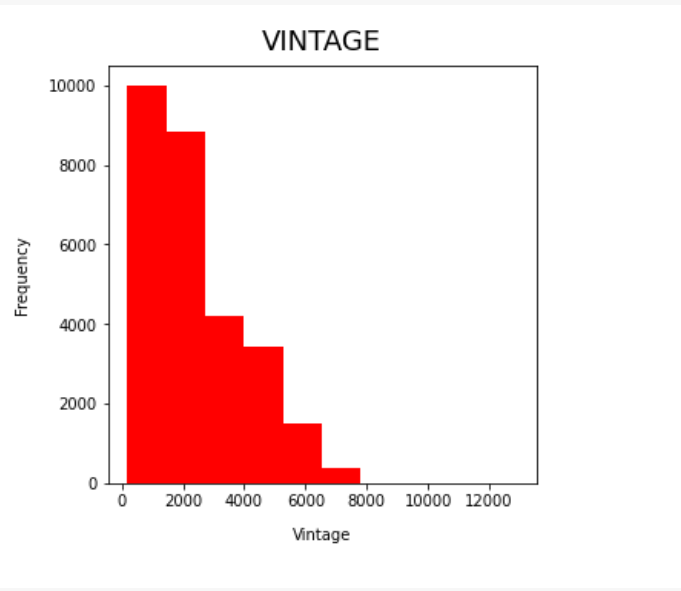
**Exploratory data analysis**

In [statistics](https://en.wikipedia.org/wiki/Statistics), exploratory data analysis is an approach to analyse [data sets](https://en.wikipedia.org/wiki/Data_set) and summarize their main characteristics, often with visual methods. A [statistical model](https://en.wikipedia.org/wiki/Statistical_model) can be used or not, but primarily EDA is for seeing what the data can tell us beyond the formal modeling or hypothesis testing task.

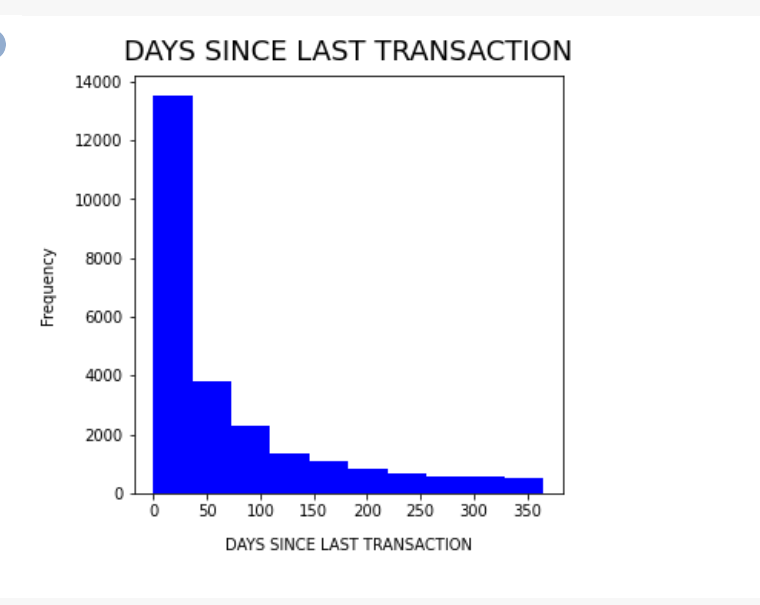
Histogram:

A histogram is used to provide a visual interpretation of numerical data by showing the number of data points that fall within a specified range of values.

1)Analysis of the vintage of all the customers has been done.

