

## **Segunda entrega proyecto final**

Samuel Fernando De Dios Pérez

Robert Daniel Fonseca Lesmez

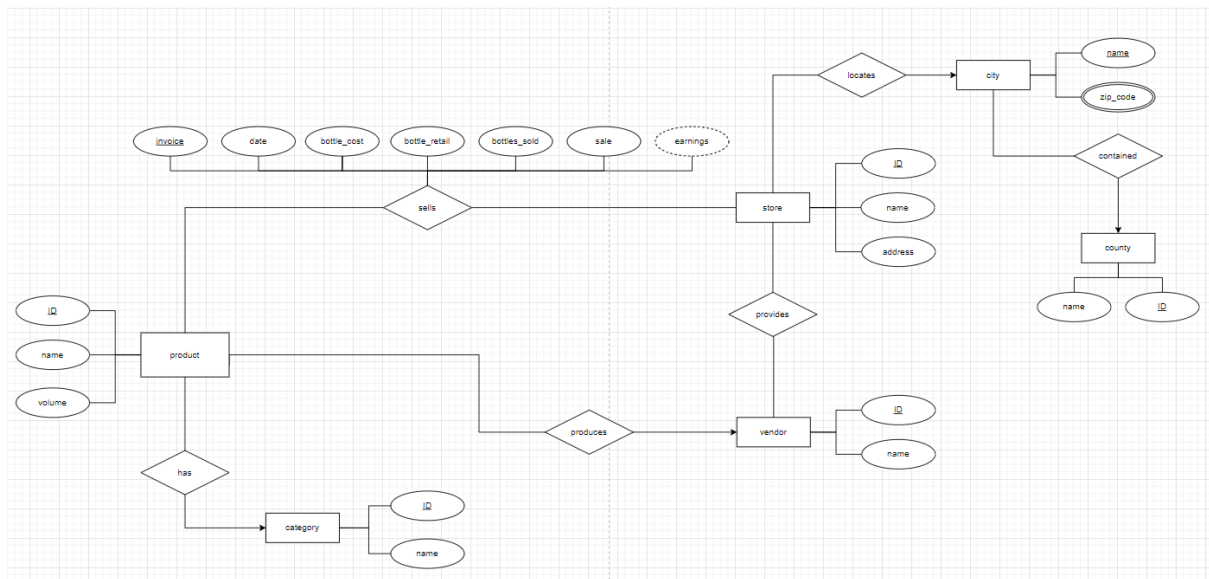
Lukas Morera Torres

Siguiendo con la línea de trabajo propuesta en la primera entrega, se corrigieron errores presentes en los diagramas originales. Las acciones tomadas fueron las siguientes:

- Se eliminó la relación entre las tablas “product” y “provides” debido a que complicaba la creación de las tablas y la resolución de los escenarios de análisis. Además, esta relación era irrelevante, ya que la conexión se podía hacer mediante consultas utilizando otras tablas.
- Los atributos “sale\_price” y “cost” se trasladaron de la tabla “product” a la tabla “sells” debido a un problema con la llave primaria, que consistía en que los precios de venta y los costos cambiaban según los años y la tienda, por lo que el mismo producto podía tener diferente información en estos atributos. Además, para facilitar la comprensión de estos atributos, sus nombres pasaron de “sale\_price” a “bottle\_retail” y de “cost” a “bottle\_cost”.
- Se eliminó el atributo 'location' de la tabla 'store' debido a la complejidad de su carga en PostgreSQL, dada la forma en que estaba planteado en la base de datos original y porque las coordenadas estaban invertidas. Además, este atributo no aportaba a la solución de los escenarios de análisis.
- Se eliminó la restricción “unique” que estaba aplicada al atributo “name” de la tabla “product”. Esto se hizo debido a que existen productos con el mismo nombre pero con diferentes volúmenes.

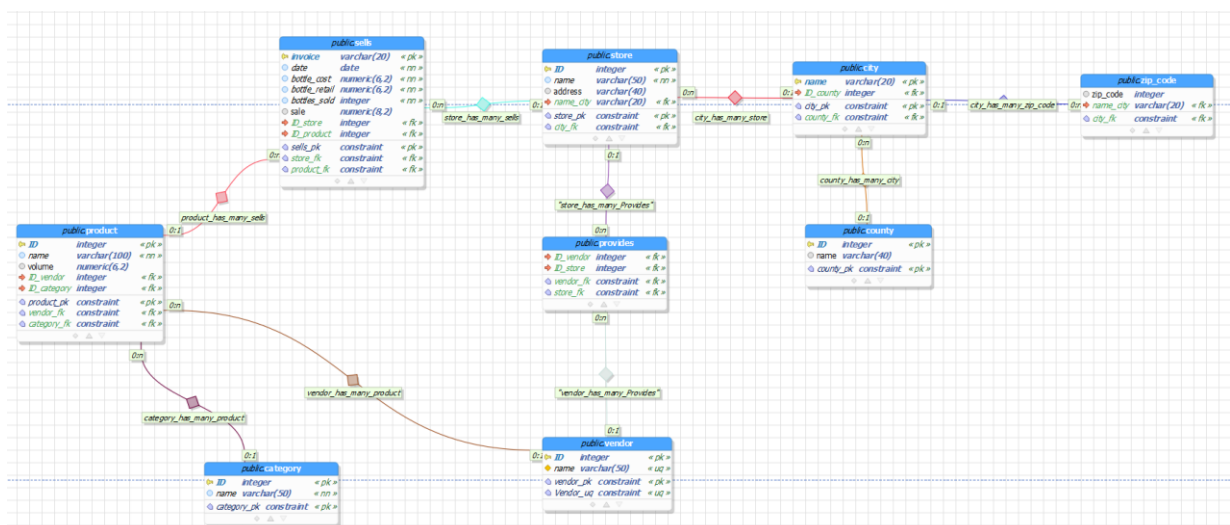
Los resultados fueron:

