Market Basket Project on E-Commerce

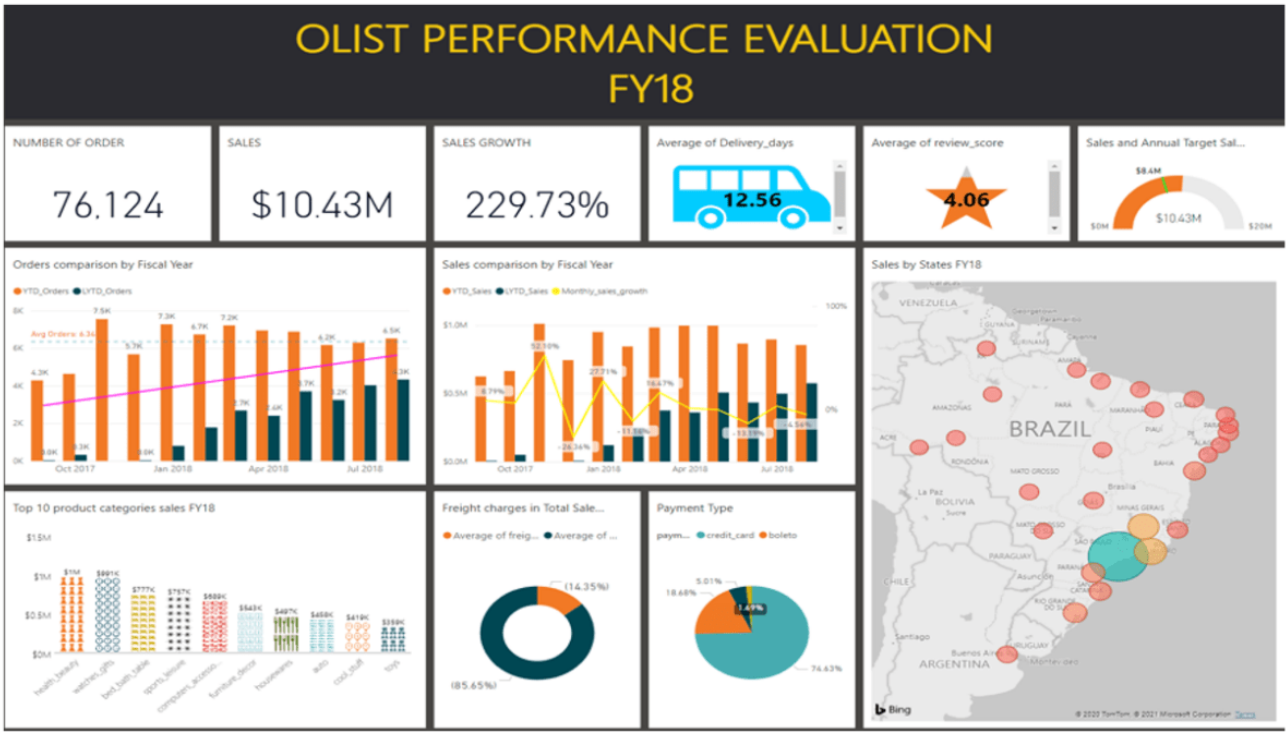
Brazilian E-Commerce Public Dataset by Olist

Wireframe Documentation

Homepage

Being able to make accurate predictions of future revenue can be hugely important for businesses. For example, if we can forecast the next month's revenue, we can inform the decision-making process across all areas of the business, from purchasing decisions,, marketing activity, staffing levels or which invoices to pay and when.

In this notebook, we'll use this monthly sales dataset to build a forecast model to predict the next year's sales revenue of the Company Olist, which operates in Brazil.



Joining data:

The original data is used to derive multiple datasets by joins and manipulations. The Data manipulation and combining jupyter notebook contains the step by step process and explanations.

customer_data - Maps customers to their locations in latitude and longitude
customer_order - Maps each order to the customer and the products bought
transaction_data - Transactional dataset
customer_payment - Information on purchases and payment information aggregated for each customer
delivery_data - Maps each customer's order to the seller fulfilling it and the seller's location
product_reviews - Maps the reviews to the products
Preliminary data analysis
The notebook Preliminary Data analysis contains the detailed analysis. This notebook visualizes and summarizes the original and the combined datasets, to find trends, patterns or faults. This analysis gives a holistic view of the dataset.

