Week 10 deliverable

I. Team members details

Group name: ***Data science Geeks***

Names: Tasnime Hamdeni

Refka Mejri

Email : hamdeni.tasnime@gmail.com

refka.mejri@enit.utm.tn

Country: Tunisia

Tunisia

College: National Engineering School of Tunis

National Engineering School of Tunis

Specialization: Data science

II. Github repository link

https://github.com/tess92/Bank-Marketing-Campaign-

III. Problem description

We want to get some insight from the data of a bank called ABC that 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 or other Financial Institution).

Business need: Buying Product for customer.

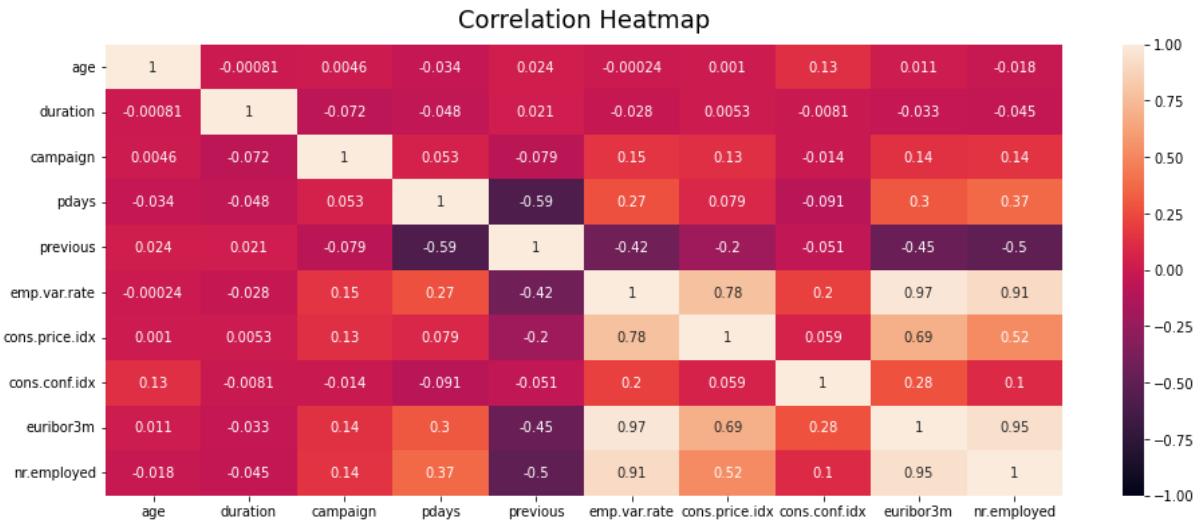
Method: using ML model to help companies shortlist customers whose chances of buying products is more so that their marketing channel can focus on them.

IV. EDA performed on the data

The exploratory data analysis performed on the bank marketing campaign client's data were very useful in terms of understanding their main characteristics. Visual method were performed to draw meaningful patterns and insights. We have also involved the preparation of the data by removing irregularities in the data.

- 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.

Features correlations



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

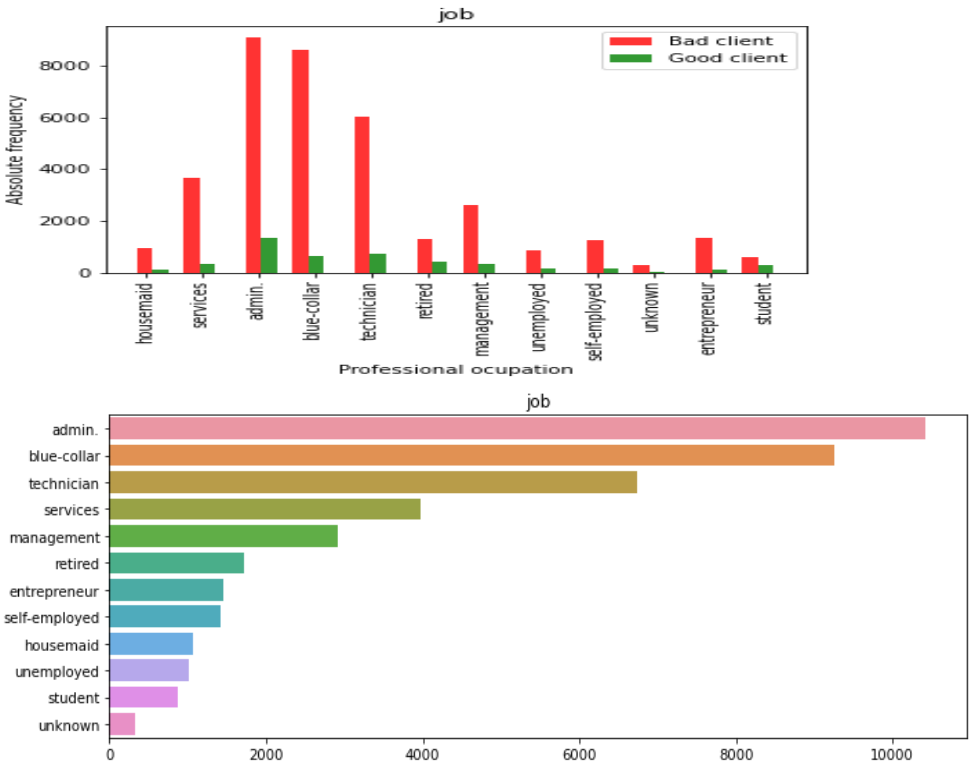
Categorical Attributes Analysis

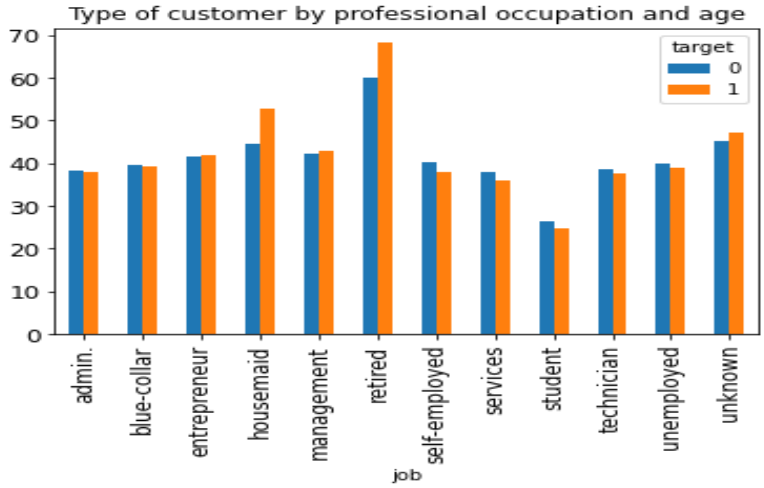
Categorical features are :

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



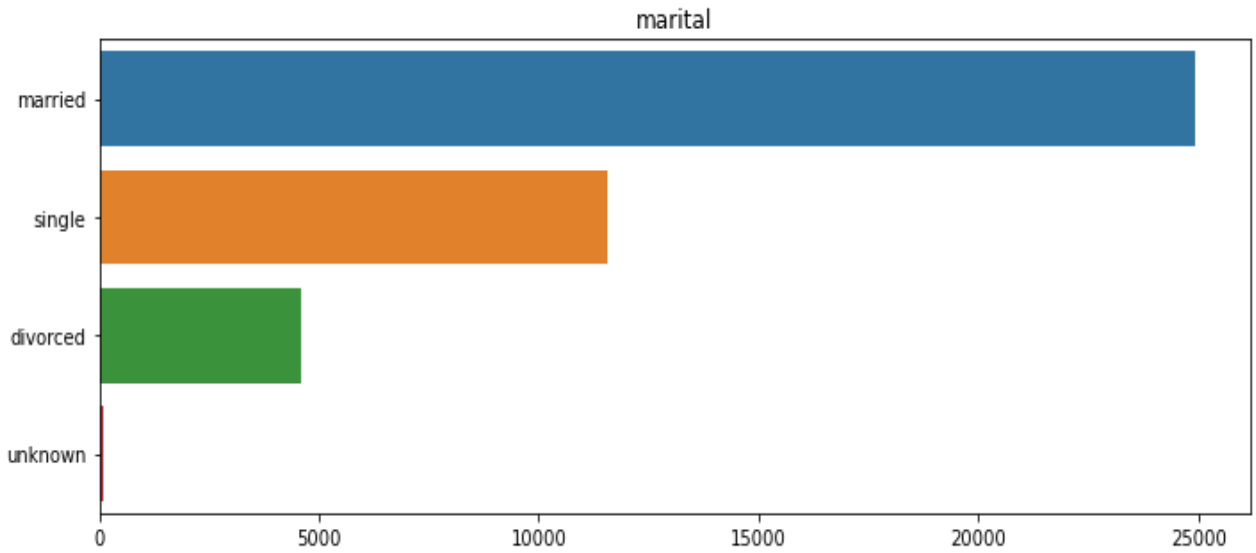


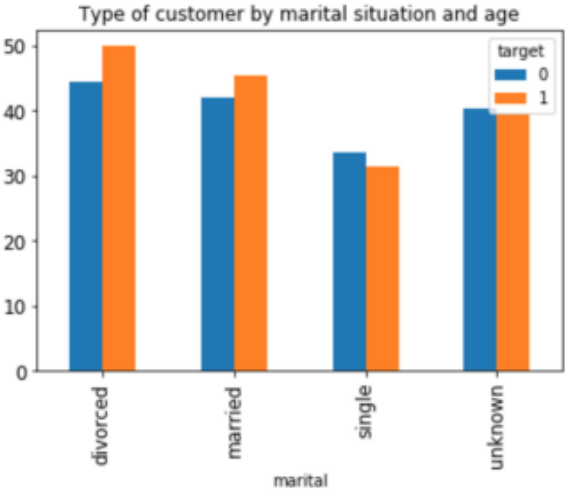
From the second graph, we can conclude that people with admin, blue_collar and technician jobs are more contacted by the bank.It can be explained by their social situation and the bank expectation.

From the first plot, we see student and retired are the more subscribed according to the total numbers of people by job category.

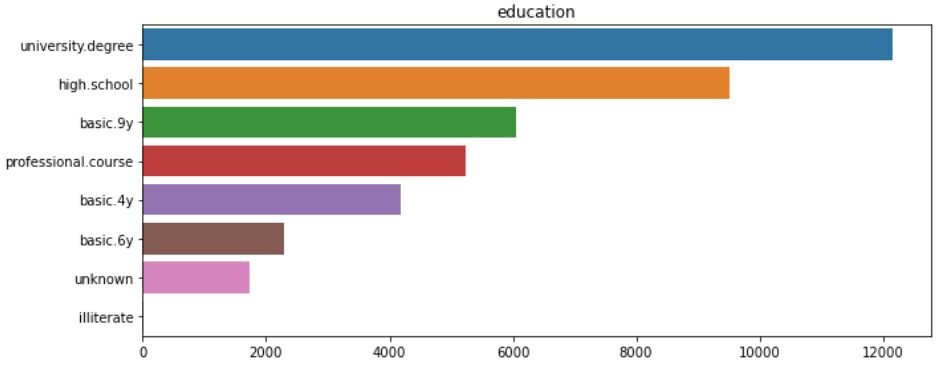
Below from the ttype of customer by job and age, we can observe that retired and housemaid are the oldest customers and the more classe that accepted to subscribe to the term deposit.

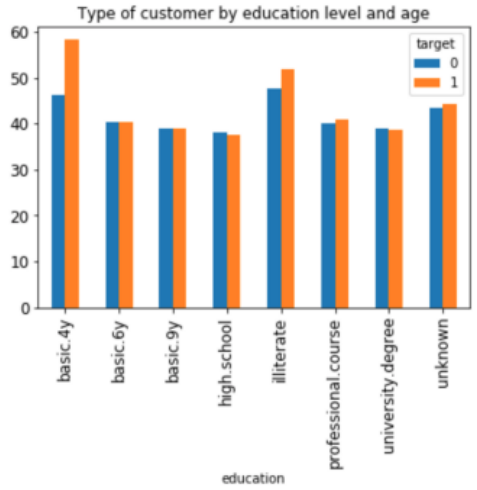
- 2 Customer repartition per marital situation and age analysis
-

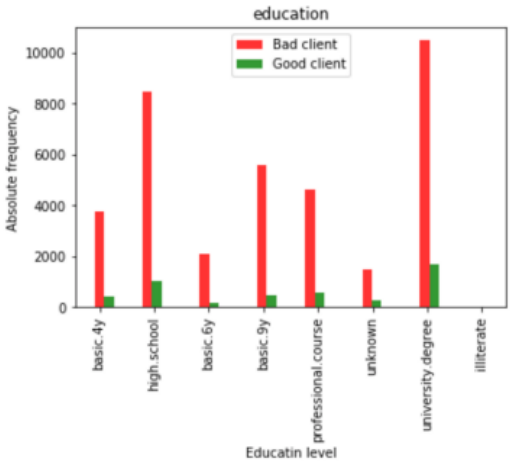




The married classe are the dominated classe contacted and subscribed to the term deposit



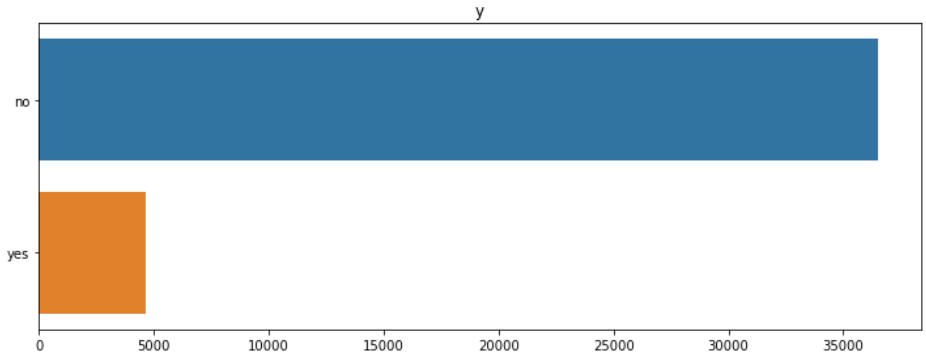


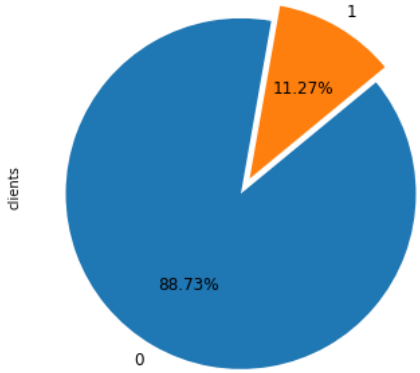


The more educated class are the dominated class in term of bank contact and also for deposit subscription.

- 3 Target repartition and data balencement

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

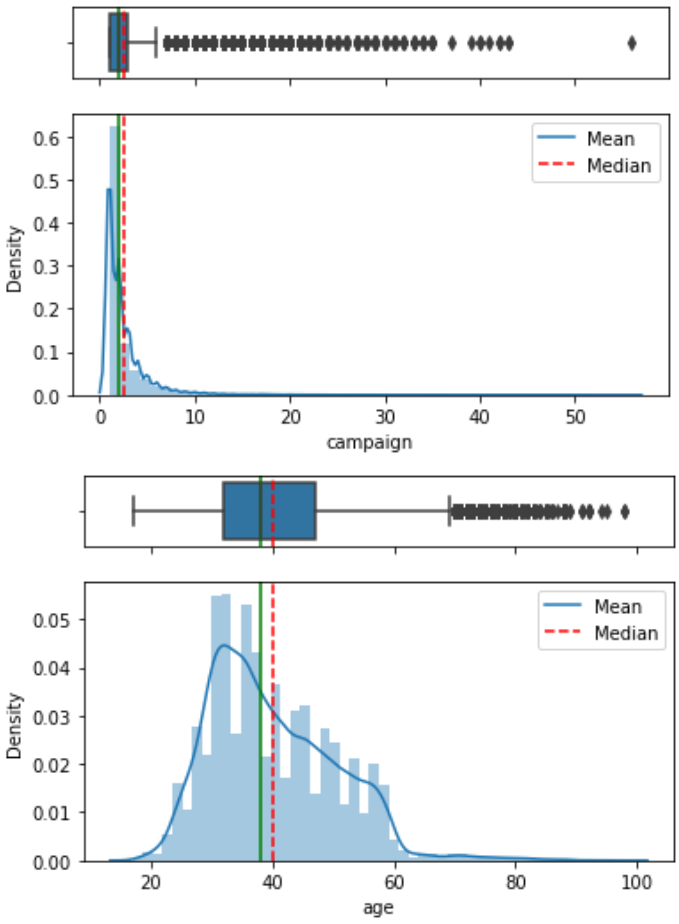




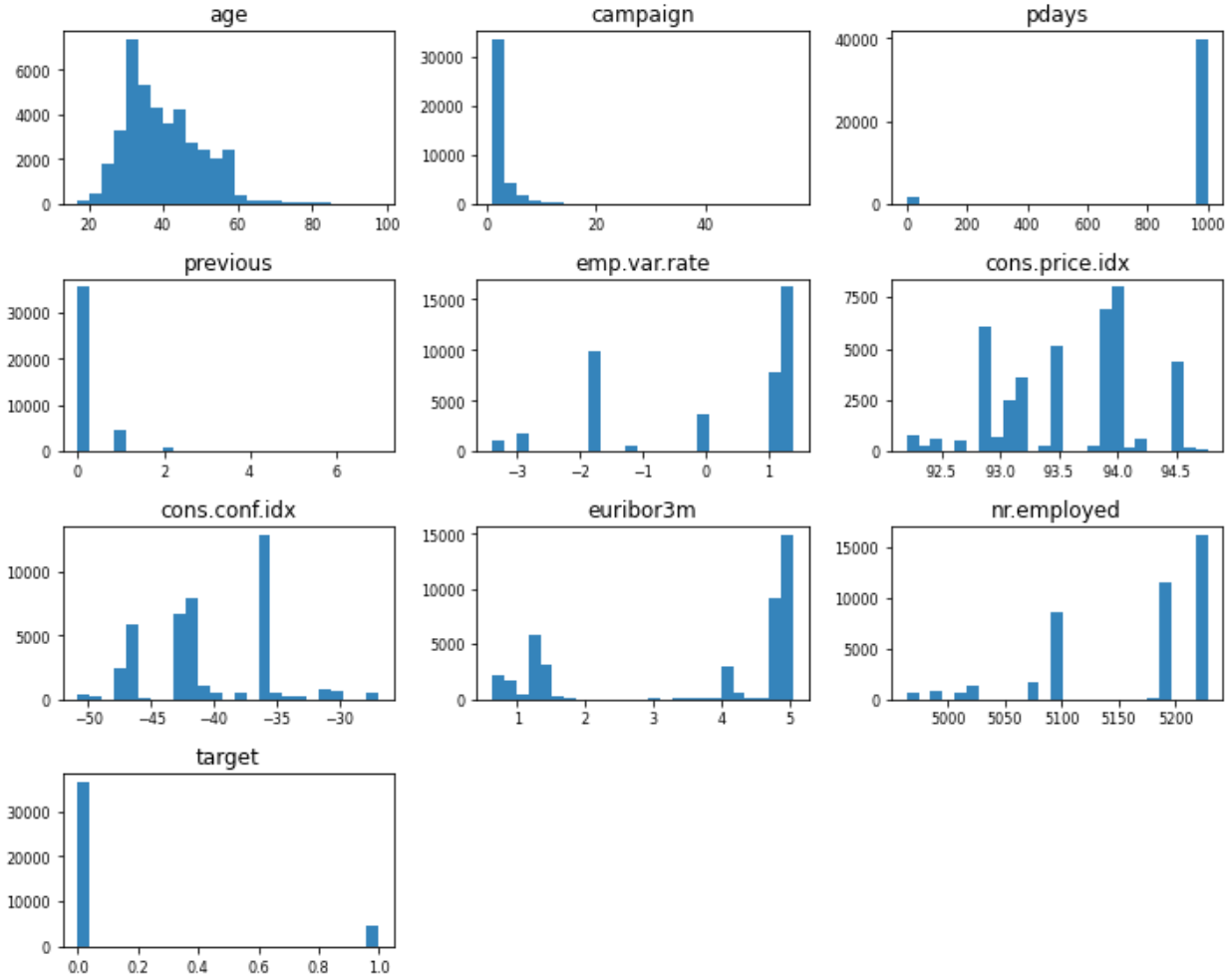
Numerical Attributes Analysis

- 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:



V. Final recommendations

We recommend the use of univariate, bivariate and multivariate analysis to evaluation the impact of each column and each risk factor on the output variable **y.**