# Instacart Basket Analysis

## Introduction:

Instacart is a well known and widely used online grocery shopping service looking to uncover more information on their sales patterns. This project from the CareerFoundry coursework was geared to introduce python to data analyst students.

## Goal:

Assist Instacart stakeholders to understand the variety of customers in their database and their purchasing behaviors in order to develop targeted marketing strategies. The goal for this analysis is to ensure that Instacart targets the right customer profile with the appropriate products. They have specifically requested data regarding busiest times of day/days of the week, price range groupings on their products in order to simplify them, and departments/products that are most popular (most products sold).

## Steps and Skills:

- Python/Anaconda
- Data Cleaning and data wrangling
- Deriving variables, merging
- Grouping datasets
- Aggregating data
- Population flows
- Visualizations in Python matplotlib, Seaborn, Scipy (line chart, bar chart, histogram, pie chart, scatterplot, stacked bar chart)
- Interpret Results and Summarize findings/insights.
- Reporting in Excel

instacart

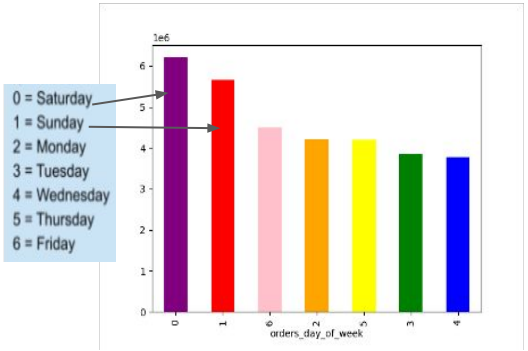
"The Instacart Online Grocery Shopping Dataset 2017", Accessed from www.instacart.com/datasets/grocery-shopping-2017 via Kaggle on [4/20/24].

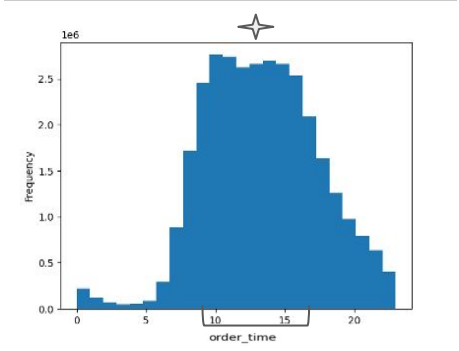
Note: the customer data was fabricated for the learning purposes of this course.

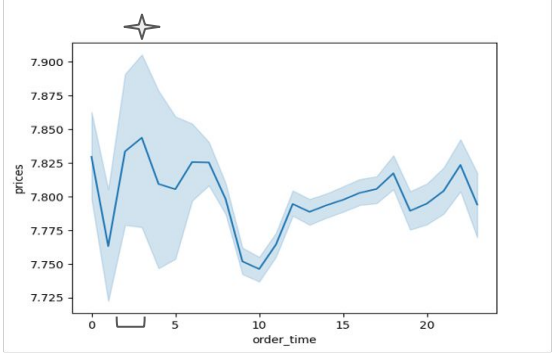
## Tools Used:

# Instacart Basket Analysis

The analysis shows that **Saturday and Sunday** are the busiest days of the week.
Peak hours are between **9am-5pm** with **10am,** the busiest hour of the day. People tend to spend most around **2-3am.**

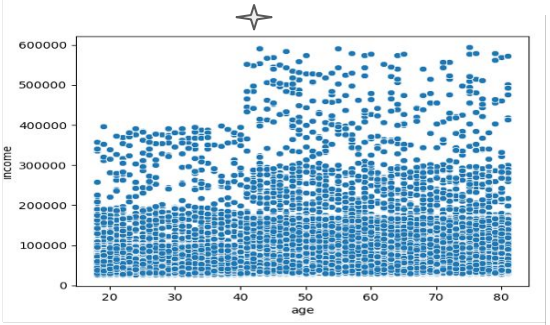


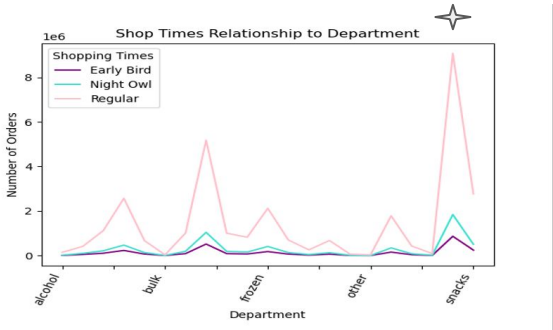




There is a correlation here between age and spending power represented in this scatterplot as there is a **definitive jump in income at age 40**, there is a huge jump in income from a max of $400,000 to $600,000.