***Diagrama entidad-relación***



## Normalización diagrama relacional

Para este proceso, se utilizó la base de datos original en formato CSV y se empleó Excel para convertirla en una tabla. A continuación, se aplicó el proceso de normalización paso a paso para finalmente generar el diagrama relacional que se había adelantado previamente. El resultado obtenido es el siguiente:



## Carga de información en base de datos

Una vez que se normalizó tanto la tabla en Excel como el diagrama relacional, se transfirió toda la información a pgAdmin4 como una base de datos con el siguiente proceso:

1. Se definieron las tablas según las condiciones establecidas en el diagrama relacional previamente normalizado. Se tuvieron en cuenta todos los atributos, llaves primarias, llaves foráneas y restricciones como "unique" en caso de que estuvieran presentes.

2. Cada una de las tablas resultantes de la normalización se dividió en archivos separados en formato CSV UTF-8. Esto se hizo para evitar problemas al leer los archivos en pgAdmin4.
3. Una vez que se obtuvieron todos los archivos, se utilizó la función "copy" dentro de pgAdmin4. Esta función permite copiar los datos de un archivo CSV y pegarlos en una tabla de la base de datos. Se realizó una instrucción para cada tabla.
4. Finalmente, se ejecutaron todas las instrucciones para completar el proceso de carga de información y obtener la base de datos creada con todos los datos cargados en SQL.

### **Análisis de escenarios**

Con base en los datos que tenemos a nuestra disposición e información estadística obtenida de estudios anteriores realizados a la base de datos, se proponen los siguientes escenarios de análisis:

- Para observar qué tipo de licor es más consumido en el estado de Iowa, se consulta el número de ventas realizado por cada categoría de licor.

*Metodología:* Se unen las tablas "sells", "product" y "category", donde se selecciona el nombre de la categoría y se cuenta el número de ventas, agrupando los datos por el nombre de la categoría.

*Utilidad:* Esta consulta permite realizar un análisis comparativo entre las diferentes categorías de licor para comparar las ventas de cada una.

- Para analizar el total de ingresos en el estado de Iowa por ventas de licor cada año, se consultan los ingresos anuales por ventas de licor.

*Metodología:* Se utiliza únicamente la tabla "sells", donde se selecciona el año y se realiza la suma de los ingresos, agrupando los datos por año (usando extract).

*Utilidad:* Esta consulta permite crear un diagrama comparativo de los ingresos por año, para comprender los cambios culturales respecto al consumo de licor anualmente.

- Para analizar las ganancias por ventas de licor de las tiendas por condado, se consultan las ganancias por condado.

*Metodología:* Se unen las tablas "sells", "store", "city" y "county", y se selecciona el nombre del condado y las ganancias. Las ganancias se calculan multiplicando el número de botellas vendidas por la diferencia entre el precio de compra y el precio de venta en las tiendas. Todo esto se agrupa por el nombre del condado.

*Utilidad:* Esta consulta permite hacer una comparativa entre los diferentes condados o entre los precios de compra y venta.

- Con el objetivo de identificar el mes en el que más se compra licor, se consulta el mes con mayor volumen de ventas.

*Metodología:* Se utiliza únicamente la tabla "sells", de la cual se extrae el mes y se cuenta el número de ventas, agrupando los datos por mes (usando extract).

*Utilidad:* Esta consulta es útil para identificar tendencias en la cultura de Iowa y determinar en qué época del año se realiza un mayor número de compras de licor.

- Con el fin de comparar los precios de cada categoría en las tiendas, se hace una consulta relacionada a la diferencia de los costos de algunas categorías.

*Metodología:* Se unen las tablas "category", "product" y "sells", seleccionando el nombre de la categoría y la suma de los costos, agrupando los datos por el nombre de la categoría.

*Utilidad:* Esta consulta permite realizar una comparativa de costos entre categorías, lo que facilita el análisis de las categorías más costosas y menos costosas.

- Según estadísticas de algunos sitios especializados en comercio en Iowa, como se mencionó en las reglas de negocio, el vodka es el licor más consumido en Iowa. Por lo tanto, se realiza una consulta relacionada con la comparación de ingresos de las categorías de vodka.

*Metodología:* Se unen las tablas "category", "product" y "sells", seleccionando el nombre de la categoría y la suma de los ingresos. Se filtran solo las categorías que contengan 'VODKA' en su nombre.

*Utilidad:* Esta consulta brinda datos más precisos sobre el licor más consumido en Iowa, permitiendo ver que variedades de vodka son más compradas.

## **Anexos**

Link del repositorio: [https://github.com/Maicken052/Proyecto\\_Datos.git](https://github.com/Maicken052/Proyecto_Datos.git)