

Fiat500_vehicle selection using LinearRegression

In [2]:

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

In [3]:

```
dt=pd.read_csv(r"C:\Users\prajapath Arjun\Downloads\fiat500_VehicleSelection_Dataset.csv")
dt
```

Out[3]:

	ID	model	engine_power	age_in_days	km	previous_owners	lat	
0	1	lounge	51	882	25000	1	44.907242	8.611
1	2	pop	51	1186	32500	1	45.666359	12.241
2	3	sport	74	4658	142228	1	45.503300	11.417
3	4	lounge	51	2739	160000	1	40.633171	17.634
4	5	pop	73	3074	106880	1	41.903221	12.495
...
1533	1534	sport	51	3712	115280	1	45.069679	7.704
1534	1535	lounge	74	3835	112000	1	45.845692	8.666
1535	1536	pop	51	2223	60457	1	45.481541	9.413
1536	1537	lounge	51	2557	80750	1	45.000702	7.682
1537	1538	pop	51	1766	54276	1	40.323410	17.568

1538 rows × 9 columns



In [4]:

```
dt=dt[['engine_power', 'price']]
dt.columns=['Engine', 'price']
```

In [5]:

```
dt.head(10)
```

Out[5]:

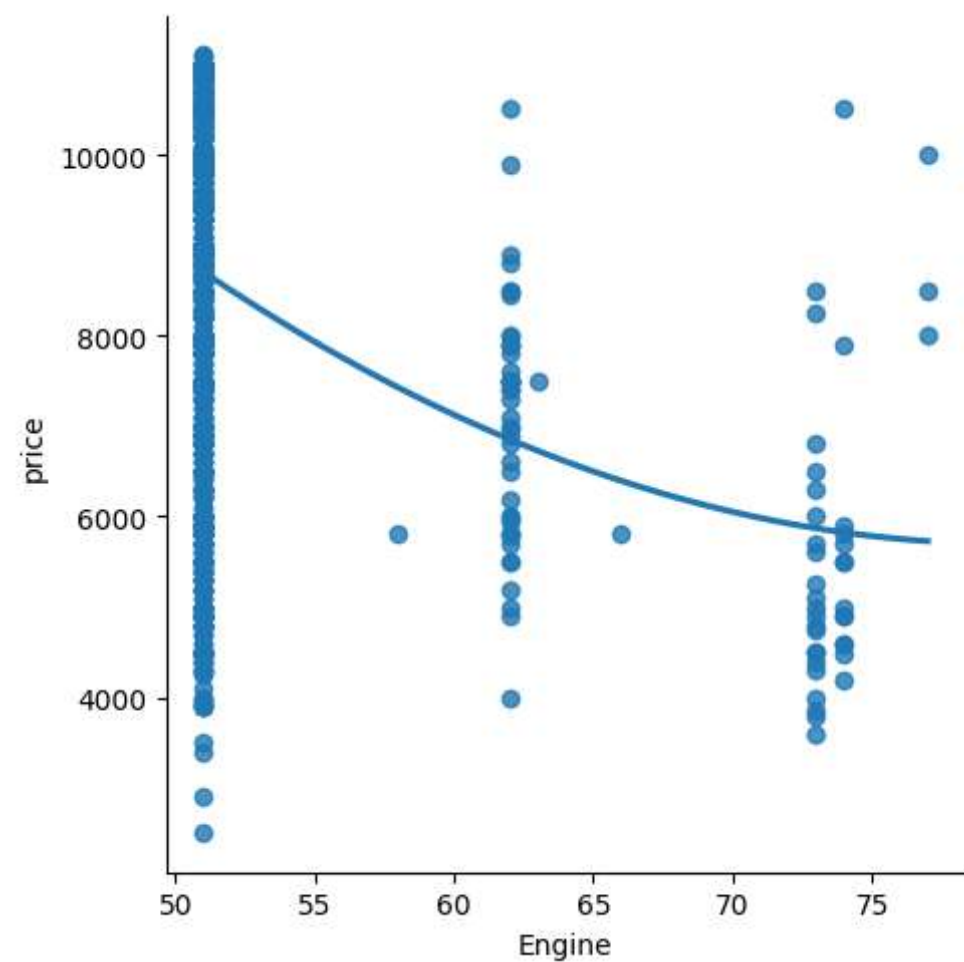
	Engine	price
0	51	8900
1	51	8800
2	74	4200
3	51	6000
4	73	5700
5	74	7900
6	51	10750
7	51	9190
8	73	5600
9	51	6000

