wholesale-customer

November 17, 2024

1 Analysis by Prisca

[ ]:

1.1 Description: Data about wholesale customer purchasing behavior.

1.2 Features: Annual spending on various products (e.g., milk, grocery, frozen foods).

1.3 The data set Multivariate refers to clients of a wholesale distributor. It includes the annual spending in monetary units (m.u.) on diverse product categories

1.4 Dataset Source: Available on the UCI Machine Learning Repository.

1.5 i will be using clustering and self organizing map to analyze this data

[ ]:

1.6 columns explainations:

1.7 Region Represents the geographic region of the customer.

1.7.1 1: Lisbon.

1.7.2 2: Oporto.

1.7.3 3: Other regions.

1.8 Channel Represents the type of customer.

1.8.1 1: Hotel/Restaurant/Cafe (HoReCa).

1.8.2 2: Retail (e.g., shops or individual customers

1.9 Fresh:The annual spending (in monetary units) on fresh products, such as vegetables, fruits, meat, and seafood,

1.10 Milk:The annual spending (in monetary units) on milk and dairy products.

1.11 Grocery: The annual spending (in monetary units) on grocery items, including non-perishables like canned goods or pantry staples.

1.12 Frozen:The annual spending (in monetary units) on frozen foods, such as ice cream, frozen vegetables, and frozen meals.

1.13 Detergents_Paper:The annual spending (in monetary units) on cleaning products and paper products, such as detergents, tissues, and paper towels.

1.14 Delicassen:The annual spending (in monetary units) on delicatessen items, such as prepared foods, specialty products, or high-end foods often found in deli sections.

[ ]:

1.15 the aim of this analysis

1.15.1 1. Identify customer segments (e.g., bulk buyers vs. low spenders).

1.15.2 2. Optimize marketing efforts.

1.15.3 3. Tailor product offerings to specific customer needs.

```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns  # import libraries
import warnings
warnings.filterwarnings('ignore')
```

```python
df=pd.read_csv(r'C:\Users\USER\Documents\dataset\Wholesale customers data.csv')
#load Dataset
```

```
[50]: df.head()
```

```
[50]:    Channel  Region  Fresh   Milk  Grocery  Frozen  Detergents_Paper  Delicassen
      0        2       3  12669   9656     7561     214              2674        1338
      1        2       3   7057   9810     9568    1762              3293        1776
      2        2       3   6353   8808     7684    2405              3516        7844
      3        1       3  13265   1196     4221    6404               507        1788
      4        2       3  22615   5410     7198    3915              1777        5185
```

```
[160]: df.shape # size of the dataset
```

```
[160]: (440, 8)
```

```
[55]: df.duplicated().sum() # no duplicate values
```

```
[55]: 0
```

