Instacart Basket Analysis

Scenario: A well-established company wants to discover more information about its client base. Assuming every demographic cannot be targeted in the same way, they are contemplating a marketing campaign that will customize ad messages to specific groups in hopes that they will have a positive effect on sales.

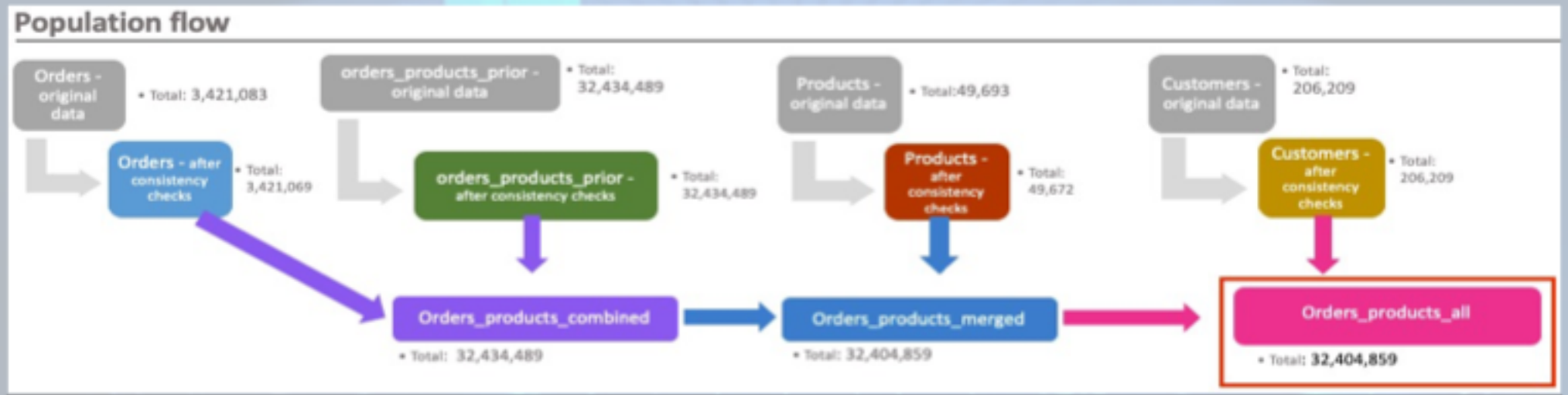
Objective: Investigate, clean, wrangle, subset, and merge datasets using Python - Pandas, Numpy, Matplotlib, and Seaborn in Jupyter Notebooks. Detail steps and findings in user-friendly Excel report.

Key Questions: *What are the busiest days of the week, and hours of the day? *What times of the day do people spend the most money? *Can products be neatly grouped into categories in terms of price range? **What types of products are the most popular? *Can customers be neatly grouped into categories in terms of their loyalty to Instacart? *How do ordering habits differ based on a customer's loyalty status? *How do demographics affect customers' ordering habits?

Data Sets: Orders and products, departments, orders products prior, customer

Instacart Basket Analysis

There were multiple datasets in this project that required cleaning, wrangling, merging, and derivations, to make the analyses that were called for.



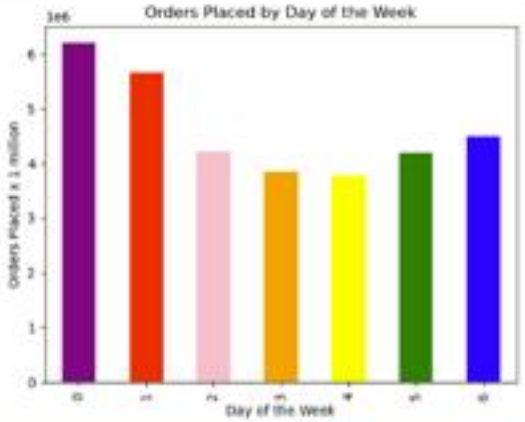
1.) The grey boxes in the first row of the population flow represent the original data sets as they were when downloaded. In the Total fields the count of the rows was added when the data set was imported into Jupyter.

