**Problem Statement: Energy Efficiency in Commercial Buildings Using Seattle Building Energy Benchmarking Data**

**1. Project Title**

**Predicting Energy Efficiency of Commercial Buildings Using Seattle Energy Benchmarking Data**

**2. Background**

Buildings are one of the largest consumers of energy in urban areas, and improving energy efficiency in commercial buildings is a critical step toward reducing overall energy consumption and emissions. By analyzing historical energy usage data from commercial buildings, it is possible to identify patterns and factors that affect energy efficiency. Predictive models can be developed to forecast energy consumption and help building managers take proactive measures to optimize energy use.

The Seattle Building Energy Benchmarking dataset provides detailed information about the energy usage, building characteristics, and environmental data from various commercial buildings in Seattle. By leveraging this data, we can predict energy consumption and uncover key factors that influence a building’s energy efficiency.

**3. Problem Statement**

The goal of this project is to develop a machine learning model that predicts the energy efficiency (measured as **Energy Use Intensity - EUI**) of commercial buildings based on building characteristics such as size, type, and usage patterns. The project will also identify the key factors contributing to energy efficiency and provide recommendations for improving energy consumption across different types of buildings.

Key questions this project aims to answer:

* Can we predict the energy efficiency (EUI) of commercial buildings based on building characteristics and operational data?
* What are the key factors (e.g., building type, age, square footage) that influence energy efficiency?
* How can machine learning models help building owners reduce energy consumption and meet sustainability goals?

**4. Objectives**

* **Primary Objective**: To build a machine learning model that predicts a building's energy efficiency (EUI) based on building attributes such as size, year built, and primary use.
* **Secondary Objectives**:
  + Identify which building characteristics are most important in determining energy efficiency.
  + Explore the relationship between building features (e.g., square footage, occupancy, building type) and energy consumption.
  + Provide actionable insights for building managers to improve energy efficiency.

**5. Scope**

* **Data Source**: The Seattle Building Energy Benchmarking dataset contains energy usage, building characteristics, and environmental data for commercial buildings in Seattle.
  + Link to Dataset
* **Target Variable**: **Energy Use Intensity (EUI)** is the key target variable to predict, as it measures the energy consumed per square foot of building area.
* **Modeling Approach**: The project will use regression models such as **Random Forest**, **XGBoost**, and **Gradient Boosting** to predict energy efficiency, with a focus on feature importance analysis.

**6. Challenges**

* **Missing Data**: Some building attributes (e.g., year built, square footage) or energy data might be missing, requiring imputation.
* **Outliers**: Some buildings might show extreme energy consumption values that could skew model results.
* **Multicollinearity**: Some features (e.g., building type and square footage) might be highly correlated, which could impact model performance.

**7. Success Criteria**

* **Model Accuracy**: A successful model should predict energy efficiency with a high degree of accuracy, evaluated using metrics such as **Mean Absolute Error (MAE)** and **Root Mean Squared Error (RMSE)**.
* **Feature Importance**: The project should identify key building characteristics that influence energy consumption, providing actionable insights for building management.
* **Actionable Insights**: Recommendations for reducing energy consumption based on model results, which building managers can implement to improve efficiency.

**8. Deliverables**

* **Machine Learning Models**: A trained and validated model for predicting energy efficiency (EUI).
* **Feature Importance Analysis**: A detailed analysis of the most important building characteristics affecting energy efficiency.
* **Code & Documentation**: A Jupyter Notebook containing all steps, from data preprocessing to model evaluation, with thorough documentation.

import pandas as pd

from statsmodels.stats.outliers\_influence import variance\_inflation\_factor

from sklearn.model\_selection import train\_test\_split

from sklearn.linear\_model import LinearRegression

from sklearn.metrics import mean\_squared\_error, mean\_absolute\_error

from sklearn.ensemble import RandomForestRegressor

from sklearn.ensemble import GradientBoostingRegressor

from sklearn.linear\_model import Ridge, Lasso

from sklearn.model\_selection import GridSearchCV

from sklearn.model\_selection import RandomizedSearchCV

import numpy as np

from sklearn.ensemble import GradientBoostingRegressor

from sklearn.ensemble import RandomForestRegressor

from sklearn.metrics import mean\_squared\_error, mean\_absolute\_error

from sklearn.ensemble import GradientBoostingRegressor

import matplotlib.pyplot as plt

import seaborn as sns