Deriving Brazilian E-Commerce Insights with Olist

ZHANG Jieyuan, LIU Chentao, KWONG Mei Lin

Overview

Olist is a company operating in the Brazilian E-commerce segment aiming to be the largest online shopping platform in Brazil.

With fast growth and wide range of product offered, this tool aims to provide Olist a clear visualisation of its business performance from various aspects with different data granularity to facilitate future growth and identify opportunities. Insights at product segment level also enables Olist to further improve its supply chain and campaign management.

Issues and Problems

1. How is the current business performance in terms of customer activity, order value and number of orders?
2. What are the insights to drive further sales uplift in the future?

Motivations

Brazil is currently the largest e-commerce market in Latin-America and the 13th largest in the world despite a low internet penetration of 61%. This suggests an immense potential for growth in the Brazilian E-commerce landscape.

Using the Olist data from 2016-2018, our group aims to:

1. Learn about the Brazilian E-commerce market/consumers preferences/purchase trends
2. Discover Olist's potential areas of growth and improvement
3. Apply visualisation/analytics techniques to deliver the information

Approaches

We aim to analyze the activities and performance of Olist through the following metrics with a focus on Sao Paulo (SP) state.

1. Gross Merchandise Value (GMV) (derived from the sum of value of purchases made)
2. Number of orders
3. Number of customers

We take the below analysis approaches to construct relevant visualisations and display them in a dashboard using R shiny:

Descriptive Analysis
↓
Geovisual Analysis
↓
Sales Forecast

Results

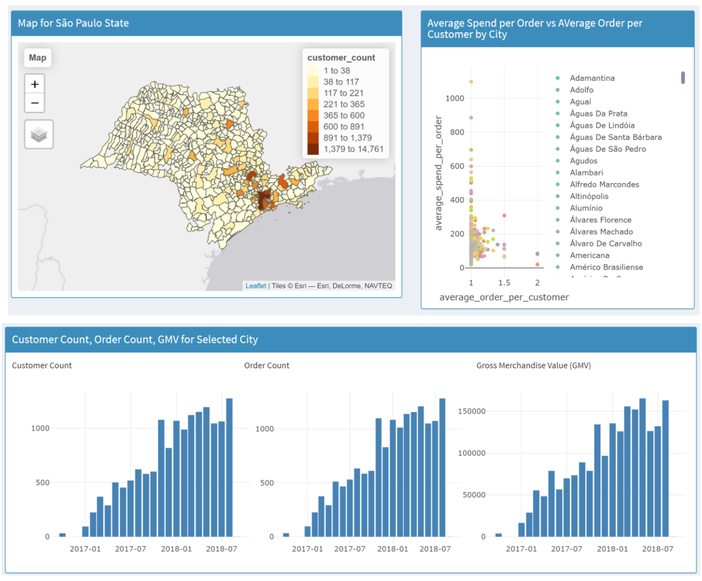
Section 1: Descriptive Analysis

Users can have a quick glance of the business performance metrics e.g. total orders/total gross merchandise value by product category or customer city.

By Category: Trends and ranking by categories are shown in line plot and bar plots enabling user to understand top categories that brought in volume (order counts) vs those that contributed the highest value (GMV) hence deriving insights to identify key product categories to focus.

By Customer City: Growth trend of order and GMV can be compared across multiple cities in Brazil to identify key markets for expansion.

Section 2: Geovisual Analysis



Section 2 focuses on the Sao Paulo state.

Interactive Choropleth Map: Thematic map is used to enable user to compare the performance metrics across different cities for selected period. User may toggle to view the map by customer count, order count or GMV. Geobr, Tmap packages are used.

Scatterplot: Comparing average spend per order vs average order per customer across the different cities. Olist may want to focus on building customer base in cities with high spend and high order per customers, assuming ceteris paribus in other considerations.

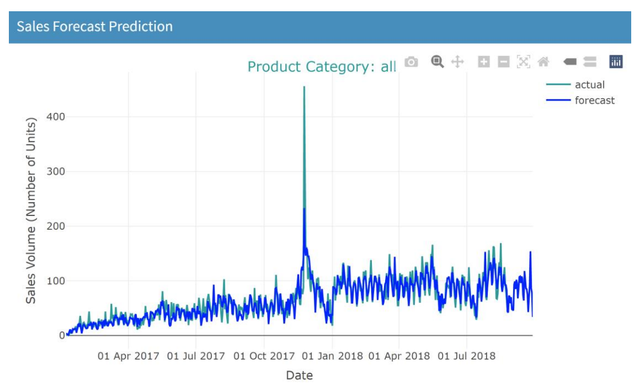
Bar plot: Comparing trends across different metrics for selected city(ies).

Results Con't

Section 3: Sales Forecast

Lastly, users are able to forecast the sales volume of a category of choice for the following one month, and project the sales volume for a given day of the month for a particular category.

The in-built Prophet package model was used for our sales forecast, which makes use of a time-series additive model which accounts for yearly, weekly, seasonal and holiday effects.



Insights

1. The product category *bed_bath_table* had the highest total orders while it was not the highest GMV contributor in 2017. This could be due to its low unit price. In contrast, *health_beauty* and *watch_gift's* order count grew exponentially and became the top 2 contributors in GMV. Hence, olist may want to focus on fast growing categories with high value to achieve fastest gain in GMV.

2. The city of Sao Paulo and its surrounding cities placed the most orders within the state of Sao Paulo. This could be due to people from Sao Paulo cities and nearby having higher purchasing power.

3. Leveraging the scatter plot in section 2, the order per customer in most cities is ~1-2 although some cities may have a high average spend per order. Two key areas of focus for Olist may be building customer loyalty and getting them to make repeated purchases.

4. The actual sales didn't match the forecast near the end of November in 2017. This is likely due to the occurrences of sale events (black Friday sales). Predictive models may not work for special/outlier events if they were not considered during the training of model.

Future Work

We plan to extend the scope the project to include the data after 2018. Specifically, we would like to compare consumers' preferences and purchase trends pre-Covid and mid-Covid to see the impact of the pandemic (occurred Dec. 2019) on the E-commerce market.

Singapore Management University

School of Computing and Information System