**Customer Churn Analysis Project (2025)**

**Dataset**

* Telco Customer Churn Dataset (Kaggle)

**Tasks**

* Clean and preprocess customer subscription data.
* Conduct exploratory data analysis (EDA) for churn patterns.
* Build a predictive model to classify customers at risk of churn.
* Evaluate model performance with accuracy, precision, recall, and confusion matrix.
* Analyze key drivers of churn through feature importance.
* Visualize churn behavior across critical factors like contract type and services.
* Provide actionable recommendations to reduce churn and improve retention.

**Skills Learned**

* Data cleaning and preprocessing with Pandas.
* Exploratory data analysis using Matplotlib and Seaborn.
* Encoding categorical variables for machine learning.
* Logistic regression modeling and model evaluation.
* Feature importance analysis.
* Business insight generation and visualization.

**Outcome**

Built a logistic regression model achieving 80% accuracy to predict customer churn. Identified key churn drivers such as contract type, phone service, and tech support. Developed business strategies targeting high-risk customer segments to improve retention.

**Purpose**

To understand why customers leave a telecom service provider and leverage data analytics techniques to predict churn and recommend actions to reduce attrition.

**Objectives**

* Clean and preprocess telecom customer data for analysis.
* Explore distributions and relationships influencing churn.
* Build a machine learning model to predict churn behavior.
* Evaluate the model’s performance and interpret feature importance.
* Provide practical recommendations based on data-driven insights.

**Approach**

The project began by importing and inspecting the dataset, checking for missing data and inconsistencies. Data was cleaned by converting numeric columns and encoding categorical variables. A logistic regression model was trained on 80% of the data and tested on the remaining 20%, stratified to preserve churn distribution. Model metrics and confusion matrix were used for evaluation. Feature importance was extracted to identify the main influencers of churn. Visualizations were created to illustrate churn patterns across contract types, phone service status, and billing methods. Finally, business recommendations were formulated to reduce churn by targeting at-risk groups.

**Key Insights**

* Month-to-month contract customers show the highest churn rates.
* Phone service availability reduces churn risk.
* Paperless billing customers need targeted communication to prevent churn.
* Services like tech support and online security help retain customers.
* Senior citizens and customers with dependents have distinct churn behaviors.

**Visual Summary**

* Bar charts displaying churn distribution by contract type, phone service, and paperless billing.
* Confusion matrix illustrating predictive accuracy of the logistic regression model.

**Business Recommendations**

* Promote long-term contracts through incentives and loyalty programs.
* Enhance phone service packages and customer support quality.
* Simplify billing processes and communicate benefits of paperless billing.
* Invest in tech support and security services to strengthen customer trust.
* Tailor retention campaigns specifically for senior citizens and families.

**Conclusion**

This project demonstrates how data analysis and machine learning can uncover critical factors behind customer churn, enabling telecom providers to make informed decisions to boost retention and profitability.