mayanksharma.analytics@gmail.com
+91 9560812017

CAR PRICE PREDICTION PROJECT

Know Your Car's Worth

mayanksharma-ds.github.io

MAYANK SHARMA

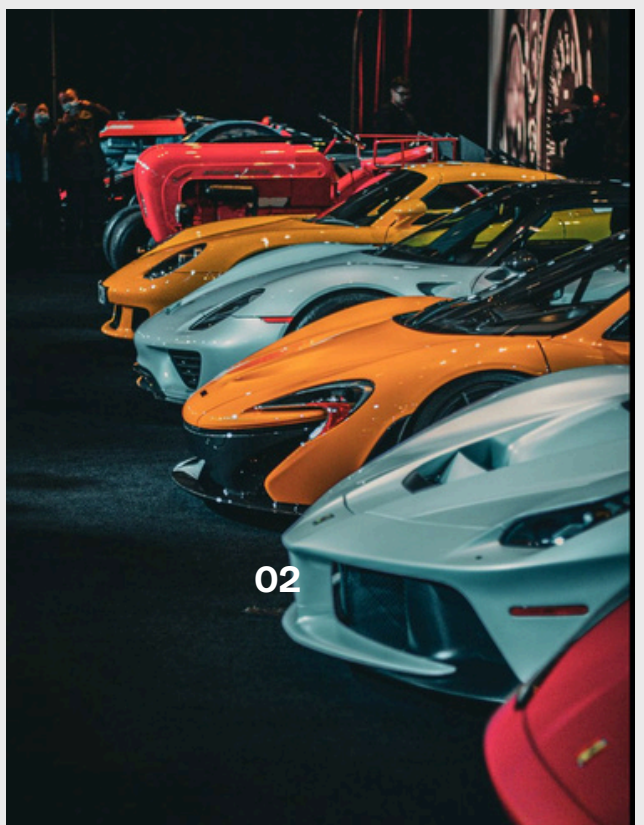
MAYANK SHARMA
ASPIRING DATA SCIENTIST

INTRODUCTION

The Car Price Prediction project is a machine learning–based solution designed to estimate the market value of used cars based on their key features. With the rapid growth of online car marketplaces, accurately predicting a car's price has become essential for both buyers and sellers to make informed decisions.

In this project, historical car data is analyzed and preprocessed to identify important factors such as brand, model, year of manufacture, fuel type, transmission, mileage, and engine specifications that influence car prices. Various machine learning techniques are applied to build a predictive model that learns patterns from the data and provides reliable price estimations.

To make the model accessible and user-friendly, it is deployed as an interactive Streamlit web application, allowing users to input car details and instantly receive predicted prices. This project demonstrates the complete end-to-end machine learning workflow—from data preprocessing and model training to deployment—making it a practical and real-world application of data science concepts.



OBJECTIVE

The objective of this project is to develop a machine learning model that can accurately predict the price of used cars based on their key attributes. The project focuses on understanding price-influencing factors and delivering a user-friendly solution through a web application.

Key Objectives:

- Identify important features that affect car pricing.
- Build and evaluate a reliable car price prediction model.
- Deploy the model using Streamlit for real-time predictions.

MAYANK SHARMA
ASPIRING DATA SCIENTIST

MAYANK SHARMA

ASPIRING DATA SCIENTIST

IMPORT REQUIRED LIBRARIES

•[1]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.preprocessing import OneHotEncoder
from sklearn.compose import make_column_transformer
from sklearn.pipeline import make_pipeline
from sklearn.metrics import r2_score
```

MAYANK SHARMA
ASPIRING DATA SCIENTIST

DATASET IMPORT

[2]:

```
car=pd.read_csv('Cleaned_Car_data.csv')
```

[3]:

