

SUMMER BOOTCAMP PROJECT

SIVAM VISNU

-2024

List of Tables

*First five rows of dataset.....	3
*Last five rows of dataset.....	3
*Statistical summary.....	4
*Statistical summary (top 10% spenders, fresh category).....	12
*Statistical summary (top 10% spenders, milk category).....	12
*Statistical summary (top 10% spenders, grocery category).....	13
*Statistical summary (top 10% spenders, frozen category).....	13
*Statistical summary (top 10% spenders, detergents paper category).....	14
*Statistical summary (top 10% spenders, delicatessen category).....	14
*Statistical summary of high spenders.....	15
*Statistical summary of low spenders.....	15
*Correlation matrix(Delicatessen and Frozen category).....	16

List of Figures

*Data description.....	4
*Null value count.....	4
*Boxplot of features(before corrections).....	5
*Boxplot of features(after corrections).....	6
*Average spending on each category.....	7
*Total spending in each region.....	8
*Average spending(Grocery) in each region.....	9
*Variation in spending for different categories.....	10
*Spending on Detergents Paper in each region.....	10
*Correlation heatmap(Fresh and Frozen category).....	11
*Correlation heatmap(Grocery and Milk category).....	16
*Average spending on Milk and Fresh in each region.....	17
*Growing spending trends on Fresh vegetables.....	17
*Standard deviation on total spending in each region.....	18

Problem Statement/ Objective

A wholesale distributor operating in different regions of Portugal has information on annual spending of several items in their stores across different regions and channels. The data consists of 440 large retailers' annual spending on 6 different varieties of products in 3 different regions (Lisbon, Oporto, Other) and across different sales channel (Hotel, Retail).

Data Description

1. Buyer/Spender- ID's of customers
2. Region- Region of the distributor
3. Fresh- spending on Fresh Vegetables
4. Milk- spending on milk
5. Grocery- spending on grocery
6. Frozen- spending on frozen foods
7. Detergents_paper- spending on detergents and toilet paper
8. Delicatessen- spending on instant foods

Data Cleaning and Preprocessing

	0	1	2	3	4
Buyer/Spender	1	2	3	4	5
Channel	Retail	Retail	Retail	Hotel	Retail
Region	Other	Other	Other	Other	Other
Fresh	12669	7057	?	13265	22615
Milk	9656	9810	8808	1196	5410
Grocery	7561	9568	7684	4221	7198
Frozen	214.0	1762.0	2405.0	6404.0	3915.0
Detergents_Paper	2674.0	3293.0	3516.0	507.0	1777.0
Delicatessen	1338.0	1776.0	7844.0	1788.0	5185.0

	435	436	437	438	439
Buyer/Spender	436	437	438	439	440
Channel	Hotel	Hotel	Retail	Hotel	Hotel
Region	Other	Other	Other	Other	Other
Fresh	29703	39228	14531	10290	2787
Milk	12051	1431	15488	1981	1698
Grocery	16027	764	30243	2232	2510
Frozen	13135.0	4510.0	437.0	1038.0	65.0
Detergents_Paper	182.0	93.0	14841.0	168.0	477.0
Delicatessen	2204.0	2346.0	1867.0	2125.0	52.0

- The Shape of the dataframe is (440,9)

