Data Science Final Project :

Bank Marketing Campain prediction Using Machine Learning

Data Science Final Project :Week9

Plan:

- ***Project details***
- ***Project Process***
- ***Problem Description***
- ***Data Preprocessing(Cleaning and Transformation)***

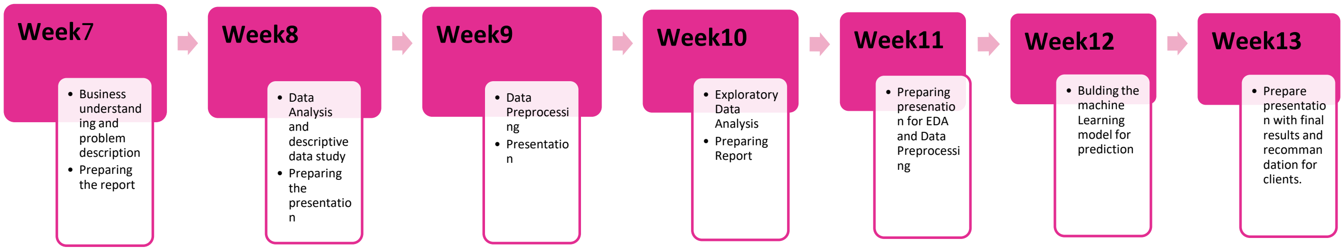
  1/ Duplication, transformation

  2/Missing Values and Outliers
- ***Data Storage location : Github link reposiroty***

Project Informations:

Team Members Informations	**Name of Group** : **Data Scientist Geeks**
	Member 1:***Refka Mejri*** - Tunisia/National Engineering School of Tunis
	Member 2 :***Tasnime Hamdeni*** - Tunisia/National Engineering School of Tunis

> ***Project Roadmap and Process:***
> *This process is prepared according to the needs of the company and the submission of each week.*



Week7
- Business understanding and problem description
- Preparing the report

Week8
- Data Analysis and descriptive data study
- Preparing the presentation

Week9
- Data Preprocessing
- Presentation

Week10
- Exploratory Data Analysis
- Preparing Report

Week11
- Preparing presenation for EDA and Data Preprocessing

Week12
- Bulding the machine Learning model for prediction

Week13
- Prepare presentation with final results and recommandation for clients.

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 Data Glacier Company to help them with a AI model for prediction.

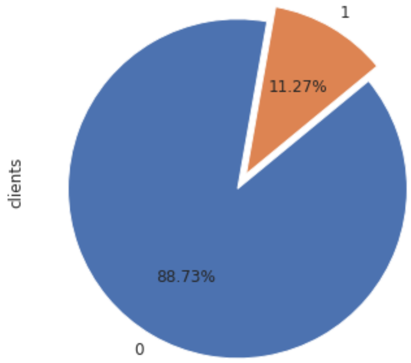
The Data Glacier company give us as a data scientist team this mission.
We will develop a Ml Model to shortlist customer whose chances of buying the product is more so that their marketing channel can focus only on those customers whose chances of buying the product is more.

Data Preprocessing :

- Duplication verification **:**
  `Using data_dup.shape in my code I conclude that we have 12 columns (12, 21)duplicated to remove.`

  #Drop duplication
  data = bank_additional_full.drop_duplicates() data.shape `(41176, 21)`

- Replacing the dots with underscores for better working with different variables:
  In this step, I choose to replace dots with underscores for better variables manipulations.
- Convert the target to binary (see code week9)
- Verify if the data is imbalenced or no for the target



The dataset is imbalenced with the class 0(no) more higher than the classe 1(yes)

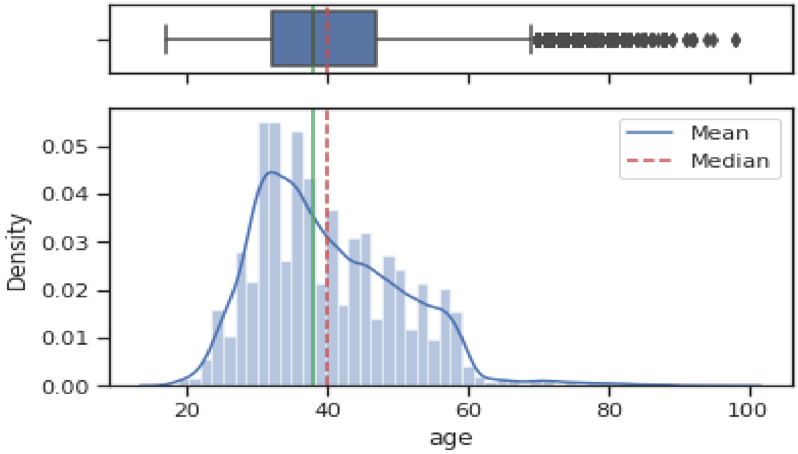
2/Missing Values and Outliers

- ### *Missing Values*

We dont have missing values in our datasets.
If we have missing values we can drop them or use inference methode to handel and deal with them.

- ### *Handling ouliers*

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.



After analyzing the real word data logicaly and verifying the different values,Histogram, mean of those attributes we assume and decide to treat them as normal variables.

In fact the graph of age and the plot as seen has a normal behavior, so we will not drop them.

Data storage location: https://github.com/RefkaaMejri/refka