Car Dheko Used Car Price Prediction

Capstone project –3

By

S.Vinodha

Objective

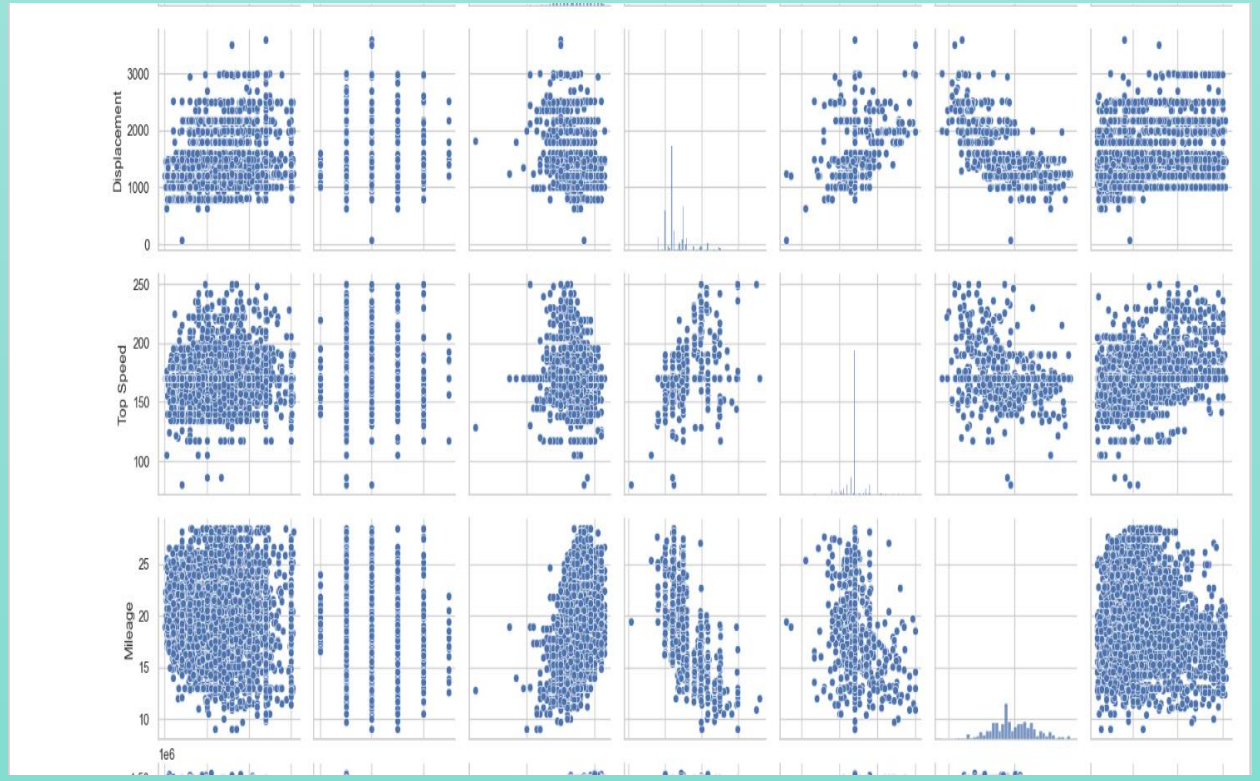
- To create an accurate and user-friendly streamlit tool that predicts the prices of used cars based on various features.
- This tool is an interactive web application for both customers and sales representatives to use seamlessly.
- It allows users to input car details and receive an estimated price instantly.

