Vehicle Price Prediction (vehicle.ipynb)

- **Problem Statement**: The project aims to predict the price of a vehicle based on its features. This is a regression problem.
- **Dataset Used**
  :
  - The dataset was loaded from CAR DETAILS FROM CAR DEKHO.csv.
  - The target variable is 'selling_price'.
  - Features include 'name', 'year', 'km_driven', 'fuel', 'seller_type', 'transmission', 'owner'.

- **Methodology and Approach**
  :
  - **Data Preprocessing**
    :
    - The 'name' column (car model) was dropped as it has too many unique values for simple encoding.
    - Categorical features ('fuel', 'seller_type', 'transmission', 'owner') were converted into numerical representations using pd.get_dummies() (one-hot encoding).
    - No explicit feature scaling was performed on the numerical features ('year', 'km_driven') before training the final models, though this can be important for some regression algorithms.

  - **Model Training**
    :
    - The data was split into training (70%) and testing (30%) sets.
    - Several regression models were trained and evaluated:
      - Linear Regression (LinearRegression)
      - Lasso Regression (Lasso)
      - Decision Tree Regressor (DecisionTreeRegressor)
      - Random Forest Regressor (RandomForestRegressor)

  - **Model Evaluation**
    :

    - Models were evaluated using the R-squared (R2) score and Mean Squared Error (MSE).

- **Results and Conclusion**
  :
  - **Linear Regression**: R2 score of 0.44.
  - **Lasso Regression**: R2 score of 0.44.
  - **Decision Tree Regressor**: R2 score of 0.53.
  - **Random Forest Regressor**: R2 score of 0.60.
  - The Random Forest Regressor provided the highest R2 score, indicating it explained the most variance in the selling price among the models tested.
  - The Random Forest Regressor model was saved to vehicle_price_model.pkl.