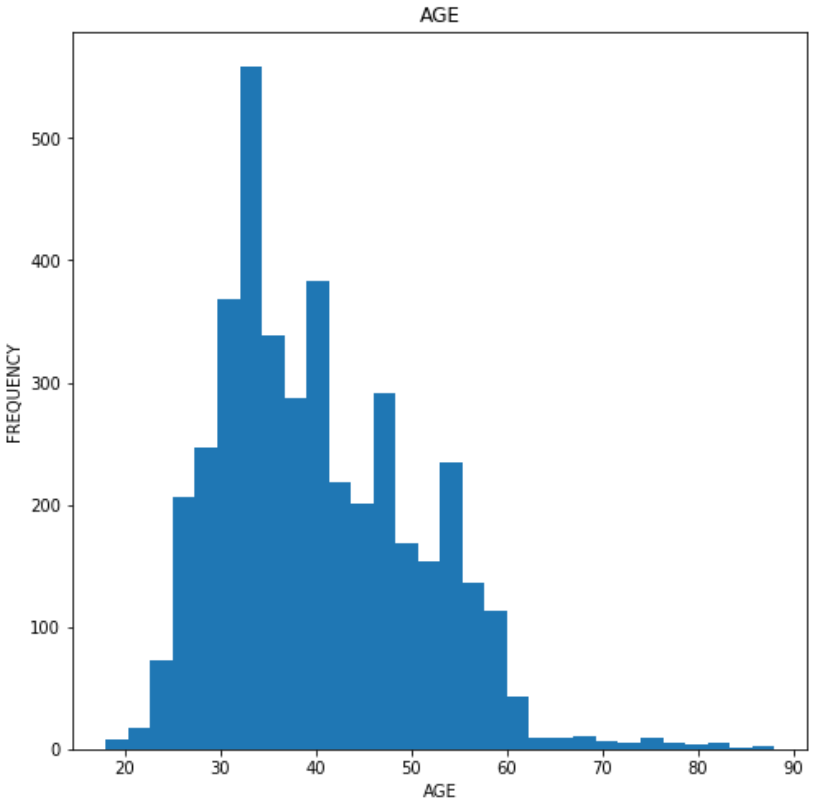
Data Preparation

1) This task was addressed by first putting the bank.csv file in the home directory. Then in jupyter notebook this .csv was imported and read using the pandas and made sure that correct arguments were used.
2) Made sure that the loaded data was equivalent by displaying the dataframe. Ensured that each attribute had proper data type and changed for those which needed appropriate data types.
3) Checked for typos, i.e., checked for any spelling errors and corrected them by using the replace function.
4) Removed white spaces from the data by using the strip function.
5) Demonstrated how to cast text data to lower case by first converting the text data into upper case and then converted to lower case by using upper and lower functions accordingly.
6) Checked for impossible values in each attribute and removed them by using the drop command. Removed age rows where age was greater than 100, dropped campaign where the values were 999.0, dropped pdays which was in negative value, dropped emp.var.rate which was in negative.
7) Checked for missing values or null values and replaced them by the mean the attribute accordingly.

Data Exploration

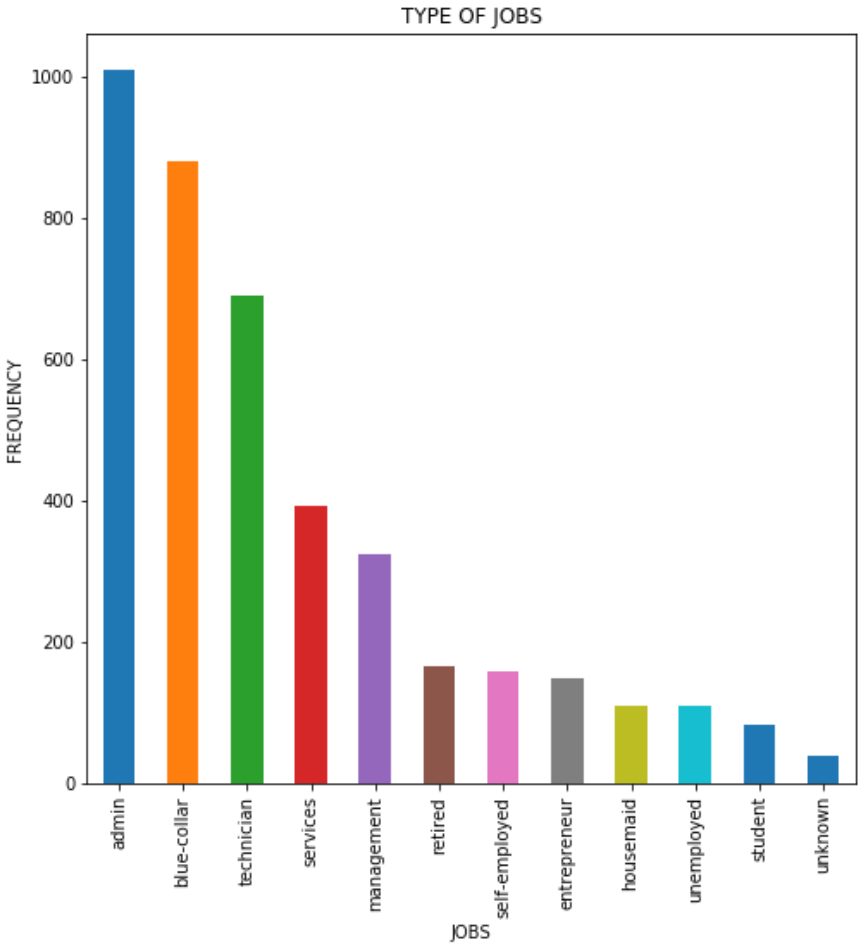
1) Step 1

1.1. HISTOGRAM FOR AGE ATTRIBUTE



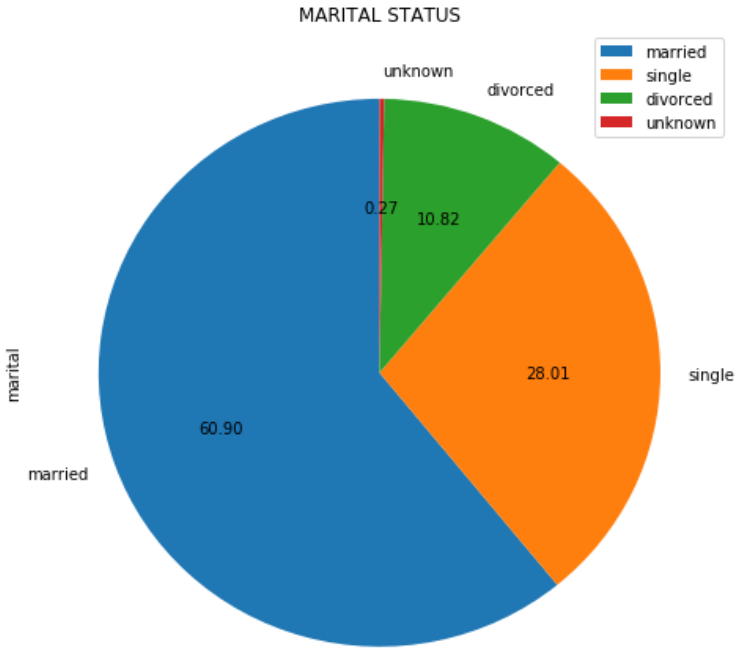
Histogram provides a better visualization for age attribute. For example it shows that there are more people between age 32 and 45. X-axis gives the age and y-axis gives the number of people. In this way we can predict various desired outcomes

