CHURN ANALYSIS REPORT

Dataset Overview

- The dataset contains **demographic, service, and account information** of Telco customers.
- Key features: gender, tenure, InternetService, Contract, MonthlyCharges, Churn, etc.
- Target variable: **Churn** (Yes/No)

Key Analytical Steps

1. **Data Cleaning & Preparation**
    - Identified and converted object-type numeric columns.
    - Checked and handled missing or inconsistent values (e.g., TotalCharges).
    - Removed duplicates and ensured data consistency.
2. **Univariate Analysis**
    - Distribution of categorical variables (e.g., gender, contract type).
    - Histogram plots and count plots for customer characteristics.
3. **Bivariate Analysis**
    - Correlation heatmap for numerical variables.
    - Relationship between Churn and categorical features like Contract, PaymentMethod, InternetService.
4. **Churn Drivers Identified**

- Customers with **month-to-month contracts**, **electronic check payments**, and **fiber optic internet** had significantly higher churn.
- Longer tenure and higher customer satisfaction were associated with **lower churn**.

5. **Visualizations Used**
   - Count plots, KDE plots, histograms, correlation heatmap, and box plots.
   - Applied color-coded EDA to clearly distinguish churn vs non-churn customers.

Key Insights

- **Contract Type** is one of the strongest predictors of churn; long-term contracts correlate with higher retention.
- **Monthly Charges** and **Internet Service Type** impact churn probability.
- Customer engagement drops significantly for users with low tenure and high monthly bills.

Data Preprocessing

- Loaded cleaned dataset from the EDA phase.
- Removed unnecessary columns and handled class imbalance using **SMOTEENN** (a hybrid of oversampling and undersampling).
- Split the dataset into **training and testing sets** using train_test_split.

Modeling Approach

1. **Model Used:**
   - **Decision Tree Classifier**

- Chosen for its interpretability and ability to handle categorical variables without one-hot encoding.

2. **Evaluation Metrics:**
    - **Recall Score**: Prioritized to capture true churn cases.
    - **Confusion Matrix**: Assessed classification balance.
    - **Classification Report**: Precision, recall, F1-score for both classes.
3. **Results:**
    - The model effectively identified churned customers while maintaining acceptable accuracy for non-churn cases.
    - Applied **feature importance** to understand key drivers in model decision-making.