Deliverable Week 9

Group Name: Destined Data Team
Specialization: Data Science

Team members:

1. **Name:** Praneetha Rajupalepu
Email: praneetha.724@gmail.com
Country: Canada
Company: Modest Tree
Specialization: Data Science

2. **Name:** Selaelo Kholofelo Ramokgopa
Email: sly.kholo@gmail.com
Country: South Africa
College: University of Johannesburg
Specialization: Chemical Engineering

3. **Name:** Surya Chandra
Email: ksuryachandra619@gmail.com
Country: Germany
College: Otto von Guericke University
Specialization: Electrical Engineering and Information Technology

4. Name: Cagla Yucel
Email: caglasipahi@gmail.com
Country: Switzerland
College: Georgia Institute of Technology
Specialization: Analytics

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 will help them in understanding whether a particular costumer will buy their product or not (based on customer's past interaction with bank and other Financial Institutions).

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/PraneethaRajupalepu/Bank-DataScience-Project

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.

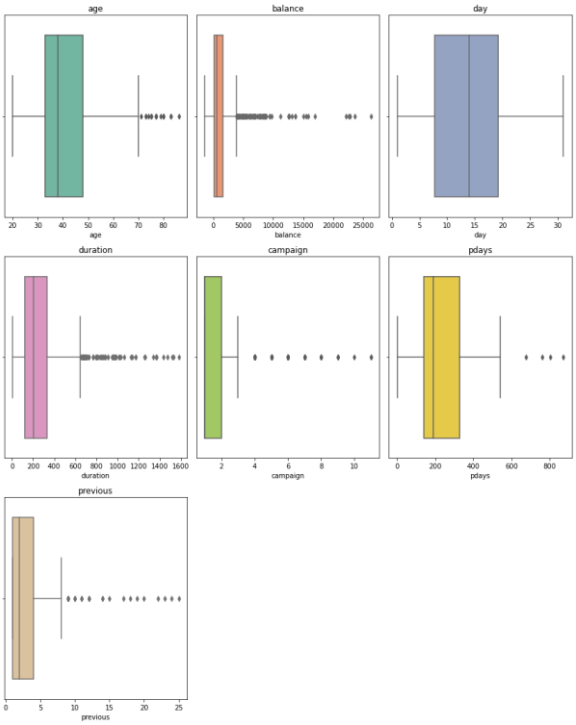

Figure 1 - Before Transformation

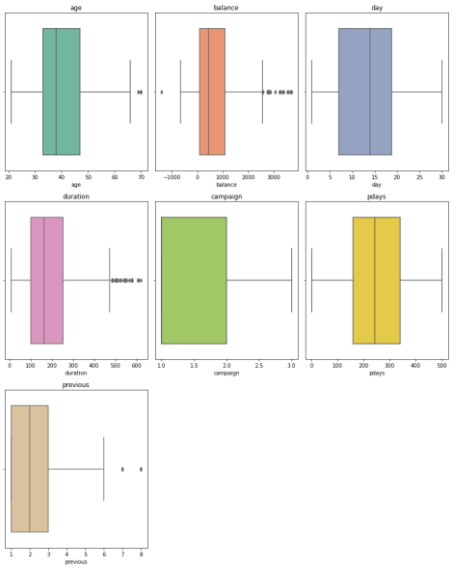

Figure 2 - After Transformation

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.