Deliverables Week: 8
Bank Marketing Campaign

Group Name: Evolve Data
Name: Fredricka David
E-mail: frederickadavid@gmail.com
Country: Namibia
Company: Omdena
Specialization: Data Science**Github Repo Link: https://github.com/Fredricka23/Bank-Marketing-Campagain---Week-7**

Problem Description:

ABC Bank wants to sell it's term deposit product to customers and before launching the product to customers they want to develop a model which help them in understanding whether a particular customer will buy their product or not (based on customer's past interaction with bank or other Financial Institution).

Data Understanding:

We have information about 41188 clients in a csv file of size 583 Mb. For each client, 21 attributes are available in the data. Some demographic data such as:

1. age (numeric)
2. job : type of job (categorical)
    - admin
    - blue-collar
    - technician
    - services
    - management
    - retired
    - entrepreneur
    - self-employed
    - housemaid
    - unemployed
    - student
    - unknown
3. marital : marital status (categorical)
    - married
    - single
    - divorced
    - unknown
4. education (categorical)
    - university Degree

- high school
- basic.9y
- professional Course
- basic.4y
- basic.6y
- unknown
- illiterate

5. default: has credit in default?(categorical)
   - yes
   - no
   - unknown
6. housing: has housing loan?(categorical)
   - yes
   - unknown
   - no
7. loan: has personal loan?(categorical)
   - yes
   - no
   - unknown
8. contact (categorical): communication type
   - cellular
   - telephone
9. month:(categorical):last contact month of the year
10. day_of _week:(categorical): last contact day of the week (in working days)
11. duration :(numeric): last contact duration, in seconds
12. campaign :(numeric): number of contacts performed during this campaign and for this client
13. pdays :(numeric): number of days that passed by after the client was last contacted from a previous campaign
14. previous: (numeric): number of contacts performed before this campaign and for this client
15. poutcome: (categorical): outcome of the last campaign marketing
    - nonexistent
    - failure
    - success
16. emp.var.rate (numeric): employment variation rate - quarterly indicator
17. cons.price.idx (numeric): consumer price index - monthly indicator
18. cons.conf.idx (numeric): consumer confidence index - monthly indicator

19. euribor3m (numeric): Euro Interbank Offered 3-month rate - daily indicator
20. nr.employed (numeric): number of employees - quarterly indicator
21. y - has the client subscribed to a term deposit? (binary: 'yes','no')

Type of Data

1. Multivariate analysis of the data

There is correlation between the numeric variables. There is high correlation between employment variation rate, number of employees, Euro Interbank Offered 3-month rate and the consumer confidence index.

2. Categorical vs Categorical

Most of the features are categorical.

It would be suitable to do the comparison using feature engineering after data cleaning.

3. Numeric vs Categorical analysis:

There are 9 numeric columns and 11 categorical columns. Showing the correlation between them gives an insight of the data.

Problems in the data (number of NA values, outliers, skewed etc)

- Outliers - We calculate the z-score to detect outliers and we can also draw the box plot to check the eventual existence of outliers.
- NA values: There are no NA values in the data.

Approaches to overcome problems like NA value, outlier etc and why?

Removing outliers - We can remove outliers either by calculating the z-score or by calculating the interquartile range.

Using Z-score -the z-score is a numerical value that quatifies the relationship to the mean of of the values in our data. Z-score is measured in terms of standard deviations from the mean.