Car Dheko Used Car Price Prediction

Capstone project –3

By

S.Vinodha

Objective

- To create an accurate and user-friendly streamlit tool that predicts the prices of used cars based on various features.

- This tool is an interactive web application for both customers and sales representatives to use seamlessly.

- It allows users to input car details and receive an estimated price instantly.

Data Processing

Import and concatenate:
- Imported all unstructured city's dataset and converted it into a structured format.
- Assigned all rows in 'City' column with the name of the respective city.
- Concatenated all datasets a single dataset.

Handling Missing Values:
- Identified and filled or removed missing values in the dataset.using techniques like mean, median, or mode imputation..

Standardising Data Formats:
- Changing data to correct data types and format.

Encoding Categorical Variables
- Converted categorical features into numerical values using label encoding .
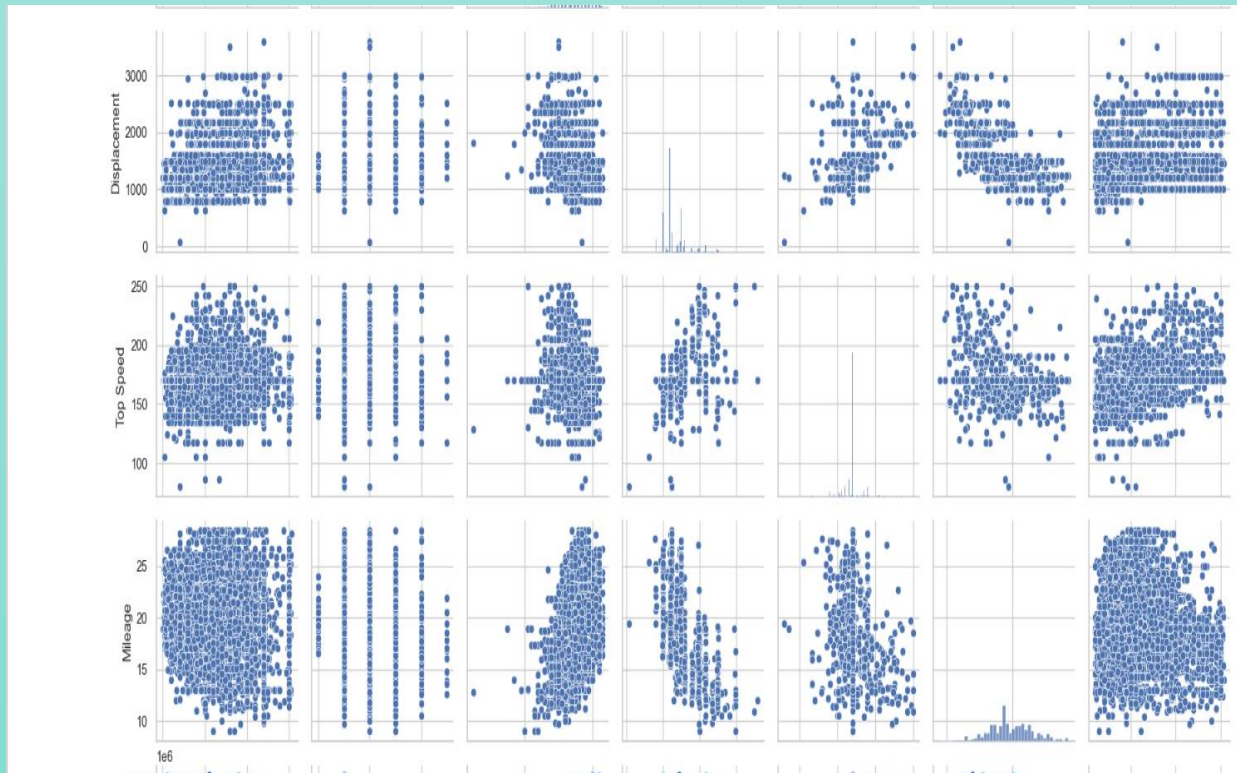
Normalizing Numerical Features
- Scaling numerical features to a standard range, usually between 0 and 1. using Min-Max Scaling.
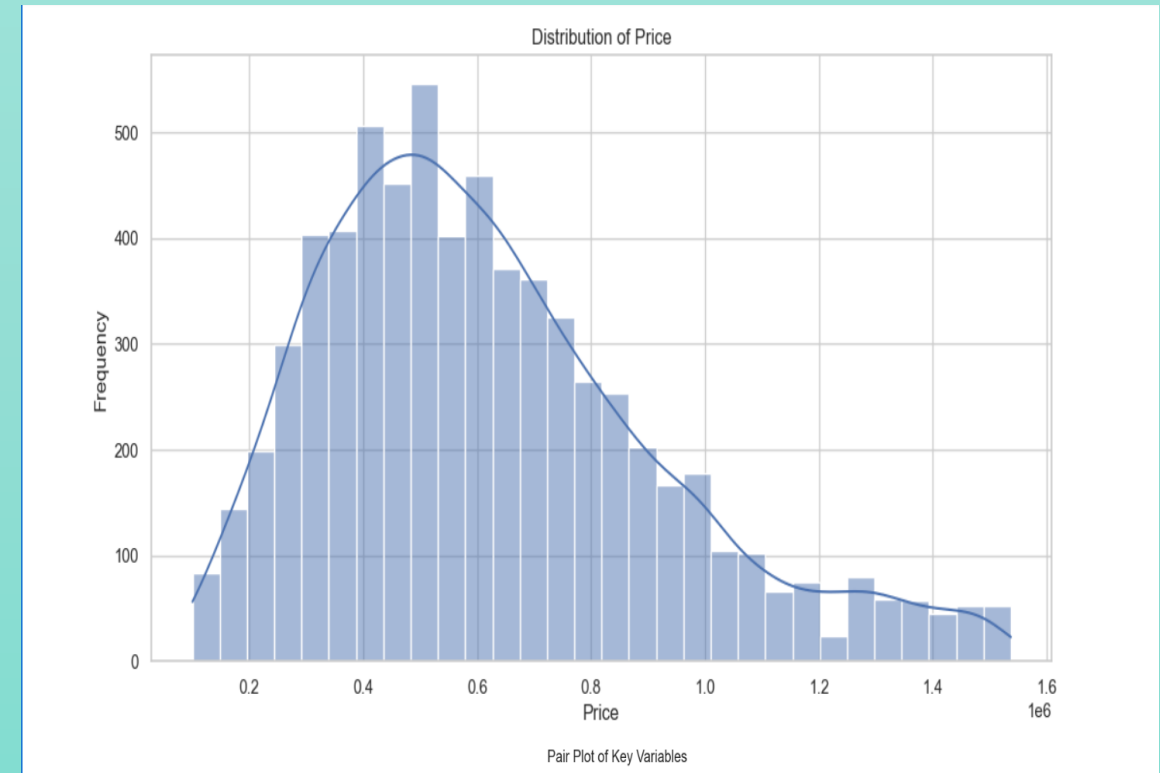
Removing Outliers
- Identified and removed outliers in the dataset to avoid skewing the model.Using IQR (Interquartile Range) method .

