasto_transformado ADD dia_ultimo_contacto SMALLINT;
ALTER TABLE etl.tumores_pasto_transformado ADD cod_mes_ultimo_contacto SMALLINT;
ALTER TABLE etl.tumores_pasto_transformado ADD mes_ultimo_contacto SMALLINT;
ALTER TABLE etl.tumores_pasto_transformado ADD ano_ultimo_contacto SMALLINT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 136 msec.
--COMMIT
--Query returned successfully in 122 msec.
--Separar la fecha de ultimo contacto en los campos creados para la fecha de muerte cuando el estado vital
es MUERTO
BEGIN;
UPDATE etl.tumores_pasto_transformado SET
cod_dia_muerte = CAST(split_part(fecha_ultimo_contacto, '-', 1) AS SMALLINT),
dia_muerte = CAST(split_part(fecha_ultimo_contacto, '-', 1) AS SMALLINT),
cod_mes_muerte = CAST(split_part(fecha_ultimo_contacto, '-', 2) AS SMALLINT),
mes_muerte = CAST(split_part(fecha_ultimo_contacto, '-', 2) AS SMALLINT),
ano_muerte = CAST(split_part(fecha_ultimo_contacto, '-', 3) AS SMALLINT)
WHERE estado_vital = 'MUERTO';
COMMIT;
--UPDATE 18975
--Query returned successfully in 707 msec.
--COMMIT
--Query returned successfully in 839 msec.
--Asignar el valor de 0 a los campos cod_dia_muerte, dia_muerte, cod_mes_muerte, mes_muerte y
ano_muerte cuando
--el estado vital es VIVO
BEGIN;
UPDATE etl.tumores_pasto_transformado SET
cod_dia_muerte = 0, dia_muerte = 0, cod_mes_muerte = 0, mes_muerte = 0, ano_muerte = 0
WHERE estado_vital = 'VIVO';
COMMIT;
--UPDATE 19857
--Query returned successfully in 558 msec.
--COMMIT
--Query returned successfully in 585 msec.
--Reemplazar el valor 99 del atributo cod_dia_muerte por el valor de 32 estableciendo un orden para el
atributo
--dia_muerte
BEGIN;
```

```
UPDATE etl.tumores_pasto_transformado SET cod_dia_muerte = 32 WHERE cod_dia_muerte = 99;
COMMIT;
--UPDATE 1493
--Query returned successfully in 242 msec.
--COMMIT
--Query returned successfully in 175 msec.
--Reemplazar los valores nulos del atributo cod_dia_muerte por el valor de 33 estableciendo un orden para
el
--atributo dia_muerte
BEGIN;
UPDATE etl.tumores_pasto_transformado SET cod_dia_muerte = 33 WHERE cod_dia_muerte IS NULL;
COMMIT;
--UPDATE 2513
--Query returned successfully in 145 msec.
--COMMIT
--Query returned successfully in 165 msec.
--Reemplazar el valor 99 del atributo cod_mes_muerte por el valor de 13 estableciendo un orden para el
atributo
--mes_muerte
BEGIN;
UPDATE etl.tumores_pasto_transformado SET cod_mes_muerte = 13 WHERE cod_mes_muerte = 99;
COMMIT;
--UPDATE 1385
--Query returned successfully in 82 msec.
--COMMIT
--Query returned successfully in 123 msec.
--Reemplazar los valores nulos del atributo cod_mes_muerte por el valor de 14 estableciendo un orden para
el
--atributo mes_muerte
BEGIN;
UPDATE etl.tumores_pasto_transformado SET cod_mes_muerte = 14 WHERE cod_mes_muerte IS NULL;
COMMIT;
--UPDATE 2513
--Query returned successfully in 144 msec.
--COMMIT
--Query returned successfully in 123 msec.
--Separar la fecha de ultimo contacto en los nuevos campos creados para la fecha de ultimo contacto para
todos los
--estados vitales
BEGIN;
UPDATE etl.tumores_pasto_transformado SET
cod_dia_ultimo_contacto = CAST(split_part(fecha_ultimo_contacto, '-', 1) AS SMALLINT),
dia_ultimo_contacto = CAST(split_part(fecha_ultimo_contacto, '-', 1) AS SMALLINT),
cod_mes_ultimo_contacto = CAST(split_part(fecha_ultimo_contacto, '-', 2) AS SMALLINT),
mes_ultimo_contacto = CAST(split_part(fecha_ultimo_contacto, '-', 2) AS SMALLINT),
ano_ultimo_contacto = CAST(split_part(fecha_ultimo_contacto, '-', 3) AS SMALLINT);
COMMIT;
--UPDATE 41345
--Query returned successfully in 1 secs 143 msec.
--COMMIT
--Query returned successfully in 1 secs 233 msec.
--Reemplazar el valor 99 del atributo cod_dia_ultimo_contacto por el valor de 32 estableciendo un orden
para el
--atributo dia_ultimo_contacto
BEGIN;
```

```
UPDATE etl.tumores_pasto_transformado SET cod_dia_ultimo_contacto = 32 WHERE cod_dia_ultimo_contacto = 99;
COMMIT;
--UPDATE 4262
--Query returned successfully in 155 msec.
--COMMIT
--Query returned successfully in 85 msec.
--Reemplazar el valor 99 del atributo cod_mes_ultimo_contacto por el valor de 13 estableciendo un orden para el
--atributo mes_ultimo_contacto
BEGIN;
UPDATE etl.tumores_pasto_transformado SET cod_mes_ultimo_contacto = 13 WHERE cod_mes_ultimo_contacto = 99;
COMMIT;
--UPDATE 4020
--Query returned successfully in 147 msec.
--COMMIT
--Query returned successfully in 203 msec.
--Crear los atributos de tipo TEXT
--trimestre_muerte, trimestre_ultimo_contacto, semestre_muerte y semestre_ultimo_contacto
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ADD trimestre_muerte TEXT;
ALTER TABLE etl.tumores_pasto_transformado ADD semestre_muerte TEXT;
ALTER TABLE etl.tumores_pasto_transformado ADD trimestre_ultimo_contacto TEXT;
ALTER TABLE etl.tumores_pasto_transformado ADD semestre_ultimo_contacto TEXT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 77 msec.
--COMMIT
--Query returned successfully in 172 msec.
--Actualizar el atributo trimestre_muerte discretizando el atributo mes_muerte en 4 grupos de meses, grupo 0 para
--Desconocido, grupo 5 (No aplica) para los registros 0 (estado vital VIVO) y grupo 6 (Sin informacion) para los
--registros nulos (estados vitales PERDIDO EN EL SEGUIMIENTO y DESCONOCIDO) de acuerdo a la siguiente relacion:
--0 -> Desconocido -> mes 99
--1 -> Trimestre 1 (de enero a marzo) -> meses 1 al 3
--2 -> Trimestre 2 (de abril a junio) -> meses 4 al 6
--3 -> Trimestre 3 (de julio a septiembre) -> meses 7 al 9
--4 -> Trimestre 4 (de octubre a diciembre) -> meses 10 al 12
--5 -> No aplica -> mes 0
--6 -> Sin información -> mes nulo
BEGIN;
UPDATE etl.tumores_pasto_transformado SET trimestre_muerte = CASE
WHEN mes_muerte = 99 THEN 'Desconocido'
WHEN mes_muerte BETWEEN 1 AND 3 THEN 'Trimestre 1 (de enero a marzo)'
WHEN mes_muerte BETWEEN 4 AND 6 THEN 'Trimestre 2 (de abril a junio)'
WHEN mes_muerte BETWEEN 7 AND 9 THEN 'Trimestre 3 (de julio a septiembre)'
WHEN mes_muerte BETWEEN 10 AND 12 THEN 'Trimestre 4 (de octubre a diciembre)'
WHEN mes_muerte = 0 THEN 'No aplica'
ELSE 'Sin información' END;
COMMIT;
--UPDATE 41345
--Query returned successfully in 1 secs 366 msec.
--COMMIT
```

```
--Query returned successfully in 1 secs 2 msec.
--Actualizar el atributo trimestre_ultimo_contacto discretizando el atributo mes_ultimo_contacto en 4 grupos de meses
--y grupo 0 para Desconocido, de acuerdo a la siguiente relacion:
--0 -> Desconocido -> mes 99
--1 -> Trimestre 1 (de enero a marzo) -> meses 1 al 3
--2 -> Trimestre 2 (de abril a junio) -> meses 4 al 6
--3 -> Trimestre 3 (de julio a septiembre) -> meses 7 al 9
--4 -> Trimestre 4 (de octubre a diciembre) -> meses 10 al 12
BEGIN;
UPDATE etl.tumores_pasto_transformado SET trimestre_ultimo_contacto = CASE
WHEN mes_ultimo_contacto = 99 THEN 'Desconocido'
WHEN mes_ultimo_contacto BETWEEN 1 AND 3 THEN 'Trimestre 1 (de enero a marzo)'
WHEN mes_ultimo_contacto BETWEEN 4 AND 6 THEN 'Trimestre 2 (de abril a junio)'
WHEN mes_ultimo_contacto BETWEEN 7 AND 9 THEN 'Trimestre 3 (de julio a septiembre)'
WHEN mes_ultimo_contacto BETWEEN 10 AND 12 THEN 'Trimestre 4 (de octubre a diciembre)'
ELSE 'Desconocido' END;
COMMIT;
--UPDATE 41345
--Query returned successfully in 764 msec.
--COMMIT
--Query returned successfully in 1 secs 5 msec.
--Actualizar el atributo semestre_muerte discretizando el atributo mes_muerte en 2 grupos de meses, el grupo 0 para
--Desconocido, el grupo 3 (No aplica) para los registros 0 (estado vital VIVO) y el grupo 4 (Sin información) para los
--registros nulos (estados vitales PERDIDO EN EL SEGUIMIENTO y DESCONOCIDO), de acuerdo a la siguiente relacion:
--0 -> Desconocido -> mes 99
--1 -> Semestre 1 (de enero a junio) -> meses 1 al 6
--2 -> Semestre 2 (de julio a diciembre) -> meses 7 al 12
--3 -> No aplica -> mes 0
--4 -> Sin información -> mes nulo
BEGIN;
UPDATE etl.tumores_pasto_transformado SET semestre_muerte = CASE
WHEN mes_muerte = 99 THEN 'Desconocido'
WHEN mes_muerte BETWEEN 1 AND 6 THEN 'Semestre 1 (de enero a junio)'
WHEN mes_muerte BETWEEN 7 AND 12 THEN 'Semestre 2 (de julio a diciembre)'
WHEN mes_muerte = 0 THEN 'No aplica'
ELSE 'Sin información' END;
COMMIT;
--UPDATE 41345
--Query returned successfully in 1 secs 134 msec.
--COMMIT
--Query returned successfully in 1 secs 282 msec.
--Actualizar el atributo semestre_ultimo_contacto discretizando el atributo mes_ultimo_contacto en 2 grupos de meses
--y el grupo 0 para Desconocido, de acuerdo a la siguiente relacion:
--0 -> Desconocido -> mes 99
--1 -> Semestre 1 (de enero a junio) -> meses 1 al 6
--2 -> Semestre 2 (de julio a diciembre) -> meses 7 al 12
BEGIN;
UPDATE etl.tumores_pasto_transformado SET semestre_ultimo_contacto = CASE
WHEN mes_ultimo_contacto = 99 THEN 'Desconocido'
WHEN mes_ultimo_contacto BETWEEN 1 AND 6 THEN 'Semestre 1 (de enero a junio)'
WHEN mes_ultimo_contacto BETWEEN 7 AND 12 THEN 'Semestre 2 (de julio a diciembre)'
```