1.2. BAR GRAPH FOR JOB ATTRIBUTE



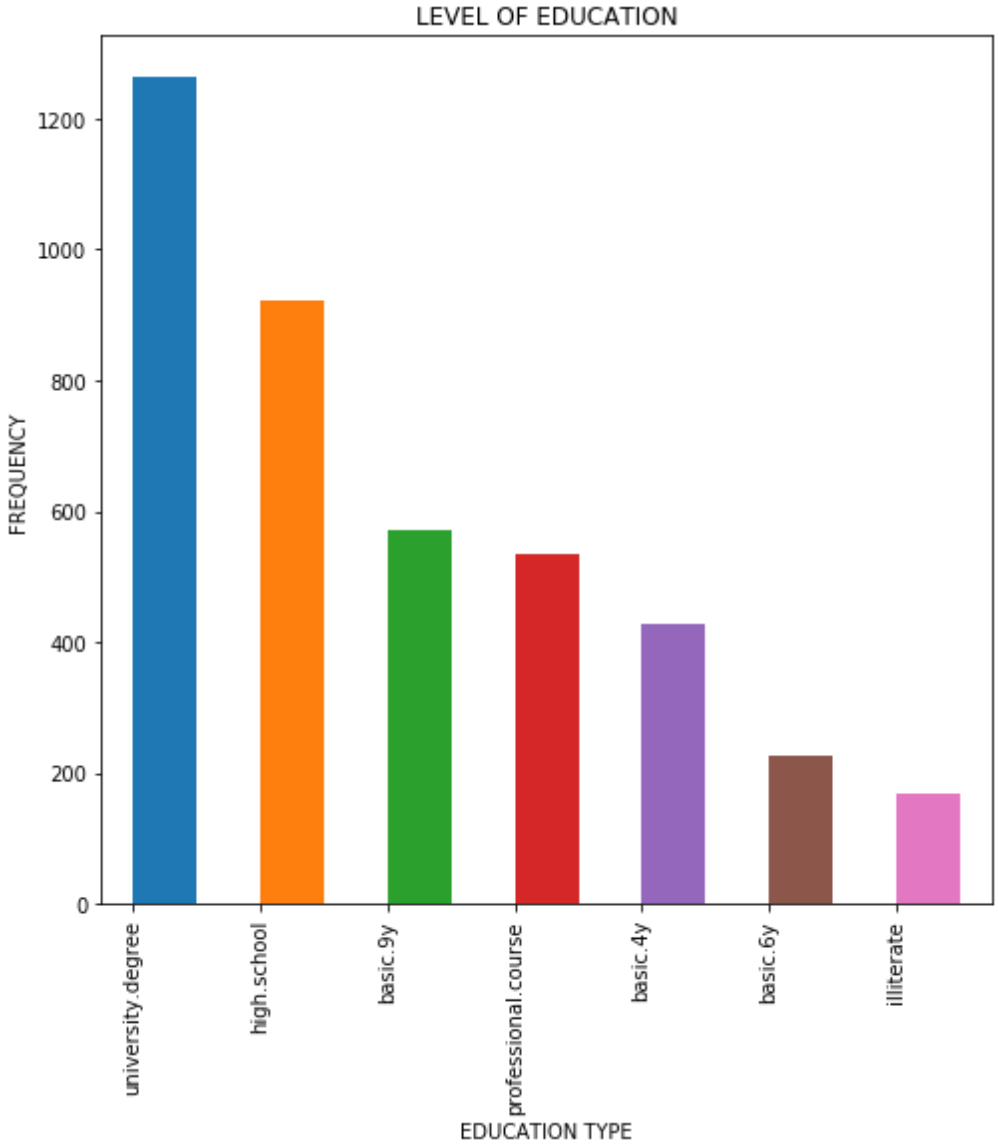
Bar Chart provides a good representation of number of people in every job type. . X-axis gives the type of jobs and y-axis gives the number of people It shows that admin job type have more people.

1.3. CHART FOR MARITAL STATUS



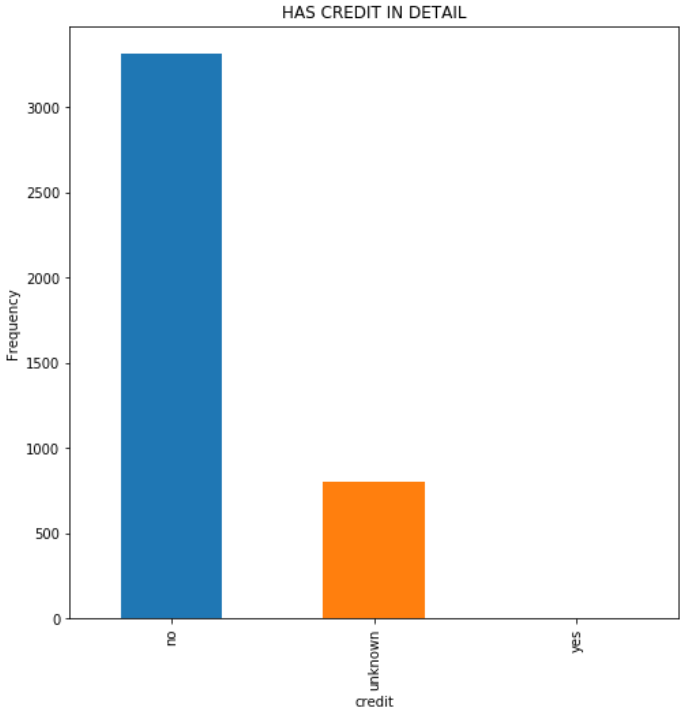
Pie chart provides a better representation for marital status. It clearly shows the percentage of people in every category. For example, there are 60.90 of the people married.