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 440 entries, 0 to 439
Data columns (total 9 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Buyer/Spender         440 non-null    int64
1   Channel               437 non-null    object
2   Region               434 non-null    object
3   Fresh                440 non-null    object
4   Milk                 440 non-null    int64
5   Grocery              440 non-null    int64
6   Frozen               437 non-null    float64
7   Detergents_Paper     439 non-null    float64
8   Delicatessen         438 non-null    float64
dtypes: float64(3), int64(3), object(3)
memory usage: 31.1+ KB
```

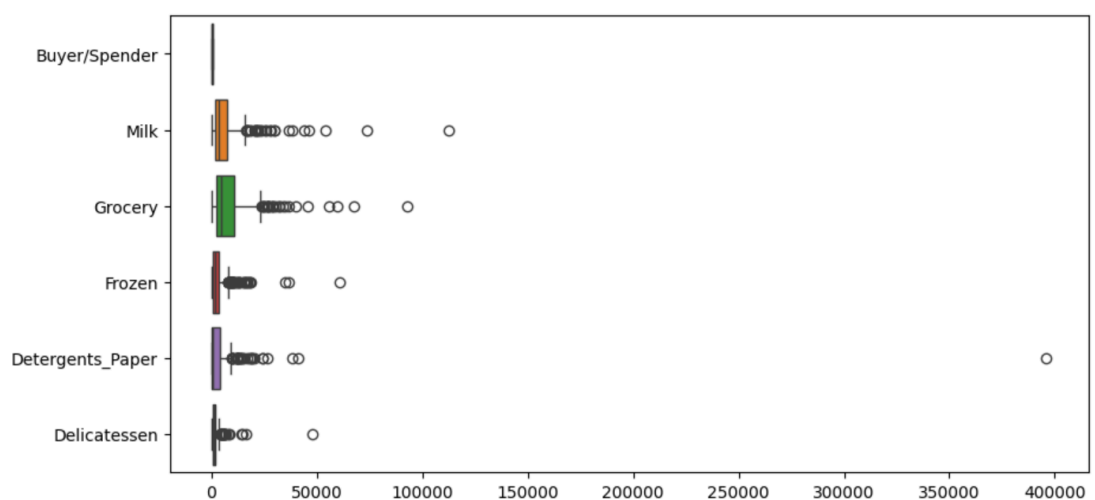
	Buyer/Spender	Milk	Grocery	Frozen	Detergents_Paper	Delicatessen
count	440.000000	440.000000	440.000000	437.000000	439.000000	438.000000
mean	220.500000	6035.779545	7951.277273	3085.638444	3773.747153	1531.057078
std	127.161315	8964.929649	9503.162829	4867.744145	19364.886053	2825.044262
min	1.000000	1.000000	3.000000	25.000000	3.000000	3.000000
25%	110.750000	1525.250000	2153.000000	744.000000	256.500000	411.250000
50%	220.500000	3641.000000	4755.500000	1535.000000	813.000000	971.000000
75%	330.250000	7217.500000	10655.750000	3570.000000	3956.000000	1822.750000
max	440.000000	112400.000000	92780.000000	60869.000000	396100.000000	47943.000000