Data Processing

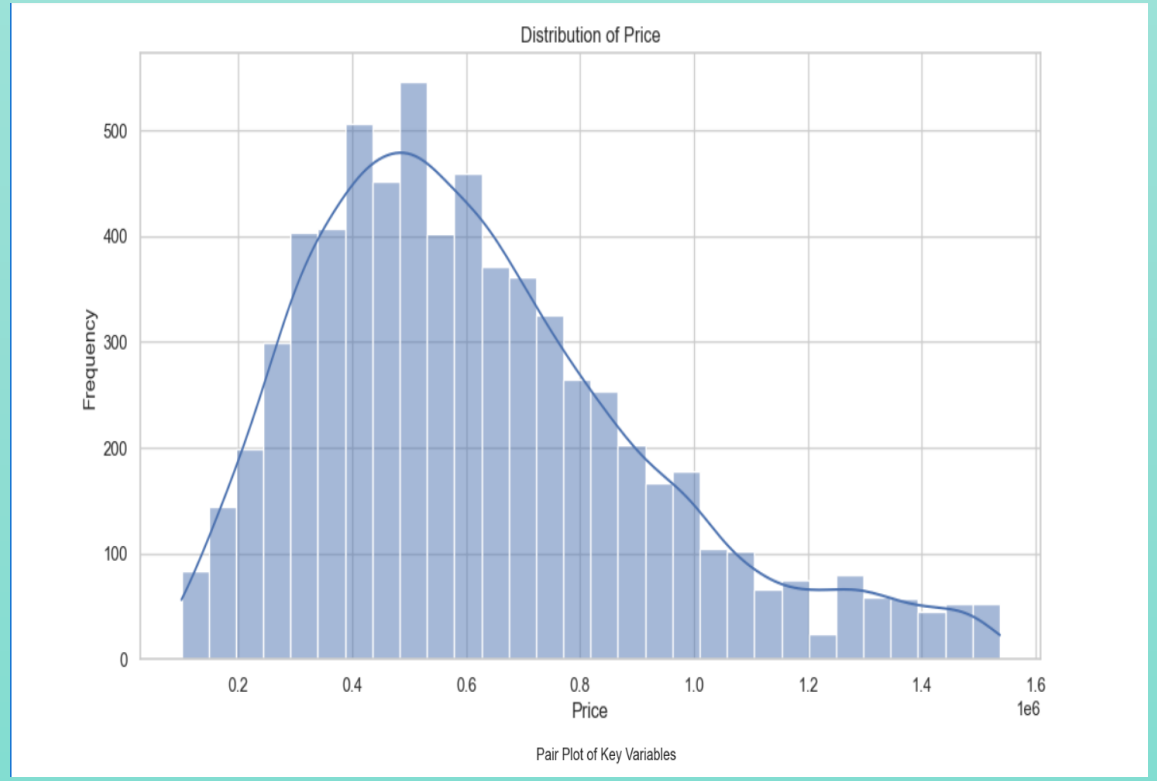
Import and concatenate:	•Imported all unstructured city's dataset and converted it into a structured format. •Assigned all rows in 'City'column with the name of the respective city. •Concatenated all datasets a single dataset.
Handling Missing Values:	• Identified and filled or removed missing values 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

