Super Mart Case Study

Developing an Analytical Case Study

By: Kirthi Chandra

BUSINESS OBJECTIVE

To analyze and address the operational inefficiencies in the client's supply chain management system in order to optimize processes, reduce costs, and improve overall efficiency.

HIGH- LEVEL APPROACH

By conducting exploratory data analysis (EDA), variate analysis, and correlation analysis, we aim to address specific business questions and define clear objectives for the data analysis project.

APPROACH TO PROJECT:

1. Identify the Demographic Factors: We will **focus on customers** who have completed their Graduation, are married, do not have children or teenagers, hold Serbian citizenship, and are currently employed, as they tend to spend more.

2. Analyse Spending Habits: We will study the purchase channel and transaction of our target customers to understand how much and how often they spend on our products.

3. Campaign Acceptance Analysis: We will assess the **responses received** for the fifth campaign and determine the group with the highest number of responses. We will also look into the channels through which they primarily received campaign information.

4. Education Level Analysis: By categorizing customers **based on their educational qualifications**, such as PhD, Graduates, and Masters, we will identify the highest spenders and understand the relationship between education level and spending.

5. Analysis of Unique Spenders: We will examine the **spending habits of customers** who exhibit YOLO (You Only Live Once) or absurd behavior. This will help us understand their motivations and preferences that lead to higher spending.

6. Geographic Spending Analysis: We will analyse customer **data from different countries** like the United States (US), Canada (CA), and Germany (GER) to determine the countries with the highest spending. This will help us understand regional variations in spending habits.

		0	1	2	3	4
Other Details	ID	1826	1	10476	1386	5371
	Year_Birth	1970	1961	1958	1967	1989
	Education	Graduation	Graduation	Graduation	Graduation	Graduation
	Marital_Status	Divorced	Single	Married	Together	Single
	Income	$84,835.00	$57,091.00	$67,267.00	$32,474.00	$21,474.00
	Kidhome	0	0	0	1	1
	Teenhome	0	0	1	1	0
	Dt_Customer	2014-06-16 00:00:00	2014-06-15 00:00:00	2014-05-13 00:00:00	2014-11-05 00:00:00	2014-08-04 00:00:00
	Recency	0	0	0	0	0
Type of Product	MntWines	189	464	134	10	6
	MntFruits	104	5	11	0	16
	MntMeatProducts	379	64	59	1	24
	MntFishProducts	111	7	15	0	11
	MntSweetProducts	189	0	2	0	0
	MntGoldProds	218	37	30	0	34
Source of Purchase	NumDealsPurchases	1	1	1	1	2
	NumWebPurchases	4	7	3	1	3
	NumCatalogPurchases	4	3	2	0	1
	NumStorePurchases	6	7	5	2	2
	NumWebVisitsMonth	1	5	2	7	7
Campaign	AcceptedCmp3	0	0	0	0	1
	AcceptedCmp4	0	0	0	0	0
	AcceptedCmp5	0	0	0	0	0
	AcceptedCmp1	0	0	0	0	0
	AcceptedCmp2	0	1	0	0	0
	Response	1	1	0	0	1
	Complain	0	0	0	0	0
	Country	SP	CA	US	AUS	SP

2240 Rows	28 Columns	47 - Duplicates	1% – Null Values

Cleaning:

1. Income Column – rename-dtype-float
2. ID – Column Removal
3. Converting Year_Birth to Age Column
4. 47 Duplicates Removed
5. Adding Total_spends column