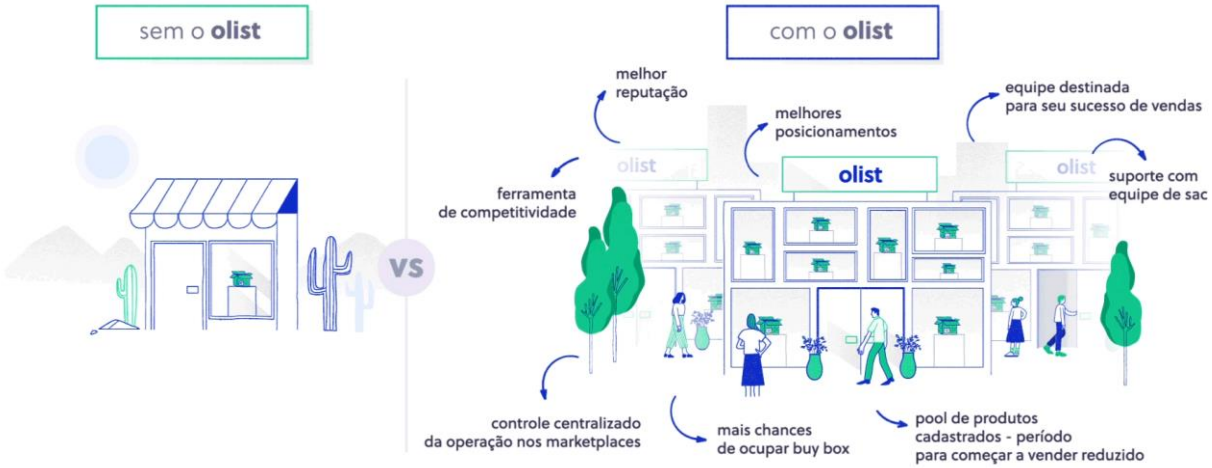
Association rule mining
Using the transactional dataset created from the original dataset, association rule mining is performed using frequent pattern algorithms - FP growth trees. The model returns frequent item sets with a confidence threshold of 10%. The Association rule mining notebook contains the detailed explanation for this task.

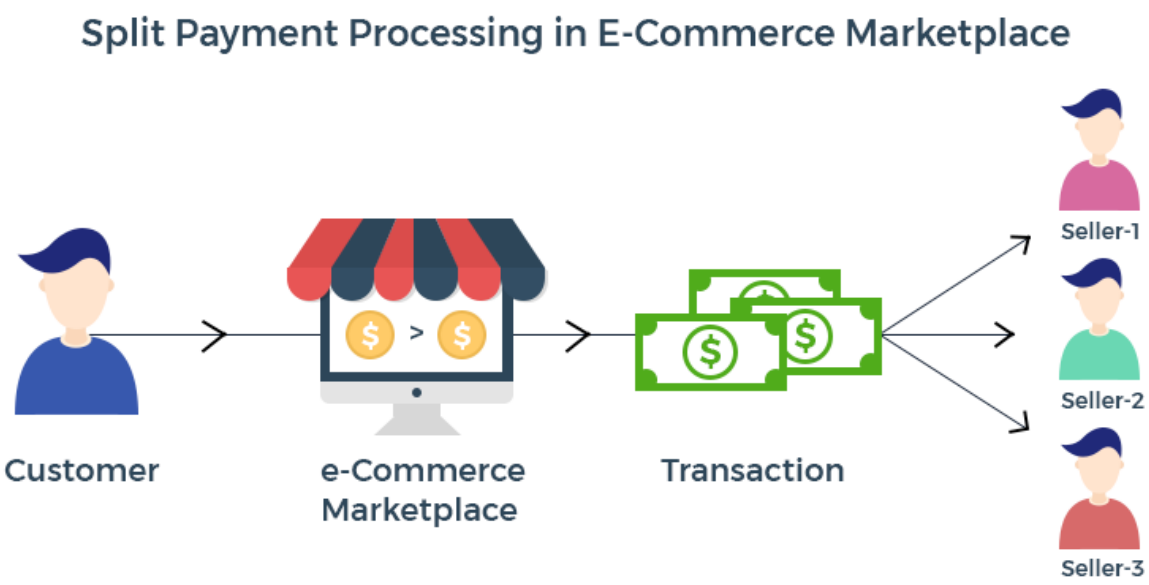
Product analysis
The objective of this analysis is to find the most popular products, popular product categories and category wise popular products in the Olist ecosystem. Further, the delivery times and product characteristics such as - description lenght, number of photos etc, are compared to popularity to find correlations in the data. The Product analysis notebook contains the detailed code.

Reviews sentiment analysis
Sentiment analysis is carried out on the reviews offered by customers. The notebook contains Supervised and Unsupervised methods for sentiment analysis, Reviews Sentiment Analysis. The supervised technique uses the rating provided as label for sentiment analysis, whereas the unsupervised technique lexicons- AFINN, TextBlob to perform the analysis.

Freight value prediction
The freight value is the shipping value associated with each order. The Freight value prediction notebook contains detailed model building steps to predict the shipping value for an order, given the distance between seller and customer, the dimensions and weight of the product. Further, the model is also build and deployed on AzureML as a web service. A webapp is created using Dash plotly, to query the model for freight value predictions.

Web application
The web application is deployed on Heroku, built in Dash plotly. The application provides a front end to the freight value predictor model built on the AzureML platform. The web application has the following file structure.





__ app.py
__ layout.py
__ requirements.txt
__ Procfile
__ Resources
  |__ API_key
  |__ URL
  |__ ship.png

The schema below is of the web service built on AzureML.

