**Telecom Churn Analysis**

Abstract:

Customer churn is a critical concern for telecom companies due to its direct impact on revenue and customer retention costs. This project focuses on exploring patterns and factors contributing to customer churn using a real-world telecom dataset. Through systematic exploratory data analysis (EDA), we identify key trends, potential causes of churn, and data patterns that can inform retention strategies and future predictive modeling.

Introduction:

Telecommunication providers continuously face the challenge of retaining customers in a competitive market. Customer churn—the act of customers discontinuing service—represents a significant loss if not understood and mitigated. This project aims to analyze customer data to uncover insights into behaviors and attributes that distinguish churned users from retained ones. Understanding these patterns is a critical step before applying any predictive machine learning models.

Tools Used Python:

Core programming language used

Pandas: Data manipulation and analysis

NumPy: Numerical computations

Matplotlib & Seaborn: Data visualization

Jupyter Notebook / MYSQL workbench

Steps Involved in Building the Project:

1. Data Loading and Cleaning

* Loaded the dataset telecom\_churn.csv containing 243,553 rows and 14 columns.
* Verified no missing values or duplicates.
* Converted date\_of\_registration to datetime format.

1. Data Understanding

* Columns include customer attributes (age, gender, telecom partner, etc.) and service usage (calls, SMS, data).
* Target variable churn indicates whether a customer has left the service.

1. Univariate Analysis

* Numerical Features: Most users are in their 30s, have 0–2 dependents, and limited-service usage. Calls, SMS, and data usage are right-skewed.
* Categorical Features: Imbalance among telecom partners and state distribution. Gender distribution is roughly even.
* Churn: The dataset is highly imbalanced, with a majority of customers not having churned.

1. Visual Analysis

* Used histograms and boxplots to understand distributions.
* Count plots highlighted categorical frequency distributions.
* Churn distribution plot revealed the need for imbalance handling in future modeling.

1. Data Preparation for Modeling (Next Step)

* Identified that features like usage and demographics can be used for prediction.
* Recommended encoding categorical variables and scaling numeric ones for machine learning.
* Suggested balancing techniques like SMOTE or class weights due to churn imbalance.

Conclusion:

This exploratory data analysis laid the foundation for understanding customer behavior in the telecom sector. It revealed critical patterns related to churn, such as service usage intensity and demographic influence. With clean, well-understood data, the next step would be to build a churn prediction model using classification algorithms. Overall, this project helps align data insights with business strategies aimed at reducing customer churn.

[P-2\EDA OF TELECOM CHURN.ipynb](P-2/EDA%20OF%20TELECOM%20CHURN.ipynb)