

UNDERSTANDING CUSTOMER SHOPPING BEHAVIOR

Project Write Up – Customer Segmentation & Order Forecasting Model

Summary

More Americans than ever before are choosing to shop for groceries from the comfort of their couch as the coronavirus pandemic sweeps the US and forces most of the country into lockdown. This is driving a significant spike in demand for online grocery delivery and pickup services and has seen the major players in the market, such as Walmart, Amazon, Instacart, and Target, scramble to service their new audiences in the past few weeks. Analysts say that the effects of this unusual period are likely to be long-lasting as more shoppers who have never used online services become increasingly used to these.

The increasing demand for online grocery shopping makes it difficult to fulfill user demands. A knowledge of consumer behavior can help these firms be on top of their inventory management and work in an optimum manner. For this, I have tried to build a recommendation system for Instacart which will help them predict the future orders of their consumers and manage their inventory responsibly.

A series of exploratory analysis was done to analyze user's behavior and understand their shopping patterns and unsupervised clustering method was used to cluster users. The segmentation helps understand the demand on a higher level. To dig into the user level, a forecasting model is built to predict order composition of each user. Here, statistical machine learning algorithms for predictive analysis have been used.

The outcome of this study will divide Instacart users into different segments based on their shopping behavior and predict future orders of these users. This information can then be used to handle value-chain, from inventory to manpower requirement, in an efficient way.

Project Objective

- Purpose/ Business Statement:

The purpose of the project is to optimize effectiveness of online grocery shopping market. Two major questions that are tried to answer are -

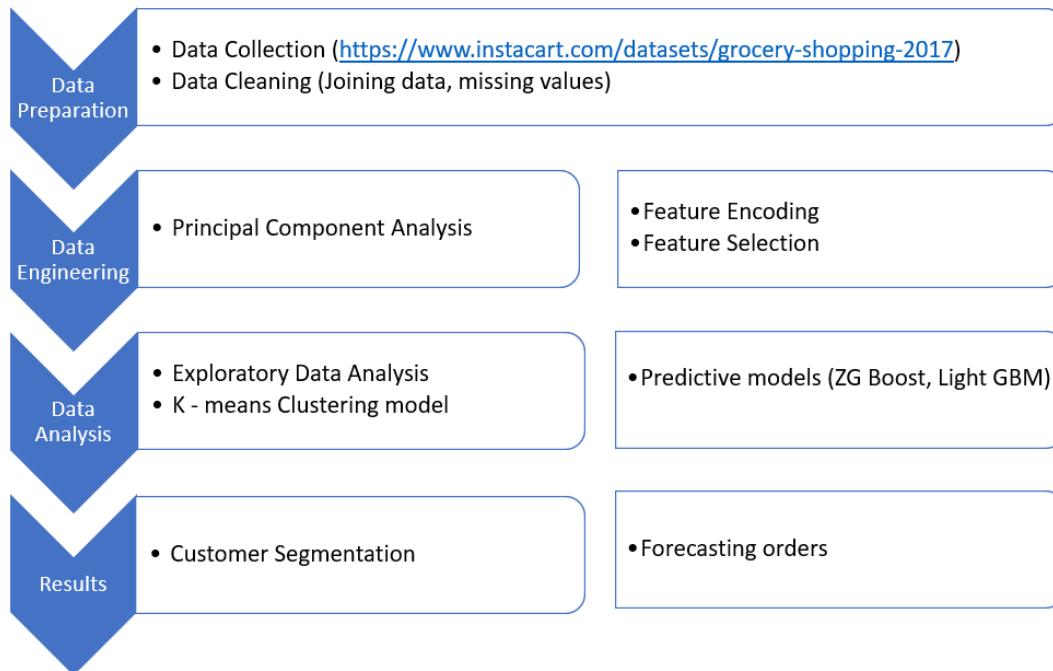
- How to predict the manpower requirement for order delivery?
- How to manage inventory in an efficient way?

- Numerical interpretation of the business problem:

The 2 questions were handled differently based on the nature of the question.

- Question1 – Customer segmentation was done depending on the pattern of their ordering. An unsupervised machine learning model was built to cluster all the customers into different buckets based on the frequency of their ordering.
- Question 2 – Customer segmentation was done to understand the composition of the orders and a predictive model was built to forecast future orders of the consumers.

Methodology



- Data Collection:

The data was collected from “<https://www.instacart.com/datasets/grocery-shopping-2017>”, an open source data provided by Instacart.

- Data Exploration:

- The dataset comprises of over 3 million grocery orders from more than 200,000 Instacart users. Each user has between 4 and 100 orders. Also the week and hour of day the order was placed, and a relative measure of time between orders.
- Data was explored by creating graphs between the available features like order composition, time of order, frequency of order, etc.
- These graphs help understand the behavior of the consumers and gain insights on the data.

- Data Transformation:

- The data was cleaned and structured in a format that can be used to build supervised and unsupervised models.
- For unsupervised model, the dimension of the data was reduced using principal component analysis and clusters were visualized based on these new components.
- For supervised learning, the data was transformed into numerical format using hot encoder and dimensionality was reduced by using feature selection.

This document is purely made for the project presentation and to be used internally only.

CUSTOMER SEGMENTATION & ORDER FORECASTING MODEL DOCUMENTATION BY SAKSHI JOSHI

- Modelling or tool building:

The tools were built using python libraries like pandas, kmeans, sklearn, etc. There are 2 major tools – clustering tool and predictive tool. The functionality of both these tools differ based on the business problem.

- The clustering tool was used to segment datapoints based on the principal components. The input to this tool was a dataset with 2 features that were derived using PCA model. The model is built using k-mean clustering algorithm which clusters data points into similar groups.
- The predictive tool was used to forecast orders for each user. The model was given a dataset with selected features, these features were selected based on feature importance. This model was built on advanced statistical predictive models like xgboost and light gbm. The performance of both these models were analyzed to choose to the best model for prediction.

Results

- The result of both the models will be shared differently.
- The output of the clustering model will a list of users divided into multiple clusters based on their characteristics. This output will give information about the eating and ordering habits of the consumers.
- The output of the forecasting model will be a prediction of future orders and its composition for every user which will determine the overall requirement of the inventory.