EDWARD SOMGAL

DATA ANALYTICS CASE STUDY

Instacart Grocery Basket Analysis

- **Objective**

As an Analyst at Instacart, an online grocery store that operates through an app. Instacart already has very good sales, but they want to uncover more information about their sales patterns. My task was 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.

- **Context**

The Instacart stakeholders are most interested in the variety of customers in their database along with their purchasing behaviors. They assume they can't target everyone 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. My analysis will inform what this strategy might look like to ensure Instacart targets the right customer profiles with the appropriate products.

- **Tools Used**

Python using Jupyter

Excel

- **Project Scale**

2 weeks

- **Primary Stakeholders**

CareerFoundry Data Analytics Course

Key Questions from Sales and Marketing team

What are days and hours with most orders?	Time when the most money spent?	What are the products that are most popular?
Analyze different types of customers based on	Their loyalty	Their Region
Their age and family status	Demographic classification	Income

Analytical Techniques

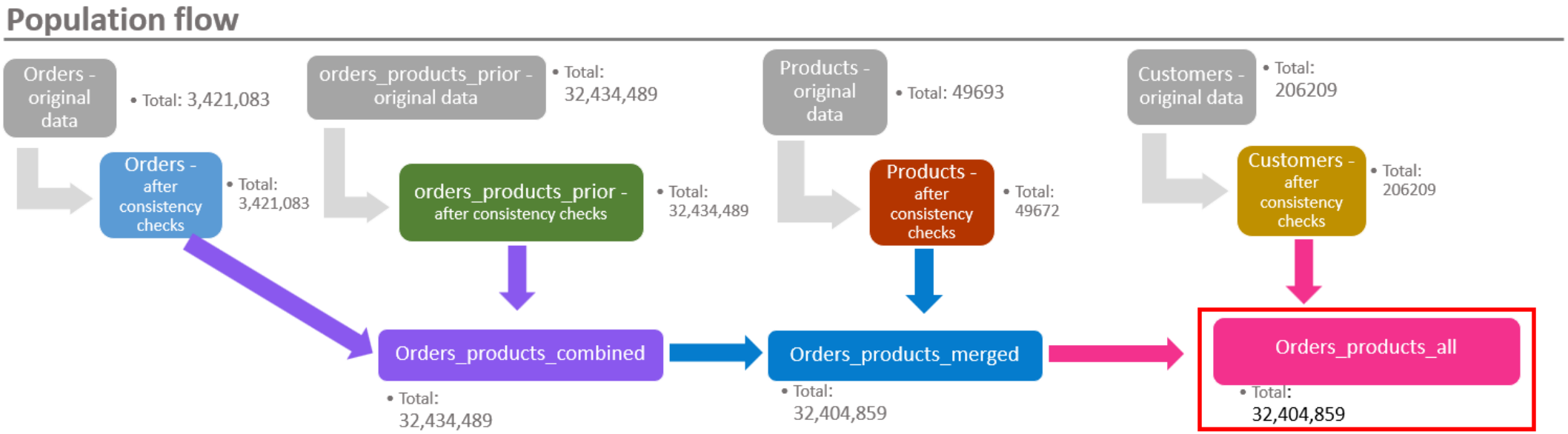
Data Preparation and Analysis	› Data loading › Data Cleaning › Data transformation
Data Exploration and Analysis	› Data exploration › Data filtering and grouping › Aggregation › Data Wrangling › Data Merging › Delivering variables › Data exporting
Data Visualization	› Creating histograms, bar and line charts, scatterplots
Reporting	› Reporting in Excel › Population Flows

Data Info

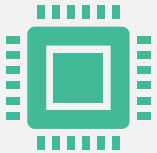
- Instacart, a company from the US, made their data available online for research and training. The customer data has been anonymized for the purpose of this project.
- Instacart customer dataset was downloaded from Kaggle
- Datasets – Orders, Products, Customers were used to merge into one final dataset

Working with the Datasets

- After importing the libraries and the dataset, I started with cleaning the original datasets before merging everything to final dataset. Here are the steps that I followed:
  - Cleaned orders dataset
  - Cleaned orders_products_prior dataset
  - Merged above two datasets
  - Transpose Department Dataset
  - Merge datasets



Challenges

		Solution
	While working with the Python for the very first time, I did a typical mistake of overwriting the dataset.	I downloaded and reloaded dataset to rework. I was cautious when using overwrite and replace functions.
	After merging, dataset was a large file to handle given the basic configuration of my laptop. This made project to slow down significantly	I used data optimization by sampling data to frac=0.1, deleting unnecessary columns, reducing dtypes, closed all unnecessary apps while working with Jupyter.

