Project Title: Wholesale customer purchasing behavior

Problem statement

This project seek to use self-organizing map to identify wholesale customer purchasing behavior.

Aims and Objectives

The aim of this analysis:

1. Identify customer segments (e.g., bulk buyers vs. low spenders).
2. Optimize marketing efforts.
3. Tailor product offerings to specific customer needs.

This project is set to use the self-organizing map to clusters high volume spending customers and less purchased products to enable the marketers to strategize.

Dataset Overview

The dataset used for this project is from UCI machine learning repository (Wholesale customers - UCI Machine Learning Repository)

Its features includes Annual spending on various products (e.g., milk, grocery, frozen foods).The data set is multivariate with 440 rows and 8 columns.

It includes the annual spending in monetary units (m.u.) on diverse product categories.

Tools used:

- Pandas
- Matplotlib
- Seaborn
- Sklearn
- Self-organizing map
- Numpy
- Jupyter notebook

Methodology

This project adapted the quantitative approach to data analysis in processing and analyzing data, which includes:

Data cleaning:

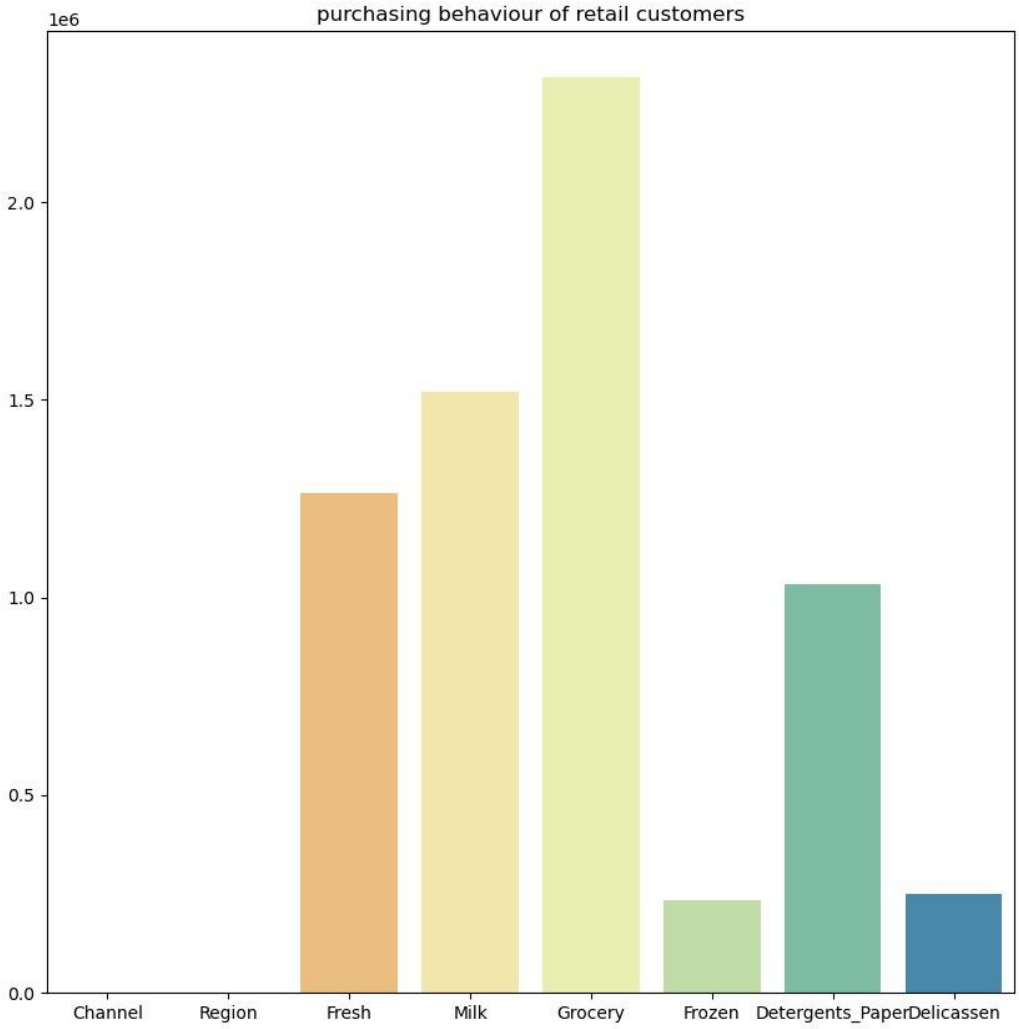
No null values were found and no duplicated values detected, the dataset was clean and the original data size was maintained.

