**4. ANÁLISIS DE DATOS**

**4.1. Preparación de Datos**

**4.1.1. Extracción de Datos**La fase de extracción de datos es crucial en la preparación de datos, ya que garantiza la disponibilidad de información desde diversas fuentes para construir un conjunto de datos robusto y completo.En este proyecto, se extrajeron datos desde varias fuentes clave, incluyendo datos históricos y otros generados internamente mediante herramientas como Mockaroo, lo cual permitió obtener una perspectiva integral:

* **Base de Datos de Ventas**: Esta base de datos contiene información detallada sobre transacciones, productos vendidos y clientes. Incluye datos como fechas de transacción, cantidades vendidas, tipos de productos y perfiles de clientes, lo cual resulta esencial para entender las dinámicas de compra y segmentar el mercado.
* **Archivos CSV**: Utilizando herramientas como Mockaroo, se generaron archivos CSV para simular datos específicos y complementar aquellos que no estaban disponibles en otras fuentes. Esto permite añadir datos controlados y personalizados que enriquecen el análisis.
* **API Externas**: A través de APIs externas, se accedió a información adicional relevante, como datos de mercado o tendencias de consumo. Estas fuentes externas ayudan a contextualizar los datos internos con tendencias globales o nacionales, ofreciendo una perspectiva más amplia para el análisis.

La herramienta **Pentaho Data Integration (PDI)** fue utilizada para realizar la extracción de datos. PDI permite configurar conexiones seguras con cada sistema de origen y ejecutar la extracción de datos de manera automatizada, minimizando errores y asegurando la consistencia de los datos recolectados.

**4.1.2. Transformación de Datos**La transformación de datos es la etapa en la que se preparan los datos para ser analizados, realizando una serie de operaciones para mejorar la calidad y usabilidad de los mismos:

* **Limpieza de Datos**: Se eliminan registros duplicados, se corrigen errores de formato o de valores y se gestionan valores nulos. Este proceso es esencial para asegurar que los datos sean precisos y no presenten inconsistencias que puedan afectar los resultados del análisis.
* **Normalización**: Consiste en ajustar los datos a formatos estandarizados. Por ejemplo, las fechas se convierten a un formato uniforme, y los datos monetarios se alinean en una sola moneda o tipo de cambio, lo que facilita la comparación y el análisis de la información.
* **Enriquecimiento**: En esta etapa, se derivan nuevos datos a partir de los existentes. Por ejemplo, se calcula el total de ventas a partir de los datos de cantidad y precio unitario. Este proceso permite generar nuevas variables que pueden ser útiles para el análisis y que, de otro modo, no estarían disponibles.
* **Agregación**: Los datos se resumen a niveles específicos según las necesidades del análisis, como ventas por sucursal, categoría de producto o región. Esta agregación ayuda a identificar patrones de alto nivel y facilita la interpretación de grandes volúmenes de datos.

**Pentaho Data Integration (PDI)** facilita la transformación de datos mediante un entorno visual en el que se pueden aplicar estas operaciones de manera intuitiva y controlada, lo que asegura la calidad de los datos antes de su carga final.

**4.1.3. Carga de Datos**La carga de datos ,en la cual los datos transformados se almacenan en la base de datos de destino para su posterior análisis.

El modelo de datos organiza la información en elementos descriptivos y cuantitativos que juntos permiten realizar un análisis completo de las transacciones. Los elementos descriptivos contienen detalles estáticos, como los datos de clientes, productos y fechas, proporcionando el contexto necesario para comprender y relacionar la información. Por otro lado, los elementos cuantitativos capturan los valores específicos de cada transacción, como las cantidades y montos de ventas, que sirven como base para identificar patrones de consumo y tendencias a lo largo del tiempo.

