**Summary Report: Petrol Consumption Prediction Analysis**

**Objective:**

The goal of this analysis was to predict **Petrol Consumption** using a regression model based on a dataset with various features. We applied two machine learning models: **Decision Tree Regressor** and **Random Forest Regressor**, and evaluated their performance.

**Key Steps:**

1. **Data Preprocessing**:
   * The dataset was loaded, and the target variable, **Petrol\_Consumption**, was separated from the feature columns.
   * The dataset was split into training (80%) and testing (20%) sets.
   * Missing values and potential outliers were addressed through basic preprocessing steps.
2. **Model Selection**:
   * Two models were trained:
     + **Decision Tree Regressor**: A simple, interpretable model based on splitting data at decision points.
     + **Random Forest Regressor**: An ensemble method that uses multiple decision trees to improve performance and robustness.
3. **Model Evaluation**:
   * **Mean Absolute Error (MAE)**, **Mean Squared Error (MSE)**, and **R-squared (R²)** were used to evaluate the models' performance on the test set.

**Results:**

* **Decision Tree Regressor**:
  + **MAE**: 94.3
  + **MSE**: 17,347.7
  + **R²**: -1.59 (negative, indicating poor model fit)
* **Random Forest Regressor**:
  + **MAE**: 53.96
  + **MSE**: 6,835.46
  + **R²**: -0.019 (still negative but better than Decision Tree)

**Key Findings:**

1. **Model Performance**:
   * The **Random Forest Regressor** outperformed the **Decision Tree Regressor** in terms of **MAE**, **MSE**, and **R²**. It provided better predictions but still showed poor performance with negative R² values, indicating the need for further improvement.
2. **Model Limitations**:
   * Both models performed poorly, as indicated by the negative R² scores, meaning they did not explain much of the variation in petrol consumption.
   * **Overfitting** or **underfitting** could be an issue, suggesting that the models either capture noise or fail to generalize.
3. **Data Quality**:
   * The negative R² scores imply that there may be issues with the data, such as insufficient feature selection, need for better preprocessing, or the need for more relevant features to predict petrol consumption effectively.

**Conclusions:**

* The **Random Forest** model slightly outperformed the **Decision Tree** model, but both models require improvement.
* **Feature engineering**, **hyperparameter tuning**, and **cross-validation** are necessary steps to improve model performance.
* Exploring alternative models (e.g., **Gradient Boosting** or **Neural Networks**) and enhancing the dataset could potentially lead to better predictions.

**Next Steps:**

1. **Data Enhancements**: Perform further preprocessing such as scaling, handling missing data, and creating new features.
2. **Model Tuning**: Optimize hyperparameters using techniques like **Grid Search** or **Randomized Search**.
3. **Exploring Advanced Models**: Consider models like **Gradient Boosting** or **XGBoost** for better results.
4. **Cross-Validation**: Implement **k-fold cross-validation** to validate the models' performance more robustly.