Report: Customer Churn Prediction

1. Import Libraries

Import required libraries like Pandas, NumPy, Seaborn, Matplotlib, and Scikit-learn.

2. Load the Dataset

Use `pd.read_csv()` to load the telecom churn dataset (e.g., `customer_churn_large.csv`).

3. Explore the Dataset

- View shape, columns, data types (`df.info()`)
- Check for null values and duplicates
- Use `df.describe()` for summary statistics

4. Data Cleaning

- Handle missing or invalid values in `TotalCharges`
- Convert data types if needed
- Drop or impute missing rows

5. Exploratory Data Analysis (EDA)

- Visualize churn counts using countplots
- Analyze relationships using histograms, boxplots, etc.
- Understand correlation between features and churn

6. Feature Encoding

- Convert categorical columns to numerical using `pd.get_dummies()` or `LabelEncoder`

7. Split Data

- Split the data into features (X) and target (y)
- Use `train_test_split()` to divide into training and test sets

8. Train Model

- Use a classification algorithm like `RandomForestClassifier`
- Fit the model on training data

9. Evaluate Model

- Predict on test data
- Use metrics like Accuracy, Confusion Matrix, Precision, Recall, and F1-Score

10. Export the Model

- Save the trained model using `joblib.dump(model, 'churn_model.pkl')`

11. Optional Enhancements

- Deploy the model using Flask or Streamlit
- Create dashboards using Tableau or Power BI