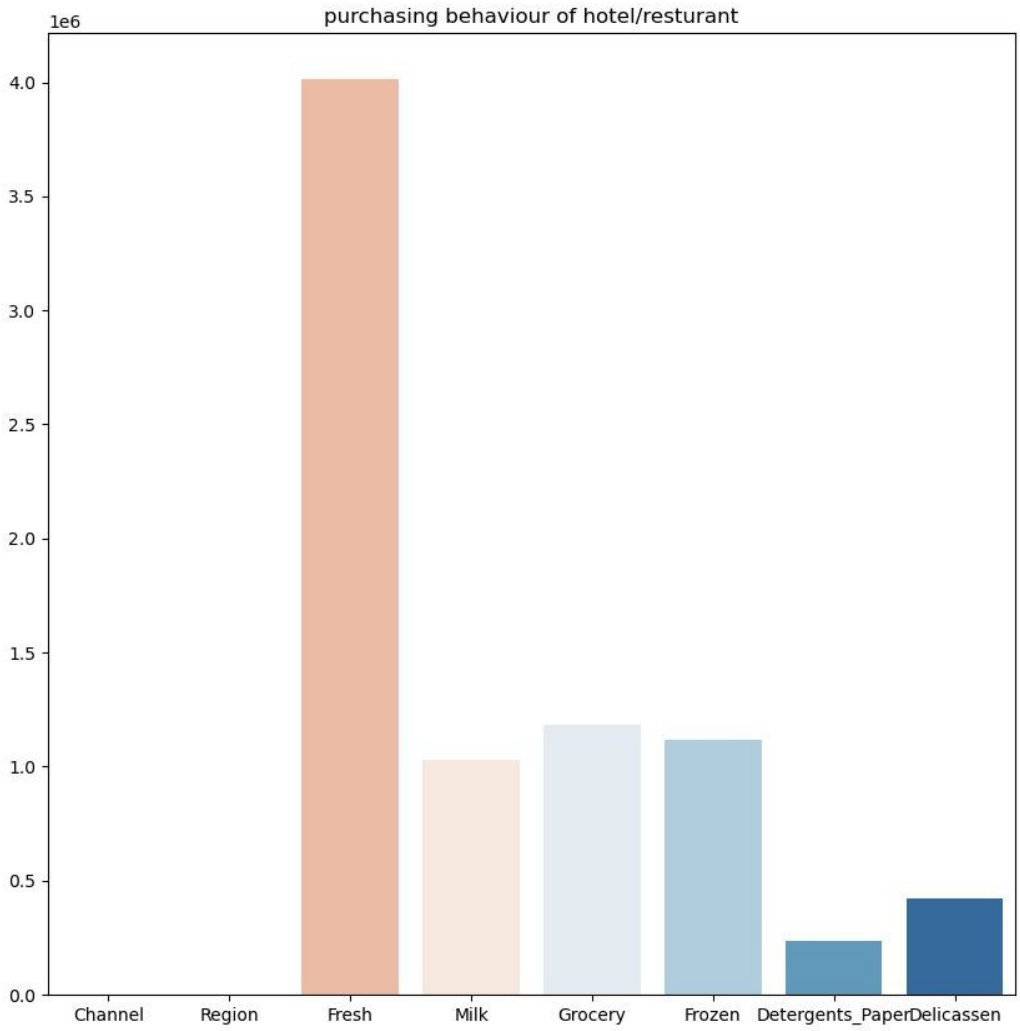
No further cleaning was required.

