In [1]:

```
import numpy as np
import pandas as pd
import seaborn as sb
import matplotlib.pyplot as plt
from sklearn import preprocessing ,svm
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.linear_model import Ridge,Lasso
from sklearn.preprocessing import StandardScaler
```

In [2]:

```
dv=pd.read_csv(r"C:\Users\poorn\Downloads\fiat500_VehicleSelection_Dataset.csv")
dv.head(10)
```

Out[2]:

	ID	model	engine_power	age_in_days	km	previous_owners	lat	lon
0	1	lounge	51	882	25000	1	44.907242	8.611560
1	2	pop	51	1186	32500	1	45.666359	12.241890
2	3	sport	74	4658	142228	1	45.503300	11.417840
3	4	lounge	51	2739	160000	1	40.633171	17.634609
4	5	pop	73	3074	106880	1	41.903221	12.495650
5	6	pop	74	3623	70225	1	45.000702	7.682270
6	7	lounge	51	731	11600	1	44.907242	8.611560
7	8	lounge	51	1521	49076	1	41.903221	12.495650
8	9	sport	73	4049	76000	1	45.548000	11.549470
9	10	sport	51	3653	89000	1	45.438301	10.991700

In [3]:

```
dv.info()
```

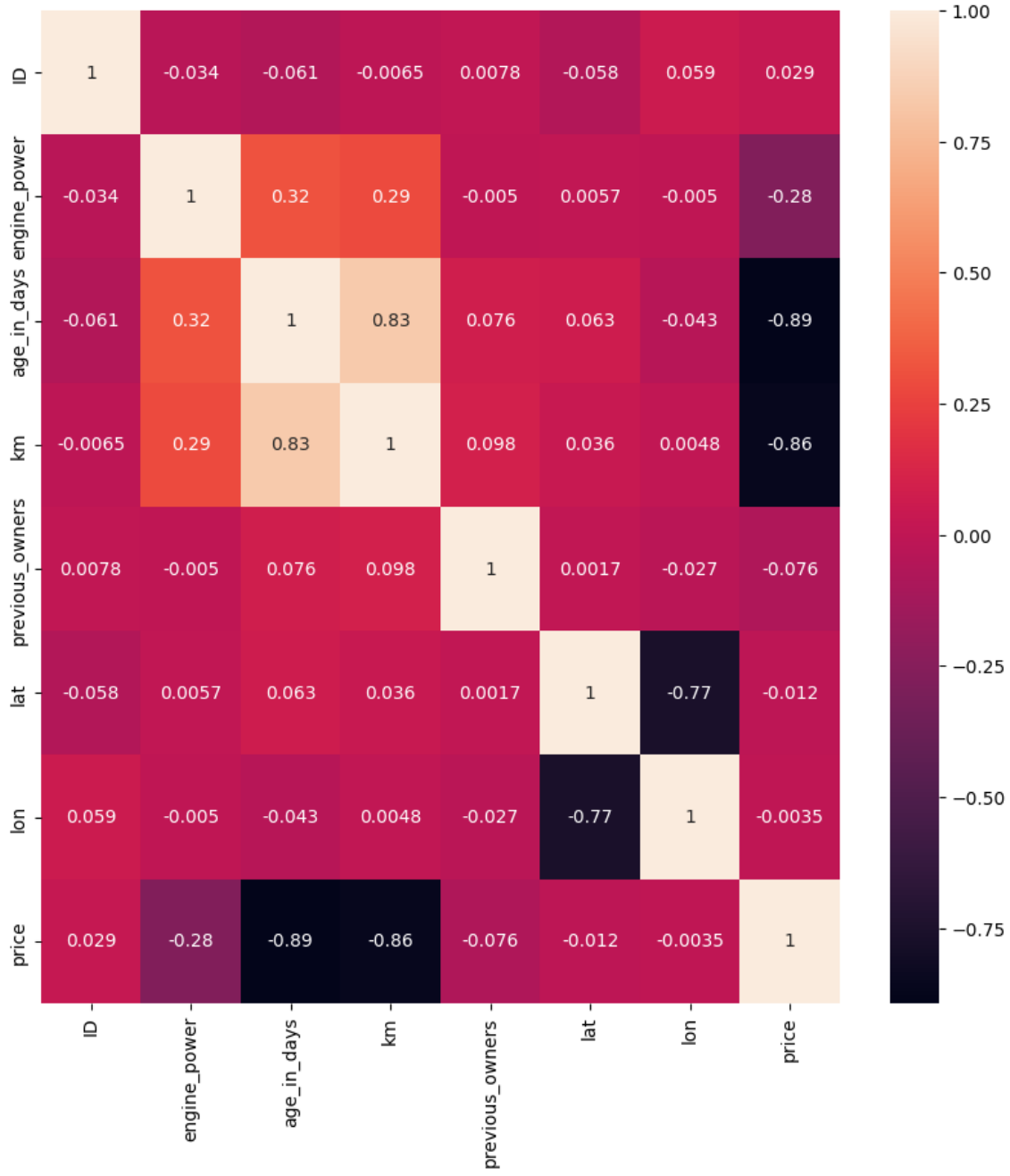
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1538 entries, 0 to 1537
Data columns (total 9 columns):
 #   Column           Non-Null Count  Dtype  
---  ------           --------------  -----  
 0   ID               1538 non-null   int64  
 1   model            1538 non-null   object 
 2   engine_power     1538 non-null   int64  
 3   age_in_days      1538 non-null   int64  
 4   km               1538 non-null   int64  
 5   previous_owners  1538 non-null   int64  
 6   lat              1538 non-null   float64
 7   lon              1538 non-null   float64
 8   price            1538 non-null   int64  
dtypes: float64(2), int64(6), object(1)
memory usage: 108.3+ KB
```

In [4]:

