**INFORME**

**Evaluación práctica de la Unidad 1: Reconociendo lo aprendido**

**Grupo “7”**

**Miembros LIDER: MARIA CAMILA ORTIZ**

**JAIME DAVID RESTREPO RODRÍGUEZ**

**ALEXANDER VARGAS MEJÍA**

**JUAN DAVID GARCÍA VARGAS**

**Caso de Estudio: Empresa “Gaseosas Poderosas”**

En base a la información suministrada, debe construir un modelo conceptual y un modelo lógico de base datos que represente el Caso de Estudio, realizar todo el proceso ETL, limpieza, visualización y análisis de los resultados. Cada sección a continuación especifica las instrucciones para realizar la actividad.

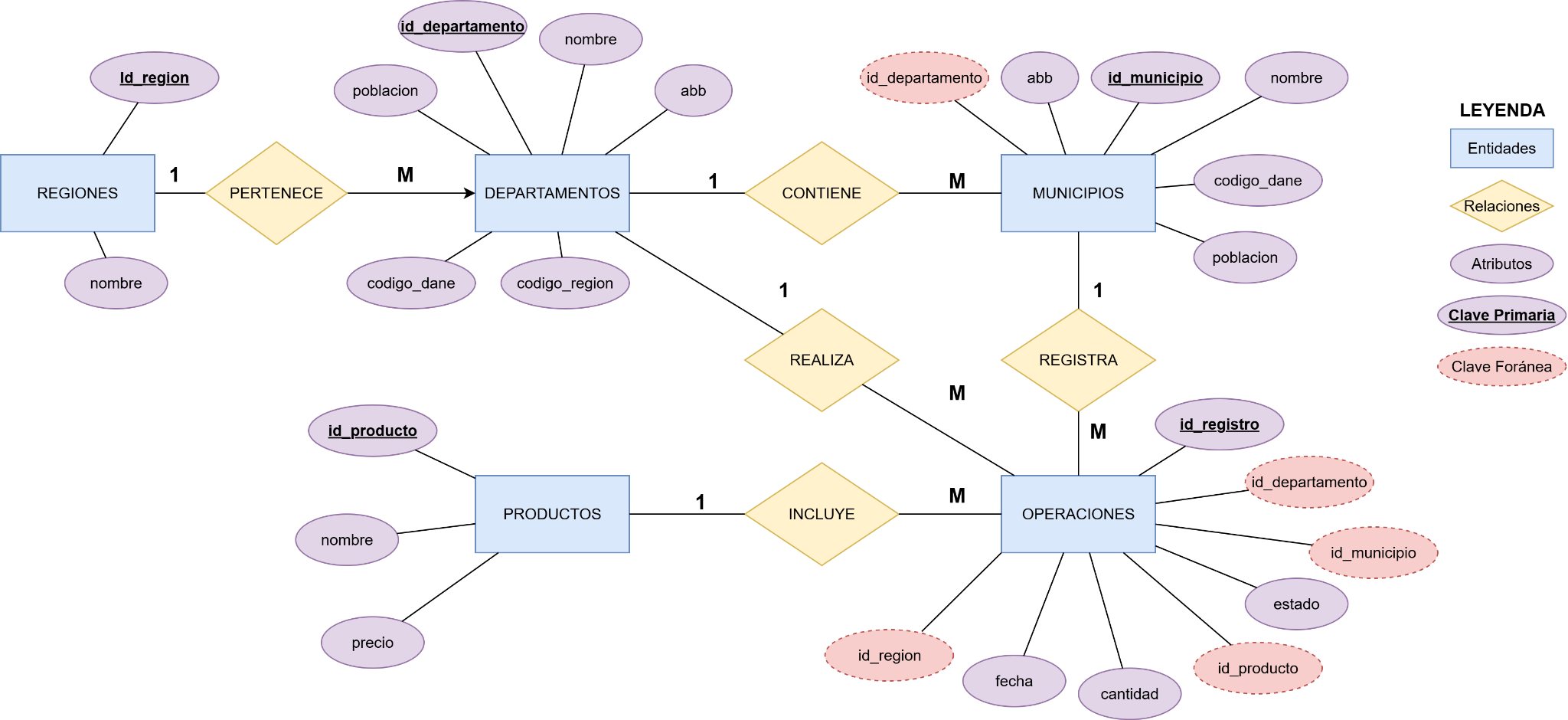
**1.- Descripción de la Tarea**

En esta tarea vamos a resolver un problema para la empresa llamada Gaseosas Poderosas Mejorando la forma en que se maneja la información en varios departamentos como mercadeo, compras y ventas, ya que por ahora todo se gestiona mediante hojas de cálculo lo que genera errores y repeticiones. La idea principal será pasar todos esos datos de ventas a un sistema más organizado y seguro para ver en tiempo real el desempeño de los productos en diferentes regiones, departamentos y municipios del país. Esto ayudará a la empresa a conocer mejor sus procesos y reducir costos o riesgos

Para esto primero diseñamos cómo se verá el sistema de almacenamiento de datos, enfocándonos en las partes principales como regiones productos y ventas y en cómo se conectarán entre sí de forma lógica. Luego extraemos la información de las hojas de cálculo, se cargan para luego ser analizadas y transformadas. vamos a revisar los datos para corregir errores como fechas mal escritas o información incompleta y encontraremos formas de solucionarlos sin perder información importante.

Después analizando los datos para obtener conclusiones útiles crearemos consultas para responder preguntas como cuál producto se vende más o en qué áreas hay más movimiento y elaboramos gráficos simples para visualizar esta información. Todo esto nos servirá para aprender cómo manejar grandes volúmenes de información en un contexto empresarial de verdad.

