Project 4:

Instacart Analysis

PROJECT 4: MARKETING STRATEGY FOR AN ONLINE GROCERY STORE

Overview: Instacart stakeholders are interested in the variety of customer in their database along with their purchasing behaviors. The objective of this task is to perform an initial data and exploratory analysis of Instacart data in order to derive insights and suggest strategies for better targeted marketing. I decided to focus on three employment types: student, worker, and retired.

Data

- Instacart Online Grocery Shopping Dataset (2017).
- Customer dataset provided by CareerFoundry.

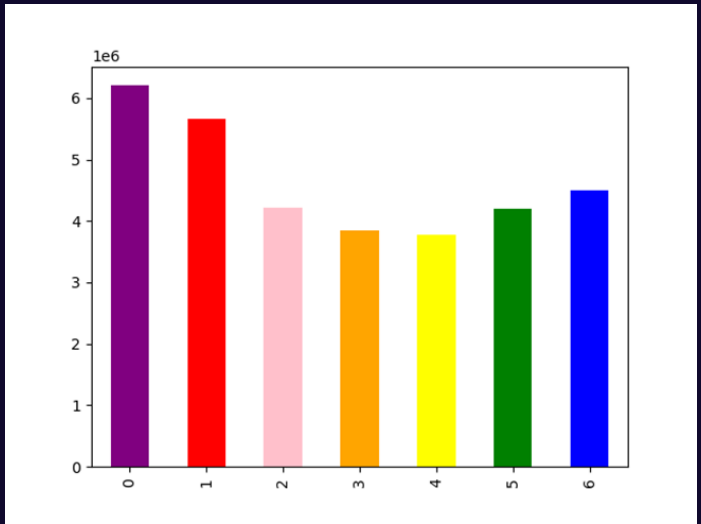
Skills

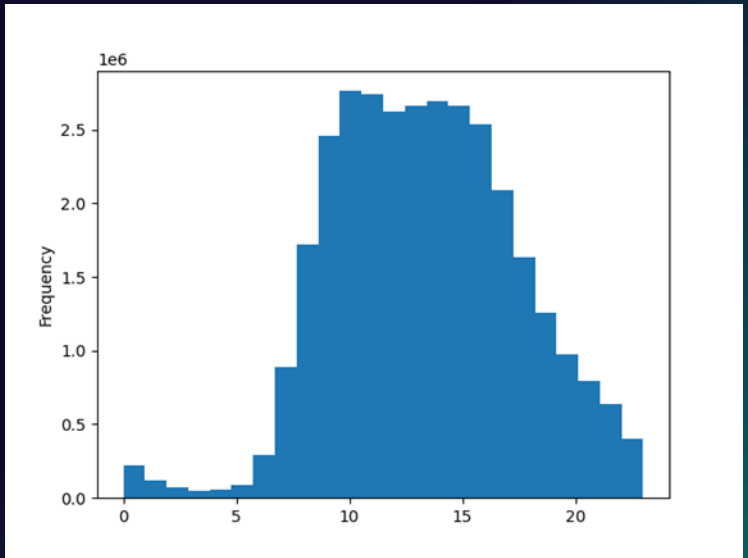
- Python
- Data wrangling
- Data merging
- Deriving variables
- Grouping and aggregating data
- Population flows

