**Project 1**

**Abstract**

**Instacart Market Basket Analysis Prediction**

Whether you shop from scrupulous planned grocery lists or let whimsy guide your browsing, our unique food rituals define who we are. Instacart, a grocery ordering delivery app aim to make it easy to fill your ice chest with your personal favourites when you need them. Instacart’s data science team plays a big part in providing this delightful shopping experience. Instacart open-sourced this data of 3 Million Instacart Orders.

**Tools / Skills Used**

* Python Programming
* Jupyter Notebook
* Google Colab
* Pandas
* Numpy
* Matplotlib
* Seaborn
* Exploratory Data Analysis
* Data Visualization
* Machine Learning

**Problem Statement**

Instacart Market Basket Prediction: In this data science project, we are going to use this anonymized data on customer orders over time to predict which previously purchased products will be in a user’s next order.

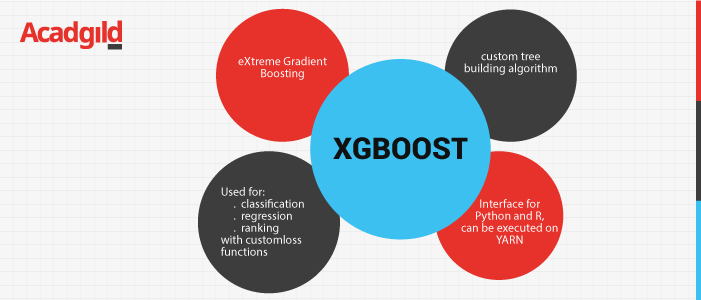
**Modelling**

From the owners of Instacart as mentioned by them they use XGBoost as one of their models to predict what will be in user’s next order we will also be going ahead with it.

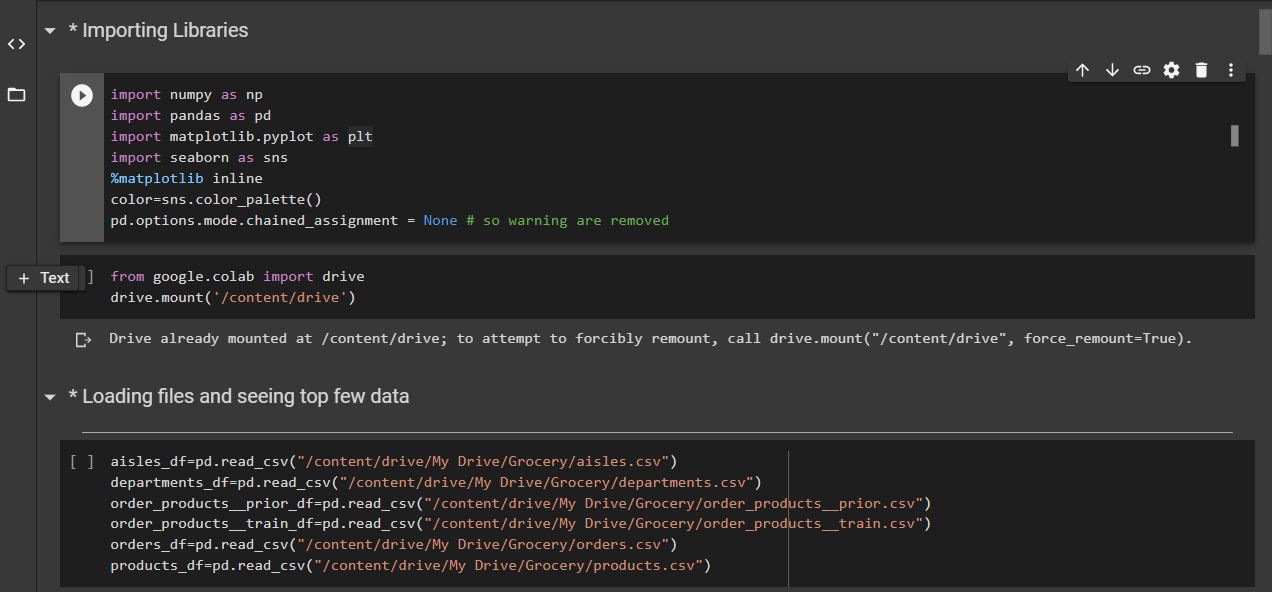
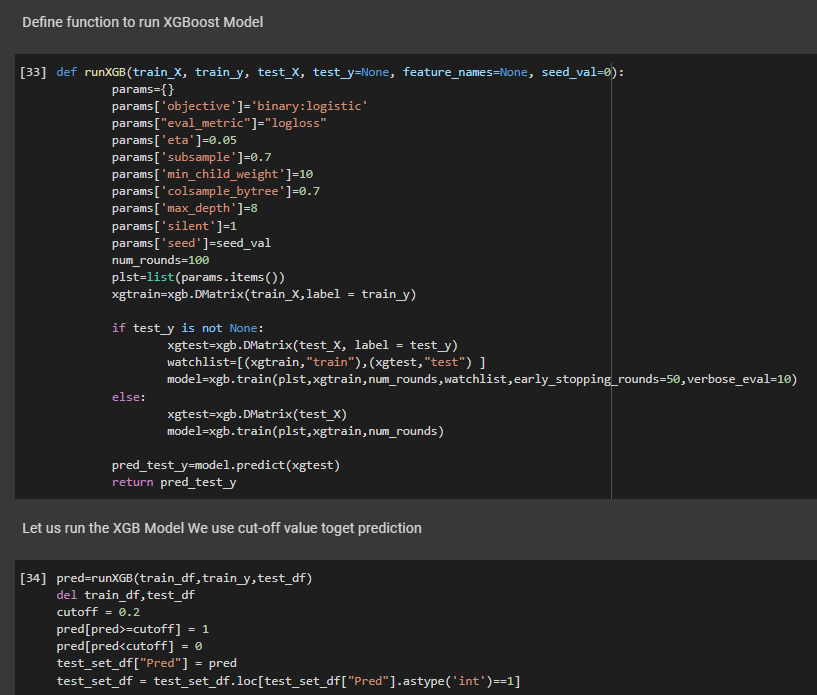
**XGBOOST**: Its name stands for **eXtreme Gradient Boosting**.  XGBoost is a scalable and accurate implementation of gradient boosting machines and it has proven to push the limits of computing power for boosted trees algorithms as it was built and developed for the sole purpose of model performance and computational speed. Specifically, it was engineered to exploit every bit of memory and hardware resources for tree boosting algorithms.

