DELIVERABLE WEEK - 9

Group Name : Walther Rathenau Team
Specialization : Data Science

Team members:

1. Name : Mohammed Maqsood
Email : mohammedmaqsood48@gmail.com
Country : Germany
College : Otto von Guericke University
Specialization : Chemical and Energy Engineering

2. Name : Davangam Sreedhar Saikiran
Email : saikiran.davangams@gmail.com
Country : Germany
College : Otto von Guericke University
Specialization : Chemical and Energy Engineering

Problem description

ABC Bank wants to sell it's 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).

Here we are using different approaches to clean and transform the data in order to solve the above-mentioned problem.

Data Understanding

The objective of data understanding was to understand attributes of the data by identifying key characteristics in the data such as data volume, total number of variables in the data, understanding problems with the data such as missing values, duplicates outliers, skewness etc.

Data Cleaning and Transformation

The dataset was checked for missing values, duplicates, outliers, skewness

No missing values or duplicates were found. Summary statistics such as the mean, standard deviation, distribution, and kurtosis, skewness were checked.

Github repo link:

https://github.com/Maqsood8/Group-Project-Bank-Marketing.git

Results and approaches

- EDA was performed for continuous and categorical variables

- Suggested methods for dealing with outliers

    - Method 1: Flooring and capping using Interquantile Range (IQR)
      Outliers are removed by dropping any values that are below 25th percentile and above 75th percentile.

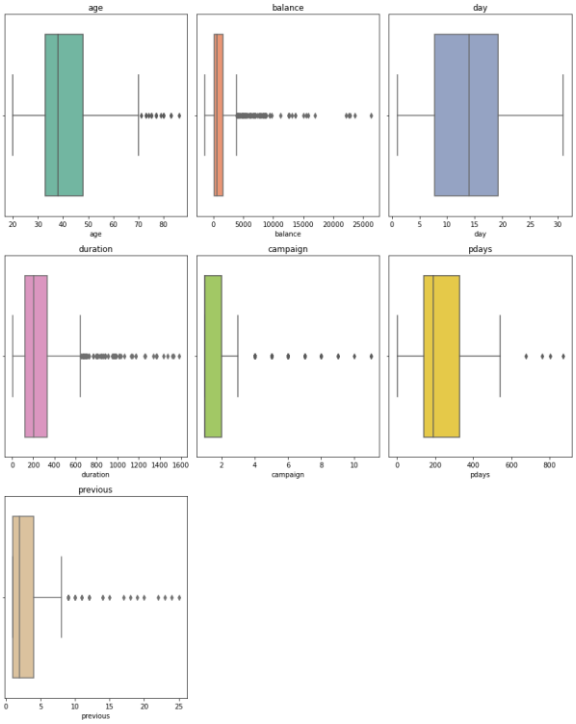

Figure 1 - Before Transformation

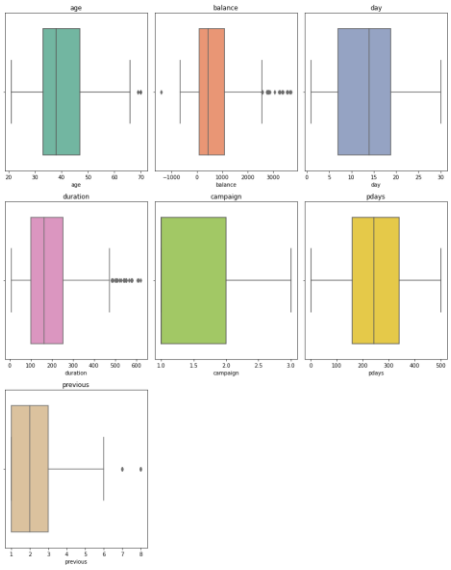

Figure 2 - After Transformation

After transformation of the data using the interquantile range, there is less outliers.

- Method 2: Binning of outliers

  Continuous variables were binned so that outlier values are converted to count values.

- We decided to keep outliers. A different method of dealing with outliers without removing them is to use the Minkowski error which reduces the impact that outliers will have on the model.

- Weight of Evidence (WOE) and Information Value (IV)

  We identified variables that had significant predictive power for modeling using WOE and IV.