Team Member Details:

Name	Email	Country	College/Company	Specialization
Jane Condon	jane.condon@stonybrook.edu	United States	Stony Brook University	Data Science
Jeffery Su	Jefferysu556@gmail.com	Canada	University of Toronto	Data Science
Seyed Mojtaba Hejazi	Mojtaba.hejazi991368@gmail.com	Italy		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).

Data Understanding

The dataset contains 17 columns and 45,211 rows (in the full dataset). There is a binary categorical target variable “y,” which indicates whether a client has subscribed for a term deposit. There are 16 input variables, 7 of which are numerical and 9 of which are categorical. There are no missing values present within the dataset.

Problems With the Data

Many of the numerical variables in the dataset are skewed. The variables ‘balance’, ‘duration’, ‘campaign’, and ‘pdays’ are strongly skewed to the right. The ‘previous’ variable may also be skewed to the right. The ‘age’ variable appears to be slightly skewed as well. There are also outliers present within many of the numerical variables (‘age’, 'balance', 'duration', 'campaign', 'pdays', and 'previous').

Approaches to Overcoming Problems With the Data

There are no missing/NA values in the dataset, so there is no need to deal with missing data. To fix the skewness, we are trying a variety of techniques, such as log transformation, square root transformation, box-cox transformation. We can fix the outliers by removing them or using mean/median imputation. However, most of the transformations that we used to fix the skewness (mentioned above) also deemphasize outliers, so we may not need to remove them or impute them with the median/mean.

Github Repository: https://github.com/janecondon/Data-Glacier-Internship-Group-Project.git