```
ELSE 'Desconocido' END;
COMMIT;
--UPDATE 41345
--Query returned successfully in 742 msec.
--COMMIT
--Query returned successfully in 1 secs 217 msec.
-----------------------------------------------------------------------------------------------------------------------------------------
--34. Campo cod_fuente
--Modificar el tipo de dato a texto
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN cod_fuente SET DATA TYPE TEXT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 1 secs 246 msec.
--COMMIT
--Query returned successfully in 101 msec.
-----------------------------------------------------------------------------------------------------------------------------------------
--35. Campo fuente
--Modificar el tipo de dato a texto
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN fuente SET DATA TYPE TEXT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 128 msec.
--COMMIT
--Query returned successfully in 145 msec.
-----------------------------------------------------------------------------------------------------------------------------------------
--36. Campo cod_tratamiento
--Modificar el tipo de dato a SMALLINT
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN cod_tratamiento SET DATA TYPE
SMALLINT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 1 secs 19 msec.
--COMMIT
--Query returned successfully in 439 msec.
-----------------------------------------------------------------------------------------------------------------------------------------
--37. Campo tratamiento
--Modificar el tipo de dato a texto
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN tratamiento SET DATA TYPE TEXT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 115 msec.
--COMMIT
--Query returned successfully in 67 msec.
-----------------------------------------------------------------------------------------------------------------------------------------
--38. Campo cod_estado_enfermedad
--Modificar el tipo de dato a char(1)
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN cod_estado_enfermedad SET DATA
TYPE CHAR(1);
COMMIT;
--ALTER TABLE
--Query returned successfully in 1 secs 40 msec.
```

```
--COMMIT
--Query returned successfully in 492 msec.
-----------------------------------------------------------------------------------------------------------------------
--39. Campo estado_enfermedad
--Modificar el tipo de dato a texto
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN estado_enfermedad SET DATA
TYPE TEXT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 122 msec.
--COMMIT
--Query returned successfully in 75 msec.
-----------------------------------------------------------------------------------------------------------------------
--40. Campo fecha_tratamiento
--Modificar el tipo de dato a texto
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN fecha_tratamiento SET DATA TYPE
TEXT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 119 msec.
--COMMIT
--Query returned successfully in 120 msec.
--Crear los atributos
cod_dia_tratamiento,dia_tratamiento,cod_mes_tratamiento,mes_tratamiento,ano_tratamiento de
--tipo SMALLINT
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ADD cod_dia_tratamiento SMALLINT;
ALTER TABLE etl.tumores_pasto_transformado ADD dia_tratamiento SMALLINT;
ALTER TABLE etl.tumores_pasto_transformado ADD cod_mes_tratamiento SMALLINT;
ALTER TABLE etl.tumores_pasto_transformado ADD mes_tratamiento SMALLINT;
ALTER TABLE etl.tumores_pasto_transformado ADD ano_tratamiento SMALLINT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 88 msec.
--COMMIT
--Query returned successfully in 170 msec.
--Separar la fecha de tratamiento en los campos creados
BEGIN;
UPDATE etl.tumores_pasto_transformado SET
ano_tratamiento = CAST(SUBSTRING(fecha_tratamiento FROM 1 FOR 4) AS SMALLINT),
cod_mes_tratamiento = CAST(SUBSTRING(fecha_tratamiento FROM 5 FOR 2) AS SMALLINT),
mes_tratamiento = CAST(SUBSTRING(fecha_tratamiento FROM 5 FOR 2) AS SMALLINT),
cod_dia_tratamiento = CAST(SUBSTRING(fecha_tratamiento FROM 7 FOR 2) AS SMALLINT),
dia_tratamiento = CAST(SUBSTRING(fecha_tratamiento FROM 7 FOR 2) AS SMALLINT);
COMMIT;
--UPDATE 41345
--Query returned successfully in 1 secs 57 msec.
--COMMIT
--Query returned successfully in 866 msec.
--Reemplazar el valor 99 del atributo cod_dia_tratamiento por el valor de 32 estableciendo un orden para el
--atributo dia_tratamiento
BEGIN;
UPDATE etl.tumores_pasto_transformado SET cod_dia_tratamiento = 32 WHERE cod_dia_tratamiento =
99;
```

```
COMMIT;
--UPDATE 32467
--Query returned successfully in 771 msec.
--COMMIT
--Query returned successfully in 615 msec.
--Reemplazar el valor 99 del atributo cod_mes_tratamiento por el valor de 13 estableciendo un orden para el
--atributo mes_tratamiento
BEGIN;
UPDATE etl.tumores_pasto_transformado SET cod_mes_tratamiento = 13 WHERE cod_mes_tratamiento
= 99;
COMMIT;
--UPDATE 31767
--Query returned successfully in 711 msec.
--COMMIT
--Query returned successfully in 405 msec.
--Crear los atributos trimestre_tratamiento y semestre_tratamiento de tipo TEXT
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ADD trimestre_tratamiento TEXT;
ALTER TABLE etl.tumores_pasto_transformado ADD semestre_tratamiento TEXT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 74 msec.
--COMMIT
--Query returned successfully in 146 msec.
--Actualizar el atributo trimestre_tratamiento discretizando el atributo mes_tratamiento en 4 grupos de meses
y el
--grupo 0 para Desconocido, de acuerdo a la siguiente relacion:
--0 -> Desconocido -> mes 99
--1 -> Trimestre 1 (de enero a marzo) -> meses 1 al 3
--2 -> Trimestre 2 (de abril a junio) -> meses 4 al 6
--3 -> Trimestre 3 (de julio a septiembre) -> meses 7 al 9
--4 -> Trimestre 4 (de octubre a diciembre) -> meses 10 al 12
BEGIN;
UPDATE etl.tumores_pasto_transformado SET trimestre_tratamiento = CASE
WHEN mes_tratamiento = 99 THEN 'Desconocido'
WHEN mes_tratamiento BETWEEN 1 AND 3 THEN 'Trimestre 1 (de enero a marzo)'
WHEN mes_tratamiento BETWEEN 4 AND 6 THEN 'Trimestre 2 (de abril a junio)'
WHEN mes_tratamiento BETWEEN 7 AND 9 THEN 'Trimestre 3 (de julio a septiembre)'
WHEN mes_tratamiento BETWEEN 10 AND 12 THEN 'Trimestre 4 (de octubre a diciembre)'
ELSE 'Desconocido' END;
COMMIT;
--UPDATE 41345
--Query returned successfully in 2 secs 43 msec.
--COMMIT
--Query returned successfully in 1 secs 305 msec.
--Actualizar el atributo semestre_tratamiento discretizando el atributo mes_tratamiento en 2 grupos de meses
y el
--grupo 0 para Desconocido, de acuerdo a la siguiente relacion:
--0 -> Desconocido -> mes 99
--1 -> Semestre 1 (de enero a junio) -> meses 1 al 6
--2 -> Semestre 2 (de julio a diciembre) -> meses 7 al 12
BEGIN;
UPDATE etl.tumores_pasto_transformado SET semestre_tratamiento = CASE
WHEN mes_tratamiento = 99 THEN 'Desconocido'
WHEN mes_tratamiento BETWEEN 1 AND 6 THEN 'Semestre 1 (de enero a junio)'
WHEN mes_tratamiento BETWEEN 7 AND 12 THEN 'Semestre 2 (de julio a diciembre)'
```

```
ELSE 'Desconocido' END;
COMMIT;
--UPDATE 41345
--Query returned successfully in 1 secs 271 msec.
--COMMIT
--Query returned successfully in 1 secs 495 msec.
-----------------------------------------------------------------------------------------------------------------------------------------
--41. Campo orden_tratamiento
--Modificar el tipo de dato a char(1)
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN orden_tratamiento SET DATA TYPE
CHAR(1);
COMMIT;
--ALTER TABLE
--Query returned successfully in 829 msec.
--COMMIT
--Query returned successfully in 713 msec.
-----------------------------------------------------------------------------------------------------------------------------------------
--42. Campo cod_departamento
--Modificar el tipo de dato a texto
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN cod_departamento SET DATA TYPE
TEXT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 1 secs 17 msec.
--COMMIT
--Query returned successfully in 718 msec.
-----------------------------------------------------------------------------------------------------------------------------------------
--43. Campo departamento
--Modificar el tipo de dato a texto
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN departamento SET DATA TYPE
TEXT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 119 msec.
--COMMIT
--Query returned successfully in 81 msec.
-----------------------------------------------------------------------------------------------------------------------------------------
--44. Campo cod_municipio
--Modificar el tipo de dato a texto
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN cod_municipio SET DATA TYPE
TEXT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 1 secs 143 msec.
--COMMIT
--Query returned successfully in 982 msec.
-----------------------------------------------------------------------------------------------------------------------------------------
--45. Campo municipio
--Modificar el tipo de dato a texto
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN municipio SET DATA TYPE TEXT;
COMMIT;
```

```
--ALTER TABLE
--Query returned successfully in 125 msec.
--COMMIT
--Query returned successfully in 120 msec.
-------------------------------------------------------------------------------------------------------------------------------
--46. Campo cod_zona
--Modificar el tipo de dato a char(1)
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN cod_zona SET DATA TYPE CHAR(1);
COMMIT;
--ALTER TABLE
--Query returned successfully in 590 msec.
--COMMIT
--Query returned successfully in 1 secs 48 msec.
-------------------------------------------------------------------------------------------------------------------------------
--47. Campo zona
--Modificar el tipo de dato a texto
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN zona SET DATA TYPE TEXT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 180 msec.
--COMMIT
--Query returned successfully in 130 msec.
--Estandarizar el atributo a mayusculas
BEGIN;
UPDATE etl.tumores_pasto_limpio SET zona = UPPER(zona);
COMMIT;
--UPDATE 41345
--Query returned successfully in 498 msec.
--COMMIT
--Query returned successfully in 1 secs 644 msec.
-------------------------------------------------------------------------------------------------------------------------------
--48. Campo cod_barrio_comuna
--Modificar el tipo de dato a texto
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN cod_barrio_comuna SET DATA TYPE TEXT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 180 msec.
--COMMIT
--Query returned successfully in 732 msec.
-------------------------------------------------------------------------------------------------------------------------------
--49. Campo barrio_comuna
--Modificar el tipo de dato a texto
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN barrio_comuna SET DATA TYPE TEXT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 184 msec.
--COMMIT
--Query returned successfully in 890 msec.
--Crear los atributos barrio_vereda y comuna_corregimiento de tipo TEXT
```

