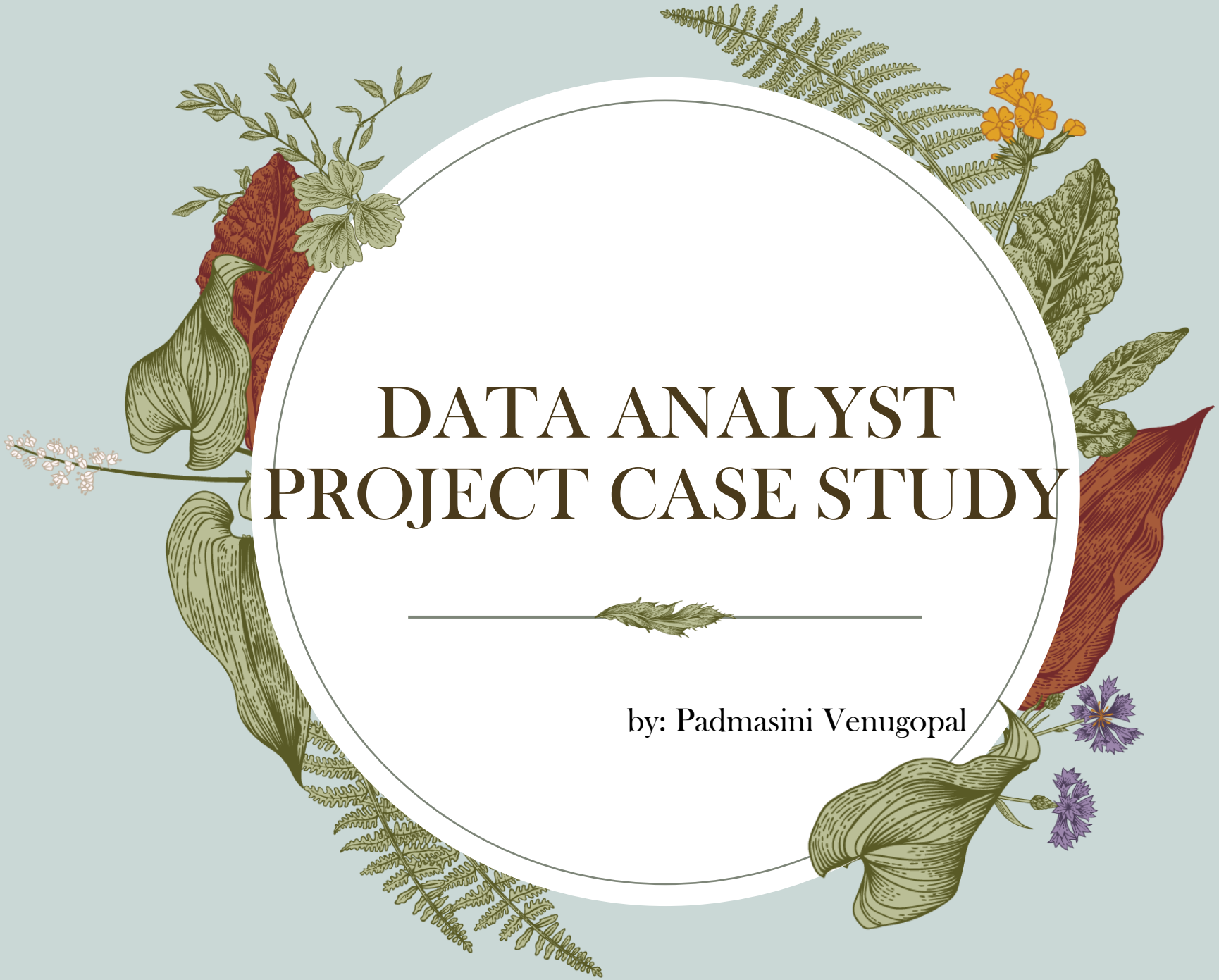


DATA ANALYST PROJECT CASE STUDY

by: Padmasini Venugopal

INTRODUCTION

Instacart, is an online grocery store that operates through an app. Performing an initial data and exploratory analysis of some of their data in order to derive insights and suggest strategies is the primary objective of this analysis.

Project Overview

DATA AND PROJECT

It's an open-source data set from Instacart.

- Data: customers departments,orders, products
- Data dictionary: source
- Project brief: source

DURATION

It took almost 4 weeks for me to

complete the entire analysis.

