Instacart Basket Analysis

Haley Bustle

November 2023

Project Tools and Resources

- Project Duration: Four weeks
- Project Resources
  - The datasets were provided by Career Foundry and were altered for educational purposes only. There were no citations provided on when, where, or how the data was collected
- Tools Utilized
  - Microsoft Office: Word for project brief, data summary, and final project report; Excel used for data changes, statistical analysis, final report
  - Python and Jupyter for data analysis, wrangling, merging, and creation of visualizations

Project References and Data Sources

- Project Resources, please note these are very large dataset that take up a large amount of space:
  - Data Dictionary
  - Customer Dataset
  - Orders & Products Dataset
  - Orders & Products Prior Dataset
  - Departments Dataset
- Final Report on GitHub
- Final Excel Report
  - This report contains how the datasets were cleaned, analyzed, and merged to the final dataset. It also contains visualizations created with Python and final recommendations.

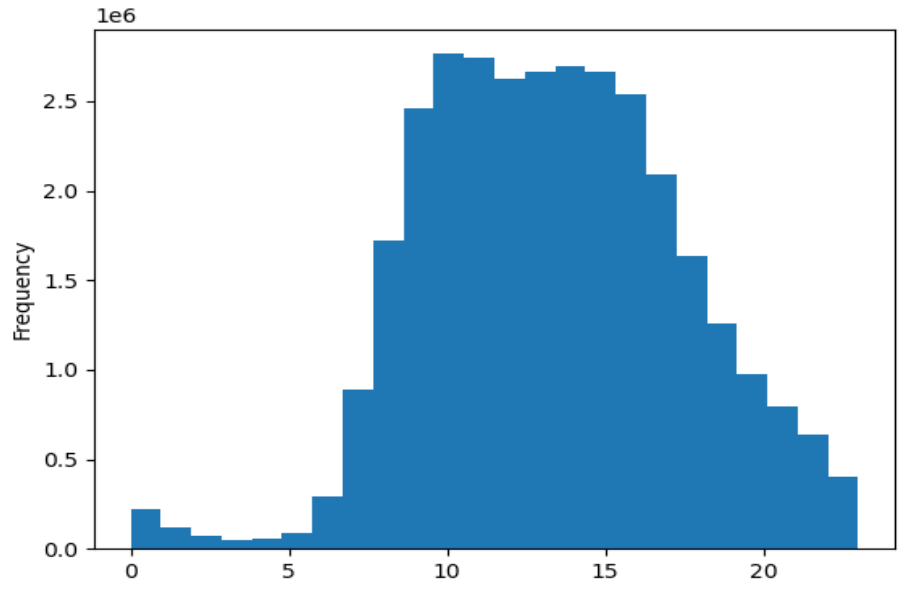
What is the Instacart Basket Analysis Project?

- Project to study a real company using fictionalized data provided by Career Foundry
- Aim of project is to better understand customer spending habits and peak shopping hour
  - Identify busiest day and hour of day for orders
  - Analyze customer behavior to develop rewards program
- Objective of project is to leverage findings to enhance customer experience, optimizing the timing for advertising, sales offers, and coupons

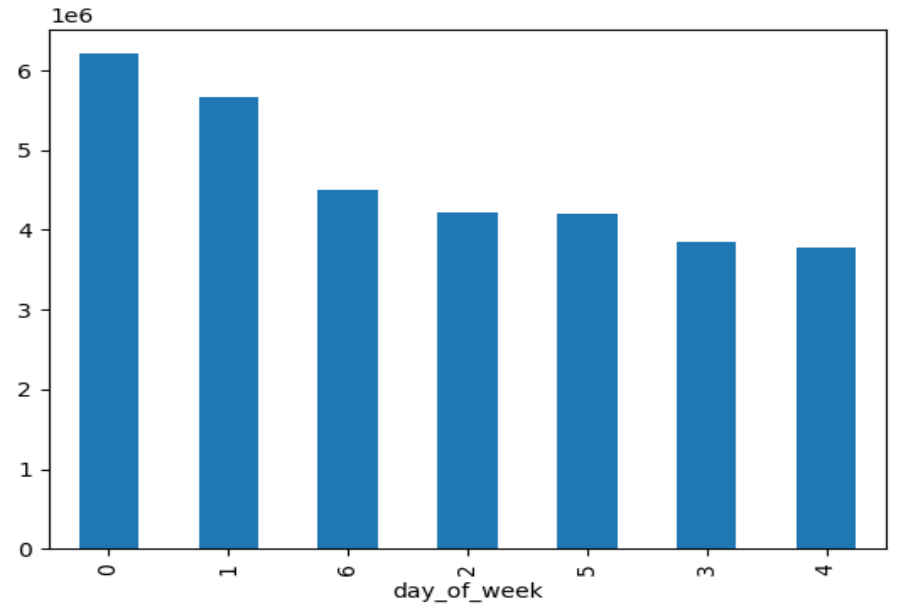
Project Challenges

- Data limitations was the largest challenge of this project
  - Data was manipulated by Career Foundry without a detailed change log
  - Absence of data timestamps, challenging the relevancy of project usage to real world application
- Dataset size was a machine challenge of this project
  - Final report required integration of five distinct datasets; each dataset required its own processing and analysis
  - Extensive size of dataset strained machine memory