```

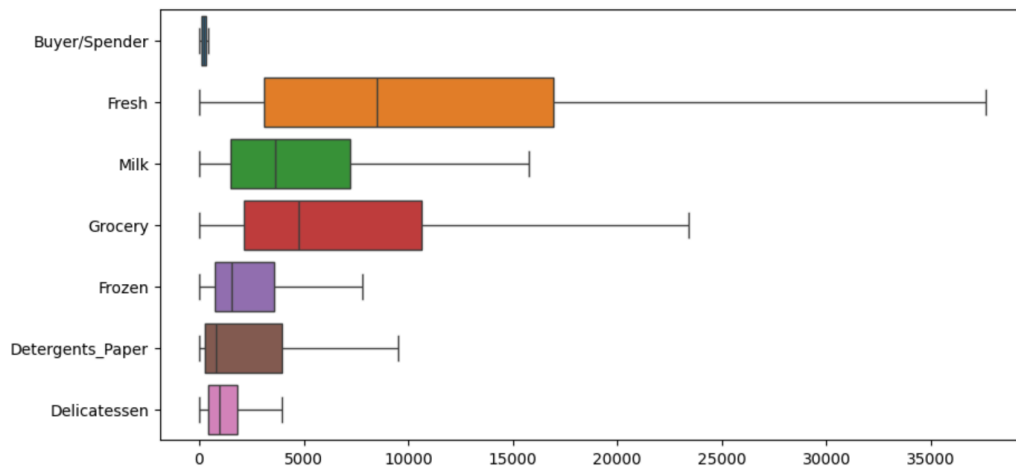
Buyer/Spender      0
Channel            3
Region             6
Fresh              0
Milk               0
Grocery            0
Frozen             3
Detergents_Paper   1
Delicatessen        2
Total Spending      0
dtype: int64

```

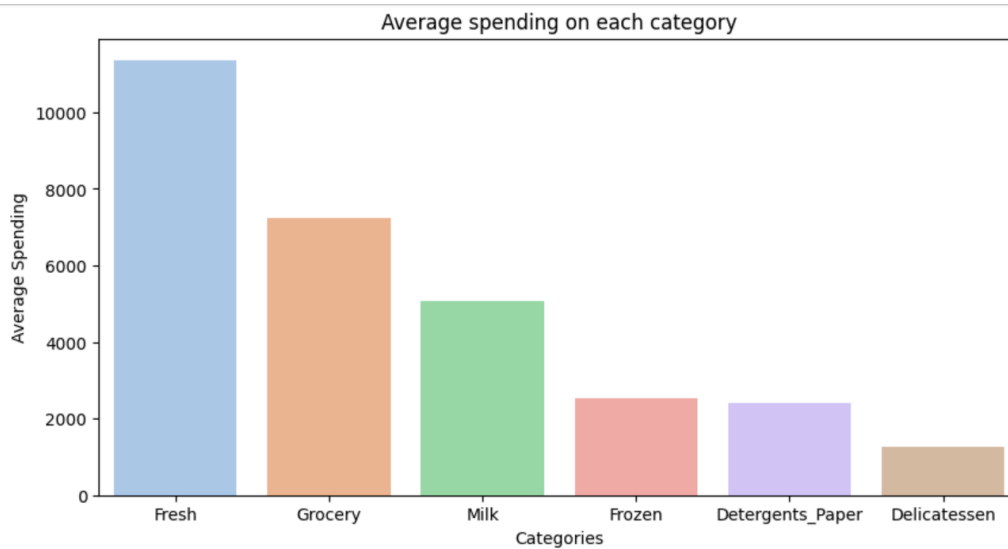
- It has 0 duplicate values



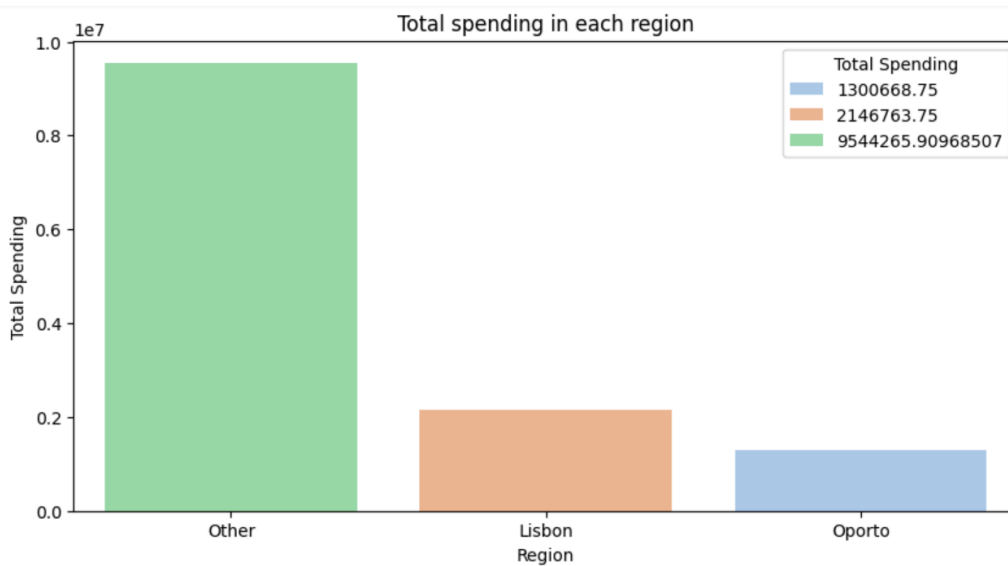
- Detergents_Paper Feature have an outlier with value more than 350000

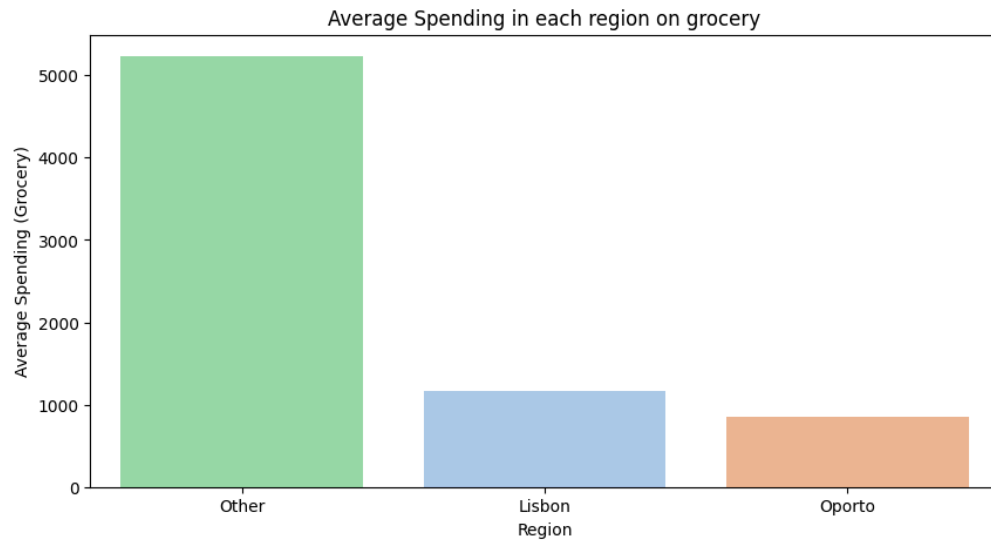


Spending Analysis



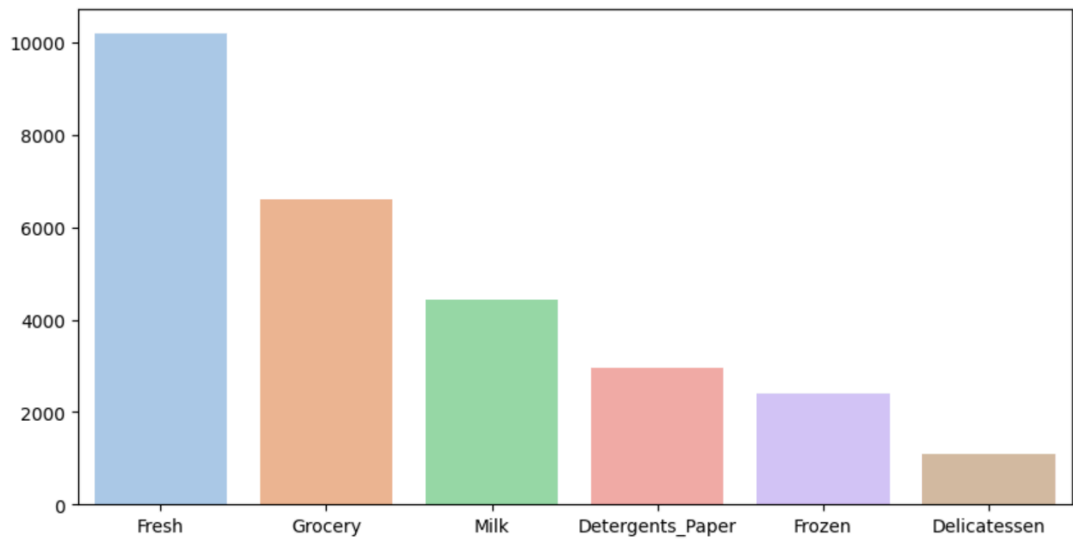
Regional Demand

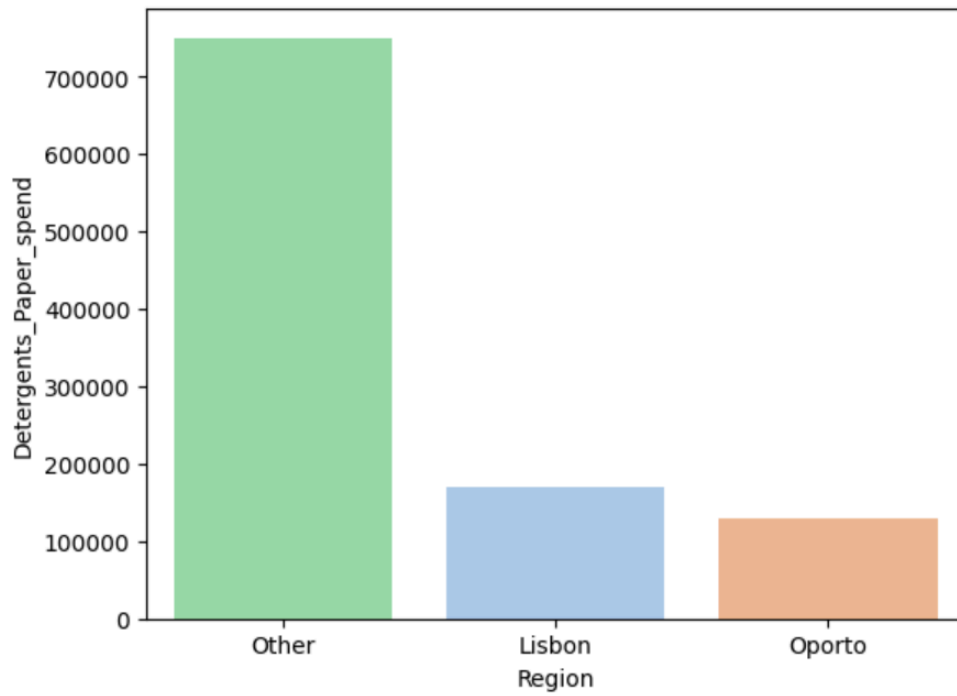




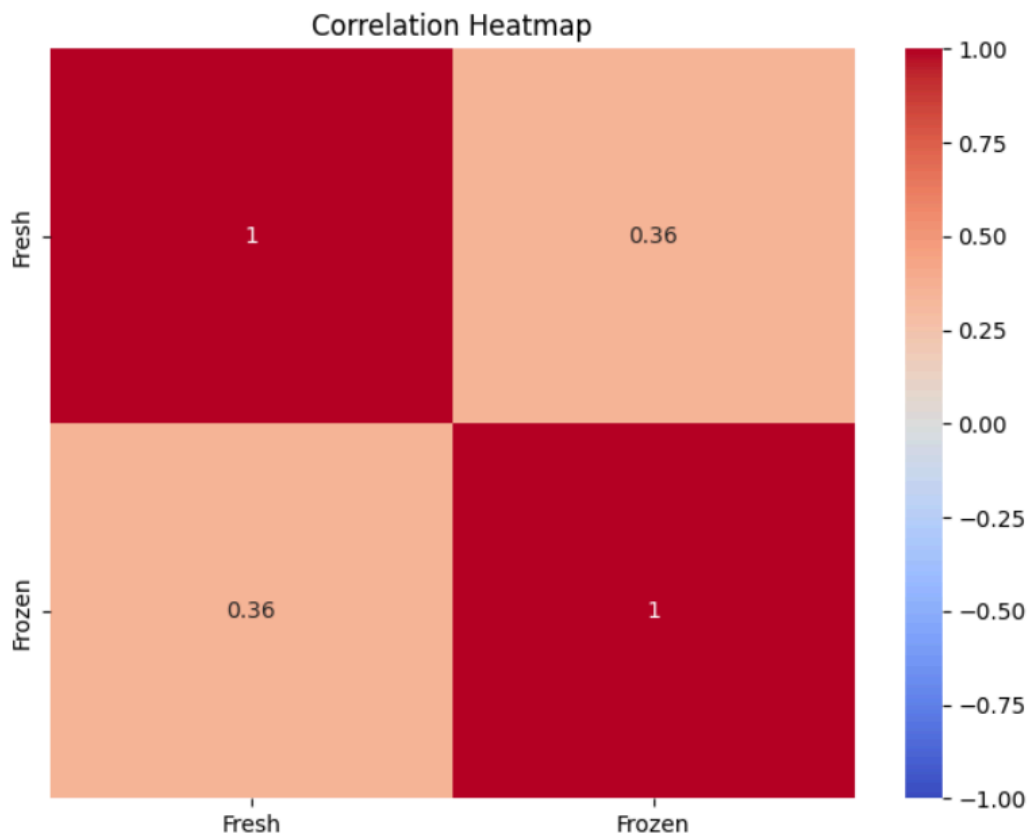
Category Preferences

Category which shows the most variation in spending among buyers is: Fresh





- This correlation indicates that there is a moderate positive relationship between Fresh and Frozen spending. Customers who spend more on Fresh items are likely to also spend more on Frozen items, but the relationship is not very strong, suggesting that other variables might also be influencing the spending patterns.



Customer Segmentation

Fresh

	Buyer/Spender	Fresh	Milk	Grocery	Frozen	Detergents_Paper	Delicatessen	Total Spending
count	44.000000	44.000000	44.000000	44.000000	44.000000	44.000000	44.000000	44.000000
mean	211.568182	34266.613636	5782.437500	7237.403409	4467.613636	1342.965909	1926.931818	55023.965909
std	131.158416	3891.692400	5031.556791	6089.871474	2790.905690	2132.859038	1229.882191	14299.436532
min	13.000000	27167.000000	286.000000	471.000000	287.000000	20.000000	3.000000	34698.000000