1.4. BAR GRAPH FOR EDUCATION ATTRIBUTE



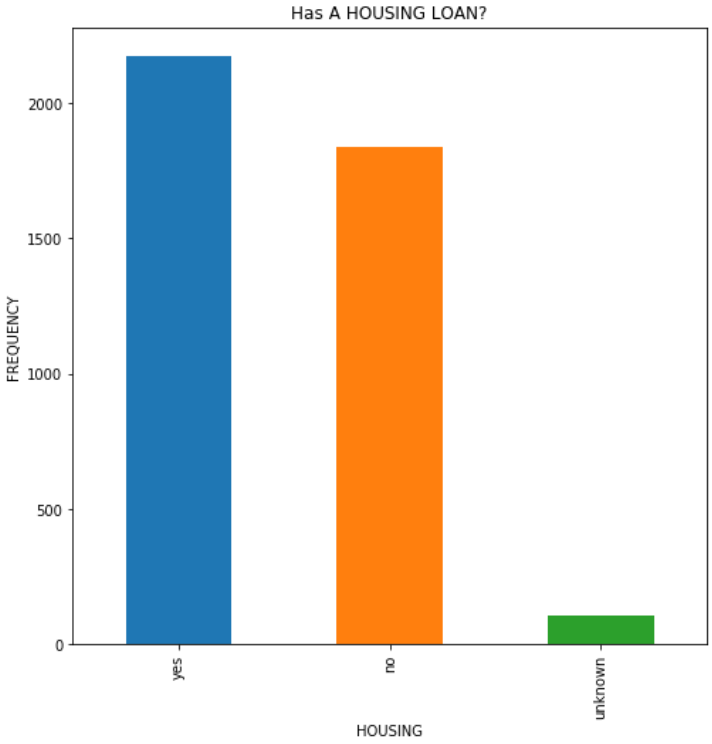
Bar graph gives clear presentation number of people in every education type. . X-axis gives the Education Type and y-axis gives the number of people
For example, there are many people in university degree.

1.5. BAR GRAPH FOR DEFAULT ATTRIBUTE



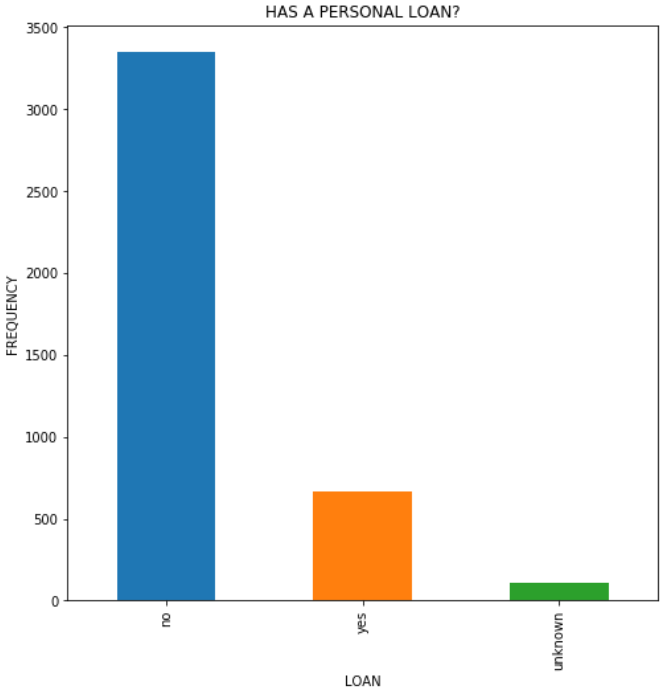
Since it is a categorical attribute, plotting a bar graph gives a better presentation and tells clearly how many people have credit. . X-axis gives if there is credit and y-axis gives the number of people

1.6. GRAPH BAR FOR HOUSING ATTRIBUTE



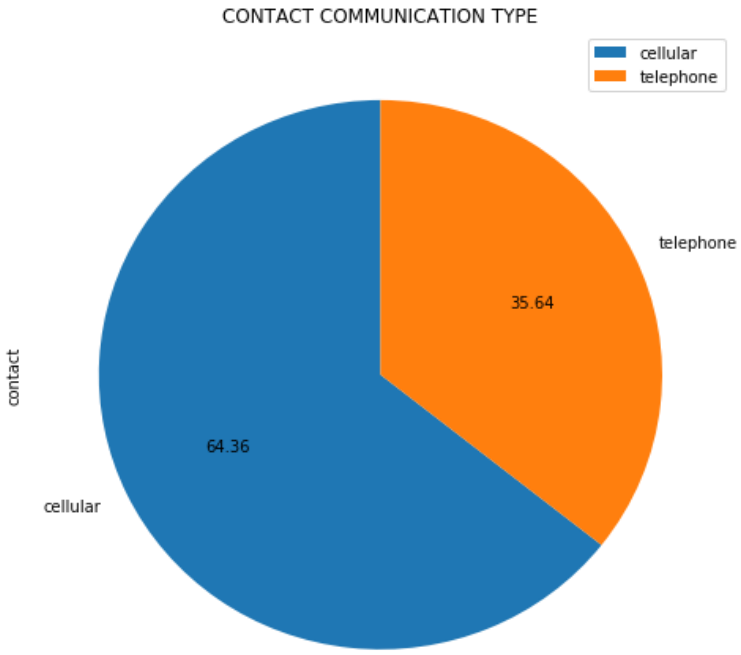
Since it is a categorical attribute, plotting a bar graph gives a better presentation and tells clearly how many people have housing loan. . X-axis gives if people have loan and y-axis gives the number of people.

1.7. BAR GRAPH FOR LOAN ATTRIBUTE



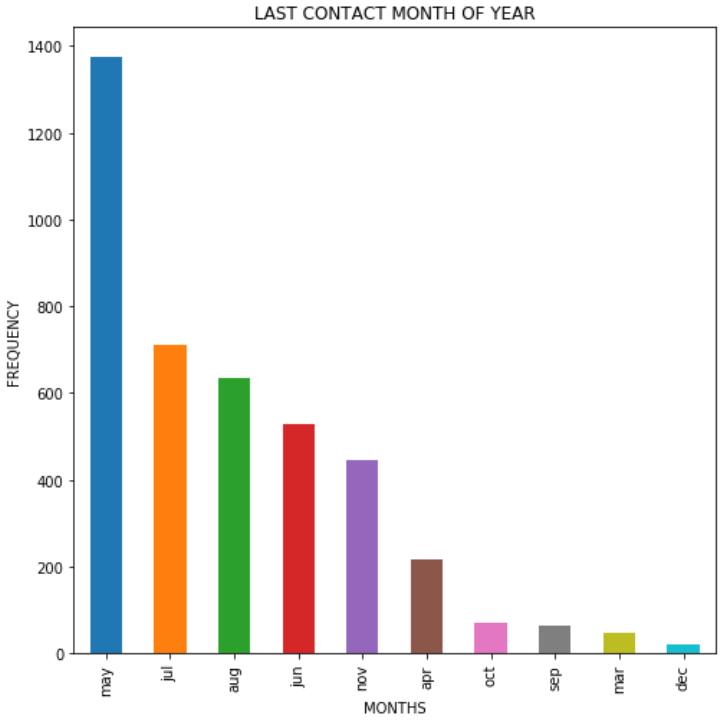
Since it is a categorical attribute, plotting a bar graph gives a better presentation and tells clearly how many people have personal loan.

