Project 4
Instacart
Grocery Basket
Analysis

INSTACART GROCERY BASKET ANALYSIS

Analysis of Sales Pattern and Customer Segmentation

TOOL : JUPYTER

LANGUAGE : PYTHON

Instacart, 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.

Project Goal : To derived insights and suggest strategies for better customer segmentation based on the provided criteria.

Skills : Data wrangling & Subsetting , Data consistency checks, Combining & Exporting Data ,Deriving new variables, Grouping Data and Aggregation, & Data Visualization.

Data Source : open-source data sets from Instacart from year 2017
The consumer data and the prices of the products were both fabricated for learning purposes. Some of the datasets contain over 32M observations.

In order to draw insights, as expected by Instacart stakeholders, transformation procedures like deriving new columns using if-statements, loc and for-loops functions, as well as grouping and aggregating methods were applied.

Ques 3 The Instacart officers are interested in comparing customer behavior in different geographic areas. Create a regional segmentation of the data.

```
In [25]:  ▶  # Creating filter using else/if for each region

          region = []

          for state in allcomb_df1['state']:
              if (state == 'Maine') or (state == 'New Hampshire') or (state == 'Vermont') or (state == 'Massachusetts') or (state == 'R
                  region.append('Northeast')
              elif (state == 'Idaho') or (state == 'Montana') or (state == 'Wyoming') or (state == 'Nevada') or (state == 'Utah') or (s
                  region.append('West')
              elif (state == 'Wisconsin') or (state == 'Michigan') or (state == 'Illinois') or (state == 'Indiana') or (state == 'Ohio'
                  region.append('Midwest')
              elif (state == 'Delaware') or (state == 'Maryland') or (state == 'District of Columbia') or (state == 'Virginia') or (sta
                  region.append('South')
              else:
                  region.append('Stateless')
```

Fig 1 : Regional Segmentation of Customer Data using else/if

6.The team now wants to target different types of spenders in their marketing campaigns. This can be achieved by looking at the prices of the items people are buying. Create a spending flag for each user based on the average price across all their orders using the following criteria:

If the mean of the prices of products purchased by a user is lower than 10, then flag them as a "Low spender." If the mean of the prices of products purchased by a user is higher than or equal to 10, then flag them as a "High spender."

```
In [25]:  ▶  # Aggregating data with agg()
             # Calculating the mean,min and max of prices col grouped by user_id
             ords_prods_merge.groupby('user_id').agg({'prices': ['mean', 'min', 'max']}).head(10)
```

Out[25]:

		prices	
user_id	mean	min	max
1	6.367797	1.0	14.0
2	7.515897	1.3	14.8
3	8.197727	1.3	14.4
4	8.205556	1.4	14.6
5	9.189189	3.2	14.8
6	8.471429	1.8	19.6

Fig 2 : Aggregating Price data to target different kind of spenders in their marketing campaigns

Finally, for **Data visualizations** of my analysis, I used bar charts, horizontally and stacked bar charts, pie charts, line charts, histograms, and scatterplots and created them in **Jupyter**.
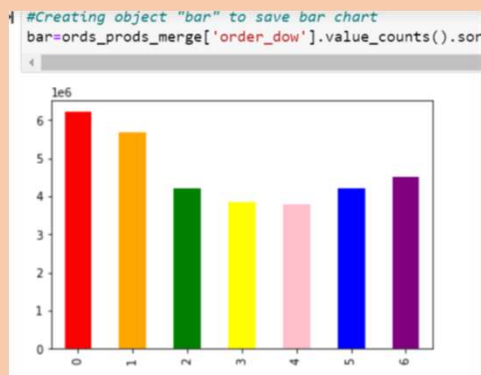

