**Vehicle Price Prediction Project**

**Project Overview**  
This project analyzes a dataset of used vehicles and leverages machine learning to predict their selling prices. By examining features like vehicle specifications, mileage, and engine type, it provides valuable insights into price determinants while developing an accurate prediction framework.

**Objectives**

1. Perform data exploration to understand key factors influencing vehicle prices.
2. Preprocess and clean data to ensure quality for modeling.
3. Develop and evaluate machine learning models for robust price prediction.
4. Derive actionable insights from data analysis for the automotive market.

**Dataset Overview**  
The dataset is sourced from [Kaggle](https://www.kaggle.com/), a platform for data science and machine learning projects. It includes various vehicle attributes:

* **Features:** Car model, fuel type, mileage, engine specifications, and price.
* **Size:** (Details to be confirmed after processing).

**Key Features:**

* **Mileage:** Fuel efficiency measure.
* **Engine:** Displacement capacity of the vehicle.
* **Torque:** Key indicator of engine performance.
* **Seats:** Vehicle seating capacity.

**Data Preprocessing**

* **Handling Missing Data:** Columns like mileage, engine, and seats had missing entries that were removed.
* **Feature Engineering:**
  + Extracted numerical data from text fields like torque (e.g., splitting into numeric torque and RPM values).
  + Created new columns for better feature representation (e.g., torque\_rpm).
* **Normalization:** Scaled numerical features to ensure uniformity.
* **Encoding:** Transformed categorical variables into numerical formats for compatibility with models.

**Modeling and Evaluation**

* **Implemented Models:**
  + Random Forest Regressor
  + Decision Tree Regressor
* **Performance Metrics:**
  + R-squared, Mean Absolute Error (MAE), and Mean Squared Error (MSE) were used to evaluate model accuracy.

**Key Insights and Findings**

1. **Feature Importance:** Engine displacement and mileage had a high impact on price predictions.
2. **Data Cleaning Impact:** Proper handling of missing values and feature transformations improved model accuracy.
3. **Model Performance:**
   * Random Forest outperformed other models, providing the most reliable predictions.

**Conclusion**  
This project demonstrates how machine learning can be applied to predict vehicle prices effectively. The insights gained from data exploration and feature analysis offer actionable knowledge for buyers, sellers, and manufacturers. Future improvements could include integrating more diverse datasets and fine-tuning models for enhanced accuracy.