25%	100.000000	30562.750000	2054.250000	2493.250000	1726.250000	211.500000	1067.750000	46103.437500
50%	218.500000	36832.000000	3954.500000	5428.500000	4494.500000	601.500000	1821.500000	52072.250000
75%	295.500000	37642.750000	7265.500000	8578.250000	7809.000000	1324.000000	2880.250000	60602.031250
max	437.000000	37642.750000	15755.875000	23409.875000	7809.000000	9505.250000	3940.000000	98035.750000

Milk

	Buyer/Spender	Fresh	Milk	Grocery	Frozen	Detergents_Paper	Delicatessen	Total Spending
count	44.000000	44.000000	44.000000	44.000000	44.000000	44.000000	44.000000	44.000000
mean	196.500000	11724.988636	15042.286932	18611.653409	2930.409091	6599.909091	2077.000000	56986.247159
std	126.616524	11586.796672	1145.466730	6383.194773	2541.921406	3619.303961	1429.678839	15452.800767
min	12.000000	85.000000	12653.000000	1660.000000	33.000000	51.000000	3.000000	26953.000000
25%	75.000000	4109.250000	14580.500000	16381.750000	957.500000	3904.500000	803.750000	49111.656250
50%	192.500000	6301.000000	15755.875000	21550.500000	1824.500000	8678.000000	1584.000000	55459.500000
75%	307.750000	14802.000000	15755.875000	23409.875000	4490.250000	9505.250000	3940.000000	65090.281250
max	438.000000	37642.750000	15755.875000	23409.875000	7809.000000	9505.250000	3940.000000	98035.750000

Grocery

	Buyer/Spender	Fresh	Milk	Grocery	Frozen	Detergents_Paper	Delicatessen	Total Spending
count	44.000000	44.000000	44.000000	44.000000	44.000000	44.000000	44.000000	44.000000