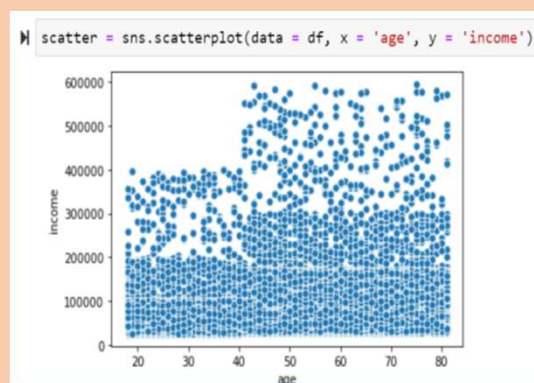Fig 1 : Created order frequency bar chart for column day of week

Fig 2 : Created Scatter Plot to show relationship between Income & Age

Fig 3 : Created Pie Chart to show the customer mix by Loyalty

Finally, the entire analysis process was documented in form of Excel Reporting that contains population flow, describes all wrangling and deriving operations, shows visualizations of results along with recommendations for the new marketing strategies.
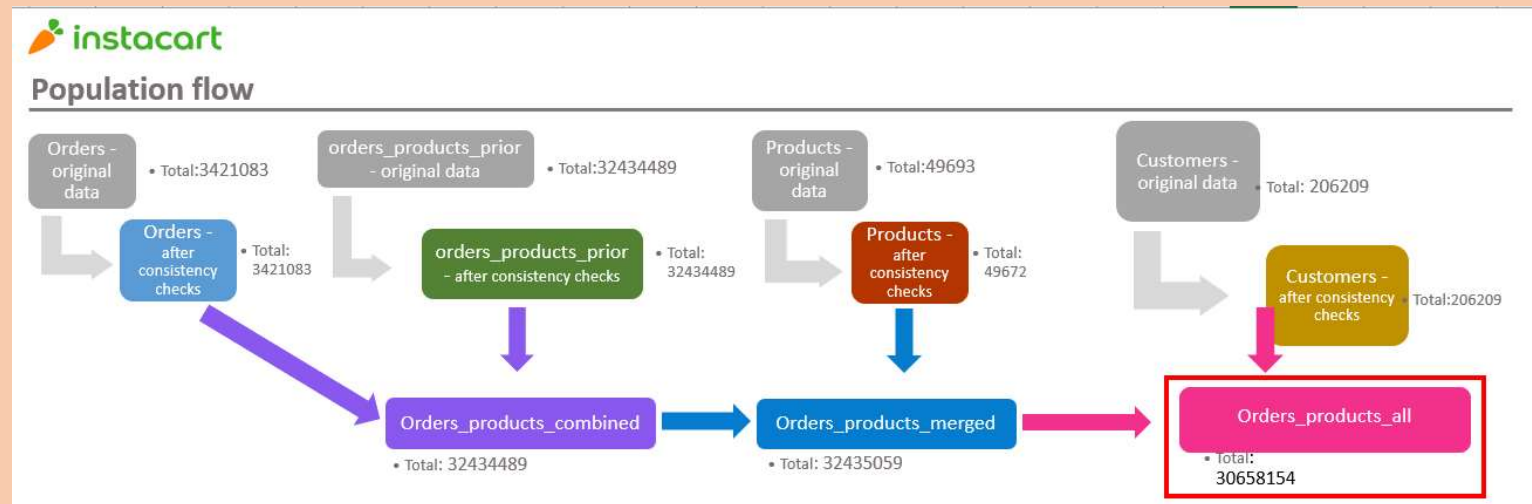

Fig : Population flow for the Instacart Project

Instacart Grocery Basket Analysis

Recommendations

- Saturday and Sunday are the busiest days, and 10 a.m to 3 p.m are the busiest hours. So the slowest days and slowest hours are perfect for targeting advertisements and promotions (upcoming) so that families can plan ahead for weekend deals. Slowest days can be celebrated for special weekday deals.

- Focusing on struggling departments is going to increase the loyalty customer base. Deals and special days can be planned for those struggling departments after discussing with vendors and that should definitely target the specific customer set and give a sales boost.

- 33.2 % customers are loyal customers and 51.3% as our regular customers. To erase this thin line between Regular and Loyal customers , We need to reward our regular customers with some reward points or may be a mail coupon for their next purchase, to convert them to loyal customers and increase our loyalty base

Please find the full excel reporting , project scripts with data link here