2 Exploratory analysis

```
[161]: df_grouping=df.groupby('Channel').mean() # overall filetring
```

```
[57]: df_grouping
```

```
[57]:           Region          Fresh          Milk       Grocery       Frozen  \
      Channel
      1        2.510067   13475.560403   3451.724832   3962.137584  3748.251678
      2        2.612676    8904.323944  10716.500000  16322.852113  1652.612676

              Detergents_Paper   Delicassen
      Channel
      1              790.560403  1415.956376
      2             7269.507042  1753.436620
```
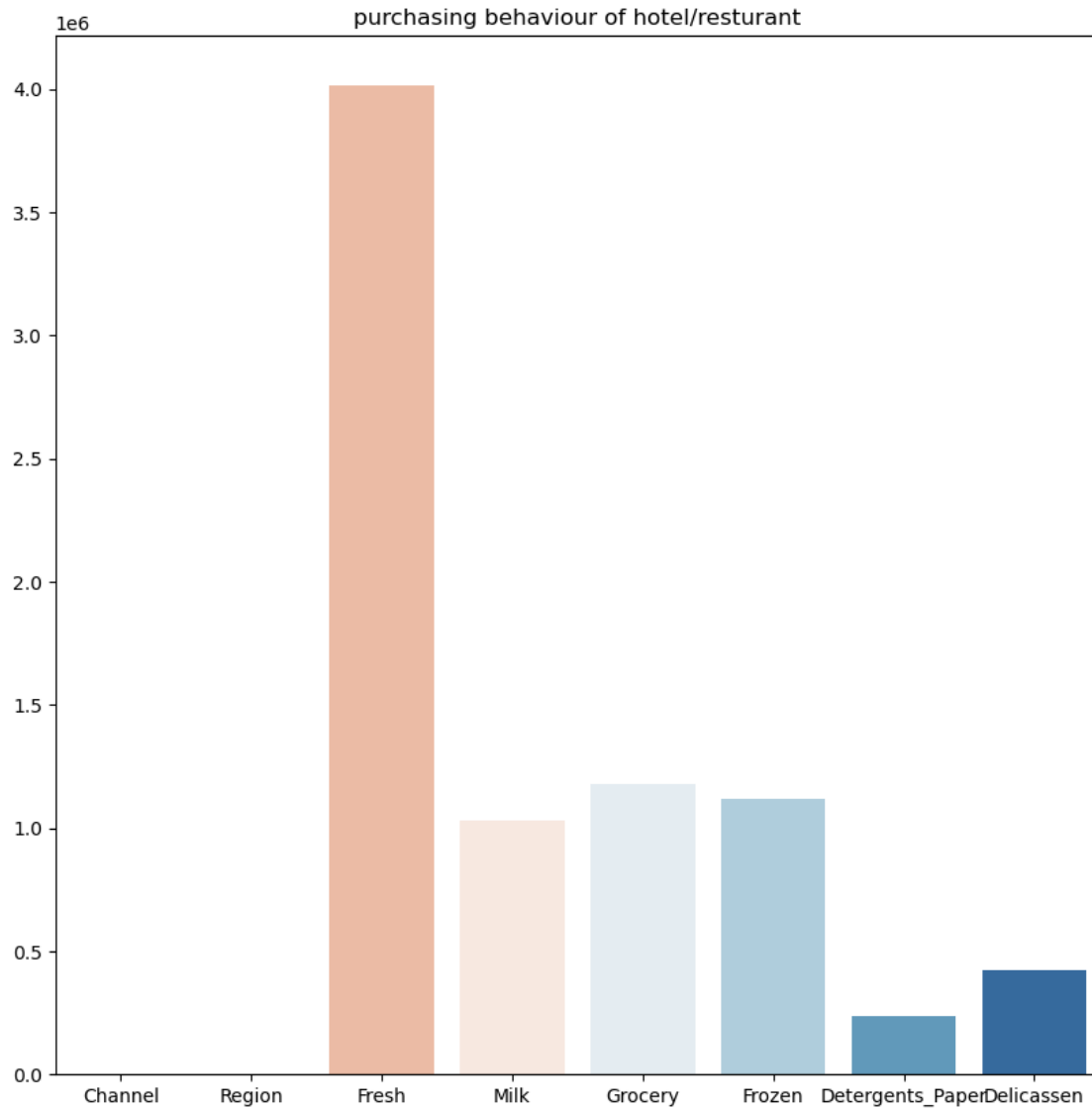
3 High spending vs Low spending

```
[58]: df_1=df[df['Channel']==1].sum()
      plt.figure(figsize=(10,10))
      sns.barplot(df_1,ci=None,palette='RdBu')
      plt.title('purchasing behaviour of hotel/resturant') # hotels are the low␣
      ↪spenders
```
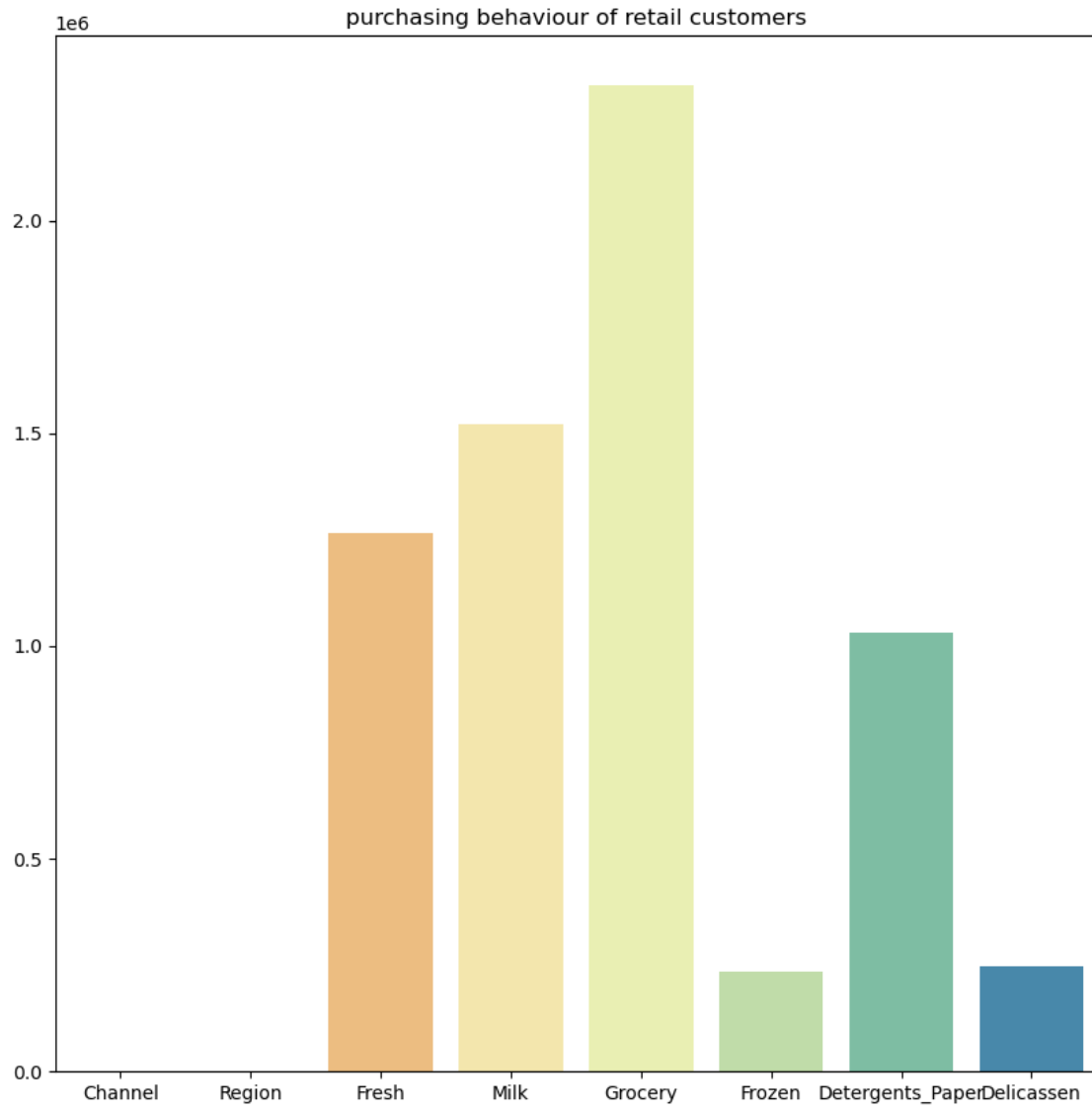
```
[58]: Text(0.5, 1.0, 'purchasing behaviour of hotel/resturant')
```

purchasing behaviour of hotel/resturant

[ ]:

```
[59]: df_2=df[df['Channel']==2].sum()
      plt.figure(figsize=(10,10))
      sns.barplot(df_2,palette='Spectral')
      plt.title('purchasing behaviour of retail customers')  # retails are the high␣
        ↪spenders
```

