Data guide:

About the data :

The dataset has information of 100k orders from 2016 to 2018 made at multiple marketplaces in Brazil. Its features allow viewing an order from multiple dimensions: from order status, price, payment and freight performance to customer location, product attributes and finally reviews written by customers. We also released a geolocation dataset that relates Brazilian zip codes to lat/lng coordinates. This is real commercial data, it has been anonymised, and references to the companies and partners in the review text have been replaced with the names of Game of Thrones great houses.

**Customers Dataset**

This dataset has information about the customer and its location. Use it to identify unique customers in the orders dataset and to find the orders delivery location.

At our system each order is assigned to a unique customer*id. This means that the same customer will get different ids for different orders. The purpose of having a customer* unique\_id on the dataset is to allow you to identify customers that made repurchases at the store. Otherwise you would find that each order had a different customer associated with.

**Geolocation Dataset**

This dataset has information Brazilian zip codes and its lat/lng coordinates. Use it to plot maps and find distances between sellers and customers.

**Order Items Dataset**

This dataset includes data about the items purchased within each order.

Example:

The order\_id = 00143d0f86d6fbd9f9b38ab440ac16f5 has 3 items (same product). Each item has the freight calculated accordingly to its measures and weight. To get the total freight value for each order you just have to sum.

**The total order\_item value is:** 21.33 \* 3 = 63.99

**The total freight value is:** 15.10 \* 3 = 45.30

**The total order value (product + freight) is:** 45.30 + 63.99 = 109.29

**Payments Dataset**

This dataset includes data about the orders payment options.

**Order Reviews Dataset**

This dataset includes data about the reviews made by the customers.

After a customer purchases the product from Olist Store a seller gets notified to fulfill that order. Once the customer receives the product, or the estimated delivery date is due, the customer gets a satisfaction survey by email where he can give a note for the purchase experience and write down some comments.

**Order Dataset**

This is the core dataset. From each order you might find all other information.

**Products Dataset**

This dataset includes data about the products sold by Olist.

**Sellers Dataset**

This dataset includes data about the sellers that fulfilled orders made at Olist. Use it to find the seller location and to identify which seller fulfilled each product.

**Coupon Dataset**

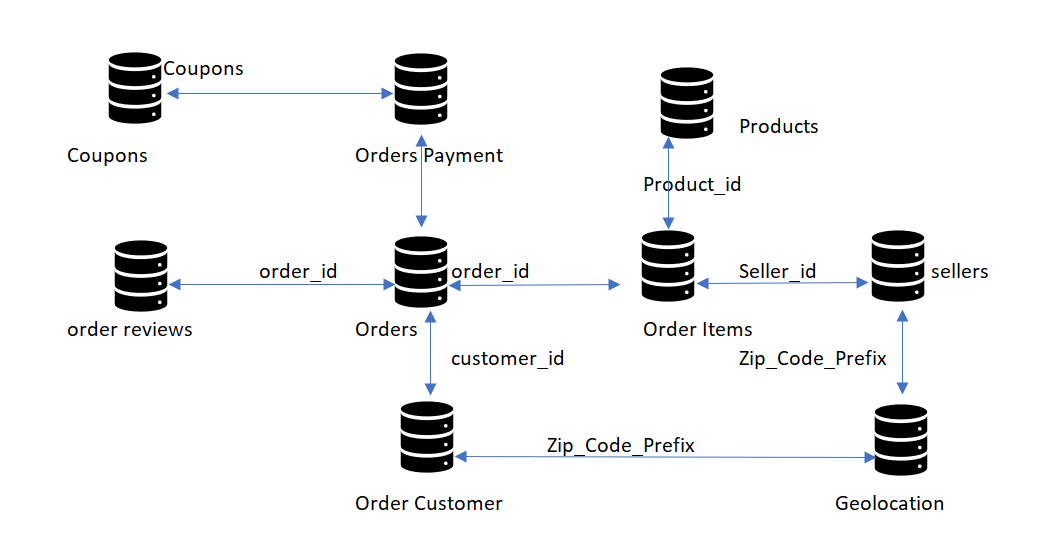
This dataset includes Coupons which were related to various orders. This lists the various discount percentage. The sales value in the orders are after the discount has been applied. All coupons are assumed to be active.

**Category Name Translation**

Translates the productcategoryname to english.

An order might have multiple items.

Each item might be fulfilled by a distinct seller.



Few hints and helps:

1. There can be discrepancies in the data. Please take a wise/best decision for analysis.
2. For your problem statement all data may not be required. Its upto your creativity how you want to use the data.
3. Black Friday in Brazil always happens on the fourth Friday of November

# Python program to Find day of

# the week for a given date

import datetime

import calendar

def findDay(date):

    born = datetime.datetime.strptime(date, '%d %m %Y').weekday()

    return (calendar.day\_name[born])

# Driver program

date = '03 02 2019'

print(findDay(date))

1. Merge, join, concatenate and compare : <https://pandas.pydata.org/pandas-docs/stable/user_guide/merging.html>
2. <https://www.youtube.com/watch?v=QUT1VHiLmmI> – Python