**Why use XGBoost?**

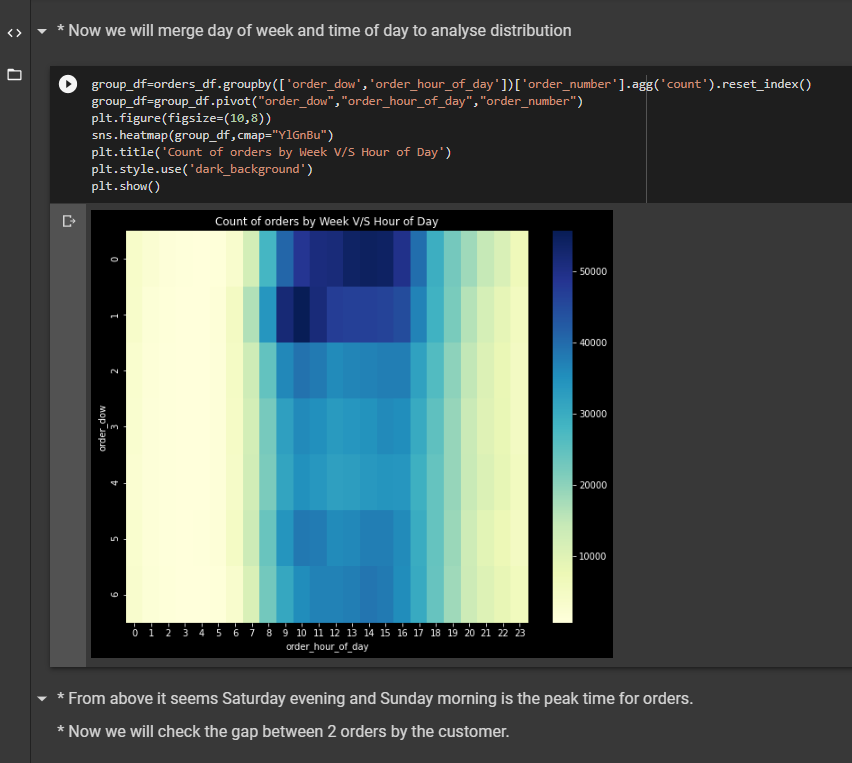
As we already mentioned, the key features of this library rely on model performance and execution speed.



**Code** **Snippets**

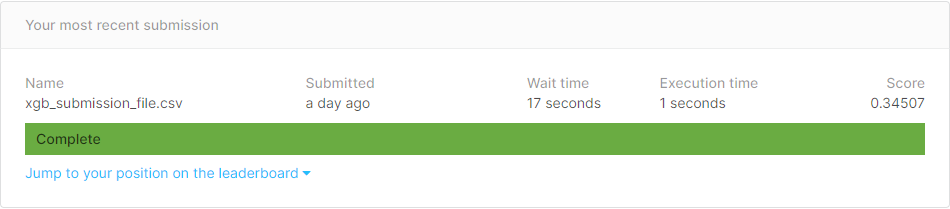
****

**Visualization Snippets**



**Conclusion / Result**

Our code run on **XGBoost** and submission file on **Kaggle** scored **0.34507** on the Leaderboard.



**Future Scope**

There is scope of improvement with better techniques and making the score reach near to **0.409** which is highest as on 25/05/2020 on Kaggle.