**Analysis of Marketing Data**

**Section 01**

1. There are 24 missing entries in the Income column. These seem to have been intentionally left empty. The subset of the dataset with missing entries in the entry’s column indicates that most of those customers are highly educated and married. Thus, I chose to fill in the missing values with median income of the customers.

There was an outlier in the income column which seems to have been intentionally and erroneously entered. Thus, I decided to delete the single instance.

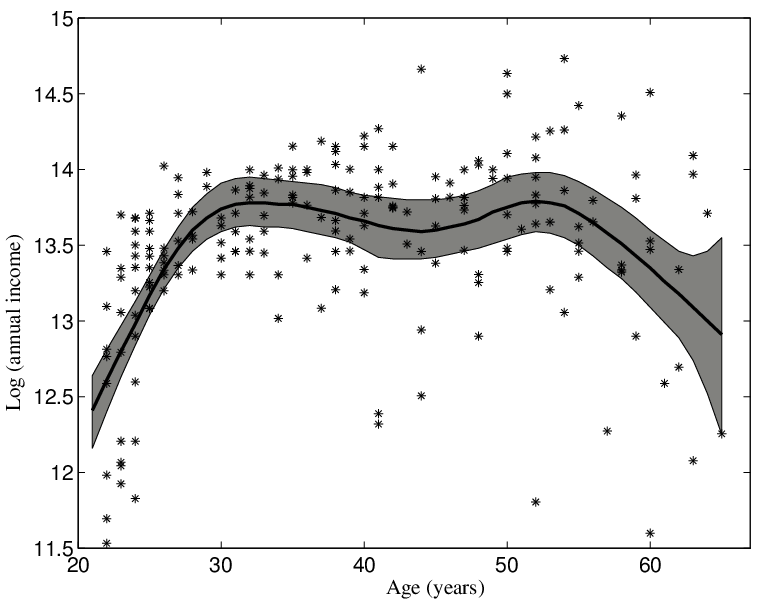
Having transformed the year of birth into an age column three outliers were found which seem too extreme. Those outliers were deleted.

47 duplicates were found in the dataset.

1. The customer’s date of enrollment with the company was transformed into three separate columns of Month, Day and Year.
2. It is well known that there is a strong correlation between age and income in the world’s population. Income generally increases with age up until around retirement age. After retirement age it generally decreases. This pattern does not hold true in our dataset. What income a person has seems to be completely random relative to age. One hypothesizes that this dataset is largely compromised of high earning individuals and this should always be kept in mind.

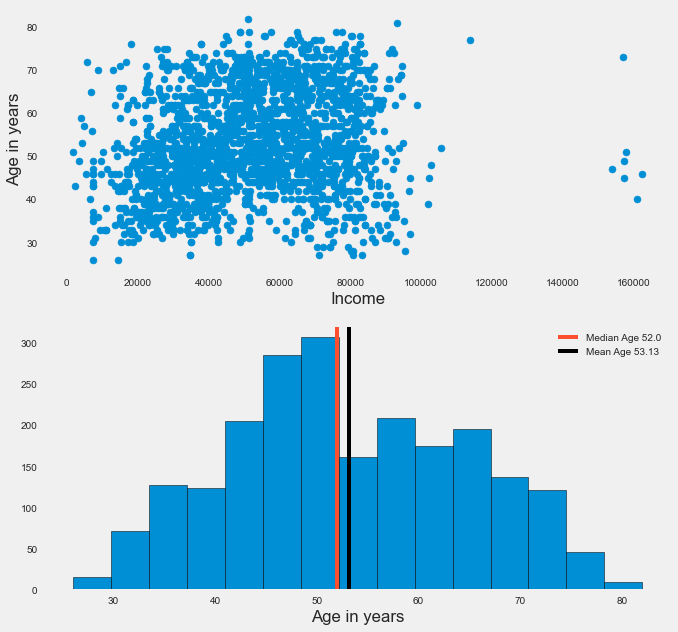
Additionally, it seems the dataset represents an older population. The average age of Spain is 44,9 years (Spain constitutes 48,88% of customers in the dataset) and the average age of customers in the dataset is 53,13 years. While 56,42% are between the age of 45 years and 65 years old.