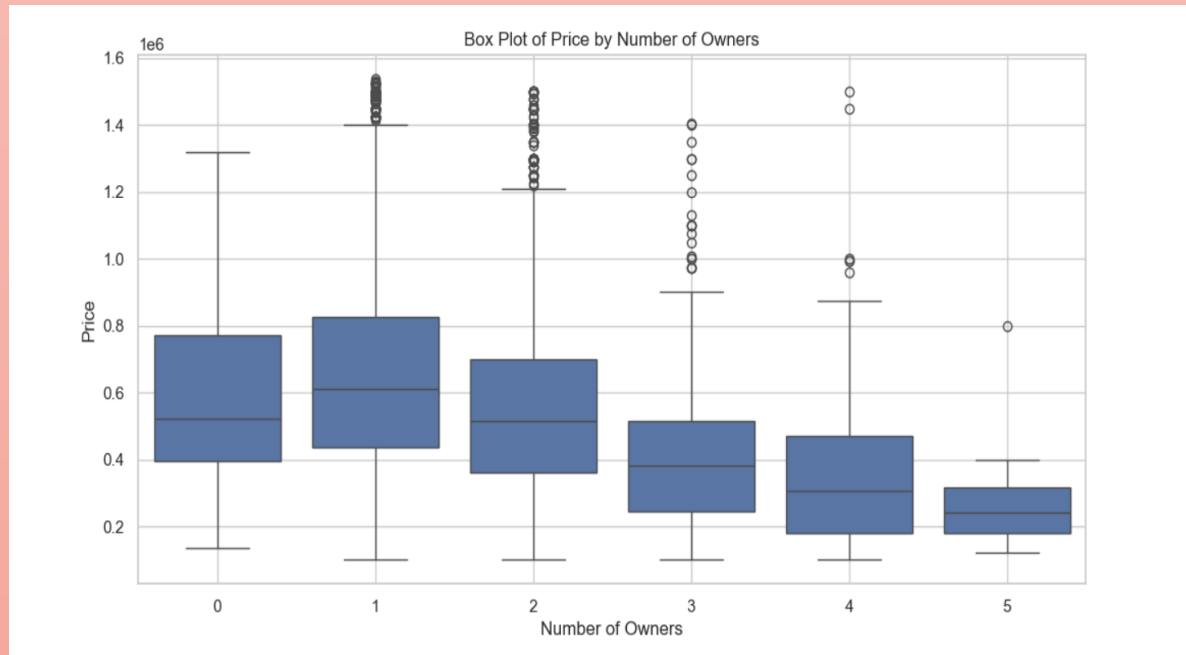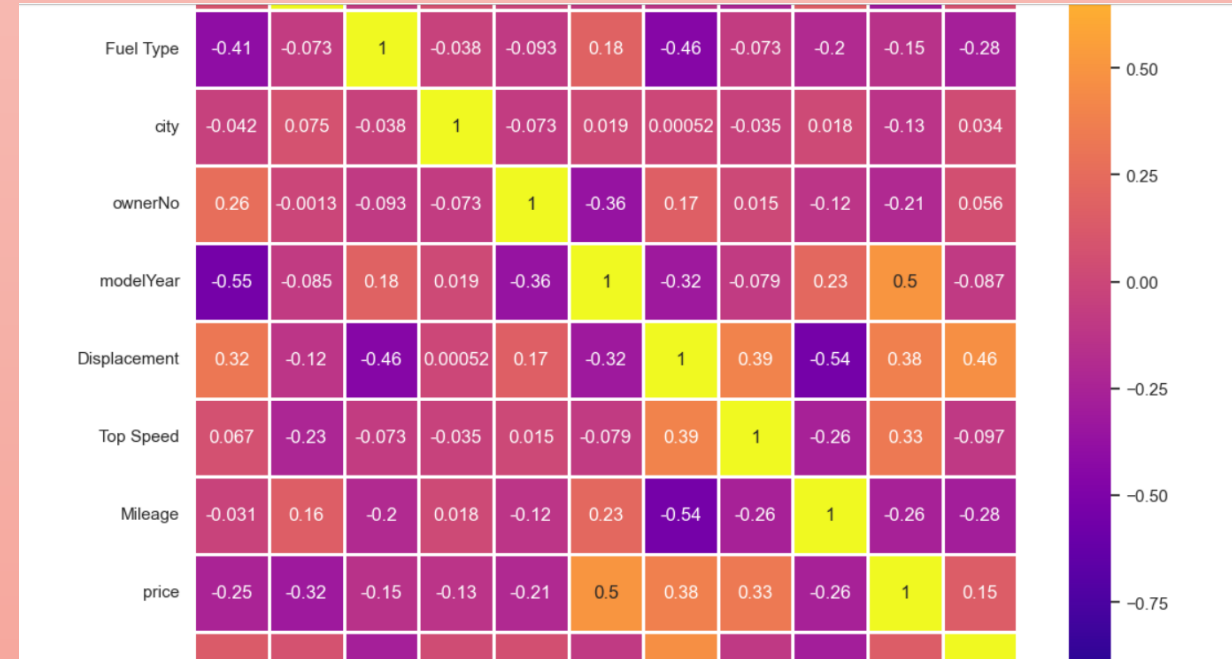