[59]: Text(0.5, 1.0, 'purchasing behaviour of retail customers')

purchasing behaviour of retail customers

3.1 findings:

3.2 1.the both customer base have very low spending on item Delicassen.

3.3 2. retailers/individuals are the customers with high puchasing power, that is they spend more .

3.4 3. hotel/resturant purchase frozen foods in a very high quantity.

3.5 recommendation:

3.6 1 .the marketing team can niche or narrow their focus on products that the individuals customer barely purchase such includes frozen and Delicassen products, they marketing team can include bonus to these product or implement discounts.

3.7 2.the hotels purchases only frozen foods in high quantity and lowest of Delicassen, the market should put effort to communicate benefits of other products to the hotels by emailing or adverts.

3.8 3.they should also figure out the position of theses products in the minds of the customers by gving out questionaires, this way they can know what and how to communicate the customers.

3.9 conclusion:

3.10 the individuals customer base spend alot in buying products
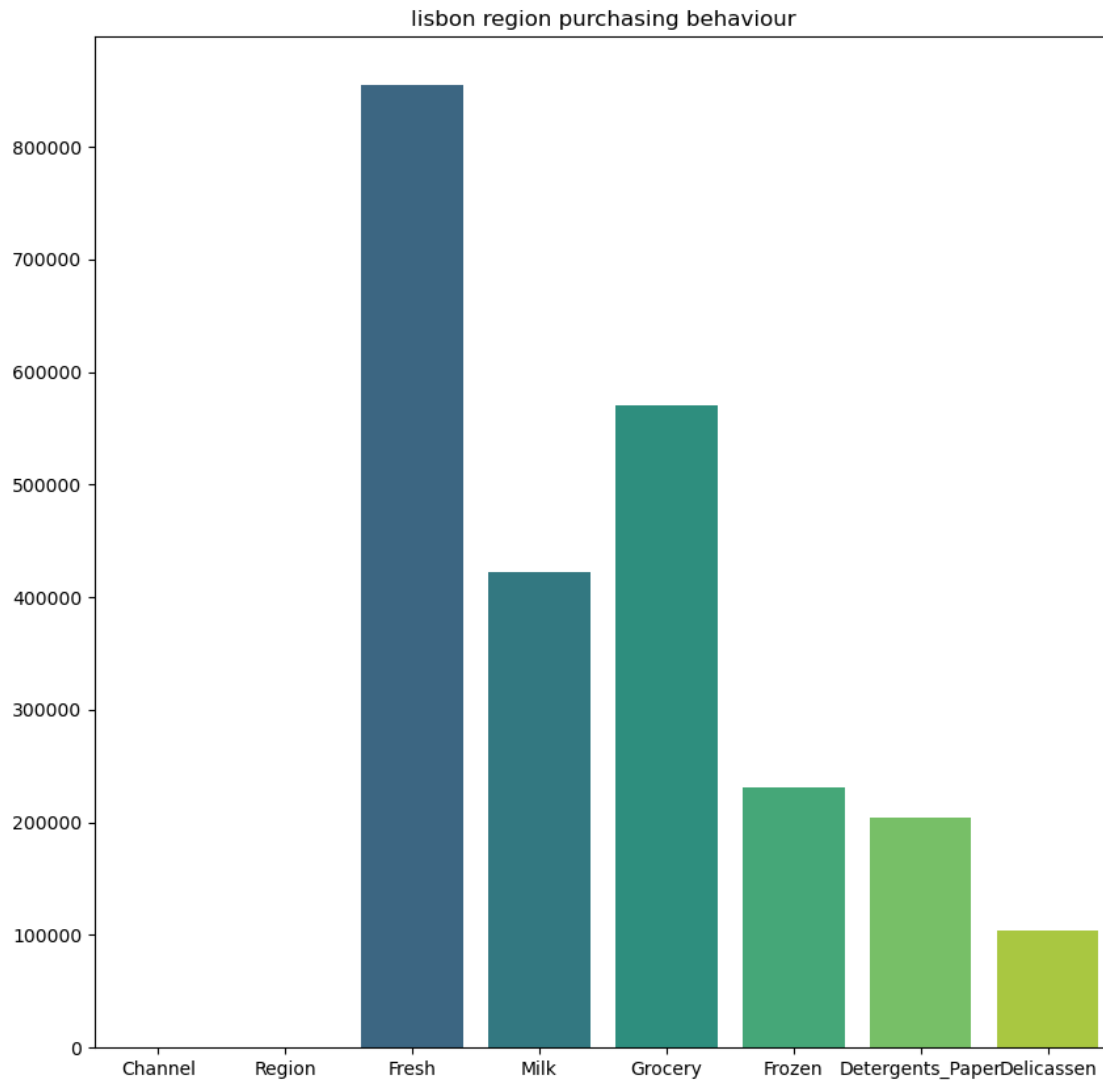
[ ]:

[ ]:

3.11 Regional customer base

```
[162]: region_1=df[df['Region']==1].sum()
       plt.figure(figsize=(10,10))
       sns.barplot(region_1,palette='viridis')
       plt.title('lisbon region purchasing behaviour')
```
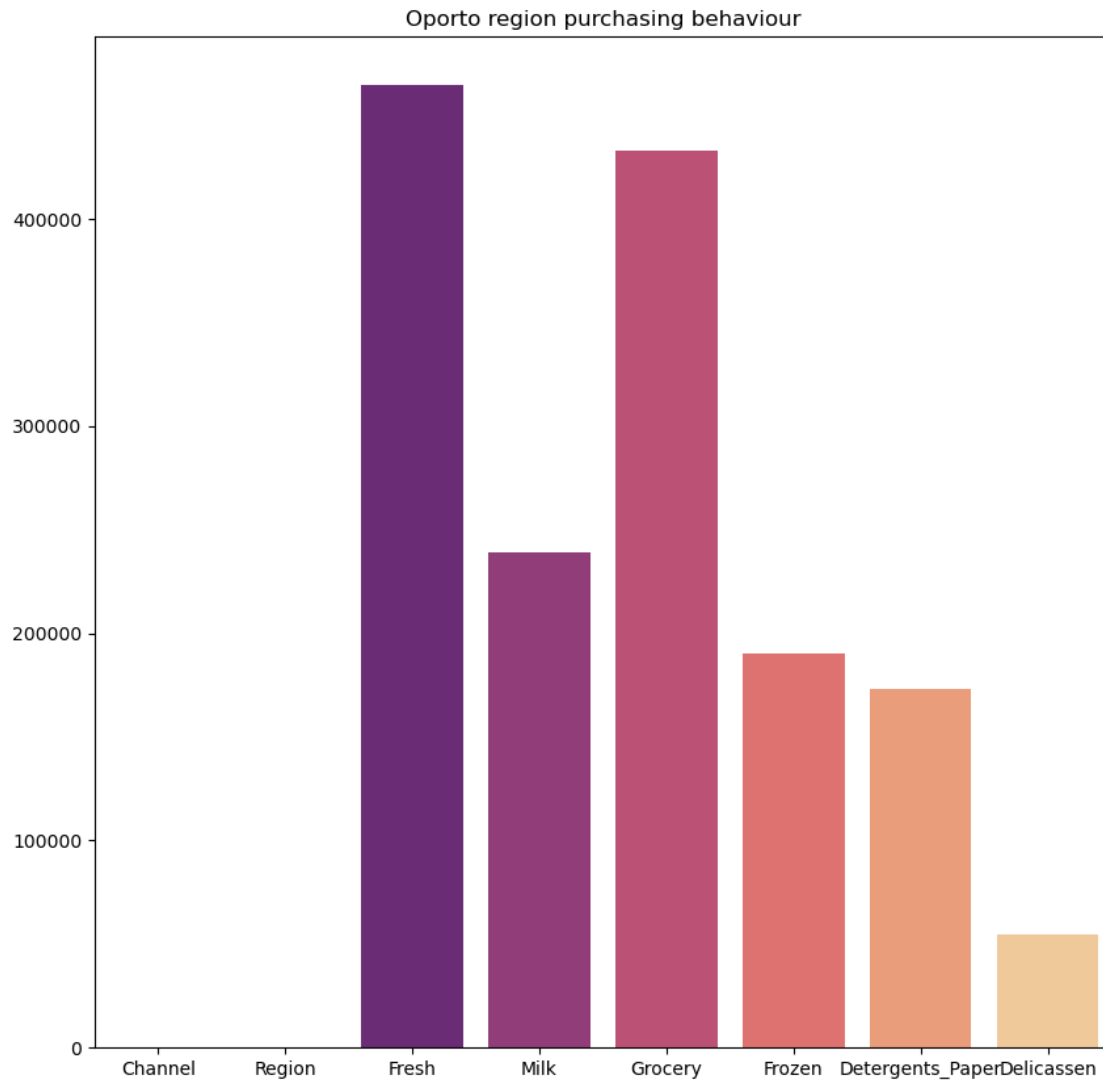
```
[162]: Text(0.5, 1.0, 'lisbon region purchasing behaviour')
```

lisbon region purchasing behaviour

```
region_2=df[df['Region']==2].sum()
plt.figure(figsize=(10,10))
sns.barplot(region_2, palette='magma')
plt.title('Oporto region purchasing behaviour')
```