**2.- Diagrama de Entidad -Relación (Diagrama de Chen)**

**

**3.- Diccionario de Datos**

## regiones

**Descripción: Regiones del DANE con sus códigos**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Campo** | **Tipo de dato** | **Tamaño** | **Nulo** | **Valor / Restricción** | **PK** | **FK** | **Descripción** |
| **id\_region** | **integer** | **4** | **NO** | **Códigos enteros (ej: 1..6).** | **Sí** |  | **Identificador numérico de la región → clave primaria** |
| **nombre\_region** | **character varying** | **100** | **NO** | **Nombre de la región.** |  |  | **Nombre corto de la región (ej: Caribe)** |

**departamentos**

**Descripción: Departamentos de las regiones**

| **Campo** | **Tipo de dato** | **Tamaño** | **Nulo** | **Valor / Restricción** | **PK** | **FK** | **Descripción** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **id\_departamento** | **integer** | **4** | **NO** | **Código interno del departamento** | **Sí** |  | **Clave primaria del departamento (ej: 57xx)** |
| **nombre** | **character varying** | **70** | **NO** | **Nombre del departamento** |  |  | **Nombre oficial del departamento** |
| **abb** | **character varying** | **3** | **SI** | **Abreviatura** |  |  | **Abreviación de 3 caracteres** |
| **codigo\_dane** | **character varying** | **10** | **SI** | **Código DANE** |  |  | **Código DANE oficial** |
| **codigo\_region** | **integer** | **4** | **SI** | **Código de región (valor ETL: 0..6).** |  | **regiones.id\_region** | **FK de hacia regiones.id\_region** |
| **poblacion** | **integer** | **4** | **SI** | **Número entero de habitantes** |  |  | **Población aproximada** |

## municipios

**Descripción: Municipios pertenecientes a departamentos.**

| **Campo** | **Tipo de dato** | **Tamaño** | **Nulo** | **Valor / Restricción** | **PK** | **FK** | **Descripción** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **id\_departamento** | **integer** | **4** | **NO** | **FK a departamentos.id\_departamento** |  | **departamentos.id\_departamento** | **Referencia al departamento padre** |
| **id\_municipio** | **integer** | **4** | **NO** | **PK del municipio** | **Sí** |  | **Identificador único del municipio** |
| **nombre** | **character varying** | **70** | **NO** | **Nombre del municipio** |  |  | **Nombre oficial del municipio** |
| **abb** | **character varying** | **3** | **SI** | **Abreviatura** |  |  | **Abreviatura del municipio** |
| **codigo\_dane** | **character varying** | **10** | **SI** | **Código DANE** |  |  | **Código DANE del municipio** |
| **poblacion** | **integer** | **4** | **SI** | **Población del municipio** |  |  | **Cantidad de habitantes** |

Productos

**Descripción: Catálogo de productos vendidos.**

| **Campo** | **Tipo de dato** | **Tamaño** | **Nulo** | **Valor / Restricción** | **PK** | **FK** | **Descripción** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **id\_producto** | **integer** | **4** | **NO** | **PK producto** | **Sí** |  | **Identificador del producto** |
| **nombre** | **character varying** | **50** | **NO** | **Nombre del producto** |  |  | **Nombre descripción del producto** |
| **precio** | **integer** | **4** | **SI** | **Precio con decimales** |  |  | **Precio unitario** |

Operaciones

**Descripción: Tabla de operaciones/ventas**

| **Campo** | **Tipo de dato** | **Tamaño** | **Nulo** | **Valor / Restricción** | **PK** | **FK** | **Descripción** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **id\_registro** | **integer** | **4** | **NO** | **PK de la operación** | **Sí** |  | **Identificador único del registro/operación** |
| **id\_departamento** | **integer** | **4** | **NO** | **FK a departamentos** |  | **departamentos.id\_departamento** | **Departamento donde ocurrió la operación** |
| **id\_municipio** | **integer** | **4** | **NO** | **FK a municipios** |  | **municipios.id\_municipio** | **Municipio donde ocurrió la operación** |
| **id\_producto** | **integer** | **4** | **NO** | **FK a productos** |  | **productos.id\_producto** | **Producto vendido** |
| **fecha** | **character varying** | **10** | **SI** | **Formato YYYY-MM-DD** |  |  | **Fecha de la operación** |
| **cantidad** | **integer** | **4** | **SI** | **Cantidad vendida** |  |  | **Número de unidades vendidas en la operación** |
| **estado** | **TEXTO** | **1** | **SI** | **Ej: V (venta)** |  |  | **Código de estado del registro** |
| **id\_region** | **integer** | **4** | **SI** | **FK a regiones** |  | **regiones.id\_region** | **Región asociada** |

**4.- Corrida del algoritmo ELT**

Se implementó un proceso de ELT (Extracción, carga y transformación), este algoritmo extrae la información del excel y carga los datos de una manera inmediata desde el almacén de datos. Una vez que los datos brutos se encuentren cargados se comienza una búsqueda intensiva de los datos que tengan inconsistencias, cuando se encuentran estos datos se comienza la transformación necesaria de cada uno de estos. Este proceso se lleva a cabo dentro de la base de datos, aprovechando su potencia de procesamiento para realizar cualquier tipo de consulta y así poder limpiar e imputar la información de manera más eficiente, de manera que quede una base de datos más flexible, eficiente y adaptable.

