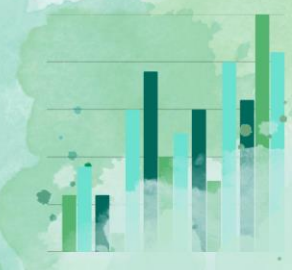
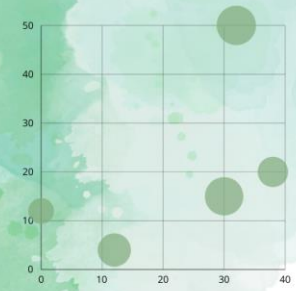
Instacart



Project Goals

Instacart is an online grocery store that operates through an app. Even though their sales are going well, they want clear insights in their customer behavior and sales patters.

My task is to perform an initial data and exploratory analysis of their orders, products and customers and suggest marketing strategies by answering key question regarding their price ranges, customer profiling and advertising. Both the process and answers will be included in a final excel report.



Data that I used

Python: final merged dataset:
32404859 rows, 32 columns
Regions: United States, also devided in Midwest, Northeast, South, West
Other data: customer characteristics, order details, department details and product information
Created: coding, merged datasets, derived and aggregated columns and visuals, final excel report
Data Sources: Dataset 2017, Data Dictionary, CareerFoundry Customers Datset

Skills that I applied

Python, Anaconda, Jupyter Notebook and the following libraries: pandas, numpy, os, matplotlib.pyplot, seaborn and scipy
Data wrangling and subsetting
Data cleaning
Combining and exporting data
Deriving new variables
Grouping data
Aggregating variables
Visualization in Python
Coding etiquette
Reporting in Excel
Population flows

instacart

The process

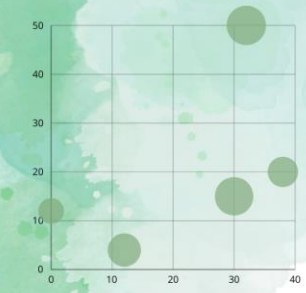
Customer profiling: children

```
# creating category with/without children:
customer_profiling_merged.loc[customer_profiling_merged['children'] == 0, 'children_group'] = 'without children'
```

```
customer_profiling_merged.loc[customer_profiling_merged['children'] >= 1, 'children_group'] = 'with children'
```

As this is my first project working in Python, I first learned about the programming languages Python and R and their characteristics and advantages, followed by setting up the Python coding environment using Anaconda. From here I could launch Jupyter Notebook and add the necessary Python libraries. I learned about the different data types and which queries to use to clean the data. I enjoy using df.describe() from pandas, to see all the important descriptive statistics that give me a strong overview of a dataset and to be able to spot irregularities.

From Instacart I was given datasets about the products, the departments, the orders and the customers of Instacart. To maximize the use of data I merged the orders and products datasets, so I was able to create useful new columns for my analysis. With the use of if statements in combination with loc or For-Loops I could make a segmentation in prices and discover the busiest days and hours. I aggregated the orders per customer using groupby and transform, which resulted in getting the max orders per customer and their loyalty and spending level and how frequently they place orders.
With this information I could merge it together with the customers and departments dataset, to create different customer profiles.




```
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 np.nan
```

I also created the visualizations of the derived columns in Python, using Seaborn, Matplotlib and Scipy. Because of their clear insights in customer behaviour, they provided clear answers on the key questions for the marketing strategy. I completed the project by adding the visualizations and answers in a final excel report, with a detailed explanation of every step. Including data citation, population flow, consistency checks, data wrangling and colum derivation