Outliers:
Income – 1 Number
Age – 3 Numbers
Country – 3 Number

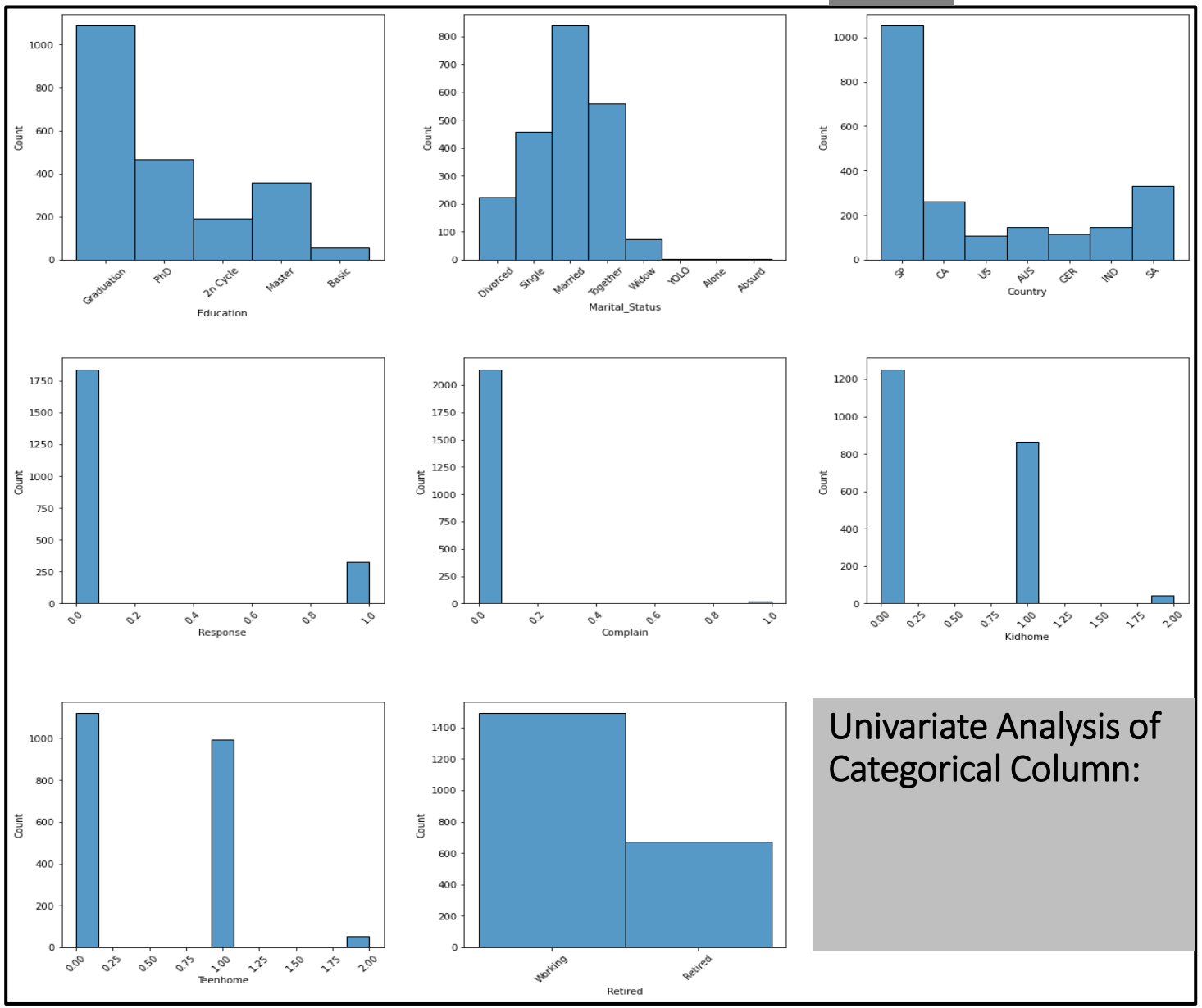
Output:

2162---------- Rows
29 ----------Columns
0----------Duplicates
0---------- Null Values

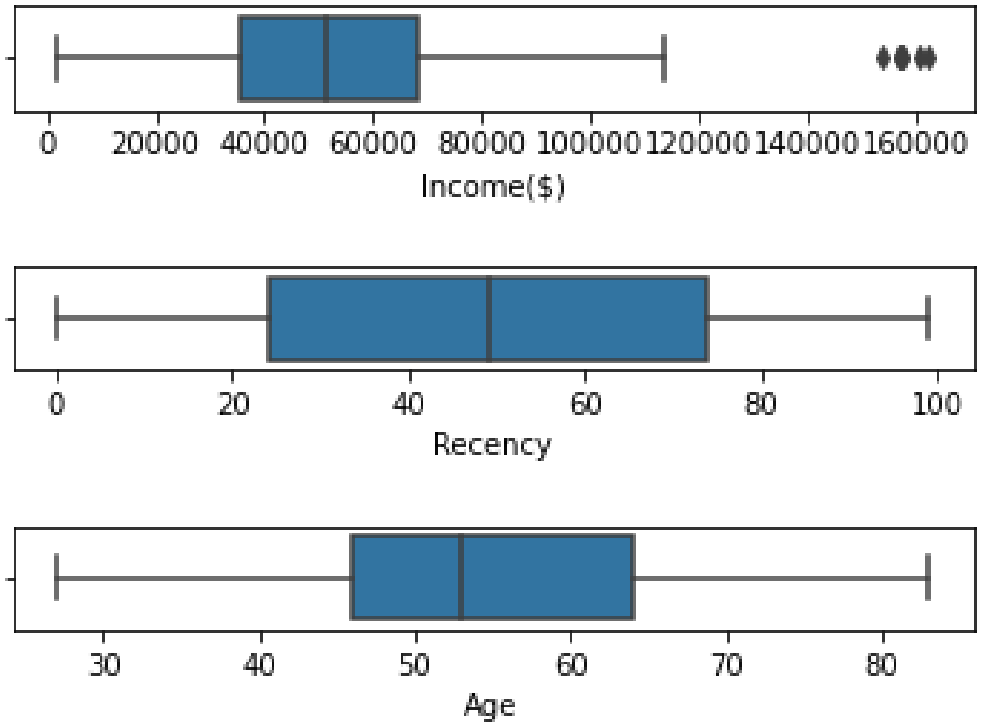
	Mean	Median	Mode
Education	NaN	NaN	Graduation
Marital_Status	NaN	NaN	Married
Income($)	51982.285156	51381.5	7500.0
Kidhome	0.442183	0.0	0.0
Teenhome	0.506013	0.0	0.0
Dt_Customer	2013-07-12 14:53:50.342275840	2013-07-12 00:00:00	2012-08-31 00:00:00
Recency	49.065217	49.0	56.0
MntWines	305.1605	175.5	2.0
MntFruits	26.270583	8.0	0.0
MntMeatProducts	167.254857	68.0	7.0
MntFishProducts	37.509713	12.0	0.0
MntSweetProducts	27.00185	8.0	0.0
MntGoldProds	44.044403	24.0	1.0
NumDealsPurchases	2.325624	2.0	1.0
NumWebPurchases	4.09667	4.0	2.0
NumCatalogPurchases	2.667438	2.0	0.0
NumStorePurchases	5.800648	5.0	3.0
NumWebVisitsMonth	5.323312	6.0	7.0
Country	NaN	NaN	SP
Age	54.115171	53.0	47.0
Retired	NaN	NaN	Working
total_spends	607.241906	397.0	46.0

