Car Price Prediction

Building a Car Price Prediction Model – A Machine Learning Approach

Objective

The goal is to develop a machine learning model that can accurately estimate the selling price of used cars. You will handle data preprocessing, perform exploratory analysis, train a predictive model, and evaluate its accuracy using statistical metrics.
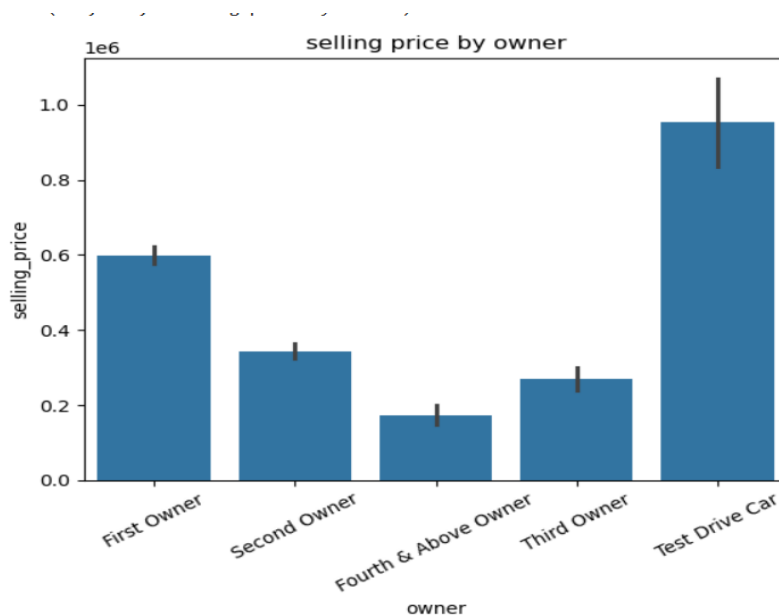
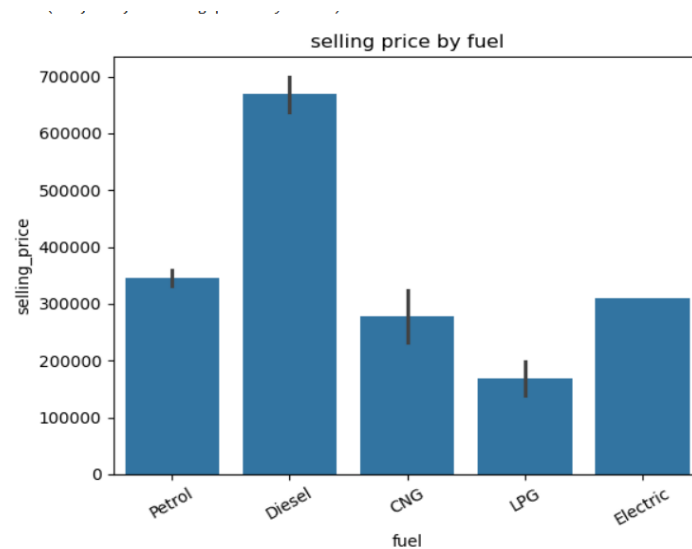Numerical columns :

Year , km_driven

Categorical columns:

Name , fuel , seller_type , transmission , owner

Target column : selling_price

Model Interpretation

The graph above shows how different types of owners influence the selling price of car , as 'owner' is a categorical column we will use encoding.

selling price by fuel

Above graph shows categorical column 'fuel' also influencing the selling price.

this heatmap shows the correlation between all numerical columns signifying that km_driven and year corelate to selling price.

correlation heatmap
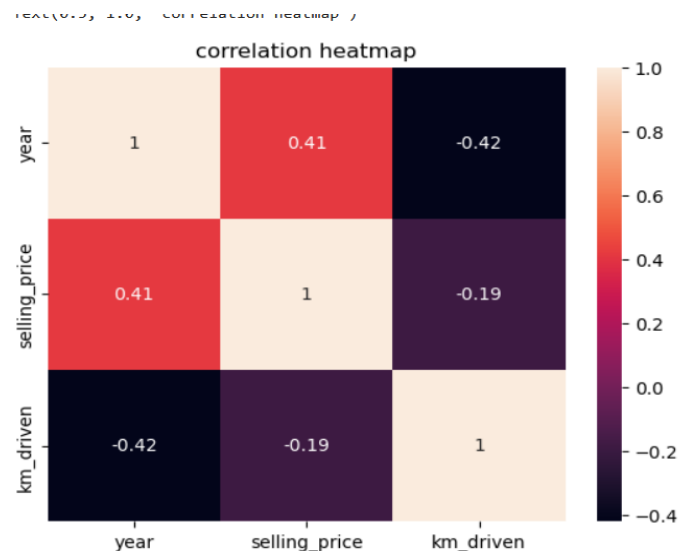
after learning about the data columns and visual insights we preprocess the data=>

```python
from sklearn.preprocessing import LabelEncoder

label = LabelEncoder()
df['owner'] = label.fit_transform(df['owner'])
df['fuel'] = label.fit_transform(df['fuel'])
df['transmission'] = label.fit_transform(df['transmission'])
df['seller_type'] = label.fit_transform(df['seller_type'])
```

Above code labels the categorical columns.
Then feature scaling is done using standardscaler.

Then data is splitted into 80:training data and 20:test data

```
from sklearn.model_selection import train_test_split
```

```
X=new_df.iloc[:,2:]
y=new_df.iloc[:,1]
```

```
X_train,X_test,y_train,y_test = train_test_split(X,y,test_size=0.2,random_state=42)
```

Then model is trained using linear regression

```
from sklearn.linear_model import LinearRegression
lr = LinearRegression()
```

```
lr.fit(X_train,y_train)
```

```
▼   LinearRegression  ⓘ ⓘ

LinearRegression()
```

Values are predicted on test data

```
y_pred = lr.predict(X_test)
```

```
y_pred[0]
```

```
590689.1356533766
```

Conclusion=>

In this project we developed a car price prediction model using features which were categorical and numerical. Data was first preprocessed through encoding and feature scaling. Then using linear regression we trained the model on training data after splitting it in 80:20 ratio of training data and testing data. Then it got evaluated using mean square error and R2 score with fair accuracy.