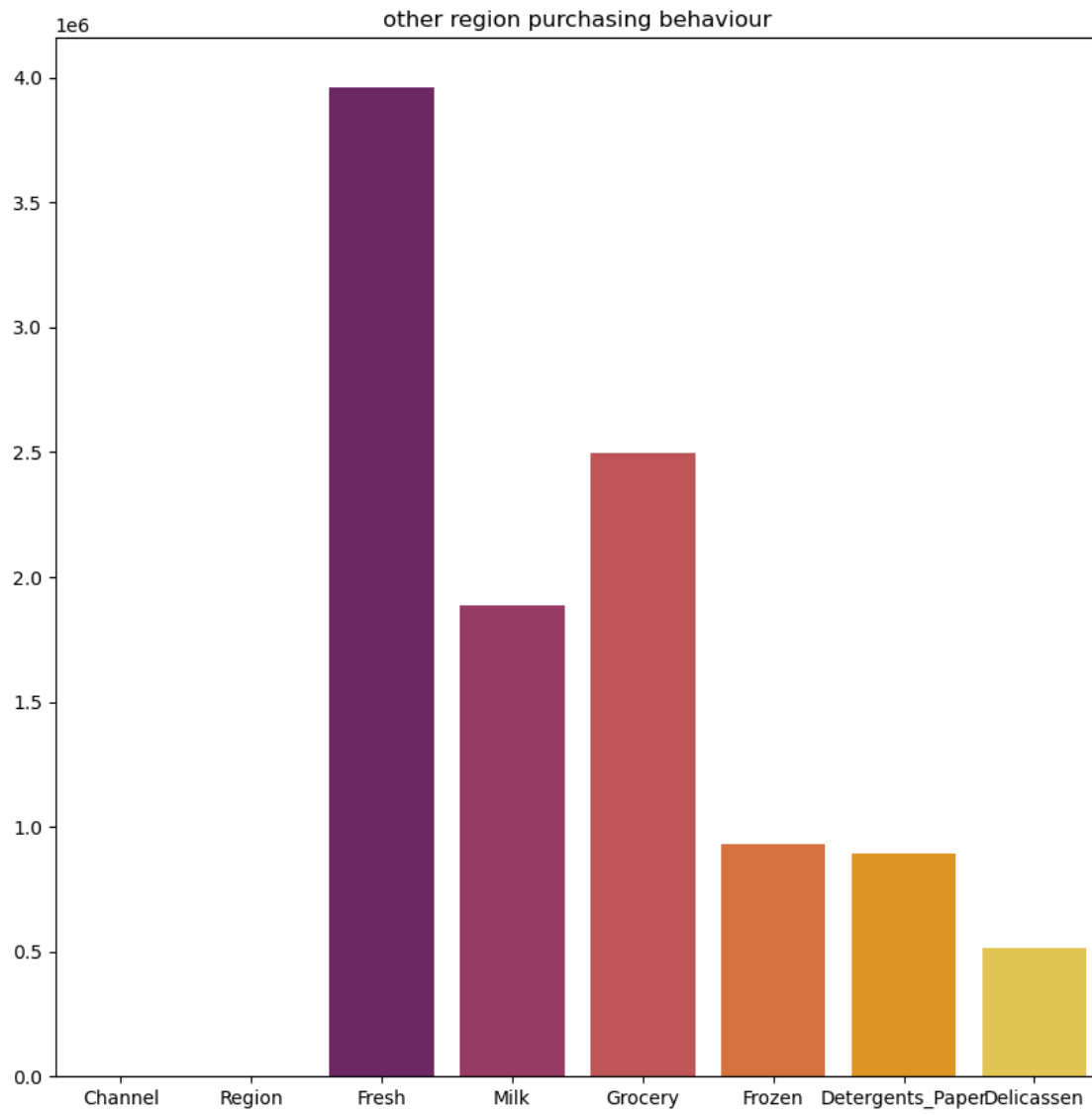[62]: Text(0.5, 1.0, 'Oporto region purchasing behaviour')

```
[ ]:
```

```
[64]: region_3=df[df['Region']==3].sum()
      plt.figure(figsize=(10,10))
      sns.barplot(region_3,palette='inferno')
      plt.title('other region purchasing behaviour')
```

```
[64]: Text(0.5, 1.0, 'other region purchasing behaviour')
```

other region purchasing behaviour

[ ]:

4 findings:

4.1 All regions puchases fresh food and grocery in high quantity but very little of Delicassen items.

4.2 products like frozen foods, detergent and delicassen are in very low quantity accross all regions.

5 Recommendations:

5.1 i suggest that the marketing team to focus their advert on these products with low purchases especially Delicassen.

5.2 communicates these products to customers to emails and advert.

5.3 they can promote the products by implementing discounts, bonuses and many packages.

6 Conclusion

6.1 All regions purchases fresh foods,milk and grocery in high quantity and the rest in very low quantity.

[ ]:

[ ]:

7 clustering using self organizing map

```
[81]: from minisom import MiniSom
      from sklearn.preprocessing import MinMaxScaler
```

```
[152]: x=df[['Region','Fresh', 'Milk', 'Grocery', 'Frozen',
             'Detergents_Paper', 'Delicassen']]
       y=df['Channel'] # feature selection
```

```
[83]: scaler=MinMaxScaler(feature_range=(0,1))
```

```
[153]: x_scaled=scaler.fit_transform(x) #scaling the features
```

```
[154]: som=MiniSom(x=15,y=15,input_len=7,sigma=1.0,learning_rate=0.5) #initializing
       ↪the model
```

```
[86]: # initialize the weights
      som.random_weights_init(x_scaled)
```

```
[87]: # train the model
      som.train_random(data=x_scaled,num_iteration=100)
```

```python
[88]:  # visualization
       from pylab import bone,pcolor,plot,colorbar
```

```python
[89]:  plt.figure(figsize=(10,10))
       bone()
       pcolor(som.distance_map().T)
       colorbar()
       markers=['o','d']
       colors=['g','r']
       for i, xi in enumerate(x_scaled):
           w=som.winner(xi)
           plot( w[0] +0.5, w[1] +0.5, markers[y[i] %
        ↪len(markers)],markeredgecolor=colors[y[i] % len(colors)],
        ↪markersize=15,markerfacecolor=None)
```
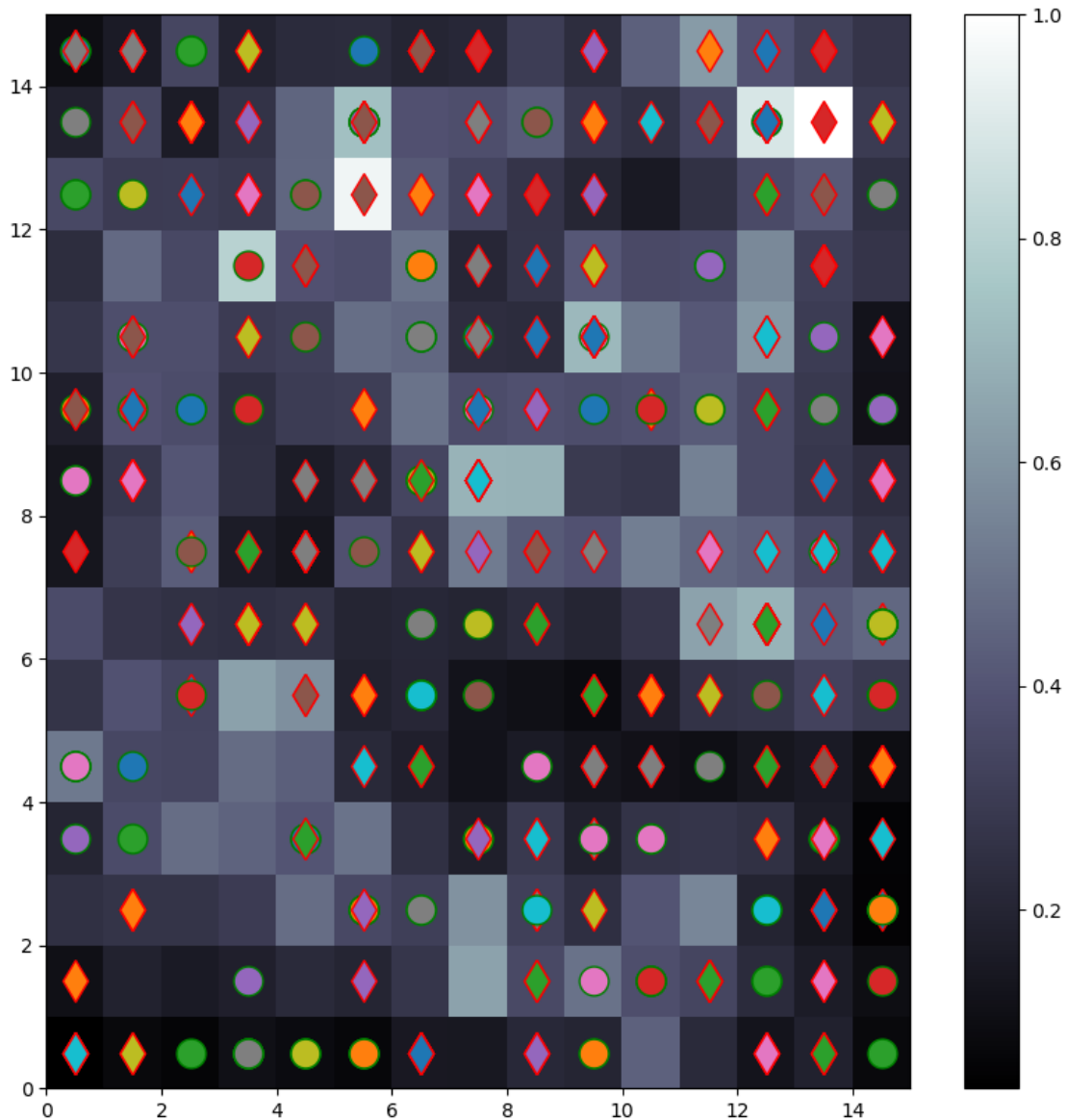
7.1 marker 'o' represent hotel which stands for **1**, and 'd' represent retailers which stands for **2** and this project is to determine the high spenders and on what they spend on.

7.2 from this map, it shows that marker 'D' that is diamond shaped, that is the white box are the obvious spenders.

```
[ ]:
```

```
[139]: mappings=som.win_map(x_scaled) # getting the values inside the cluster D
```

```
[93]: len(mappings)
```

```
[93]: 161
```

```
[155]: # it is grouped into 2 clusters
       spenders1=mappings[(13,13)]
       spenders2=mappings[(5,12)]
```

```
[123]: spender_df=np.concatenate([spenders1,spenders2],axis=0) # combining the mapping␣
       ↪together
```

```
[157]: spender_inv=scaler.inverse_transform(spender_df) # reversing the scaling
```

```
[156]: data=pd.DataFrame(spender_inv) # converting the clusters to df
```

```
[130]: data
```

```
[130]:      0        1        2        3        4       5        6
       0  3.0  36847.0  43950.0  20170.0  36534.0   239.0  47943.0
       1  2.0  32717.0  16784.0  13626.0  60869.0  1272.0   5609.0
```

