

## **CAPSTONE PROJECT PROPOSAL**

### **Instacart Market Basket Analysis: Predicting Existing Customer's Next Order List**

By: Pamela Ubaldo Augustine

#### **Introduction**

Ordering groceries online provides the convenience of quality products delivered to your door steps without spending hours in the store and ability to track previous orders. US online grocery sales amounted to 7 billion USD in 2015 and projected to go up to 18 billion USD in 2020. In 2015, US consumers visit an average of 2.2 websites shopping groceries online and a third of these consumers place grocery orders once a month according to [Statista](#). Availability of features that removes friction in completing on an online purchase such as free shipping, one click ordering, product recommendations can definitely increase sales. Thus, development of new website features that make checking out easier is lucrative.

#### **Problem Statement**

Instacart, #1 in Forbes most promising company list in 2015, is conducting an Instacart Market Basket Analysis Kaggle competition to predict products that existing customers will purchase again. This capstone project aims to predict Instacart customer's next order list. This is a great web store feature than can make Instacart's check out process smoother by providing more convenient shopping experience for busy customers, saving them time and energy for grocery planning and store trip.

#### **Data Source**

The data used in this project is acquired from Instacart Market Basket Analysis Kaggle competition in Kaggle.com (<https://www.kaggle.com/c/instacart-market-basket-analysis/data>) and also available in Instacart website (<https://www.instacart.com/datasets/grocery-shopping-2017>).

#### **Methodology**

Analysis will be done using Python, Pandas, Matplotlib, Scikit-Learn packages in IPython Notebook 3.7. Exploratory Data Analysis will be used to get a generalized prediction model. The following will be analyzed about the data

- 1) Most to least ordered products and from which department and aisle
- 2) Ranking of Time and day of the week when most to least order happen

- 3) For each customers the following will be asked a) number of items in customers cart, b) from department and aisle, c) frequency of order, frequency of order for each product, d) time of the week or month customer is most likely to order

### **Deliverables**

Employing machine learning, this project aims to predict the a) number range of items in customer's cart, from which department and aisle b) products in customer's next order, and c) range of time that customer is most likely to order, which are features that can make Instacart check out process smoother.

A full written report of the data analysis will be provided including the code in IPython notebook used. Slide deck will be included explaining key steps and summary of the project.