mean	187.954545	8739.017045	12151.701705	22281.181818	2088.204545	8243.971591	1875.882524	55379.959228
std	116.049700	10236.193741	4454.071193	1462.146367	2031.173467	2283.654594	1369.415870	14459.409456
min	24.000000	37.000000	1266.000000	19172.000000	36.000000	239.000000	3.000000	30578.000000
25%	76.500000	1357.000000	9418.250000	21162.750000	758.500000	7506.750000	714.500000	46743.187500
50%	179.000000	5074.000000	14234.000000	23409.875000	1365.000000	9505.250000	1437.500000	53354.750000
75%	302.750000	12140.500000	15755.875000	23409.875000	2770.000000	9505.250000	2992.250000	62788.750000
max	438.000000	37642.750000	15755.875000	23409.875000	7809.000000	9505.250000	3940.000000	98035.750000

Frozen

	Buyer/Spender	Fresh	Milk	Grocery	Frozen	Detergents_Paper	Delicatessen	Total Spending
count	44.000000	44.000000	44.000000	44.000000	44.000000	44.000000	44.000000	44.000000
mean	239.886364	19723.897727	5033.892045	5981.201705	7805.522727	890.301136	1882.704545	41317.519886
std	132.228522	11392.338776	4819.802776	5364.072425	19.333654	1555.474522	1213.047087	19874.210057
min	23.000000	3.000000	333.000000	683.000000	7683.000000	15.000000	46.000000	15731.000000
25%	110.750000	10141.500000	1880.250000	2514.750000	7809.000000	244.250000	905.500000	27460.250000
50%	259.500000	18408.000000	3488.500000	4604.500000	7809.000000	440.500000	1724.000000	34431.000000
75%	339.250000	29933.250000	5375.500000	7026.000000	7809.000000	850.500000	2700.500000	50475.562500
max	436.000000	37642.750000	15755.875000	23409.875000	7809.000000	9505.250000	3940.000000	98035.750000