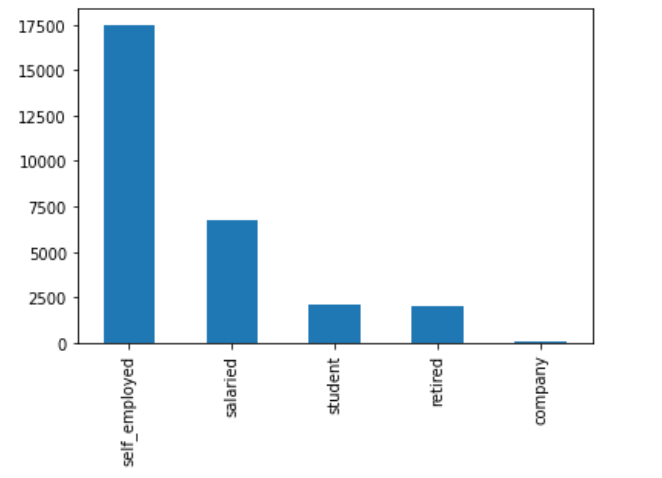


2) Plotting the days since last transaction for different customers and analysing it with the number of customers.



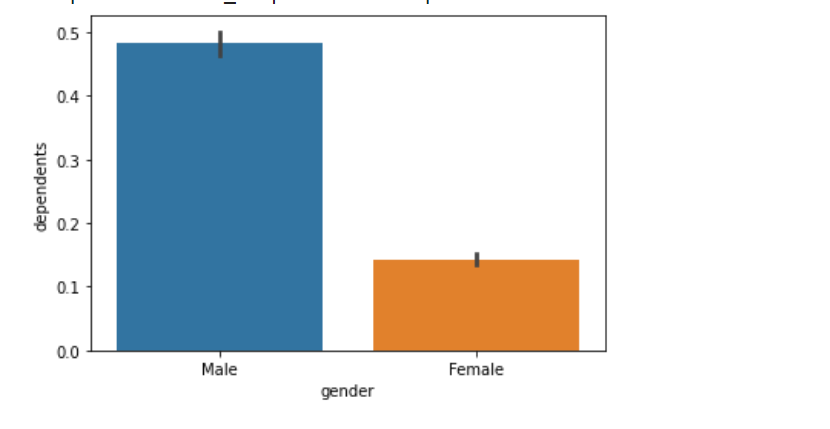
Bar Chart:

A bar chart or bar graph is a chart or graph that presents categorical data with rectangular bars with heights or lengths proportional to the values that they represent.

1)A bar chart with x and y axis as occupation and frequency respectively.  


On observing the bar chart, we can see that the number of customers in the bank who are self-employed are the greatest and company is the least.

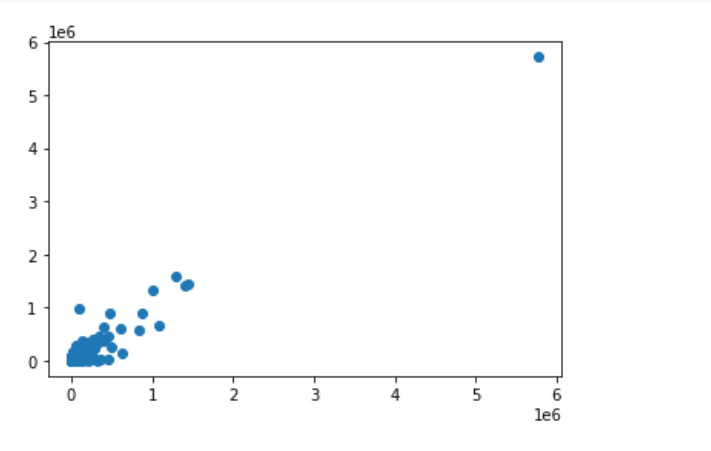
2)The following bar chart has been plotted with the gender on x axis and dependents on the y axis



We observe that men have more dependents than women.

Scatter Plot

A scatter plot shows whether two variables are correlated and shows a pattern in the relationship that cannot be seen by just looking at the data.

