Instacart Market Basket Analysis

Project Description

Instacart is grocery ordering and delivery application, by serving the customers to order groceries from participating retailers like Sam's Club, Kroger, Aldi etc. with the shopping being done by a personal shopper.

For this project we are using the anonymized data of customer order on Instacart, which they have open sourced on Kaggle. Here we are interested to predict the following -

- Which previously purchased products will be in a user's next order?
- Which product the user will try for the first time?
- What are the patterns in user consumption?
- What factors influence the purchasing decision (time of the day, day of the week)

Instacart essentially falls in the e-commerce industry, hence creating an efficient and smooth experience for a user would be a priority given the fact that every company tries their best for the same. If we try to find the solution to the questions mentioned above, it will help Instacart develop a systematic user experience and could be their USP (Unique Selling Proposition).

Dataset

The source of the dataset - https://www.kaggle.com/c/instacart-market-basket-analysis/data.

The dataset is anonymized and contains a sample of over 3 million grocery orders from more than 200,000 Instacart users. For each user, the orders provided between 4 and 100, with the sequence of products purchased in each order. It also contains the week and hour of day the order was placed, and a relative measure of time between orders.

Teammates and work division

Sidhant Kapoor – Exploratory Data Analysis (EDA), Project report.

Vedant Taneja – Data Pre-processing and manipulation, Project presentation.

Avinash Panigrahi – Model implementation, Project presentation.

Atharva Tipre – Data visualization, Model Implementation Project Proposal.

Methodologies

- We shall use Pandas and NumPy for data manipulation and analysis
- To explore data and estimate statistical models, we shall use Stats models
- To build recommender systems and regression algorithms, we shall use scikit-learn library
- Finally, we will build plots demonstrating data patterns using matplotlib/seaborn library