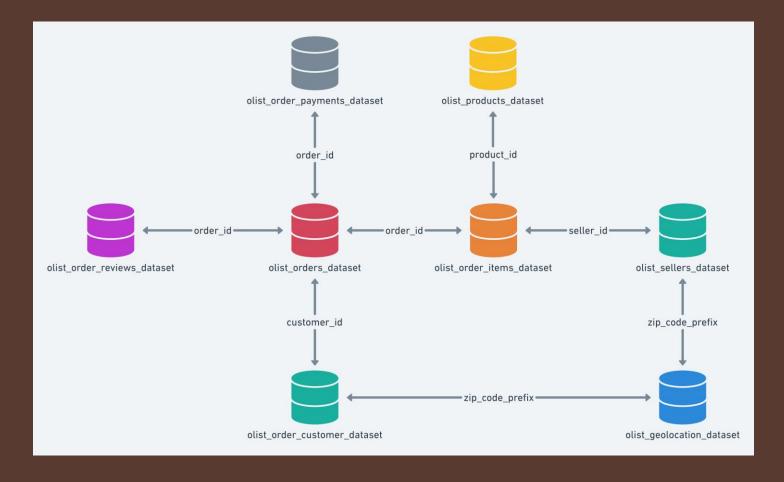


Olist Dataset BigQuery and Data Visualization

Chiao-I Lin

Introduction

 Brazilian E-Commerce Public Dataset by Olist Welcome! This is a Brazilian ecommerce public dataset of orders made at Olist Store.



• Data source: https://www.kaggle.com/datasets/olistbr/brazilian-ecommerce

Outline

- Tools
- List of Visualization
- Visualization



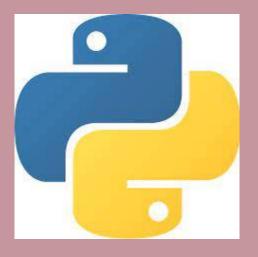


Tools

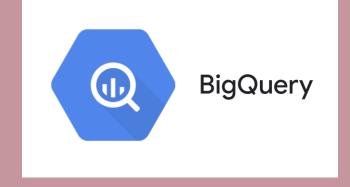
- Jupyter
- Python
- Pandas
- Numpy
- Matplotlib.pyplot
- Seaborn
- GoogleBigCuery















About this project

SQL code

- Using google BigQuery to organize data
- And applying Seaborn and Matplotlib to visualize the data
- The structureof code and output in Jupyter notnook look like the chart in right

Code for data visualization

data visualization

Table obtained by SQL

Jupyter notnook

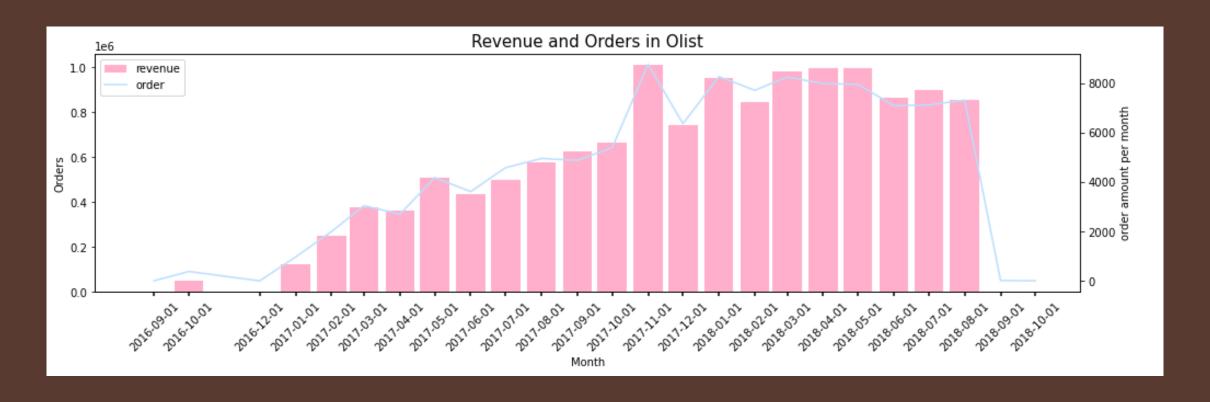
```
SELECT DATETIME_TRUNC(order_purchase_timestamp, MONTH) AS date_time
                  EXTRACT(YEAR FROM order_purchase_timestamp) AS year,
EXTRACT(MONTH FROM order_purchase_timestamp) AS month
                  COUNT(*) AS order per month.
                   SUM(price) AS revenue
      FROM `nod-sql-copy.olist.orders`
      LEFT JOIN `nod-sql-copy.olist.order_items` USING(order_id)
       plt.figure(figsize = (20,4))
     plt.righte(righte - (20,47))
sas.barplot(data = df, x = 'date_time', y = 'order_per_month')
plt.xticks(ticks = [ 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24], labels = df.date_time.dt.date, rotation = 45)
plt.ylabel('order amount per month')
      plt.xlabel('Month')
       plt.title('Order amount per month', fontsize = 15)
       plt.show()
       plt.figure(figsize = (20,6))
      sms.limeplot(data = df, x = 'date_time', y = 'revenue')
""plt.xtickg(ticks = [ 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 7, 18, 18, 19, 20, 21, 22, 23, 24], labels = df.date_time.dt.date, rotation = 45)"""
       plt.ylabel('Revenue')
      plt.xlabel('Month')
      plt.title('Revenue per month', fontsize = 15)
       plt.show()
Therefore the current project will focus on year 2017
```