1.8. PIE CHART FOR CONTACT ATTRIBUTE



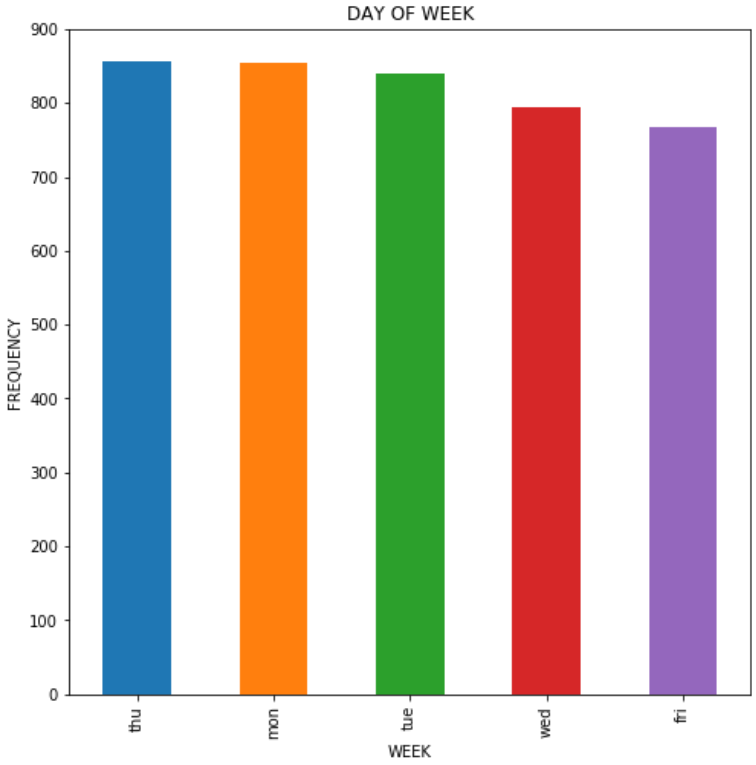
Pie chart gives a clear presentation of communication type, which shows the percentage of people using cellular and telephone. For example, 64.36% use cellular type of phone which is more than half the population.

1.9. BAR GRAPH FOR MONTH ATTRIBUTE



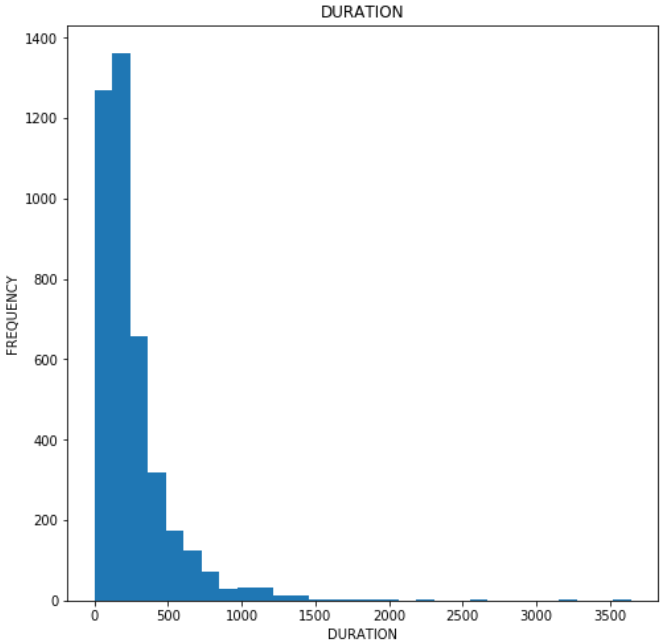
Bar plot for this attribute gives a clear representation of number people contacted every month. . X-axis gives the month and y-axis gives the number of contacts made.

1.10. BAR GRAPH FOR DAY OF WEEK ATTRIBUTE



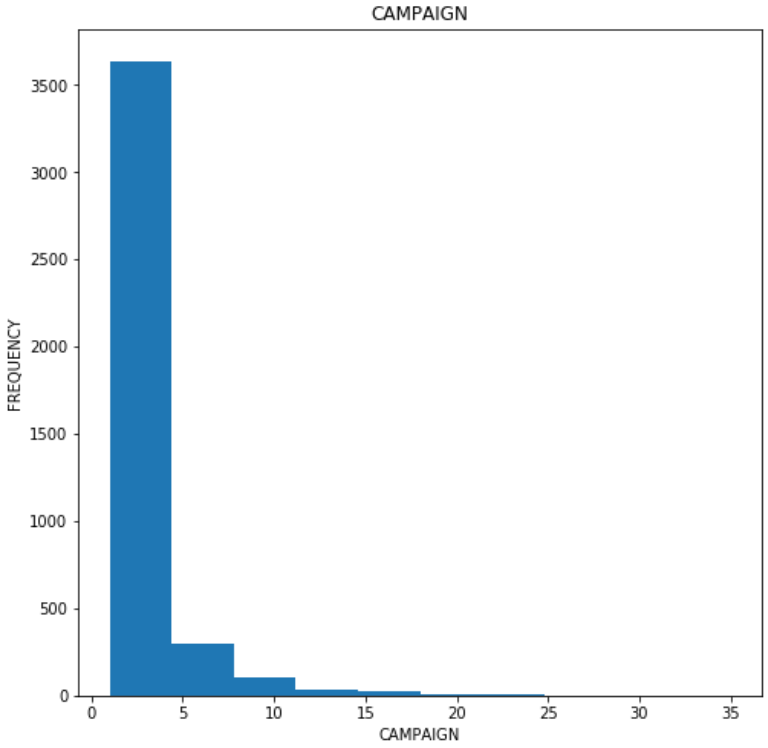
Bar plot for this attribute clear representation of number of people last contacted on a day. . X-axis gives the day of week and y-axis gives the number of people last contacted.

