Instacart Customer Analysis

by Lisa Greer

Overview:

- Instacart already has very good sales, but they want to uncover more information about their sales patterns. The task is to perform an initial data and exploratory analysis of some of their data in order to derive insights and suggest strategies for better segmentation based on the provided criteria.

Purpose:

- The Instacart stakeholders are most interested in the variety of customers in their database along with their purchasing behaviors. They assume all customers can't be targeted using the same methods, and they're considering a targeted marketing strategy. They want to target different customers with applicable marketing campaigns to see whether they have an effect on the sale of their products. This analysis will inform what this strategy might look like to ensure Instacart targets the right customer profiles with the appropriate products.

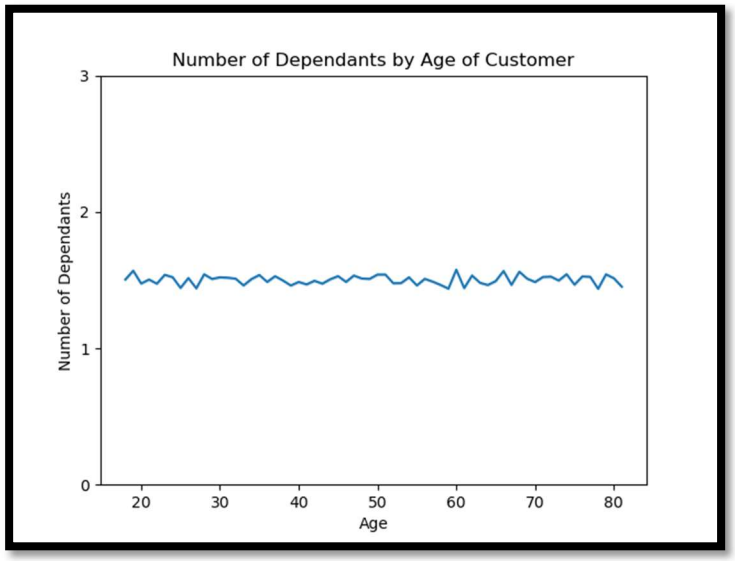
Objective:

- The stakeholders would like to be able to answer the following key questions:
    - What are the busiest days of the week and hours of the day?
    - Are there particular times of the day when people spend the most money?
    - What would be effective price range groupings?
    - Are certain departments more popular than others?
    - Do demographics such as region or number of dependents affect sales?
    - How does customer loyalty affect sales?
    - Are there any other insights to be gleaned from the data?

Tools, Skills, Methodologies:

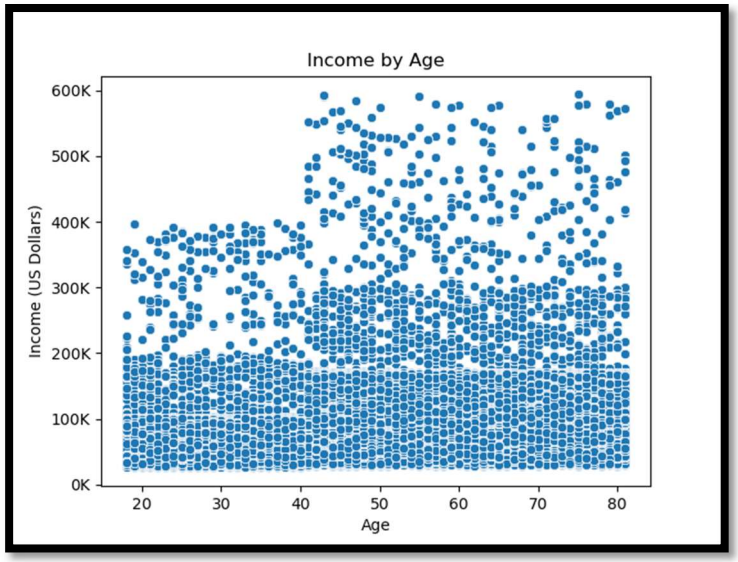
- A number of open-source data sets from Instacart and from CareerFoundry were used to complete this analysis.
- A Data Dictionary can be found here: [The Instacart Online Grocery Shopping Dataset 2017 Data Descriptions · GitHub]
- Python was used to clean, manage, and wrangle the data so that it could be analyzed using statical techniques and visualized using Python libraries.

