Instacart Grocery Basket

Overview

Objective

Instacart is 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. The objective of this study is to analyze customer behavior and find ways for growth.

Context

The Instacart stakeholders are most interested in their database's variety of customers and purchasing behaviours. 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 affect the sale of their products.

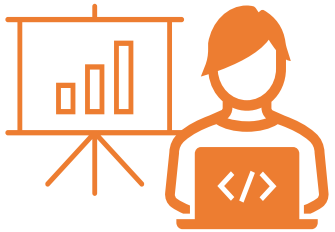
Questions to answer to the Marketing Team

1. What are the busiest days of the week and hours of the day
2. Which are the hours where most of the money is spent
3. Are certain type of products more popular than the others
4. Analyse customers based on loyalty, age, and demographic classification

Data

The dataset is fabricated for this exercise. It comprises information on Orders, Products and Departments. The full details of the data are available here

Skills



- Data wrangling
- Data merging
- Deriving variables
- Grouping data
- Aggregating data
- Visualization in Python
- Reporting in Excel
- Population flows

Tools

python jupyter NumPy pandas Matplotlib seaborn

GitHub

Click here

Analysis

Cleaned, merged and analysed 32 million rows of records

pandas

Programming

```
# Defining a function

def price_label(row):

  if row['prices'] <= 5:
    return 'Low-range product'
  elif (row['prices'] > 5) and (row['prices'] <= 15):
    return 'Mid-range product'
  elif row['prices'] > 15:
    return 'High range'
  else: return 'Not enough data'
```

Ifelse + function

NumPy

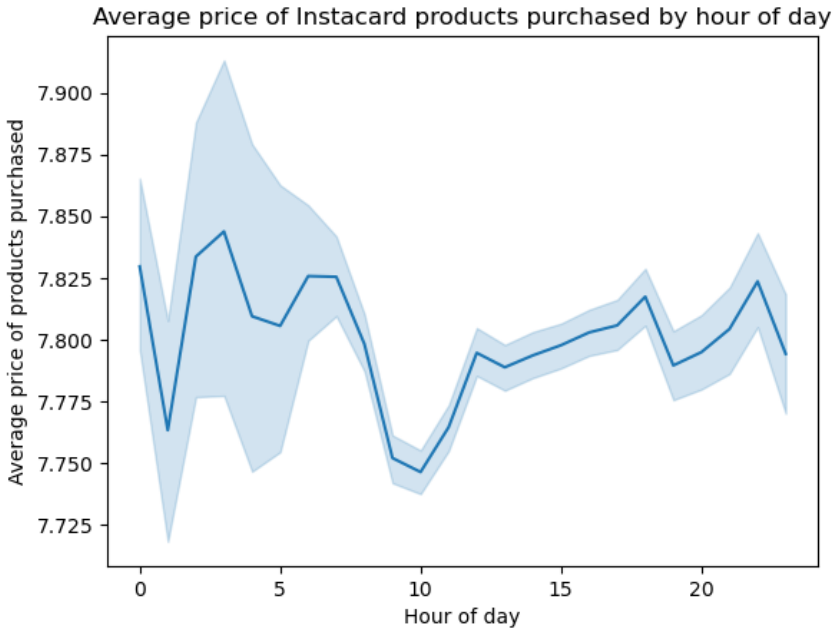
```
result = []

for value in df_orders_products_merged['orders_day_of_week']:
  if value == 0:
    result.append('Busiest day')
  elif value == 4:
    result.append('Least busy')
  else:
    result.append('Regularly busy')
```

Ifelse + for loop

```
df_orders_products_merged.loc[df_orders_products_merged['prices'] > 15, 'price_range_loc'] = 'High-range product'

df_orders_products_merged.loc[(df_orders_products_merged['prices'] <= 15) & (df_orders_products_merged['prices'] > 5), 'price_range_loc'] = 'Mid-range

df_orders_products_merged.loc[df_orders_products_merged['prices'] <= 5, 'price_range_loc'] = 'Low-range product'

df_orders_products_merged['price_range_loc'].value_counts(dropna = False)
```

Ifelse + loc()