1.11. HISTOGRAM FOR DURATION ATTRIBUTE



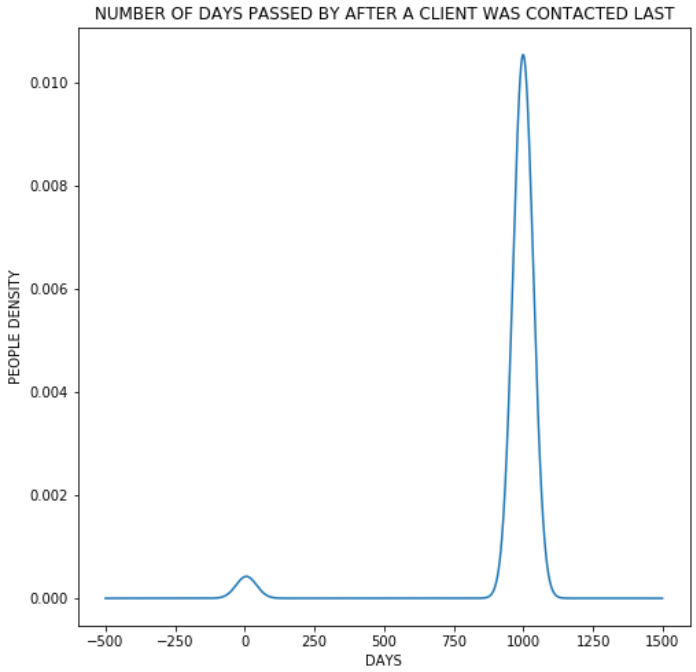
Histogram provides a clear visualization of duration time, which says the duration of call in seconds of last contact. . X-axis gives the duration in seconds and y-axis gives the number of people.

1.12. HISTOGRAM OF CAMPAIGN ATTRIBUTE



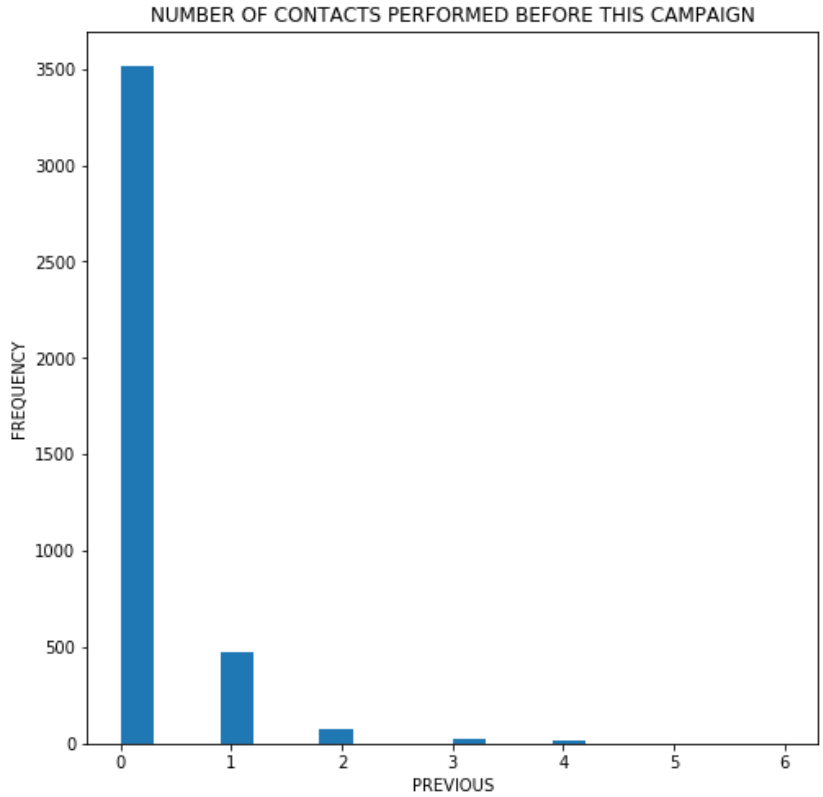
Histogram represents clearly on number of people last contacted in the campaign. . X-axis gives the campaign and y-axis gives the number of people.

1.13. DENSITY PLOT FOR PDAYS ATTRIBUTE



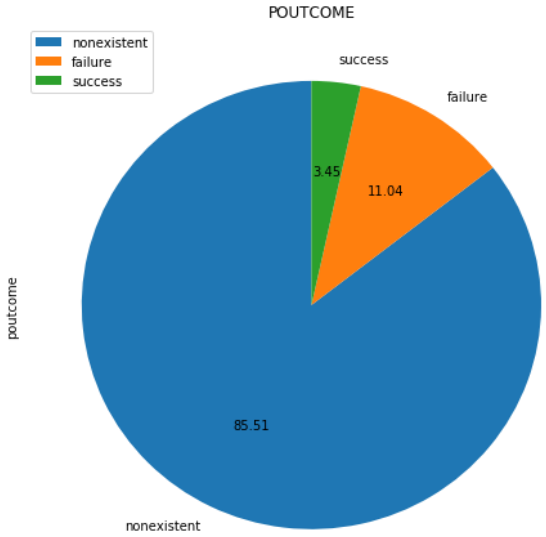
Density plot gives the number of days passed by after a client was contacted last. . X-axis gives days and y-axis gives the density of people.

1.14. HISTOGRAM PLOT FOR PREVIOUS ATTRIBUTE



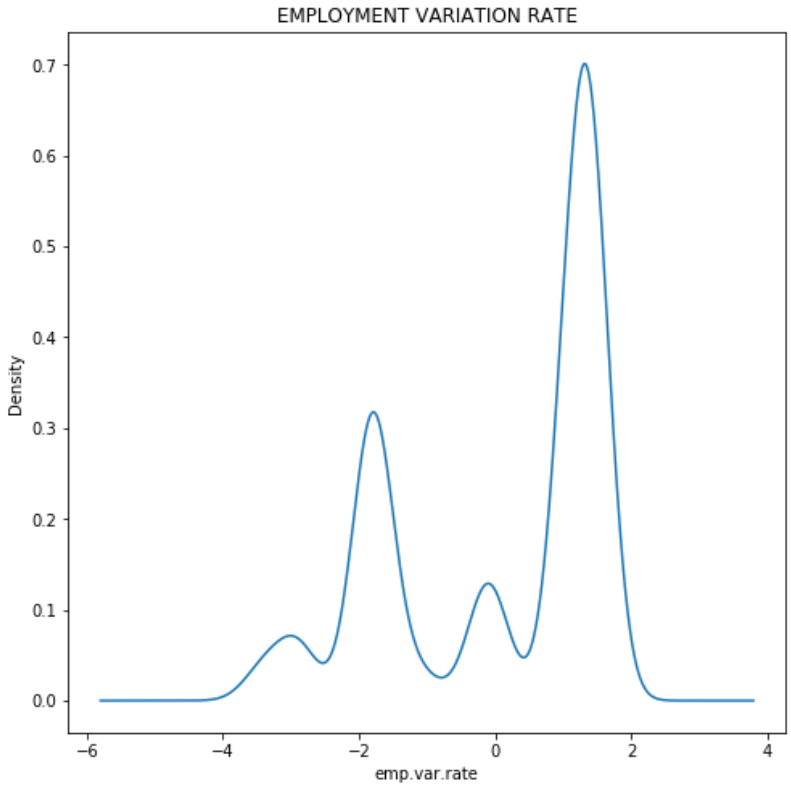
Histogram plot provides the number of contacts performed before this campaign. . X-axis gives number of contacts and y-axis gives the number of people.