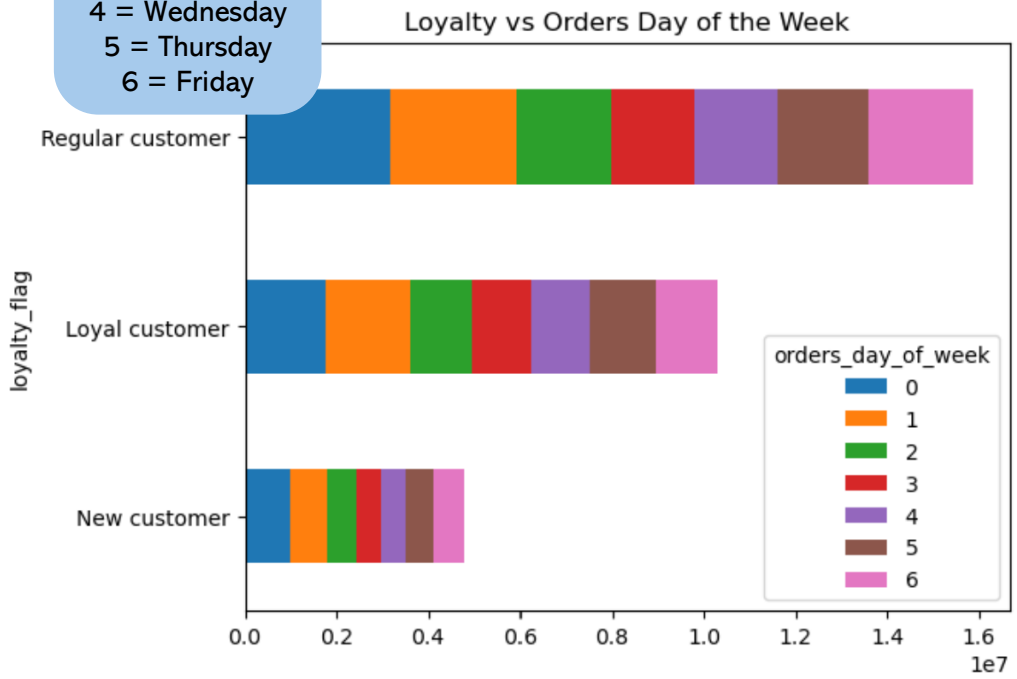
instacart

Analysis

Instacart has a healthy number of loyal customers and enough new customers coming to the platform.

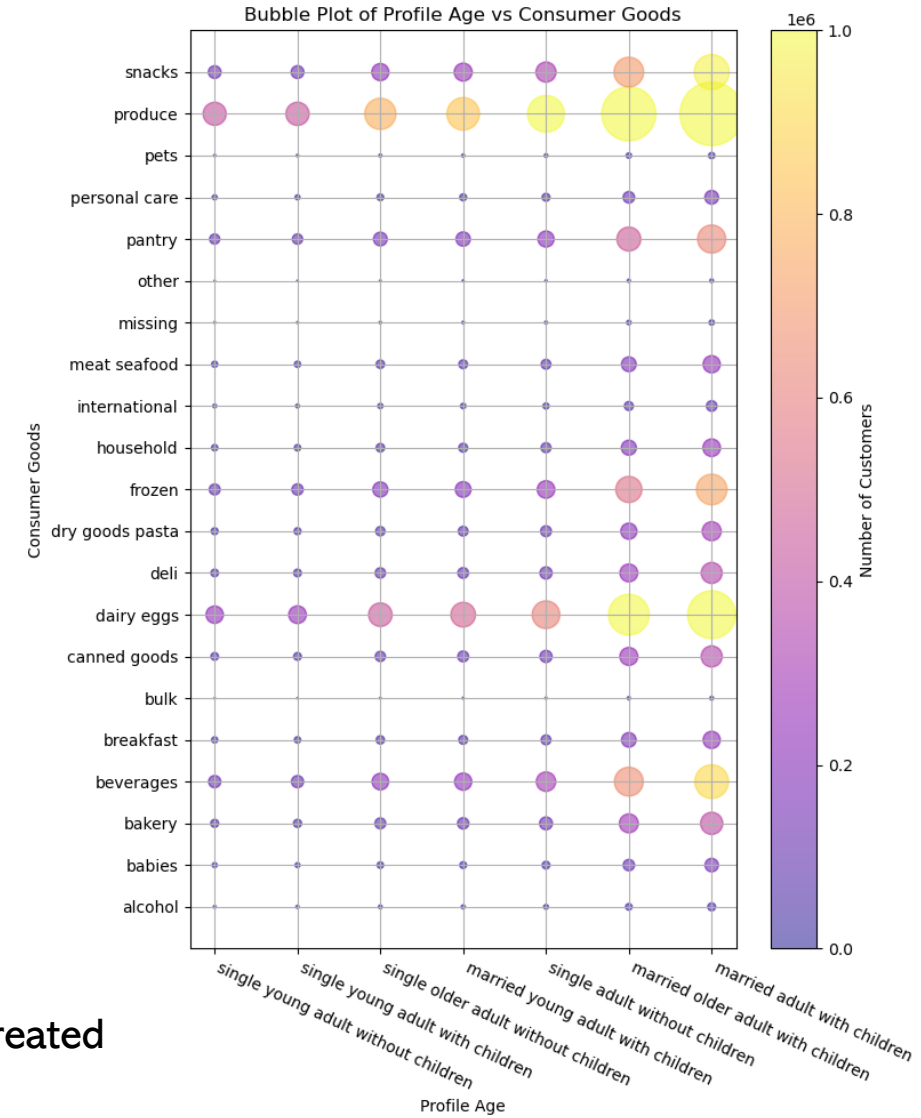
The most popular days for placing orders are Saturday, Sunday and Friday.
The most popular hours are between 9am and 4pm.