Visualizations



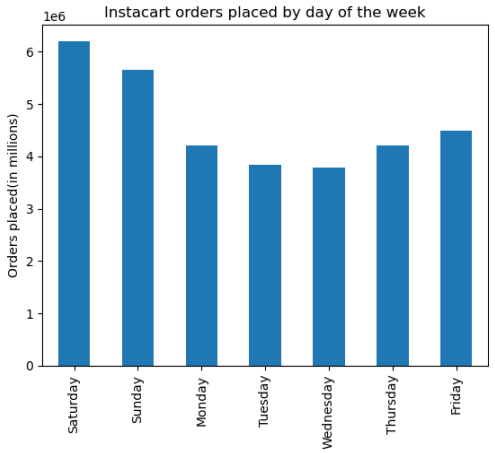
Matplotlib

seaborn

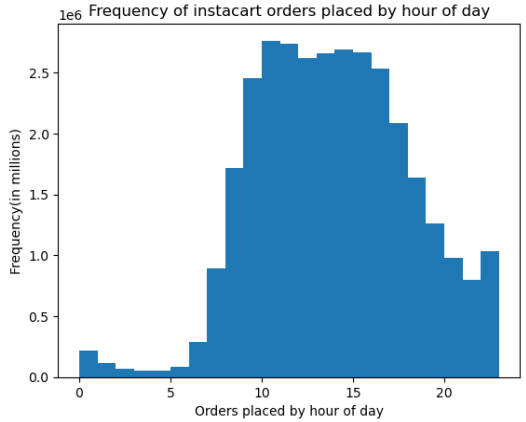
Challenges

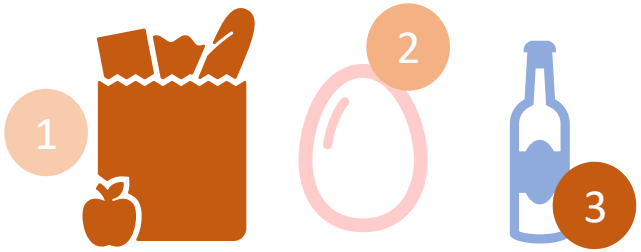
This project was both my favourite and most challenging one. I enjoyed working with such an extensive dataset, and even though it was demanding, I developed the skill of manipulating data through programming in Python. On the other hand, developing client reports took me some time to understand how to simplify and organise complex results so my audience could better grasp what I wanted to express. But with the help of my great tutor and mentor, I was able to cross the barrier.

