- Team member details:

Name	Email	Country	College	Specialization
Jagriti Bablani	bablani.jagriti@gmail.com	India	Manipal University	Data Science

- Project Problem:

Bank wants to use ML model to shortlist customer whose chances of buying the product is more so that their marketing channel (tele marketing, SMS/email marketing etc) can focus only to those customers whose chances of buying the product is more.

- Data Understanding

Total number of observations	4521 rows
Total number of files	3
Total number of features	17
Base format of the file	csv
Size of the data	451KB

- What type of data you have got for analysis?

It's a classification problem. The data is related with direct marketing campaigns of a Portuguese banking institution. The marketing campaigns were based on phone calls. Often, more than one contact to the same client was required, in order to access if the product (bank term deposit) would be ('yes') or not ('no') subscribed.
The classification goal is to predict if the client will subscribe (yes/no) a term deposit (variable y).

- What are the problems in the data (number of NA values, outliers, skewed etc)

There are no NA values in the given data set. Some variable like Age, Duration and balance has outliers. Duration (last contact duration, in seconds (numeric)) and balance have outliers and are right skewed which needs to be dealt. And the data is imbalance which could play a major rule in predictions

- What approaches you are trying to apply on your data set to overcome problems like NA value, outlier etc and why?

Since there are no null values, no imputation is needed. Some of the variables have outliers which can be dealt by using transformation or scaling which depends on the variable and for imbalance under sampling or oversampling technique can be used.

GitHub link: https://github.com/jagriti02-2000