Inference:

1. Customers are mostly **Graduated**, **married**, with **no kids and no teens** at home.
2. There is a **gap between** the mean, median and mode of Income indicating a **wide spread of data**.
3. The averagely **maximum** bought product is **wine** followed by **meat**.
4. **Store** purchase is **mostly** utilized followed by web purchases.
5. Most of the citizens are from **Serbia** SP with an average **age of 54 years**. They are mostly not retired (if 60 is age of retirement).
6. There is a gap between Mean, Median and Mode of **total_spends which is a combination** of all the products available.

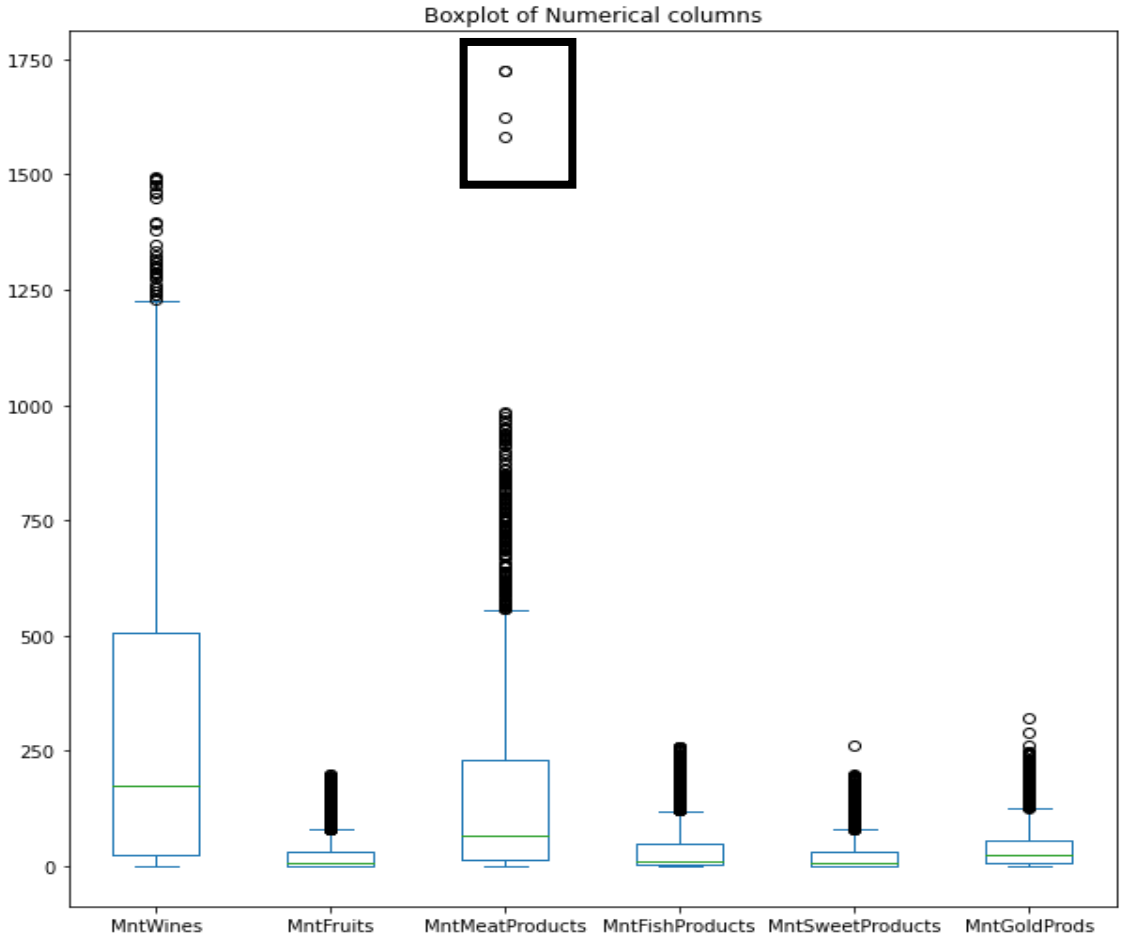


Univariate Analysis of Categorical Column:

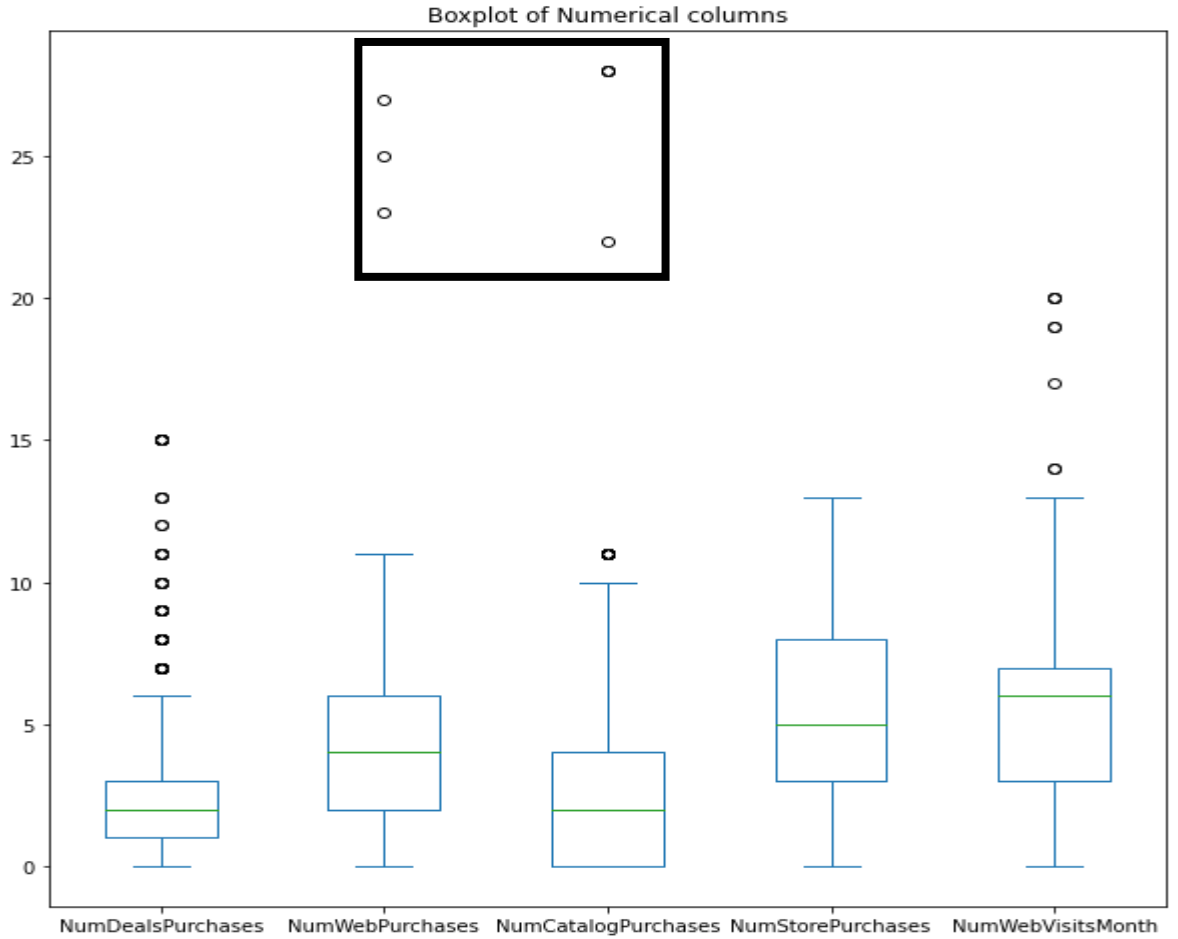


Univariate Analysis of Numerical Column: (Part1)

Inference:
There are still some outliers in Income column, but **no changes to be done**.