The line chart shows that **snack items are the most frequently purchased** for all shoppers, no matter the time. Recommend snack options to always be included in advertisements at **all hours of the day.**



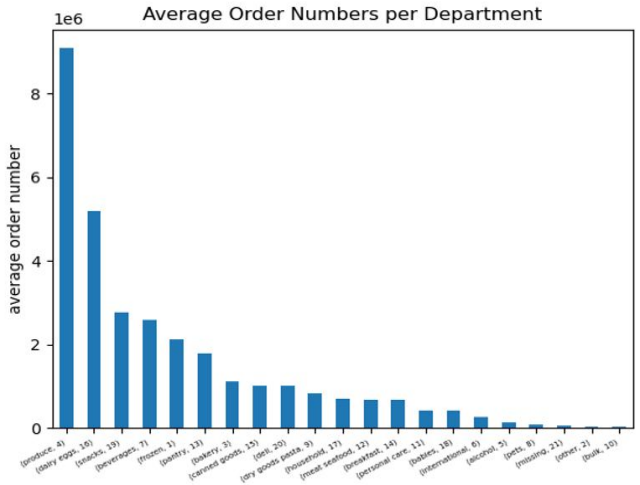



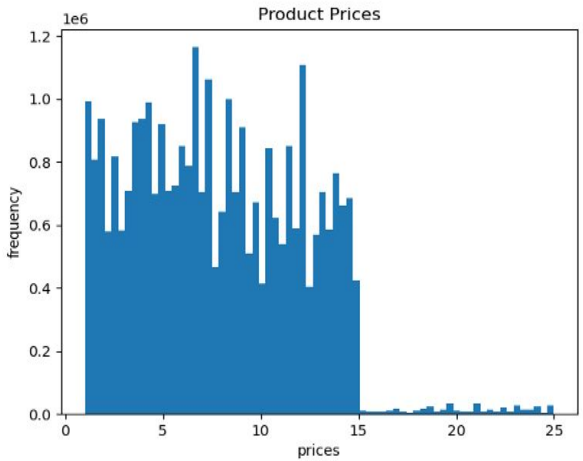
# Instacart Basket Analysis

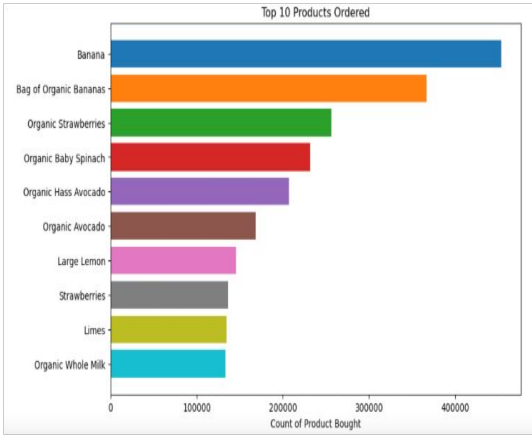
-**Produce, dairy/eggs, snacks**, **beverages**, **frozen**, and **pantry** departments are the **most popular departments** (with least popular departments being international, alcohol, and pets).

- **Most products cost** between **$1-$15**, while a few are priced higher in **$15-$25** range. This information will help to figure out appropriate and simpler price range groupings, as majority of items or lower priced. Aim for $5 increments for simplify price groupings

-**The top 10 products:** ranked with regular bananas at the top, followed by organic bananas, organic strawberries, organic baby spinach, organic avocados, lemon, regular strawberries, limes, and organic whole milk (interesting to note the popularity of **organic** products).