**4.1.4. Detalles de Carga de Datos en la Base de Datos**Para la carga de datos, se aplicaron dos enfoques diferentes en función del tipo de tabla, optimizando tanto la eficiencia como la flexibilidad de la carga:

**4.1.4.1Carga de la Tabla de Hechos** :

* **Método**: Para esta tabla, se utilizó **Pentaho**, que ofrece una automatización completa en la carga de datos y permite aplicar transformaciones directamente durante el proceso.
* **Ventajas**:
  + **Automatización**: Pentaho permite automatizar la carga de datos desde múltiples fuentes, manteniendo la integridad y consistencia en los datos cargados.
  + **Transformaciones**: Se aplican transformaciones, como cálculos y limpieza, en el mismo proceso de carga, asegurando que los datos en la tabla de hechos sean precisos y listos para el análisis.
  + **Manejo de Altos Volúmenes de Datos**: Gracias a sus capacidades avanzadas de Pentaho es ideal para manejar grandes volúmenes de datos, lo cual es esencial en un entorno donde el seguimiento de inventario requiere procesar datos masivos de forma rápida y eficiente.
* **Carga de Otras Tablas** :
  + **Método**: Las tablas de dimensión, que incluyen información sobre países, ciudades, productos y categorías, fueron cargadas mediante la instrucción **INSERT INTO** en SQL.
  + **Proceso**:
    - **Manual o Semi-Automático**: En algunos casos, los datos fueron ingresados de forma manual o con scripts SQL, lo que permite un control directo sobre el proceso de carga.
    - **Flexibilidad**: Este método permite personalizar la carga de datos en función de las necesidades de negocio, ya que algunas dimensiones no requieren una carga automatizada debido a su naturaleza estática o su volumen relativamente bajo.

**4.1.2. Transformación de Datos**

En este caso no se realizó la transformación de datos, ni depuración de datos atípicos, ni la técnica MAECI, ya que no podemos agregar un dato a un producto que no tuvo ninguna venta en el día.

