Exploratory data analysis

Correlations between the bank products. The correlation between the bank products is very useful to understand which products were frequently bought together. There is a strong correlation between Pension and Payroll Acccount, between Pension and Direct Debit and Payroll_Acc and Direct Debit. On the other hand, there is a negative correlation between Payroll_Acc and Current_Acc, Pensions and Current_Acc, meaning that these products are not frequently bought together.

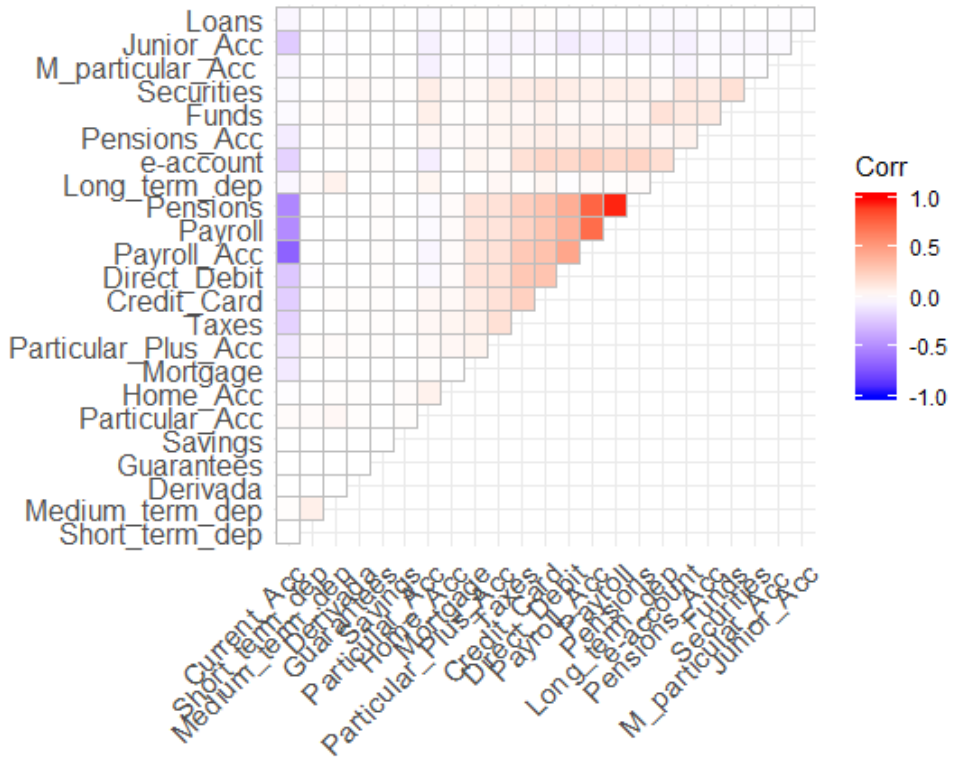

Figure 3. The Correlation between the bank products

Age distribution of the active customers. Shows that there are few products for people under 20, which can be explained by the junior account, more than 200000 customers between 20

and 25 years old and between 40 and 45 years old. There are very few customers close to 100 years old, but the they will not be considered outliers because the information are not incorrectly recorded, and they will not affect our analysis as it focuses more on their products than their profile.