The Findings:

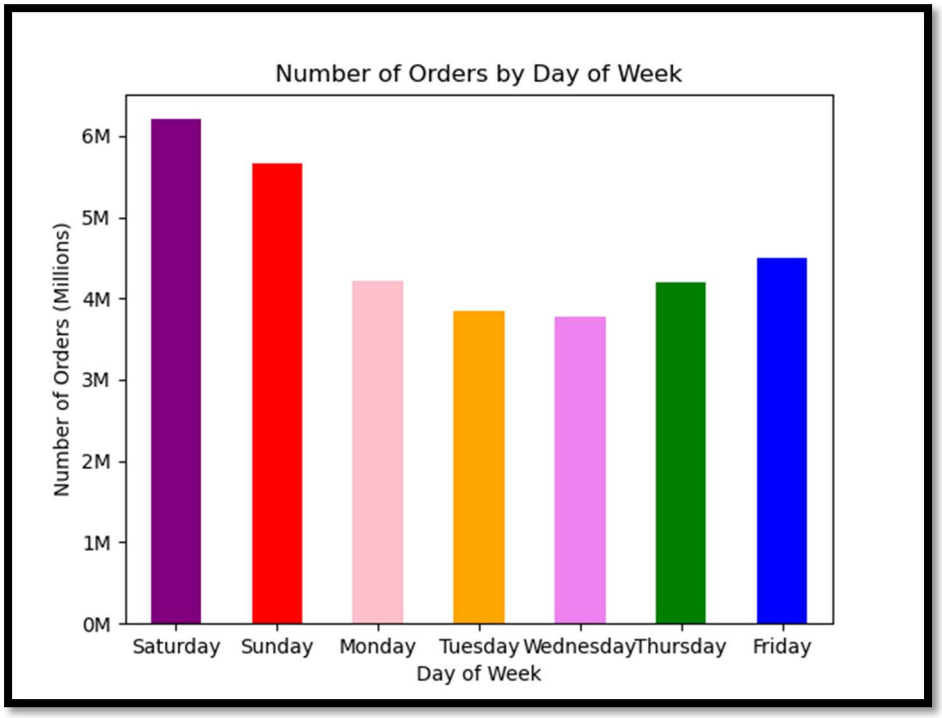


This chart indicates that there is little correlation between age and number of dependents.

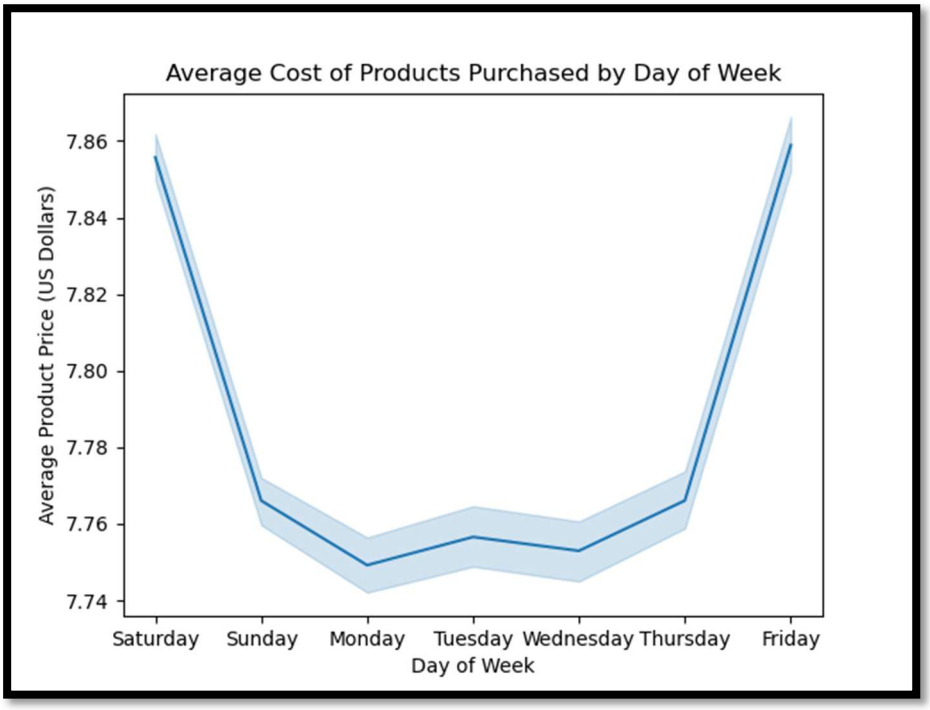
The average of ~1.5 dependents is regardless of age.