```
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ADD barrio_vereda TEXT;
ALTER TABLE etl.tumores_pasto_transformado ADD comuna_corregimiento TEXT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 142 msec.
--COMMIT
--Query returned successfully in 172 msec.
--Actualizar el atributo barrio_vereda a partir del campo barrio_comuna dependiendo de la zona, acorde a la siguiente
--relacion:
--1. zona DESCONOCIDA -> barrio_vereda = Barrio o Vereda desconocido
--2. zona URBANA -> barrio_vereda = barrio_comuna
--3. zona RURAL -> barrio_vereda extrae el nombre de la Vereda del atributo barrio_comuna
BEGIN;
UPDATE etl.tumores_pasto_transformado SET barrio_vereda = CASE
WHEN zona = 'DESCONOCIDA' THEN 'Barrio o Vereda desconocido'
WHEN zona = 'URBANA' THEN barrio_comuna
WHEN zona = 'RURAL' THEN TRIM(SUBSTRING(barrio_comuna FROM 1 FOR POSITION('(' IN barrio_comuna) - 1))
ELSE 'Barrio o Vereda desconocido' END;
COMMIT;
--UPDATE 41345
--Query returned successfully in 1 secs 142 msec.
--COMMIT
--Query returned successfully in 1 secs 105 msec.
--Actualizar el valor 'Rural desconocido' a 'Vereda desconocida' del atributo barrio_vereda
BEGIN;
UPDATE etl.tumores_pasto_transformado SET
barrio_vereda = 'Vereda desconocida' WHERE barrio_vereda = 'Rural desconocido';
COMMIT;
--UPDATE 93
--Query returned successfully in 76 msec.
--COMMIT
--Query returned successfully in 73 msec.
--Actualizar el valor 'Urbano desconocido' a 'Barrio desconocido' del atributo barrio_vereda
BEGIN;
UPDATE etl.tumores_pasto_transformado SET
barrio_vereda = 'Barrio desconocido' WHERE barrio_vereda = 'Urbano desconocido';
COMMIT;
--UPDATE 121
--Query returned successfully in 82 msec.
--COMMIT
--Query returned successfully in 75 msec.
--Actualizar el atributo comuna_corregimiento a partir de los campos cod_comuna_corregimiento y barrio_comuna
--dependiendo de la zona, acorde a la siguiente relacion:
--1. zona DESCONOCIDA -> comuna_corregimiento = Comuna o Corregimiento desconocido
--2. zona URBANA -> comuna_corregimiento = cod_comuna_corregimiento
--3. zona RURAL -> comuna_corregimiento extrae el nombre del Corregimiento del atributo barrio_comuna (esta entre
--parentesis)
BEGIN;
UPDATE etl.tumores_pasto_transformado SET comuna_corregimiento = CASE
WHEN zona = 'DESCONOCIDA' THEN 'Desconocido'
WHEN zona = 'URBANA' THEN cod_comuna_corregimiento
```

```
WHEN zona = 'RURAL' THEN TRIM(SUBSTRING(barrio_comuna FROM '\((.*?)\)'))
ELSE 'Desconocido' END;
COMMIT;
--UPDATE 41345
--Query returned successfully in 1 secs 607 msec.
--COMMIT
--Query returned successfully in 902 msec.
--Actualizar los valores del atributo comuna_corregimiento dependiendo del tipo de dato (numero o texto) por su
--descripcion acorde a la siguiente relacion y teniendo en cuenta de excluir la zona DESCONOCIDA puesto que para este
--caso el atributo conserva el valor de Desconocido:
--99 -> Comuna desconocida
--Desconocido y Zona RURAL -> Corregimiento desconocido
--Desconocido y Zona DESCONOCIDA -> Comuna o Corregimiento desconocido
--Numeros de comuna -> Concatenar la palabra Comuna antes del numero de la comuna
--Corregimiento -> Concatenar la palabra Corregimiento antes del nombre del corregimiento (comuna_corregimiento)
BEGIN;
UPDATE etl.tumores_pasto_transformado SET comuna_corregimiento = CASE
WHEN comuna_corregimiento = '99' THEN 'Comuna desconocida'
WHEN comuna_corregimiento = 'Desconocido' AND zona = 'RURAL' THEN 'Corregimiento desconocido'
WHEN comuna_corregimiento = 'Desconocido' AND zona = 'DESCONOCIDA' THEN 'Comuna o Corregimiento desconocido'
WHEN comuna_corregimiento ~ '^[0-9]+$' THEN 'Comuna ' || comuna_corregimiento
ELSE 'Corregimiento ' || comuna_corregimiento END;
COMMIT;
--UPDATE 41345
--Query returned successfully in 1 secs 257 msec.
--COMMIT
--Query returned successfully in 1 secs 28 msec.
-----------------------------------------------------------------------------------------------------------------------------------
--50. Campo cod_comuna_corregimiento
--Modificar el tipo de dato a texto
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN cod_comuna_corregimiento SET DATA TYPE TEXT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 1 secs 50 msec.
--COMMIT
--Query returned successfully in 1 secs 136 msec.
-----------------------------------------------------------------------------------------------------------------------------------
--51. Campo cod_barrio_vereda
--Modificar el tipo de dato a texto
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN cod_barrio_vereda SET DATA TYPE TEXT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 1 secs 3 msec.
--COMMIT
--Query returned successfully in 819 msec.
```

```
--
================================================================================
================================================================================
--------------------------------------------------------------------------------------------------------------------------
--TRANSFORMACIONES REFERENTES A INDICADORES
--------------------------------------------------------------------------------------------------------------------------
--52. Campo fi (frecuencia absoluta de incidencia) que tambien representa la cantidad de casos de cancer
--Crear el atributo fi de tipo INTEGER
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ADD fi INTEGER;
COMMIT;
--ALTER TABLE
--Query returned successfully in 134 msec.
--COMMIT
--Query returned successfully in 118 msec.
--Actualizar el atributo fi con valor de 1 para todos los casos de cancer, es decir todos los diferentes tumores
--registrados sin tener en cuenta las fuentes asociadas a cada tumor
BEGIN;
UPDATE etl.tumores_pasto_transformado AS t1 SET fi = CASE
WHEN t2.id = 1 THEN 1 ELSE 0 END
FROM (SELECT ctid, ROW_NUMBER() OVER (PARTITION BY tumor ORDER BY ctid) AS id FROM
etl.tumores_pasto_transformado) AS t2
WHERE t1.ctid = t2.ctid;
COMMIT;
--UPDATE 41345
--Query returned successfully in 1 secs 940 msec.
--COMMIT
--Query returned successfully in 2 secs 3 msec.
--------------------------------------------------------------------------------------------------------------------------
--53. Campo proy_poblacion_mundial_estandar (proyeccion de poblacion mundial estandar)
--Crear el atributo proy_poblacion_mundial_estandar de tipo INTEGER
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ADD proy_poblacion_mundial_estandar INTEGER;
COMMIT;
--ALTER TABLE
--Query returned successfully in 150 msec.
--COMMIT
--Query returned successfully in 87 msec.
--Actualizar para todos los casos (variable fi en 1) el atributo proy_poblacion_mundial_estandar a partir del
valor
--de proyeccion de poblacion mundial estandar (poblacion) registrado en la tabla auxiliar
(etl.tab_aux_pobmundialestandar)
--dependiendo del grupo etario
BEGIN;
UPDATE etl.tumores_pasto_transformado t1 SET proy_poblacion_mundial_estandar = t2.poblacion
FROM etl.tab_aux_pobmundialestandar t2 WHERE t1.grupo_etario1 = t2.grupo_etario1;
COMMIT;
--UPDATE 41345
--Query returned successfully in 983 msec.
--COMMIT
--Query returned successfully in 989 msec.
--------------------------------------------------------------------------------------------------------------------------
--54. Campo cantidad_pacientes
--Crear el atributo cantidad_pacientes de tipo INTEGER
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ADD cantidad_pacientes INTEGER;
```

```
COMMIT;
--ALTER TABLE
--Query returned successfully in 142 msec.
--COMMIT
--Query returned successfully in 75 msec.
--Actualizar para todos los casos (variable fi en 1) el atributo cantidad_pacientes con valor de 1 para todos los
--diferentes pacientes registrados sin tener en cuenta el tumor ni las fuentes asociadas a cada tumor
BEGIN;
UPDATE etl.tumores_pasto_transformado AS t1 SET cantidad_pacientes = CASE
WHEN t2.id = 1 THEN 1 ELSE 0 END
FROM (SELECT ctid, ROW_NUMBER() OVER (PARTITION BY paciente ORDER BY ctid) AS id FROM etl.tumores_pasto_transformado) AS t2
WHERE t1.ctid = t2.ctid;
COMMIT;
--UPDATE 41345
--Query returned successfully in 1 secs 532 msec.
--COMMIT
--Query returned successfully in 1 secs 323 msec.
-----------------------------------------------------------------------------------------------------------------------------------
--55. Campo cantidad_caso_fuente que representa la cantidad de reportes que han hecho las diferentes fuentes de los
--diferentes casos de cancer
--Crear el atributo cantidad_caso_fuente de tipo INTEGER
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ADD cantidad_caso_fuente INTEGER;
COMMIT;
--ALTER TABLE
--Query returned successfully in 72 msec.
--COMMIT
--Query returned successfully in 94 msec.
--Actualizar el atributo cantidad_caso_fuente con valor de 1 para todos los registros de la tabla que representan
--los reportes de diferentes fuentes asociadas a cada caso de cáncer
BEGIN;
UPDATE etl.tumores_pasto_transformado SET cantidad_caso_fuente = 1;
COMMIT;
--UPDATE 41345
--Query returned successfully in 1 secs 213 msec.
--COMMIT
--Query returned successfully in 1 secs 358 msec.
-----------------------------------------------------------------------------------------------------------------------------------
--56. Campo cdmv (casos con un diagnostico morfologicamente verificado)
--Crear el atributo cdmv de tipo INTEGER
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ADD cdmv INTEGER;
COMMIT;
--ALTER TABLE
--Query returned successfully in 158 msec.
--COMMIT
--Query returned successfully in 75 msec.
--Actualizar el atributo cdmv con valor de 1 para aquellos casos (variable fi en 1) en los que el campo
--cod_base_diagnostico es igual a 5 (citologia y hematologia) o 7 (histologia de un tumor primario)
BEGIN;
UPDATE etl.tumores_pasto_transformado SET cdmv = CASE
WHEN fi = 1 AND cod_base_diagnostico IN ('5','7') THEN 1 ELSE 0 END;
```

```
COMMIT;
--UPDATE 41345
--Query returned successfully in 978 msec.
--COMMIT
--Query returned successfully in 1 secs 662 msec.
-----------------------------------------------------------------------------------------------------------------------------
--57. Campo cscd (casos conocidos solo por certificado de defuncion)
--Crear el atributo cscd de tipo INTEGER
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ADD cscd INTEGER;
COMMIT;
--ALTER TABLE
--Query returned successfully in 156 msec.
--COMMIT
--Query returned successfully in 85 msec.
--Actualizar el atributo cscd con valor de 1 para aquellos casos (variable fi en 1) en los que el campo
--cod_base_diagnostico es igual a 0 (solo certificado de defuncion)
BEGIN;
UPDATE etl.tumores_pasto_transformado SET cscd = CASE
WHEN fi = 1 AND cod_base_diagnostico = '0' THEN 1 ELSE 0 END;
COMMIT;
--UPDATE 41345
--Query returned successfully in 2 secs 109 msec.
--COMMIT
--Query returned successfully in 1 secs 70 msec.
-----------------------------------------------------------------------------------------------------------------------------
--58. Campo mccpm (muertes por cancer que fueron certificadas por personal medico)
--Crear el atributo mccpm de tipo INTEGER
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ADD mccpm INTEGER;
COMMIT;
--ALTER TABLE
--Query returned successfully in 92 msec.
--COMMIT
--Query returned successfully in 75 msec.
--Actualizar el atributo mccpm con valor de 1 para aquellos casos (variable fi en 1) en estado vital MUERTO
(codigo 2)
--en los que el campo cod_profesion_expedidor es igual a 1 (medico tratante), 2 (medico no tratante) y 3
(medico
--legista)
BEGIN;
UPDATE etl.tumores_pasto_transformado SET mccpm = CASE
WHEN fi = 1 AND cod_estado_vital = '2' AND cod_profesion_expedidor IN ('1','2','3') THEN 1 ELSE 0
END;
COMMIT;
--UPDATE 41345
--Query returned successfully in 1 secs 355 msec.
--COMMIT
--Query returned successfully in 1 secs 8 msec.
-----------------------------------------------------------------------------------------------------------------------------
--59. Campo ccopd (casos de cancer de origen primario desconocido - localizacion C809)
--Crear el atributo ccopd de tipo INTEGER
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ADD ccopd INTEGER;
COMMIT;
--ALTER TABLE
```