8   Cluster interpretation

```
[141]: spender1_inv=scaler.inverse_transform(spenders1)
       spender2_inv=scaler.inverse_transform(spenders2)
```

```
[151]: sns.barplot(spender1_inv)
       plt.title('other region spending behaviour') #visualizing first cluster
```

```
[151]: Text(0.5, 1.0, 'other region spending behaviour')
```

other region spending behaviour

```
[ ]:
```
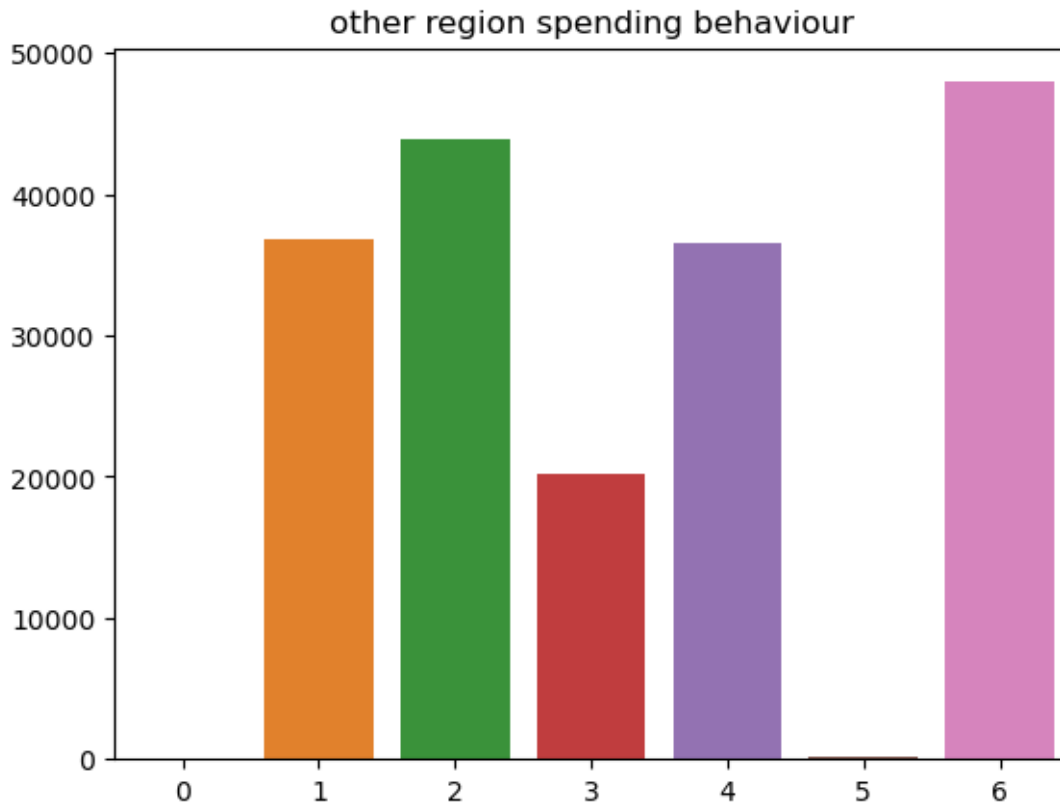
```
[158]: sp=[36847.0,        43950.0,        20170.0,        36534.0,        239.
       ↪0,        47943.0]
       plt.pie(sp,labels=['Fresh', 'Milk', 'Grocery', 'Frozen',
               'Detergents_Paper', 'Delicassen'],autopct='%1.1f%%')
       plt.title('other region spending behaviour') # getting the % spent on each
       ↪products
```

```
[158]: Text(0.5, 1.0, 'other region spending behaviour')
```
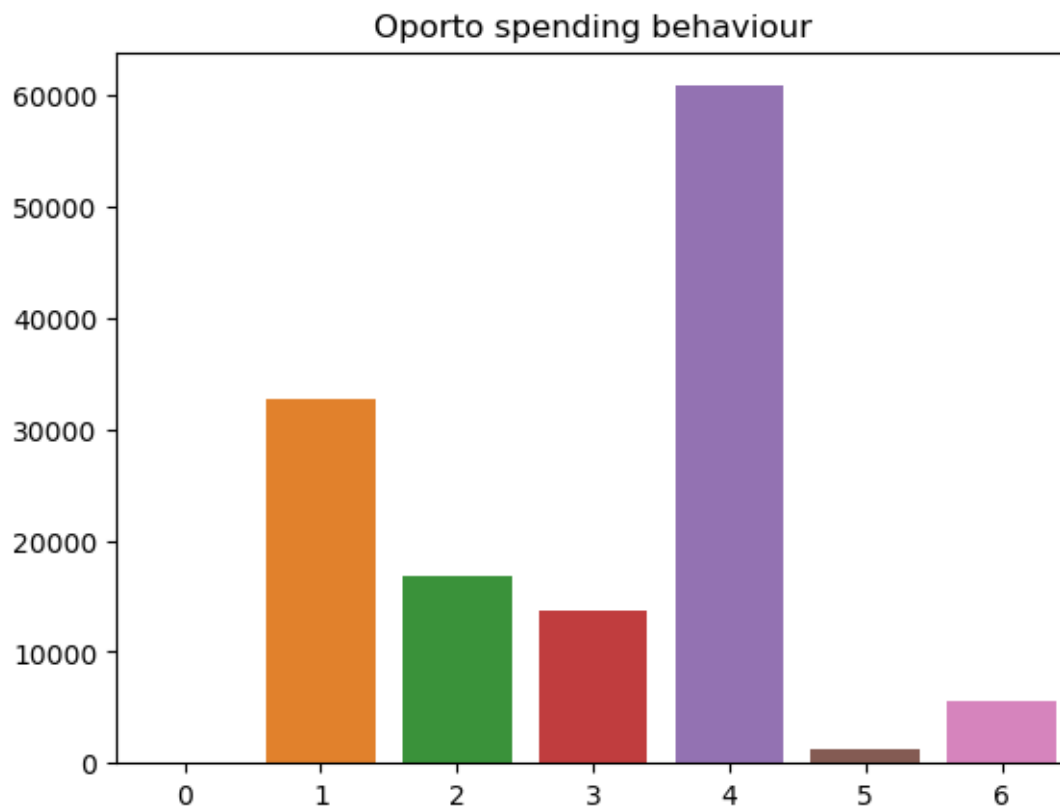
14