1.15. PIE CHART FOR POUTCOME ATTRIBUTE



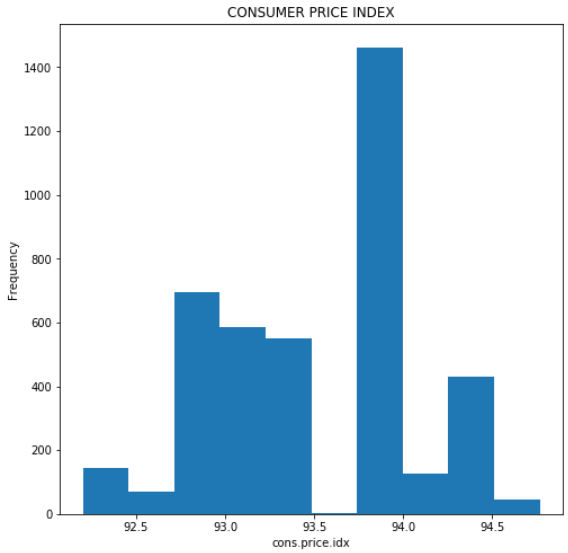
Pie chart provides the clear percentage of outcome of the previous marketing campaign. For example, only 3.45% is the success rate of poutome.

1.16. DENSITY PLOT FOR EMPLOYMENT VARIATION RATE



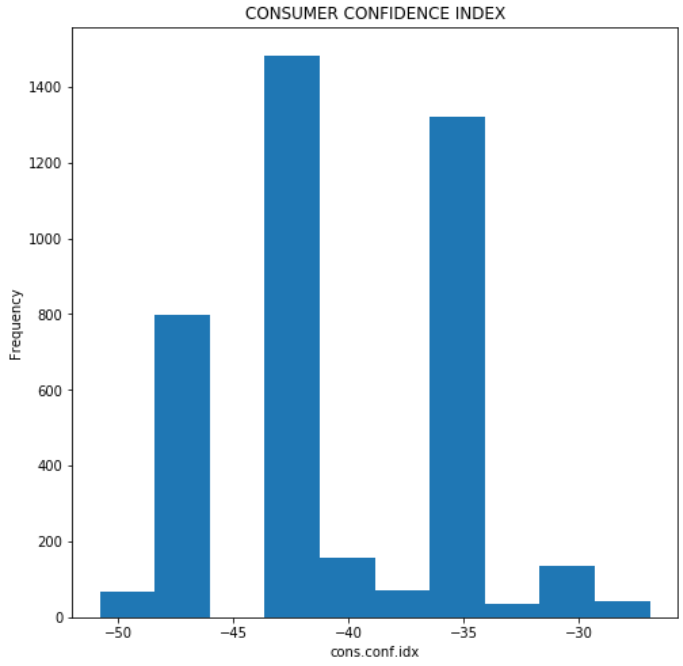
Density plot gives a clear representation of employment variation rate, which becomes easy for visualization and to predict the outcome. . X-axis gives employment variation rate and y-axis gives the density of people.

1.17. HISTOGRAM PLOT FOR CONS.PRICE.IDX



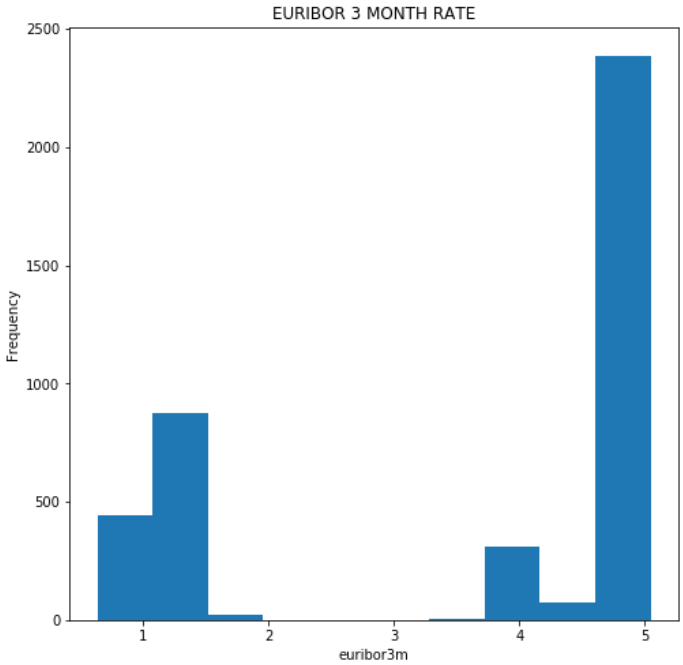
Consumer price index gives the changes in the price level and a histogram clearly plots the attribute. . X-axis gives consumer price index and y-axis gives the number of people.

1.18. HISTOGRAM PLOT FOR CONS.CONF.IDX



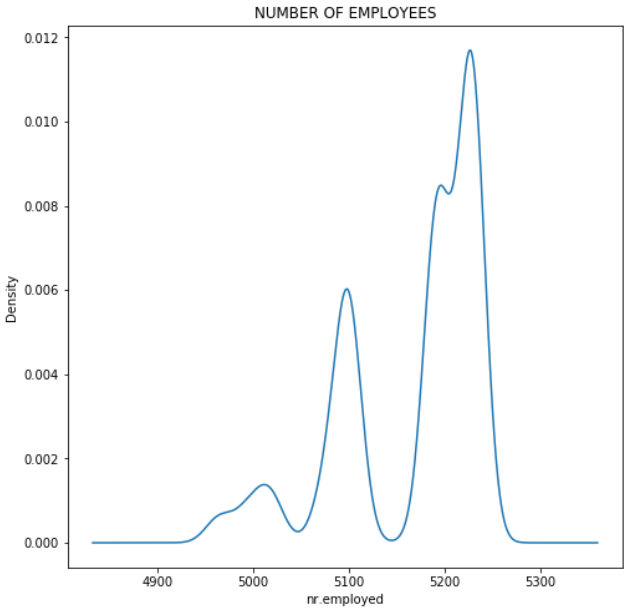
Depending upon customers savings and spendings this chart gives the consumer confidence index which is monthly indicator.. X-axis gives the consumer confidence index and y-axis gives the number of people.