966 120312.87

List of visualization

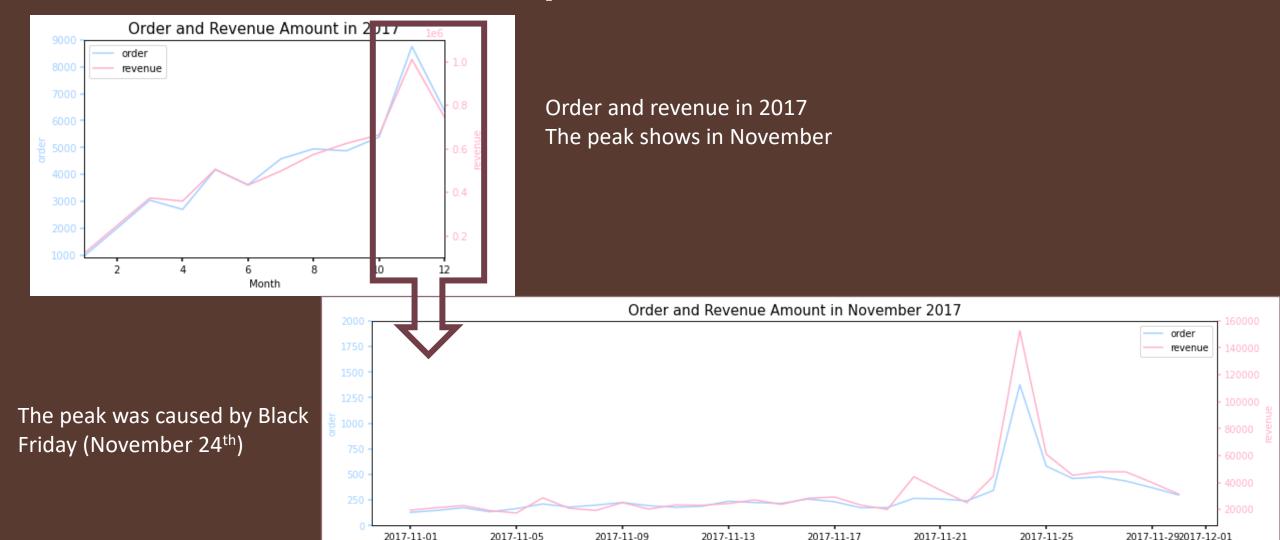
- Data: data per year
- Percentage change of order number/revenue per month in 2017
- Monthly Revenue in 2017: KDE plot
- Cancellation per month in 2017
- 2017 best monthly seller
- 2017 best selling state
- Monthly order per state
- Majority payment method in 2017
- Most popular product category

Data: data per year



Some data in 2016 and 2018 are missing. Therefore, the current project will focus on year 2017

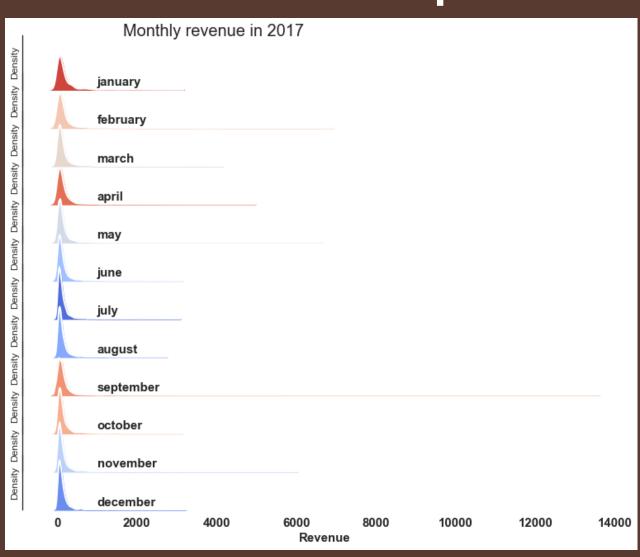
Percentage change of order number/revenue per month in 2017