```
--Query returned successfully in 91 msec.
--COMMIT
--Query returned successfully in 67 msec.
--Actualizar el atributo ccopd con valor de 1 para aquellos casos (variable fi en 1) en los que el campo cod_localizacion
--es igual a 809 (C80.9  Sitio primario desconocido)
BEGIN;
UPDATE etl.tumores_pasto_transformado SET ccopd = CASE
WHEN fi = 1 AND cod_localizacion = '809' THEN 1 ELSE 0 END;
COMMIT;
--UPDATE 41345
--Query returned successfully in 2 secs 283 msec.
--COMMIT
--Query returned successfully in 1 secs 41 msec.
-----------------------------------------------------------------------------------------------------------------------------
--60. Campo mcopm (muertes de cancer de origen primario desconocido - localizacion C809)
--Crear el atributo mcopm de tipo INTEGER
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ADD mcopm INTEGER;
COMMIT;
--ALTER TABLE
--Query returned successfully in 93 msec.
--COMMIT
--Query returned successfully in 72 msec.
--Actualizar el atributo mcopm con valor de 1 para aquellos casos (variable fi en 1) en estado vital MUERTO (codigo 2)
--en los que el campo cod_localizacion es igual a 809 (C80.9  Sitio primario desconocido)
BEGIN;
UPDATE etl.tumores_pasto_transformado SET mcopm = CASE
WHEN fi = 1 AND cod_estado_vital = '2' AND cod_localizacion = '809' THEN 1 ELSE 0 END;
COMMIT;
--UPDATE 41345
--Query returned successfully in 1 secs 310 msec.
--COMMIT
--Query returned successfully in 1 secs 34 msec.
-----------------------------------------------------------------------------------------------------------------------------
--61. Campo edad_muerte
--Crear el atributo edad_muerte de tipo SMALLINT
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ADD edad_muerte SMALLINT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 166 msec.
--COMMIT
--Query returned successfully in 70 msec.
--Actualizar el atributo edad_muerte con el calculo de la edad para los registros con estado_vital igual a 'MUERTO',
--incluyendo el valor 999 para edad DESCONOCIDA y con el valor de 997 para los registros con estado_vital igual a
--'VIVO' y el valor de 998 para otros estados vitales ('PERDIDO EN EL SEGUIMIENTO' y 'DESCONOCIDO')
BEGIN;
UPDATE etl.tumores_pasto_transformado SET edad_muerte = CASE
WHEN estado_vital = 'MUERTO' THEN (
CASE
        WHEN fecha_ultimo_contacto = '99-99-9999' THEN 999
```

```
        WHEN fecha_nacimiento = '99-99-9999' THEN (
                CASE
                        WHEN edad_paciente = 999 THEN 999
                        ELSE (
                                CASE
                                        WHEN fecha_diagnostico LIKE '99-99-%' OR
fecha_ultimo_contacto LIKE '99-99-%' THEN
                                                edad_paciente + ano_muerte - ano_diagnostico
                                        WHEN fecha_diagnostico LIKE '99-%' OR
fecha_ultimo_contacto LIKE '99-%' OR fecha_diagnostico IS NULL OR fecha_ultimo_contacto IS NULL
THEN
                                                CASE
                                                        WHEN ano_diagnostico <= ano_muerte
AND mes_diagnostico <= mes_muerte THEN
                                                                edad_paciente + ano_muerte -
ano_diagnostico
                                                        WHEN ano_diagnostico < ano_muerte AND
mes_diagnostico > mes_muerte THEN
                                                                edad_paciente + ano_muerte -
ano_diagnostico - 1
                                                        ELSE 999
                                                END
                                        ELSE
                                edad_paciente + EXTRACT(YEAR FROM
AGE(fecha_ultimo_contacto::timestamp,fecha_diagnostico::timestamp::DATE))
                                END
                        )
                END
        )
        WHEN fecha_nacimiento LIKE '99-99-%' OR fecha_ultimo_contacto LIKE '99-99-%' THEN
                ano_muerte - CAST(SUBSTRING(fecha_nacimiento FROM 7 FOR 10) AS INTEGER)
        WHEN fecha_nacimiento LIKE '99-%' OR fecha_ultimo_contacto LIKE '99-%' OR
fecha_nacimiento IS NULL OR fecha_ultimo_contacto IS NULL THEN
                CASE
                        WHEN CAST(SUBSTRING(fecha_nacimiento FROM 7 FOR 10) AS
INTEGER) <= ano_muerte
                                AND CAST(SUBSTRING(fecha_nacimiento FROM 4 FOR 2) AS
INTEGER) <= mes_muerte THEN
                                ano_muerte - CAST(SUBSTRING(fecha_nacimiento FROM 7 FOR 10)
AS INTEGER)
                        WHEN CAST(SUBSTRING(fecha_nacimiento FROM 7 FOR 10) AS
INTEGER) < ano_muerte
                                AND CAST(SUBSTRING(fecha_nacimiento FROM 4 FOR 2) AS
INTEGER) > mes_muerte THEN
                                ano_muerte - CAST(SUBSTRING(fecha_nacimiento FROM 7 FOR 10)
AS INTEGER) - 1
                        ELSE 999
                END
        ELSE
                EXTRACT(YEAR FROM
AGE(fecha_ultimo_contacto::timestamp,fecha_nacimiento::timestamp::DATE))
END
)
WHEN estado_vital = 'VIVO' THEN 997
ELSE 998 END;
COMMIT;
```

```
--UPDATE 41345
--Query returned successfully in 2 secs 23 msec.
--COMMIT
--Query returned successfully in 1 secs 572 msec.
-----------------------------------------------------------------------------------------------------------------------------
--62. Campo esperanza_vida (esperanza de vida)
--Crear el atributo esperanza_vida de tipo NUMERIC
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ADD esperanza_vida NUMERIC(4,1);
COMMIT;
--ALTER TABLE
--Query returned successfully in 167 msec.
--COMMIT
--Query returned successfully in 194 msec.
--Actualizar el atributo esperanza_vida con el valor de la esperanza de vida asociada al campo edad_paciente puesto
--que corresponde con la edad de incidencia. Cabe resaltar que tendra un valor de 999.0 para aquellos registros con
--edad Desconocida
BEGIN;
UPDATE etl.tumores_pasto_transformado t1 SET esperanza_vida = t2.esperanza_vida
FROM etl.tab_aux_esperanzadevida t2 WHERE t1.edad_paciente = t2.edad;
COMMIT;
--UPDATE 41345
--Query returned successfully in 1 secs 216 msec.
--COMMIT
--Query returned successfully in 1 secs 606 msec.
-----------------------------------------------------------------------------------------------------------------------------
--63. Campo avmp (años de vida saludable perdidos por muerte prematura)
--Crear el atributo avmp de tipo NUMERIC
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ADD avmp NUMERIC(4,1);
COMMIT;
--ALTER TABLE
--Query returned successfully in 94 msec.
--COMMIT
--Query returned successfully in 162 msec.
--Actualizar el atributo avmp con el calculo de los años acorde a la formula del indicador, la resta entre el valor
--del atributo esperanza_vida sumado el valor del atributo edad_paciente, y el valor del atributo edad_muerte para los
--casos (variable fi en 1) con estado_vital igual a 'MUERTO', incluyendo el valor 999 para edades DESCONOCIDAS y el
--valor de 997 para los registros con estado_vital igual a 'VIVO' y el valor de 998 para otros estados vitales
--('PERDIDO EN EL SEGUIMIENTO' y 'DESCONOCIDO'). Cabe resaltar que para el resto de registros se asigna un valor de
--0.0 que no aporta en nada al valor del indicador
BEGIN;
UPDATE etl.tumores_pasto_transformado SET avmp = CASE
WHEN fi = 1 AND estado_vital = 'MUERTO' THEN
(CASE WHEN edad_muerte = 999 OR esperanza_vida = 999 THEN 999 ELSE edad_paciente + esperanza_vida - edad_muerte END)
WHEN fi = 1 AND estado_vital = 'VIVO' THEN 997
WHEN fi = 1 AND estado_vital NOT IN ('VIVO','MUERTO') THEN 998
ELSE 0 END;
COMMIT;
```

```
--UPDATE 41345
--Query returned successfully in 1 secs 416 msec.
--COMMIT
--Query returned successfully in 1 secs 721 msec.
-----------------------------------------------------------------------------------------------------------------------------------
--64. Campo oac (oportunidad de la atencion en cancer)
--Crear el atributo oac de tipo SMALLINT
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ADD oac SMALLINT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 70 msec.
--COMMIT
--Query returned successfully in 109 msec.
--Actualizar el atributo oac con el calculo de los dias transcurridos entre la fecha de diagnostico y la fecha del
--primer tratamiento acorde a la formula del indicador, la resta entre la fecha del primer tratamiento y la fecha de
--diagnostico para los casos (variable fi en 1), incluyendo el valor de 0 no solo para cuando hayan 0 dias en el
--calculo si no tambien para aquellos registros donde no es posible hacer el calculo porque alguna de las dos fechas
--es totalmente desconocida en el respectivo formato (99-99-9999 fecha de diagnostico y 99999999 fecha de tratamiento)
BEGIN;
UPDATE etl.tumores_pasto_transformado SET oac = CASE
WHEN fi = 1 AND (fecha_diagnostico = '99-99-9999' OR fecha_tratamiento = '99999999') THEN 0
ELSE
        CASE
                WHEN fi = 1 AND (fecha_diagnostico LIKE '99-99-%' OR fecha_tratamiento LIKE '%9999') THEN (ano_tratamiento - ano_diagnostico)*365
                ELSE
                        CASE
                                WHEN fi = 1 AND (fecha_diagnostico LIKE '99-%' OR fecha_tratamiento LIKE '%99') THEN (ano_tratamiento - ano_diagnostico)*365 + (mes_tratamiento - mes_diagnostico)*30
                                ELSE
                                        CASE
                                                WHEN fi = 1 THEN TO_DATE(CONCAT(dia_tratamiento,'-',mes_tratamiento,'-',ano_tratamiento), 'DD-MM-YYYY') - TO_DATE(fecha_diagnostico, 'DD-MM-YYYY')
                                                ELSE 0
                                        END
                        END
                END
END;
COMMIT;
--UPDATE 41345
--Query returned successfully in 2 secs 692 msec.
--COMMIT
--Query returned successfully in 3 secs 34 msec.
-----------------------------------------------------------------------------------------------------------------------------------
--65. Complemento de transformacion de los campos dia_diagnostico y mes_diagnostico posterior al calculo de edad_muerte
--Modificar el tipo de dato a texto para el atributo dia_diagnostico
BEGIN;
```