Customer Insights

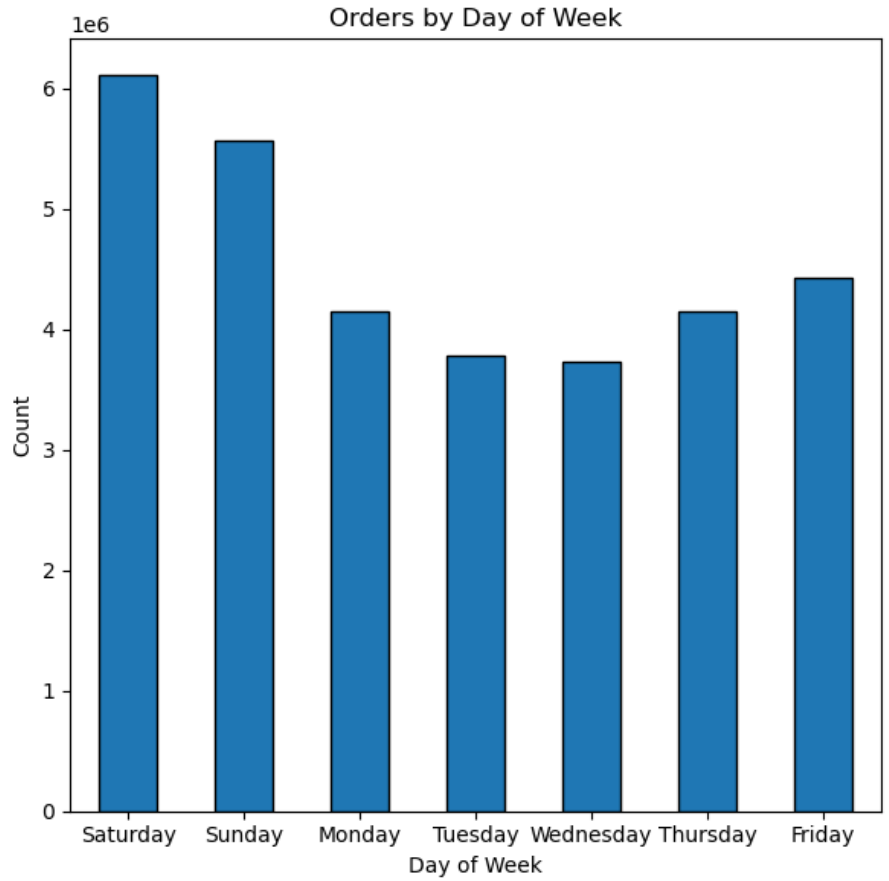
Highest orders on Saturdays

Most orders between 9 am to 5 pm

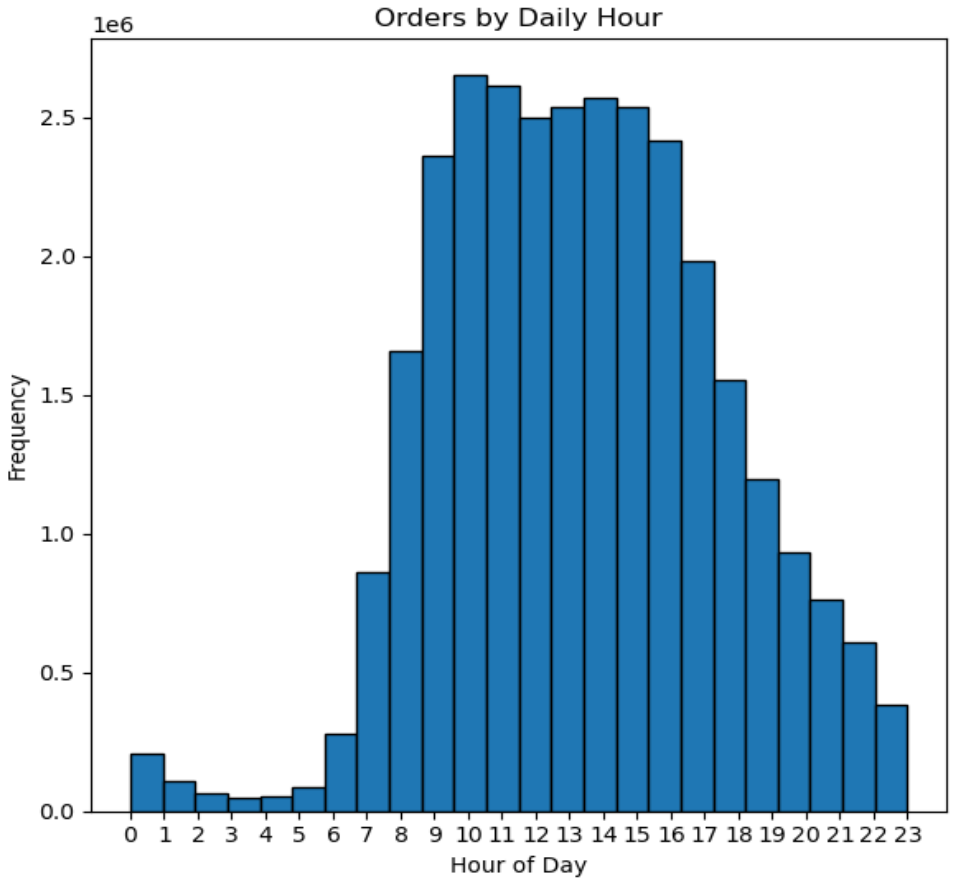
On an average Loyal Customers order every 6 days