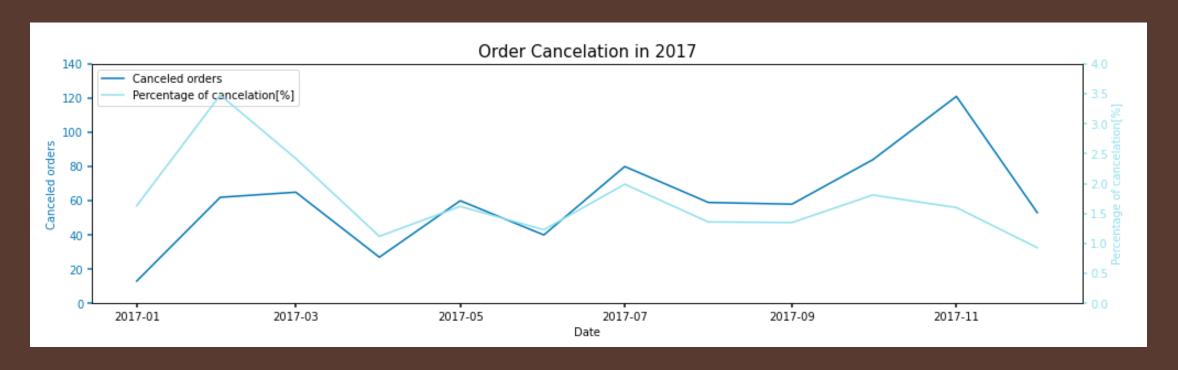
Monthly Revenue in 2017: KDE plot

It seems like the mean of revenue per order are similar in each month.

The increasement of revenue may caused by the order amount.



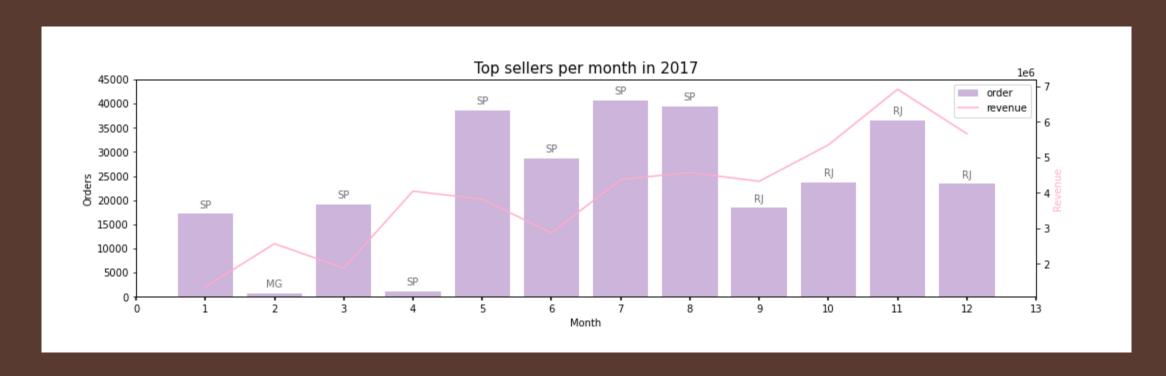
Cancellation per month in 2017



Finding:

- In February percentage of cancelation is relatively heigh.
- In November the amount of canceled order is heigh but the total order is also heigh. Therefore, the percentage of cancelation is not that heigh

Popular product category in 2017



The chart show the top seller per month in 2017.

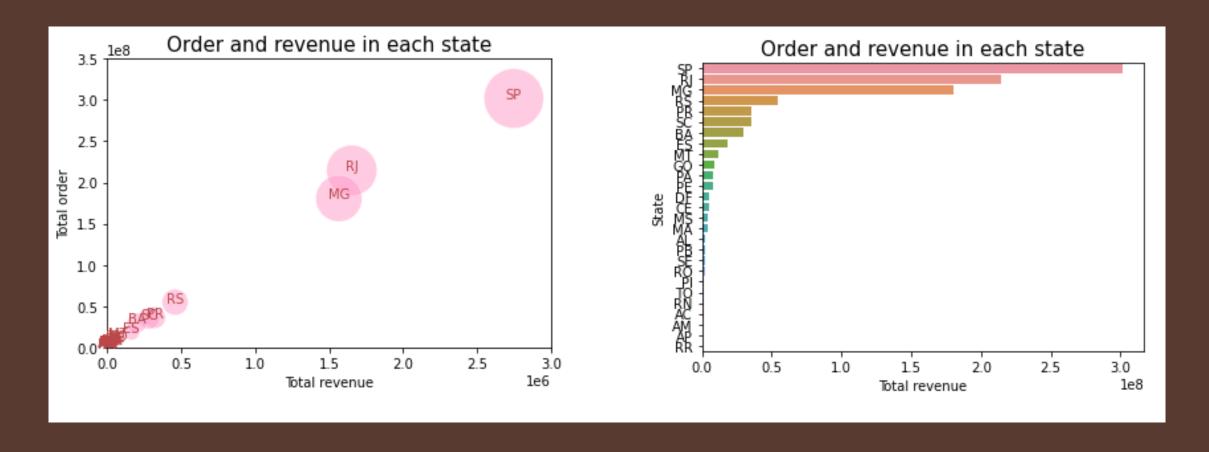
The ids of top sellers show in the table