TOOLS USED

- Microsoft Excel
- Python –Jupyter Notebook
- Panda
- Anaconda

TECHNIQUES APPLIED

- Data cleaning: wrangling and Subsetting
- Data consistency check
- Combining and exporting data
- Deriving new variables
- Grouping data and Aggregating variables
- Python visualization and Excel report

THE PROCESS

1. THE DATA PREPARATION PHASE

To streamline the Instacart sales data, I conducted a thorough scan for duplicates, missing numbers, and other anomalies. I subsequently eliminated extraneous columns and duplicates from the data set.

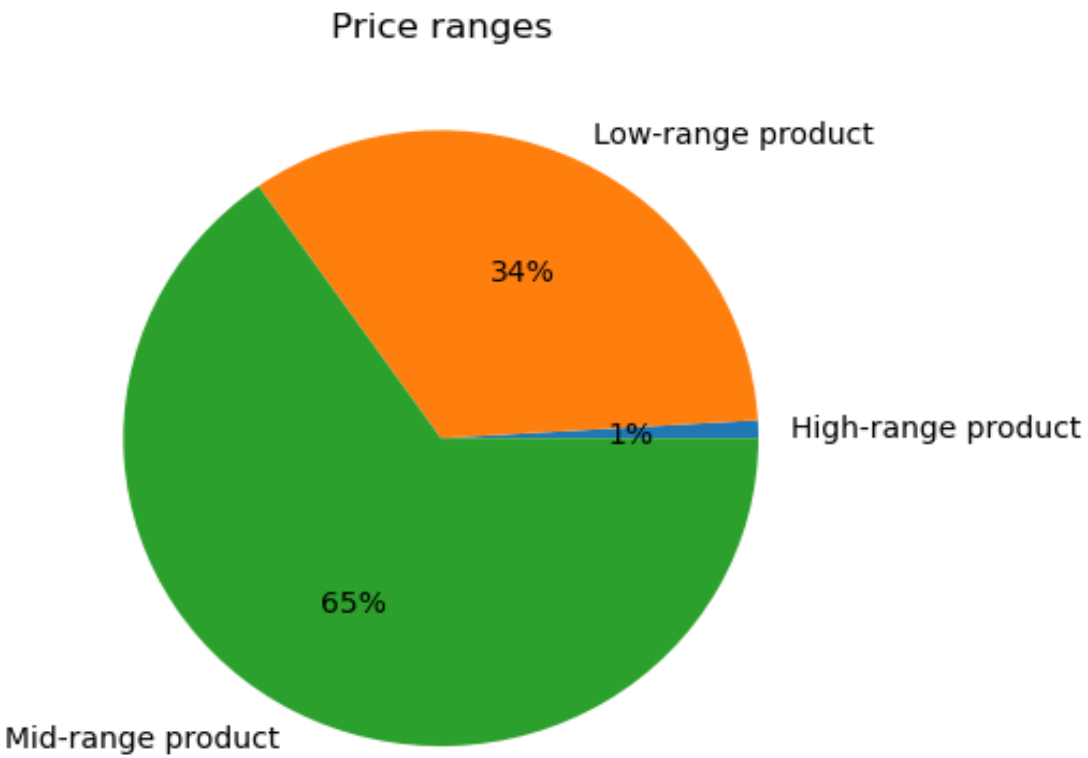
2. THE ANALYSIS PHASE

After receiving answers to my queries from the management team, I proceeded with my investigation. I merged necessary data using Python, created subgroups as new columns using the LOC function, and performed the analysis by checking for the different probabilities.

3. THE RESULTS

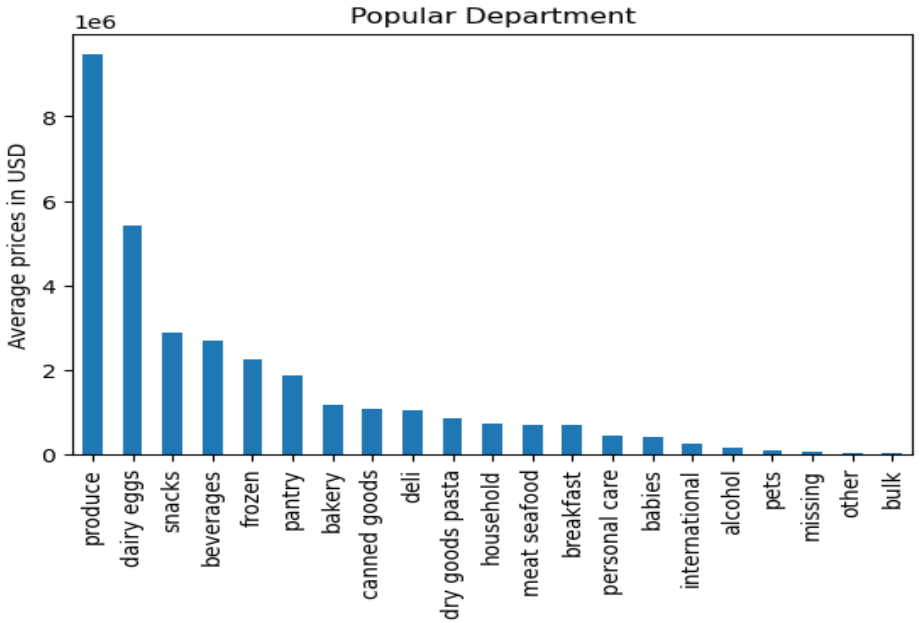
Finally, after all the analysis the final presentation was created using Excel, and the visualizations were created using Python. The stakeholders were advised based on the results of the analysis with a clear disclaimer regarding the limitations of the data.

