

INSTACART

Grocery Basket Analysis

CASE STUDY



Table of contents:

1. OVERVIEW
2. OBJECTIVE & GOAL
3. KEY QUESTIONS
4. PROJECT TASKS & DELIVERABLES
5. DATA
6. MY WORKFLOW
7. KEY INSIGHTS
8. KEY RECOMMENDATIONS
9. CHALLENGES

OVERVIEW

Instacart is an online grocery store that operates through a mobile app. It has already achieved significant sales success. However, Instacart aims to gain a deeper understanding of its sales patterns and customer base.



OBJECTIVE

Analyzing Instacart's data, identify customer segments, understand their buying habits, and recommend tailored marketing strategies to maximize sales by catering to the unique needs and preferences of each segment.



GOAL

The main goal is to support Instacart in developing a targeted marketing strategy. This strategy aims to tailor marketing campaigns to different customer segments based on specific criteria. By doing so, Instacart intends to assess whether these targeted efforts can influence product sales positively.

Key questions

Temporal

When are the busiest days and hours?
Do customers spend more money at specific times of the day?

Products

How can we group products by price ranges?
Which product categories get the most orders?

Customers

What's the distribution of customer loyalty levels?
Do ordering habits differ by region, age, or family status?

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

Vice President of Marketing

Importing and
exporting data

Importing libraries

Data Consistency
Checks

Data Wrangling &
Subsetting

Data merging

Data aggregating

Deriving Variables

Project Tasks & Deliverables

While working on this project, I delved into Python, creating successive elements that ultimately came together into the final project.

Data



The project was based on open-source data made available online by Instacart: 'The Instacart Online Grocery Shopping Dataset 2017,' which was accessed from www.instacart.com/datasets/grocery-shopping-2017 via Kaggle*.

For my analysis, I utilized the following datasets: 'orders', 'products', 'orders_products_prior', 'departments' and 'customers'.

My workflow

My workflow covers data preparation, cleaning, integration, and visualization steps, which are crucial in deriving meaningful insights for decision-making

08 / 16

1. Importing libraries and data sets: 'orders' and 'products' data sets.

This step involves loading necessary libraries and importing the initial data sets.

2. Renaming column names for intuitive names.

Making column names more understandable and intuitive for easier analysis.

3. Changing data types.

Ensuring that data types are consistent and appropriate for analysis.

4. Importing 'departments' data set.

5. Transposing 'departments' data set.

Restructuring the data set from horizontal orientation to vertical.

6. Conducting descriptive statistics.

Exploring basic statistics to understand the data.

7. Addressing missing values from 'product_name' column ('products' data set) and 'days_since_prior_order' column ('orders' data set).

Handling missing data in the specified columns.

8. Removing duplicate rows from 'products' data set.

Eliminating duplicate records.

9. Combining 3 data sets together: 'orders', 'orders_products', and 'departments'.

Merging multiple data sets to create a comprehensive data set.

10. Creating new columns to answer business questions.

Deriving new features or variables to address specific business questions ('price_range', 'busiest day of the week', 'busiest period of the day', 'loyalty_flag', 'spending_flag', 'frequency flag').