```
car
```

[3]:

	Unnamed: 0	name	company	year	Price	kms_driven	fuel_type
0	0	Hyundai Santro Xing	Hyundai	2007	80000	45000	Petrol
1	1	Mahindra Jeep CL550	Mahindra	2006	425000	40	Diesel
2	2	Hyundai Grand i10	Hyundai	2014	325000	28000	Petrol
3	3	Ford EcoSport Titanium	Ford	2014	575000	36000	Diesel
4	4	Ford Figo	Ford	2012	175000	41000	Diesel

MAYANK SHARMA
ASPIRING DATA SCIENTIST

CHECKING RELATIONSHIP OF COLUMNS

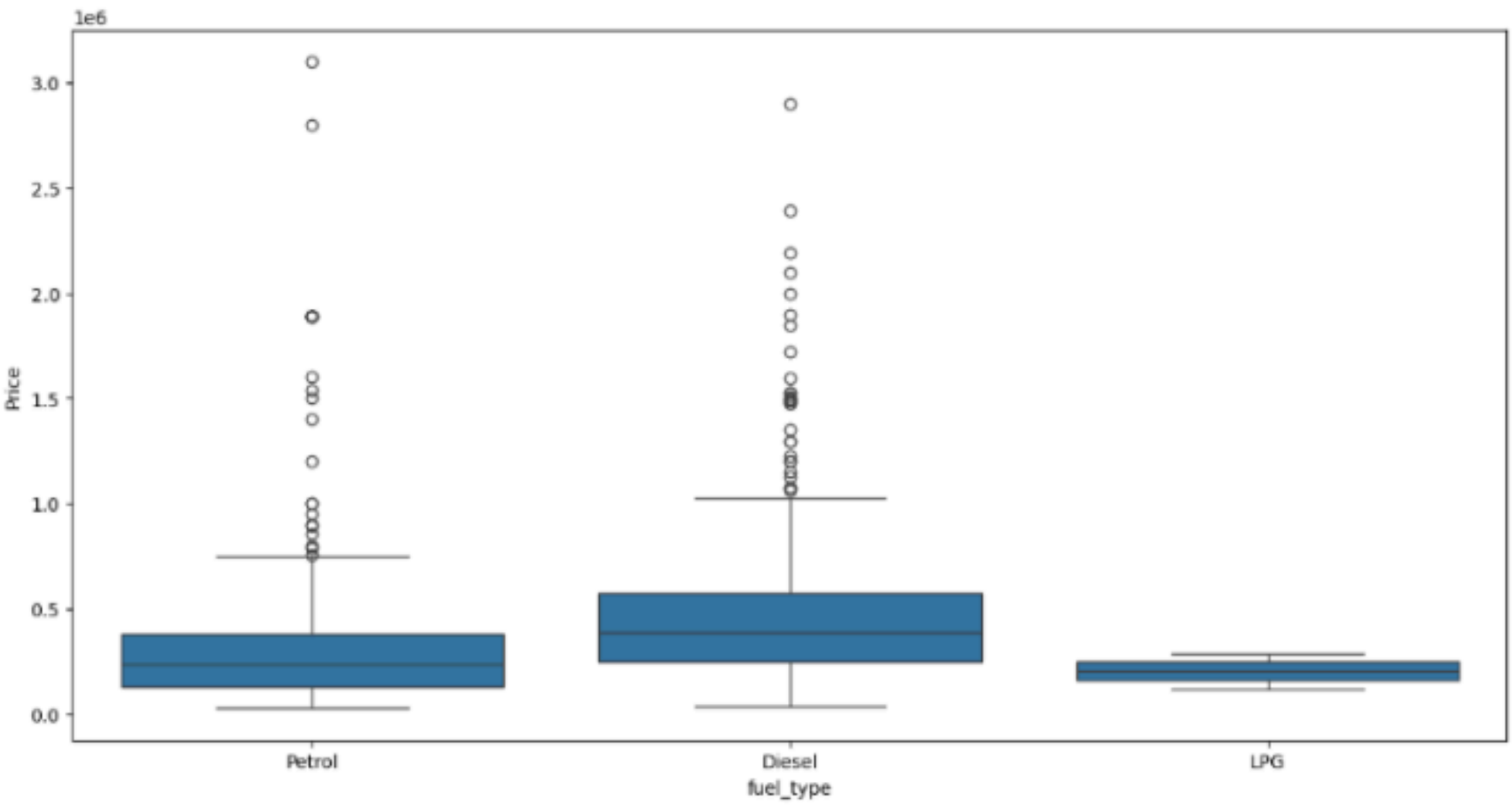
Checking relationship of Fuel Type with Price

[13]:

```
plt.subplots(figsize=(14,7))
sns.boxplot(x='fuel_type',y='Price',data=car)
```

[13]:

```
<Axes: xlabel='fuel_type', ylabel='Price'>
```



MAYANK SHARMA
ASPIRING DATA SCIENTIST

CHECKING RELATIONSHIP OF COLUMNS

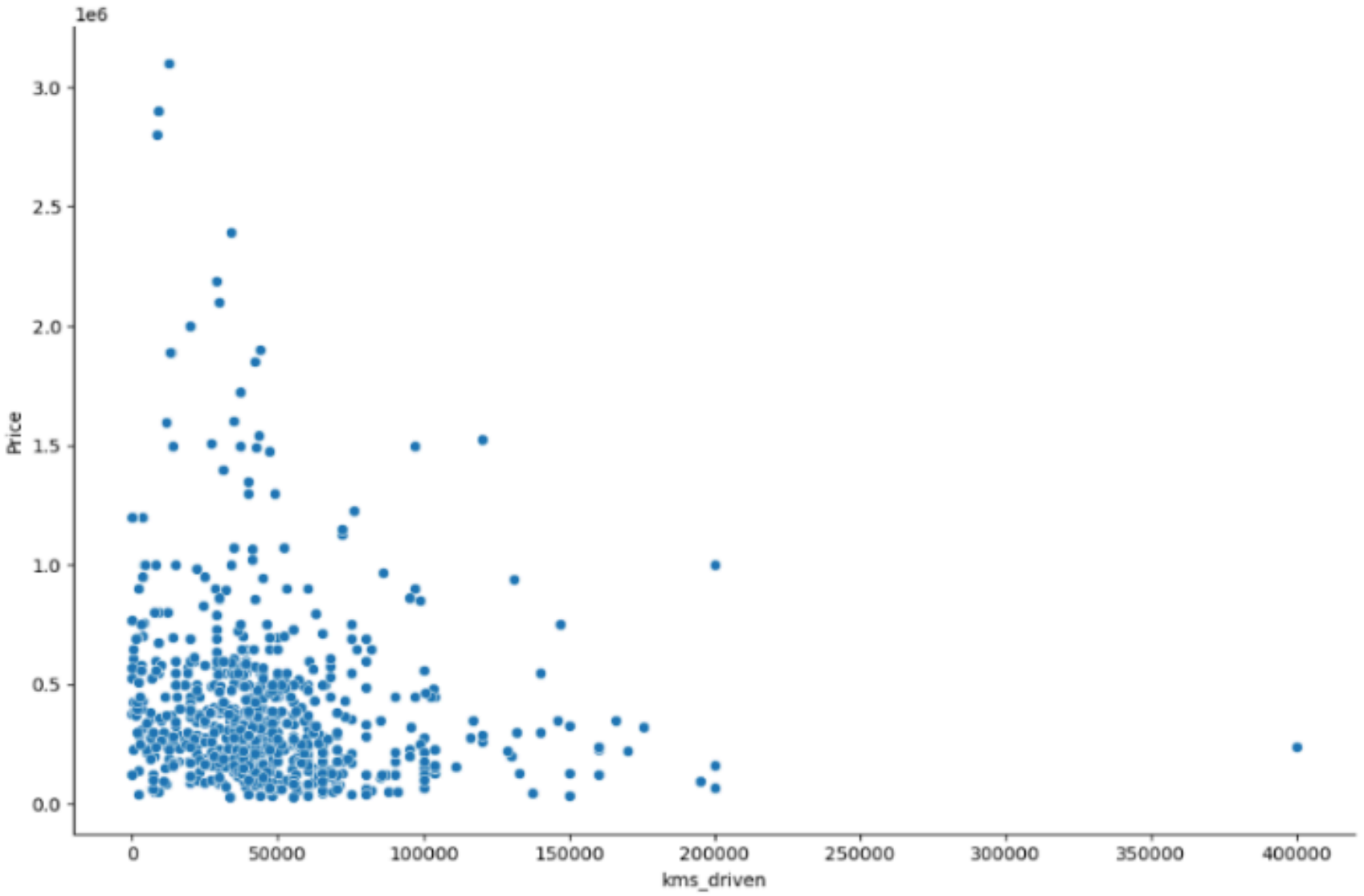
Checking relationship of kms_driven with Price

[12]:

```
sns.relplot(x='kms_driven',y='Price',data=car,height=7,aspect=1.5)
```

[12]:

```
<seaborn.axisgrid.FacetGrid at 0x1dec206dd30>
```



MAYANK SHARMA
ASPIRING DATA SCIENTIST

FEATURES & LABELS

Extracting Training Data

[15]:

```
X=car[['name','company','year','kms_driven','fuel_type']]
y=car['Price']
```

TRAIN TEST SPLIT

Applying Train Test Split

•[18]:

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

MAYANK SHARMA
ASPIRING DATA SCIENTIST

CATEGORICAL FEATURE ENCODING

Creating an OneHotEncoder object to contain all the possible categories

[21]:

```
ohe=OneHotEncoder()
ohe.fit(X[['name','company','fuel_type']])
```

[21]:

OneHotEncoder

▶ Parameters

MAYANK SHARMA
ASPIRING DATA SCIENTIST

COLUMN TRANSFORMER FOR FEATURE PROCESSING

Creating a column transformer to transform categorical columns

•[22]:

```
trans=make_column_transformer((OneHotEncoder(categories=ohe.categories_),
                               ['name','company','fuel_type']),
                              remainder='passthrough')
```

MAYANK SHARMA
ASPIRING DATA SCIENTIST

MAYANK SHARMA
ASPIRING DATA SCIENTIST

MODEL

Linear Regression Model

[23]:

```
lr=LinearRegression()
```

MAYANK SHARMA
ASPIRING DATA SCIENTIST

MAKING A PIPELINE

Making a pipeline

[24]:

```
pipe=make_pipeline(column_trans,lr)
```

TRAINING THE MODEL

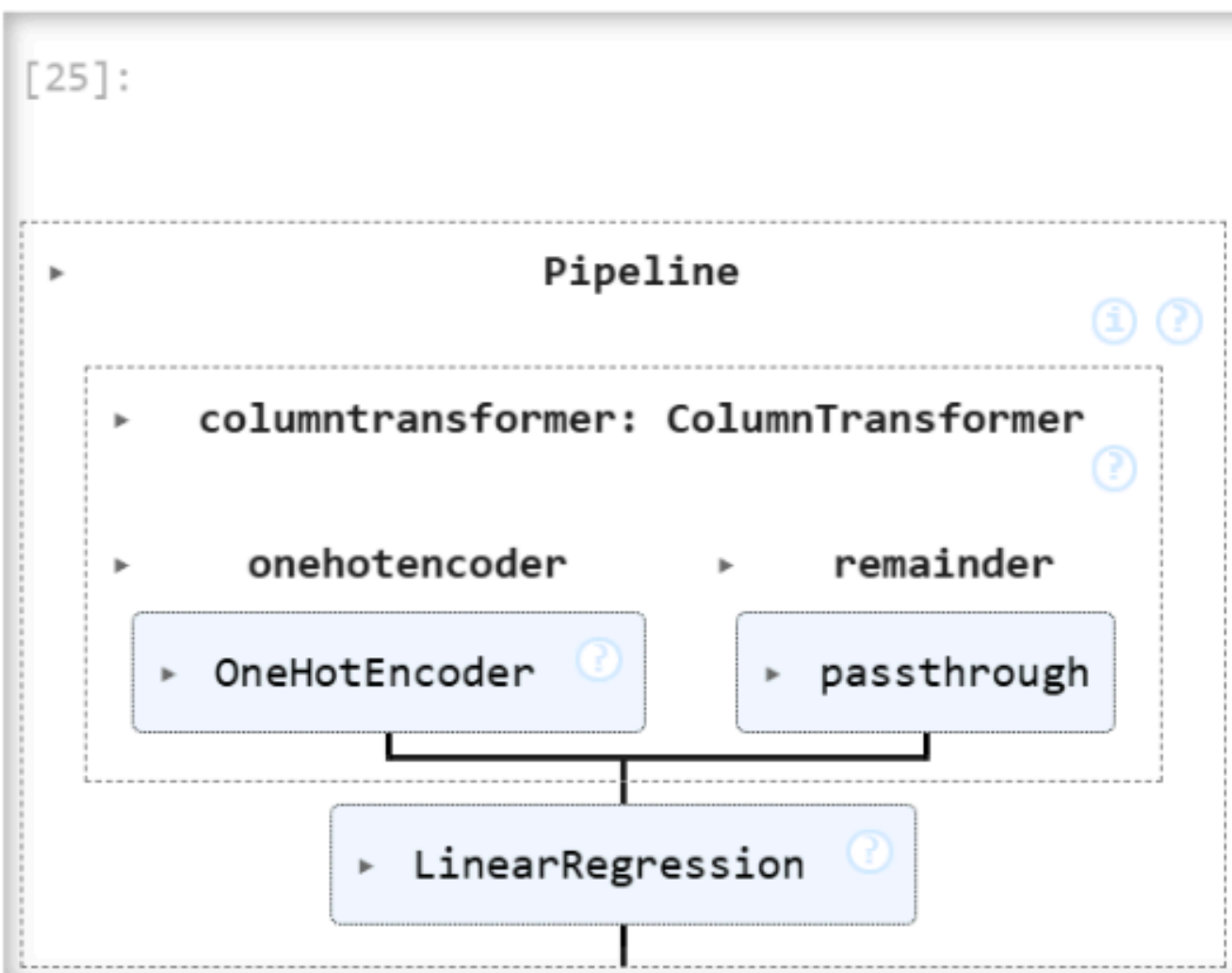
