Bank Marketing Dataset Project Report

Objective

To explore and analyze a dataset from a bank marketing campaign, where the goal is to predict whether a client will subscribe to a term deposit (column `'y'`).

Data Overview

The dataset consists of 21 columns with a mix of categorical and numerical features.
Features include client demographics (e.g., age, job, marital status), contact information (e.g., contact type, month), and economic indicators (e.g., euribor3m, emp.var.rate).

Missing Values and Data Types

- **No missing values** were found in the dataset.
- Data types:
    - **Numerical features**: int64, float64
    - **Categorical features**: object (e.g., job, marital, education)

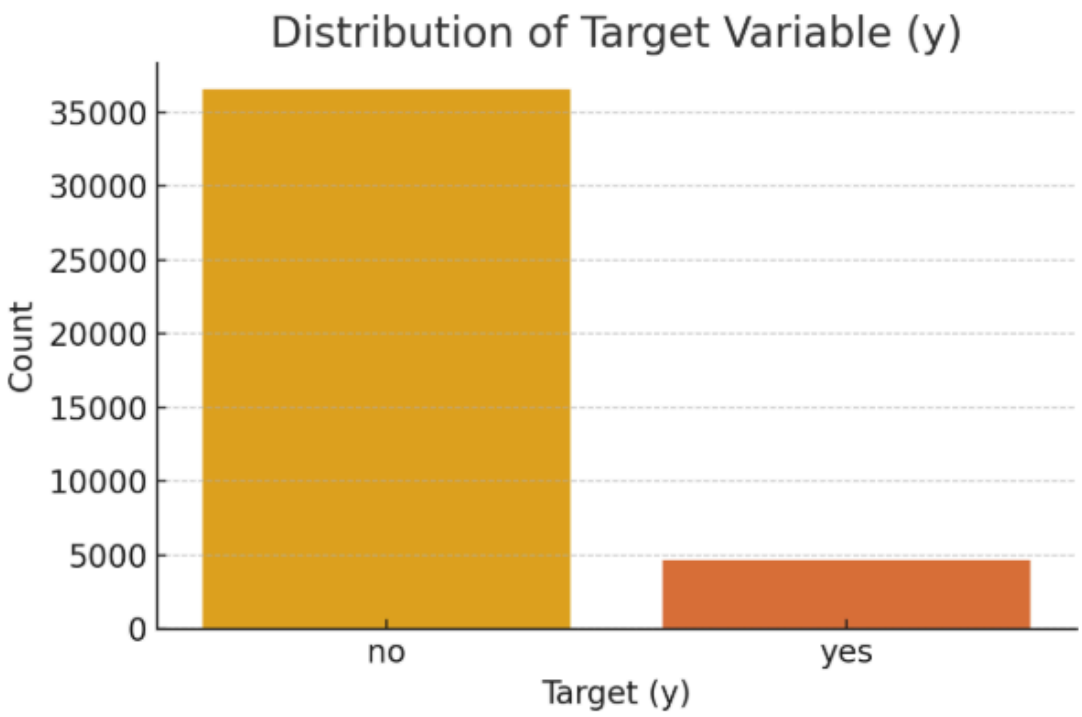
Summary Statistics

Descriptive statistics were generated using the `.describe()` method:

- Provided mean, standard deviation, min, max, and quartile values for numeric columns.
- Helped identify the distribution and scale of each feature.