```
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN dia_diagnostico SET DATA TYPE
TEXT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 841 msec.
--COMMIT
--Query returned successfully in 996 msec.
--Actualizar el atributo dia_diagnostico a DESCONOCIDO para los registros donde el atributo es igual a 99
BEGIN;
UPDATE  etl.tumores_pasto_transformado  SET
dia_diagnostico = 'DESCONOCIDO' WHERE dia_diagnostico::INTEGER = 99;
COMMIT;
--UPDATE 1134
--Query returned successfully in 220 msec.
--COMMIT
--Query returned successfully in 139 msec.
--Modificar el tipo de dato a texto para el atributo mes_diagnostico
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN mes_diagnostico SET DATA TYPE
TEXT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 801 msec.
--COMMIT
--Query returned successfully in 786 msec.
--Actualizar los valores del atributo mes_diagnostico por su descripcion acorde a la siguiente relacion:
--99 -> DESCONOCIDO
--1 -> ENERO
--2 -> FEBRERO
--3 -> MARZO
--4 -> ABRIL
--5 -> MAYO
--6 -> JUNIO
--7 -> JULIO
--8 -> AGOSTO
--9 -> SEPTIEMBRE
--10 -> OCTUBRE
--11 -> NOVIEMBRE
--12 -> DICIEMBRE
BEGIN;
UPDATE etl.tumores_pasto_transformado SET mes_diagnostico = CASE
WHEN mes_diagnostico = '99' THEN 'DESCONOCIDO'
WHEN mes_diagnostico = '1' THEN 'ENERO'
WHEN mes_diagnostico = '2' THEN 'FEBRERO'
WHEN mes_diagnostico = '3' THEN 'MARZO'
WHEN mes_diagnostico = '4' THEN 'ABRIL'
WHEN mes_diagnostico = '5' THEN 'MAYO'
WHEN mes_diagnostico = '6' THEN 'JUNIO'
WHEN mes_diagnostico = '7' THEN 'JULIO'
WHEN mes_diagnostico = '8' THEN 'AGOSTO'
WHEN mes_diagnostico = '9' THEN 'SEPTIEMBRE'
WHEN mes_diagnostico = '10' THEN 'OCTUBRE'
WHEN mes_diagnostico = '11' THEN 'NOVIEMBRE'
WHEN mes_diagnostico = '12' THEN 'DICIEMBRE'
ELSE 'DESCONOCIDO' END;
COMMIT;
```

```
--UPDATE 41345
--Query returned successfully in 1 secs 288 msec.
--COMMIT
--Query returned successfully in 772 msec.
-----------------------------------------------------------------------------------------------------------------------------
--66. Complemento de transformacion de los campos dia_muerte, mes_muerte, dia_ultimo_contacto, mes_ultimo_contacto,
--ano_muerte y ano_ultimo_contacto posterior al calculo de edad_muerte
--Modificar el tipo de dato a texto para los atributos dia_muerte y dia_ultimo_contacto
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN dia_muerte SET DATA TYPE TEXT;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN dia_ultimo_contacto SET DATA TYPE TEXT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 1 secs 396 msec.
--COMMIT
--Query returned successfully in 1 secs 487 msec.
--Actualizar el atributo dia_muerte a DESCONOCIDO para los registros donde el atributo es igual a 99, a NO APLICA para
--los registros donde el atributo es igual a 0 y a SIN INFORMACIÓN para los registros donde el atributo es nulo, de lo
--contrario, conservar su valor
BEGIN;
UPDATE etl.tumores_pasto_transformado SET dia_muerte = CASE
WHEN dia_muerte ~ '^[0-9]+$' AND dia_muerte::INTEGER NOT IN (99,0) THEN dia_muerte
WHEN dia_muerte::INTEGER = 99 THEN 'DESCONOCIDO'
WHEN dia_muerte::INTEGER = 0 THEN 'NO APLICA'
ELSE 'SIN INFORMACIÓN' END;
COMMIT;
--UPDATE 41345
--Query returned successfully in 1 secs 313 msec.
--COMMIT
--Query returned successfully in 987 msec.
--Actualizar el atributo dia_ultimo_contacto a DESCONOCIDO para los registros donde el atributo es igual a 99
BEGIN;
UPDATE etl.tumores_pasto_transformado SET dia_ultimo_contacto = 'DESCONOCIDO' WHERE dia_ultimo_contacto::INTEGER = 99;
COMMIT;
--UPDATE 4277
--Query returned successfully in 232 msec.
--COMMIT
--Query returned successfully in 295 msec.
--Modificar el tipo de dato a texto para los atributos mes_muerte y mes_ultimo_contacto
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN mes_muerte SET DATA TYPE TEXT;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN mes_ultimo_contacto SET DATA TYPE TEXT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 1 secs 605 msec.
--COMMIT
--Query returned successfully in 1 secs 556 msec.
--Actualizar los valores del atributo mes_muerte por su descripcion acorde a la siguiente relacion:
```

```
--99 -> DESCONOCIDO
--0 -> NO APLICA
--1 -> ENERO
--2 -> FEBRERO
--3 -> MARZO
--4 -> ABRIL
--5 -> MAYO
--6 -> JUNIO
--7 -> JULIO
--8 -> AGOSTO
--9 -> SEPTIEMBRE
--10 -> OCTUBRE
--11 -> NOVIEMBRE
--12 -> DICIEMBRE
--null -> SIN INFORMACIÓN
BEGIN;
UPDATE etl.tumores_pasto_transformado SET mes_muerte = CASE
WHEN mes_muerte = '99' THEN 'DESCONOCIDO'
WHEN mes_muerte = '0' THEN 'NO APLICA'
WHEN mes_muerte = '1' THEN 'ENERO'
WHEN mes_muerte = '2' THEN 'FEBRERO'
WHEN mes_muerte = '3' THEN 'MARZO'
WHEN mes_muerte = '4' THEN 'ABRIL'
WHEN mes_muerte = '5' THEN 'MAYO'
WHEN mes_muerte = '6' THEN 'JUNIO'
WHEN mes_muerte = '7' THEN 'JULIO'
WHEN mes_muerte = '8' THEN 'AGOSTO'
WHEN mes_muerte = '9' THEN 'SEPTIEMBRE'
WHEN mes_muerte = '10' THEN 'OCTUBRE'
WHEN mes_muerte = '11' THEN 'NOVIEMBRE'
WHEN mes_muerte = '12' THEN 'DICIEMBRE'
ELSE 'SIN INFORMACIÓN' END;
COMMIT;
--UPDATE 41345
--Query returned successfully in 1 secs 355 msec.
--COMMIT
--Query returned successfully in 992 msec.
--Actualizar los valores del atributo mes_ultimo_contacto por su descripcion acorde a la siguiente relacion:
--99 -> DESCONOCIDO
--1 -> ENERO
--2 -> FEBRERO
--3 -> MARZO
--4 -> ABRIL
--5 -> MAYO
--6 -> JUNIO
--7 -> JULIO
--8 -> AGOSTO
--9 -> SEPTIEMBRE
--10 -> OCTUBRE
--11 -> NOVIEMBRE
--12 -> DICIEMBRE
BEGIN;
UPDATE etl.tumores_pasto_transformado SET mes_ultimo_contacto = CASE
WHEN mes_ultimo_contacto = '99' THEN 'DESCONOCIDO'
WHEN mes_ultimo_contacto = '1' THEN 'ENERO'
WHEN mes_ultimo_contacto = '2' THEN 'FEBRERO'
```

```
WHEN mes_ultimo_contacto = '3' THEN 'MARZO'
WHEN mes_ultimo_contacto = '4' THEN 'ABRIL'
WHEN mes_ultimo_contacto = '5' THEN 'MAYO'
WHEN mes_ultimo_contacto = '6' THEN 'JUNIO'
WHEN mes_ultimo_contacto = '7' THEN 'JULIO'
WHEN mes_ultimo_contacto = '8' THEN 'AGOSTO'
WHEN mes_ultimo_contacto = '9' THEN 'SEPTIEMBRE'
WHEN mes_ultimo_contacto = '10' THEN 'OCTUBRE'
WHEN mes_ultimo_contacto = '11' THEN 'NOVIEMBRE'
WHEN mes_ultimo_contacto = '12' THEN 'DICIEMBRE'
ELSE 'DESCONOCIDO' END;
COMMIT;
--UPDATE 41345
--Query returned successfully in 986 msec.
--COMMIT
--Query returned successfully in 973 msec.
--Modificar el tipo de dato a texto para los atributos ano_muerte y ano_ultimo_contacto
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN ano_muerte SET DATA TYPE TEXT;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN ano_ultimo_contacto SET DATA TYPE TEXT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 1 secs 743 msec.
--COMMIT
--Query returned successfully in 1 secs 643 msec.
--Actualizar los valores 0, 9999 y nulos del atributo ano_muerte por su descripcion acorde a la siguiente relacion:
--0 -> NO APLICA
--9999 -> DESCONOCIDO
--null -> SIN INFORMACIÓN
BEGIN;
UPDATE etl.tumores_pasto_transformado SET ano_muerte = CASE
WHEN ano_muerte = '0' THEN 'NO APLICA'
WHEN ano_muerte = '9999' THEN 'DESCONOCIDO'
WHEN ano_muerte IS NULL THEN 'SIN INFORMACIÓN'
ELSE ano_muerte END;
COMMIT;
--UPDATE 41345
--Query returned successfully in 1 secs 267 msec.
--COMMIT
--Query returned successfully in 1 secs 140 msec.
--Actualizar los valores 9999 del atributo ano_ultimo_contacto a DESCONOCIDO
BEGIN;
UPDATE etl.tumores_pasto_transformado SET ano_ultimo_contacto = 'DESCONOCIDO'
WHERE ano_ultimo_contacto = '9999';
COMMIT;
--UPDATE 2941
--Query returned successfully in 219 msec.
--COMMIT
--Query returned successfully in 117 msec.
-----------------------------------------------------------------------------------------------------------------------------
--67. Complemento de transformacion de los campos dia_tratamiento, mes_tratamiento y ano_tratamiento posterior al
--calculo de oac
```

