Why Churn Analysis?

Customer churn is one of the biggest challenges businesses face today. Identifying why customers leave and predicting who is likely to churn allows companies to proactively engage them with personalized retention strategies. Understanding why customers reduce or stop purchasing whether due to pricing shifts, fulfillment issues, product availability, or competition is essential. With advanced data analytics, an organization can identify at-risk customers early and proactively engage them with tailored strategies, such as volume-based incentives, improved lead times, or customized packaging solutions.

This project demonstrates how advanced analytics, and predictive modeling can empower airlines to improve customer retention, loyalty programs, and overall profitability. By leveraging data, we can:

- Segment customers by purchasing behavior, frequency, and volume to identify high value clients showing signs of disengagement.
- Analyze key churn drivers, such as delivery delays, pricing sensitivity, or service responsiveness, using historical sales and CRM data.
- Optimize forecasting models to predict future churn and allocate resources accordingly, especially in peak seasons or for strategic accounts.
- Enhance sales and marketing strategies using customer lifetime value (CLV) to focus on retention where it matters most.

This project isn’t just about numbers it’s about understanding the human side of customer behavior and using data to create a seamless, rewarding experience that keeps customers engaged. Through a combination of machine learning, customer segmentation, and marketing analytics, this initiative offers actionable insights to boost customer satisfaction, increase retention, and sustain long-term profitability in the highly competitive market.

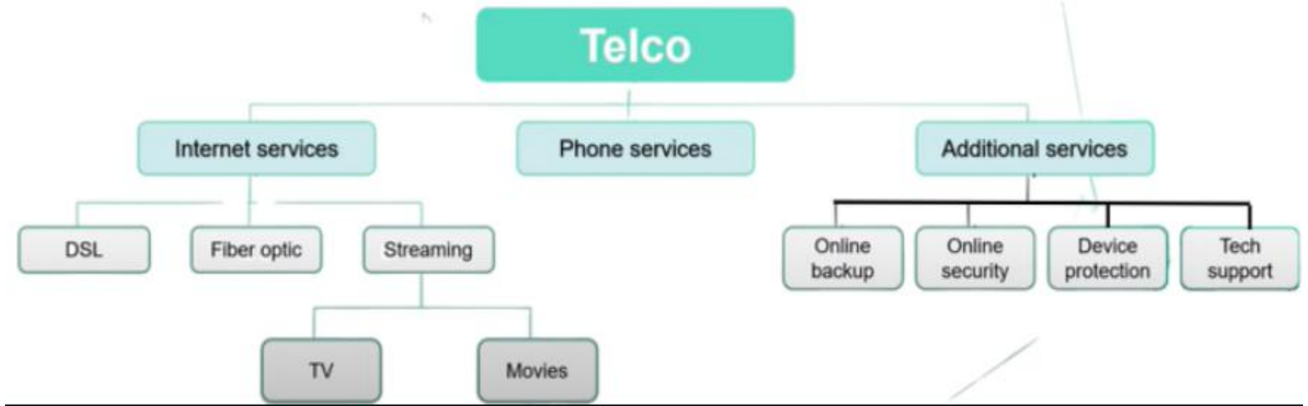
Project Overview

Project Flow

- **Data Source**: CSV file (Customer Churn Data)
- **ETL Processing**: MySQL Workbench (staging area and ODBC Driver connection with Power BI)
- **Database**: MySQL Workbench
- **Data Analysis & Predictive Analysis**: SQL Query execution, DAX Queries, Python code for Predictive Analysis in Jupyter Notebook
- **Visualization**: Power BI Dashboard

Key Features

- **Data Processing, Transformation & Cleaning** → Handles missing values, data types, and duplicates.
- **Query Execution & Insights** → Runs SQL queries in MySQL Workbench for data analysis.
- **Predicting Churn with Machine Learning** → Uses Random Forest Algorithm to Predict churn using Python
- **Seamless Power BI Integration** → Real-time visualization & business insights.



Key Questions

1) **Customer Demographics & Segmentation**
   - How does the number of referrals correlate with customer retention?
   - Are certain states experiencing a higher churn rate than others?
2) **Subscription & Service Analysis:**
   - Do customers with premium support have lower churn rates compared to those without?
   - What is the most common internet type among churned customers?
3) **Churn Behavior & Reasons:**
   - What are the top churn reasons among customers?
   - How many customers left due to service-related reasons vs. pricing-related reasons?
4) **Financial & Revenue Impact:**
   - What is the average monthly charge of churned vs. non-churned customers?
   - Do customers with higher total charges tend to stay longer?
5) **Predictive Insights & Retention Opportunities:**
   - Can we predict churn based on tenure in months?
   - Are there specific combinations of services that lead to lower churn rates?

Key Findings

- **High-risk groups** include unmarried customers (27%), Fiber Optic users (41%), and those on month-to-month contracts (88% Churn rate), meaning that we have a high possibility to convert 1-time users into permanent users by targeting promotional and marketing strategies.
- **Regional concerns** exist, with Jammu & Kashmir (58.47% churn) and Assam (41.09%) showing the highest rates.
- **Competitor influence is strong**, with better devices (16.69%) and better offers (15.82%) being the top churn reasons.
- **Premium support reduces churn significantly** (15.41% vs. 31.70%), making it a key retention tool.
- **High monthly charges ($73 vs. $61)** are linked to higher churn, suggesting affordability concerns.
- **Bundled services & security features improve retention**, emphasizing cross-selling opportunities.

Predicting Customer Churn with ML

1) **Data Preprocessing & Feature Engineering:**
   - We first drop irrelevant columns like Customer_ID, Churn_Category, and Churn_Reason since they don't contribute to prediction.
   - Next, categorical variables such as Gender, Contract Type, and Payment Method are label-encoded to convert them into numerical form for model compatibility.
   - The target variable Customer_StaturesultsStayed = 0 and Churned = 1.

2) **Model Training with Random Forest:**
   - The dataset is split into training (80%) and testing (20%) using train_test_split.
   - A Random Forest Classifier with 100 estimators is trained on the dataset.
   - Post-training, we evaluate the model using a confusion matrix and classification report, checking metrics like precision, recall, and F1-score.

3) **Feature Importance Analysis:**
   - We extract feature importance scores from the trained model to understand which factors contribute the most to churn.
   - A bar plot is generated to visualize the top features, helping us interpret the model and improve feature selection.

4) **Making Predictions on New Data:**
   - We load new customer data from an Excel file and apply the same preprocessing steps (dropping unnecessary columns and encoding categorical variables using previously saved encoders).
   - Predictions are made using the trained Random Forest model, and customers predicted to churn (Customer_Status_Predicted = 1) are filtered.
   - The results are saved as a CSV file for further business action.

5) **Results:**
   - The model gave an accuracy of 85%