Action Plan for the Customer Churn Prediction Project

Project Goal Definition

Build a model to accurately predict churn and surface top drivers for actionable retention strategies.

Data Acquisition

Use the Telco Customer Churn dataset (e.g., Kaggle). Fields typically include customerID, gender, SeniorCitizen, Partner, Dependents, tenure, PhoneService, InternetService, Contract, PaperlessBilling, PaymentMethod, MonthlyCharges, TotalCharges, and Churn.

Environment Setup

Python: Pandas, NumPy, Matplotlib, Seaborn, scikit■learn; optional imbalanced■learn for SMOTE.

EDA & Hypothesis Testing

Profile churn by contract type, payment method, tenure, and charges. Use chi■square for categorical associations and t■tests/Mann–Whitney for numerical differences between churned vs. non■churned groups.

Feature Engineering & Preprocessing

One■hot encode categoricals, scale numeric features as needed; handle class imbalance via class weights/SMOTE; guard against leakage; create features like tenure buckets, charge■to■tenure ratios.

Model Selection, Training & Evaluation

Train baseline Logistic Regression; compare with Decision Tree, Random Forest, Gradient Boosting, and SVM. Use stratified split and cross■validation. Evaluate with Precision, Recall, F1, ROC■AUC, and Confusion Matrix; optimize thresholds for business goals.

Optimization & Explainability (Optional)

Tune hyperparameters (Grid/Random Search). Use feature importance, permutation importance, and SHAP for explainability.

Deployment (Optional)

Create a Streamlit/Flask app: inputs → preprocessing → model → churn probability + top drivers; include retention playbook suggestions.

Documentation & Timeline

Deliver notebooks, cleaned data, final report, and (optional) web app.
Week 1: Setup & data understanding
Week 2: EDA & hypotheses
Week 3: Modeling & tuning
Week 4: Evaluation & reporting