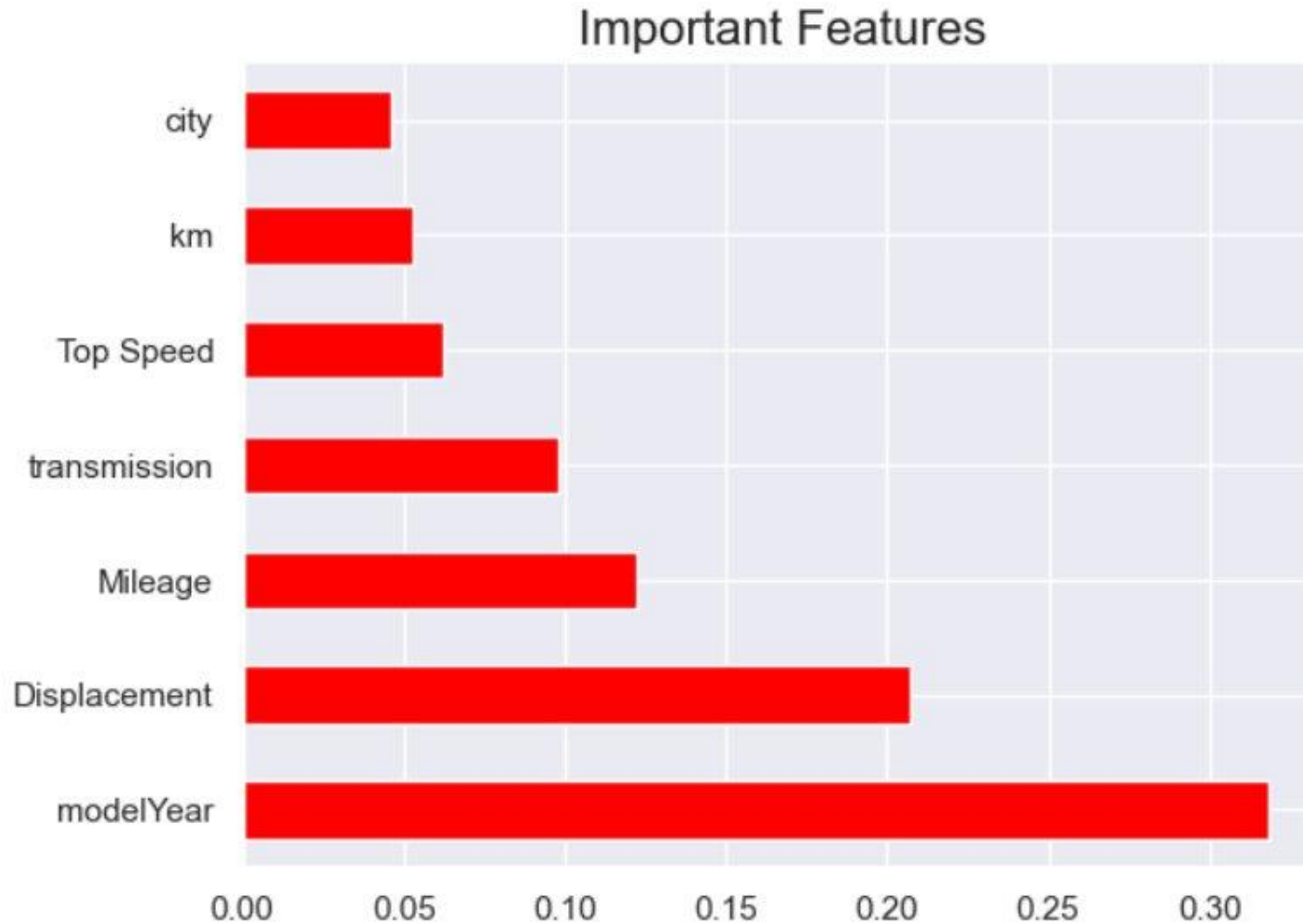# Exploratory Data Analysis (EDA)