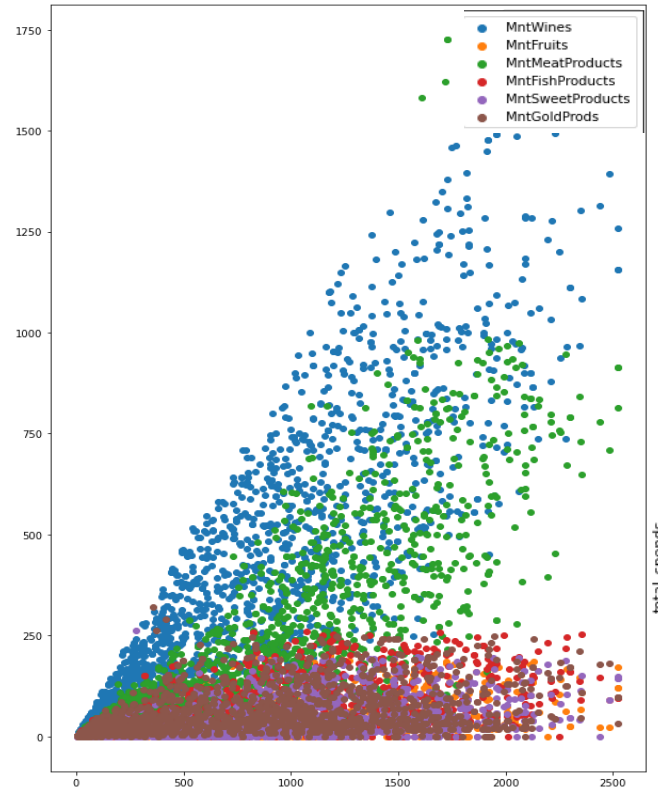


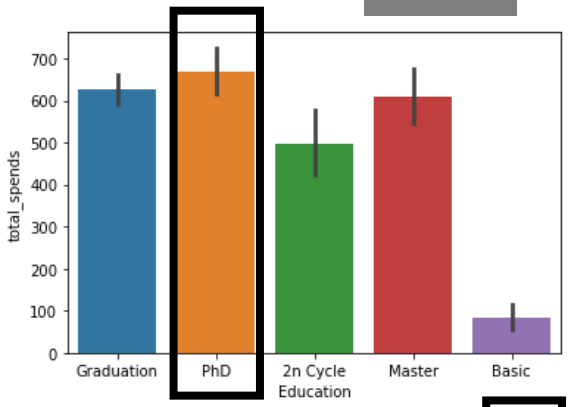
Univariate Analysis of Categorical Column: (Part 2)

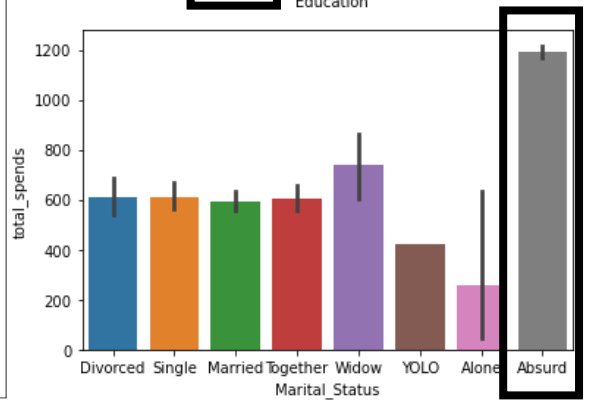


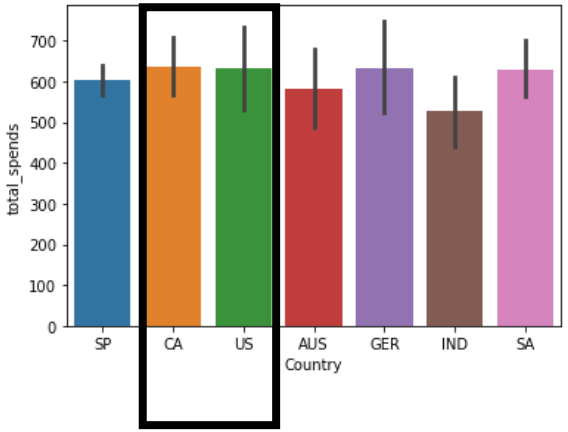
Univariate Analysis of Categorical Column: (Part 3)

Inference: There are outliers in products and mode of purchase but I am **not removing as the outliers may tend to show** new possibilities.