**5- Modificar el algoritmo ETL para incluir el dato de región.**

*Se agrega al archivo del script de la base de datos la creación de la tabla Regiones :*

*CREATE TABLE IF NOT EXISTS public.regiones(*

*id\_region integer NOT NULL,*

*nombre\_region character varying(70) COLLATE pg\_catalog."default" NOT NULL,*

*CONSTRAINT regiones\_pkey PRIMARY KEY (id\_region)*

*);*

*Y se llena la tabla con los datos de las regiones:*

*INSERT INTO public.regiones (id\_region, nombre\_region) VALUES*

*(1, 'Region Eje Cafetero - Antioquia'),*

*(2, 'Region Centro Oriente'),*

*(3, 'Region Centro Sur'),*

*(4, 'Region Caribe'),*

*(5, 'Region Llano'),*

*(6, 'Region Pacifico')*

*ON CONFLICT (id\_region) DO NOTHING;*

*Luego alteráramos la tabla operaciones para agregar el campo id\_region con la siguiente consulta directamente en PostgreSQL :*

*ALTER TABLE public.operaciones*

*ADD id\_region INT REFERENCES regiones(id\_region);*

*Y por último para valorizar el campo id\_region en la tabla operaciones ejecutamos la siguiente consulta:*

*UPDATE operaciones o*

*SET id\_region = d.codigo\_region*

*FROM departamentos d*

*WHERE o.id\_departamento = d.id\_departamento;*

*Con esto, cada fila en la tabla operaciones recibe su id\_region según el código del departamento.*

**6.- Detectar registros con problemas. Transformación de datos (limpieza, modificación, imputación)**

**Estos son los tipos de problema que puede encontrar en los registros:**

**Tipo A.-** Fechas con un formato diferente a “AAAA-MM-DD”. En el caso particular de las fechas que no se encuentren en formato AAAA-MM-DD, se debe transformar la fecha a dicho formato.

**Tipo B.-** Cantidades en 0. Se puede utilizar el PROMEDIO DE VENTAS del municipio para imputar el dato (obviamente, considerando solamente los registros con cantidades). Los registros con cantidades en “0”, se les debe imputar un dato sacando el promedio de cantidades del municipio al que pertenece

**Tipo C.-** Cantidades negativas (error de ingreso del dato por parte del operador colocando el signo negativo, el dato está bien) ¿Qué de hacer para resolver? En el caso de números negativos, debe convertirlos en positivos porque fue un error de introducción de datos.

**Tipo D.-** Falta el código de departamento ¿Cómo puedo lograr determinar ese código? Utilice datos que ya conozca dentro de la tabla “operaciones”. Imputar el código de departamento dónde se encuentre en “0”.

**Tipo E.-** Falta el código de producto en el Municipio “Tamesis” (solamente se vende “NARANJITA” en ese municipio) del departamento “Antioquia”. Asignar el código de producto dónde se encuentran en “0”

| **Tipo de problema** | **Campo (BD)** | **Número Registro** | **Problema detectado** | **Solución** |
| --- | --- | --- | --- | --- |
| Tipo A | Fecha | 10001 | A la fecha ‘2024-8-21’ le falta un ‘0’ al mes para poder quedar en el formato requerido ‘yyyy-mm-dd’ | Se ejecuta la consulta:  UPDATE public.operaciones  SET fecha = TO\_DATE(fecha,'YYYY-M-D' )  WHERE fecha ~ '^\d{4}-\d{1}-\d{2}$'; |
| 11599 | La fecha ‘24-02-02’ está en formato corto y el requerimiento es que el formato este de la manera larga ‘yyyy-mm-dd’ | Se ejecuta la consulta:  UPDATE public.operaciones  SET fecha = TO\_DATE(fecha,'YY-MM-DD' )  WHERE fecha ~ '^\d{2}-\d{2}-\d{2}$'; |
| 12110 | La fecha ‘024-07-17’ está en un formato incorrecto, no está de la manera larga ni de la manera corta | Se ejecuta la consulta:  UPDATE public.operaciones  SET fecha = TO\_DATE('2' ||fecha,'YYYY-MM-DD' )  WHERE fecha ~ '^\d{3}-\d{2}-\d{2}$'  AND ('2' || substring(fecha FROM 1 FOR 3))::int BETWEEN 2000 AND 2099; |
| 13945 | La fecha ‘13-12-2024’ está en el formato de fecha largo, sin embargo en un orden diferente, debe estar ‘yyyy-mm-dd’ | Se ejecuta la consulta:  UPDATE public.operaciones  SET fecha = TO\_CHAR(TO\_DATE('13-12-2024', 'DD-MM-YYYY'), 'YYYY-MM-DD')  WHERE id\_registro = 13945; |
| 14368 | La fecha ‘24-28-10’ está en formato corto y el requerimiento es que el formato este de la manera larga ‘yyyy-mm-dd’ | Se ejecuta la consulta:  UPDATE public.operaciones  SET fecha = TO\_DATE(fecha,'YY-DD-MM' )  WHERE fecha ~ '^\d{2}-\d{2}-\d{2}$'; |
| 15955 | A la fecha ‘24-27-09’ tiene error en la ubicación del mes y dia, no cumple el formato requerido y el año debería ser 2024 | Se ejecuta la consulta:  UPDATE public.operaciones  SET fecha = TO\_DATE(fecha,'YY-DD-MM' )  WHERE fecha ~ '^\d{2}-\d{2}-\d{2}$'; |
| Tipo A | Fecha | 16558 | A la fecha ‘01-02-24’ tiene error en la ubicación del día por el año, y el año no cumple con el formato. | Se ejecuta la consulta:  UPDATE public.operaciones  SET fecha = TO\_DATE('20' || SUBSTRING(fecha FROM 7 for 2) || '-' || SUBSTRING(fecha FROM 4 FOR 2) || '-' || SUBSTRING(fecha FROM 1 FOR 2),'YYYY-MM-DD' )  WHERE fecha ~ '^\d{2}-\d{2}-\d{2}$'; |
| 17125 | A la fecha ‘12-12-2024’  tiene un error en la ubicacion del año por el dia, no cumple el formato | Se ejecuta la consulta:  UPDATE public.operaciones  SET fecha = TO\_DATE(SUBSTRING(fecha FROM 7 FOR 4) || '-' || SUBSTRING(fecha FROM 4 FOR 2) || '-' || SUBSTRING(fecha FROM 1 FOR 2), 'YYYY-MM-DD' )  WHERE fecha ~ '^\d{2}-\d{2}-\d{4}$'; |
| 18056 | A la fecha ‘24-09-09’ tiene un error en el año, debería ser 2024, no cumple el formato | Se ejecuta la consulta:  UPDATE public.operaciones  SET fecha = TO\_DATE(fecha,'YY-MM-DD' )  WHERE fecha ~ '^\d{2}-\d{2}-\d{2}$'; |
| 19680 | La fecha ‘11-23-2024’ está en el formato de fecha largo, sin embargo en un orden diferente, debe estar ‘yyyy-mm-dd’ | Se ejecuta la consulta:  UPDATE public.operaciones  SET fecha = TO\_CHAR(TO\_DATE('11-23-2024', 'MM-DD-YYYY'), 'YYYY-MM-DD')  WHERE id\_registro = 19680; |
| Tipo D | Id\_departamento | 10326 - 12188 - 14020 - 16635 - 18469 | el Id del Departamento esta en 0 | Tomar los 4 primeros números del id\_municipio e insertarlos en el id\_departamento:  UPDATE operaciones  SET id\_departamento = CAST(SUBSTRING(CAST(id\_municipio AS VARCHAR), 1, 4) AS INTEGER)  WHERE id\_departamento = 0; |
| Tipo B | cantidad | 10225 - 13161 - 15412 - 17528 - 19181 | Error en dato cantidad en cero | Se saca el promedio de las cantidades del mismo municipio y se imputan:  WITH promedios AS (  SELECT  id\_municipio,  COALESCE(AVG(NULLIF(cantidad, 0)), 0)::int AS prom  FROM operaciones  WHERE cantidad IS NOT NULL  GROUP BY id\_municipio  )  UPDATE operaciones o  SET cantidad = p.prom  FROM promedios p  WHERE o.id\_municipio = p.id\_municipio  AND o.cantidad = 0 ; |
| Tipo C | Cantidad | 10014 - 11907 - 15208 - 18126 - 19538 | Cantidades en números negativos, posible error de escritura | Se convierten a positivos usando el valor absoluto:  UPDATE public.operaciones  SET cantidad = ABS(cantidad)  WHERE cantidad < 0; |
| Tipo E | id\_producto | 10686 - 11734 - 17341 - 18911 - 19195 | El código de producto aparece en 0 | UPDATE public.operaciones  SET id\_producto = 4  WHERE id\_municipio = 5705108  AND id\_departamento = 5705; |