Project Insight: Understanding Customer Activity

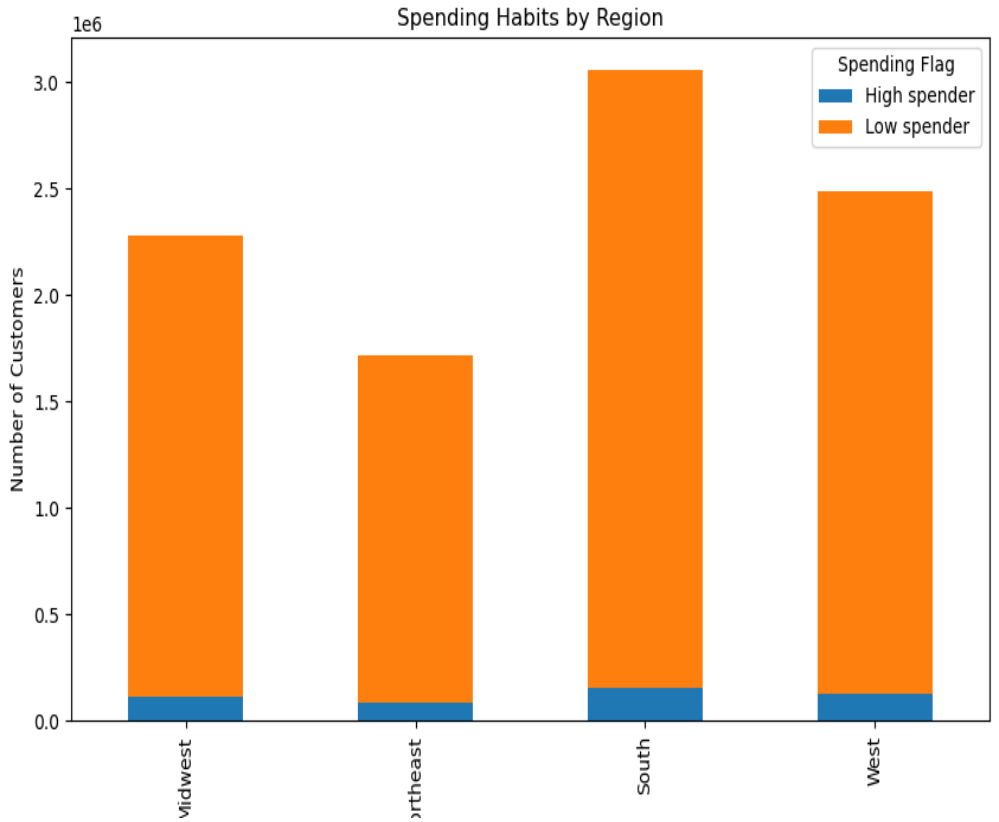


- The times are identified in US military time (ie, 5 is 0500 or 5:00am, 15 is 1500 or 3:00pm)
- Busiest hour of the day starts at 9/0900/9am and persists through 16/1600/4:00pm. Recommend increase advertising at 0600 and continuing through at least 1800 (6:00pm). Include promotions and special offers for



- The days are identified on the x-axis as such: 0 (Saturday), 1 (Sunday), 2 (Monday), 3 (Tuesday), 4 (Wednesday), 5 (Thursday), and 6 (Friday)
- Busiest day is Saturday (0), then Sunday (1). Recommend advertising through peek hours on Sunday and persisting through Monday (2). Persist customer reward promotions Thursday (5) through Monday (2)

Project Insights: U.S. Regional Customer Analysis



- Customers grouped based on their spending, then separated into regions
  - Low spender: mean price of products purchased is lower than 10 (US $)
  - High spender: mean price of products purchased is equal or greater to 10 (US $)
- Visualization created within Python; axis as follows:
  - X-axis shows regions of the country in U.S.
  - Y-axis shows number of customers in millions
- Concluded most customers are in South followed closely by the West, and least customers are in Northeast
  - Continue improving current marketing strategies and store availability to improve and continue positive sales trend in South, West, and Midwest

Project Conclusions and Remarks

- Project & Analysis Challenges
  - Size of final integrated dataset surpassed capabilities of excel and tableau public to create meaningful visualizations
  - Visualizations created with Python limited in presentation & functionality
- Analysis Development
  - Identifies potential for deeper analysis into customer habits to better personalize marketing and advertisement strategies
  - Findings from analysis could be used in predictive analysis and better experience for customers
  - Underscores importance of developing more efficient methods of large dataset collection, integration and analysis

THANK YOU