# Definición del problema.

El proyecto se enfoca en abordar un desafío común en la gestión de productos de consumo masivo: la predicción de la demanda estacional. El problema específico es la dificultad que enfrentan las empresas para anticipar los cambios en el comportamiento de compra de los consumidores, especialmente durante las diferentes temporadas del año (por ejemplo, vacaciones, festividades, cambios climáticos, etc.). Estos cambios afectan las ventas y, si no son correctamente previstos, pueden resultar en problemas como exceso o falta de inventario, pérdidas de oportunidades de venta y costos innecesarios.

El objetivo del proyecto es utilizar datos históricos de ventas para identificar patrones estacionales y prever la demanda futura, lo que permitirá a las empresas tomar decisiones más informadas sobre la gestión de inventarios y estrategias comerciales.

# Definición de la pregunta de negocio y alcance del proyecto.

La pregunta del negocio principal que guiara este proyecto es:

***¿Cómo podemos predecir la demanda estacional de productos de consumo masivo, teniendo en cuenta factores históricos de ventas y comportamiento de compra, para optimizar la gestión de inventario y mejorar la eficiencia operativa?***

Este proyecto tiene un alcance limitado a productos de consumo masivo vendidos a través de plataformas de e-commerce (en este caso, Olist Store), durante un período determinado (3 años inicialmente, pero ajustable). El modelo debe ser capaz de anticipar cuáles productos tendrán mayor demanda en diferentes épocas del año, considerando patrones estacionales y de comportamiento del consumidor. La solución está diseñada para empresas que buscan mejorar su gestión de inventarios y maximizar la disponibilidad de productos según las necesidades de sus clientes.

# Definición de los conjuntos de datos a emplear.

Para el desarrollo del proyecto de predicción de demanda estacional, se utilizarán datos históricos de la plataforma Olist Store. Estos datos proporcionan información detallada sobre los productos vendidos, las transacciones realizadas, los clientes, los vendedores y las reseñas. A continuación, se describen las principales tablas que serán empleadas en el análisis:

#### Tabla geo

Esta tabla contiene información geográfica relevante sobre los códigos postales de los clientes y vendedores.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ****Nombre de la Columna**** | ****Descripción**** | ****Tipo de Dato**** | ****Ejemplo del Dato**** | ****Clave**** |
| **cep\_prefix** | Prefijo del CEP (Código Postal) | string | "64091" | Clave primaria |
| **city** | Nombre de la ciudad | string | "teresina" |  |
| **uf** | Unidad Federativa (Estado) | string | "PI" |  |
| **lat** | Latitud de la ubicación | doble | -5.0874608 |  |
| **lon** | Longitud de la ubicación | doble | -42.8049571 |  |

#### Tabla customers

Información detallada sobre los clientes que realizaron compras en Olist Store.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ****Nombre de la Columna**** | ****Descripción**** | ****Tipo de Dato**** | ****Ejemplo del Dato**** | ****Clave**** |
| **customer\_id** | Identificador único del cliente | string | "f2a1d75b74d9ec748" | Clave primaria |
| **customer\_unique\_id** | Identificador único y persistente del cliente | string | "15ee900ec703c9a10" |  |
| **customer\_zip\_code\_prefix** | Prefijo del código postal del cliente | string | "68590" | Clave foránea (geo.cep\_prefix) |
| **customer\_city** | Ciudad del cliente | string | "jacunda" |  |
| **customer\_state** | Estado del cliente | string | "PA" |  |

#### Tabla order\_items

Esta tabla contiene los detalles de los ítems incluidos en cada pedido realizado en Olist Store.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ****Nombre de la Columna**** | ****Descripción**** | ****Tipo de Dato**** | ****Ejemplo del Dato**** | ****Clave**** |
| **order\_id** | Identificador único del pedido | string | "8ac26cb70a17887c..." | Clave foránea (orders.order\_id) |
| **order\_item\_id** | Identificador del ítem dentro del pedido | integer | 1 | Clave primaria compuesta junto a order\_id |
| **product\_id** | Identificador único del producto | string | "4lebbrb7a41c44632..." | Clave foránea (products.product\_id) |
| **seller\_id** | Identificador único del vendedor | string | "7a67c85e85bbc2e85..." | Clave foránea (sellers.seller\_id) |
| **shipping\_limit\_date** | Fecha límite para el envío | timestamp | "2017-05-22 16:05:44" |  |
| **price** | Precio del producto | doble | 109.99 |  |
| **freight\_value** | Valor del envío | doble | 18.02 |  |