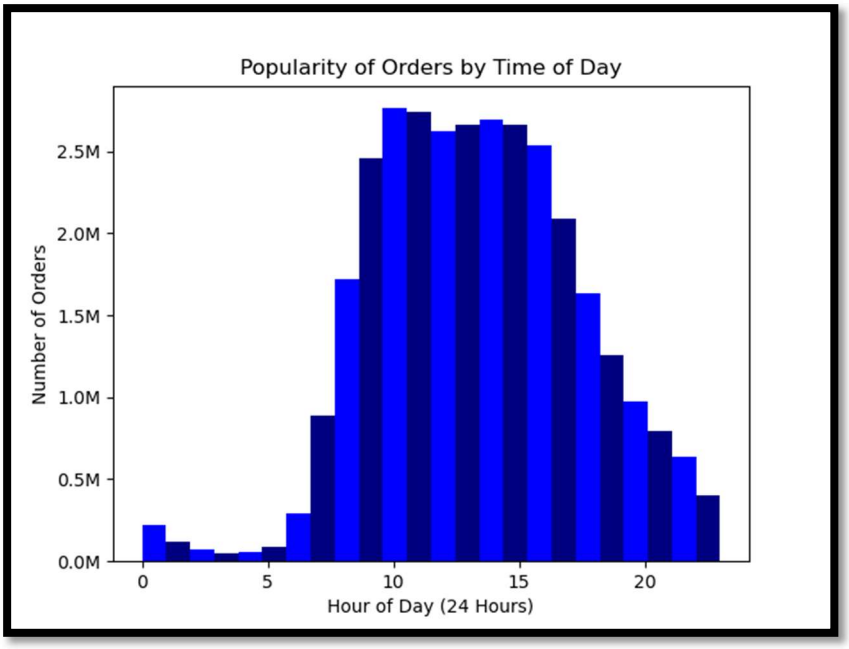
This chart indicates a strong correlation between age and income growth. Around 40 years of age there is a significant increase in income amongst the Instacart customers.



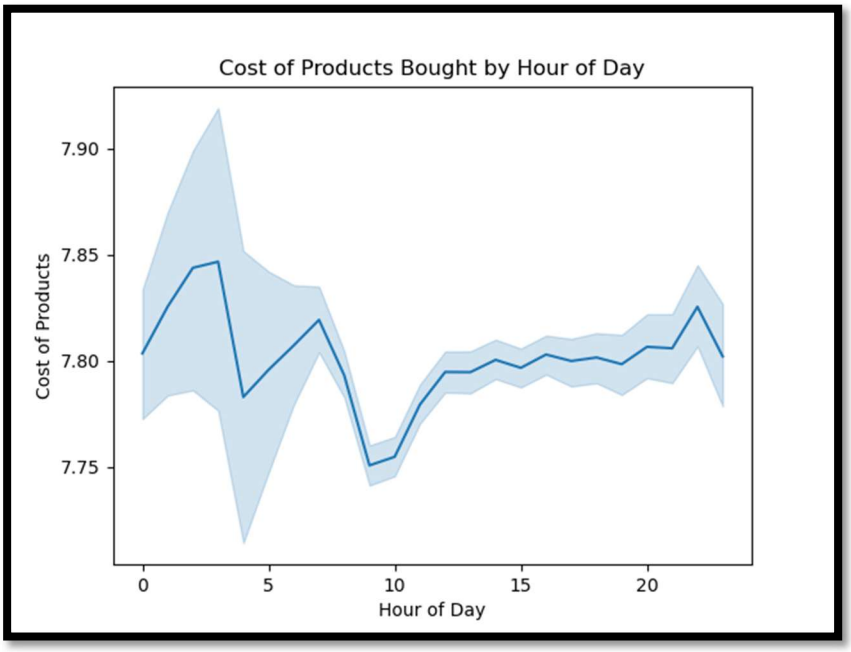
This chart indicates which days of the week are most popular for orders.



This chart evaluates if more expensive products are ordered on certain days. The ~12 cent difference from the low end of the chart to the high end would indicate a negligible difference in cost by day of the week.