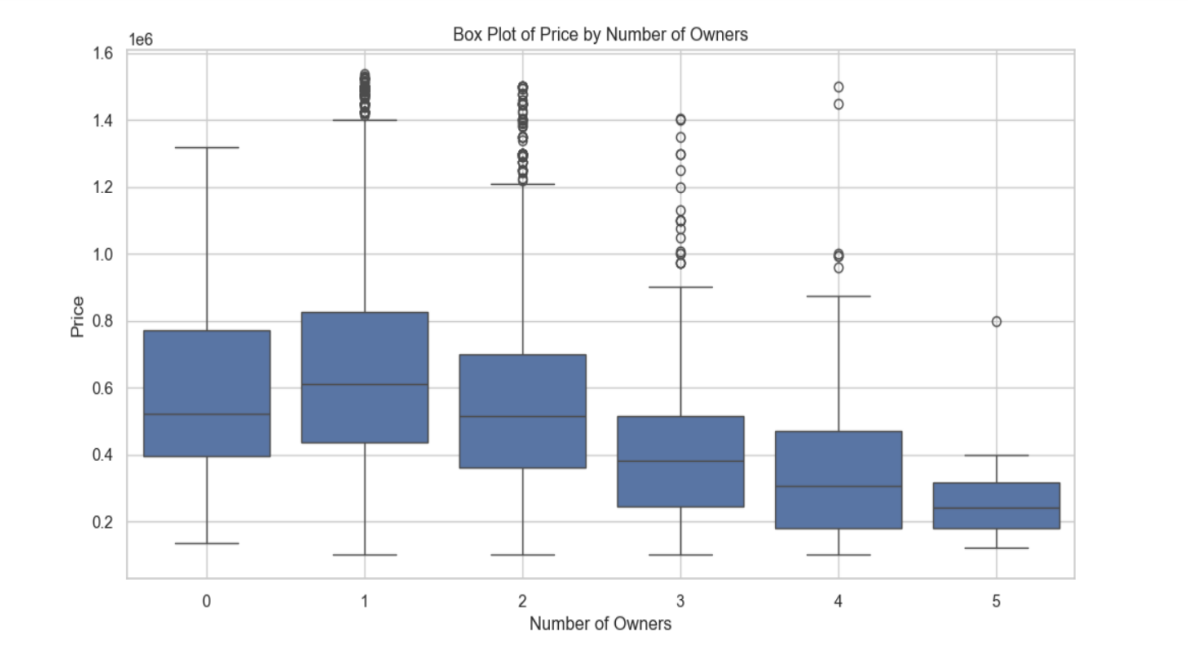


histograms

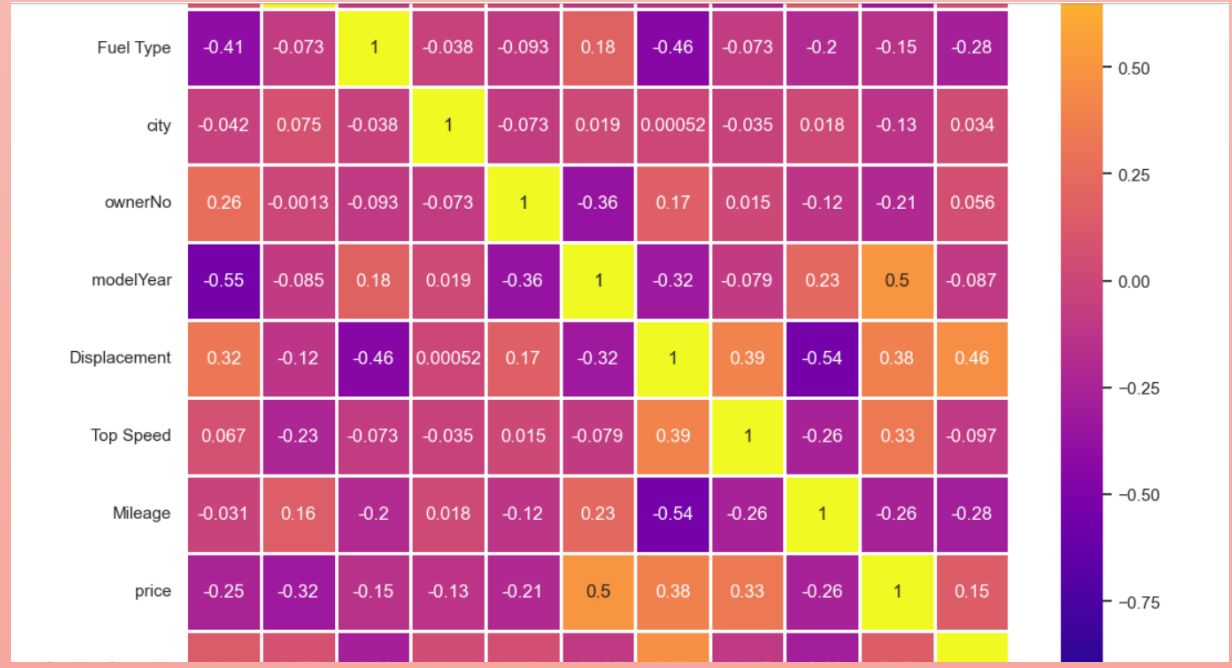


Data Visualization

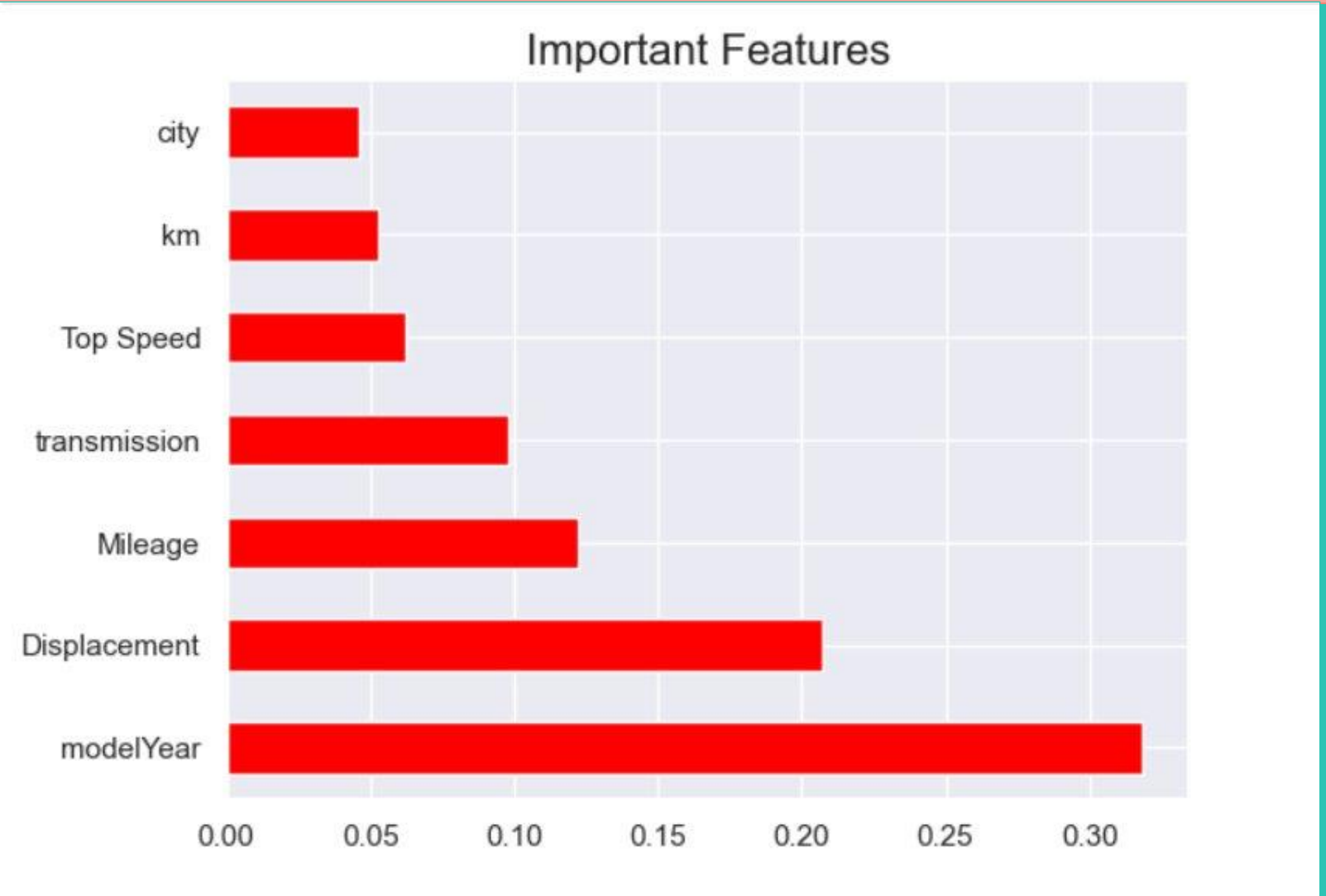
box plots



correlation heatmaps



Feature Selection



Model Development

Models	Mean Squared Error	Root Mean Squared Error	Explained Variance Score	R-Square Score / Accuracy
Linear Regression	3.331581e+10	182526.179664	0.644241	0.643943
Support Vector R regression	9.534279e+10	308776.281499	0.000399	-0.018961
Decision Tree Regressor	2.272008e+10	150731.818026	0.757229	0.757183
Random Forest Regressor	1.297649e+10	113914.400493	0.861461	0.861316
Ridge	3.331562e+10	182525.669239	0.644242	0.643945

Cross-validation techniques -to ensure robust performance.

```
 Cross-Validation Mean Squared Errors:
  [1.25626804e+10 1.28313024e+10 1.46369160e+10 1.40721487e+10 1.28349167e+10]

 Cross-Validation Root Mean Squared Errors:
  [112083.36367925 113275.33905078 120983.1227751  118626.08775661 113291.29142121]

Average RMSE from Cross-Validation: 115651.84093659182


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 from Random Search:

```
{'n_estimators': 500, 'min_samples_split': 2, 'min_samples_leaf': 1,
'max_features': 'sqrt', 'max_depth': 30}
```

Best Cross-Validated RMSE from Random Search:

112623.18859536767

RandomForestRegressor with the best hyperparameters

```
Test RMSE with tuned parameters: 116263.94748006128

Final Model Evaluation Metrics:

Mean Squared Error: 13517305483.646446

Root Mean Squared Error: 116263.94748006128

Explained Variance Score: 0.8555917573357397

R-Squared Score: 0.8555359378024255
```

Best model – RandomForestRegressor – saved as .pkl

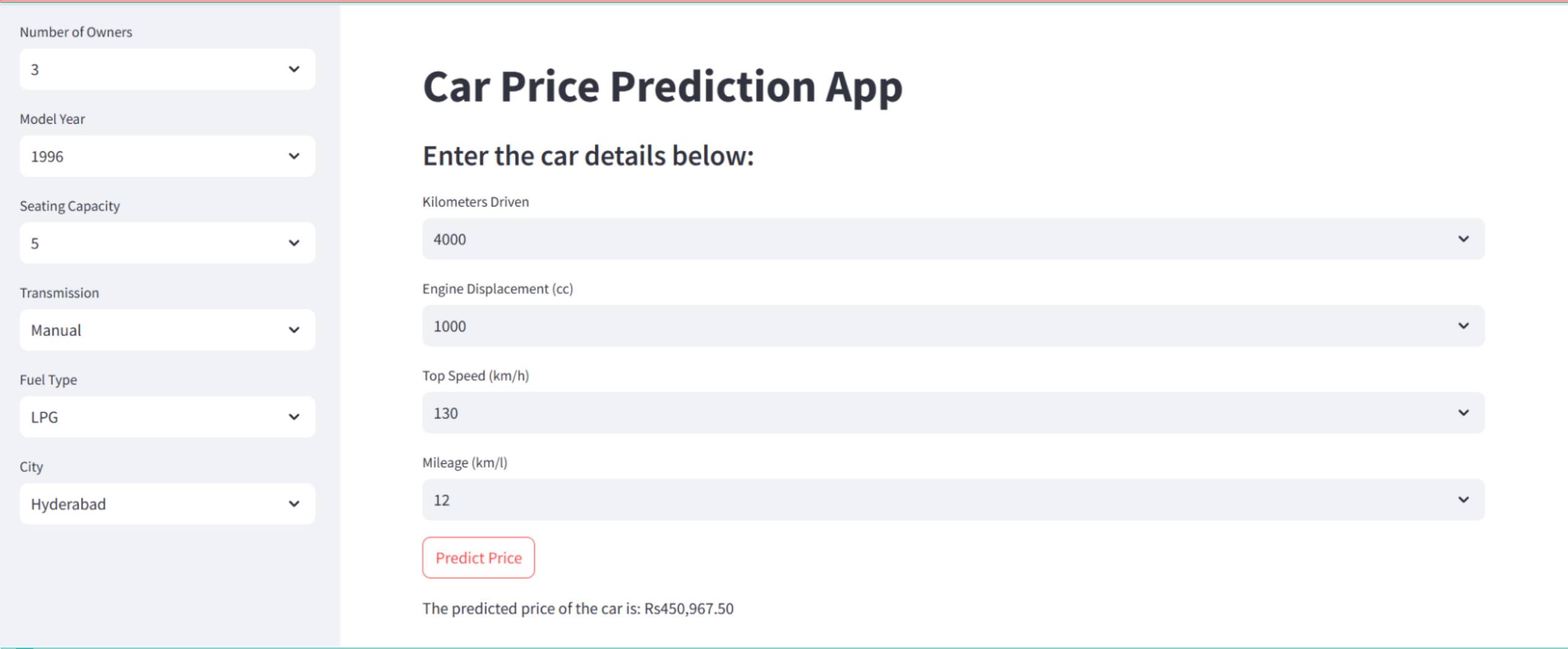
Mean Squared Error: 10681370686.478922

Root Mean Squared Error: 103350.71691323153

Explained Variance Score: 0.8811581155264866

R-Square Score: 0.8801226975069154

Streamlit Application



Thank you