Exploratory Data Analysis (EDA)

1. **Determining the High spender and low spenders among channels**

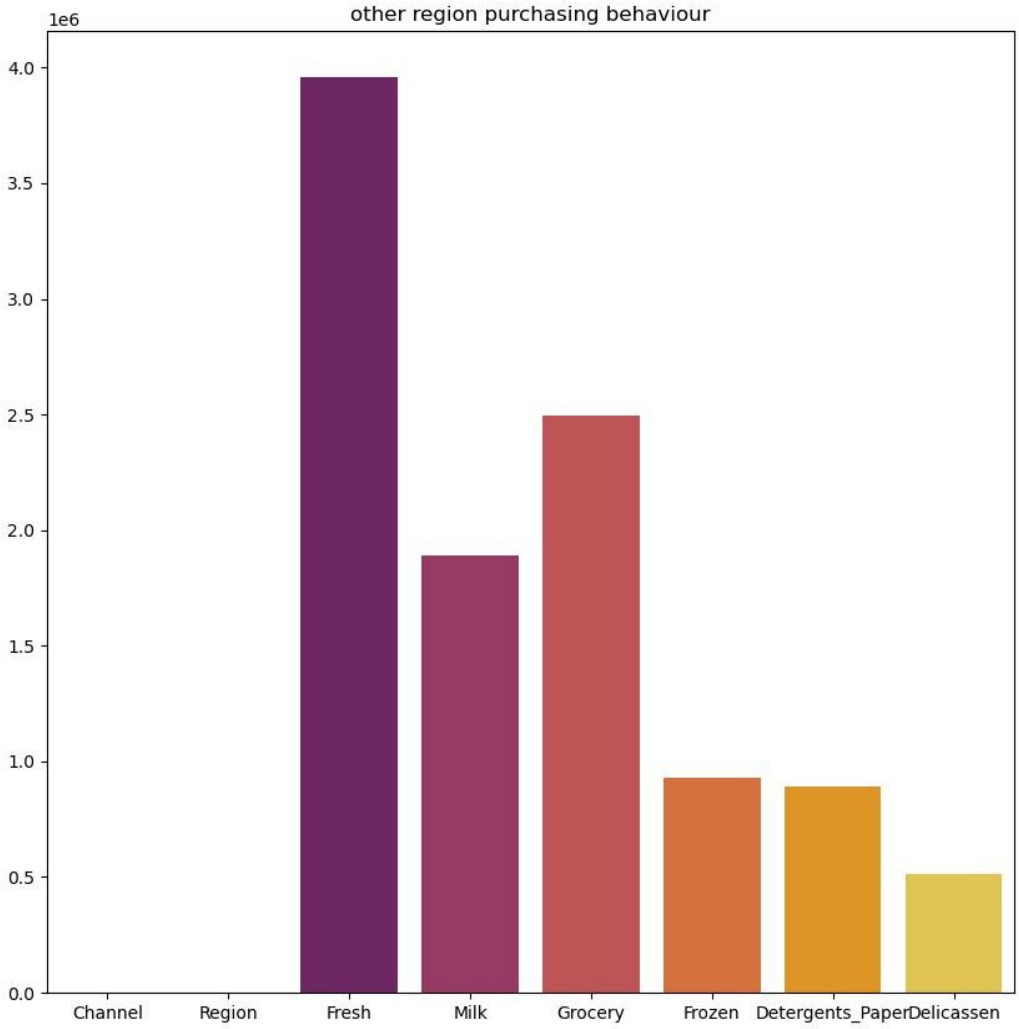
   The bar plot tells us, that the both customer base (hotel and retail customers) have very low spending on item Delicassen. Retailers/individuals are the customers with high purchasing power as they spend more while Hotel/restaurant purchases only frozen foods in a very high quantity. The retails customers buy grocery in a very large quantity and verylittle of frozen foods and other products moderately as shown in the chart below.