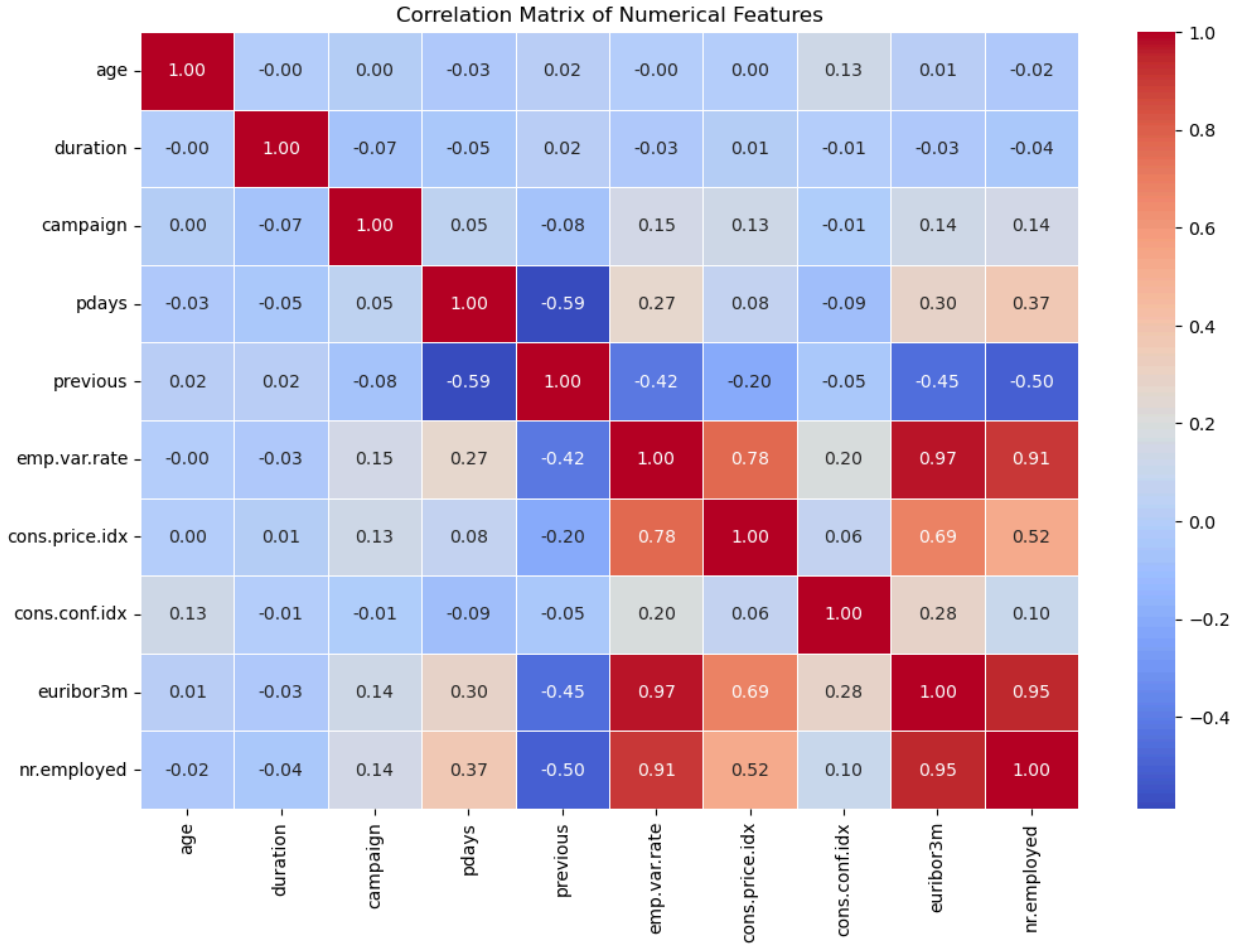
Target Variable Distribution

- The target variable `'y'` (yes/no) indicates whether a client subscribed to a term deposit.
- Distribution is **imbalanced**, with the majority class being `'no'`.
- This imbalance is important for any future classification models.



Correlation Matrix

- A heatmap was used to visualize correlations between numerical features.
- Strong correlations were observed between:
    - `euribor3m`, `emp.var.rate`, and `nr.employed` (strong economic indicators)
- These insights are valuable for feature selection and understanding market behavior.



Insights and Conclusion

- **Economic factors** have a strong influence on the success of marketing efforts.
- **Customer demographics and contact behavior** can be used for better targeting.
- The **class imbalance** in the target variable requires attention if moving to predictive modeling.
- This **Exploratory Data Analysis (EDA)** provides a strong foundation for:
    - Customer segmentation
    - Targeted marketing strategies
    - Predictive analytics and machine learning