

4. Analysis

4.1 Univariate Analysis

A univariate analysis was conducted on key features, especially the target variable (churn)

Through this analysis, we uncovered crucial insights about our data, including its distribution and class imbalance.

Appropriate visualization and interpretation were also provided.

4.2 Bivariate Analysis

Through bivariate analysis, we compared relationships between two variables/features contained in the data.

The main focus was on how each independent feature correlated with our target variable.

This process enabled us to uncover crucial patterns in our data.

4.3 Multivariate Analysis

We also investigated the relationship between more than two features and their contribution to our project objective.

This enabled a deeper understanding and the uncovering of key patterns.

Heatmap visualization best depicted how each feature correlated with our target variable.

4.4 Feature Engineering

To enhance model performance and interpretability, we applied targeted feature engineering techniques to the SyriaTel customer churn dataset. These transformations aimed to capture customer behavior more effectively and prepare the data for classification modeling.

Feature Creation

We derived a new variable to represent customer usage patterns better
Total Charges: Aggregated total day charge, total night charge, and total intl charge into a single total charges feature to reflect overall billing impact
This enabled us to investigate amounts or charges lost during churn

Feature Encoding

To prepare categorical variables for modeling, we applied feature encoding techniques that convert non-numeric data into numerical formats. To achieve this objective, we used Standard Scaler.

This ensures compatibility with machine learning algorithms and preserves the informational value of each category.

4.5 MODELING

Model Selection: Since we were dealing with a Classification problem, the machine learning models considered included :

- a) Logistic Regression (Baseline, binary classification)
- b) Decision Tree Classifier
- c) Random Forest

Class imbalance: The SMOTE technique was used to handle class imbalance

Model Evaluation: Evaluation metrics used in our modeling included,

- a) Accuracy Score
- b) Confusion Matrix
- c) Precision, Recall, F1 score