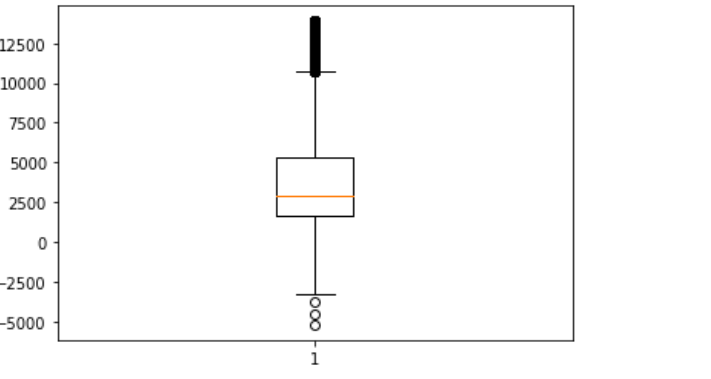


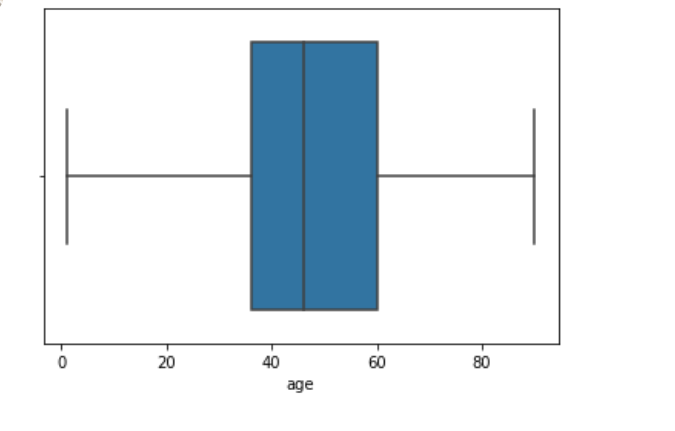
The scatter plot has been plotted between the quantities of current month balance and previous month balance.

Box Plot:

The box plot has been used to summarize the value age in the dataset in the form of an interval scale.

Implementing box plot using mathplotlib.

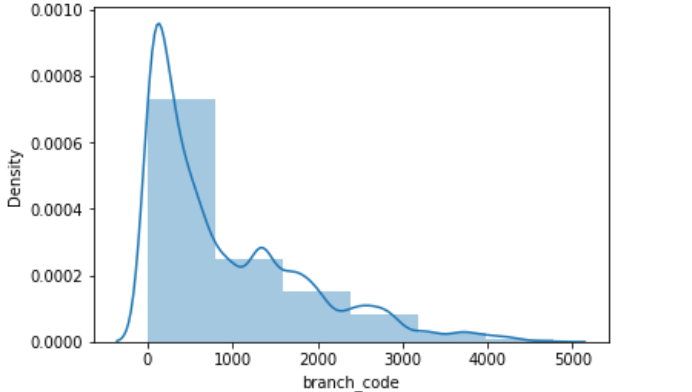


Implementing box plot using seaborn

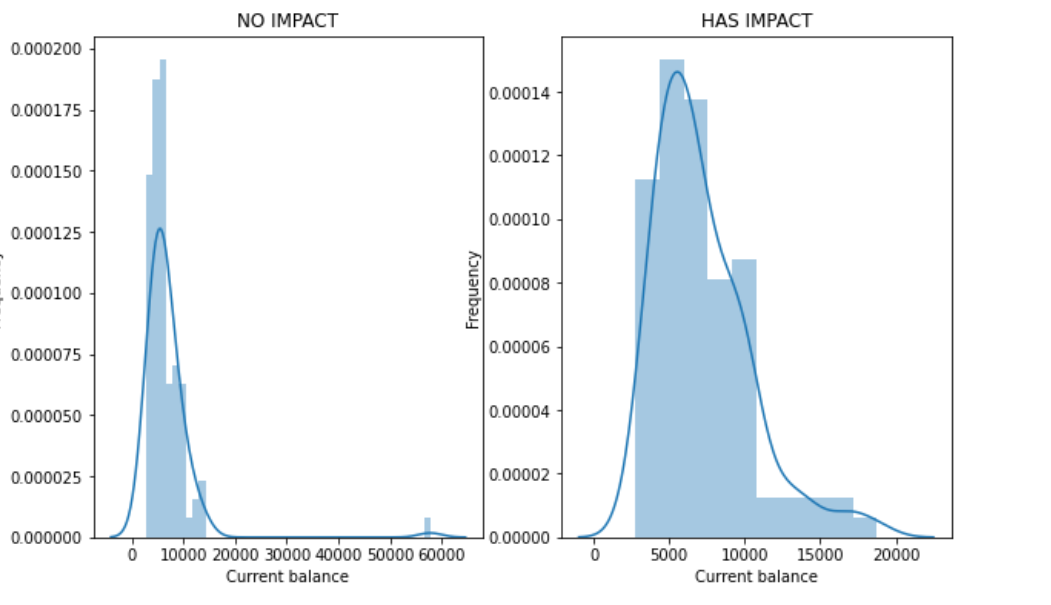
Seaborn Distplot.