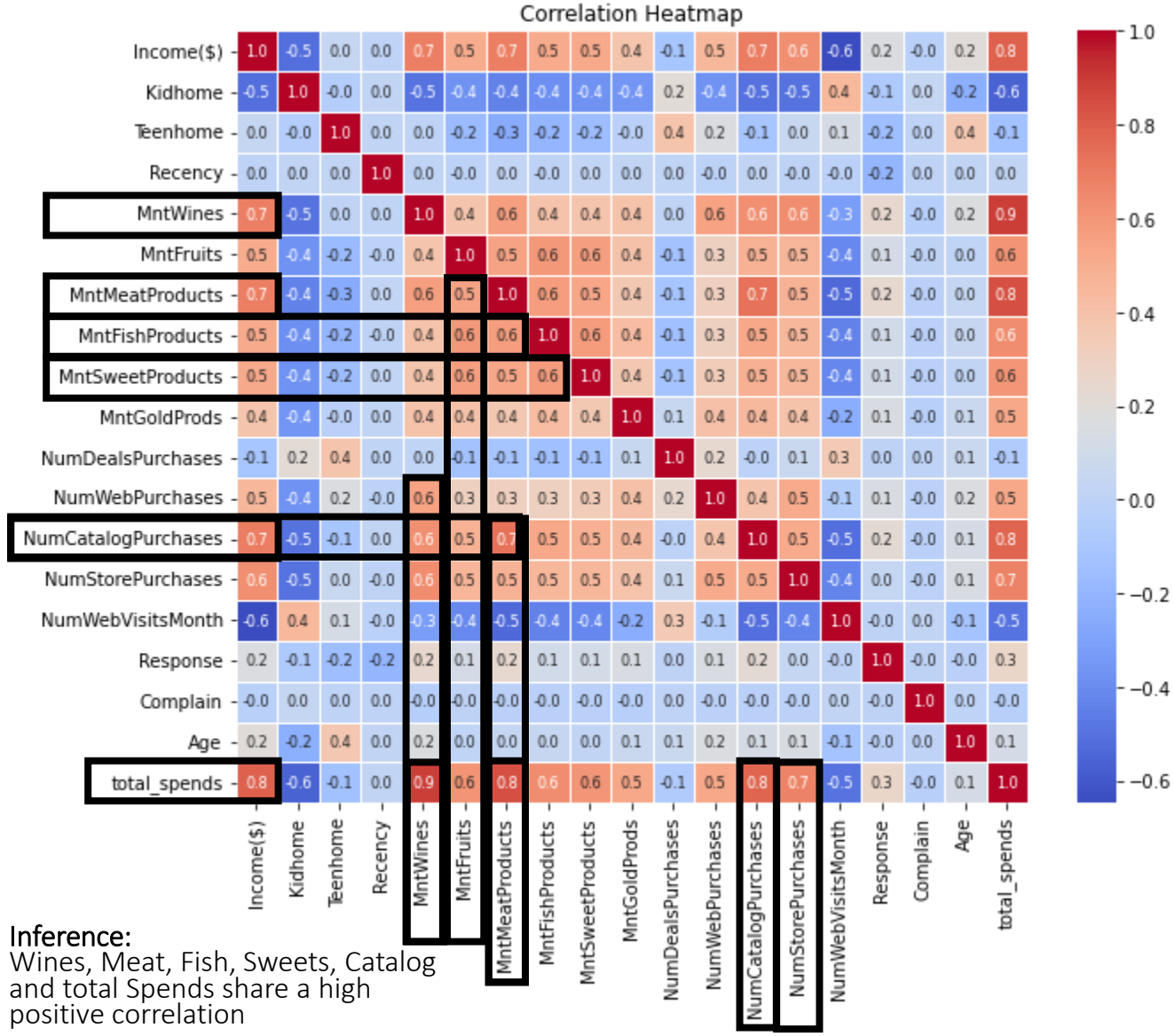








Bivariate Analysis on Products:
Wines played a major role followed by Meat



Inference:
Wines, Meat, Fish, Sweets, Catalog and total Spends share a high positive correlation

average and Median of 'total spends' of customers?

[Web Visit per month <10]

```
median 401.000000
mean 609.965116
```

[Web Visit per month >=10]

```
median 52.000000
mean 119.333333
```

[Num Store Purchase<10]

```
median 221.500000
mean 481.563483
```

[Num Store Purchase >=10]

```
median 1124.000000
mean 1192.863874
```

[Num Deals Purchases <10]

```
median 393.00000
mean 604.01028
```

[Num Deals Purchases>=10]

```
median 826.000000
mean 921.590909
```

Inference:

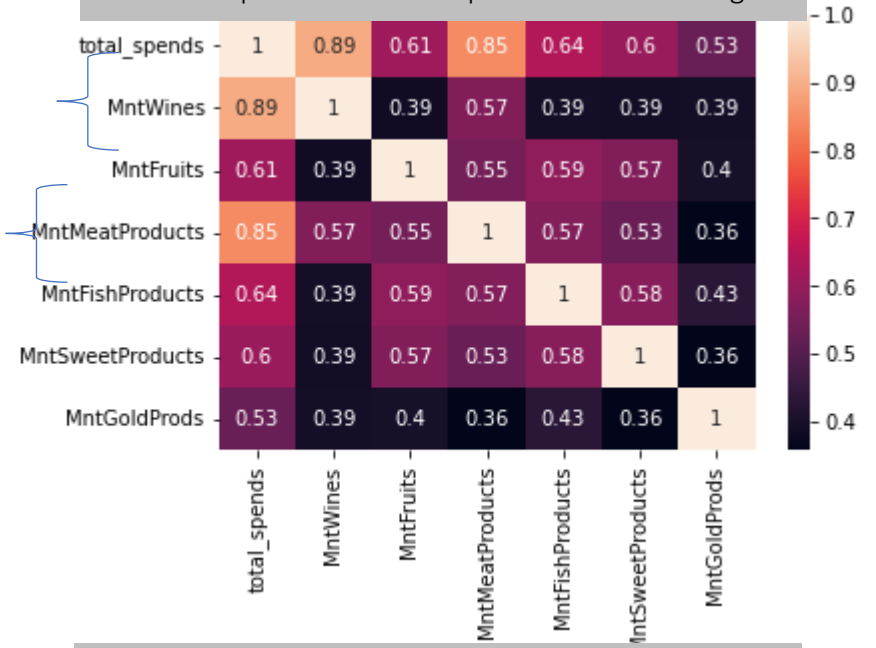
Wines, Meat, Catalogs, web and income are highly corelative to total spends

Customers acceptance campaign ?

