📓 Full Final Report: Telecom Customer Churn Analysis

🧠 Project Title: Telecom Customer Churn Prediction

🔍 Overview

Customer churn prediction involves identifying customers likely to cancel a service based on usage patterns. This project aims to help telecom companies proactively address churn by understanding key drivers and implementing targeted interventions. Reducing churn is essential as acquiring new customers is significantly more expensive than retaining existing ones.

❗ Why is it Important?

Customer churn affects profitability across industries. Each departing customer means lost revenue and additional costs for acquisition. Being able to predict when customers are likely to leave and proactively engaging them can result in substantial savings and improved customer satisfaction.

📂 Dataset Summary

- **Source**: Telco Customer Churn Dataset (Kaggle)
- **Total Customers**: 7043
- **Churn Rate**: ~27% (~1869 customers)
- **Monthly Revenue**: $16,056,169
- **Estimated Revenue Lost to Churn**: ~$2,862,927 (18%)

🧪 Steps Involved in the Project

1. **Data Analysis (EDA)**
2. **Data Preprocessing**
3. **Feature Engineering**
4. **Feature Selection** (SelectKBest)
5. **Model Building** (ML Algorithms)
6. **Hyperparameter Tuning** (RandomSearchCV)
7. **Model Deployment** (Pickle Serialization)

📊 Exploratory Data Analysis (EDA)

- Churn rate distribution and comparison between churned vs non-churned customers.
- Revenue estimation based on monthly charges and churn rate.
- Univariate plots: histograms, boxplots, Q-Q plots.
- Bivariate analysis with pairplots and segment-based visualizations (e.g., SeniorCitizen, Contract type).
- Correlation matrix and feature importance for numeric predictors.

⚙️ Preprocessing & Feature Engineering

- Converted `TotalCharges` to numeric and handled missing values.
- Categorical vs Numerical feature separation.
- Potential next steps include:
  - One-Hot Encoding of categorical features.
  - Train-Test Split.
  - Feature scaling (e.g., MinMaxScaler, StandardScaler).
  - Model Evaluation with classification metrics (accuracy, precision, recall, AUC).
  - Model comparison: Logistic Regression, Random Forest, XGBoost, etc.

🔎 Insights

- **High churn rate** among customers with **Month-to-Month contracts**.
- Customers with **short tenure** and **high monthly charges** are more likely to churn.
- **Senior citizens** have a slightly higher churn rate.
- **Electronic payment methods** correlate with increased churn.
- Loss in revenue is significant and justifies urgent intervention.

☑ Recommendations

- Offer **incentives for long-term contracts**.
- Provide **personalized offers** to short-tenure customers to improve loyalty.
- **Improve service quality and customer support**, especially for older and high-paying clients.
- Implement **early warning systems** using churn probabilities for proactive retention campaigns.

🧰 Tools Used

- Python
- Pandas
- Matplotlib & Seaborn
- Scikit-learn
- SciPy
- Jupyter Notebook