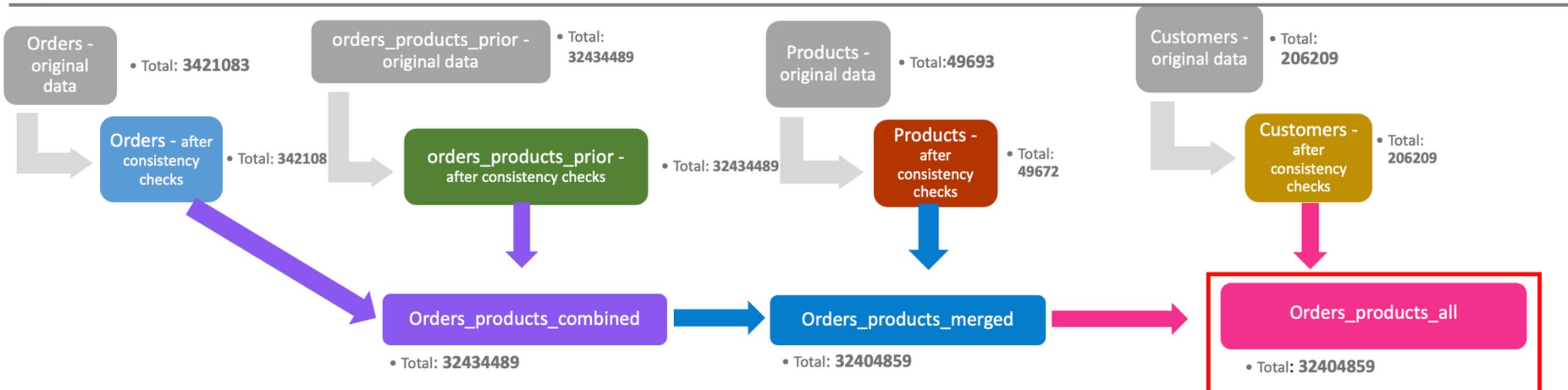
My workflow

- 11. Importing 'customers' data set.**
- 12. Renaming column names** for intuitive names.
- 13. Checking for duplicates.**
- 14. Removing columns** with Personal Identifiable Information (PII).
Protecting sensitive data by removing or masking PII.
- 15. Changing data types.**
- 16. Combining 'customers' data set** with previously merged data.
Integrating customer data into the existing dataset.
- 17. Creating new columns** to answer business questions:
'regions', 'age group', 'income category', 'household type'.
- 18. Aggregating** the maximum, mean, and minimum values of variables related to usage frequency and expenditure at the customer-profile level.
- 19. Creating visualizations** to answer business questions.
Creating charts, graphs, and plots to visualize data to extract insights.
- 20. Creating a final report.**
Summarizing the entire analysis process, including population flow, data consistency, data wrangling steps, column derivations, visualizations, and recommendations.

My workflow



Population flow



Key findings

Saturday

busiest day of the week

10 a.m.

busiest hour of the day

3 a.m.

hour when customer
spend the most money



Key findings

49,627

different products with prices ranging from \$1 to \$25.

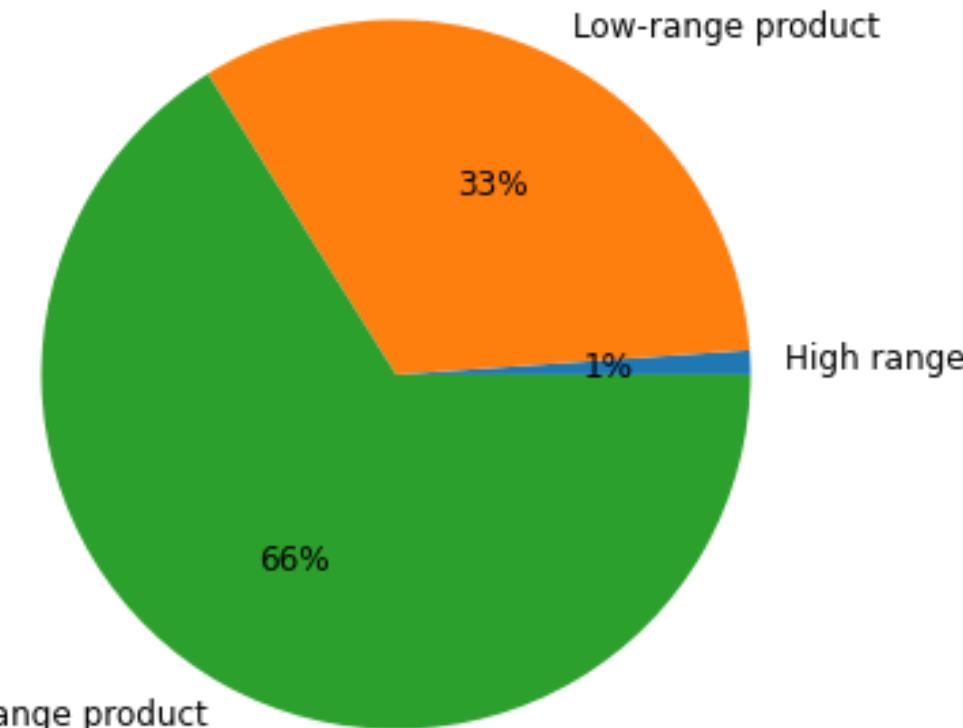
3 price range groups

Low-range under \$5, Mid-range \$5-\$15, High- range over \$15.