1.19. HISTOGRAM PLOT FOR EURIBOR3M ATTRIBUTE



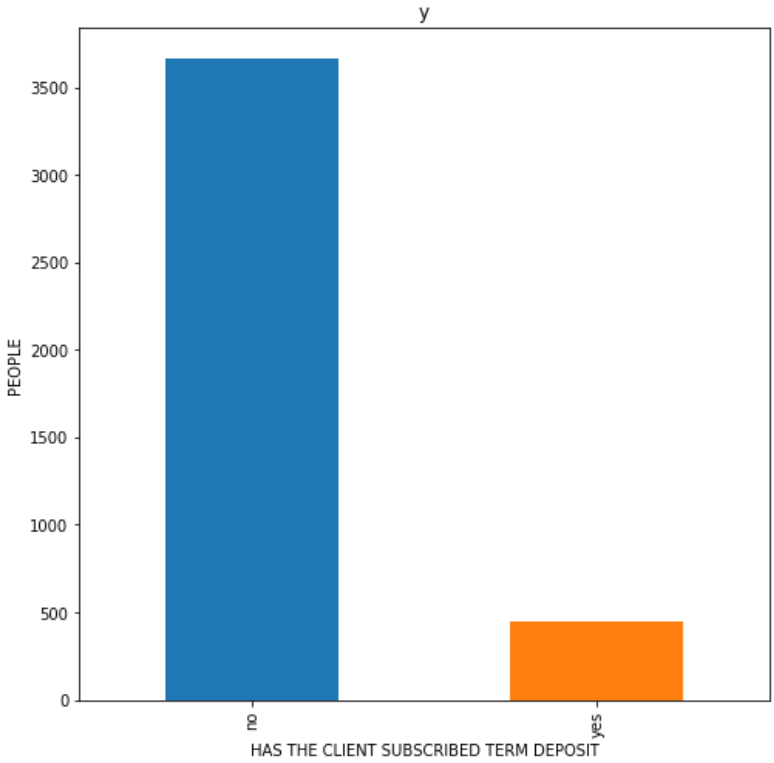
This histogram gives the averaged interest rates of 3 months on a daily indicator. X-axis gives euribor 3 month rate and y-axis gives the frequency.

1.20. DENSITY PLOT FOR NR.EMPLOYED ATTRIBUTE



This density plot gives the number of employees employed on a quarterly indicator. . X-axis gives number of people employed and y-axis gives density.

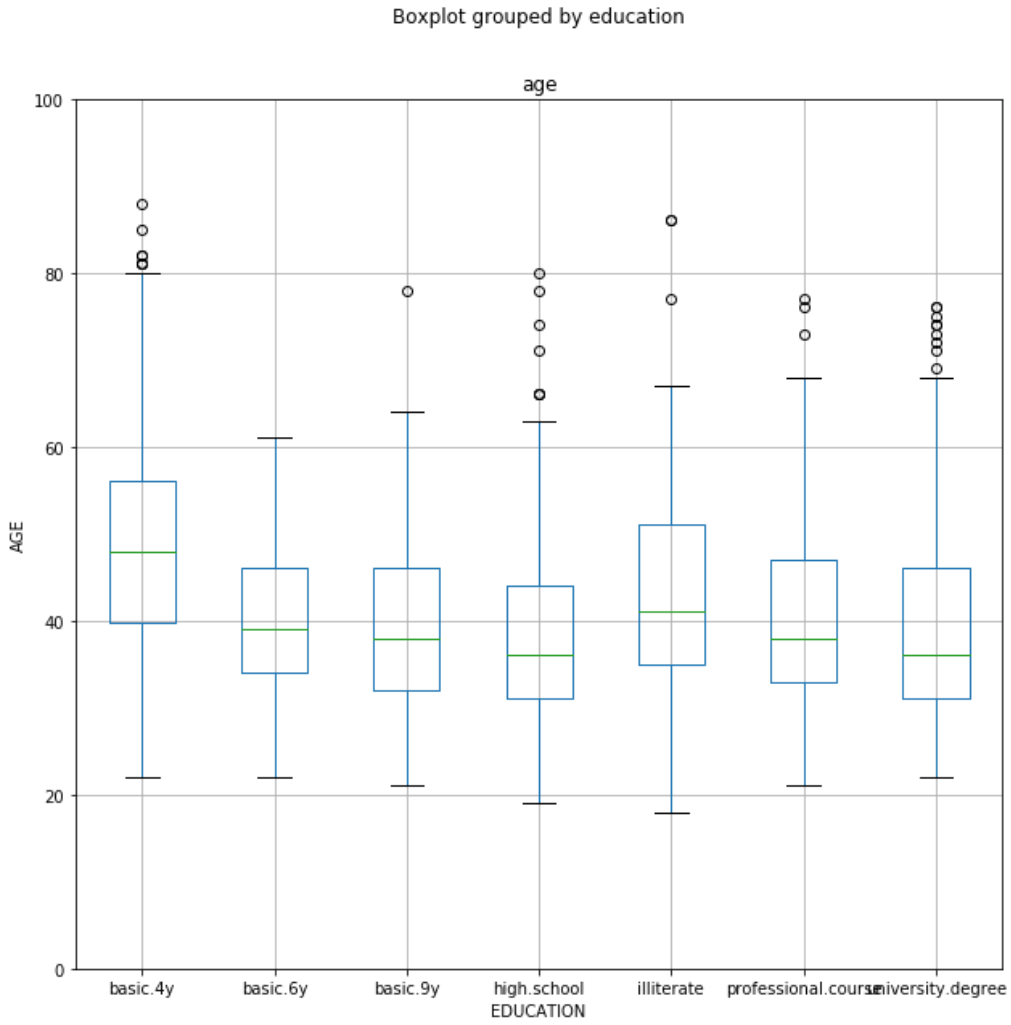
1.21. BAR GRAPH FOR y ATTRIBUTE



Being a categorical value this bar graph says how many people have subscribed to the term deposit. . X-axis gives whether the client has subscribed term deposit and y-axis gives the number of people.

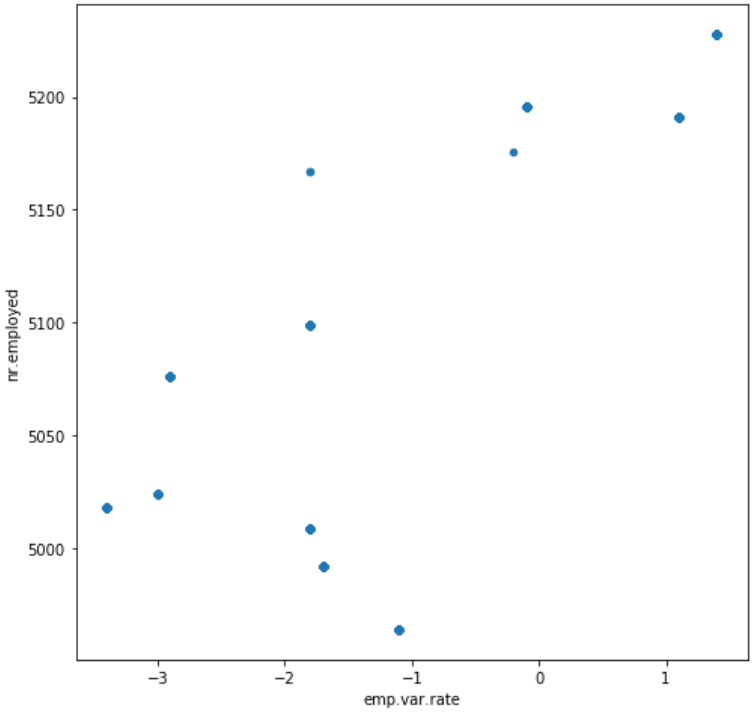
2)