Seaborn distplot depicts a histogram with a line on it.It plots a univariate distribution of the observations.

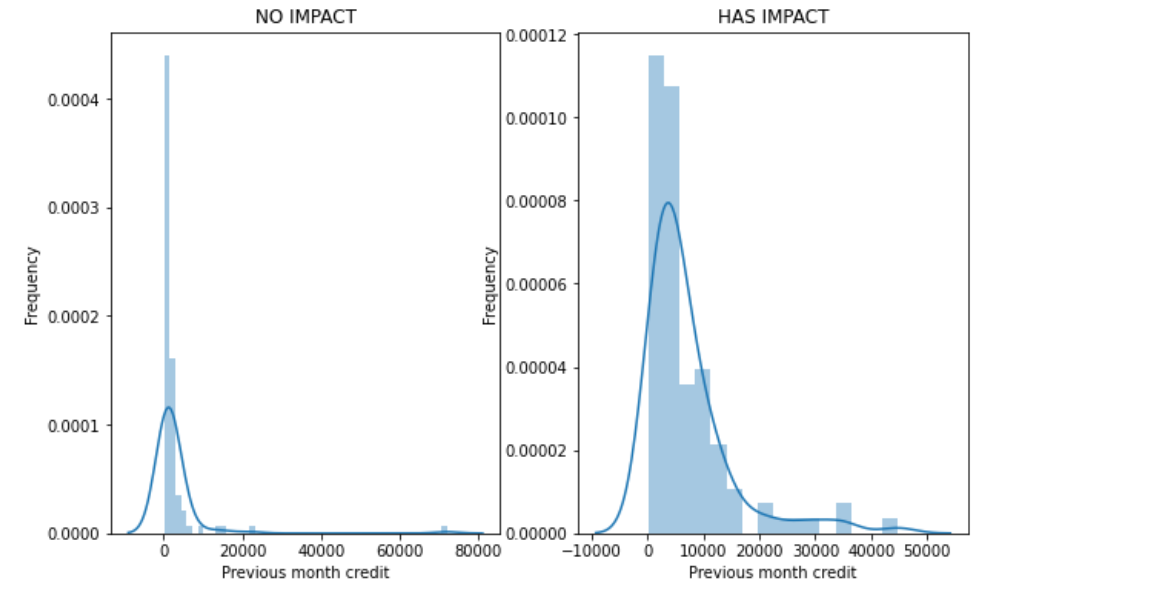
1)Distplot on branchcode.



2)Distplot fr the impact and absence of impact of the number of dependents on current balance.