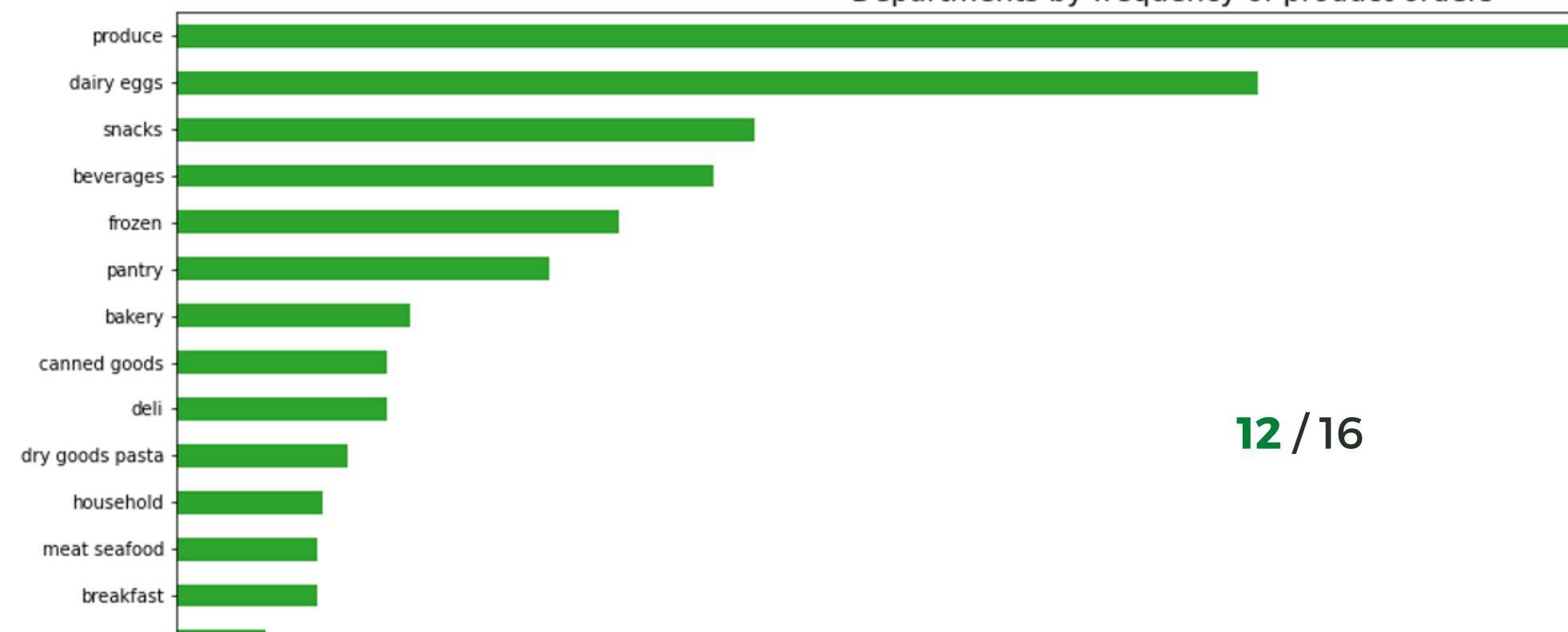
Produce

most popular department

Products by price range



Departments by frequency of product orders



Key findings

164,631

customers

Loyalty

47% regular customer

42% new customer

10% loyal customer

South

region with most of customers

65+

biggest age group

Loyal customers order every 6 days.

Upper middle-class families are the largest customer group, ordering the most products and generating the most revenue.

Most customers are low spenders.

43% of customers have upper income.

Almost a quarter of our customers are non-frequent (order every 20 days or more).

The popularity of departments remains consistently similar across all user profiles.

Key recommendations

Temporal

Offer exclusive discounts, deals, or promotions that are valid only on weekdays.

Advertise more expensive items during rush hours 10a.m.- 5p.m.

Products

Implement clear and user-friendly pricing labels for products to help customers easily identify low-, mid- or high range products.

Implement cross-promotional campaigns between popular departments (discounts on complementary products from other department).

Customers

Offer short-term coupons to incentivize non-frequent customers to shop more frequently.

Consider introducing more high-range products to meet the needs of customers with upper income.

Challenges

HARDWARE LIMITATIONS

Working with a vast database can be challenging due to hardware limitations. Fortunately, in my case, it only required extra patience because some tasks took an extensive amount of time to complete.

CONTENT SELECTION

Throughout the analysis process, I gathered a wealth of information and created numerous visualizations. The challenge I faced was selecting the most valuable ones among them. However, I'm hopeful that I successfully navigated this by making thoughtful choices.

LIMITED PYTHON SKILLS

In my current Python journey, my skills are limited, and I faced challenges in gaining certain information or creating advanced visualizations. Nonetheless, I'm dedicated to advancing my Python skills to enhance the project with new capabilities in the future.

Thank You

INSTACART GROCERY BASKET
CASE STUDY

This project was created as part of the
'Python Fundamentals' course in the CareerFoundry
Data Analyst program