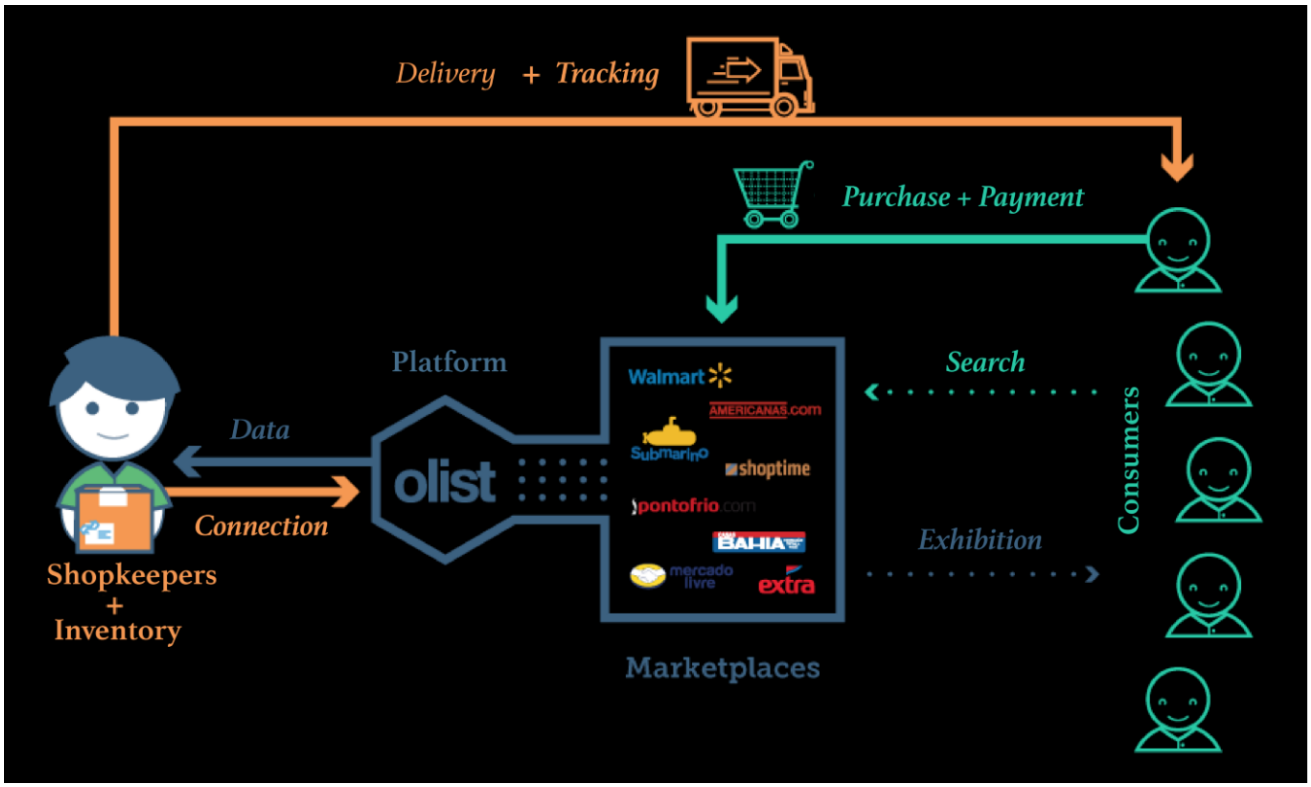
Heroku deployment
The app.py file contains the server instance and the layout.py file contains the layout to be rendered. The requirements.txt file contains the dependencies of the web application

pip freeze > requirements.txt
The Procfile is used to launch an instance of the app server onto the virtual machine. The contents of the Procfile are

web: gunicorn app:server
Once the files are ready, push the repo to Heroku master

heroku create new_app
git add .
git commit -m "intial push"
git push heroku master
heroku ps:scale web=1
The web application, deployed can be accessed on Heroku via this link. https://olist-freight-app.herokuapp.com/

NOTE:- The Web service on Azure is not currently, up and running. Kindly create a similar web service on Azure and change the API_key and URL files accordingly.



Olist's business model

CONCLUSION

- We have gone through an extensive journey for understanding the data and plotting some useful charts to clarify the concepts and get insights from data.
- Credit card is most used for payments in Olist which is true because Credit Card Company provides most amount of offers for their customers.
- There are 74 different category present in Olist from where Brazilians can shop.
- It seems transportation cost (freight value) in Olist is reasonable getting max up to 70 Brazilian real.
- Olist has nearly 97% successful delivery rate which shows there commitment towards the order and making this amount of successful deliveries will help them grow in near future.

FUTURE WORK

- In near future we can look at all the reviews and comments and will try to analyze what people thinks about Olist store.
- We can also use this dataset to map differently on graph as we have geo location of every pin code we can explore more into it.
- One more thing we can do is to list pricing of the products and tends to compare it with freight values which will be next step to analyze transportation costs.
- Last but not the least we can try for a new Olist dataset and comparing it will give us result whether it's truly a genuine dataset or not.