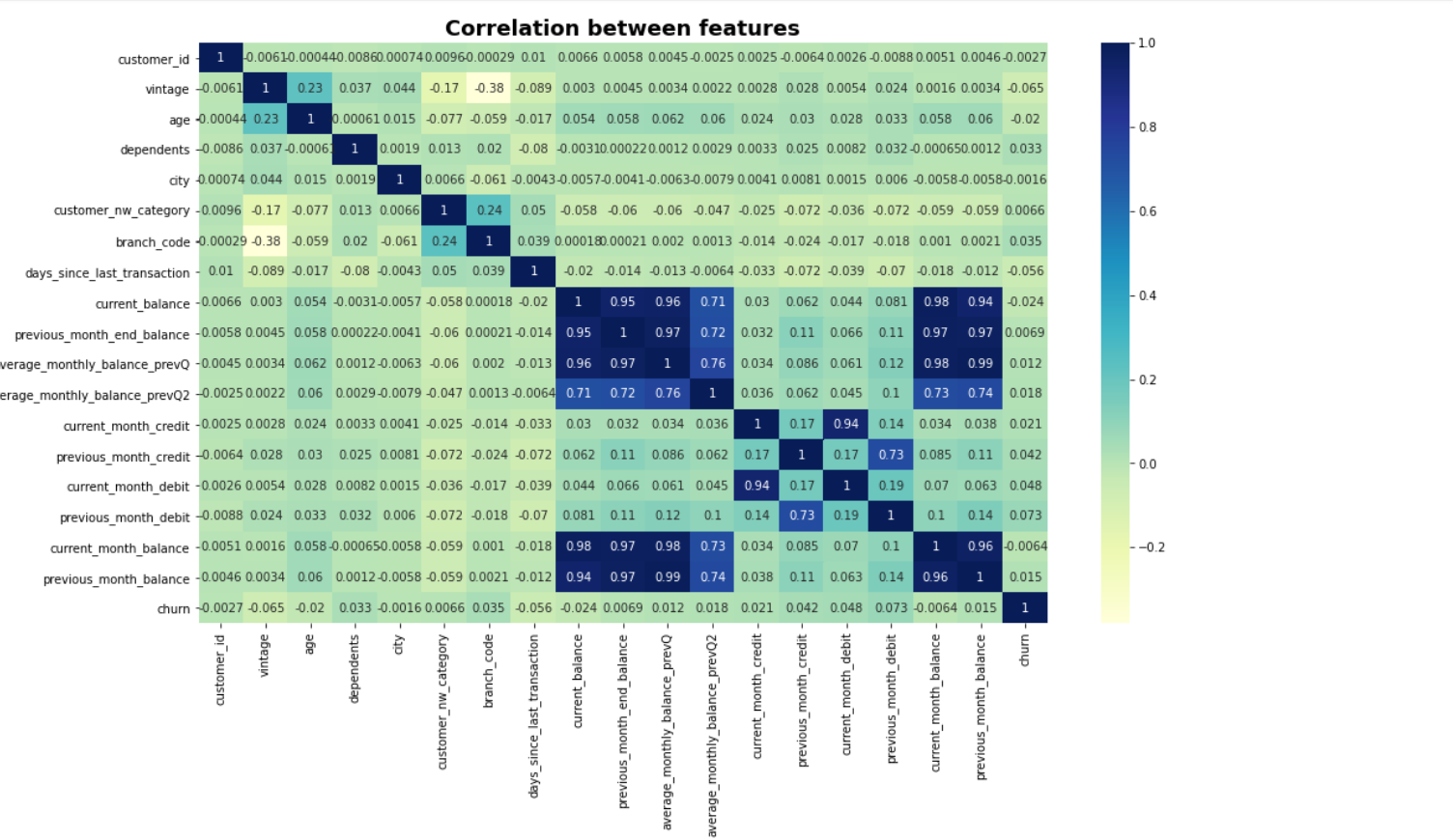


3)Distplot for the impact and absence of impact of churn on previous month credit.



Correlation between features represented using Heat Map.

A heatmap is a graphical representation of data in which data values are represented as colors



There is a high positive relation between terms such as:

current balance and average monthly balance prevQ .

previous month balance and current balance.

There is a negative correlation between:

Vintage and branch code

Dependents and days since last transaction.

RESULTS AND DISCUSSION:

The number of dependents are more for men than women among the customers.

Among all occupations,the customers who are self employed are the greatest in proportion when compared with other occupations.

The churn has a major impact on the previous month credit.