2.1. BOX PLOT FOR AGE BY EDUCATION RELATIONSHIP



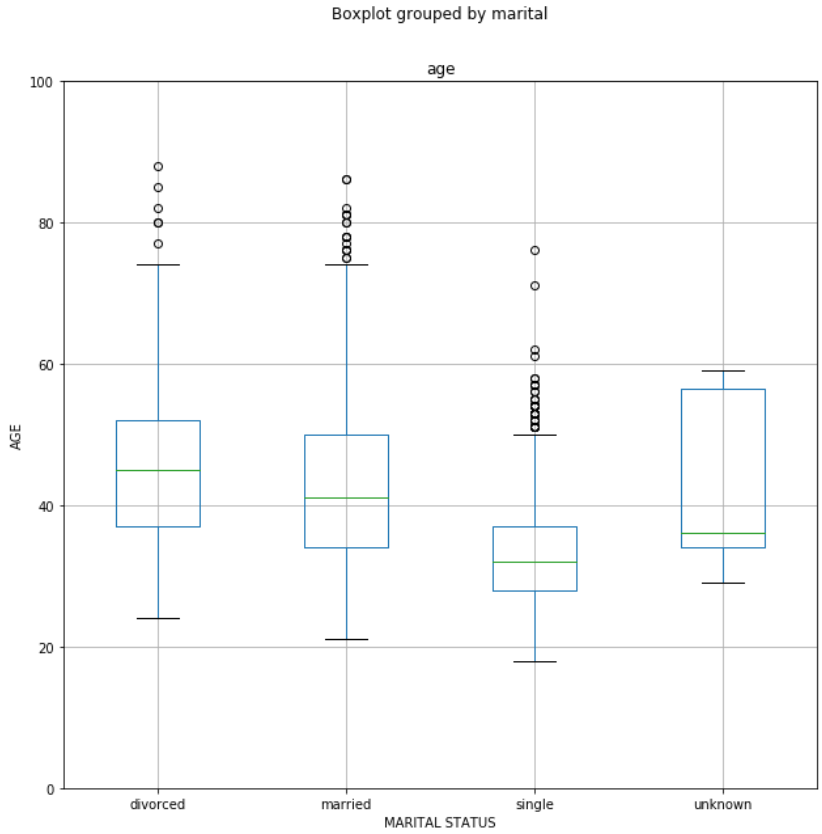
This boxplot gives the relationship between ages by education. X-axis represents the type of education and y-axis represents the age. It shows that the age group 30-50 do their university degree but also there are people who have finished only year 4 between the age group 40-55. Similarly there are many people illiterate between 35-50. It shows there are mixed age group with different education levels.

2.2. SCATTER PLOT FOR EMPLYMENT VARIATION RATE AND NUMBER OF EMPLOYEES



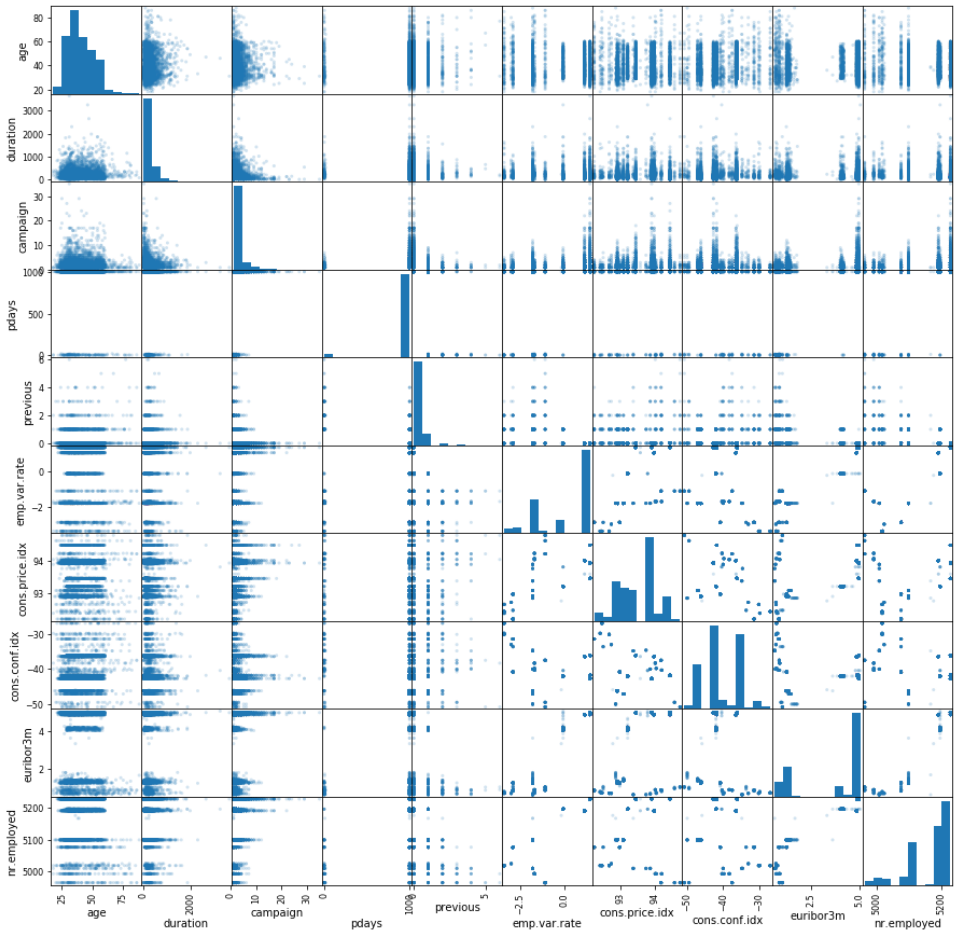
This scatter plot shows the relationship between employment variation rate and number of employees. It shows that as the employment variation rate increases, the number of employees also increase. After some point it shows that they are more than enough number of employees. Also with reduced employment variation rate, the number employees removed and added were pretty much okay. So in this way we can interpret the graph.

2.3. BOX PLOT FOR AGE BY MARITAL RELATIONSHIP



This box plot shows the relationship between age and marital status. X-axis represents the marital status and y-axis represents age. It shows that there are people between the age group 38-50 who are divorced and similarly there are people who are married in the same group. There are people who are single in the age group 25-38. Hence this way we visualize this graph.

3) SCATTER MATRIX



It is a scatter matrix representing all the numeric values.