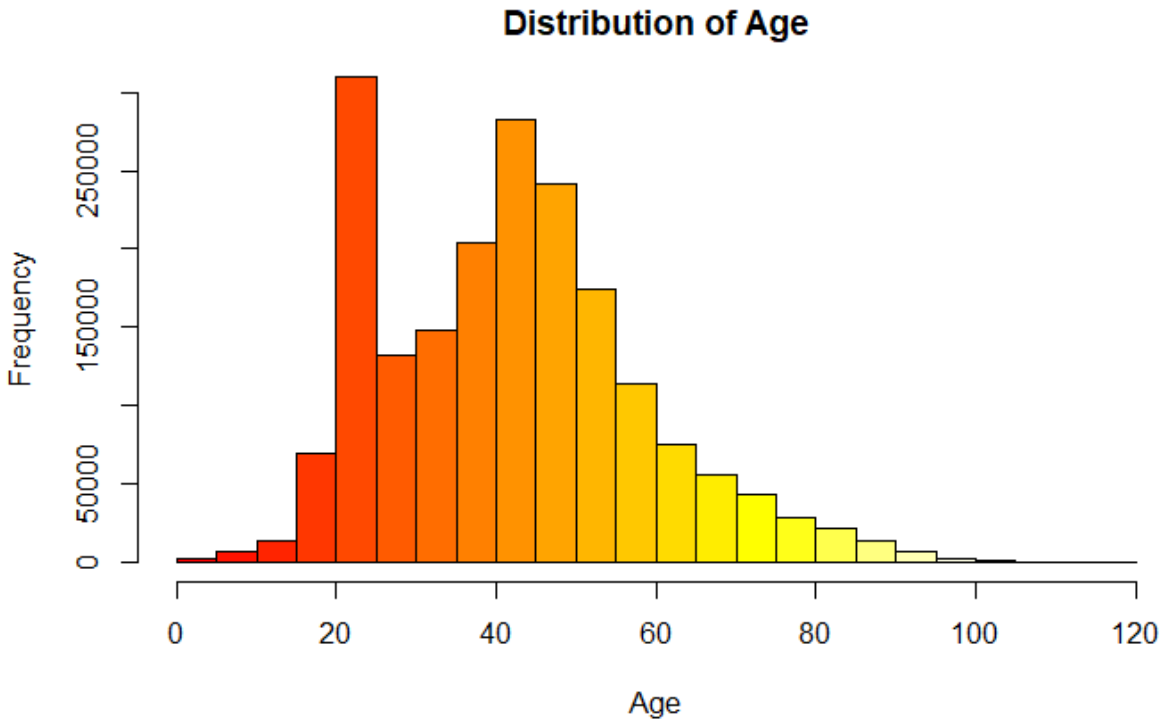

Figure 4. The age distribution for the active customers

Determining the most popular product is very useful, as recommendation can be made according to the top N list of the most popular items. It can be observed that the most common bank product is Curent Account, followed by Direct_Debit, Particular Account, e-account and ayroll Account.

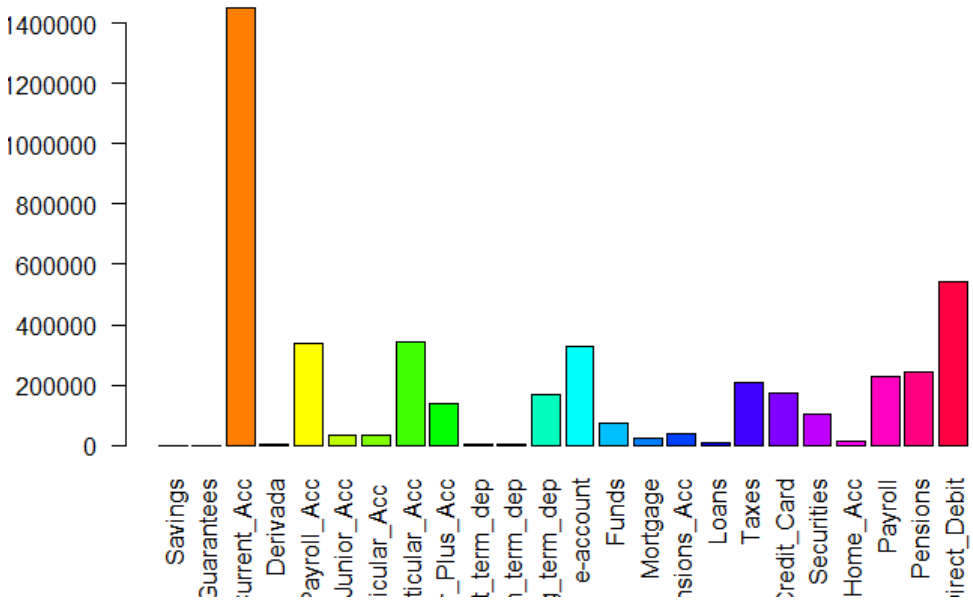

Figure 5. The representation of the most popular products

The household income can be a variable that influences how many products a customer has at the bank. A representation of the household income per age per segment reveals that the highest number of customers with high income and medium income belong to the "Particulares" segment.

