Yoana Boneva

CareerFoundry Data Analytics Graduate

Instacart Grocery Basket Analysis- Case Study

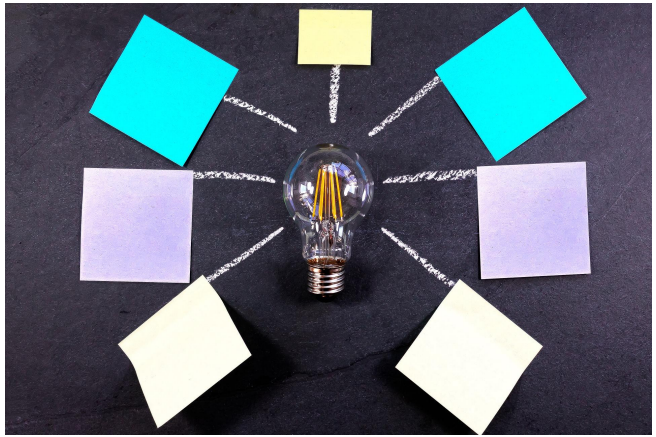
Instacart Grocery Basket Analysis- Case Study

Context

Instacart is an online grocery store that operates through an app. It already has very good sales, however we want to uncover more information about our sales patterns by performing an initial data and exploratory analysis of some of our data points. The main goal is to derive insights and suggest strategies for better segmentation based on the provided criteria.

Objective

I am a Data Analyst working for Instacart. My analysis will inform what this strategy might look like to ensure Instacart targets the right customer profiles with the appropriate products.



Python Skills Used

- Importing Libraries*
- Importing & Exporting Data
- Descriptive Analysis
- Data Wrangling Procedures
- Data Merging
- Deriving New Variables
- Grouping Data
- Aggregating Data
- Visualizations with Python Libraries*
- Reporting in Excel
- Population Flows

*Python libraries: Pandas, NumPy, Matplotlib and Seaborn (for visualizations) and SciPy (for computations).

Data set

The Instacart Online Grocery Shopping Dataset 2017, Accessed from https://www.instacart.com/datasets/grocery-shopping-2017 on 2021-10-15.

Please Meet My Stakeholders

Marcos Galán
Senior Vice President of Sales

”We need to know what part of our offering has the lowest market share and why. Based on this input, we could improve this sector and boost sales.”

Ixchel Valdía
Vice President of Marketing

“We’re always looking into improving our targeting for ad campaigns.”

Imani Jackson
Instacart Customer

“I want to receive ads, promotions, and recommendations that are relevant to the products I order regularly.”

My Stakeholder's Key Questions

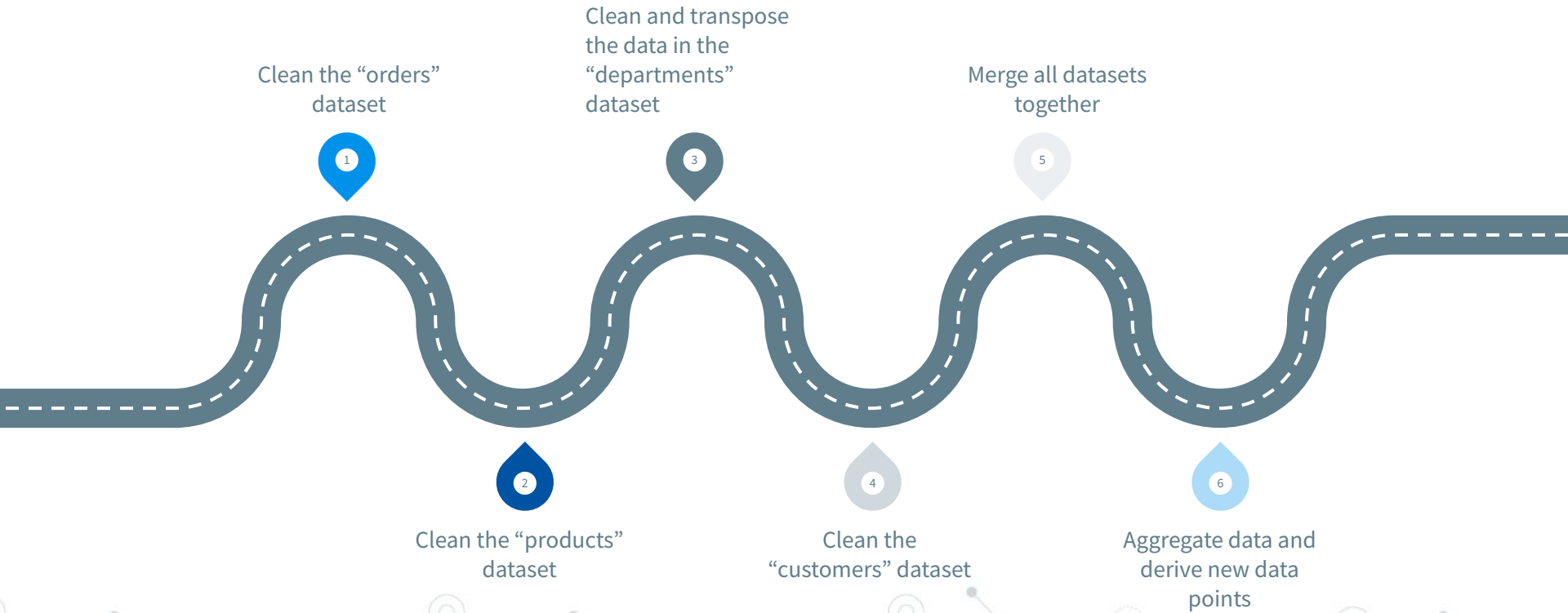
What are the busiest days of the week and hours of the day ?

