Proposal: Instacart Market Basket Analysis

Introduction:

Instacart, a grocery ordering and delivery app, aims to make it easy to fill your refrigerator and pantry with your personal favorites and staples when you need them. After selecting products through the Instacart app, personal shoppers review your order and do the in-store shopping and delivery for them. Instacart uses customers' transactional data to develop models which recommend products that users will buy again based on their previous purchases.

Objective:

Predict which previously purchased products will be in a user's next order. By providing an optimal and accurate recommendations of products to purchase, Instacart can enhance customers overall shopping experience, browsing experience, increase revenue from sales, and increase overall customer satisfaction.

Data:

- The dataset is obtained from kaggle, https://www.kaggle.com/c/instacart-marketbasket- analysis/data.
- The dataset is a relational set of files describing Instacart customers' orders over time.
- The dataset is anonymized and contains a sample of over 3 million grocery orders from more than 200,000 Instacart users.
- For each user, Instacart provides between 4 and 100 of their orders, with the sequence of products purchased in each order. Instacart also provide the week and hour of day the order was placed, and a relative measure of time between orders.

Methodology:

- Understand the problem and objective
- Data Preparation/ Data Wrangling using tool such as pandas
    - Find the relationships between tables and Join them appropriately
    - Clean the data : missing values, outliers, bad data
- Perform EDA (exploratory data analysis) to better familiarize ourselves with the data
- Find the specific patterns, outliers, Key performance indexes
- Come up with hypothesis and test them
- Use machine learning techniques to build a robust predictive models.

Deliverable:

- Recommendation of products based on the prediction model
- Codes in Python on Jupyter Notebook
- Project report explaining the problems, steps, and results