Busiest days:
Weekend

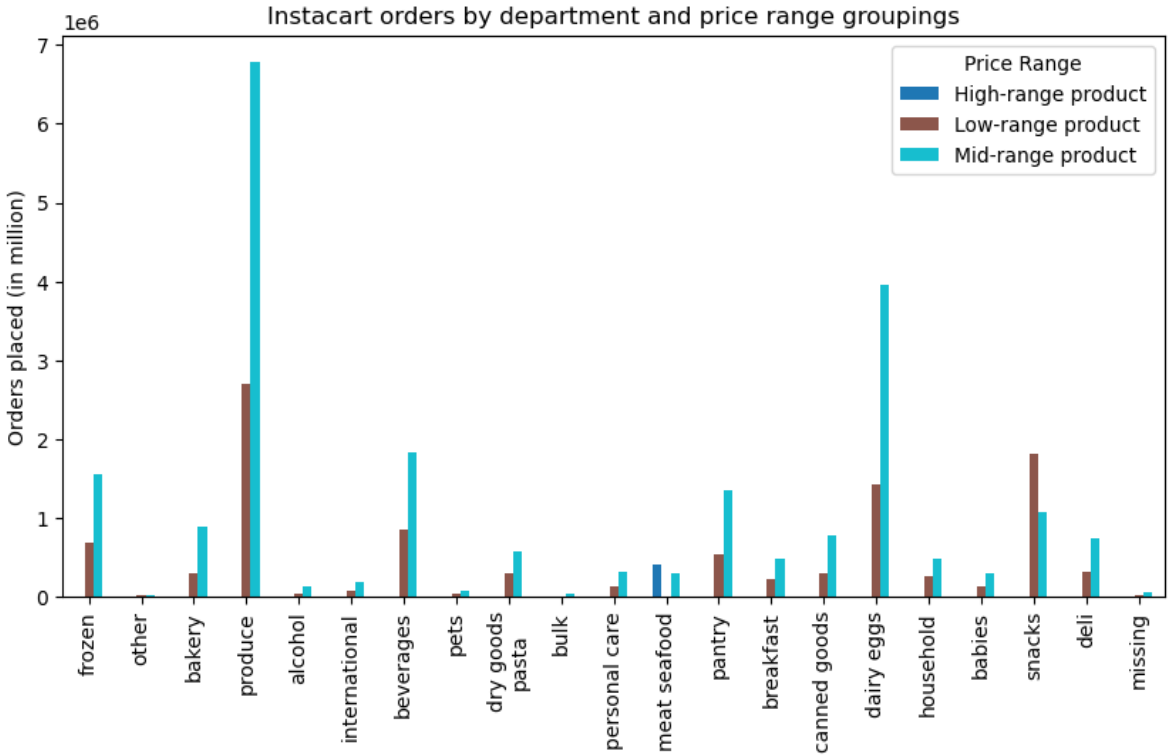


Busiest time:
9:00 to 17:00





Mid-range ítems ($5-15) in the departments: Produce, dairy eggs, beverages, and frozen foods are the most popular.

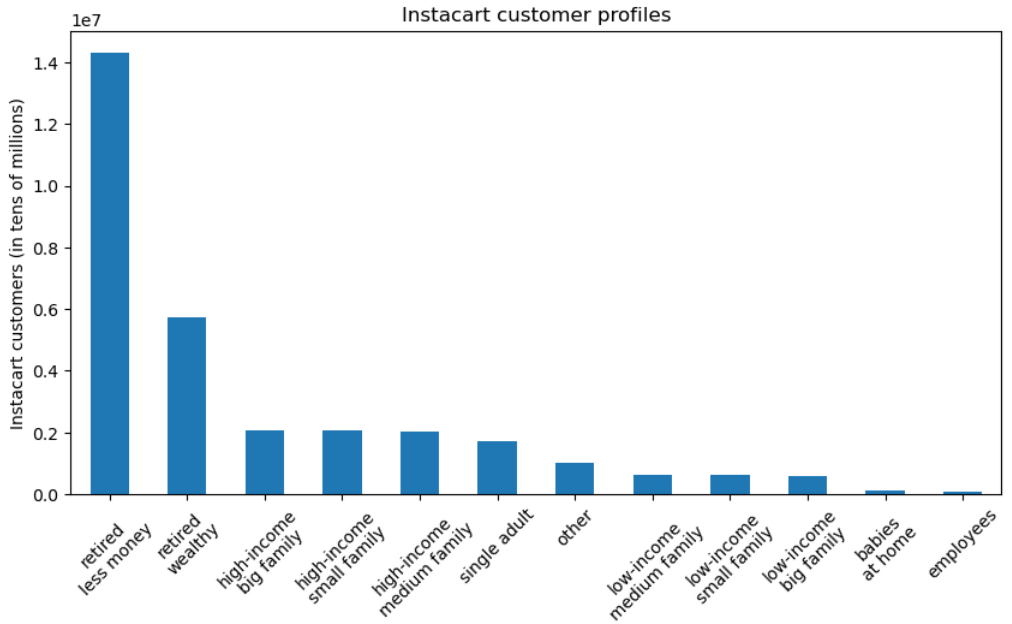


Customer habits

Customer habits

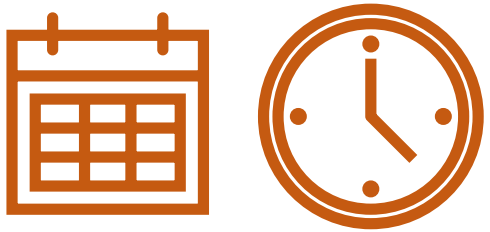
Loyal customers buy every six days and **Regular customers** every 12.

The most **popular customers** are retired people.



Recommendations

Report here



01

02

03

Target market

- Instacart could advertise more heavily on weekends to increase sales.
- Special offers strategy could be used during the week.

Times

- Invest more in marketing strategies that are focused on the busiest hours of the day (9 am to 5 pm)

Specialise

- Instacart should continue investing in products of high revenue, as in the case of the Mid and Low-range.