ANALYSIS

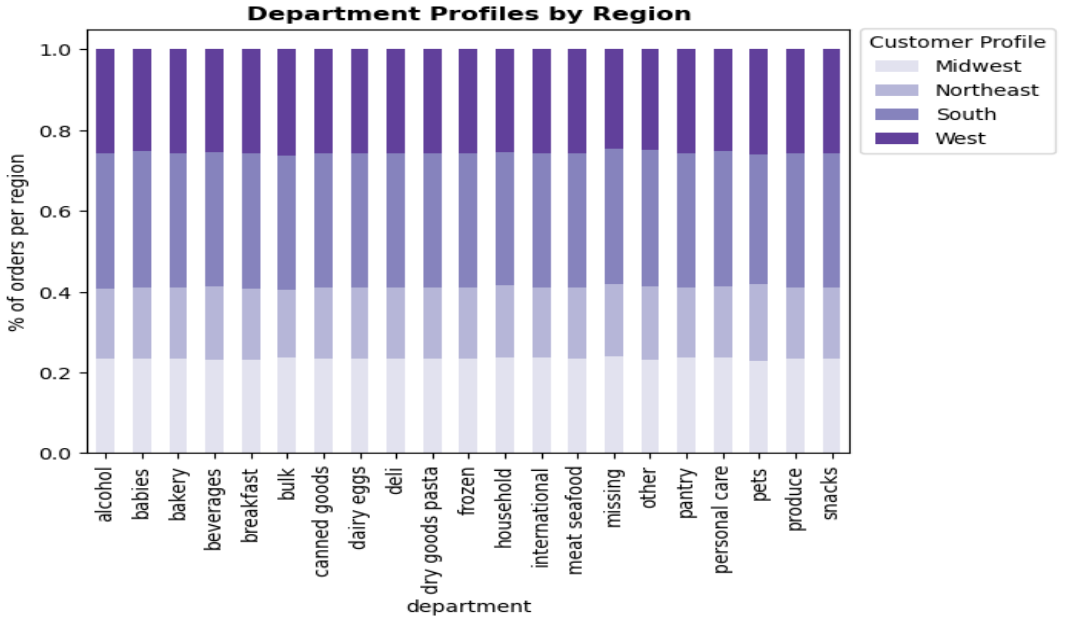


According to the pie chart,
Price ranges were categorized into 3 parts - High, Low, and Mid-range products.
According to the data, the mid-range products are higher at 65%, the low range at 34%, and about 1% of high-range products.