The chart shows the total orders of the top seller in each month (left axis)

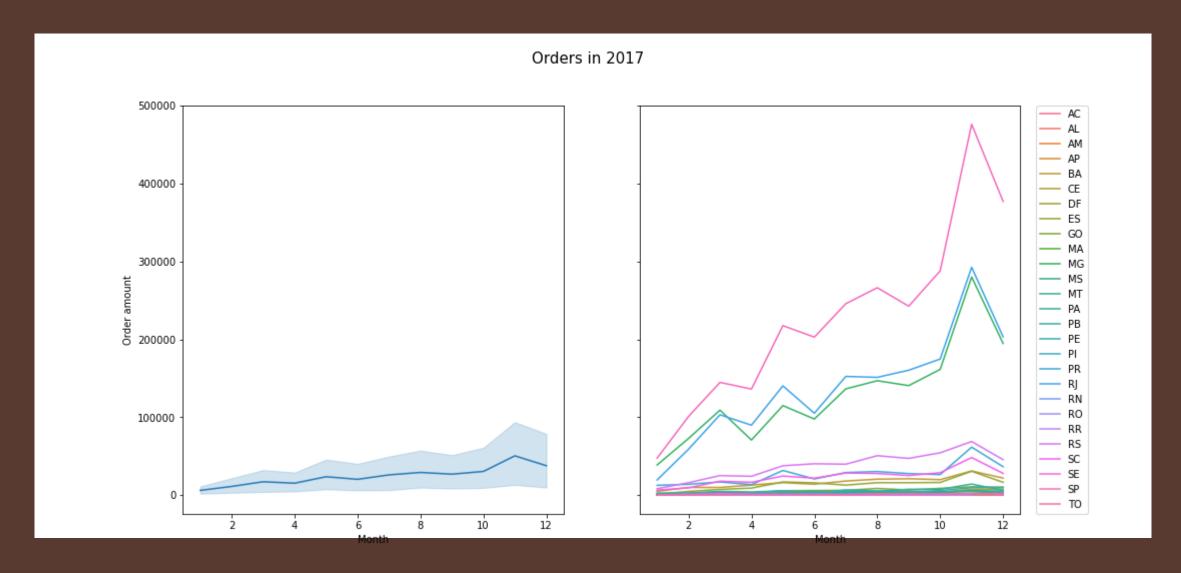
and the total revenue of the top seller in each month (right axis)

Texts above bars are the state where sellers located

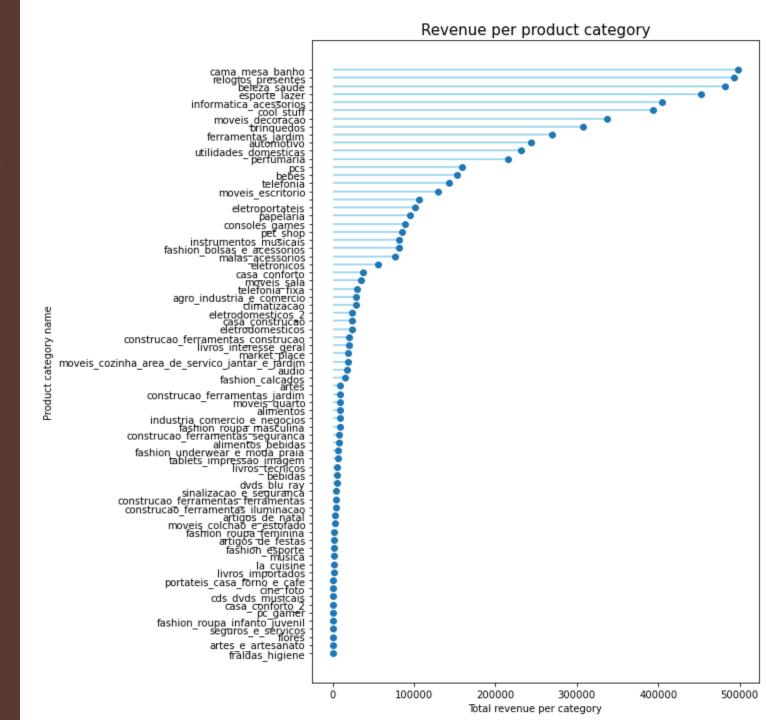
2017 best selling state



Monthly order per state



Most popular product category



Majority payment method in 2017