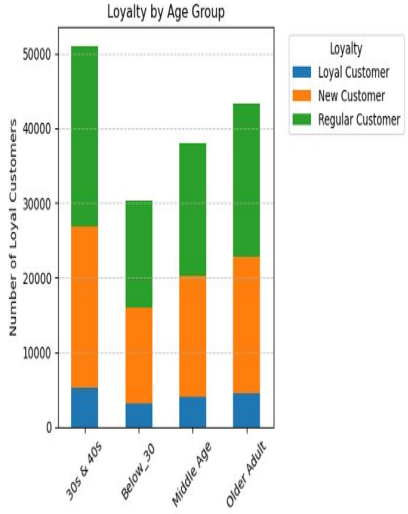


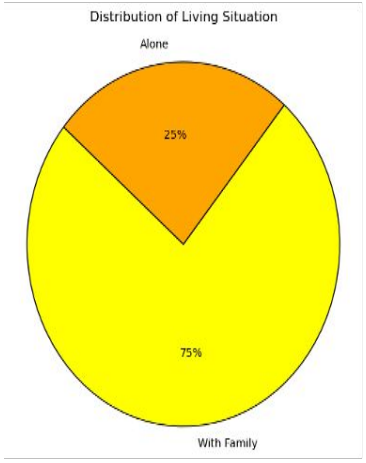



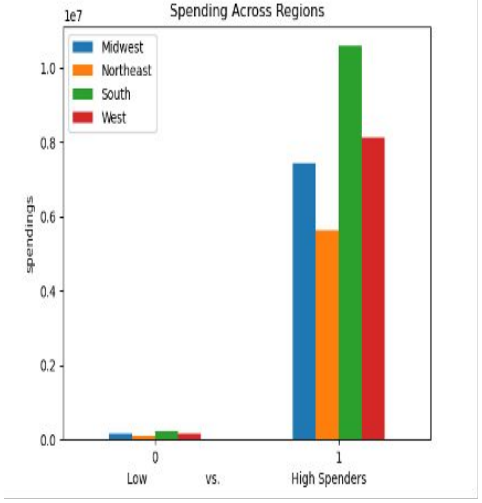



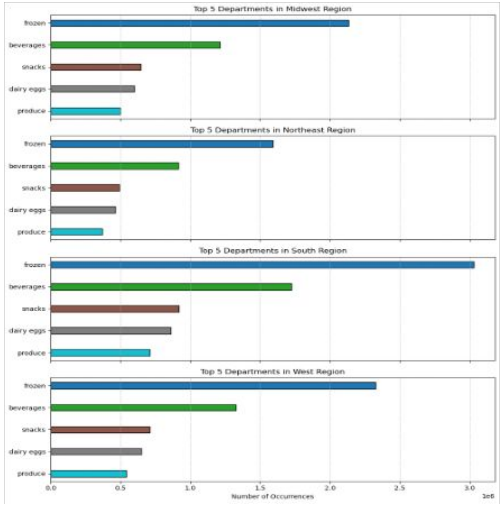
# Instacart Basket Analysis

- The **30s/40s** age group tends to rank the highest in all aspects of customers, according to the first bar chart.
  - They make up: 1) the most **regular customers**, 2) majority of **new customers**, & 3) majority of **loyal customers**.
- Most customers **have families** ( pie chart).
- The next bar chart shows that **Southern Regions** generate the most of Instacart's spendings. This is followed by Western Regions, then Midwest Regions, with Northeast spending the least. This pattern is the same for both groups of people (high and low spenders)
- These final horizontal bar charts are broken up by regions, showing the **top 5 departments for that region.** All of them show a similar distribution of department preference in the following order: **1) <u>Frozen</u>, 2) Beverages, 3) Snacks, 4) Dairy Eggs, 5) Produce**



# Instacart Basket Analysis

## Recommendations:

- Increase Ads during mid week (Tuesdays & Wednesdays). 
- Schedule Ads between 12pm and 3pm, the busiest hours for placing orders. 
- Schedule Ads for higher priced items in the early morning hours between 2-3am. $
- Advertise snacks during all hours 
- Place eye-catching poster Ads within the popular departments (Produce, dairy/eggs. Snacks, beverages, frozen, and pantry )
- Use promotional Ads in popular sections for popular items to increase amount purchased, as well as Ads for lesser populated products to direct viewing towards less popular sections (international, pet, & alcohol)


- Keep $5 price range groupings as majority of products are between $5-$15.
- Promotional Ads for organic items (placed in highly populated banana section) 
- Keep organic foods well stocked and in good shape to maintain the integrity of these sales.
- Recommend targeting 30s-40s, young family groups in advertisements - busy lifestyles (prepared food, quick snacks, food on-the-go) 
- Recommend increasing advertisements in southern regions to maintain integrity of their higher sales, as well as promotion of Instacart to the other regions to gain more customers (keeping in mind these variables and target populations in the types of Ads used. 

# Instacart Basket Analysis

## Challenges:

- Creating flagged columns or column derivations did not always go as planned.
    - **Solution**: Since I carefully kept track of each step of the wrangling process, I could trace my footsteps and was able to restart the steps prior to my error. This taught me to record each manipulation of original datasets, no matter how small or insignificant it may seem. As a result, I have learned to be diligent with organization and record keeping throughout the entire wrangling process.
- I experienced difficulty in creating certain visualizations with python using this large dataset, particularly pie charts.
    - **Solution:** I transferred categorized data into a csv file and then upload into Tableau to create a pie chart there instead.

★ This project was extremely time-consuming for me, taking much longer than the CareerFoundry time frame expectation for task completion.
    - Because of the amount of time invested, I found that this was a great learning toolfor me to fully understand python and jupyter notebook. I grew a great appreciation for python with hands on work throughout the tasks involved. As my knowledge grew, I found it fun to dissect and manipulate this large dataset to find the key insights that I was targeting. I spent time deep diving into this coursework since I knew that this process was going to be essential for my future work as a data analyst.