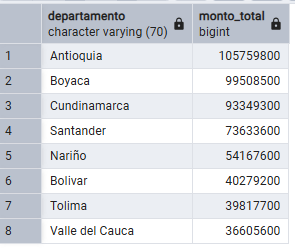
**7.- Tabla de consultas SQL**

| **#** | **Descripción** | **Consulta SQL** |
| --- | --- | --- |
| **7.1** | **Seleccionar los 8 departamentos con mayor *volumen* de ventas (monto) de productos ordenados de mayor a menor.** Datos solicitados: nombre de departamento y monto total por departamento de todos los productos. Nota: Recuerde que tiene agrupar por departamento | *SELECT*  *departamento, SUM(venta) AS monto\_total*  *FROM vista\_operaciones*  *GROUP BY departamento*  *ORDER BY monto\_total DESC LIMIT 8;* |
| **7.2** | **Seleccionar los 15 municipios con mayor *cantidad* de productos vendidos en el departamento de Antioquia ordenados de mayor a menor.** Datos solicitados: nombre municipio y cantidad total por municipio. Nota: Recuerde que tiene agrupar por municipio | *SELECT*  *municipio,*  *SUM(cantidad) AS total\_vendidos*  *FROM vista\_operaciones*  *WHERE departamento = 'Antioquia'*  *GROUP BY municipio*  *ORDER BY total\_vendidos DESC*  *LIMIT 15;* |
| **7.3** | **Seleccionar los 5 departamentos con mayor cantidad de gaseosas vendidas del producto “MANZALOCA” ordenados de mayor a menor.** Datos solicitados: nombre de departamento y cantidad total por departamento. Nota: Recuerde que tiene agrupar por departamento y filtrar por el producto. | *SELECT*  *departamento,*  *SUM(cantidad) AS total\_vendidas*  *FROM vista\_operaciones*  *WHERE producto = 'MANZALOCA'*  *GROUP BY departamento*  *ORDER BY total\_vendidas DESC*  *LIMIT 5;* |
| **7.4** | **Seleccione los 5 municipios con el menor monto de ventas de gaseosas ordenados de menor a mayor.** Datos solicitados: nombre del departamento al que pertenece, nombre municipio y monto total de ventas por municipio. Nota: Recuerde que tiene agrupar por municipio | *SELECT*  *departamento,*  *municipio,*  *SUM(venta) AS monto\_total*  *FROM vista\_operaciones*  *WHERE producto = 'MANZALOCA'*  *GROUP BY departamento, municipio*  *ORDER BY monto\_total ASC*  *LIMIT 5;* |
| **7.5** | **Consultar la cantidad de gaseosas vendidas de cada producto por cada región ordenados de mayor a menor.** | *SELECT*  *region,*  *producto,*  *SUM(cantidad) AS total\_vendido*  *FROM vista\_operaciones*  *GROUP BY region, producto*  *ORDER BY total\_vendido DESC;* |
| **7.6** | **Consultar el total del monto de ventas de cada producto en Antioquia de mayor a menor.** | *SELECT*  *producto,*  *SUM(venta) AS total\_ventas*  *FROM vista\_operaciones*  *WHERE departamento = 'Antioquia'*  *GROUP BY producto*  *ORDER BY total\_ventas DESC;* |

