Data Science Project :

Bank Marketing Campain prediction Using Machine Learning

Data Science Project :

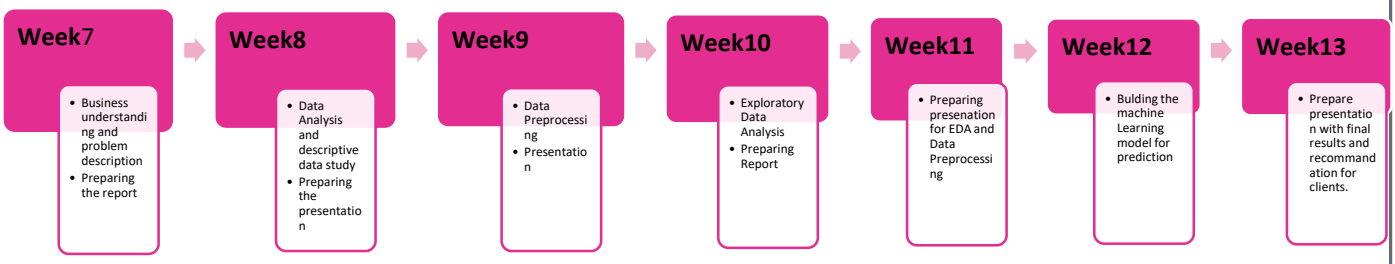
Plan:

- *Project details*
- *Project Process*
- *Problem Description*
- *Exploratory Data Analysis*
- *Recommandations*

Project Informations:

Project Roadmap and Process:

This process is prepared according to the needs of the company and the submission of each week.



Problem Understanding and Business Need of the Company:

ABC Bank wants to sell it's term deposit product to customers and before launching the product 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 and other financial Institution)

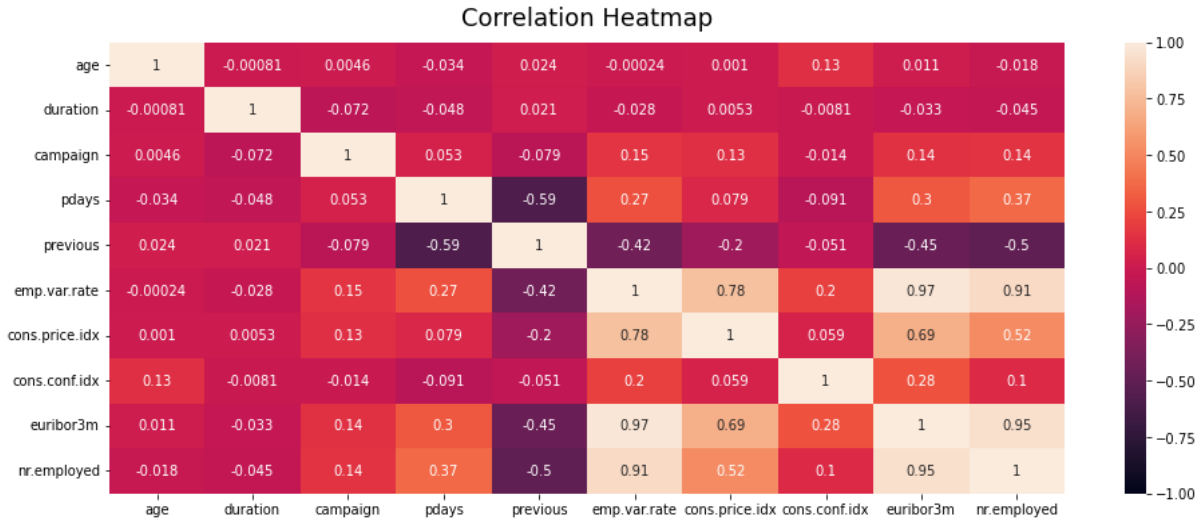
In order to achieve their goal and need they demand to help them with a AI model for prediction.

The company give me as a data scientist this mission.
I will develop a Ml Model to shortlist customer whose chances of buying the product is more so that their marketing channel can focus only on those customers whose chances of buying the product is more.

Exploratory Data Analysis

- Missing values : After verifiying missing values I conclude that we don't have missing values in our dataset.
- Duplication rows : The dataset contain 12 rows duplicated, I droped them using drop function.
- Categorical and numerical attributes analysis :
  For better analysis I started by working on categorical attributes than I moved to numerical data analysis.

Step1 Features correlations



The emp.var.rate, euriborn3m, nr.employed and cons.price.index are the 3 highly correlated attributes.

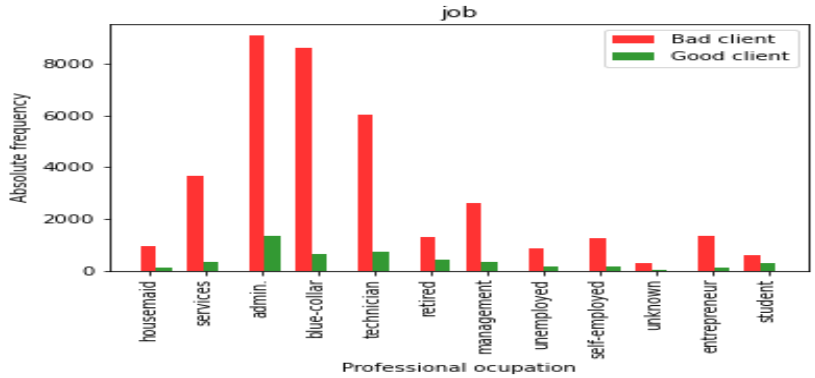
Step2 Categorical Attributes Analysis

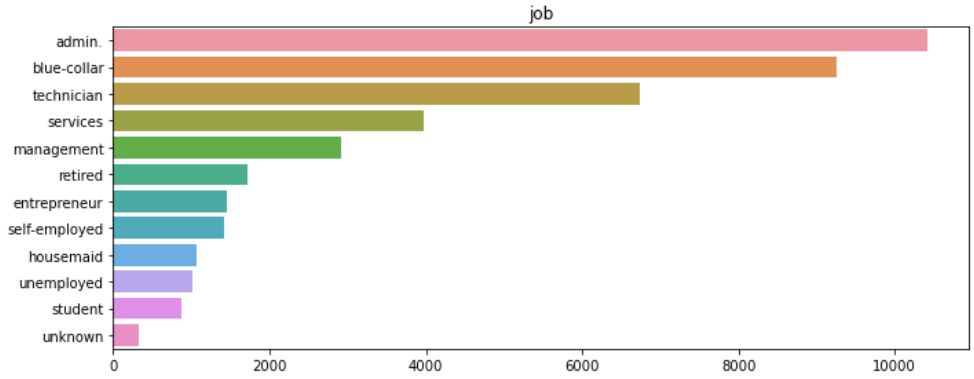
Categorical features are :

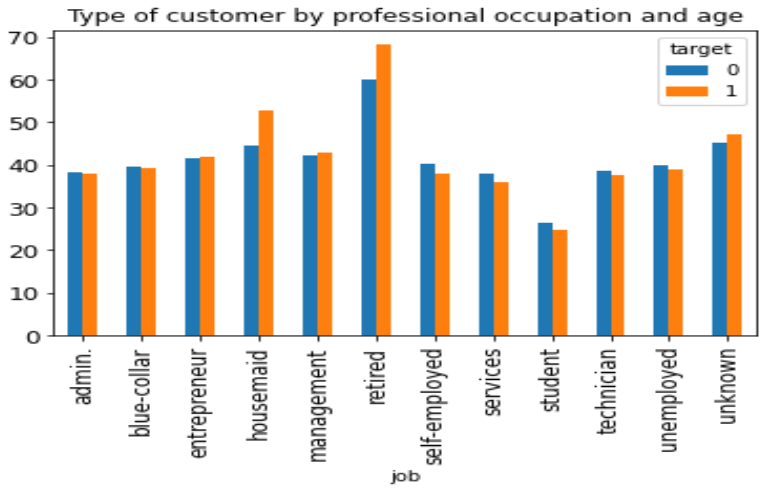
Target :y	•Job •Day of week
Education default	•Marital •Contact
Housing Month	•loan •Poutcome

Let's start by plotting visualization of attributes repartition by categories.

- 1 Customer repartition per job situation and age analysis

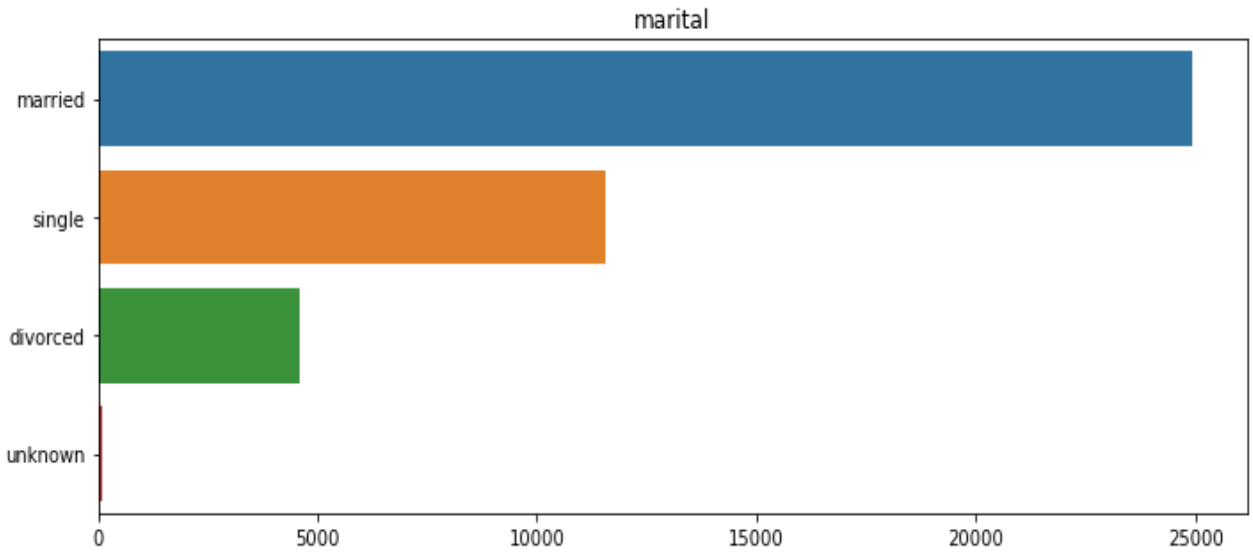


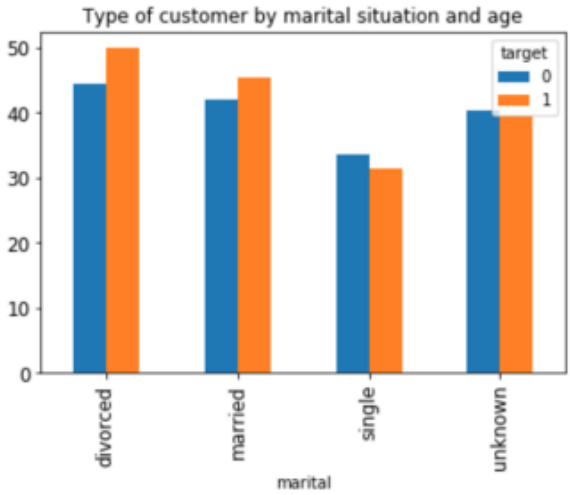




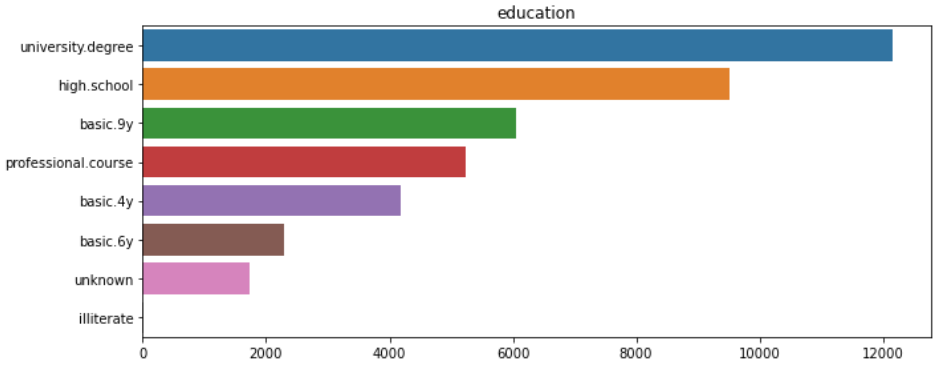
From the second graph, we can conclude that people with admin, blue_collar and technician jobs are more contacted by the bank.It can be explained by their social situation and the bank expectation.
From the first plot, we see student and retired are the more subscribed according to the total numbers of people by job category.
Below from the ttype of customer by job and age, we can observe that retired and housemaid are the oldest customers and the more classe that accepted to subscribe to the term deposit.

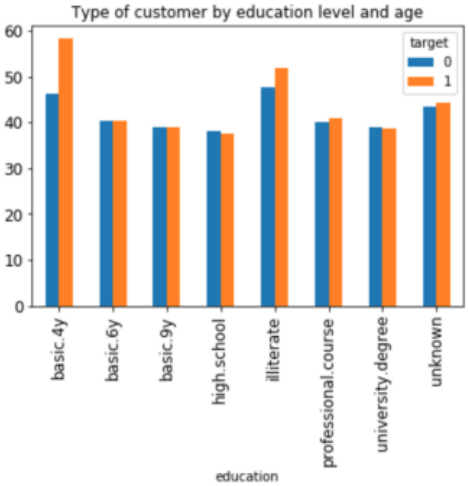
- 2 Customer repartition per marital situation and age analysis
-

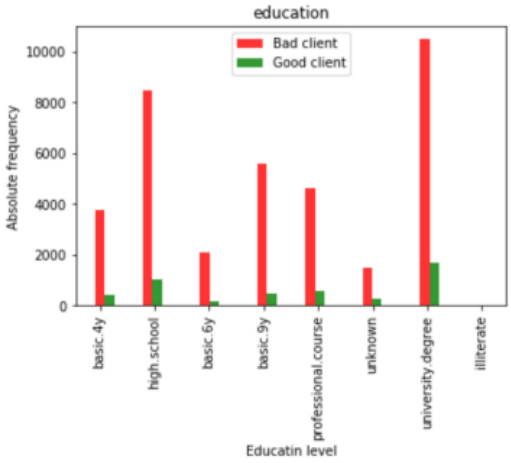




The married classe are the dominated classe contacted and subscribed to the term deposit



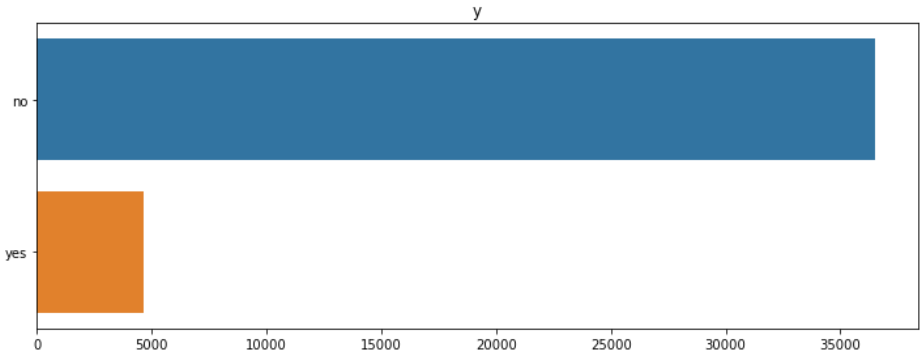


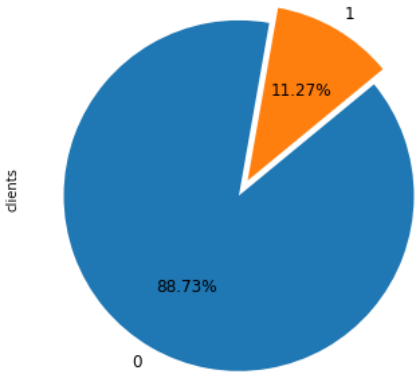


The more educated class are the dominated class in term of bank contact and also for deposit subscription.

- 3 Target repartition and data balencement

The dataset is imbalanced with the class no is higher than the class yes



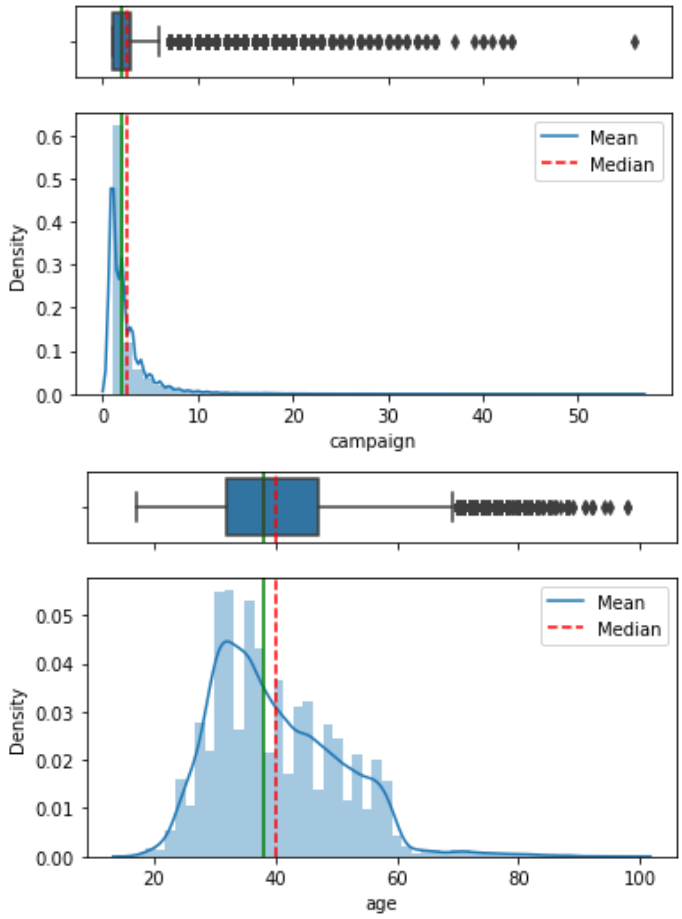


Step3 : Numerical Attributes Analysis

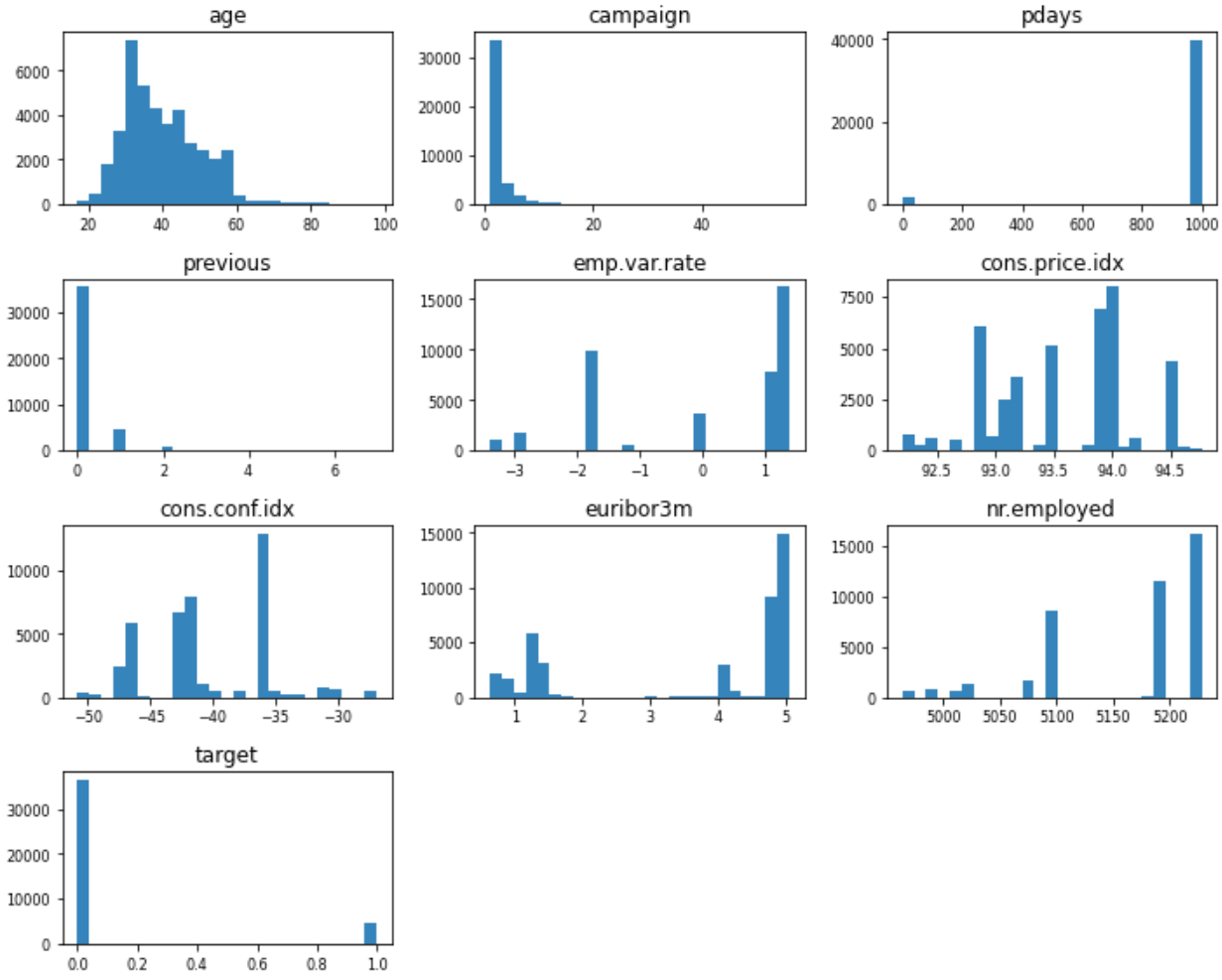
- Outlier Verification:

  From boxplot and histogram of numerical variables plotted using univariate analysis, we can see that it seems that our data have outlier on age and campain attributes.

- We can observe from boxplot that we have dispersion in our data, specially in the age and campaign attributes.I handle them using percentile method in the next week task but, actually I will not drop them.In fact we can se that they have a normal behavior as we work with a real world dataset.



- Distribution of numerical attributes, histogram plot:



-

Recommandation :

-
-