```
dat=dv
dat.drop(columns=["model"], inplace = True)
plt.figure(figsize = (10, 10))
sb.heatmap(dv.corr(), annot = True)
```

Out[4]:

<Axes: >



In [2]:

```
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call las
t)
Cell In[2], line 1
----> 1 data.drop(columns =["Radio","Newspaper"], inplace = True)
      2 #pairplot
      3 sns.pairplot(data)

NameError: name 'data' is not defined
```

In [7]:

```
dv.columns
```

Out[7]:

```
Index(['ID', 'engine_power', 'age_in_days', 'km', 'previous_owners', 'la
t',
       'lon', 'price'],
      dtype='object')
```

In [1]:

```
features = dv.columns[3:5]
target = dv.columns[-1]
#X and y values
x = dv[features].values
y = dv[target].values
#splot
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.3, random_size=100
print("The dimension of X_train is {}".format(x_train.shape))
print("The dimension of X_test is {}".format(x_test.shape))
#Scale features
scaler = StandardScaler()
x_train = scaler.fit_transform(x_train)
x_test = scaler.transform(x_test)
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call las
t)
Cell In[1], line 1
----> 1 features = dv.columns[3:5]
      2 target = dv.columns[-1]
      3 #X and y values

NameError: name 'dv' is not defined
```

In [12]:

```
#Model
lr = LinearRegression()
#Fit model
lr.fit(x_train, y_train)
#actual
actual = y_test
train_score_lr = lr.score(x_train, y_train)
test_score_lr = lr.score(x_test, y_test)
print("\nLinear Regression Model:\n")
print("The train score for lr model is {}".format(train_score_lr))
print("The test score for lr model is {}".format(test_score_lr))
```

```
---------------------------------------------------------------------------
-
NameError                                 Traceback (most recent call las
t)
Cell In[12], line 4
      2 lr = LinearRegression()
      3 #Fit model
----> 4 lr.fit(x_train, y_train)
      5 #actual
      6 actual = y_test

NameError: name 'x_train' is not defined
```

In [13]:

```
#Ridge Regression Model
ridgeReg = Ridge(alpha=10)
ridgeReg.fit(x_train,y_train)
#train and test scorefor ridge regression
train_score_ridge = ridgeReg.score(x_train, y_train)
test_score_ridge = ridgeReg.score(x_test, y_test)
print("\nRidge Model:\n")
print("The train score for ridge model is {}".format(train_score_ridge))
print("The test score for ridge model is {}".format(test_score_ridge))
```

```
---------------------------------------------------------------------------
-
NameError                                 Traceback (most recent call las
t)
Cell In[13], line 3
      1 #Ridge Regression Model
      2 ridgeReg = Ridge(alpha=10)
----> 3 ridgeReg.fit(x_train,y_train)
      4 #train and test scorefor ridge regression
      5 train_score_ridge = ridgeReg.score(x_train, y_train)

NameError: name 'x_train' is not defined
```

In [ ]:

```

```