```
--Modificar el tipo de dato a texto para el atributo dia_tratamiento
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER  COLUMN dia_tratamiento SET  DATA TYPE
TEXT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 1 secs 199 msec.
--COMMIT
--Query returned successfully in 817 msec.
--Actualizar el atributo dia_tratamiento a DESCONOCIDO para los registros donde el atributo es igual a 99
BEGIN;
UPDATE  etl.tumores_pasto_transformado  SET
dia_tratamiento = 'DESCONOCIDO' WHERE dia_tratamiento::INTEGER = 99;
COMMIT;
--UPDATE 32542
--Query returned successfully in 1 secs 165 msec.
--COMMIT
--Query returned successfully in 1 secs 398 msec.
--Modificar el tipo de dato a texto para el atributo mes_tratamiento
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN mes_tratamiento SET DATA TYPE
TEXT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 809 msec.
--COMMIT
--Query returned successfully in 1 secs 377 msec.
--Actualizar los valores del atributo mes_tratamiento por su descripcion acorde a la siguiente relacion:
--99 -> DESCONOCIDO
--1 -> ENERO
--2 -> FEBRERO
--3 -> MARZO
--4 -> ABRIL
--5 -> MAYO
--6 -> JUNIO
--7 -> JULIO
--8 -> AGOSTO
--9 -> SEPTIEMBRE
--10 -> OCTUBRE
--11 -> NOVIEMBRE
--12 -> DICIEMBRE
BEGIN;
UPDATE etl.tumores_pasto_transformado SET mes_tratamiento = CASE
WHEN mes_tratamiento = '99' THEN 'DESCONOCIDO'
WHEN mes_tratamiento = '1' THEN 'ENERO'
WHEN mes_tratamiento = '2' THEN 'FEBRERO'
WHEN mes_tratamiento = '3' THEN 'MARZO'
WHEN mes_tratamiento = '4' THEN 'ABRIL'
WHEN mes_tratamiento = '5' THEN 'MAYO'
WHEN mes_tratamiento = '6' THEN 'JUNIO'
WHEN mes_tratamiento = '7' THEN 'JULIO'
WHEN mes_tratamiento = '8' THEN 'AGOSTO'
WHEN mes_tratamiento = '9' THEN 'SEPTIEMBRE'
WHEN mes_tratamiento = '10' THEN 'OCTUBRE'
WHEN mes_tratamiento = '11' THEN 'NOVIEMBRE'
WHEN mes_tratamiento = '12' THEN 'DICIEMBRE'
```

```
ELSE 'DESCONOCIDO' END;
COMMIT;
--UPDATE 41345
--Query returned successfully in 1 secs 264 msec.
--COMMIT
--Query returned successfully in 2 secs 934 msec.
--Modificar el tipo de dato a texto para el atributo ano_tratamiento
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ALTER COLUMN ano_tratamiento SET DATA  TYPE
TEXT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 1 secs 200 msec.
--COMMIT
--Query returned successfully in 1 secs 376 msec.
--Actualizar los valores 9999 del atributo ano_tratamiento a DESCONOCIDO
BEGIN;
UPDATE etl.tumores_pasto_transformado SET ano_tratamiento = 'DESCONOCIDO'
WHERE ano_tratamiento = '9999';
COMMIT;
--UPDATE 31654
--Query returned successfully in 1 secs 180 msec.
--COMMIT
--Query returned successfully in 1 secs 851 msec.
-------------------------------------------------------------------------------------------------------------------------
--68. Campo cantidad_tumores
--Crear el atributo cantidad_tumores de tipo SMALLINT
BEGIN;
ALTER TABLE etl.tumores_pasto_transformado ADD cantidad_tumores SMALLINT;
COMMIT;
--ALTER TABLE
--Query returned successfully in 71 msec.
--COMMIT
--Query returned successfully in 1 secs 90 msec.
--Actualizar el atributo con la cantidad de tumores que tiene cada paciente
BEGIN;
UPDATE etl.tumores_pasto_transformado t1 SET cantidad_tumores = t2.cantidad_tumores FROM
(SELECT paciente,COUNT(*) cantidad_tumores FROM etl.tumores_pasto_transformado
WHERE fi = 1 GROUP BY paciente ORDER BY 1) t2 WHERE t1.paciente = t2.paciente;
COMMIT;
--UPDATE 41345
--Query returned successfully in 962 msec.
--COMMIT
--Query returned successfully in 907 msec.

-------------------------------------------------------------------------------------------------------------------------
==========================================================================
--TABLA FINAL
--SELECT * FROM etl.tumores_pasto_transformado;
-------------------------------------------------------------------------------------------------------------------------
-==========================================================================

--Con el fin de calcular las Tasas de Incidencia, Especificas por Edad y Acumuladas es necesario Transformar
la
--tabla que contiene los registros de Estadisticas de Proyecciones de Poblacion del DANE
--(etl.tab_dane_pobpasto_base)
```

```
---------------------------------------------------------------------------------------------------------------------------------------
--Crear la tabla tab_dane_pobpasto a partir de la tabla tab_dane_pobpasto_base.
BEGIN;
CREATE TABLE etl.tab_dane_pobpasto AS SELECT * FROM etl.tab_dane_pobpasto_base;
COMMIT;
--SELECT 13311
--Query returned successfully in 412 msec.
--COMMIT
--Query returned successfully in 625 msec.
---------------------------------------------------------------------------------------------------------------------------------------
--TRANSFORMACION DE LA TABLA DE PROYECCIONES DE POBLACION DEL DANE
---------------------------------------------------------------------------------------------------------------------------------------
--1. Se eliminan los registros totalizados por sexo y edad donde los campos sexo y/o edad toman el valor de 'X',
--con el fin de conservar solo los valores de proyeccion individuales por edad, sexo y año
BEGIN;
DELETE FROM etl.tab_dane_pobpasto WHERE sexo = 'X' OR edad = 'X';
COMMIT;
--DELETE 4539
--Query returned successfully in 222 msec.
--COMMIT
--Query returned successfully in 212 msec.
---------------------------------------------------------------------------------------------------------------------------------------
--2. Adicionar registros faltantes a la tabla etl.tab_dane_pobpasto
--Teniendo en cuenta que en los casos del RPCMP hay edades superiores a los 85 años y en las tablas de proyeccion
--de poblacion se registra la expresion '85 y mas', es necesario independizar los registros con los mismos valores
--asignados para el registro de '85 y mas' para los casos con edades superiores a los 85 años y bajo el respectivo
--sexo y año de diagnostico
BEGIN;
INSERT INTO etl.tab_dane_pobpasto (ano,sexo,edad,vr_cab_muni,vr_cent_pob_rural,vr_total)
SELECT DISTINCT ano_diagnostico ano,t1.sexo sexo,edad_paciente edad,vr_cab_muni,vr_cent_pob_rural,vr_total
FROM etl.tumores_pasto_transformado t1 LEFT JOIN etl.tab_dane_pobpasto t2 ON
ano_diagnostico = ano AND t1.sexo = t2.sexo AND t2.edad = '85 y más' WHERE edad_paciente > 85
ORDER BY ano_diagnostico,sexo,edad_paciente;
COMMIT;
--INSERT 0 394
--Query returned successfully in 310 msec.
--COMMIT
--Query returned successfully in 438 msec.
---------------------------------------------------------------------------------------------------------------------------------------
--3. Campo edad
--Actualizar los valores '85 y más' del atributo edad a 85
BEGIN;
UPDATE etl.tab_dane_pobpasto SET edad = '85' WHERE edad = '85 y más';
COMMIT;
--UPDATE 102
--Query returned successfully in 146 msec.
--COMMIT
--Query returned successfully in 111 msec.
---------------------------------------------------------------------------------------------------------------------------------------
--4. Campos de valores de proyeccion (vr_cab_muni, vr_cent_pob_rural y vr_total)
--Actualizar los valores de vr_cab_muni, vr_cent_pob_rural y vr_total a 0 para los registros donde la edad es
```

```
--desconocida (999)
BEGIN;
UPDATE etl.tab_dane_pobpasto SET vr_cab_muni = 0 WHERE edad = '999';
UPDATE etl.tab_dane_pobpasto SET vr_cent_pob_rural = 0 WHERE edad = '999';
UPDATE etl.tab_dane_pobpasto SET vr_total = 0 WHERE edad = '999';
COMMIT;
--UPDATE 4
--Query returned successfully in 187 msec.
--COMMIT
--Query returned successfully in 102 msec.
-----------------------------------------------------------------------------------------------------------------------
--5. Eliminar registros sobrantes de la tabla etl.tab_dane_pobpasto
--Teniendo en cuenta que en la tabla de proyecciones de poblacion del DANE se registran años para los que
no se
--encuentran casos del RPCMP (años futuros por ejemplo), es necesario eliminar estos registros para evitar
incluir
--estos valores en los calculos de proyecciones de poblacion
BEGIN;
DELETE FROM etl.tab_dane_pobpasto WHERE ano NOT IN
(SELECT DISTINCT(ano_diagnostico) FROM etl.tumores_pasto_transformado);
COMMIT;
--DELETE 2064
--Query returned successfully in 167 msec.
--COMMIT
--Query returned successfully in 171 msec.

-----------------------------------------------------------------------------------------------------------------------
=======================================================================
--TABLA FINAL
SELECT * FROM etl.tab_dane_pobpasto;
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
```

Anexo 5. Script empleado en la implementación del Modelo Estrella de la Bodega de Datos en PostgreSQL.

A continuación, se presenta el código SQL que permite crear las tablas de Dimensiones y la tabla de Hechos que conforman la Bodega de Datos diseñada.

```
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

=======================================================================
--BODEGA DE DATOS SISTEMA YACHAY - RPCMP
--REGISTRO POBLACIONAL DE CANCER DEL MUNICIPIO DE PASTO
--
=======================================================================
--MODELO MULTIDIMENSIONAL DE LOS TUMORES-FUENTE DEL MUNICIPIO DE PASTO
--DEFINICION DEL MODELO EN ESTRELLA
--
=======================================================================
-----------------------------------------------------------------------------------------------------------------------
--DIMENSION TUMORES
CREATE TABLE dw.dim_tumores(
id_dimtumor SERIAL NOT NULL PRIMARY KEY,
cod_tumor INT,
fecha_incidencia TEXT,
comportamiento TEXT,
```

```
lateralidad TEXT,
grado_diferenciacion TEXT,
base_diagnostico TEXT,
tratamiento TEXT,
estado_enfermedad TEXT);
--CREATE TABLE
--Query returned successfully in 118 msec.
-----------------------------------------------------------------------------------------------------------------------------------------------------
--DIMENSION EDAD
CREATE TABLE dw.dim_edades(
id_dimedad SERIAL NOT NULL PRIMARY KEY,
edad SMALLINT,
descripcion_edad TEXT,
cod_grupo_etario1 SMALLINT,
grupo_etario1 TEXT,
cod_grupo_etario2 SMALLINT,
grupo_etario2 TEXT,
proy_poblacion_mundial_estandar  INT);
--CREATE TABLE
--Query returned successfully in 140 msec.
-----------------------------------------------------------------------------------------------------------------------------------------------------
--DIMENSION PACIENTES
CREATE TABLE dw.dim_pacientes(
id_dimpaciente SERIAL NOT NULL PRIMARY KEY,
cod_paciente INT,
sexo TEXT,
estado_vital TEXT,
mencion_cancer TEXT,
profesion_expedidor TEXT,
edad_muerte SMALLINT,
cantidad_tumores SMALLINT);
--CREATE TABLE
--Query returned successfully in 140 msec.
-----------------------------------------------------------------------------------------------------------------------------------------------------
--DIMENSION LOCALIZACIONES
CREATE TABLE dw.dim_localizaciones(
id_dimlocalizacion SERIAL NOT NULL PRIMARY KEY,
cod_localizacion TEXT,
nombre_localizacion TEXT,
cod_cieo3 TEXT,
cod_cieo3_grupo TEXT,
cod_grupo TEXT,
nombre_grupo TEXT);
--CREATE TABLE
--Query returned successfully in 153 msec.
-----------------------------------------------------------------------------------------------------------------------------------------------------
--DIMENSION MORFOLOGIAS
CREATE TABLE dw.dim_morfologias(
id_dimmorfologia SERIAL NOT NULL PRIMARY KEY,
cod_morfologia TEXT,
nombre_morfologia  TEXT);
--CREATE TABLE
--Query returned successfully in 124 msec.
-----------------------------------------------------------------------------------------------------------------------------------------------------
--DIMENSION CIE10
CREATE TABLE dw.dim_cie10(
```