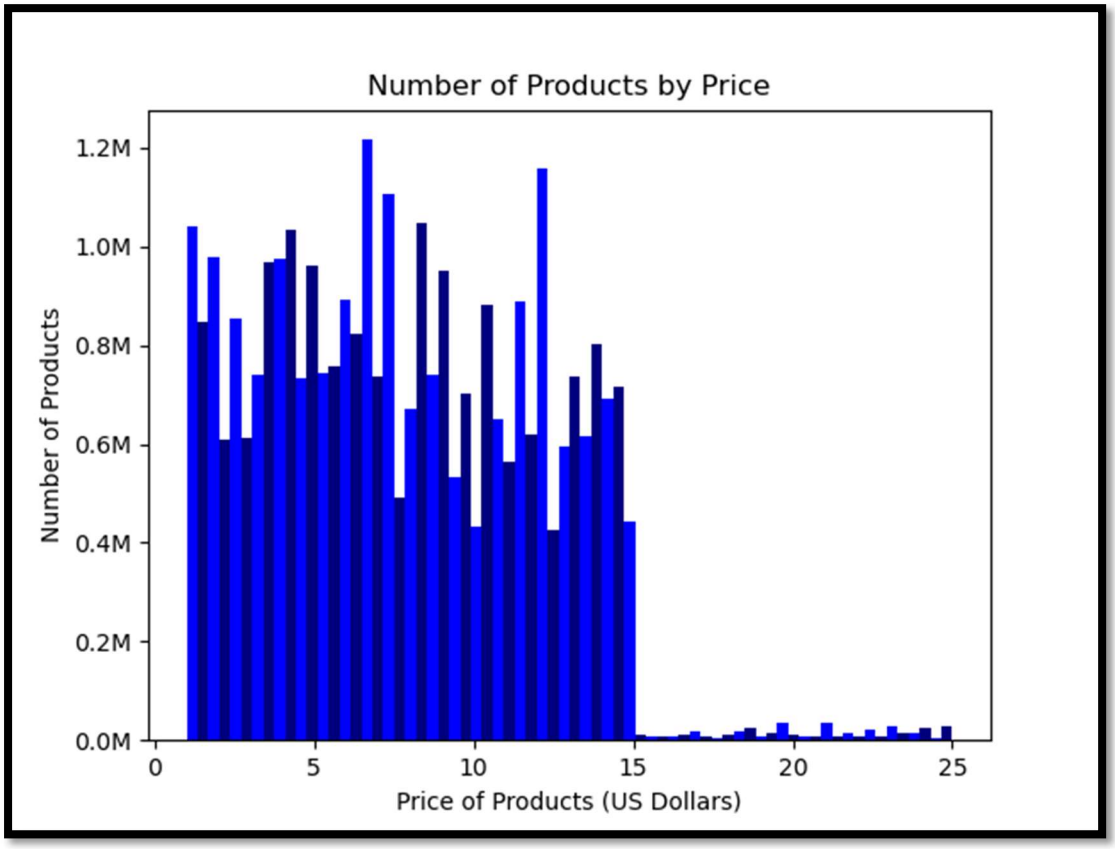


This chart indicates that there is a significant busy period for Instacart orders. 9am to 4pm would be considered busy times.