General scatter plot for Age and Income.



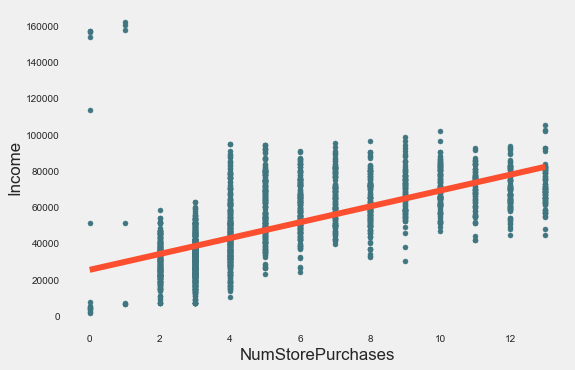
Source: https://www.researchgate.net/figure/Scatterplot-of-logincome-versus-age-for-a-sample-of-n-205-Canadian-workers-with\_fig1\_5142911

Scatter plot of Age and Income from dataset, along with the histogram of Age.

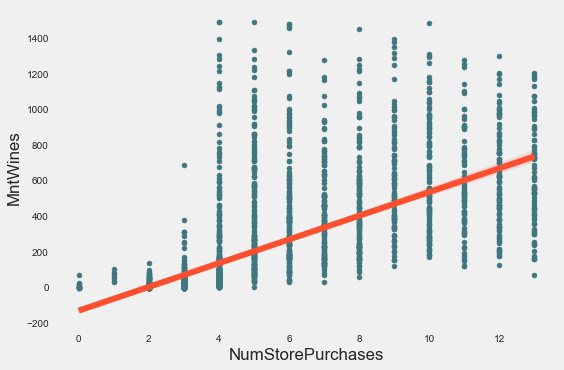


**Section 02**

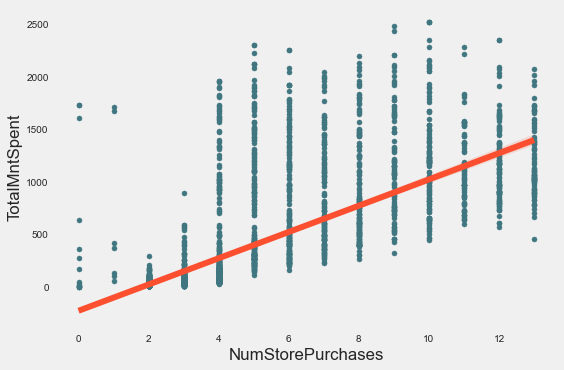
1. Income is positively correlated to number of store purchases. The higher a customer’s income the higher the probability they will buy products from the physical store. The R-Squared value of the regression line between these two variables is 0,392.



The amount spent on wines is positively correlated to the number of instore purchases. This means that there is a high probability that when a customer decides to purchase wine it will be an instore purchase. The R-Squared value of the regression line between these two variables is 0,409.



Total number of sales is positively correlated to instore purchases. This means that most customers conduct their purchases instore. The R-Squared value of the regression line between these two variables is 0,454.



1. No, the USA does not fare significantly more when it comes to the number of Total purchases. Mexico only has three customers thus its contribution to the number of sales can not be considered due to the relative scale of the dataset.
2. The seems to be no correlation between above average Gold purchasers and instore purchases. The correlation coefficient between these two attributes is 0.01395



Scatter plot of Above average gold purchasers and instore purchases.