Detergents_Paper

	Buyer/Spender	Fresh	Milk	Grocery	Frozen	Detergents_Paper	Delicatessen	Total Spending
count	44.000000	44.000000	44.000000	44.000000	44.000000	44.000000	44.000000	44.000000
mean	193.204545	7408.244318	11890.823864	20325.048295	1705.636364	9175.06250	1634.109797	52138.925138
std	115.848855	8537.095014	3931.623321	4318.787439	1712.759844	595.12105	1236.212735	14513.241462
min	29.000000	85.000000	3688.000000	6861.000000	36.000000	7572.00000	59.000000	28829.250000
25%	85.000000	1914.000000	8232.750000	18423.500000	478.250000	8960.00000	714.500000	44229.529538
50%	187.500000	5125.000000	12786.500000	23409.875000	1143.000000	9505.25000	1276.915525	49955.000000
75%	304.250000	9486.750000	15755.875000	23409.875000	2507.250000	9505.25000	2503.500000	57694.875000
max	438.000000	37642.750000	15755.875000	23409.875000	7782.000000	9505.25000	3940.000000	98035.750000

Delicatessen

	Buyer/Spender	Fresh	Milk	Grocery	Frozen	Detergents_Paper	Delicatessen	Total Spending
count	44.000000	44.000000	44.000000	44.000000	44.000000	44.000000	44.000000	44.000000
mean	175.886364	15521.659091	8975.892045	11541.349432	3716.068182	3467.801136	3684.431818	46907.201705
std	134.894139	12203.112801	5356.401080	7437.894536	2672.931719	3252.435828	352.483508	19171.476309
min	3.000000	18.000000	928.000000	1641.000000	42.000000	235.000000	2963.000000	14406.000000
25%	44.750000	4789.500000	4354.500000	4964.750000	1449.250000	735.250000	3308.500000	32960.500000
50%	162.000000	12565.500000	7382.000000	9794.500000	3242.000000	2128.500000	3940.000000	45920.500000