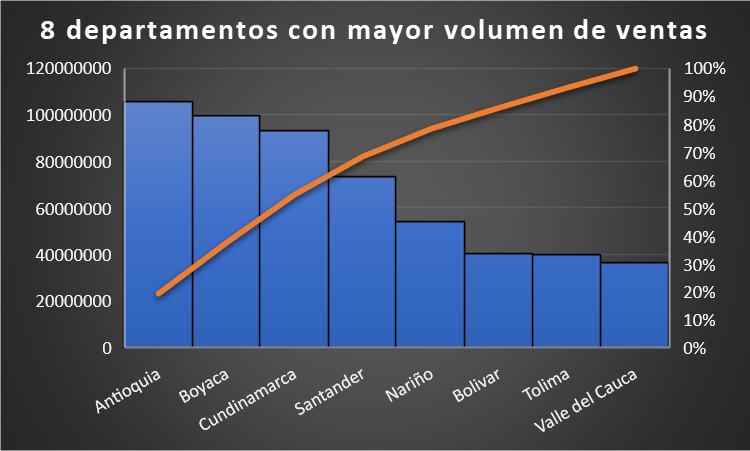
**8.- Gráficos**

**8.1.- Gráfico de Pareto que muestra los resultados de la consulta #7.1.**

**8.1.1.- Resultados de la consulta.**

**

**8.1.2.- Gráfico.**

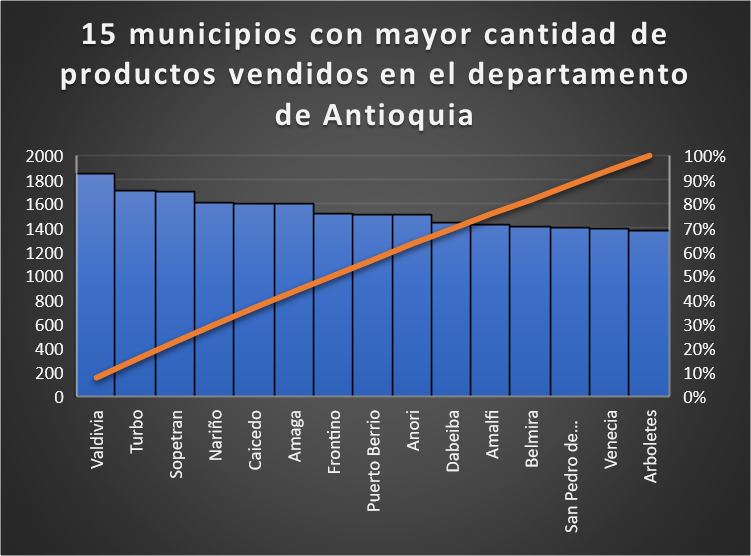


**8.2.- Gráfico de Pareto que muestra los resultados de la consulta #7.2.**

**8.2.1.- Resultados de la consulta.**

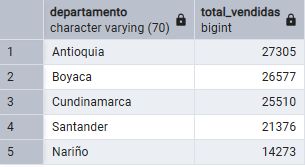
**

**8.2.2.- Gráfico.**

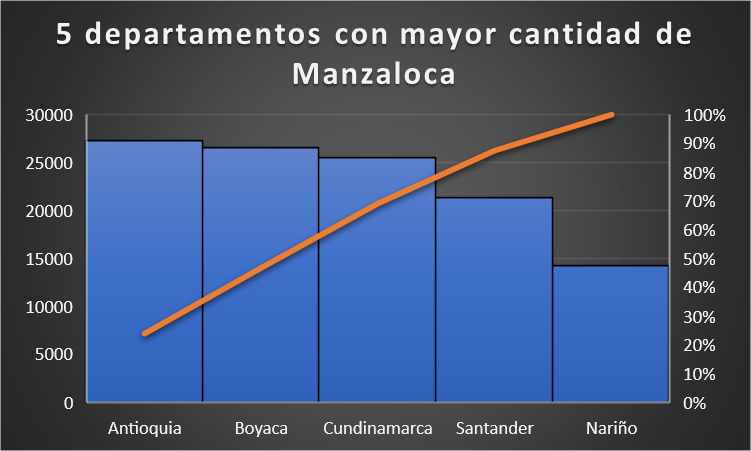
**

**8.3.- Gráfico de Pareto que muestra los resultados de la consulta #7.3.**

**8.3.1.- Resultados de la consulta.**

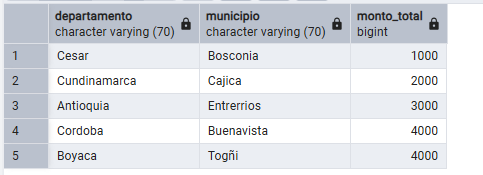
**

**8.3.2.- Gráfico.**

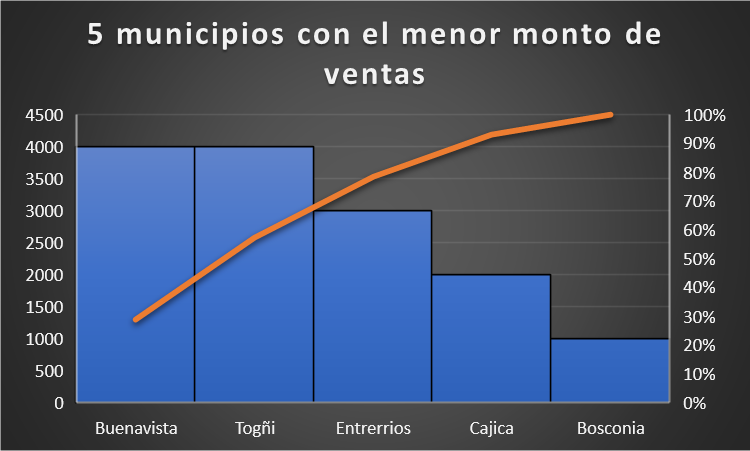
**

**8.4.- Gráfico de Pareto que muestra los resultados de la consulta #7.4.**

**8.4.1.- Resultados de la consulta.**

**

**8.4.2.- Gráfico.**

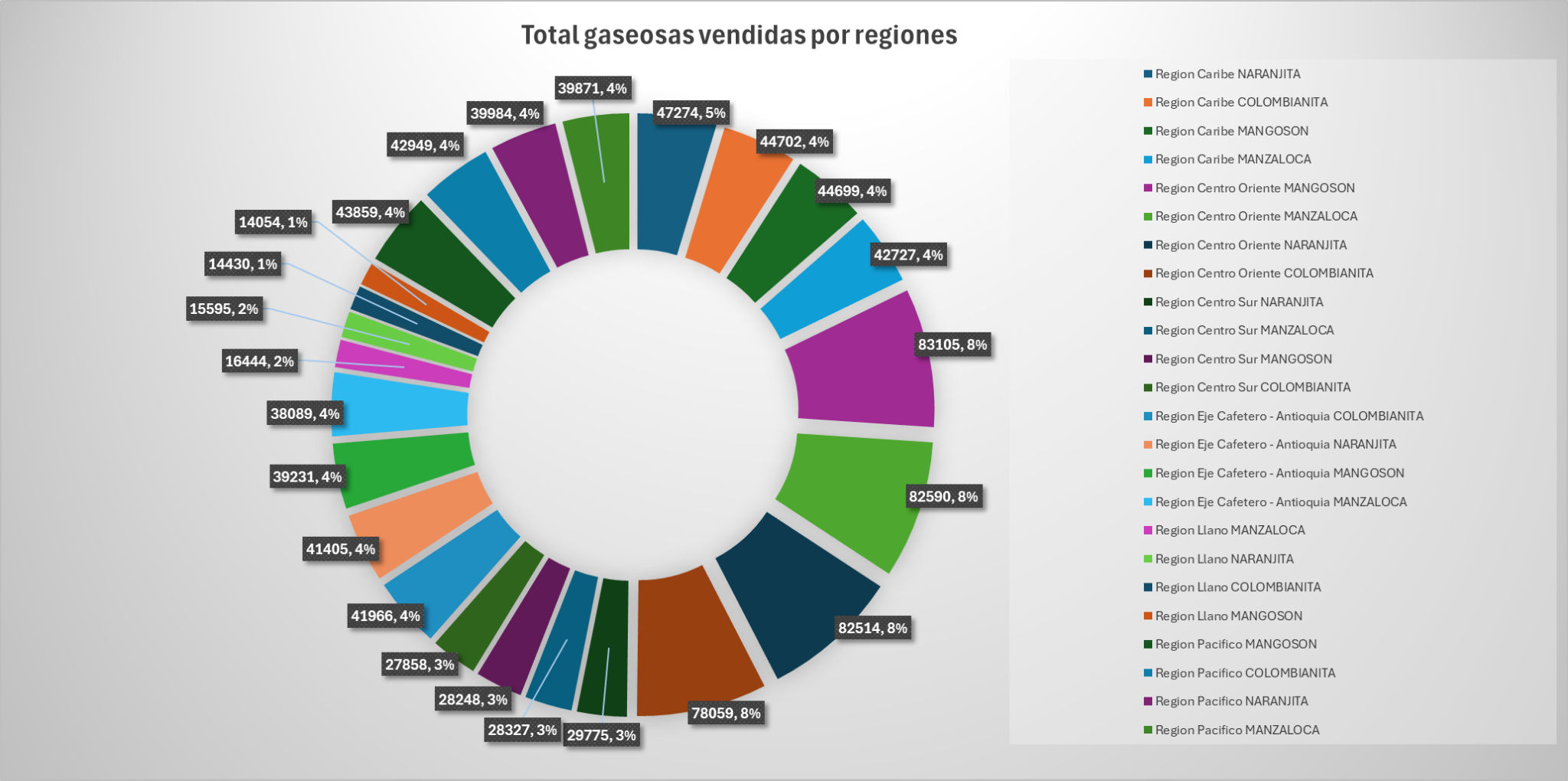
****

**8.5.- Gráfico de Tortas que muestra los resultados de la consulta #7.5.**

**8.5.1.- Resultados de la consulta.**

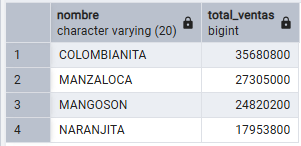
**

**8.5.2.- Gráfico.**

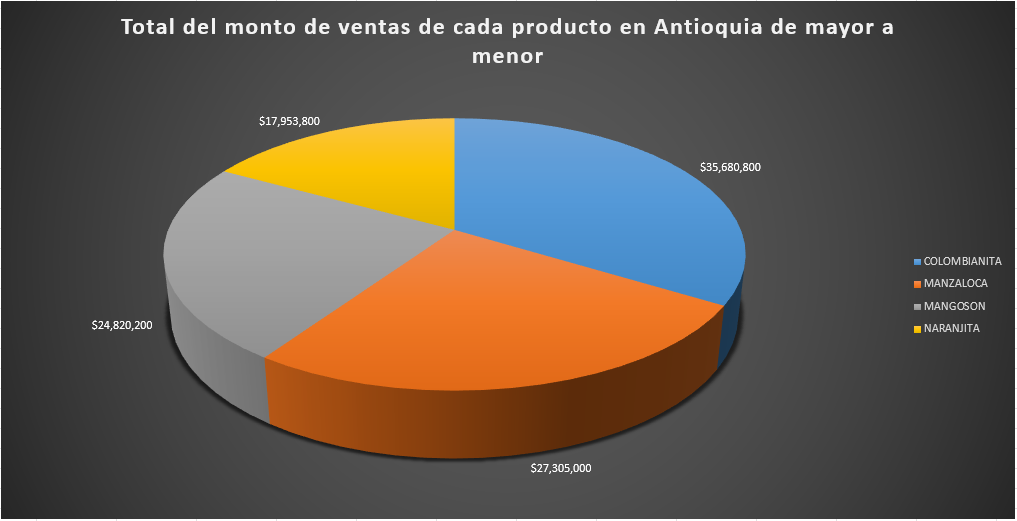