#### Tabla order\_payments

Registra los detalles de los pagos realizados en los pedidos.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ****Nombre de la Columna**** | ****Descripción**** | ****Tipo de Dato**** | ****Ejemplo del Dato**** | ****Clave**** |
| **order\_id** | Identificador único del pedido | string | "b81ef226f3fe1789b..." | Clave foránea (orders.order\_id) |
| **payment\_sequential** | Secuencia del pago dentro del pedido | integer | 1 | Clave primaria compuesta junto a order\_id |
| **payment\_type** | Tipo de pago utilizado | string | "credit\_card" |  |
| **payment\_installments** | Número de cuotas del pago | integer | 8 |  |
| **payment\_value** | Valor del pago | doble | 99.33 |  |

#### Tabla order\_reviews

Incluye información sobre las reseñas de los clientes en Olist Store.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ****Nombre de la Columna**** | ****Descripción**** | ****Tipo de Dato**** | ****Ejemplo del Dato**** | ****Clave**** |
| **review\_id** | Identificador único de la reseña | string | "f21e5dff1c6efe..." | Clave primaria |
| **order\_id** | Identificador único del pedido | string | "d6027c4fb846f61d7..." | Clave foránea (orders.order\_id) |
| **review\_score** | Calificación de la reseña (1 a 5) | string | 4 |  |
| **review\_comment\_title** | Título de la reseña (si existe) | string | NULL |  |
| **review\_comment\_message** | Mensaje o comentario de la reseña | string | "Adorei a mercadoria!" |  |
| **review\_creation\_date** | Fecha en que se creó la reseña | timestamp | "2017-09-21 00:00:00" |  |
| **review\_answer\_timestamp** | Fecha y hora en que se respondió la reseña | timestamp | "2017-09-23 11:06:26" |  |

#### Tabla orders

Contiene información sobre los pedidos realizados en la plataforma.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ****Nombre de la Columna**** | ****Descripción**** | ****Tipo de Dato**** | ****Ejemplo del Dato**** | ****Clave**** |
| **order\_id** | Identificador único del pedido | string | "995392413cee616c1..." | Clave primaria |
| **customer\_id** | Identificador único del cliente | string | "4bf2490c4245cdb25a..." | Clave foránea (customers.customer\_id) |
| **order\_status** | Estado del pedido (ej. delivered, canceled) | string | "delivered" |  |
| **order\_purchase\_timestamp** | Fecha y hora en que se realizó la compra | timestamp | "2017-09-04 22:24:05" |  |
| **order\_approved\_at** | Fecha y hora en que se aprobó el pedido | timestamp | "2017-09-04 22:43:54" |  |
| **order\_delivered\_carrier\_date** | Fecha y hora en que el transportista entregó el pedido | timestamp | "2017-09-13 17:20:04" |  |
| **order\_delivered\_customer\_date** | Fecha y hora en que el cliente recibió el pedido | timestamp | "2017-09-22 21:09:32" |  |
| **order\_estimated\_delivery\_date** | Fecha estimada de entrega del pedido | timestamp | "2017-09-27 00:00:00" |  |

#### Tabla products

Incluye detalles sobre los productos vendidos en la plataforma.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ****Nombre de la Columna**** | ****Descripción**** | ****Tipo de Dato**** | ****Ejemplo del Dato**** | ****Clave**** |
| **product\_id** | Identificador único del producto | string | "1e9e8ef04dbcff454..." | Clave primaria |
| **product\_category\_name** | Categoría del producto | string | "perfumaria" |  |
| **product\_name\_lenght** | Longitud del nombre del producto | integer | 40 |  |
| **product\_description\_lenght** | Longitud de la descripción del producto | integer | 287 |  |
| **product\_photos\_qty** | Cantidad de fotos del producto | integer | 1 |  |
| **product\_weight\_g** | Peso del producto en gramos | integer | 225 |  |
| **product\_length\_cm** | Longitud del producto en centímetros | integer | 16 |  |
| **product\_height\_cm** | Altura del producto en centímetros | integer | 10 |  |
| **product\_width\_cm** | Ancho del producto en centímetros | integer | 14 |  |