75%	280.000000	24244.750000	15755.875000	18565.000000	6163.500000	4954.250000	3940.000000	55870.843750
max	412.000000	37642.750000	15755.875000	23409.875000	7809.000000	9505.250000	3940.000000	98035.750000

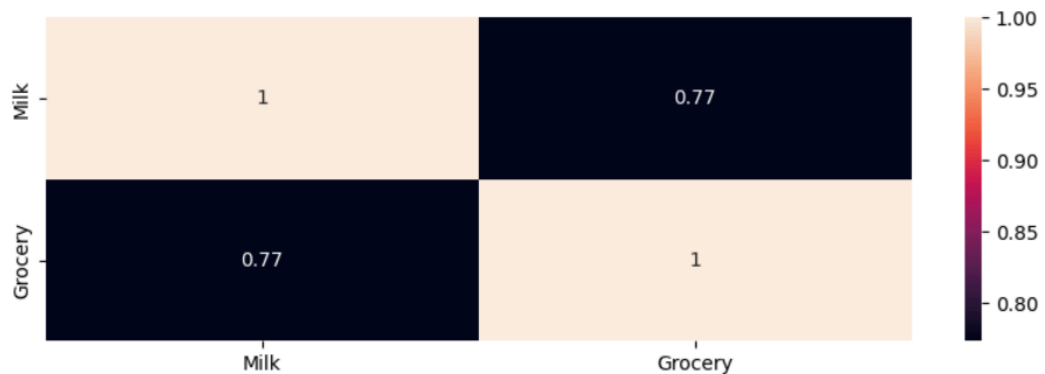
High Spenders

	Buyer/Spender	Fresh	Milk	Grocery	Frozen	Detergents_Paper	Delicatessen	Total Spending
count	88.000000	88.000000	88.000000	88.000000	88.000000	88.000000	88.000000	88.000000
mean	183.397727	20841.985795	9942.548295	14596.583807	3430.125000	4957.409091	2119.261364	55887.913352
std	132.363531	13214.032487	4979.603975	7597.390640	2637.915529	3827.303722	1261.287870	10645.642014
min	5.000000	85.000000	555.000000	764.000000	36.000000	20.000000	3.000000	44855.000000
25%	60.750000	8264.250000	5002.500000	7294.000000	1150.500000	854.000000	1067.750000	47944.562500
50%	169.000000	22327.000000	10869.500000	15191.000000	2875.000000	5003.000000	1880.500000	53354.750000
75%	283.500000	34826.000000	15733.468750	22502.250000	5275.500000	9505.250000	3183.000000	59838.031250
max	438.000000	37642.750000	15755.875000	23409.875000	7809.000000	9505.250000	3940.000000	98035.750000

Low Spenders

	Buyer/Spender	Fresh	Milk	Grocery	Frozen	Detergents_Paper	Delicatessen	Total Spending
count	88.000000	88.000000	88.000000	88.000000	88.000000	88.000000	88.000000	88.000000
mean	249.670455	3835.056818	1839.147727	2232.238636	1338.136364	515.079545	683.363636	10443.022727
std	115.133678	2854.055607	1549.258247	1495.325021	1154.385045	765.025343	687.432307	3337.573780