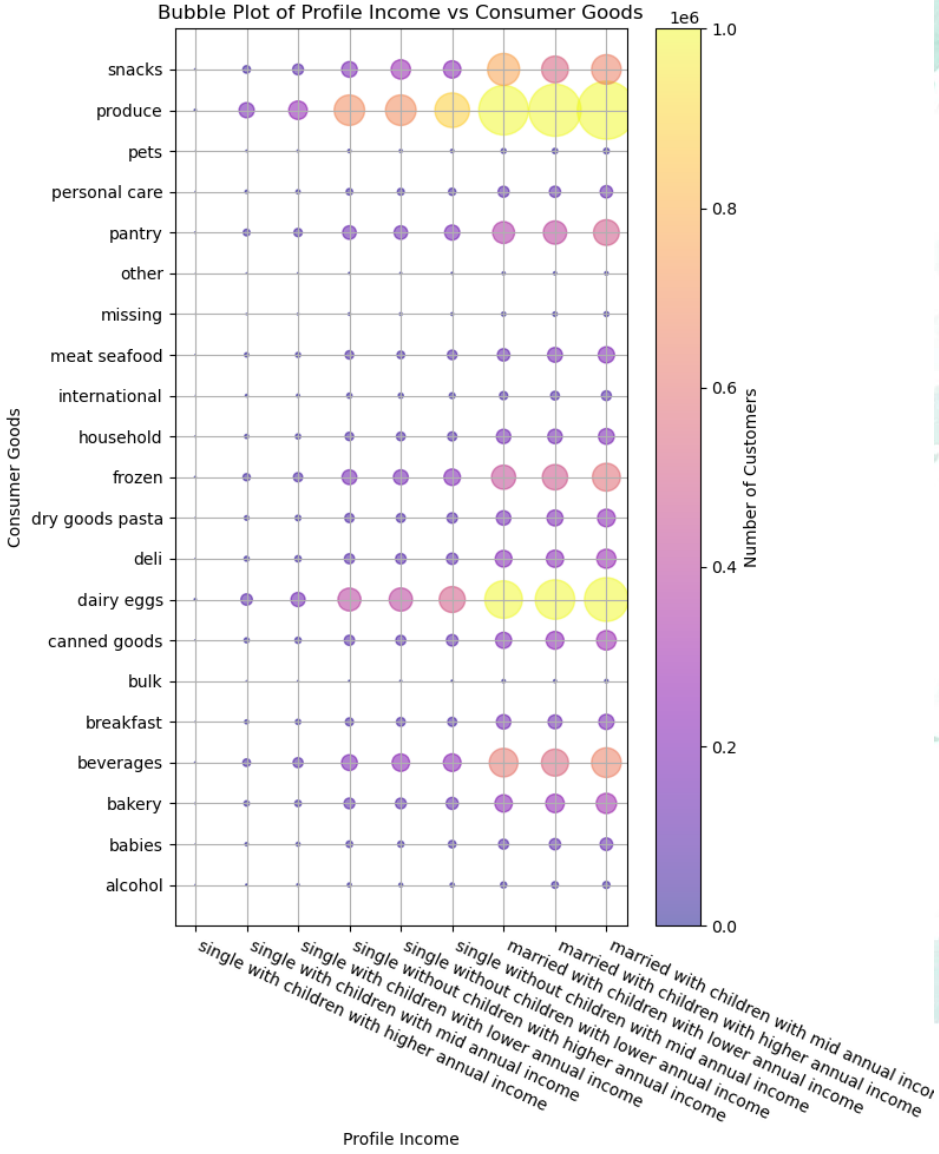
0 = Saturday
1 = Sunday
2 = Monday
3 = Tuesday
4 = Wednesday
5 = Thursday
6 = Friday