2.) The second row of boxes (colored) represents the data sets after they were manipulated, e.g., removed missing values and duplicates. In the Total fields, the count of the rows was added after conducting these operations. This offers a visual overview of how the data *flowed* throughout the data consistency checks.

3.) The third row, where the arrows are also colored, represents the merges performed between the datasets. In the Total fields, the count of the rows was added in the merged datasets, so that's what ended up as the final dataset (in the red box).

Instacart Basket Analysis

Charts created in Python with Matplotlib helped answer questions about what day of the week and time of day had the highest volume of orders, and uncover more insights.



The weekends are the busiest: Saturday, then Sunday. Friday receives roughly 1.2 million less orders than Saturday. Monday and Thursday are next, and then Tuesday and Wednesday are last.

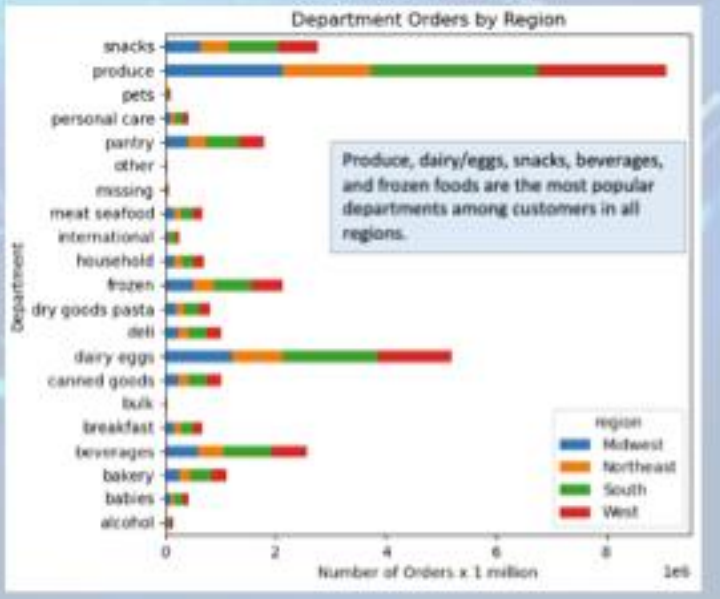
0 = Saturday
1 = Sunday
2 = Monday
3 = Tuesday
4 = Wednesday
5 = Thursday
6 = Friday

The vast majority of orders are placed between the hours of 10am and 5pm. Ordering activity is by far the lowest between the hours of 2am and 6am. There is a sharp rise between 7am and 9am. Then a gradual fall from 6pm to 1am. 10am is the most active hour. 4am is least active.

Note: 0 on the hour scale = 1am and 23 =midnight.

Sales by customer loyalty categories, products by department and demographics, also provided valuable information.

Regular customers, those who have placed between 11 and 40 orders with IC make up more than half of the customer base. It would likely benefit IC to market to these customers with targeted offers and deals on the products they have previously ordered. Loyalty programs would likely help to retain loyal and regular customers. Personalized marketing and incentivized referrals and feedback would likely prove helpful in further solidifying loyal customers. E-mail and social media marketing can help build relationships and promote sales with new customers.



Produce, dairy/eggs, snacks, beverages, and frozen foods are the most popular departments among customers in all regions.

Instacart Basket Analysis

More Data Discoveries:

- Based on a sample of the entire data, the highest expenditure per order comes at 5am. The lowest expenditure per order is at 10am.
- Mid-range products, those between $6 and $15, are the most-ordered products.
- Loyal, regular and new customers largely order from the same departments overall.
- Married couples make up the largest portion of IC customers among all age groups.
- Lower-middle to upper-middle earners (those with an income between $40k and $160k) make up the largest active IC customer base by far.

More Recommendations:

- Offer big discounts on new and different products to encourage newer customers to try out more products.
- Market products geared toward married couples.
- Market to lower-middle, middle, and upper-middle earners.

See full project in GitHub Repository here.