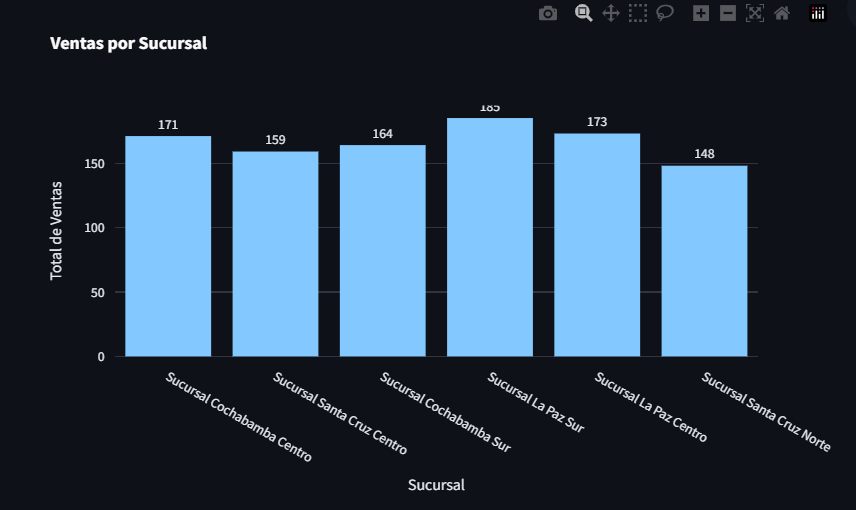
**4.2. Exploración de Datos**

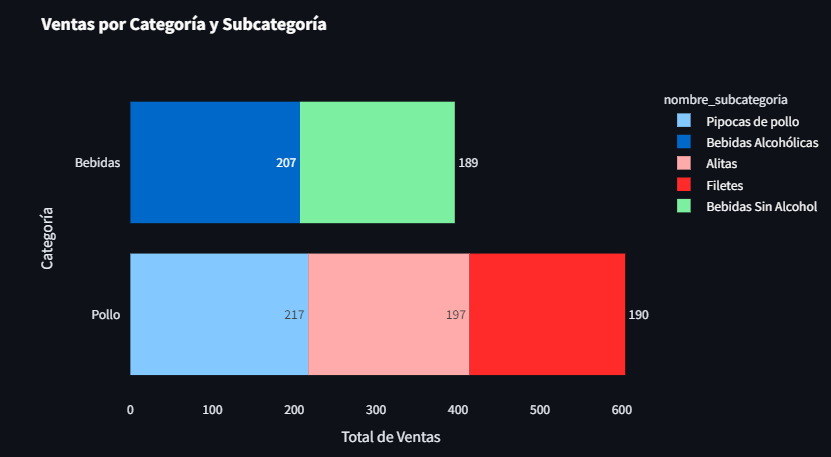
**4.2.1. Análisis Descriptivo:**

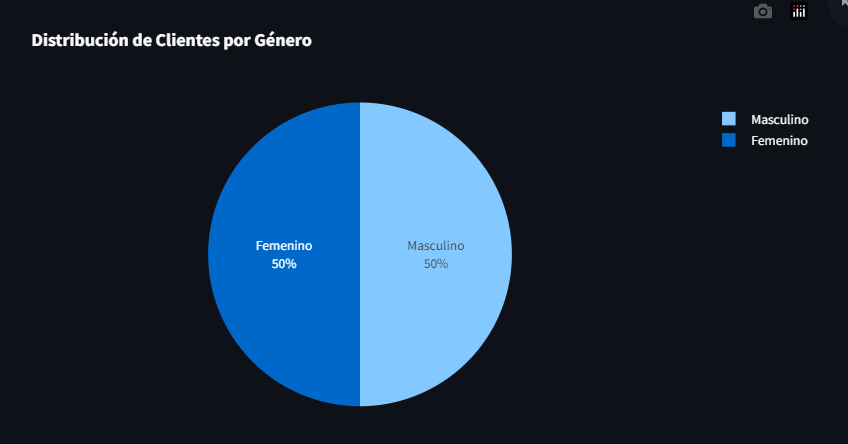
Se puede observar que en promedio los días lunes se vende 5 unidades de agua mineral, 5 unidades de alitas de pollo de 10 unidades, 6 unidades de alitas de pollo de 6 unidades, 7 unidades de gaseosa cola, esto fueron los productos que en promedio obtuvieron mayor cantidad vendida en el día lunes, más detalles en el dashboard.

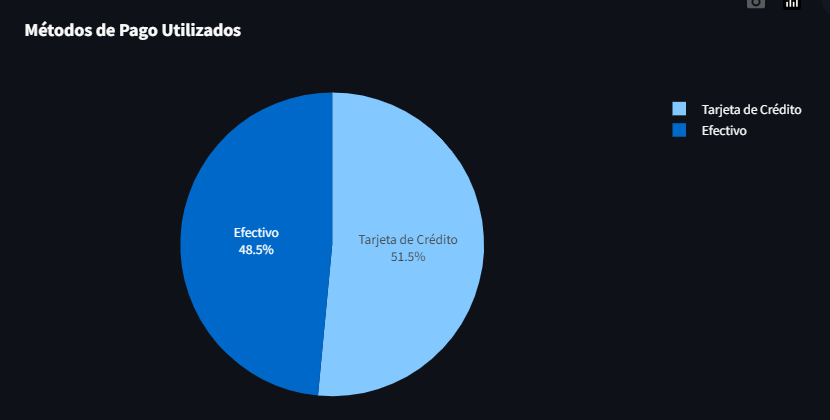
**4.3. Visualización de Datos**

**4.3. Análisis Univariado y Bivariado**

**4.3.1. Análisis Univariado**

**Se puede observar que la sucursal La Paz Sur vendió 185 unidades en el mes de octubre, seguida de la sucursal La Pas Centro con 173 unidades vendidas de nuestros productos.

Se puede observar que las pipocas de pollo fueron el producto más adquirido por nuestros clientes tomando en cuenta que es de todas las sucursales.

Nuestros clientes se clasifican por ser 50% hombres y 50% mujeres.

Se observa que el 51.5% de los clientes prefiere pagar con tarjeta de crédito, esto nos ayuda a ajustar el servicio al momento de adquirir un producto.

**4.3.2. Análisis Bivariado**

Se puede observar que el día martes en la sucursal de Cbba-Sur obtuvo mayor ingreso en ventas, seguido de sucursal en La Paz Sur los días jueves y Sucursal Cbba-Centro los días martes.