

## **Project Proposal -Instacart Sales -Group 35**

### **Introduction:**

Instacart is a grocery ordering and delivery app, personal shoppers reviewing customer orders and completing in-store shopping and deliveries. Groceries taking large proportion in daily spending and with interest in e-commerce, we choose this leading online grocery delivery company as our research target, aiming to dig out customer behavior patterns and identifying future behaviors and potential purchases.

### **Dataset Description:**

This dataset contains 6 relational tables, containing information including departments, transaction, orders, products and aisles. Order, product, department and aisle datasets give description information about the order, product, department and aisle. The order\_product dataset describes the relation between orders and products.

Specifically, some of the most important features are:

Order: give the order id, user id, order number, order time, order time interval, eval set, order dow.

Department: show a list of different departments( such as frozen, bakery, household.etc)

Product: provide the list of products with product id, product name, department id, aisle id.

Aisle: show the aisle id and name.

Order\_products: show the order id, product id, add to cart order, whether it's reordered.

### **Dataset Acquisition:**

Imported from Kaggle API

-Gained API key from Kaggle

-Python Code:

Pip install kaggle

kaggle competitions download -c instacart-market-basket-analysis

### **Processing Tasks:**

1. Merging multiple datasets: for example, combine the department, product and order-product datasets to study orders made in different departments.

2. Indexing, selection, and filtering: for example, select the orders which belong to "not-reorder" category to find shopping pattern of new customers.

3. Reshaping and reindexing: redefine the department: put all the departments about food( like bakery, meat, snacks) together as food department.

4. Manipulation: create a new dataset shows the frequencies that different kinds of products are put in cart and the how many times they occur in certain order sequences.

5. Cleaning: formatting, text processing, missing values

### **Expected Findings:**

Identify recursive purchase patterns, product purchase relationships which would be the foundation of recommender systems.

Identify the products which are likely to be ordered first.

Identify order pattern in a day( from morning to night), order pattern for regular and old customers.