Group Name: FIC

Group Member:

Name: Felipe Crespo

Email: [fcrespo12@gmail.com](mailto:fcrespo12@gmail.com)

Country: Argentina

College: UES21

Specialization: Data Science

Problem Description

XYZ insurance company is planning to create a model that can predict the likelihood of a customer making a claim based on their previous interactions with the insurance company or other insurance providers. The dataset provided contains information on customer demographics, contact details, and previous insurance history.

However, there are instances where certain features have missing values represented by the string "unknown". These instances will be treated as null values. The features with missing values include job (0.8%), marital status (0.2%), education (4.2%), default (20.9%), housing (2.4%), and loan (2.4%).

The target variable, whether a customer makes a claim or not, is imbalanced with 89% of customers not making a claim and only 11% making a claim.

Additionally, there are a lot of outliers in the features, which will need to be addressed during modeling. Finally, the "duration" feature is left-skewed and requires careful consideration as the target variable (whether a claim is made or not) may be known after the call has ended, making it less useful.

Github Repo link

https://github.com/felincre/Bank-\_Marketing\_Campaign