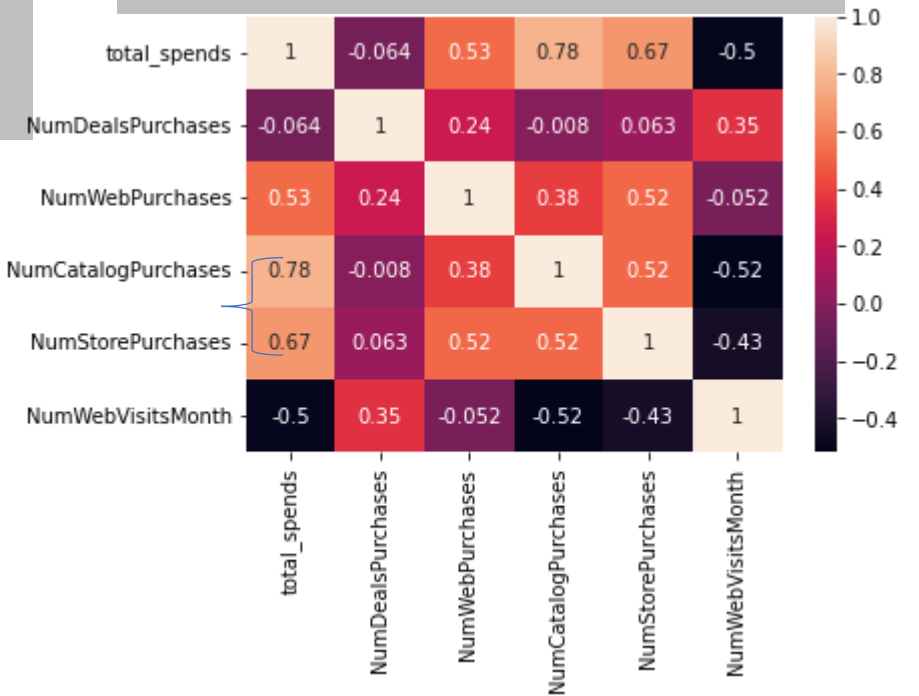
```
AcceptedCmp1 140
AcceptedCmp2 029
AcceptedCmp3 158
AcceptedCmp4 162
AcceptedCmp5 159
```

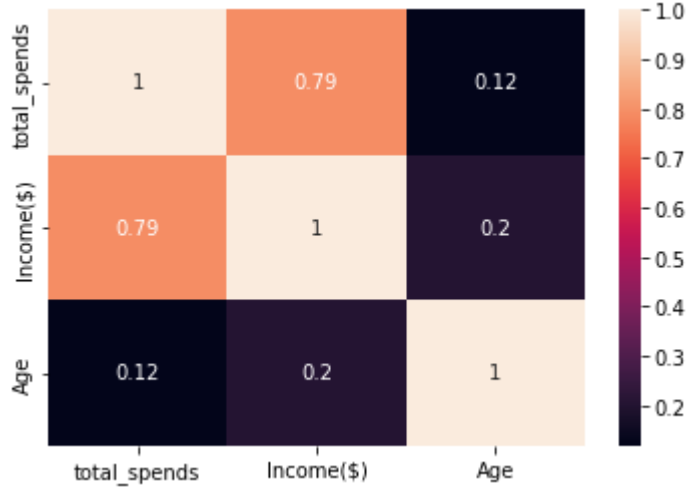
Country	Accepted Cmp1	Accepted Cmp2	Accepted Cmp3	Accepted Cmp4	Accepted Cmp5
AUS	7	0	9	6	12
CA	18	6	17	24	21
GER	7	2	10	11	8
IND	7	2	13	10	6
SA	20	4	21	20	21
SP	74	15	80	85	86
US	7	0	8	6	5

