**Telco Customer Churn Analysis**

In this project, I conducted an in-depth analysis of customer behavior using a telecom dataset to understand the patterns that lead to customer churn. I began with exploratory data analysis (EDA) to study customer demographics, service usage patterns, and payment preferences. I used various visualization techniques like histograms, boxplots, countplots, and correlation heatmaps to uncover trends and relationships in the data. Key features such as tenure, MonthlyCharges, and Contract stood out as strong indicators of churn. I cleaned the data by handling missing values in the TotalCharges column and applied label encoding to convert categorical variables into numeric form. To address the class imbalance in churn labels, I used the SMOTE technique to oversample the minority class and ensure balanced model training. After preparing the data, I built and evaluated multiple machine learning models including Decision Tree, Random Forest, and XGBoost, and assessed their performance using accuracy, precision, recall, and confusion matrix. The goal was to identify which customers are most likely to churn, so proactive retention strategies could be designed.

**Key Insights:**

* **Churn rate is around 26%**, indicating a class imbalance—most customers stay, but a significant portion leaves.
* **Customers with low tenure (0–12 months)** are far more likely to churn, showing that early engagement is critical.
* **Month-to-month contracts** have the highest churn rate, while long-term contracts reduce the risk significantly.
* **High MonthlyCharges** correlate with increased churn, possibly due to customers paying for services they don’t use.
* **Electronic check payment method** users have the highest churn rate, suggesting these customers are less committed.
* **Senior citizens** show a slightly higher tendency to churn, indicating a need for more accessible services and support.
* **Key predictive features** include Contract, tenure, MonthlyCharges, InternetService, and PaymentMethod.
* **SMOTE helped balance the dataset**, which improved model fairness and the ability to detect churners more accurately.
* **XGBoost or Random Forest** models performed best, offering reliable churn prediction based on customer behavior.