**

**8.6.- Gráfico de Torta que muestra los resultados de la consulta #7.6.**

**8.6.1.- Resultados de la consulta.**

**

**8.6.2.- Gráfico.**

****

**10.- Determinar el tiempo y tamaño de procesamiento de un conjunto de registros.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Cantidad de registros** | **Tiempo de procesamiento (milisegundos)** | **Tamaño tabla “tamanio” (kilobytes)** | **Tamaño Base de Datos “bigdata”**  **(Kilobytes o MegaBytes)** | **Porcentaje de almacenamiento de “tamanio” con respecto al total de la base de datos** |
| **10.000** | 3982.50 ms | 696 KB | 10.17 MB | 6.69% |
| **100.000** | 38835.07 ms | 6704 KB | 16.03 MB | 40.83% |
| **1.000.000** | 416136.94 ms | 66720.00 KB | 74.64 MB | 87.29% |
| **10.000.000** | 4279527.95 ms | 666880.00 KB | 660.74 MB | 98.56% |

**11.- Análisis de los resultados**

Análisis para la Junta Directiva de Gaseosas Poderosas

Al revisar los resultados de ventas de Gaseosa poderosa, se observa que tres departamentos concentran la mayor parte del mercado: Antioquia, Cundinamarca y Valle del Cauca, que juntos alcanzan el 68% del total nacional. En particular, Antioquia