Group Name: Glacier Analysis Group

Team Members Details:

Name, Email, Country, College/Company, Specialization (Data Science, NLP, Data Analyst)

1. Tarusmita Boro, tarusmita.boro@gmail.com, United Kingdom, Data Science
2. Jonas Lütticken, jonaslutticken@gmail.com, Germany, Data Science
3. Noella Mutuku, noellamutuku@gmail.com, Kenya, Data Glacier, Data Science
4. Jeffrey Joseph, jeffreyjoseph21506@gmail.com, United Kingdom, 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 helps 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.

GitHub Repo Link:

https://github.com/TaruIndia/predict-termdepositsubscription

EDA performed on the data:

1.Bank marketing dataset consists of 45211 rows.

2.We can see that we have ‘unknown’ values in categorical columns namely ‘job’, ‘education’, ’contact’ and ’poutcome’.

3.Using the mode for the columns ‘job’, ‘education’ and ’contact’

4. For ’poutcome’ we are replacing the ‘unknown’ with ‘other’ as we could see that it won't affect the outcome. We can handle it by doing one hot encoding.

5.We can see that numerical columns have outliers (especially 'pdays', 'campaign' and 'previous' columns). Possibly there are incorrect values (noisy data), so we should look closer at the data and decide how we manage the noise.

Final Recommendation:

1. Customers with 'blue-collar' and 'services' jobs are less likely to subscribe for term deposit.

2. Married customers are less likely to subscribe for term deposit.

3. Customers with 'cellular' type of contact are less likely to subscribe for term deposit.

4. People who subscribed for term deposit tend to have greater balance and age values.

5. People who subscribed for term deposit tend to have fewer number of contacts during this campaign.