MAYANK SHARMA
ASPIRING DATA SCIENTIST

Fitting the model

[25]:

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

[25]:



MAYANK SHARMA
ASPIRING DATA SCIENTIST

MODEL EVALUATION

[26]:

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

Checking R2 Score

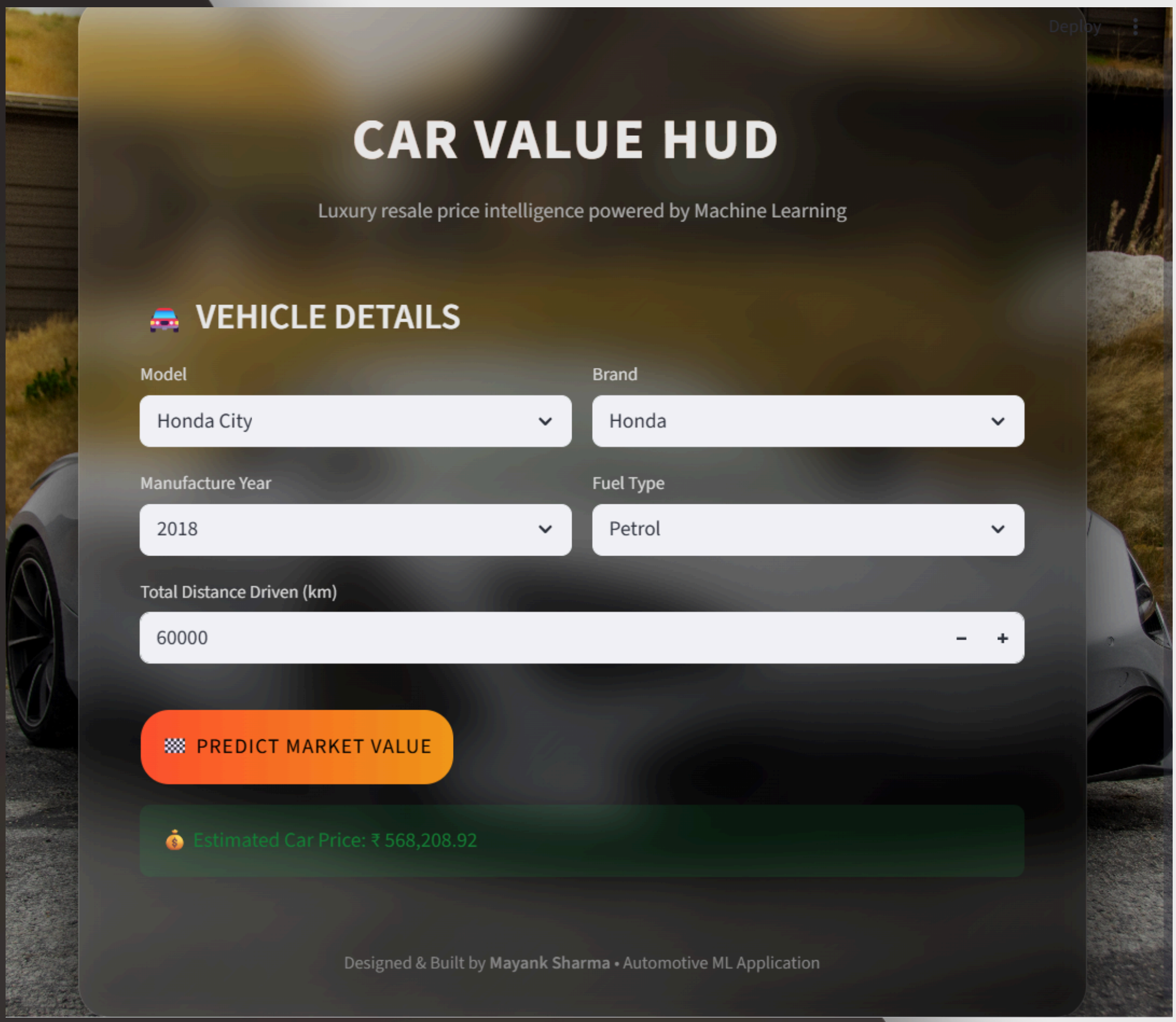
[27]:

```
r2_score(y_test,y_pred)
```

[27]:

```
0.7004095973633013
```

Streamlit UI



CONTACT

+91 9560812017

https://github.com/mayanksharma-ds

mayanksharma-ds.github.io

mayanksharma.analytics@gmail.com