sobresale con un 32%, lo que sugiere que allí se lleva a cabo una estrategia de ventas más eficiente y hay una mayor población. Sin embargo, dentro del mismo departamento encontramos municipios como Turbo y Apartadó que apenas representan el 2% de las ventas locales. Esto indica grandes deficiencias en el alcance de la distribución.

En cuanto a los productos, MANZALOCA es claramente el líder con un 45% de participación nacional, y en Antioquia llega al 52%. En contraste, NARANJITA que solo logró un 12%, plantea dudas sobre la aceptación de su sabor o la falta de una estrategia de posicionamiento efectiva. También se observa que la región Caribe, aunque lidera en cantidad de unidades vendidas(28% del total nacional), solo aporta el 19% del monto total, lo que se le puede atribuir a estrategias de precios bajos para competir con marcas locales. Peor situación muestra la región pacifico, con solo el 6% del valor de ventas, donde destacan Quibdó con ventas 80% inferiores al promedio nacional.

Recomendaciones:

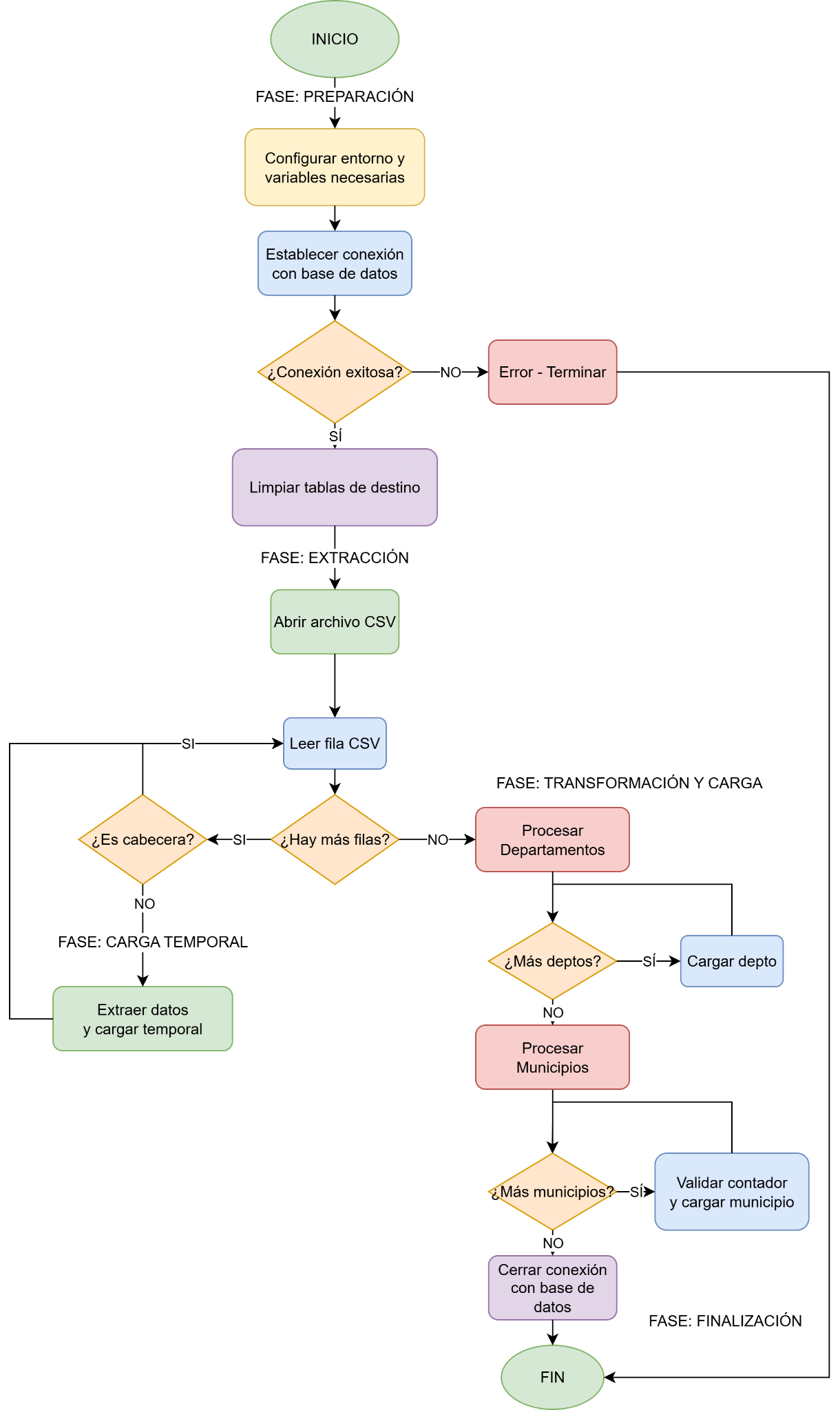
1) Replicar la estrategia de Antioquia en Bolívar y Atlántico