#### Tabla sellers

Información sobre los vendedores que ofrecen productos en Olist Store.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ****Nombre de la Columna**** | ****Descripción**** | ****Tipo de Dato**** | ****Ejemplo del Dato**** | ****Clave**** |
| **seller\_id** | Identificador único del vendedor | string | "1f50f92002b6aeb04..." | Clave primaria |
| **seller\_zip\_code\_prefix** | Prefijo del código postal del vendedor | string | "13023" | Clave foránea (geo.cep\_prefix) |
| **seller\_city** | Ciudad del vendedor | string | "campinas" |  |
| **seller\_state** | Estado del vendedor | string | "SP" |  |

#### Tabla datos product\_category

Esta tabla contiene información sobre las categorías de productos disponibles en la plataforma, tanto en portugués como en inglés.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Nombre de la Columna** | **Descripción** | **Tipo de Dato** | **Ejemplo del Dato** | **Clave** |
| **product\_category\_name** | Nombre de la categoría de productos en portugués | string | "beleza\_saude" | Clave primaria |
| **product\_category\_name\_english** | Nombre de la categoría de productos en inglés | string | "health\_beauty" |  |

A continuación, se adjunta la imagen que muestra el esquema de datos de las tablas descritas anteriormente y se identifica las relaciones que existen entre ellas.