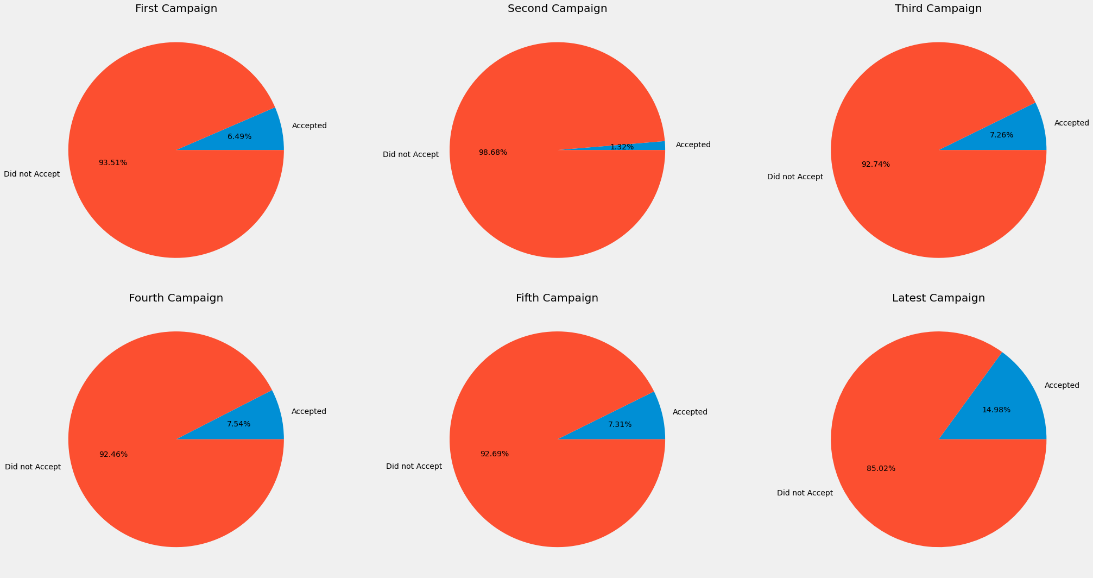
1. On average people who are Married and are PhD candidates, spend less on fish. On average this group of customers spend $26,77 on fish in the past two years. Which is statistically significantly less that the average of rest of the customers base ($38,45 in the past two years.). when one conducts a Kolmogorov–Smirnov test (Which test if two samples come from the same population) one gets an alpha and P-Value of 0,1888683 and 7,66773e-06 respectively.
2. There isn’t enough evidence to support the hypotheses that geography doesn’t influence the success of a campaign. The Following table is the result of Chi-Squared Analysis to check if geography and the success of a campaign are not independent. The decision point for Chi-squared statistic is 12,591587243743977.



**Section 03**

1. The most successful campaign was the latest campaign.

*See the graph on following page…*

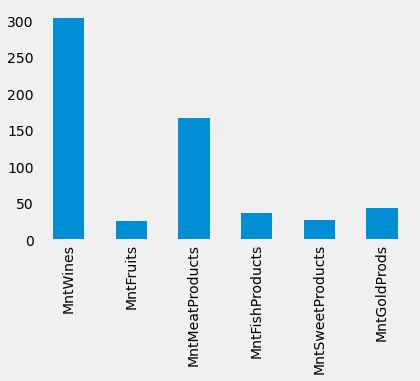


1. The most common demographic attributes of a customer are as follows:

* They are Married.
* Graduated from high school.
* Between the ages of 52,71 years and 53,54 years old.
* Most of them have no children present in the house.
* They earn between $51231,03 and $52736,63 a year.

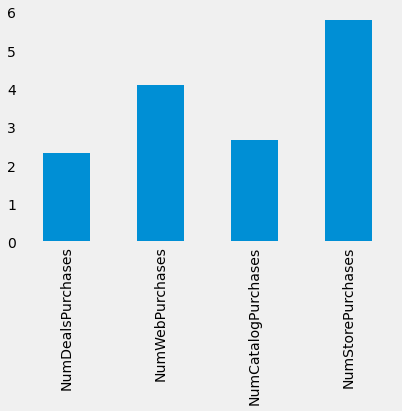
1. Wines are the best performing products.

* Spend between $292,51 and $316,19 on Wines.
* Spend between $24,79 and $27,58 on Fruits.
* Spend between $159,38 and $175,32 on Meat products.
* Spend between $35,53 and $39,37 on Fish Products.
* Spend between $25,60 and $28,51 on sweet products.
* Spend between $42,30 and $45,98 on gold products.



1. Most customers make their purchases instore, although the number of online purchases is not that far behind. Additionally:

* Number of purchases made with discount is between 2,25924 and 2,3949.
* Number of online purchases is between 4.0007 and 4.1966.
* Number of catalog purchases is between 2.5586and 2.7643.
* Number of purchases made in store between 5.6759 and 5.9046



**Additional Analysis**

**First campaign**

* Demographics -
* Purchase channels
* Behavioral patterns