```
id_dimcie10 SERIAL NOT NULL PRIMARY KEY,
cod_cie10 TEXT,
nombre_cie10 TEXT,
cod_cie10_grupo TEXT,
nombre_cie10_grupo TEXT,
grupo_general TEXT,
grupo_excepto_c44 TEXT,
cod_grupo TEXT,
nombre_grupo TEXT);
--CREATE TABLE
--Query returned successfully in 125 msec.
-----------------------------------------------------------------------------------------------------------------------------------------------------------
--DIMENSION FUENTES
CREATE TABLE dw.dim_fuentes(
id_dimfuente SERIAL NOT NULL PRIMARY KEY,
cod_fuente TEXT,
nombre_fuente TEXT);
--CREATE TABLE
--Query returned successfully in 96 msec.
-----------------------------------------------------------------------------------------------------------------------------------------------------------
--DIMENSION LUGARES
CREATE TABLE dw.dim_lugares(
id_dimlugar SERIAL NOT NULL PRIMARY KEY,
cod_departamento TEXT,
nombre_departamento TEXT,
cod_municipio TEXT,
nombre_municipio TEXT);
--CREATE TABLE
--Query returned successfully in 148 msec.
-----------------------------------------------------------------------------------------------------------------------------------------------------------
--DIMENSION ZONAS
CREATE TABLE dw.dim_zonas(
id_dimzona SERIAL NOT NULL PRIMARY KEY,
zona TEXT,
nombre_comuna_corregimiento TEXT,
nombre_barrio_vereda TEXT);
--CREATE TABLE
--Query returned successfully in 117 msec.
-----------------------------------------------------------------------------------------------------------------------------------------------------------
--DIMENSION TIEMPO DE DIAGNOSTICO
CREATE TABLE dw.dim_tiempo_diagnostico(
id_dimtiempo_diagnostico SERIAL NOT NULL PRIMARY KEY,
ano_diagnostico TEXT,
semestre_diagnostico TEXT,
trimestre_diagnostico TEXT,
cod_mes_diagnostico SMALLINT,
mes_diagnostico TEXT,
cod_dia_diagnostico SMALLINT,
dia_diagnostico TEXT);
--CREATE TABLE
--Query returned successfully in 186 msec.
-----------------------------------------------------------------------------------------------------------------------------------------------------------
--DIMENSION TIEMPO DE MUERTE
CREATE TABLE dw.dim_tiempo_muerte(
id_dimtiempo_muerte SERIAL NOT NULL PRIMARY KEY,
ano_muerte TEXT,
```

```
semestre_muerte TEXT,
trimestre_muerte TEXT,
cod_mes_muerte SMALLINT,
mes_muerte TEXT,
cod_dia_muerte SMALLINT,
dia_muerte TEXT);
--CREATE TABLE
--Query returned successfully in 125 msec.
-----------------------------------------------------------------------------------------------------------------------------------------------------------
--DIMENSION TIEMPO DE ULTIMO CONTACTO
CREATE TABLE dw.dim_tiempo_ultimo_contacto(
id_dimtiempo_ultimo_contacto SERIAL NOT NULL PRIMARY KEY,
ano_ultimo_contacto TEXT,
semestre_ultimo_contacto TEXT,
trimestre_ultimo_contacto TEXT,
cod_mes_ultimo_contacto SMALLINT,
mes_ultimo_contacto TEXT,
cod_dia_ultimo_contacto SMALLINT,
dia_ultimo_contacto TEXT);
--CREATE TABLE
--Query returned successfully in 192 msec.
-----------------------------------------------------------------------------------------------------------------------------------------------------------
--DIMENSION TIEMPO DE TRATAMIENTO
CREATE TABLE dw.dim_tiempo_tratamiento(
id_dimtiempo_tratamiento SERIAL NOT NULL PRIMARY KEY,
ano_tratamiento TEXT,
semestre_tratamiento TEXT,
trimestre_tratamiento TEXT,
cod_mes_tratamiento SMALLINT,
mes_tratamiento TEXT,
cod_dia_tratamiento SMALLINT,
dia_tratamiento TEXT);
--CREATE TABLE
--Query returned successfully in 143 msec.
-----------------------------------------------------------------------------------------------------------------------------------------------------------
--TABLA DE HECHOS (dw.fact_rpcmp)
-----------------------------------------------------------------------------------------------------------------------------------------------------------
CREATE TABLE dw.fact_rpcmp(
id_tumor INTEGER NOT NULL REFERENCES dw.dim_tumores(id_dimtumor),
id_edad INTEGER NOT NULL REFERENCES dw.dim_edades(id_dimedad),
id_tiempo_diagnostico INTEGER NOT NULL REFERENCES
dw.dim_tiempo_diagnostico(id_dimtiempo_diagnostico),
id_tiempo_tratamiento INTEGER NOT NULL REFERENCES
dw.dim_tiempo_tratamiento(id_dimtiempo_tratamiento),
id_paciente INTEGER NOT NULL REFERENCES dw.dim_pacientes(id_dimpaciente),
id_tiempo_ultimo_contacto INTEGER NOT NULL REFERENCES
dw.dim_tiempo_ultimo_contacto(id_dimtiempo_ultimo_contacto),
id_tiempo_muerte INTEGER NOT NULL REFERENCES dw.dim_tiempo_muerte(id_dimtiempo_muerte),
id_localizacion INTEGER NOT NULL REFERENCES dw.dim_localizaciones(id_dimlocalizacion),
id_morfologia INTEGER NOT NULL REFERENCES dw.dim_morfologias(id_dimmorfologia),
id_cie10 INTEGER NOT NULL REFERENCES dw.dim_cie10(id_dimcie10),
id_fuente INTEGER NOT NULL REFERENCES dw.dim_fuentes(id_dimfuente),
id_lugar INTEGER NOT NULL REFERENCES dw.dim_lugares(id_dimlugar),
id_zona INTEGER NOT NULL REFERENCES dw.dim_zonas(id_dimzona),
fi INTEGER,
cantidad_pacientes INTEGER,
```

```
cantidad_caso_fuente INTEGER,
cdmv INTEGER,
cscd INTEGER,
mccpm INTEGER,
ccopd INTEGER,
mcopm INTEGER,
esperanza_vida NUMERIC(4,1),
avmp NUMERIC(4,1),
oac INTEGER,
PRIMARY KEY (id_tumor,id_edad,id_tiempo_diagnostico,id_tiempo_tratamiento,id_paciente,
id_tiempo_ultimo_contacto,id_tiempo_muerte,id_localizacion,id_morfologia,id_cie10,id_fuente,id_lugar,i
d_zona));
--CREATE TABLE
--Query returned successfully in 125 msec.
```

Anexo 6. Script empleado en la Carga de datos a la Bodega de Datos en PostgreSQL.

A continuación, se presenta el código SQL que permite alimentar las tablas de Dimensiones y la tabla de Hechos que conforman la Bodega de Datos diseñada.

```
==========================================================================
--BODEGA DE DATOS SISTEMA YACHAY - RPCMP
--REGISTRO POBLACIONAL DE CANCER DEL MUNICIPIO DE PASTO
--
==========================================================================
--CARGA DE DATOS A PARTIR DE LOS DATOS TRANSFORMADOS DE LOS TUMORES-FUENTE
DEL MUNICIPIO DE PASTO
--
==========================================================================
--------------------------------------------------------------------------------------------------------------------------------------------
--CARGA DE DIMENSION TUMORES
INSERT INTO dw.dim_tumores
(cod_tumor,fecha_incidencia,comportamiento,lateralidad,grado_diferenciacion,base_diagnostico,tratamient
o,
estado_enfermedad)
SELECT DISTINCT tumor,fecha_diagnostico,comportamiento,lateralidad,grado_diferenciacion,
base_diagnostico,tratamiento,estado_enfermedad
FROM etl.tumores_pasto_transformado ORDER BY 1;
--INSERT 0 19619
--Query returned successfully in 504 msec.
--------------------------------------------------------------------------------------------------------------------------------------------
--CARGA DE DIMENSION EDADES
INSERT INTO dw.dim_edades
(edad,descripcion_edad,cod_grupo_etario1,grupo_etario1,cod_grupo_etario2,grupo_etario2,
proy_poblacion_mundial_estandar)
SELECT                                                                    DISTINCT
edad_paciente,descripcion_edad,cod_grupo_etario1,grupo_etario1,cod_grupo_etario2,grupo_etario2,
proy_poblacion_mundial_estandar
FROM etl.tumores_pasto_transformado ORDER BY 1;
--INSERT 0 104
--Query returned successfully in 239 msec.
--------------------------------------------------------------------------------------------------------------------------------------------
--CARGA DE DIMENSION PACIENTES
INSERT INTO dw.dim_pacientes
```

```
(cod_paciente,sexo,estado_vital,mencion_cancer,profesion_expedidor,edad_muerte,cantidad_tumores)
SELECT DISTINCT
paciente,sexo,estado_vital,mencion_cancer,profesion_expedidor,edad_muerte,cantidad_tumores
FROM etl.tumores_pasto_transformado ORDER BY 1;
--INSERT 0 18430
--Query returned successfully in 498 msec.
--------------------------------------------------------------------------------------------------------------------------------------------------------
--CARGA DE DIMENSION LOCALIZACIONES
INSERT INTO dw.dim_localizaciones
(cod_localizacion,nombre_localizacion,cod_cieo3,cod_cieo3_grupo,cod_grupo,nombre_grupo)
SELECT DISTINCT
cod_localizacion,localizacion,cod_cieo3,cod_cieo3_grupo,no_cieo3_grupo,cieo3_grupo
FROM etl.tumores_pasto_transformado ORDER BY 1;
--INSERT 0 231
--Query returned successfully in 314 msec.
--------------------------------------------------------------------------------------------------------------------------------------------------------
--CARGA DE DIMENSION MORFOLOGIAS
INSERT INTO dw.dim_morfologias (cod_morfologia,nombre_morfologia)
SELECT DISTINCT cod_morfologia,morfologia
FROM etl.tumores_pasto_transformado ORDER BY 1;
--INSERT 0 339
--Query returned successfully in 288 msec.
--------------------------------------------------------------------------------------------------------------------------------------------------------
--CARGA DE DIMENSION CIE10
INSERT INTO dw.dim_cie10
(cod_cie10,nombre_cie10,cod_cie10_grupo,nombre_cie10_grupo,grupo_general,grupo_excepto_c44,cod_
grupo,nombre_grupo)
SELECT DISTINCT
cod_cie10,cie10,cod_cie10_grupo,cie10_grupo,cie10_grupo_general,cie10_grupo_excepto_c44,cod_grupo
_cie10,grupo_cie10
FROM etl.tumores_pasto_transformado ORDER BY 1;
--INSERT 0 390
--Query returned successfully in 243 msec.
--------------------------------------------------------------------------------------------------------------------------------------------------------
--CARGA DE DIMENSION FUENTES
INSERT INTO dw.dim_fuentes (cod_fuente,nombre_fuente)
SELECT DISTINCT cod_fuente,fuente
FROM etl.tumores_pasto_transformado ORDER BY 1;
--INSERT 0 135
--Query returned successfully in 149 msec.
--------------------------------------------------------------------------------------------------------------------------------------------------------
--CARGA DE DIMENSION LUGARES
INSERT INTO dw.dim_lugares
(cod_departamento,nombre_departamento,cod_municipio,nombre_municipio)
SELECT DISTINCT cod_departamento,departamento,cod_municipio,municipio
FROM etl.tumores_pasto_transformado ORDER BY 1,3;
--INSERT 0 1
--Query returned successfully in 269 msec.
--------------------------------------------------------------------------------------------------------------------------------------------------------
--CARGA DE DIMENSION ZONAS
INSERT INTO dw.dim_zonas (zona,nombre_comuna_corregimiento,nombre_barrio_vereda)
SELECT DISTINCT zona,comuna_corregimiento,barrio_vereda
FROM etl.tumores_pasto_transformado ORDER BY 1,2;
--INSERT 0 521
--Query returned successfully in 468 msec.
--------------------------------------------------------------------------------------------------------------------------------------------------------
```