ANALYSIS

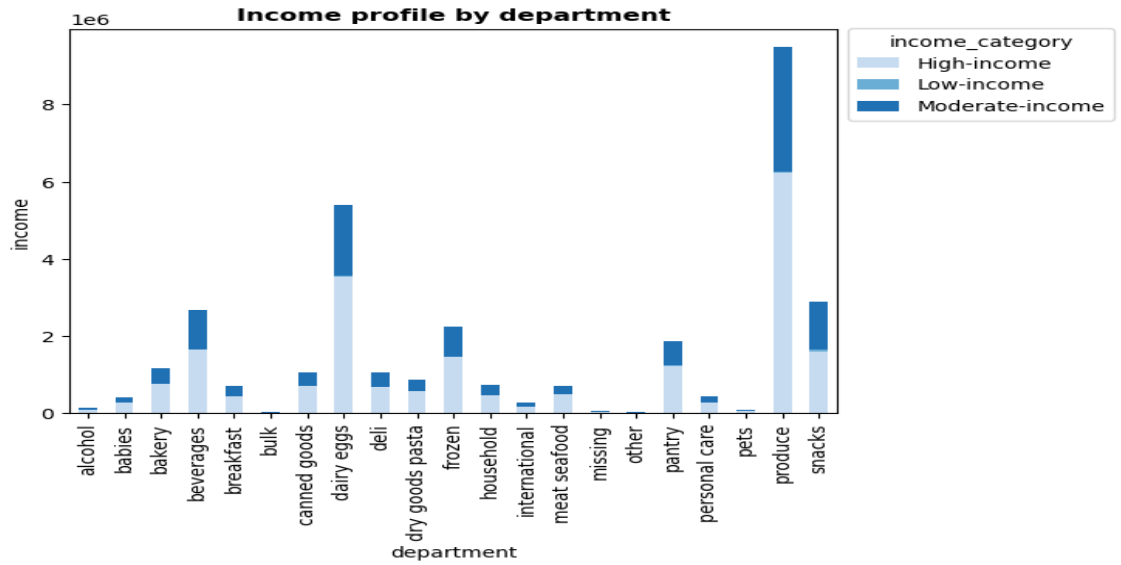


The most popular department with high purchase frequency is "produce, dairy eggs, snacks".
The least sold department is "Pets, other, bulk".

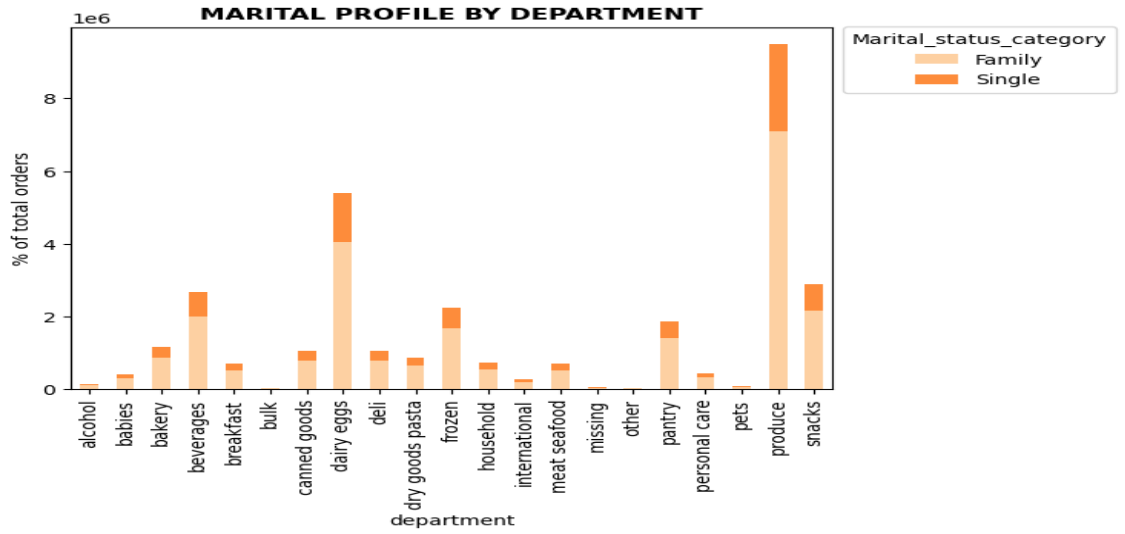


The ordering habits across all departments remain consistent for all four regions.
The south region represents the highest number of orders.

CUSTOMER BEHAVIOURAL ANALYSIS



Produce, dairy eggs, and snacks are the top three departments across all income profiles and the least three departments are pets, missing, and others.



From the stacked bar chart,
Families and singles mostly prefer produce, dairy, and snacks.

RETROSPECTIVE

WHAT WENT WELL?

I was surprised to discover a new tool I was unaware of all these years. Python just made the work so much easier when compared to all the other tools especially when the data is larger in numbers.

CHALLENGES FACED?

At first, learning Python was a huge challenge for me. It was difficult to learn a new programming language and remember all the necessary code. However, with consistent practice, I have gained some hands-on experience. Although I am not yet perfect, I am still learning and practicing.

RECOMMENDATIONS

- Create more marketing campaigns for young age groups by introducing trendy pay-later options.
- Instacart must enhance the marketing for the least sold products such as pets, babies, and alcohol.
- Family-oriented promotions by developing marketing campaigns that cater to the needs of families.
- For young mothers promoting the baby's products to them are advised.

Thank you

The Link to the Final Presentation

Click here

For further queries kindly contact
drpadmasini.v@gmail.com