min	22.000000	3.000000	1.000000	137.000000	65.000000	5.000000	3.000000	904.000000
25%	151.250000	1448.750000	865.250000	1280.250000	518.500000	153.000000	278.500000	9132.500000
50%	245.000000	3288.500000	1258.000000	2017.000000	963.500000	263.500000	494.000000	10946.500000
75%	362.250000	6233.250000	2652.250000	2736.250000	1690.250000	490.750000	971.500000	12748.750000
max	440.000000	9785.000000	8847.000000	8118.000000	5502.000000	4762.000000	3940.000000	15353.000000

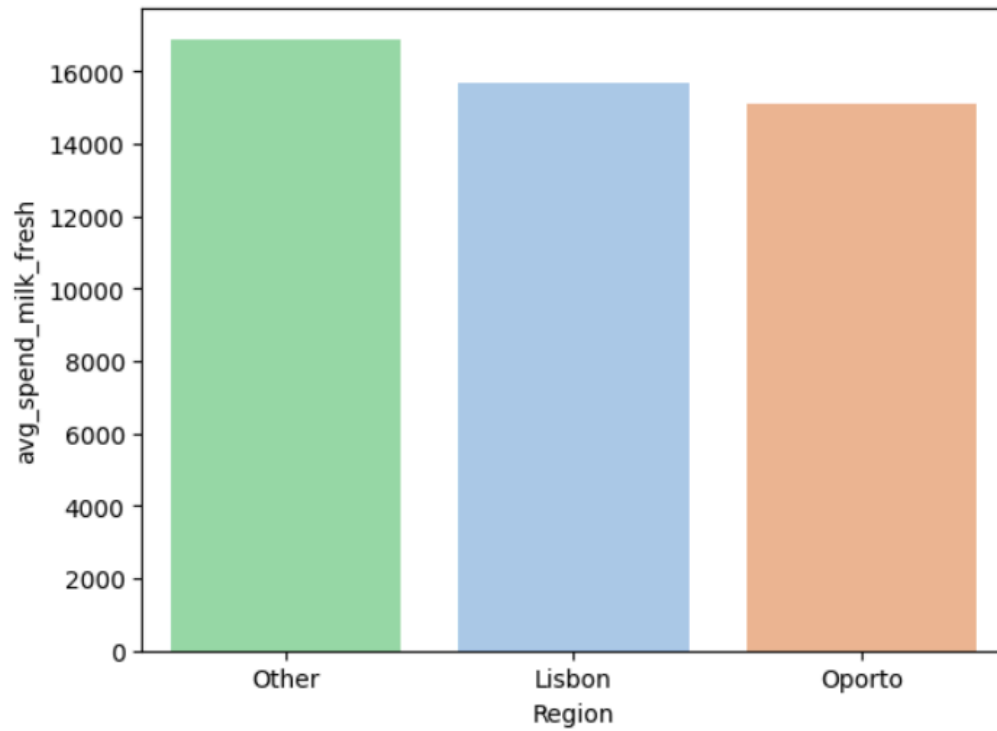
Cross-Category Analysis



- A significant and a positive correlation exists between Milk and Grocery with a correlation value of 0.77.

	Delicatessen	Frozen
Delicatessen	1.00000	0.23194
Frozen	0.23194	1.00000

- The poor positive correlation shows that a buyer who spends more on Frozen food spends more on Delicatessen too, however, the relationship is not very strong, suggesting that other variables might also be influencing the spending patterns.



- The combined average spending on Fresh and Milk for each region is:

Lisbon - 16259.140000

Oporto - 15113.130814

Other - 17671.552669

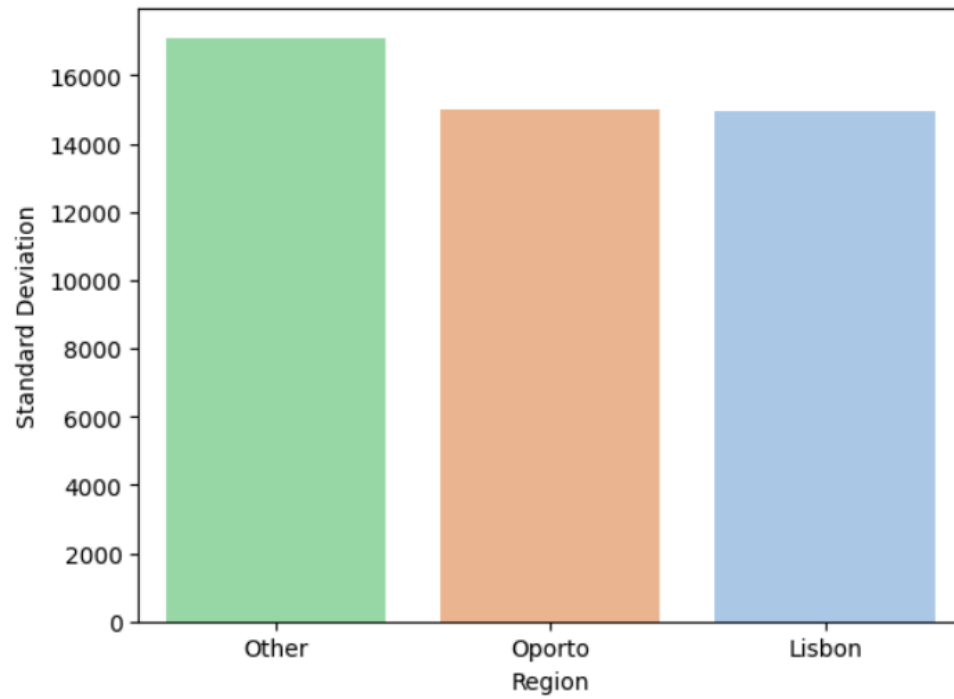
Demand Trends

```
Region
Other    11776.954905
Lisbon   10688.736667
Oporto    10054.488372
Name: Fresh, dtype: float64
```

- Lisbon Region has the fastest growing spending on Fresh Vegetables.

Buyer Insights

```
Region  std_ts_region
0  Other  17104.303013
1  Oporto  14996.166548
2  Lisbon  14948.978861
```



- Oporto has most diverse spending pattern
- Around 36 Buyer/Spender has spend consistently