**Problem Statement:**

In today’s highly competitive business landscape, customer retention is essential for long-term success. One critical approach to improving retention is churn analysis, which helps identify the reasons behind customer departures and the likelihood of future churn. By leveraging advanced data analytics and machine learning, businesses can predict which customers are most at risk of leaving and uncover the underlying factors that drive these decisions. Armed with these insights, companies can take proactive measures to enhance customer satisfaction and loyalty.

While this project is focused on churn analysis for a telecom company, the methodology is widely applicable across various industries such as retail, finance, and healthcare. By exploring methods, tools, and best practices, we aim to reduce churn, increase customer loyalty, and transform raw data into actionable insights for sustainable business success.

**Objectives:**

The objective of this project is to create an end-to-end solution, including an ETL process, and develop a comprehensive Power BI dashboard. The key objectives are:

* Visualize and analyse customer data based on:
  + Demographics
  + Geography
  + Payment & account information
  + Services utilized
* Understand the profile of customers who churned and identify potential areas for marketing campaigns.
* Develop a method to predict future customer churn.

**Required Metrics:**

* Total Customers
* Total Churn and Churn Rate
* New Joiners

**Steps**

**1: Building the ETL Process in SQL Server**

The first step involves loading the customer data from the source file (CSV or Excel) into a staging area in Microsoft SQL Server. SQL Server is more efficient at handling recurring data loads and maintaining data integrity compared to Excel.

**Key tasks:**

1. **Create a Database**: Design the database schema to hold the customer data.
2. **Import CSV into SQL Server**: Use the SQL Server Import Wizard to import the CSV data into a staging table.
   * Set customerId as the primary key.
   * Allow NULL values for all other columns.
3. **Data Exploration**:
   * Identify distinct values and patterns.
   * Check for and handle missing (NULL) values.
4. **Data Cleaning**:
   * Remove rows with NULLs, or appropriately handle missing data.
   * Insert cleaned data into a production table.
5. **Create Views**: Develop SQL views for Power BI, separating churned and active customers (train data) from new joiners (test data) for future modelling.

**2: Power BI Data Transformation**

Use Power BI to transform and clean the data for visual analysis. Apply necessary filters and create calculated columns for further insights.

**3: Power BI Measures**

Create the following custom measures within Power BI to track critical business metrics:

* **Total\_Customers**: The overall count of customers.
* **New\_Joiners**: The number of new customers.
* **Total\_Churn**: The total number of customers who churned.
* **Churn\_Rate**: The percentage of customers who left.

**4: Power BI Visualization**

Design a Power BI dashboard that visualizes the customer data at various levels (demographic, geographic, payment, etc.). This dashboard should also include:

* An analysis of the churner profile.
* Potential areas where marketing campaigns can help reduce churn.

With initial insights gathered from the visualizations, we can proceed to the next step: predictive modelling.

**5: Predict Customer Churn**

The next phase involves predicting customer churn using machine learning. The training data will consist of customers who churned and those who stayed, while the test data will be comprised of new joiners.

**Steps:**

1. **Export Views to Excel**: Import the customer data views into an Excel file from SQL Server.
2. **Churn Prediction Model (Random Forest)**:
   * Open Jupyter Notebook and build a Random Forest model to predict future churners.
   * Required Python libraries: pandas, numpy, matplotlib, seaborn, scikit-learn, and joblib.
   * Install libraries using Anaconda:

bash

Copy code

pip install pandas numpy matplotlib seaborn scikit-learn joblib

1. **Train the Model**: Train the Random Forest model using the data from customers who churned and stayed, then apply it to the new joiners.

**6: Power BI Visualization of Predicted Data**

After predicting the customers likely to churn, further analysis can be conducted in Power BI.

**Tasks:**

* **Import Predicted Data**: Load the predicted data.
* **Create New Measures**: Develop additional measures such as:
  + **Count of Predicted Churners**.
  + **Title Predicted Churners** for easy reference.
* **Churn Prediction Page**: Add a dedicated page in Power BI to visualize the predictions.