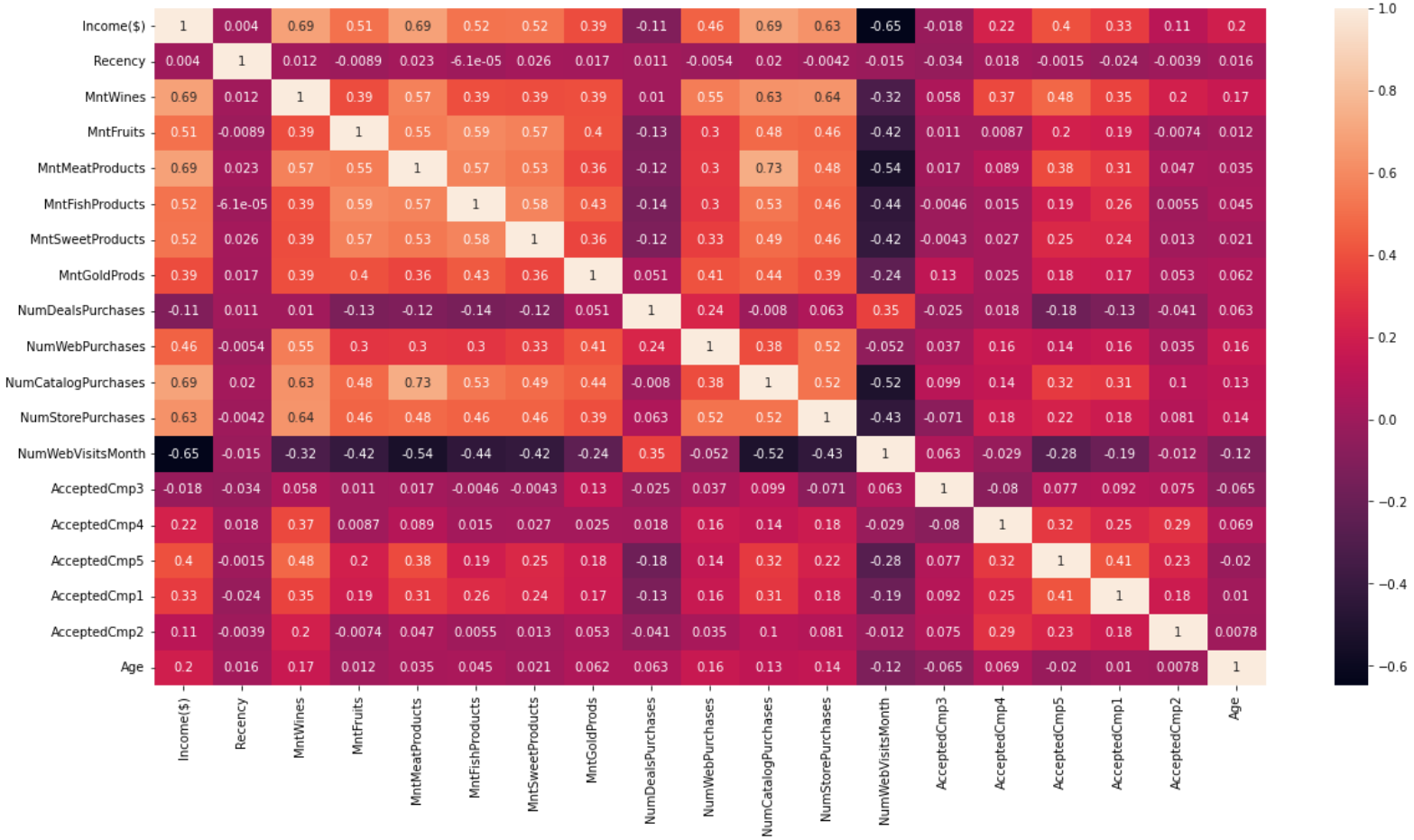
Relationship between Total Spends and Food bought:



Relationship between Total Spends and Purchase Channel:







Inference:

1. The customers who are relevant individuals and who have responded the most in the last campaign typically possess their **Graduation**, are **married**, **do not have children or teenagers**, hold **Serbian** citizenship, and are currently **employed rather than retired**.

2. The individuals have predominantly shown acceptance in the **fifth campaign**, with the **highest number of responses**, are those who primarily received information through in-store channels.

3. These are the customers who frequently purchase a larger quantity of **meat and wines** within the specific as stores.

4. The individuals who have predominantly shown acceptance in the 3rd, 4th and 5th campaign, with the highest number of responses, are those who primarily **received information through in-store channels and web visit purchases**.

5. There is a **strong positive correlation of 0.91** between the sales of meat and fish products.

Business Objectives and Approach

Business Objective
To analyze and address the operational inefficiencies in the client's supply chain management system in order to optimize processes, reduce costs, and improve overall efficiency.

High- Level Approach
By conducting exploratory data analysis (EDA), variate analysis, and correlation analysis, we aim to address specific business questions and define clear objectives for the data analysis project.

Approach to Project:

1. Identify the Demographic Factors: We will **focus on customers** who have completed their Graduation, are married, do not have children or teenagers, hold Serbian citizenship, and are currently employed, as they tend to spend more.

2. Analyze Spending Habits: We will study the purchase channel and transaction of our target customers to understand how much and how often they spend on our products.

3. Campaign Acceptance Analysis: We will assess the **responses received** for the fifth campaign and determine the group with the highest number of responses. We will also look into the channels through which they primarily received campaign information.

4. Education Level Analysis: By categorizing customers **based on their educational qualifications**, such as PhD, Graduates, and Masters, we will identify the highest spenders and understand the relationship between education level and spending.

5. Analysis of Unique Spenders: We will examine the **spending habits of customers** who exhibit YOLO (You Only Live Once) or absurd behavior. This will help us understand their motivations and preferences that lead to higher spending.

6. Geographic Spending Analysis: We will analyze customer **data from different countries** like the United States (US), Canada (CA), and Germany (GER) to determine the countries with the highest spending. This will help us understand regional variations in spending habits.

Super Mart Case Study

Developing an Analytical Case Study

By: Kirthi Chandra

Business Objective

Data Health Review

Exploratory Data Analysis (EDA)

Key Performance Index (KPI)

Questions

Business Objectives and Approach