

Sprint 11. Avaluació final per passar a projecte.

Presentació del conjunt de dades escollit: explicació i observacions generals del conjunt de dades a utilitzar per a fer el projecte final.

Nuestro conjunto de datos ha sido facilitado por la empresa Olist, el mayor especialista de plataformas online o marketplaces de Brazil. Olist ofrece a pequeños negocios de todo Brasil vender en sus marketplaces y enviar también sus productos a través de su red de logística de una forma fácil y con un contrato sencillo. Después de recibir el producto se le envía al cliente una encuesta de satisfacción.

Los datos comerciales son reales y se han anonimizado, así como las referencias a las compañías y los socios.

El conjunto de datos que utilizaremos para el proyecto se puede descargar en el siguiente link:

https://www.kaggle.com/datasets/olistbr/brazilian-ecommerce?datasetId=55151&sortBy=voteCount&select=olist_order_items_dataset.csv

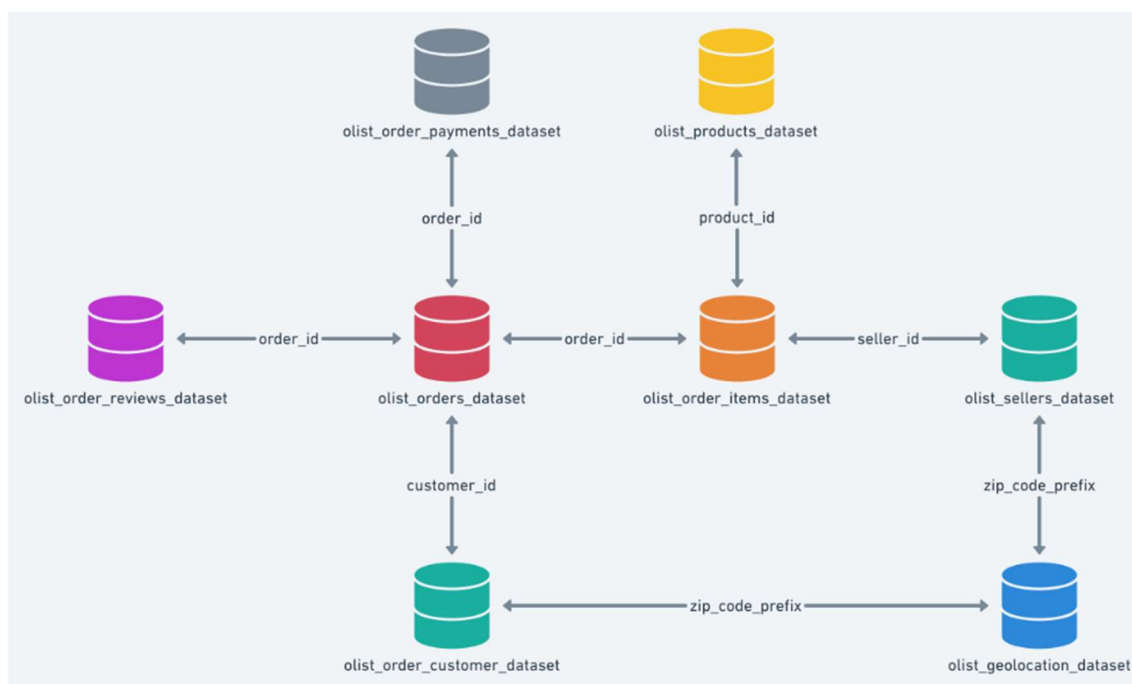
Característiques generals: sense necessitat d'entrar molt en detall, explicar les principals característiques que defineixen aquest conjunt de dades. Tipologia, sector, tipus de dades, font, context, etc.

El dataset tiene 99441 pedidos realizados desde 2016 a 2018, con la posibilidad de haber varios artículos por pedido gestionados por distintos vendedores.

La información relativa a los pedidos como el estado, el precio, el pago o el envío, así como la información relacionada con los artículos vendidos, localización de los clientes y sus opiniones sobre el servicio, se organiza en diferentes archivos:

- olist_customers_dataset.csv
- olist_geolocation_dataset.csv
- olist_order_items_dataset.csv
- olist_order_payments_dataset.csv
- olist_order_reviews_dataset.csv
- olist_orders_dataset.csv
- olist_products_dataset.csv
- olist_sellers_dataset.csv
- product_category_name_translation.csv

Los archivos se relacionan según el siguiente esquema:



El dataset ocupa 15.44 MB, tiene asignado el DOI [10.34740/kaggle/dsv/195341](https://doi.org/10.34740/kaggle/dsv/195341) y se distribuye bajo la licencia Creative Commons CC BY-NC-SA 4.0.

Definició de les variables: explicació teòrica de les principals variables que conté el conjunt de dades.

Nuestro conjunto de datos tiene múltiples variables. Hay variables que son identificadores únicos de clientes, pedidos, vendedores, artículos, opiniones, otras categóricas relacionadas con las categorías de los artículos o el tipo de pago, otras nominativas, con la localización de los clientes o el destino de los pedidos, otras numéricas con el precio o dimensiones de los artículos.

Presentació dels objectius: detallar els objectius inicials marcats de cara a extreure informació rellevant del conjunt de dades.

El objetivo del proyecto es extraer de los datos la información que nos permita aumentar ventas y aumentar ganancias. En esta línea nos proponemos los siguientes puntos:

- Predecir el precio de un producto, proponemos un algoritmo de machine learning de regresión para resolver este objetivo.
- Segmentar a los clientes según su comportamiento con el análisis RFM, recency-frequency-monetary value, en función de sus compras, la frecuencia de las compras y el valor de sus compras. Proponemos un algoritmo de machine learning para construir un modelo predictivo con la librería de Python Lifetimes para resolver este objetivo.

- Calcular el CLV, Customer Lifetime Value, de los clientes para plantear estrategias de fidelización. Proponemos un algoritmo de machine learning para construir un modelo predictivo con la librería de Python Lifetimes para resolver este objetivo.
- Predicción de ventas del siguiente año, forecasting con la librería Prophet.

También nos gustaría como objetivos complementarios:

- Optimizar el precio para un producto. Proponemos un algoritmo de machine learning de regresión para resolver este objetivo.
- Predicción de ventas diarias, necesario para saber qué superficie de almacén necesitamos, qué volumen de repartidores aproximado, etc... Proponemos un algoritmo de machine learning de regresión para resolver este objetivo.
- Predecir las compras de un cliente nuevo