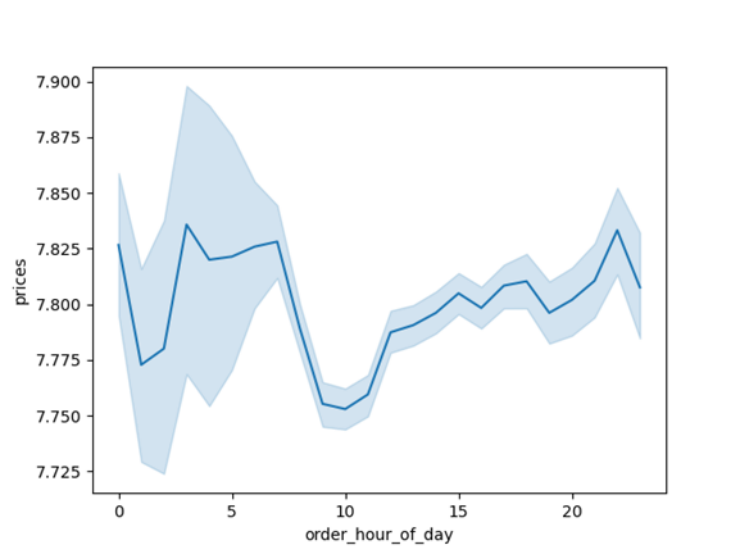
Tools

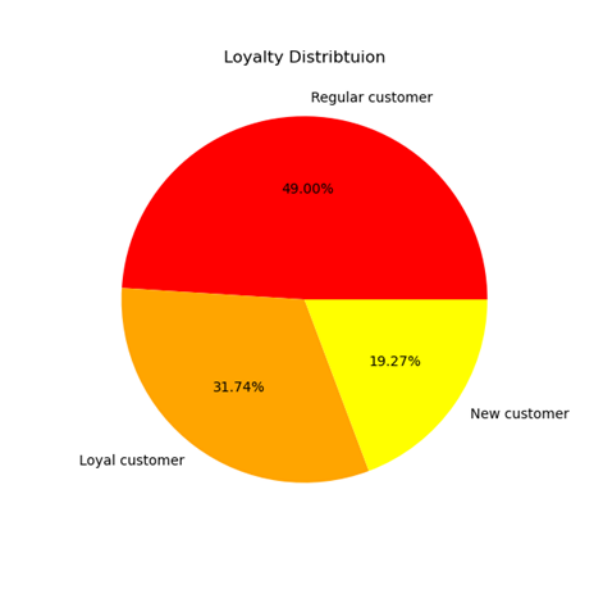
- Jupyter Notebook (Anaconda 3)
- Libraries: pandas, NumPy, seaborn, Matplotlib, SciPy
- Microsoft Excel

Project 4: Exploratory Analysis









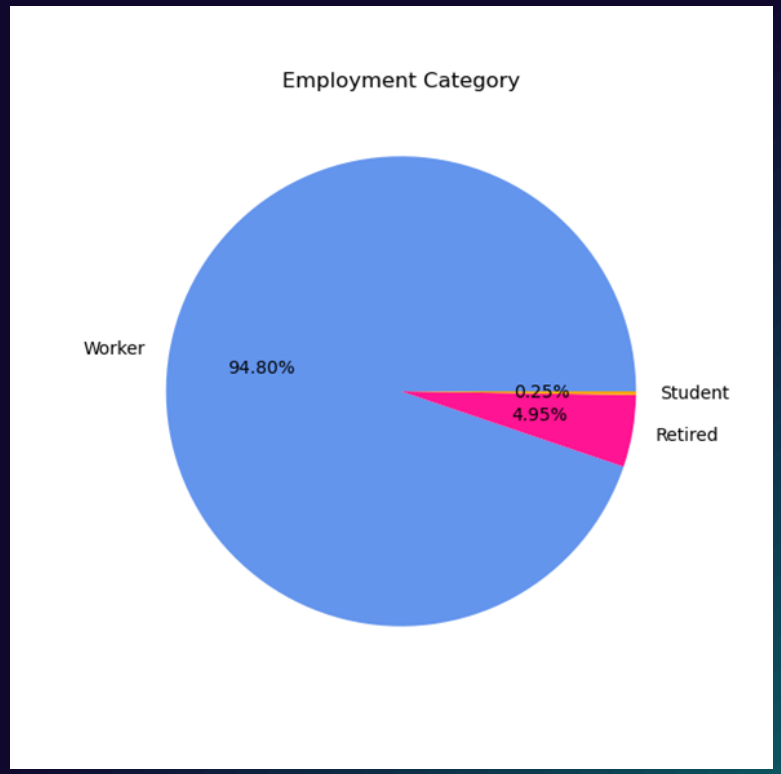
The sales team wanted to know what the busiest days of the week and hours of the day are in order to schedule ads when there are fewer orders. They also wanted to know whether there were particular times of the day when people spend the most money and what was the distribution among customers regarding loyalty.