There are few high-income customers in "Universitario" segment. By contrary, those customers who didn't belong to any segment, were classified as "Other" and they are mainly clients with lower income.

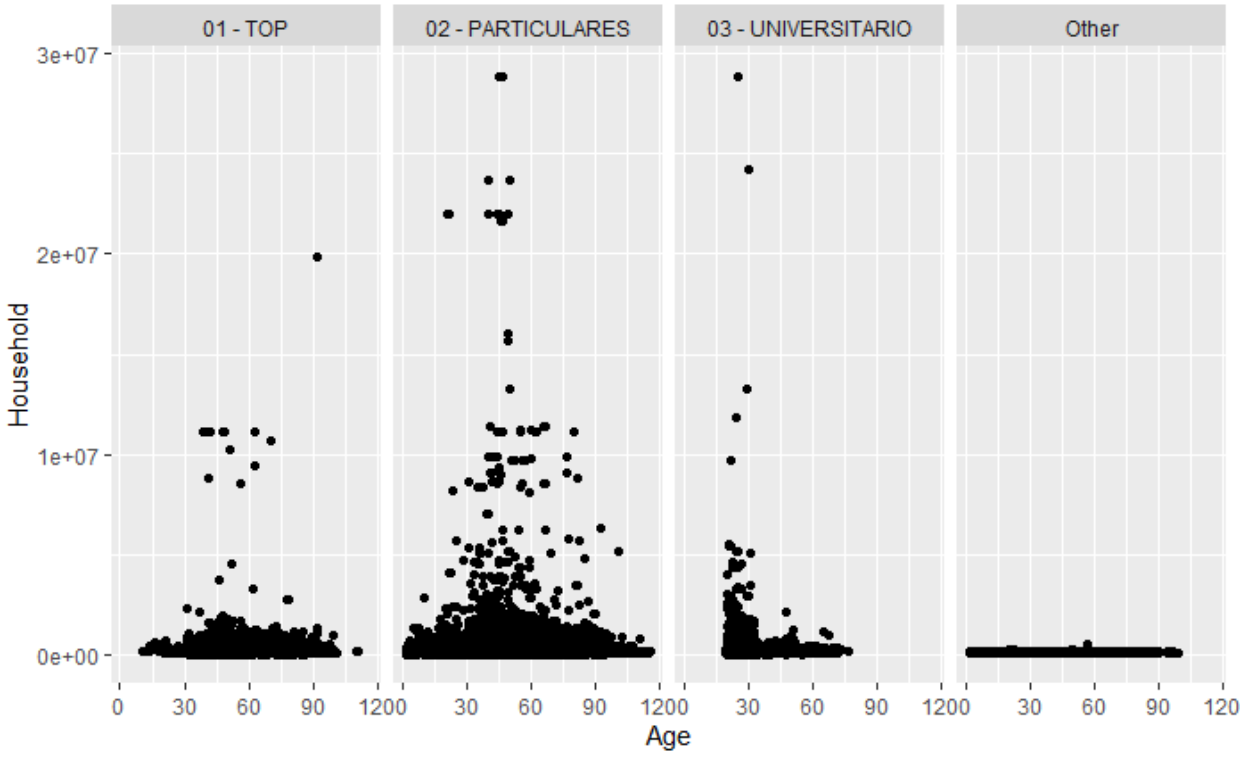

Figure 6. The representation of the household income per age per segment