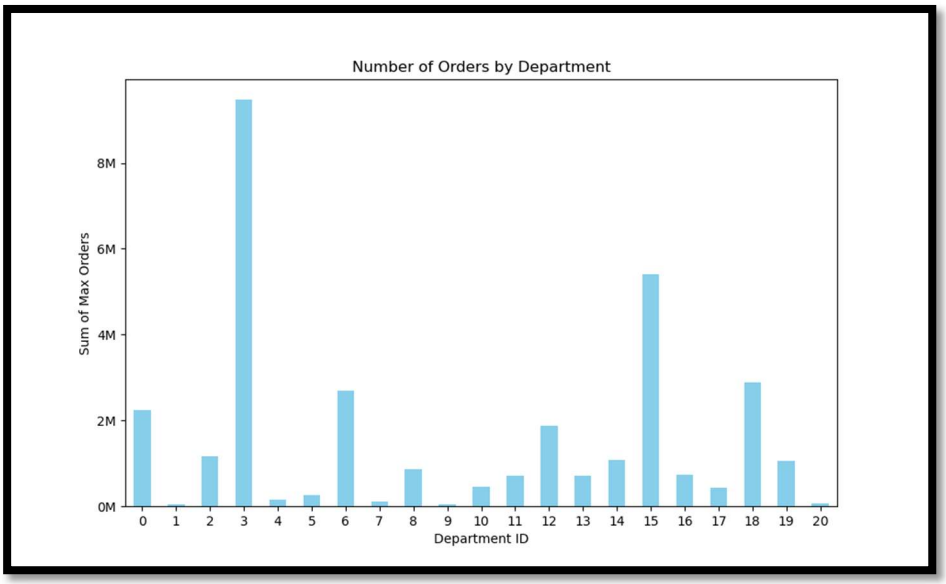


This chart evaluates if more expensive products are ordered during certain times of the day.

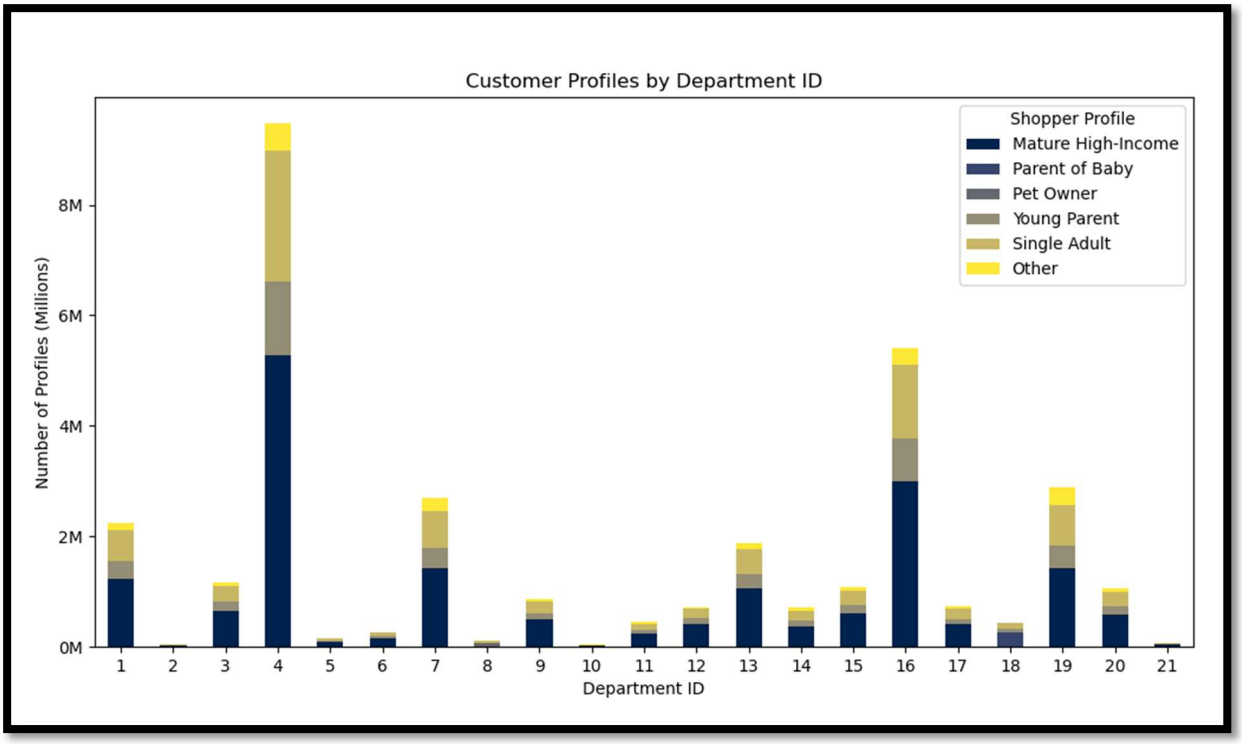
Like the chart above the ~.15 cent difference from the low end of the chart to the high end would indicate a negligible difference in cost by hour of the day.



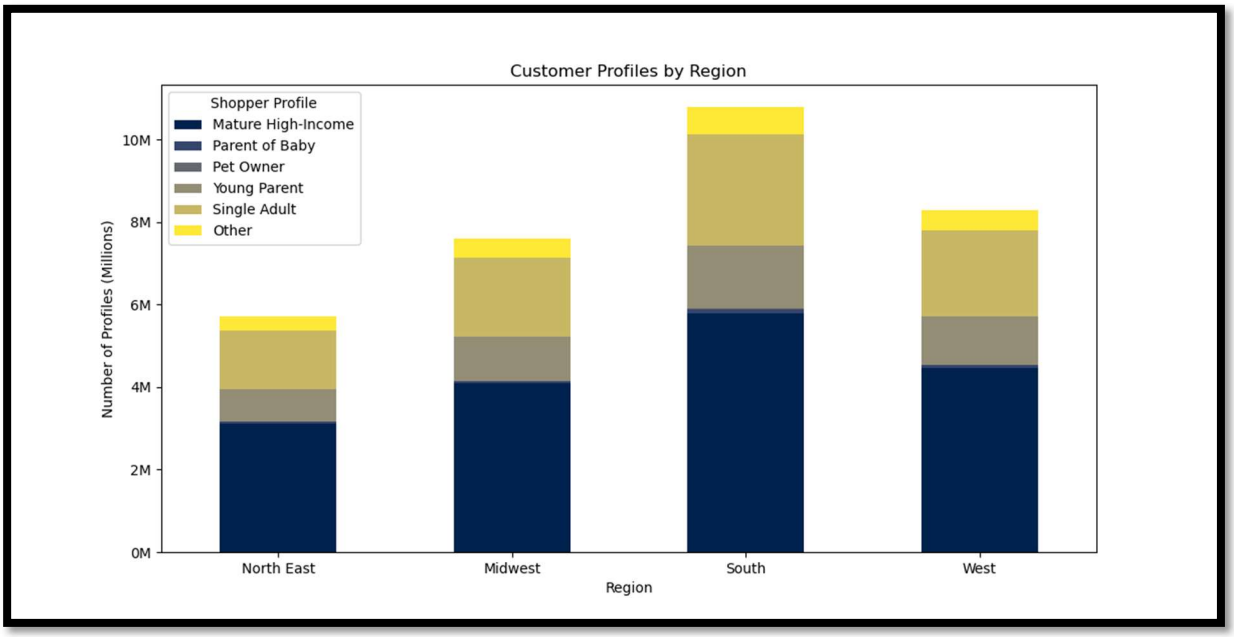
This chart indicates the distribution of cost of products offered by Instacart. Most items are ~15 dollars or less. The chart below indicates that the bakery, canned goods, and babies departments are the most popular.