other region spending behaviour
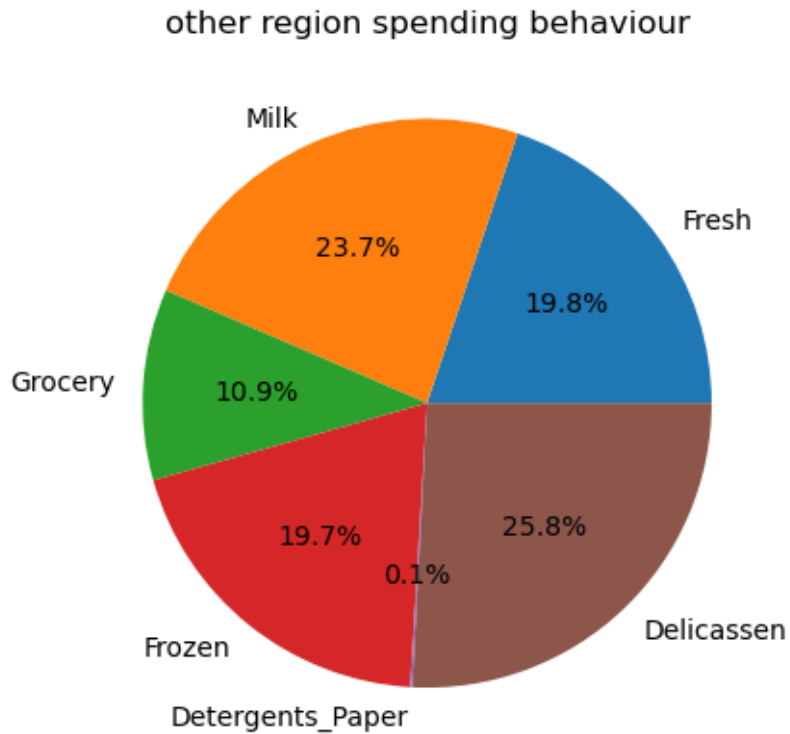
8.1 findings:

8.2 this region called 'other region', spend's very less on detergent and groceries that is 10.9% on grocery and no or 0.1% on detergent paper.

```python
[ ]:
```

```python
[150]: sns.barplot(spender2_inv)
       plt.title('Oporto spending behaviour') # visualizing second cluster
```

```
[150]: Text(0.5, 1.0, 'Oporto spending behaviour')
```

Oporto spending behaviour
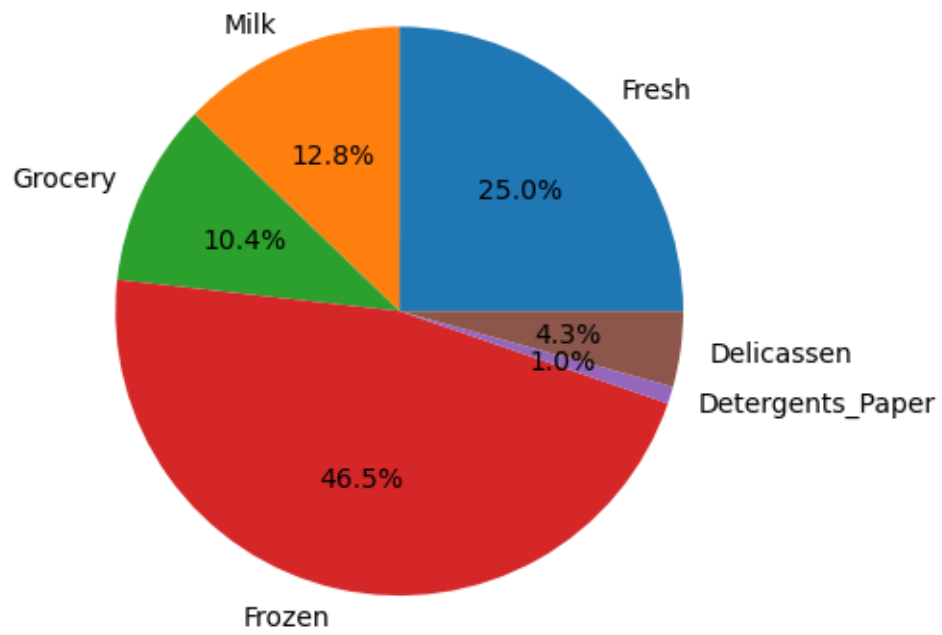
[ ]:

```
[159]: sp2=[32717.0,        16784.0,        13626.0,        60869.0           ,1272.
       ↪0,          5609.03]
       plt.pie(sp2,autopct='%1.1f%%',labels=['Fresh', 'Milk', 'Grocery', 'Frozen',
               'Detergents_Paper', 'Delicassen'])
       plt.title('Oporto spending behaviour') # getting the % of each second cluster
```

[159]: Text(0.5, 1.0, 'Oporto spending behaviour')

Oporto spending behaviour

8.3   findings:

8.4   this clusters are likely hotels, the spend 25% on fresh foods, 46% on frozen food and little or 1.0% on Detergent paper abd 4.3% on Delicassen products. The spend less on other products outside frozen foods especiall from this region Oporto.

9   Findings:

9.1   1. Both clusters spends very low on Detergent papers.

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

10   Recommendations:

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

10.2   2. the marketers should focus more to advertise these low purchase products such as detergent paper,delicassen especially to hotels/resturants in oporto and lisbon regions.

10.3   3. communicate product benefits to customers through emails.

10.4   4. include bonuses and incentives on those products.

[ ]: