Team Details

Team Heatmap				
Name	**Email**	**Country**	**College/Company**	**Specialization**
Runa Veigas	runaveigas@gmail.com	India	Manipal Academy of Higher Education (MAHE)	Data Science
Odaliz Balcazar	obalcazarm@uni.pe	Peru	National University of Engineering	Data Science
Jonatan Vignatti	j.vignatti@gmail.com	Chile	Federica Santa Maria Technical University	Data Science

Problem Description

ABC Bank wants to sell its 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).

Data Understanding

The data is related with direct marketing campaigns (phone calls) of a Portuguese banking institution. The classification goal is to predict if the client will subscribe a term deposit (variable y).

Number of rows	4521
Number of Columns	17
Number of columns with missing values	None
Now of rows with missing values	None
Total number of integers	5
Total number of categories	10
Output	Yes: The client subscribed a term deposit. No: The client has not subscribed a term deposit.

As the data is about the history of details collected from a phone call made by the bank during their previous marketing campaigns. The response variable is categorical in nature with ‘Yes’ meaning the client has subscribed a term deposit and ‘No’ meaning the client has not subscribed to the term deposit. ‘Duration’ is an important column that is said to explain about the variability in the response with a certainty. There are a total of 17 features out of which 10 are categorical and rest are numerical.

Problems in the dataset are that there are outliers in age, duration have outliers. Duration is right skewed as most of the data is near to 0. Some of the features data types need to be changed to categorical as it is given as numerical data. Also the dataset is heavily imbalanced as there are more number of ‘no’ in the dataset then the number of ‘yes’.

There are outliers in the age and duration. Since duration can vary from customer to customer this outlier will not be altered while the age does not have any significant outliers. However the data needs to be scaled especially pdays (number of days that passed by after the client was last contacted from a previous campaign)