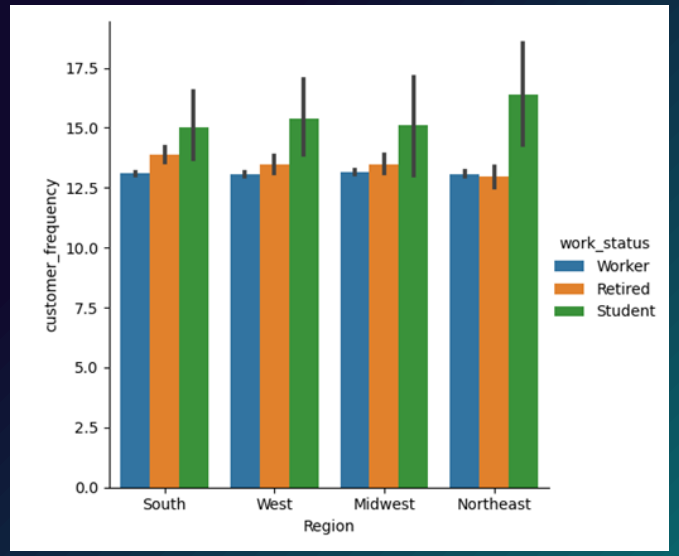
Data was cleaned, checked for consistency and wrangled. The datasets were then merged in order to group data and derive variables.

- Saturday is the busiest day of the week, and the busiest hours were 8-10 am.
- Customer are more likely to buy more expensive hours at odd hours of the morning (impulse buys?), and then lower prices during regular morning hours.
- Nearly half the customer base are regular customers

Project 4: Customer profiling Analysis

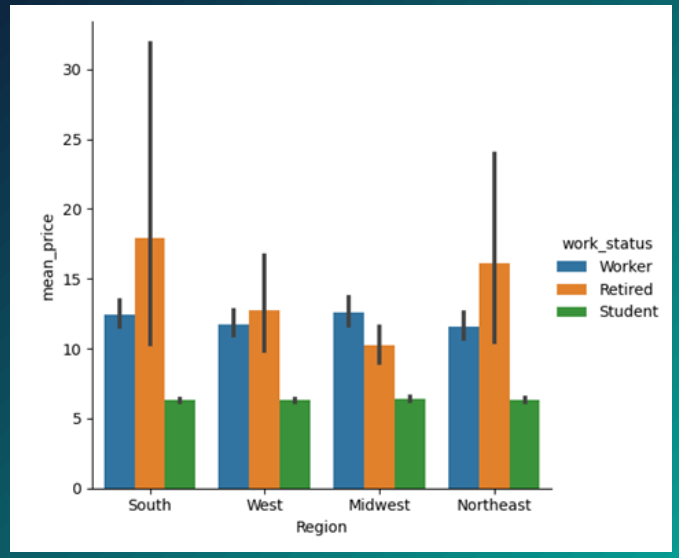
To derive a new column, I selected for number of dependents, age, income, and whether the individual shopped in the baby department. This formed three groups: retired, student, and workers.





Those likely to be college students are unlikely to be customers, making up only 0.25% of customers. However, they are also the most frequently returning customers.

Those likely retired make up only about 5% of customers, but they buy the most expensive products.