Are there particular times of the day when people spend the most money?

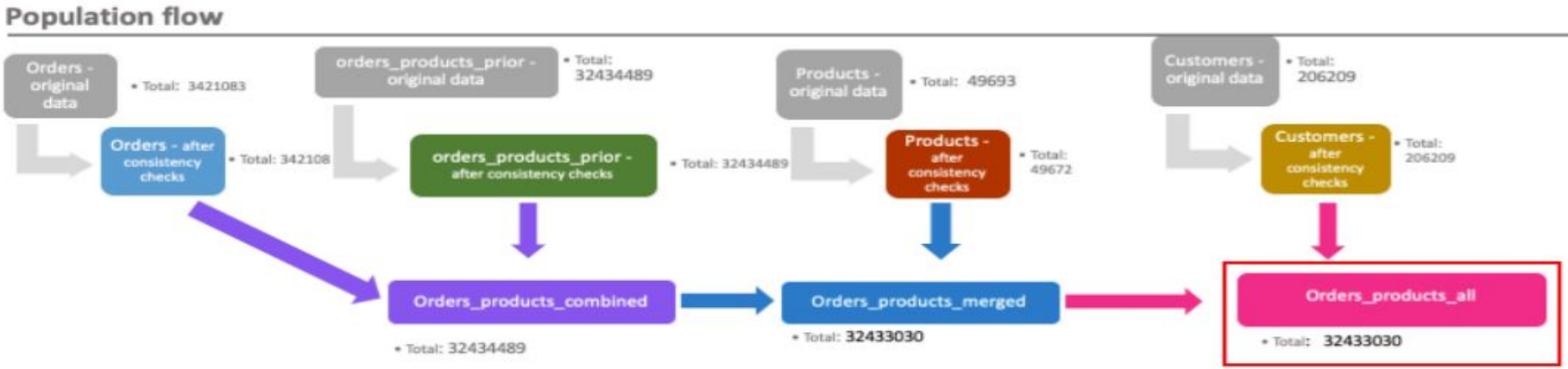
Are there certain types of products that are more popular than others?

User Analysis

Analysis Roadmap



Progress in the Population Flow



Challenges

- The size of the datasets was really big which slowed down significantly my machine.
  - I explored options for using Google Colab, which uses Google Cloud server for processing.

- I had an issue with assigning values to different categories I was trying to create with my data.
  - I found out that I was missing a line of code to assign the profile to a column.

- Creating nice visualisations from my data is really important for me. Creating them in Python in the beginning felt less smooth than Tableau for example.
  - I researched online for different color palettes and options for optimizing my visualisations beyond the basic options covered in the course.

Next Steps

- Practice the skills learned in this task (*data cleaning, data wrangling, data grouping and profiling*) on further projects.
- Explore other options for preparing an analysis in R.
- Further develop my knowledge on creating beautiful visualisations through Python libraries.

Key Findings

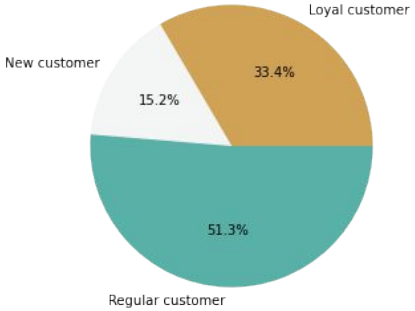
Saturday & Sunday
The most busy days of the week.

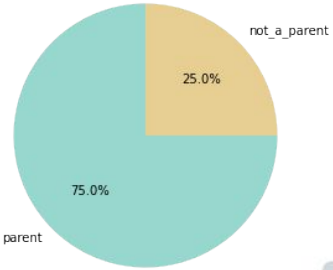
10-18h
The most busy times of the day.

6 days
Loyal customers order every 6 days.

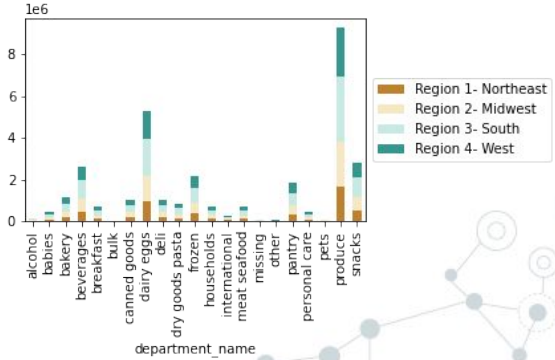
Produce
Produce is the most popular department.

We Make Data Driven Decisions





- Push more notification adds / offer more discounts for the high range priced products- during the days customers are less active (Mo, Tu, We, Th).
- Push adds for more expensive goods in the late hours.
- We have much more parents than non parents- we need to offer bigger variety of family oriented goods.
- Reward the loyal customers, push the regular once to become loyal too & retain the new ones through different loyalty campaign, collection of membership point, etc.



Thanks!

Any questions?

You can reach me at:

- ➔ yoana.boneva@hotmail.com
- ➔ LinkedIn

Or check out my GitHub profile.