Produce is the most popular department

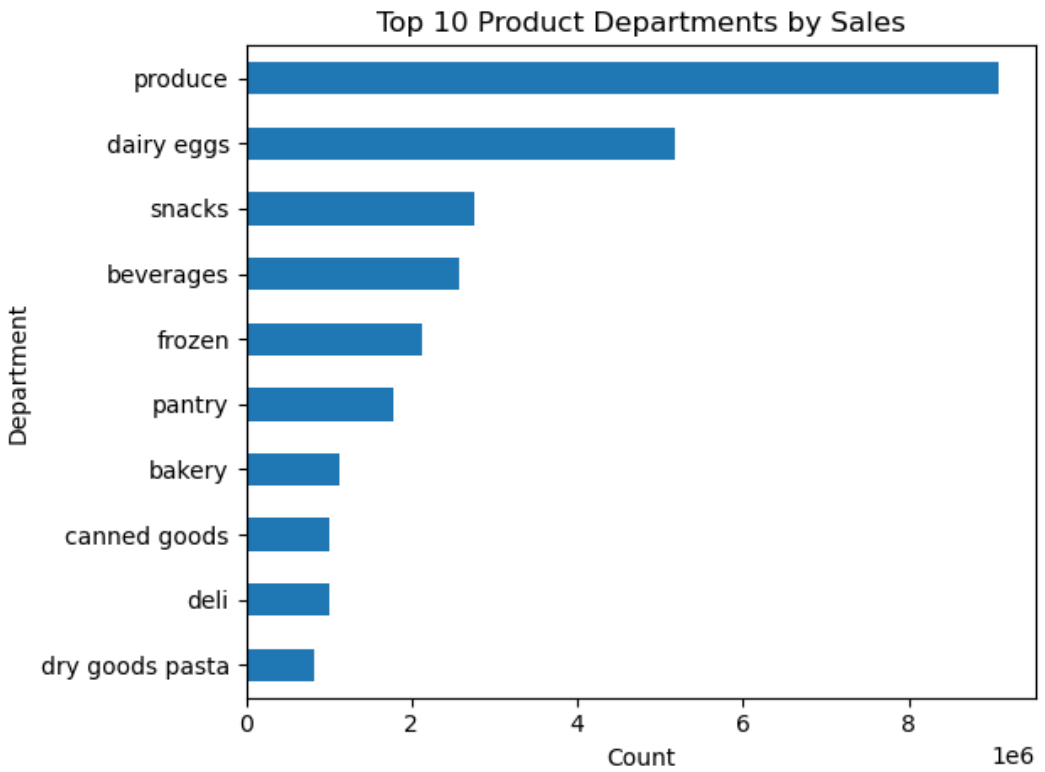
Recommendations



Weekends i.e., Saturdays and Sundays are the busiest days. The least busy period is on Tuesdays and Wednesdays. Therefore, Instacart should schedule ads and release promotions for mid-week to increase sales. Busiest hours are between 10 am to 4 pm.



Here are the ordering habits: Most customers shop low priced products, daily needs items when they are at work between 9 am to 5 pm.



loyalty_flag	mean
Loyal Customer	5.916068
New Customer	14.832121
Regular Customer	11.546425

The top five product categories are Produce, Dairy and Eggs, Snacks, Beverages, and Frozen food. This is the cash cow product for Instacart. Product team could brainstorm on how they can create a niche segment like organic product and keep premium pricing. Consider adding incentives to poorly performing departments to possible entice purchases.

Loyal Customers on an average order every 6 days, whereas Regular Customers order every 11.5 days and New Customer order once in almost 15 days. To encourage frequent orders, Instacart could consider giving vouchers or coupons which expire in one week.

Retrospective

GitHub Link

- **What went well?**
  - Successfully identified and addressed data quality issues, including an outlier product price of $99,999.
  - Effective data cleaning and wrangling enabled critical business insights.
- **Challenges Overcome**
  - Managed large dataset sizes, optimizing queries to minimize crashes and improve performance.
  - Developed time management skills, studying additional concepts to enhance efficiency.
- **Key Takeaways for Future Improvements**
  - Conduct descriptive analysis and grouping before merging datasets to reduce size.
  - Selectively merge necessary columns to optimize performance.
- **Conclusion**

Working on this Instacart Grocery Basket Analysis case study was a valuable experience. I gained insights into:

- Timely advertisement campaigns
- Target audience identification
- Best-selling product categories
- Actionable recommendations for sales and marketing teams

I'm proud to have delivered actionable business insights, enhancing the company's growth potential.

Edward Somgal

GitHub:
https://github.com/EdwardSomgal/Instacart-Market-Basket-Analysis

LinkedIn:
https://www.linkedin.com/in/edward-somgal/