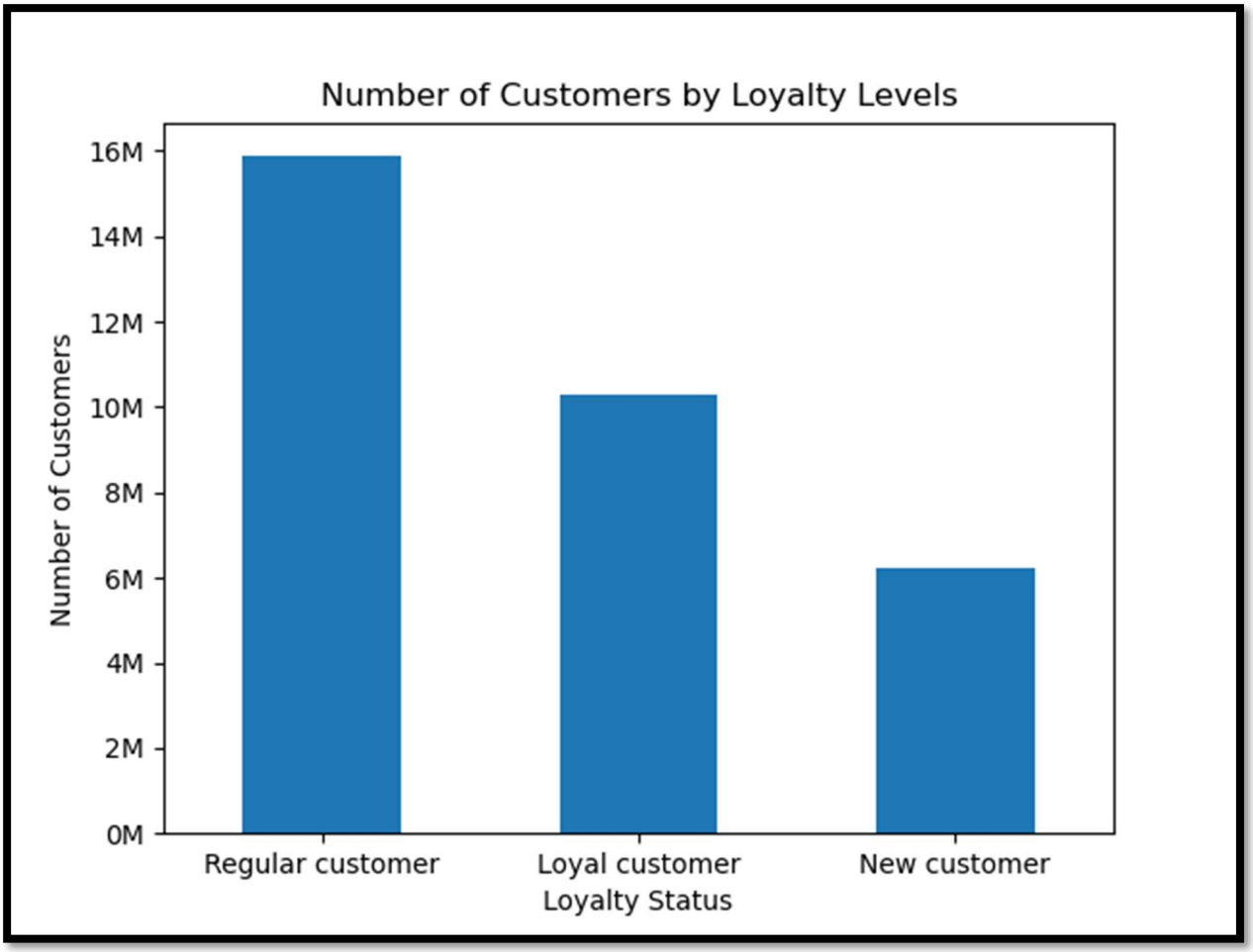
department_id	department
1	frozen
2	other
3	bakery
4	produce
5	alcohol
6	international
7	beverages
8	pets
9	dry goods pasta
10	bulk
11	personal care
12	meat seafood
13	pantry
14	breakfast
15	canned goods
16	dairy eggs
17	household
18	babies
19	snacks
20	deli
21	missing



This chart indicates the types of shoppers by department and the most popular departments. The most prolific shoppers are those in the “Mature High-Income” category.



This chart indicates shopper type by region. It indicates shopper type is fairly consistent regardless of region.



Lastly is a look at customer loyalty. There are far more regular customers than those who are considered loyal customers, but as will be explained in the conclusion loyal customers have a significantly higher number of total orders.

Conclusion:

- Busiest days of the week are Saturday, Sunday, and Friday.
- Busiest times of the day are from 9am to 4pm.
    - Promotions could be added to off days/hours to increase traffic on low days and during slow times.
- The amount of money spent by hour of day is negligibly different hour to hour.
- The majority of Instacart products are less than $15. It is a broad margin in quantity above and below $15.
    - Effective product grouping would be greater than $15 and less than $15.

- Certain departments are much more popular than others. By comparing total orders to department id it is clear the Bakery, Canned Goods, and Babies departments are three most popular.
    - Promotions or coupons could be sent to consumers to increase sales in lower selling departments/adjacencies.
- It would appear that, by ratio, region and number of dependents do not affect sales.
- It would appear most customers would be considered "Regular Customers", meaning most have ordered between 10-40 times. About 33% less is "Loyal Customers" and even fewer "Newer Customers".
    - This means targeting regular customers to move them into the loyal category might be the way to go.
- 'Pet owners', 'young low-income', and 'married with no kids' make up little to no percentage of Instacart shoppers.
    - It would be best not to target these customers or to treat them as a potential growth area.

Limitations:

- While the data was relatively clean it is a little outdated at this point (2017). Working with a more current dataset would provide insight into current, and importantly post-pandemic, trends.

Next Steps:

- Narrowing the analysis to more specific buyer profiles would be helpful.
- Cross-referencing this data with marketing data would also be useful to see if targeted advertising is proving to be effective.

Follow – Up:

- For more details about the project and data here is the link to the GitHub repository:
    - GitHub Link