scatter plots

histograms

Data Visualization

box plots

correlation heatmaps

# Feature Selection

Important Features

Model Development

Models	Mean Squared Error	Root Mean Squared Error	Explained Variance Score	R-Square Score / Accuracy
Linear Regression	3.028170e+10	174016.385379	0.653846	0.653786
Support Vector Rregression	9.068865e+10	301145.559025	0.000308	-0.036852
Decision Tree Regressor	2.144056e+10	146425.950971	0.754874	0.754868
Random Forest Regressor	1.232270e+10	111007.668662	0.859117	0.859113
Ridge	3.028137e+10	174015.422194	0.653850	0.653790

Model Performance Metrics
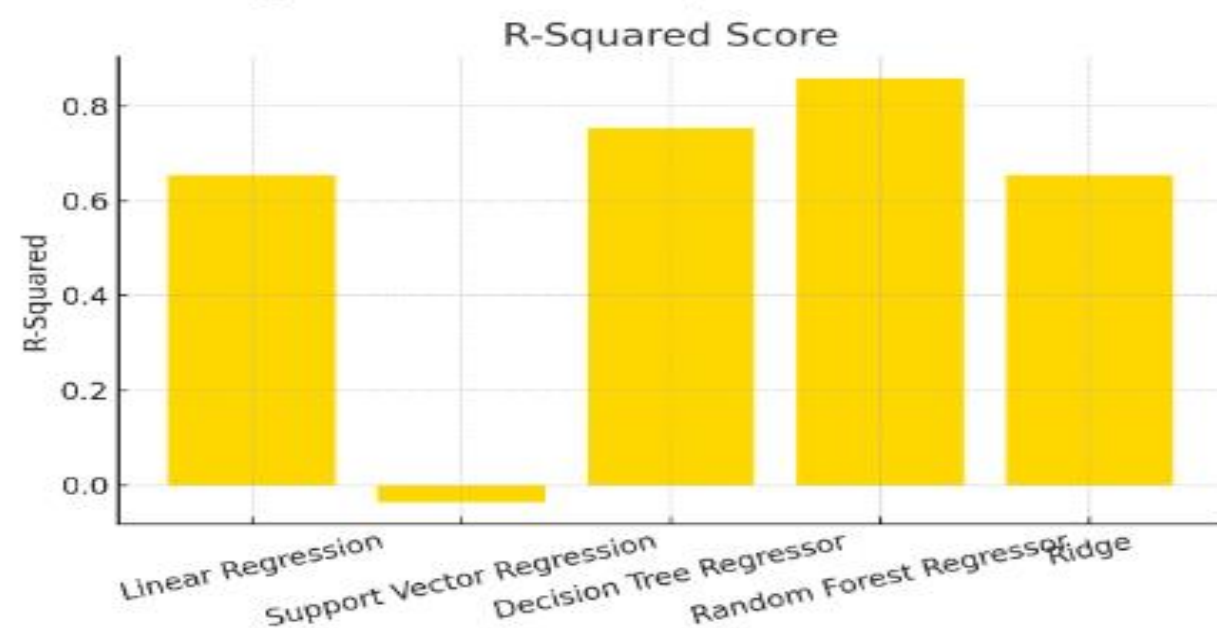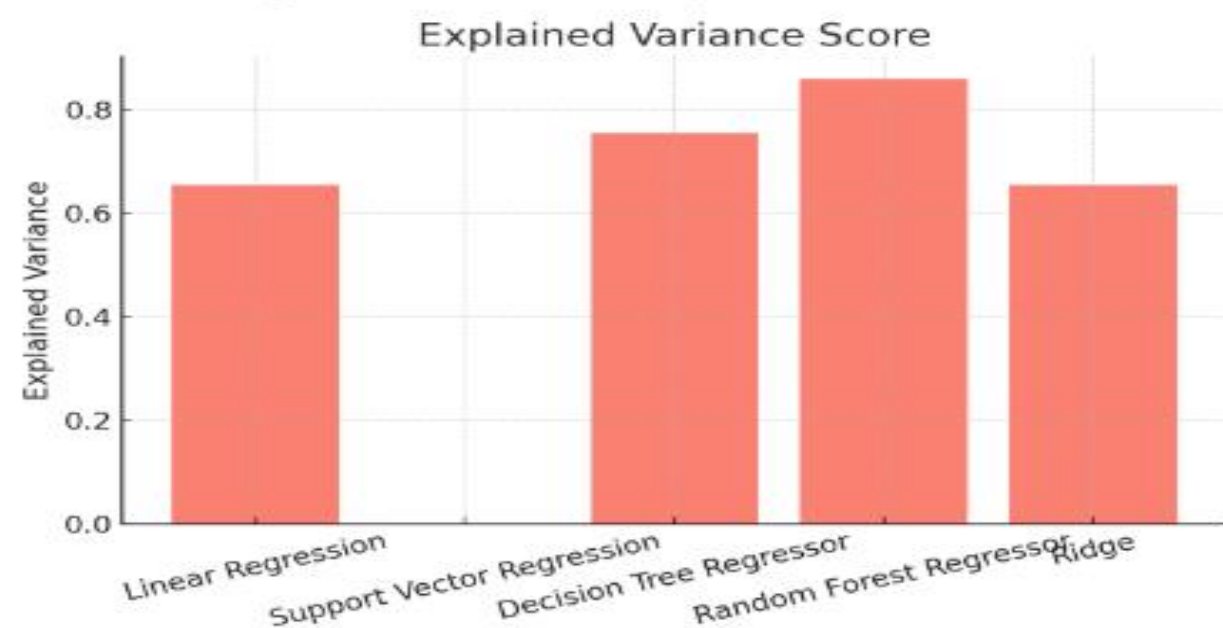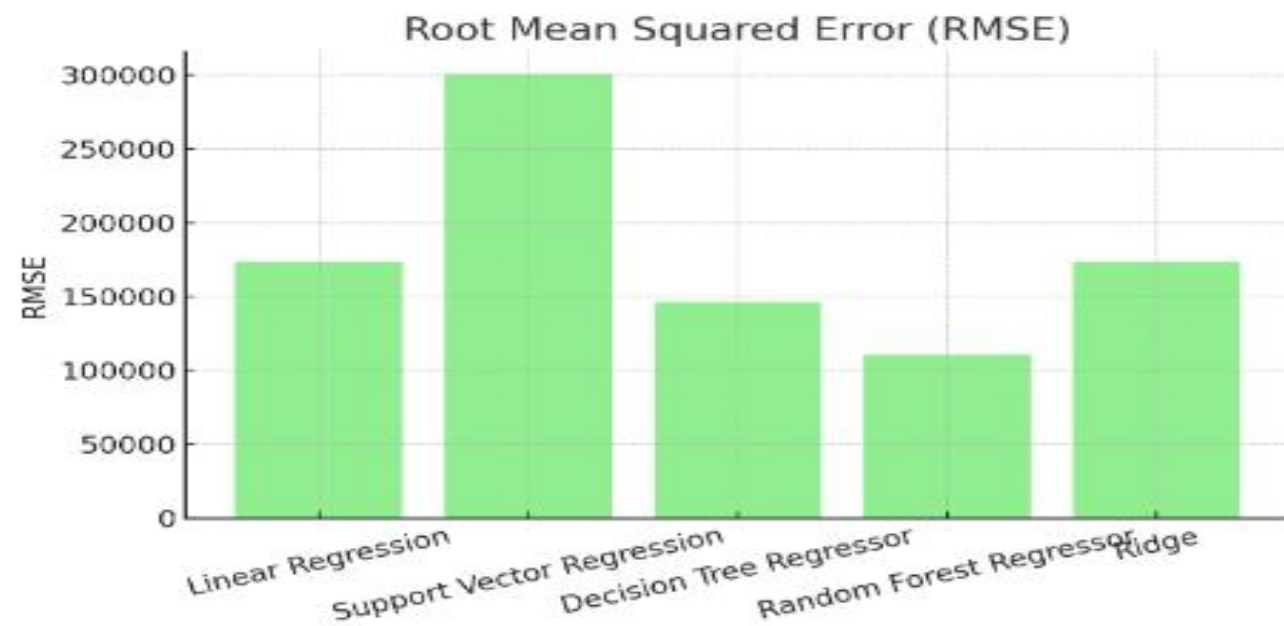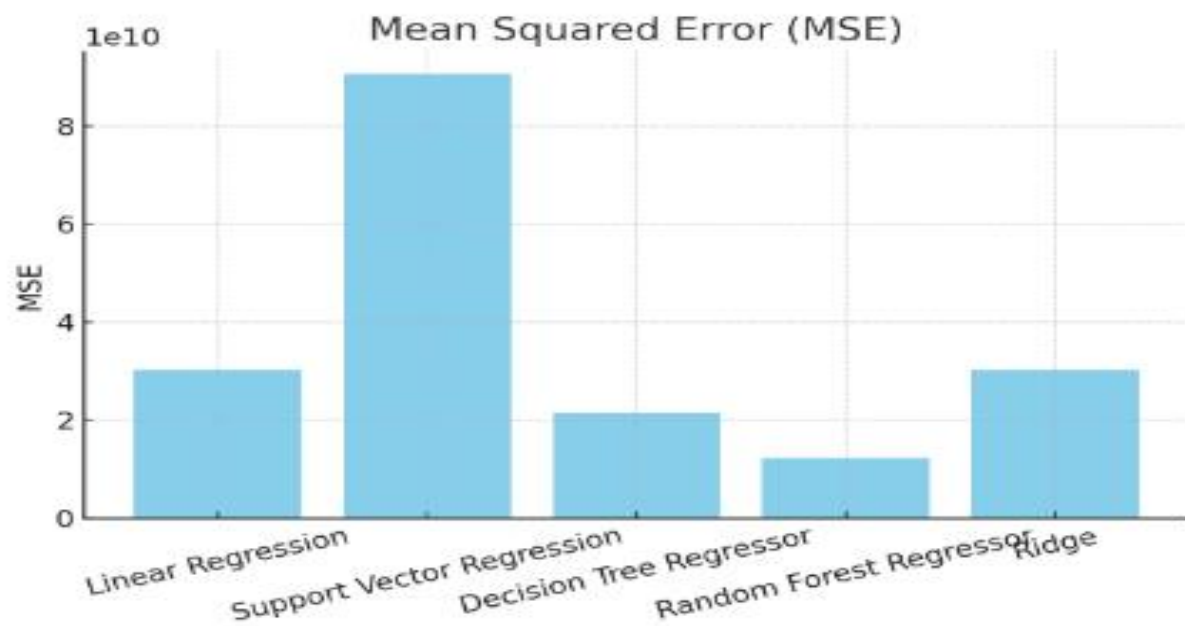
Cross-validation -to ensure robust performance.
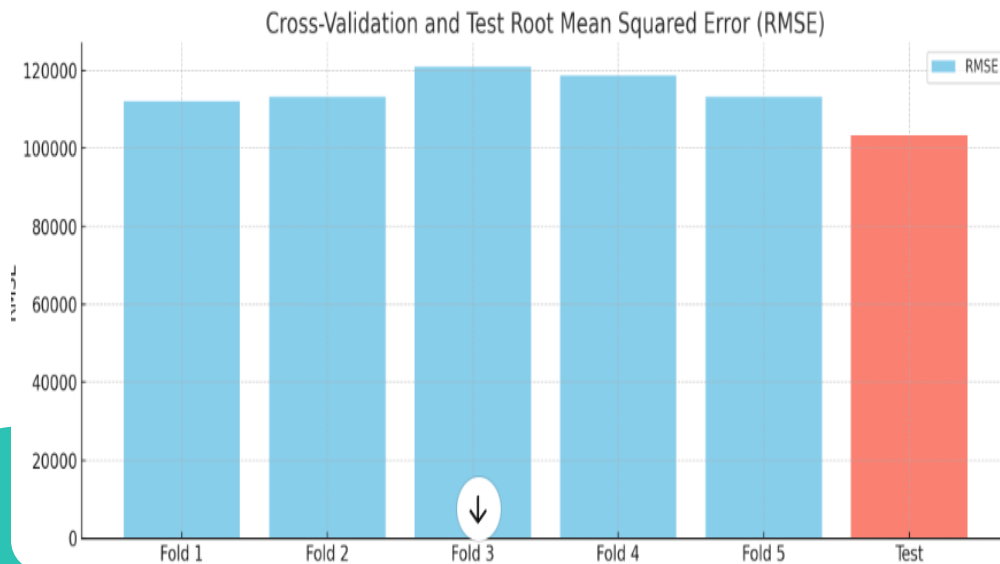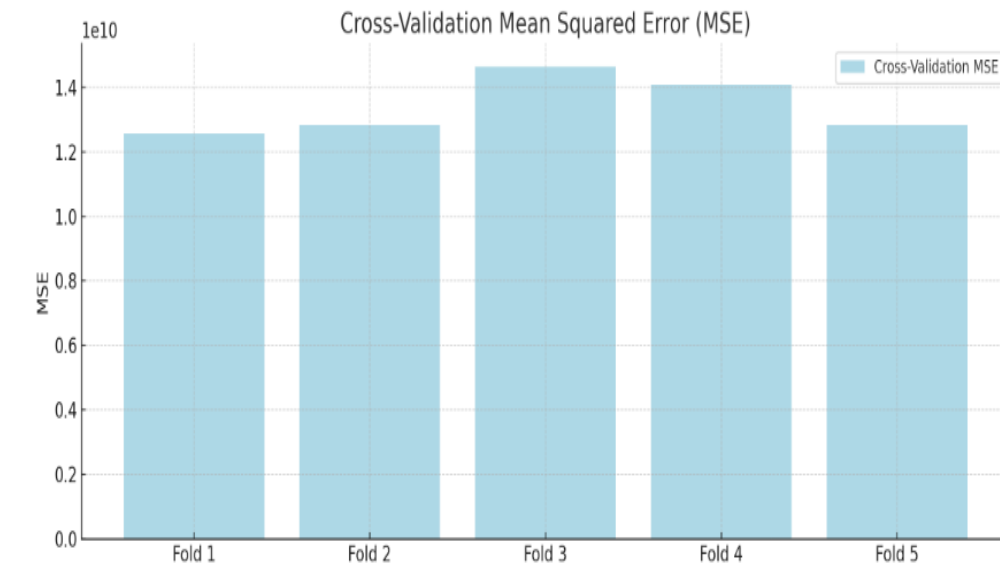
Test Set Evaluation Metrics	
Mean Squared Error	10681370686.478922
Root Mean Squared Error	103350.71691323153
Explained Variance Score	0.8811581155264866
R-Squared Score	0.8801226975069154

Hyperparameter Tuning

Best Hyperparameters	Random Search
'bootstrap'	False
'max_depth'	30
'max_features'	'sqrt'
'min_samples_leaf'	1
'min_samples_split'	5
'n_estimators'	300

RandomForestRegressor with the best hyperparameters

Mean Squared Error	10314268933.238232
Root Mean Squared Error	101559.18930967414
Explained Variance Score	0.8847389345448602
R-Square Score	0.884426900819024

Best model – RandomForestRegressor – saved as .pkl

Accuracy –88.42%

Streamlit Application

Number of Owners
3

Model Year
1996

Seating Capacity
5

Transmission
Manual

Fuel Type
LPG

City
Hyderabad

Car Price Prediction App

Enter the car details below:

Kilometers Driven

4000

Engine Displacement (cc)

1000

Top Speed (km/h)

130

Mileage (km/l)

12

Predict Price

The predicted price of the car is: Rs450,967.50

Thank you