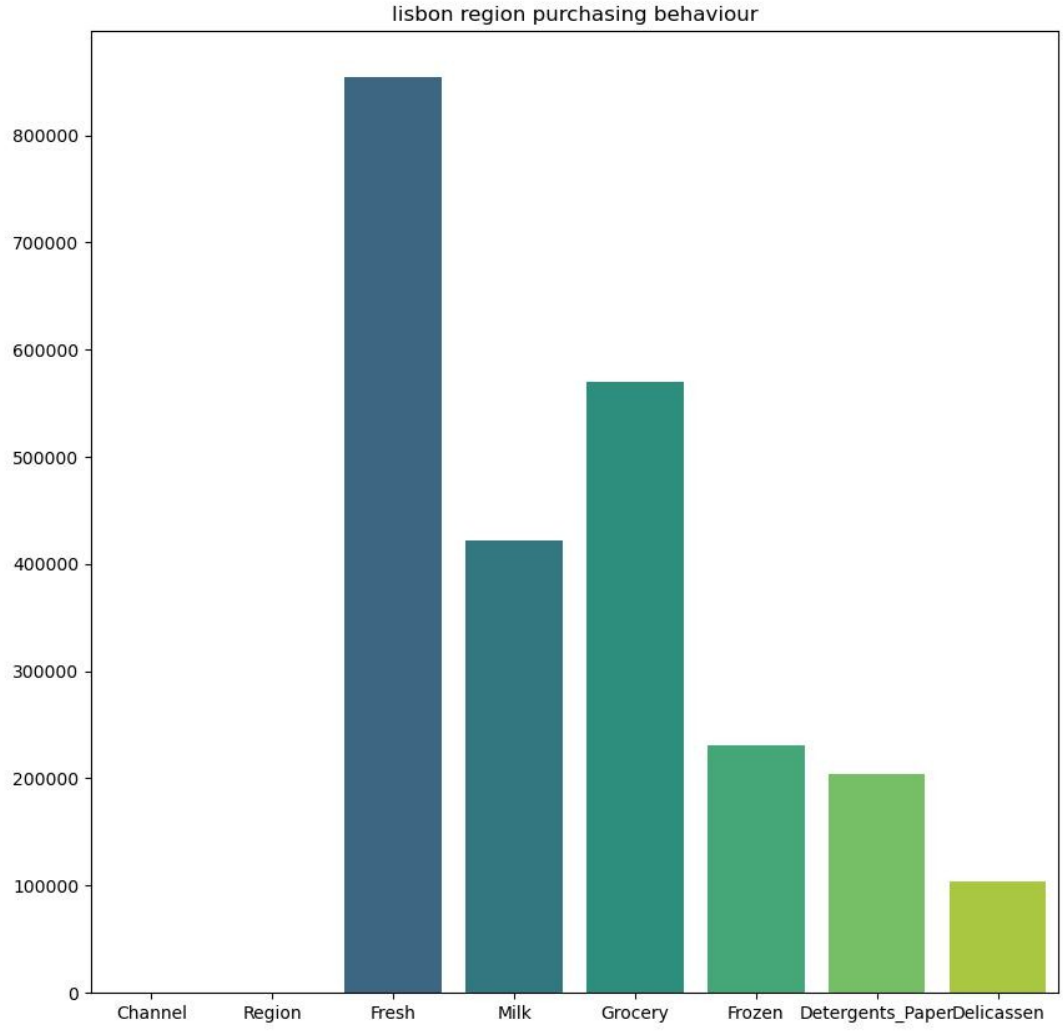

purchasing behaviour of retail customers
1e6
2.0
1.5
1.0
0.5
0.0
Channel
Region
Fresh
Milk
Grocery
Frozen
Detergents_Paper
Delicassen