instacart

Analysis





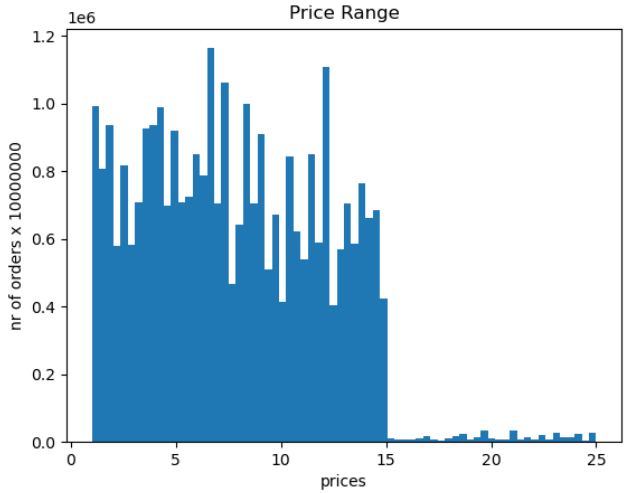
With the customer profiles I created based on age and income,
I was able to explore what the biggest target groups are, which products they buy, which regions they live in and on which days they order the most.

instacart

Recommendation

I recommend scheduling more ads on Monday until Thursday, with most ads on Tuesday and Wednesday as these are the least popular days to place an order. The most interesting hours to increase the ads are the hours between 5pm and 10pm as these are still considered to be regular hours that customers are awake.

For products of $15,- and higher, I would recommend to use this as a separate price range that you don't always have to compare with the basic price range. The number of orders for prices until $15,- are quite balanced. Therefore you can make equal price ranges.
As an extra recommendation, I would avoid round prices. The data shows that a product priced at 11,99 has almost three times more orders than the ones priced at 12,-.

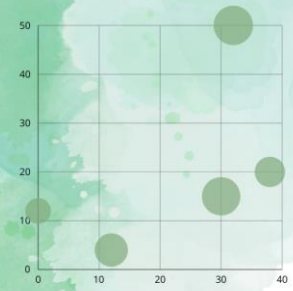


To increase the number of loyal customers, starting a loyalty program could be considered. For example giving discounts when they introduce new customers.

I would suggest to conduct more research on the customers from the customer profiles age and income that are less active on the platform., based on the results from the visualizations. Could for example promotions on consumer goods they often buy increase their overall orders? Could distribution in less active regions be improved and increase the amount and loyalty of customers?

instacart

Personal evaluation



Successes

In the histogram of price ranges, I quickly spotted that products with round prices are less often sold then products of for example ,99. Such a small change can make a big sales impact.

I needed some time so decide how to compare the customer profiles with the consumer goods, as both consist of a lot of possibilities. I did not learn how to make a bubble plot in Python, but I am happy I took the extra time to search online how to make it and I learned a lot from it.

Challenges and lessons learned

Once I started merging datasets together and the size became big, I was having a lot of trouble with saving changes. By researching the internet, I learned that next to creating lists, samples or subsets, I could easily change the size of the integer columns.

At the beginning I found it difficult to decide which customer profiles to create. Then I started to use Python source files where I could easily sketch different possibilities. As this type of file is a blank page, but where you can still see the different colors in the query, it also helped me Putting each query below each other and be sure they would all be correct, as both customer profiles exists of at least 9 options.

instacart