

## Proposal: Instacart Market Basket Analysis

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 you.

Instacart uses customers' transactional data to develop models which recommend products that users will buy again based on their previous purchases. The goal of the project is to 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.

The data can be obtained from kaggle, <https://www.kaggle.com/c/instacart-market-basket-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.

The approach of this project will follow the standard data science project life cycle. First, start with understanding the problem. Then, data preparation that includes data wrangling follows. For this problem, one of the data preparation step will be to join the corresponding tables together since the data consists of relational files. After that, Exploratory Data Analysis (EDA) will follow to get more understanding of the data and to identify patterns, outliers, etc. Then, based on the EDA, fit the prepared data to one or several predictive machine learning models to perform the prediction. After that, optimize the model for a better accuracy. Review and repeat these steps if necessary.

The deliverable of the project will be a code for the predictive model written in Python Jupyter notebook and a project report that explains the problem, the goal, the dataset, the predictive model, the steps taken and decisions that are made to create the predictive model, and the results.