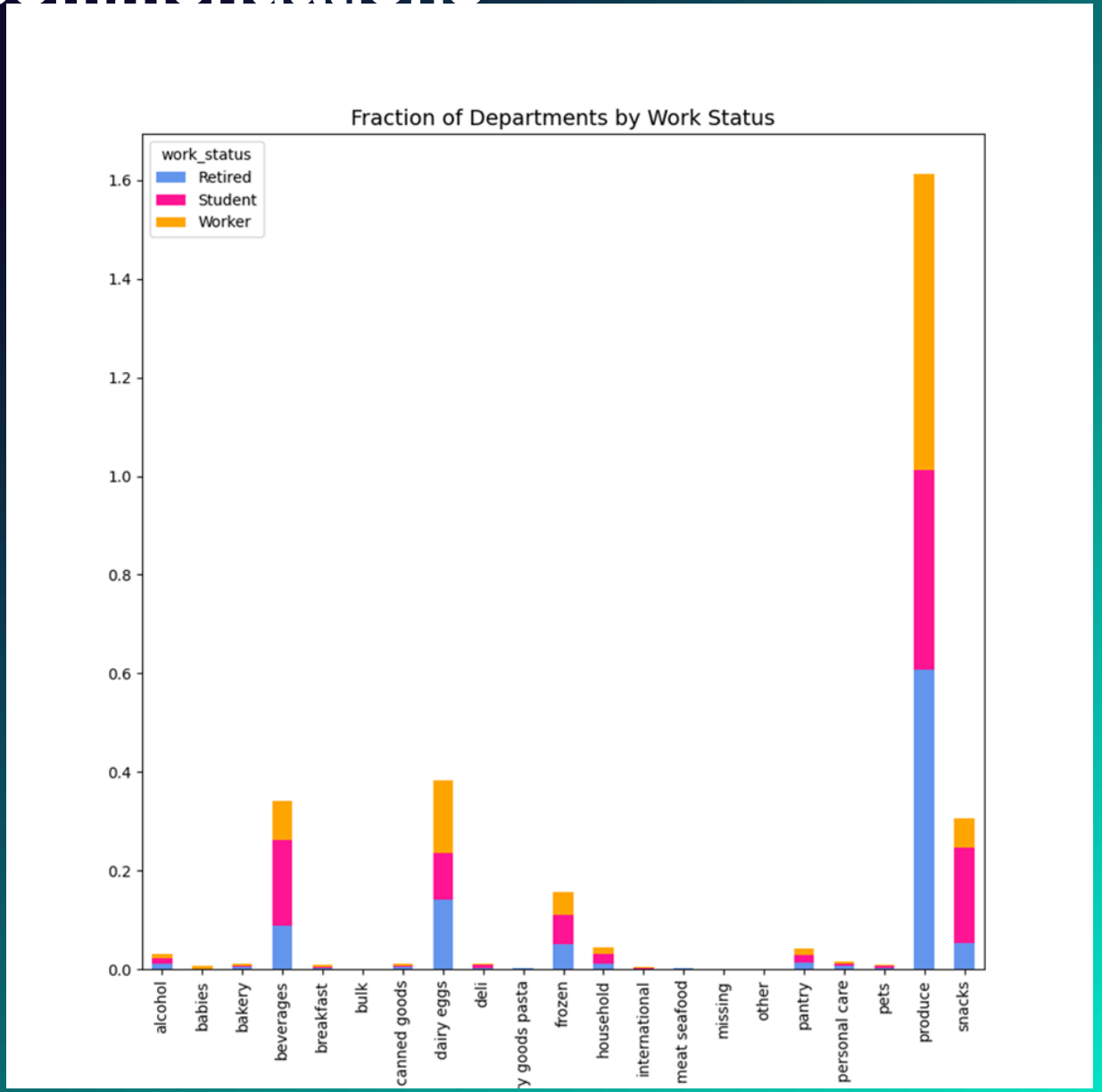
Students buy the least expensive products and there is little variation of this between regions.

Retirees in the southern region buy more expensive products than retirees in other regions, but there is a lot of variation in the spending habits of retirees.

Project 4: Additonal Insights & Recommendations

- As Wednesday is the least busy day, it would be a good day to schedule ads. However, the least busy hours are between 2-5am and scheduling ads during this time is unlikely to help as most people are sleeping. Ads should be schedules after 3pm, when orders begin to decline, but people are still awake.
- Most customers, regardless of any demographic primarily purchase produce, dairy/eggs, beverages, snacks and frozen foods.
- Students make up the majority of snack and beverage purchases so marketing to students should focus on those categories.
- Offer senior citizen and student discounts to increase the presence and frequency of purchases.

Python scripts and deliverables on Github



PROJECT 4 TAKEAWAYS

This project introduced programming for data analysts. Due to my background in with ArcMAP, I was already familiar with some Python, albeit not Jupyter notebook. I did find certain aspects of this project would have been easier in Excel. However, deriving new variables, grouping data, aggregating variables, and data visualization with Python were new key skills learned during this project.

One task during this project that was challenging was creating a new variable based on at least three others for targeting marketing. The variables I chose ended up identifying outlier groups (students and retirees), however, these two groups did still show patterns in purchasing habits and purchasing power, so I was still able to use them for this project.

Overall, I found this project fun due to the amount of creativity it allowed for when trying to create a new variable.