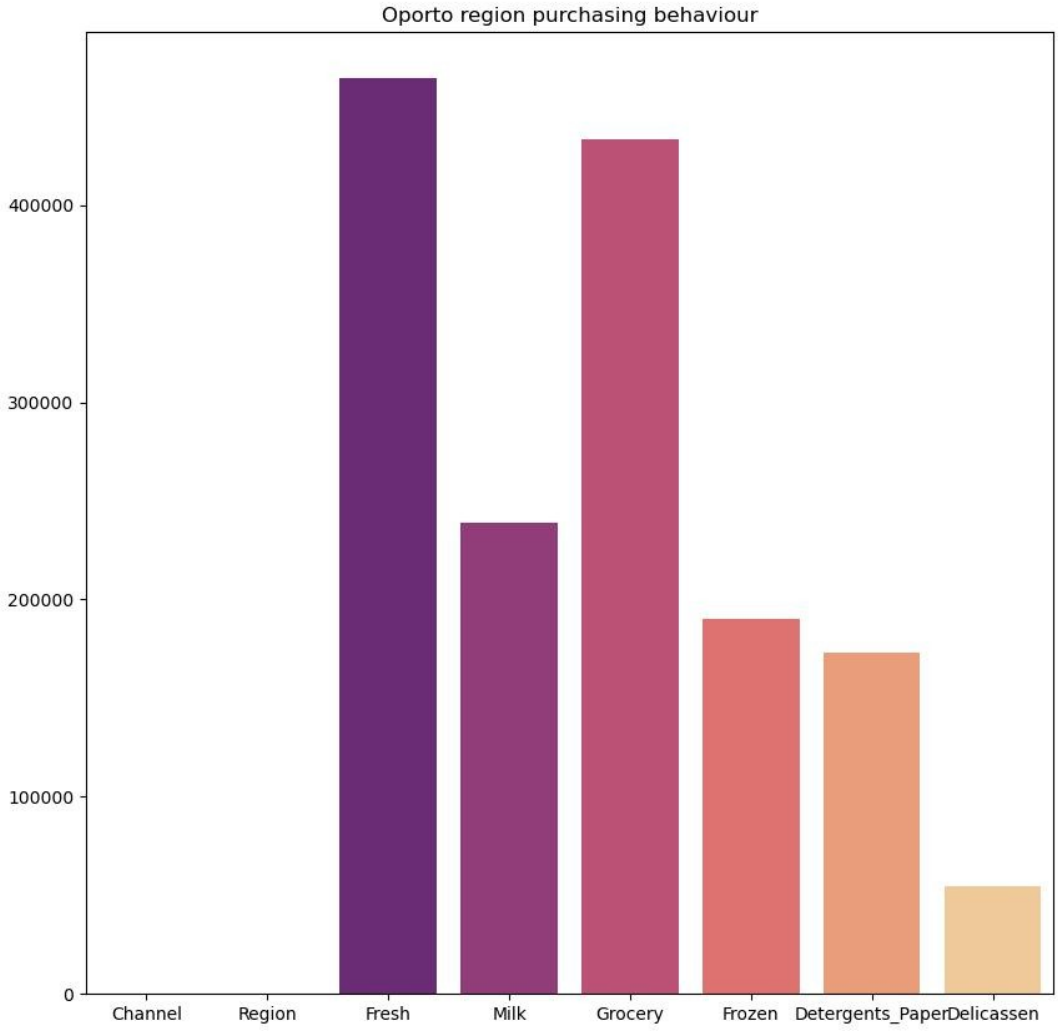


2. Identifying the purchasing behavior of customers among regions

From the chart below, all regions purchases fresh food and grocery in high quantity but very little of Delicassen items. Products like frozen foods, detergent and delicassen are in very low quantity across all regions.


other region purchasing behaviour
1e6
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Channel
Region
Fresh
Milk
Grocery
Frozen
Detergents_Paper
Delicassen


lisbon region purchasing behaviour
800000
700000
600000
500000
400000
300000
200000
100000
0
Channel
Region
Fresh
Milk
Grocery
Frozen
Detergents_Paper
Delicassen



