

## **Capstone project 1 proposal: Instacart Market Basket Analysis**

### **Background:**

Instacart is a grocery ordering and delivery app. After selecting grocery products or pantry products from the app, a personal shopper would review the order and do the in-store shopping as well as deliver the groceries for you.

### **Problem:**

Figure out which products will be in a user's next order based on the customers' historical order.

### **Data:**

Data will be collected from Kaggle: <https://www.kaggle.com/c/instacart-market-basket-analysis/data>. The dataset is an anonymized dataset with sample of over 3 million grocery orders from more than 200,000 Instacart users. For each user, there are between 4 to 100 of their orders, with sequence of products purchased in each order. The week and hour of day the order was placed is also given along with the time between the orders.

### **Approach:**

To tackle this problem, there are several data science skills that will be used. First, would be to do some data analysis such as graphs on the data, and figure out how to fill out the missing values, as well as other data cleaning techniques. Then Exploratory data analysis (EDA) is important such as plotting histogram, etc to understand the data further. Then the cleaned and understood data would be used to create predictive models, which might be logistic regression to predict 1 or 0, where 1 would be re-order that specific item and 0 otherwise.

### **Deliverables:**

The deliverable for this project would be the line of codes written down as well as a PDF describing what methods were used in exploring the data, and would also include graphs and charts to show before and after data cleaning and the model as well.