Customer Churn Prediction & Segmentation Report

June 2025

1. Objective

Build a data-driven solution to identify at-risk customers and group them into actionable segments for targeted retention campaigns.

2. Data & ETL

- Source: Telco Customer Churn dataset (7,043 records)
- CSV → SQLite ingestion, cleaned invalid TotalCharges rows, trimmed whitespace
- Derived features: avg_monthly_charge, tenure_bucket

3. Feature Engineering & Pipeline

- Numeric: tenure, MonthlyCharges, TotalCharges, avg_monthly_charge
  (strip → coercion → median impute → standard scale)
- Categorical: one-hot encode 10 service/billing fields
- Encapsulated in sklearn Pipeline, serialized via joblib

4. Churn Model

- Algorithm: LogisticRegression (class_weight='balanced')
- Validation: 80/20 stratified split
- Performance: ROC-AUC 0.82, PR-AUC 0.45 (no-skill 0.27)
- Top Drivers: month-to-month contract, fiber optic, electronic check, short tenure

5. Customer Segmentation

- Method: K-Means (k=4) via elbow & silhouette

Cluster	Tenure	Avg Charge	Churn Rate	Tactic
0	58 mo	$45	5%	Loyalty rewards
1	10 mo	$80	35%	Onboarding discount
2	30 mo	$65	15%	Bundle promotions
3	20 mo	$30	10%	Usage tips & credits

6. Deployment & Next Steps

- API: Flask /predict endpoint for churn probability
- Batch scoring: nightly CSV integration with CRM/BI
- Actions: High-risk (>=0.8): phone outreach
  Medium-risk (0.5-0.8): discount email
- Monitor drift & retrain monthly; A/B test offers

Prepared by Shaaf Chauhan