

Group Project Third Deliverable

Team Details

- Group Name: Solo Banking
- Name: Ian Paulino Velez
- Email: ianpvelez@gmail.com
- Country: United States
- College: Fordham University
- Specialization: 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 to help them understand whether a particular customer will buy their product or not (based on customer's past interaction with bank or other Financial Institution).

Data Cleansing and Transformation

Three approaches were explored in the python notebook in relation to outliers and missing values:

1. A cap was placed on outliers, eliminating extreme values but retaining the original distribution and missing values on categorical variables 'job' and 'education' were imputed using the mode and high correlation between variable values (e.g., most managers have tertiary education, so instances with missing values in the 'education' feature and 'manager' in the 'job' feature had their missing values imputed as 'tertiary'). The two features with a large number of missing values, 'poutcome' and 'contact', were dropped.
2. A cap was placed on outliers, eliminating extreme values but retaining the original distribution. Instances with missing values in 'education' and 'job' were dropped. This amounted to a total of 2,018 fewer instances. The two features with a large number of missing values, 'poutcome' and 'contact', were dropped.
3. A transformation on all variables into their WoE bin was implemented, dealing with both outliers and missing values.

Github Repo link:

<https://github.com/ianpv04/Data-Glacier-Project.git>