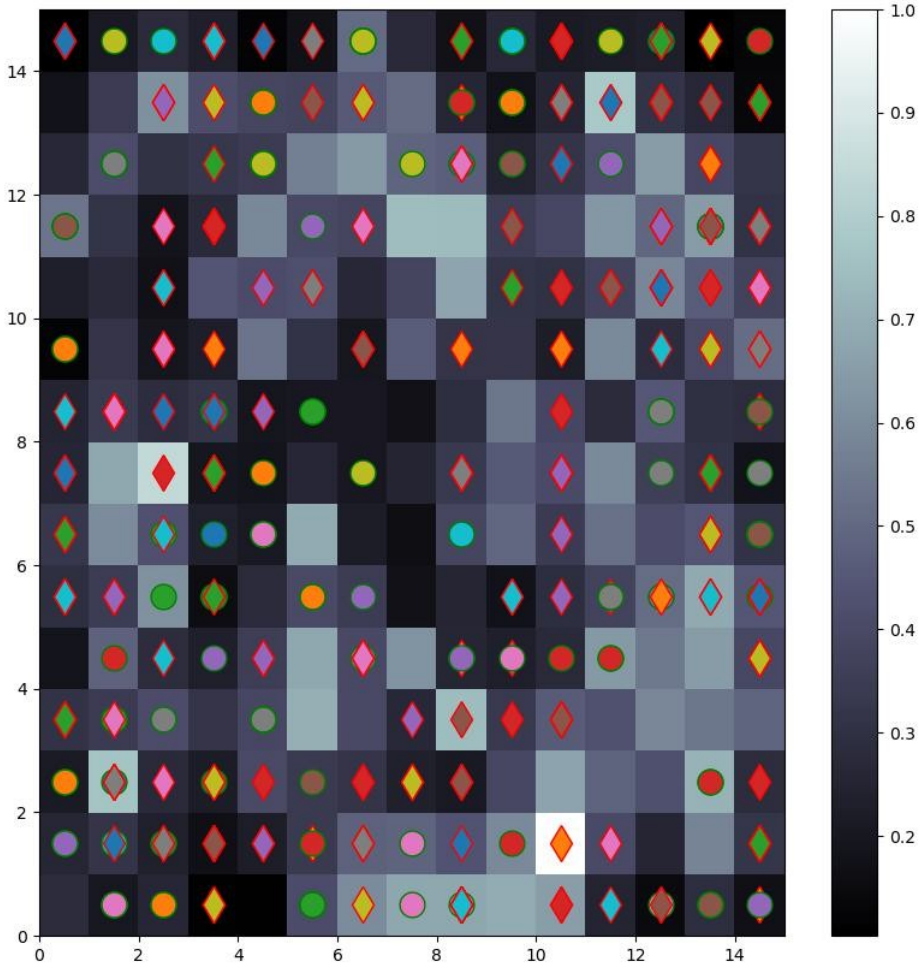
All charts were created with matplotlib and seaborn libraries in python.

Model building

I selected features for clustering into x and y, went on to scale them using MinMaxScaler from sklearn library. I initialize the model to input the size of the model chart, the input length which is the same number as my x features.

I Initialize the weights with this code som.random_weight_init (data=x,num_iteration=100) and then train the model.

After training, I visualize the result as shown below



Map interpretation

The ‘o’ marker on the chart represent the hotel customers and the ‘D’, the diamond shape represent retailers.

From the chart above, it says that the retailers are the outliers, in other words, they are the highest spender.

Clusters interpretation

I went ahead map out the values of customers in the white boxes called clusters with this code ‘mappings=som.win_map (x_scaled)’.

Findings:

1. Both clusters spends very low on Detergent papers.

2. There are 3 regions but only Oporto and other regions spends more.

Recommendations:

1. I suggest the marketing team to identify why customers don't purchase other products may be through questionnaire and suggestion box.

2. The marketers should focus on advertising these low purchase products such as detergent paper, delicassen especially to hotels/ restaurants in oporto and lisbon regions.

3. To communicate product benefits to customers through emails.

4. Include bonuses and incentives on those products.