Q1. Analysing and Predicting Customer Churn

Background

You are given a dataset containing customer information for a hypothetical subscription-based business. Your task is to explore the data, perform analysis, build a predictive model, and present insights to answer the key question: **"What factors are influencing customer churn, and how can the company reduce it?"**

Here's the link to download Dataset:

[Telco Customer Churn Dataset on Kaggle](#)

Objectives

1. **Data Exploration and Cleaning**:
   - Inspect and clean the dataset (handle missing values, outliers, and data types).
   - Explore patterns or anomalies in the data (e.g., demographic or usage trends).
2. **Exploratory Data Analysis (EDA)**:
   - Visualise key relationships in the data, such as customer age vs. churn rate, service usage vs. churn rate, etc.
   - Calculate summary statistics and identify any interesting patterns or insights.
   - Formulate hypotheses about factors influencing customer churn.
3. **Feature Engineering**:
   - Create at least three additional features (e.g., average usage per month, customer lifetime value, etc.) that could improve predictive accuracy.
4. **Predictive Modelling**:
   - Build a predictive model to classify whether a customer will churn (binary classification).
   - Test at least two different algorithms (e.g., Logistic Regression, Decision Tree, or Random Forest) and compare their performance.
   - Optimise your model for accuracy and explainability.
5. **Business Insights & Recommendations**:
   - Based on the results of your analysis, present 2-3 actionable recommendations the company could implement to reduce customer churn.
   - Explain the key drivers of churn based on your model and any patterns identified during EDA.

Deliverables

1. **Code and Documentation**:
   - Submit your code in a Jupyter Notebook or script file, with comments explaining each step.

- ○ Provide a README file describing the data, methods, and any assumptions made.
2. **Model Performance Report**:
   - ○ Include a report of model performance metrics (e.g., accuracy, precision, recall, and F1 score).
   - ○ Summarise the feature importance and interpret the results of the top-performing model.
3. **Insights Presentation**:
   - ○ Create a short slide deck (5-7 slides) to present the findings, including:
     - ■ Key factors influencing customer churn.
     - ■ Model results and performance metrics.
     - ■ Actionable recommendations for reducing churn.

Assessment Criteria

- ● **Technical Skills**: Accuracy of data cleaning, feature engineering, and modelling.
- ● **Analytical Skills**: Quality of insights drawn from EDA and hypothesis formulation.
- ● **Creativity**: Innovation in feature engineering and the thought process behind recommendations.
- ● **Communication**: Clarity of code, documentation, and presentation of insights.

Q2. Assignment: Time Series Analysis and Forecasting for Energy Consumption

Background

You are a data scientist tasked with analysing and forecasting energy consumption for a utility company. The company aims to anticipate future energy demand, identify consumption trends, and make strategic decisions based on these forecasts. **Your task is to analyse historical energy consumption data, develop a forecast model, and identify actionable insights and recommendations.**

Download Link:

 Energy Consumption Dataset (You can substitute with another dataset if preferred).

Objectives

1. **Data Exploration and Preprocessing**:
   - ○ Clean the dataset, handle missing values, and conduct basic EDA.
   - ○ Perform time series decomposition to observe trends, seasonality, and residuals.
2. **Feature Engineering**:

- ○ Create time-based features (e.g., hour, day of the week, month).
   - ○ Create lagged features to capture recent past values (e.g., energy consumption in the previous hour, day, week).
   - ○ Use temperature and weather to capture how environmental factors affect consumption.
3. **Modelling**:
   - ○ Build at least two models (e.g., SARIMA and LSTM or ARIMA and XGBoost) to forecast energy consumption for the next month.
   - ○ Use backtesting or cross-validation to evaluate model performance and choose the best model.
4. **Analysis and Insights**:
   - ○ Analyse the importance of different factors affecting consumption, such as temperature and specific hours of the day.
   - ○ Identify peak consumption periods and possible anomalies.
5. **Actionable Recommendations**:
   - ○ Based on forecasted demand, suggest two to three strategies for energy distribution and demand management.

Deliverables

1. **Code Notebook**: Submit the full code used, with comments explaining each step.
2. **Forecast Report**: Include a report on model performance, model comparisons, and any challenges encountered.
3. **Insights Presentation**: A slide deck (5-7 slides) summarising key findings, forecasted trends, and recommendations.