

Exploratory Data Analysis Report

EDA Report: Bank Marketing Campaign

1. Introduction and Problem Description

Group Name: Data Department_1

Member Details:

- Name: Minseok Kim
- Email: mxk230041@utdallas.edu
- Country: United States
- College/Company: The University of Texas at Dallas
- Specialization: Data Science

Problem Description: The objective of this project is to analyze a Portuguese bank's marketing campaign data to predict the likelihood of customers subscribing to a term deposit. This involves identifying and treating outliers in the dataset to ensure the accuracy and reliability of our predictive models, thus enabling efficient allocation of marketing resources.

2. Dataset Introduction

Data Source: The dataset is derived from a Portuguese bank's marketing campaigns, focusing on direct phone calls to offer term deposit subscriptions to potential customers.

Dataset Characteristics: Includes variables such as age, job, marital status, education, default,

Exploratory Data Analysis Report

balance, housing, loan, contact, day, month, duration, campaign, pdays, previous, and poutcome. The dataset consists of several thousands of records, each record representing a customer's details and the outcome of the marketing campaign.

3. Data Cleaning and Preprocessing

Missing Value Treatment: Initial analysis revealed minimal missing values, which were treated by imputation or removal, depending on their impact on the dataset.

Outlier Detection and Treatment:

- Z-score Method: Applied to columns with a normal distribution, identifying outliers as data points lying beyond 3 standard deviations from the mean.

- IQR Method: Calculated the Interquartile Range for each column, considering data points outside $1.5 * \text{IQR}$ from the Q1 and Q3 as outliers. This method is particularly effective for datasets with skewed distributions.

4. Exploratory Data Analysis (EDA)

Descriptive Statistics: Provides a summary of central tendencies, dispersion, and shape of the dataset's distribution.

Visualization: Histograms and box plots for individual variables to understand distributions. Scatter plots and correlation matrices to explore relationships between variables.

Analysis of Variable Relationships: Detailed investigation into how different customer attributes relate to the campaign's outcome, identifying key factors that influence a customer's decision to

subscribe to a term deposit.

5. Final Recommendations

Analysis Summary: The EDA reveals that outlier treatment significantly improves the dataset's quality, enhancing model accuracy.

Recommendations: Based on our analysis, targeted marketing strategies should focus on customers within specific age groups and job categories showing higher subscription rates. Additionally, campaign efforts should be optimized based on the insights from the duration and frequency of contacts.

6. Conclusion and Future Research Directions

Conclusion: Outlier detection and treatment using Z-score and IQR methods have proven to be crucial steps in preprocessing for predictive modeling in bank marketing campaigns.

Future Research Directions: Future studies could explore advanced outlier detection techniques and their impact on different types of predictive models. Additionally, incorporating external factors such as economic indicators could further refine customer segmentation and targeting strategies.

7. References

GitHub Repository Link: <https://github.com/NOVA-code/Week10.git>

analysis results for this project.