```
--CARGA DE DIMENSION TIEMPO DE DIAGNOSTICO
INSERT INTO dw.dim_tiempo_diagnostico
(ano_diagnostico,mes_diagnostico,cod_mes_diagnostico,semestre_diagnostico,trimestre_diagnostico,
dia_diagnostico,cod_dia_diagnostico)
SELECT * FROM
(SELECT DISTINCT(ano_diagnostico::TEXT) FROM etl.tumores_pasto_transformado)
CROSS JOIN
(SELECT DISTINCT(mes_diagnostico),cod_mes_diagnostico,semestre_diagnostico,trimestre_diagnostico
FROM etl.tumores_pasto_transformado)
CROSS JOIN
(SELECT DISTINCT(dia_diagnostico),cod_dia_diagnostico FROM etl.tumores_pasto_transformado)
ORDER BY
ano_diagnostico,semestre_diagnostico,trimestre_diagnostico,cod_mes_diagnostico,cod_dia_diagnostico;
--INSERT 0 20384
--Query returned successfully in 464 msec.
-----------------------------------------------------------------------------------------------------------------------------
--CARGA DE DIMENSION TIEMPO DE MUERTE
INSERT INTO dw.dim_tiempo_muerte
(ano_muerte,mes_muerte,cod_mes_muerte,semestre_muerte,trimestre_muerte,dia_muerte,cod_dia_muerte)
SELECT * FROM
(SELECT DISTINCT(ano_muerte) FROM etl.tumores_pasto_transformado)
CROSS JOIN
(SELECT DISTINCT(mes_muerte),cod_mes_muerte,semestre_muerte,trimestre_muerte FROM
etl.tumores_pasto_transformado)
CROSS JOIN
(SELECT DISTINCT(dia_muerte),cod_dia_muerte FROM etl.tumores_pasto_transformado)
ORDER BY ano_muerte,semestre_muerte,trimestre_muerte,cod_mes_muerte,cod_dia_muerte;
--INSERT 0 15300
--Query returned successfully in 417 msec.
-----------------------------------------------------------------------------------------------------------------------------
--CARGA DE DIMENSION TIEMPO DE ULTIMO CONTACTO
INSERT INTO dw.dim_tiempo_ultimo_contacto
(ano_ultimo_contacto,mes_ultimo_contacto,cod_mes_ultimo_contacto,semestre_ultimo_contacto,trimestre
_ultimo_contacto,
dia_ultimo_contacto,cod_dia_ultimo_contacto)
SELECT * FROM
(SELECT DISTINCT(ano_ultimo_contacto) FROM etl.tumores_pasto_transformado)
CROSS JOIN
(SELECT
DISTINCT(mes_ultimo_contacto),cod_mes_ultimo_contacto,semestre_ultimo_contacto,trimestre_ultimo_
contacto
FROM etl.tumores_pasto_transformado)
CROSS JOIN
(SELECT DISTINCT(dia_ultimo_contacto),cod_dia_ultimo_contacto FROM
etl.tumores_pasto_transformado)
ORDER BY
ano_ultimo_contacto,semestre_ultimo_contacto,trimestre_ultimo_contacto,cod_mes_ultimo_contacto,
cod_dia_ultimo_contacto;
--INSERT 0 11648
--Query returned successfully in 417 msec.
-----------------------------------------------------------------------------------------------------------------------------
--CARGA DE DIMENSION TIEMPO DE TRATAMIENTO
INSERT INTO dw.dim_tiempo_tratamiento
(ano_tratamiento,mes_tratamiento,cod_mes_tratamiento,semestre_tratamiento,trimestre_tratamiento,dia_tr
atamiento,
cod_dia_tratamiento)
```

```
SELECT * FROM
(SELECT DISTINCT(ano_tratamiento) FROM etl.tumores_pasto_transformado)
CROSS JOIN
(SELECT  DISTINCT(mes_tratamiento),cod_mes_tratamiento,semestre_tratamiento,trimestre_tratamiento
FROM
etl.tumores_pasto_transformado)
CROSS JOIN
(SELECT DISTINCT(dia_tratamiento),cod_dia_tratamiento FROM etl.tumores_pasto_transformado)
ORDER BY
ano_tratamiento,semestre_tratamiento,trimestre_tratamiento,cod_mes_tratamiento,cod_dia_tratamiento;
--INSERT 0 12480
--Query returned successfully in 348 msec.
-----------------------------------------------------------------------------------------------------------------------------------------------------
--CARGA DE TABLA DE HECHOS (dw.fact_rpcmp)
-----------------------------------------------------------------------------------------------------------------------------------------------------
INSERT INTO dw.fact_rpcmp
(id_tumor,id_edad,id_tiempo_diagnostico,id_tiempo_tratamiento,id_paciente,
id_tiempo_ultimo_contacto,id_tiempo_muerte,id_localizacion,id_morfologia,id_cie10,id_fuente,id_lugar,i
d_zona,fi,
cantidad_pacientes,cantidad_caso_fuente,cdmv,cscd,mccpm,ccopd,mcopm,esperanza_vida,avmp,oac)
SELECT t2.id_dimtumor,t3.id_dimedad,t4.id_dimtiempo_diagnostico,t5.id_dimtiempo_tratamiento,
t6.id_dimpaciente,t7.id_dimtiempo_ultimo_contacto,t8.id_dimtiempo_muerte,t9.id_dimlocalizacion,
t10.id_dimmorfologia,t11.id_dimcie10,t12.id_dimfuente,t13.id_dimlugar,t14.id_dimzona,
t1.fi,t1.cantidad_pacientes,t1.cantidad_caso_fuente,t1.cdmv,t1.cscd,t1.mccpm,t1.ccopd,t1.mcopm,t1.espera
nza_vida,
t1.avmp,t1.oac
FROM (
(SELECT * FROM etl.tumores_pasto_transformado) t1
LEFT JOIN
(SELECT * FROM dw.dim_tumores) t2
ON
(t1.tumor,t1.fecha_diagnostico,t1.comportamiento,t1.lateralidad,t1.grado_diferenciacion,t1.base_diagnostic
o,
t1.tratamiento,t1.estado_enfermedad) =
(t2.cod_tumor,t2.fecha_incidencia,t2.comportamiento,t2.lateralidad,t2.grado_diferenciacion,t2.base_diagno
stico,
t2.tratamiento,t2.estado_enfermedad)
LEFT JOIN
(SELECT * FROM dw.dim_edades) t3
ON
(t1.edad_paciente,t1.descripcion_edad,t1.cod_grupo_etario1,t1.grupo_etario1,t1.cod_grupo_etario2,t1.grup
o_etario2,
t1.proy_poblacion_mundial_estandar) =
(t3.edad,t3.descripcion_edad,t3.cod_grupo_etario1,t3.grupo_etario1,t3.cod_grupo_etario2,t3.grupo_etario2
,
t3.proy_poblacion_mundial_estandar)
LEFT JOIN
(SELECT * FROM dw.dim_tiempo_diagnostico) t4
ON
(t1.ano_diagnostico::TEXT,t1.semestre_diagnostico,t1.trimestre_diagnostico,t1.mes_diagnostico,t1.dia_dia
gnostico) =
(t4.ano_diagnostico,t4.semestre_diagnostico,t4.trimestre_diagnostico,t4.mes_diagnostico,t4.dia_diagnostic
o)
LEFT JOIN
(SELECT * FROM dw.dim_tiempo_tratamiento) t5
USING (ano_tratamiento,semestre_tratamiento,trimestre_tratamiento,mes_tratamiento,dia_tratamiento)
```

```
LEFT JOIN
(SELECT * FROM dw.dim_pacientes) t6
ON
(t1.paciente,t1.sexo,t1.estado_vital,t1.mencion_cancer,t1.profesion_expedidor,t1.edad_muerte,t1.cantidad_t
umores) =
(t6.cod_paciente,t6.sexo,t6.estado_vital,t6.mencion_cancer,t6.profesion_expedidor,t6.edad_muerte,t6.canti
dad_tumores)
LEFT JOIN
(SELECT * FROM dw.dim_tiempo_ultimo_contacto) t7
USING
(ano_ultimo_contacto,semestre_ultimo_contacto,trimestre_ultimo_contacto,mes_ultimo_contacto,dia_ulti
mo_contacto)
LEFT JOIN
(SELECT * FROM dw.dim_tiempo_muerte) t8
USING (ano_muerte,semestre_muerte,trimestre_muerte,mes_muerte,dia_muerte)
LEFT JOIN
(SELECT * FROM dw.dim_localizaciones) t9
ON
(t1.cod_localizacion,t1.localizacion,t1.cod_cieo3,t1.cod_cieo3_grupo,t1.no_cieo3_grupo,t1.cieo3_grupo) =
(t9.cod_localizacion,t9.nombre_localizacion,t9.cod_cieo3,t9.cod_cieo3_grupo,t9.cod_grupo,t9.nombre_gru
po)
LEFT JOIN
(SELECT * FROM dw.dim_morfologias) t10
ON
(t1.cod_morfologia,t1.morfologia) = (t10.cod_morfologia,t10.nombre_morfologia)
LEFT JOIN
(SELECT * FROM dw.dim_cie10) t11
ON
(t1.cod_cie10,t1.cie10,t1.cod_cie10_grupo,t1.cie10_grupo,t1.cie10_grupo_general,t1.cie10_grupo_excepto
_c44,t1.cod_grupo_cie10,t1.grupo_cie10) =
(t11.cod_cie10,t11.nombre_cie10,t11.cod_cie10_grupo,t11.nombre_cie10_grupo,t11.grupo_general,t11.gru
po_excepto_c44,t11.cod_grupo,t11.nombre_grupo)
LEFT JOIN
(SELECT * FROM dw.dim_fuentes) t12
ON
(t1.cod_fuente,t1.fuente) = (t12.cod_fuente,t12.nombre_fuente)
LEFT JOIN
(SELECT * FROM dw.dim_lugares) t13
ON
(t1.cod_departamento,t1.departamento,t1.cod_municipio,t1.municipio) =
(t13.cod_departamento,t13.nombre_departamento,t13.cod_municipio,t13.nombre_municipio)
LEFT JOIN
(SELECT * FROM dw.dim_zonas) t14
ON
(t1.zona,t1.comuna_corregimiento,t1.barrio_vereda) =
(t14.zona,t14.nombre_comuna_corregimiento,t14.nombre_barrio_vereda)
) ORDER BY id_dimtumor;
--INSERT 0 41345
--Query returned successfully in 17 secs 473 msec.
- - - -
```