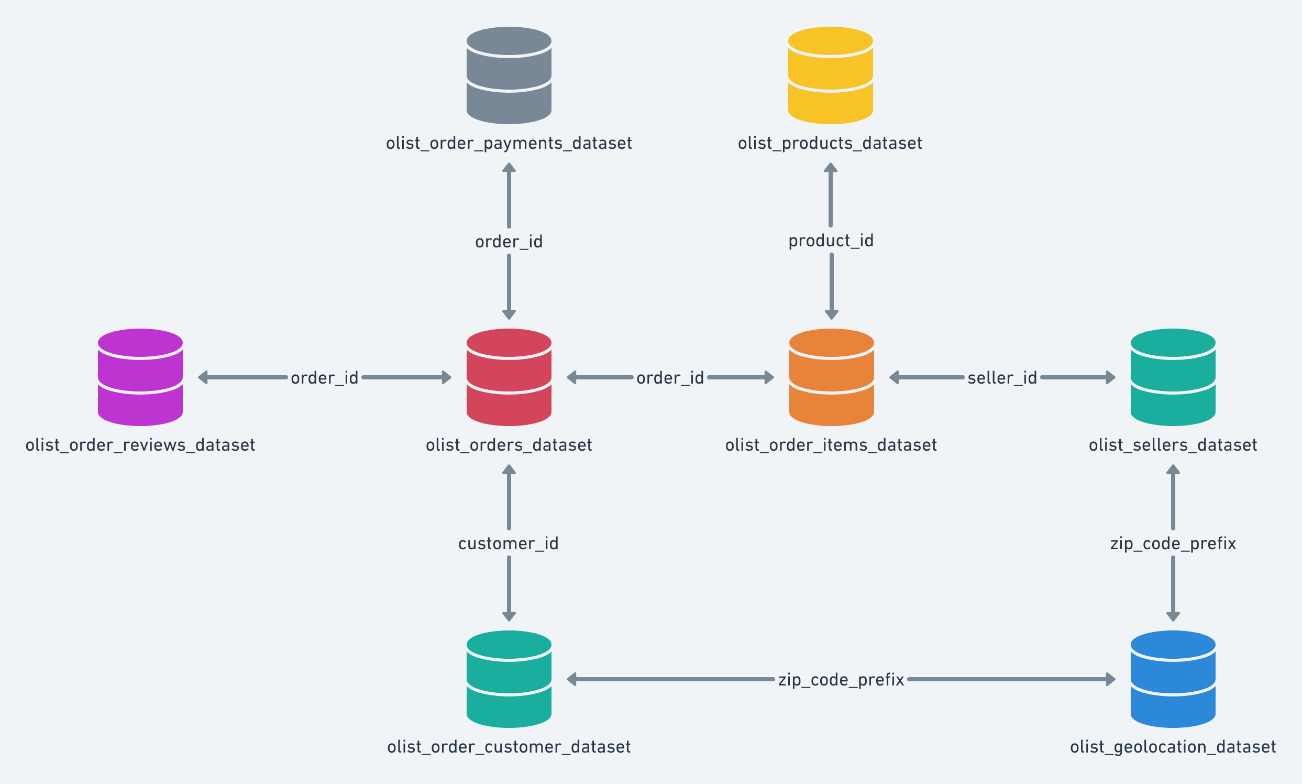


Imagen "Esquema de datos de insumos"

## 3.1 Tratamiento inicial para conexión de datos.

El primer paso en el proceso fue garantizar que los datos estuvieran preparados y organizados para un uso eficiente en los análisis. Todas las bases de datos se obtuvieron del enlace <https://www.kaggle.com/datasets/olistbr/brazilian-ecommerce>. Además, utilizamos el archivo log.csv, que contenía información actualizada de los códigos postales (CEP), adquirido a través de un vendedor por 9 dólares que realiza raspado de datos.

Para preparar los datos, utilizamos cuatro scripts, cada uno desempeñando un papel esencial en las etapas de estandarización, transformación y enriquecimiento de los datos de geolocalización. La conversión de los archivos originales en formato CSV al formato Parquet, que ofrece una estructura más eficiente para el almacenamiento y consulta de grandes volúmenes de datos, fue una de estas etapas fundamentales.

Script 1: Comenzamos identificando y tratando los archivos que necesitaban estandarización. Por ejemplo, los prefijos de los códigos postales (CEP) recibieron ceros a la izquierda para mantener la consistencia en todos los registros. El archivo log.csv se incorporó en este proceso, asegurando que su información estuviera alineada con las demás bases. Después de esta estandarización, los archivos se convirtieron y guardaron en una nueva carpeta en formato Parquet, optimizando el acceso y la manipulación de los datos para las etapas posteriores.

Script 2: A continuación, nos centramos en la transformación y unión de los datos de geolocalización. Combinamos diferentes fuentes de datos, incluyendo log.csv, estandarizando y ajustando las coordenadas de latitud y longitud para garantizar la conformidad. Este proceso fue esencial para asegurar que los datos geográficos estuvieran alineados, permitiendo el cruce con otras tablas y la realización de análisis más complejos, como análisis espaciales.

Script 3: Con los datos estandarizados, enriquecimos la información utilizando la API de Nominatim para obtener coordenadas geográficas precisas. A partir de información parcial, como los prefijos de los CEP, fue posible obtener las coordenadas exactas de cada registro, incluyendo los datos presentes en log.csv. Esto posibilitó un análisis más detallado de las ubicaciones de clientes y vendedores.

Script 4: Por último, refinamos y unificamos los datos. Leímos los archivos geolocation\_correios.parquet y geo\_coords.parquet, realizando un inner join basado en la columna cep\_prefix para asegurar la correspondencia entre las ubicaciones y las coordenadas geográficas. Después de la unión, las columnas de latitud y longitud se convirtieron al tipo de dato double, garantizando una mayor precisión en el tratamiento de las coordenadas geográficas.

Los datos enriquecidos se consolidaron y guardaron en el archivo geolocation\_correios\_coords.parquet. Con las columnas de latitud y longitud ajustadas al tipo de dato adecuado (double), aseguramos la precisión necesaria para análisis posteriores. Los datos están ahora listos para análisis espaciales, con precisión garantizada para consultas y cruces eficientes.

# Análisis Exploratorio de los datos.

# Maqueta del dashboard.

# Repositorios del proyecto.