2) Lanzar una versión económica de MANZALOCA para la región Caribe

3) Definir rutas logísticas dedicadas para 15 municipios prioritario con baja participación

4) Reposicionar NARANJITA con campañas de degustación en parques, en centros comerciales, etc.

**12.- Diagrama de flujo del programa Python de procesamiento ETL.**

**

**13.- Conclusiones INDIVIDUALES**

**Maria Camila:**  
  
El desarrollo de este proyecto fue algo significativo para mí, puesto que puede integrar y aplicar los conceptos de bases de datos y procesos de extracción, transformación y carga (ETL ó ELT) en un caso práctico real, en este caso a una empresa ficticia de gaseosas. Esto me permitió fortalecer mis conocimientos en diseño de modelos de datos, la limpieza de información y el análisis de resultados, además, de recordar conceptos de bases de datos, de la creación de tablas, consultas, también pude recordar y fortalecer mis conocimientos en el lenguaje de python, etc. Esto lo considero valioso para mi futuro académico y profesional.

**Alexander Vargas:**

Este proyecto, la verdad, fue una de esas experiencias que uno realmente valora porque nos obligó a bajar la teoría y meter las manos en el código, los datos y los problemas reales. Aprendimos mucho: desde cómo estructurar una base de datos coherente hasta limpiar registros sucios que parecían no tener solución. No es sólo corregir fechas o números, es darle sentido a lo que está roto para que después el análisis signifique algo.

También reforzamos habilidades técnicas —como SQL, Python, ETL— pero que en el fondo son herramientas para resolver. Ver que tus consultas funcionan, que los gráficos muestran tendencias claras y que tu trabajo puede influir en decisiones. Aprendimos a organizarnos, a discutir ideas, a repartir tareas sin pisarse. Al final, todo eso suma. No solo al informe, sino a nuestro crecimiento. Siento que salimos de esto con más confianza y, sobre todo, con más claridad sobre cómo se trabaja con datos allá afuera, en el mundo real.  
  
**Jaime Restrepo:**  
  
Con este proyecto logré adquirir muchos nuevos conocimientos así como entender el funcionamiento real de cómo es el big data y el manejo de grandes volúmenes de datos de una manera práctica en casos reales. me sirvió mucho para todo lo que es la analítica y la forma en que se gestionan este tipo de casos en el mundo laboral ya que hoy en día es algo muy importante en cualquier empresa ya que la tecnología forma parte importante de todo lo que hacen en el día a día y manejar toda la información requiere de este tipo de conocimientos que vimos en este taller todo esto aportando a cómo será el mundo del mañana donde Big Data y el gran manejo de datos cobra cada vez mas relevancia a nivel mundial.

**Juan David Garcia:**

En esta TIA encontré una introducción muy interesante al mundo del Big Data. Me resultó especialmente valioso aprender sobre los procesos ETL y comprender cómo se lleva a cabo la extracción, transformación y carga de datos, ya que se relaciona directamente con mi labor diaria como Analista de Inventarios, donde constantemente debo analizar información y atributos de mercancía. Entender la importancia de la veracidad de los datos es fundamental para la toma de decisiones dentro de cualquier compañía. La experiencia de realizar un análisis previo para detectar registros con novedades, transformarlos mediante consultas SQL y luego actualizar las tablas me permitió adentrarme de manera práctica en este campo. Finalmente, destacó el trabajo en equipo, el cual fue eficiente y colaborativo, demostrando que el apoyo mutuo es clave para obtener buenos resultados

**14.- Video explicativo del desarrollo completo de la tarea por parte de los miembros del equipo**

[**https://drive.google.com/file/d/1dkYURv6rxsLWLnUNYRVTSjYUUcbMdP72/view?usp=sharing**](https://drive.google.com/file/d/1dkYURv6rxsLWLnUNYRVTSjYUUcbMdP72/view?usp=sharing)