In [6]:

```
sns.lmplot(x='Engine',y='price',data=dt,order=2,ci=None)
```

Out[6]:

<seaborn.axisgrid.FacetGrid at 0x220a753d150>



In [7]:

```
dt.describe()
```

Out[7]:

	Engine	price
count	1538.000000	1538.000000
mean	51.904421	8576.003901
std	3.988023	1939.958641
min	51.000000	2500.000000
25%	51.000000	7122.500000
50%	51.000000	9000.000000
75%	51.000000	10000.000000
max	77.000000	11100.000000

In [8]:

```
dt.fillna(method='ffill',inplace=True)
```

C:\Users\prajapath Arjun\AppData\Local\Temp\ipykernel_26964\3059529803.py:

1: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

```
dt.fillna(method='ffill',inplace=True)
```

In [9]:

```
x=np.array(dt['Engine']).reshape(-1,1)
y=np.array(dt['price']).reshape(-1,1)
```

In [10]:

```
dt.dropna(inplace=True)
```

C:\Users\prajapath Arjun\AppData\Local\Temp\ipykernel_26964\735218168.py:

1: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

```
dt.dropna(inplace=True)
```

In [11]:

```
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.25)
```

In [12]:

```
regr=LinearRegression()  
regr.fit(x_train,y_train)  
print(regr.score(x_test,y_test))
```

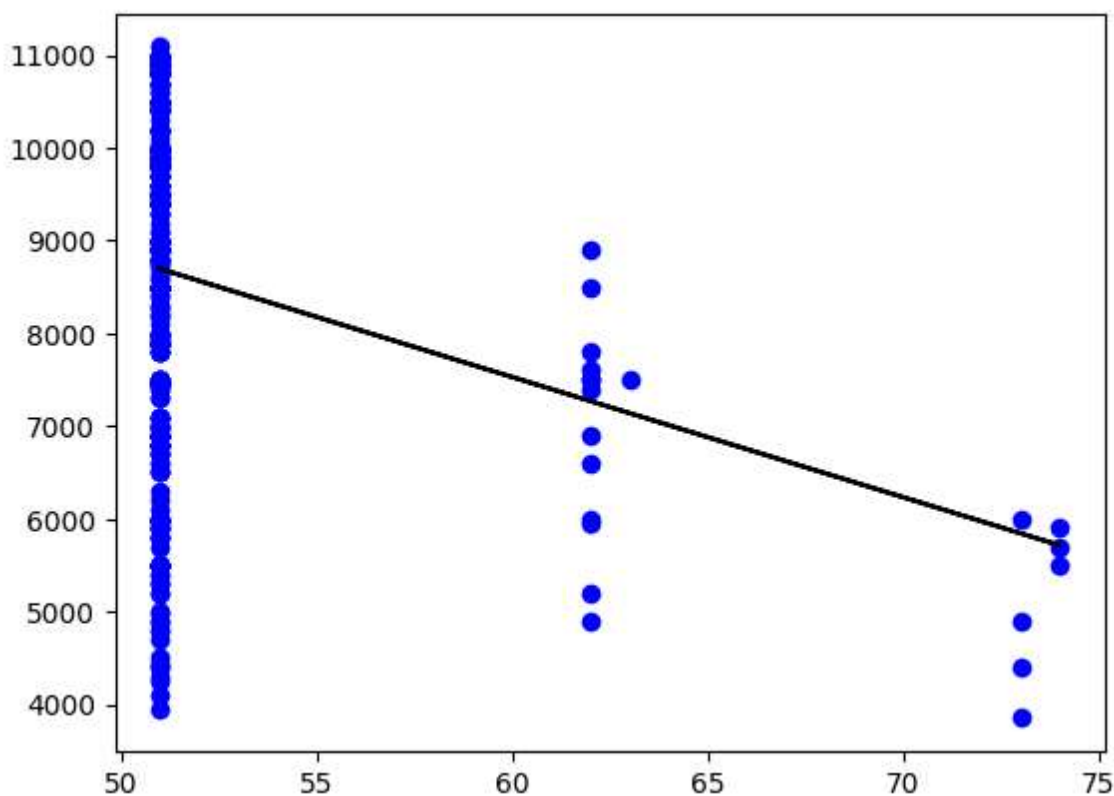
0.09111211449372236

In [13]:

```
y_pred=regr.predict(x_test)
```

In [14]:

```
plt.scatter(x_test,y_test,color='b')  
plt.plot(x_test,y_pred,color='k')  
plt.show()
```



In [15]:

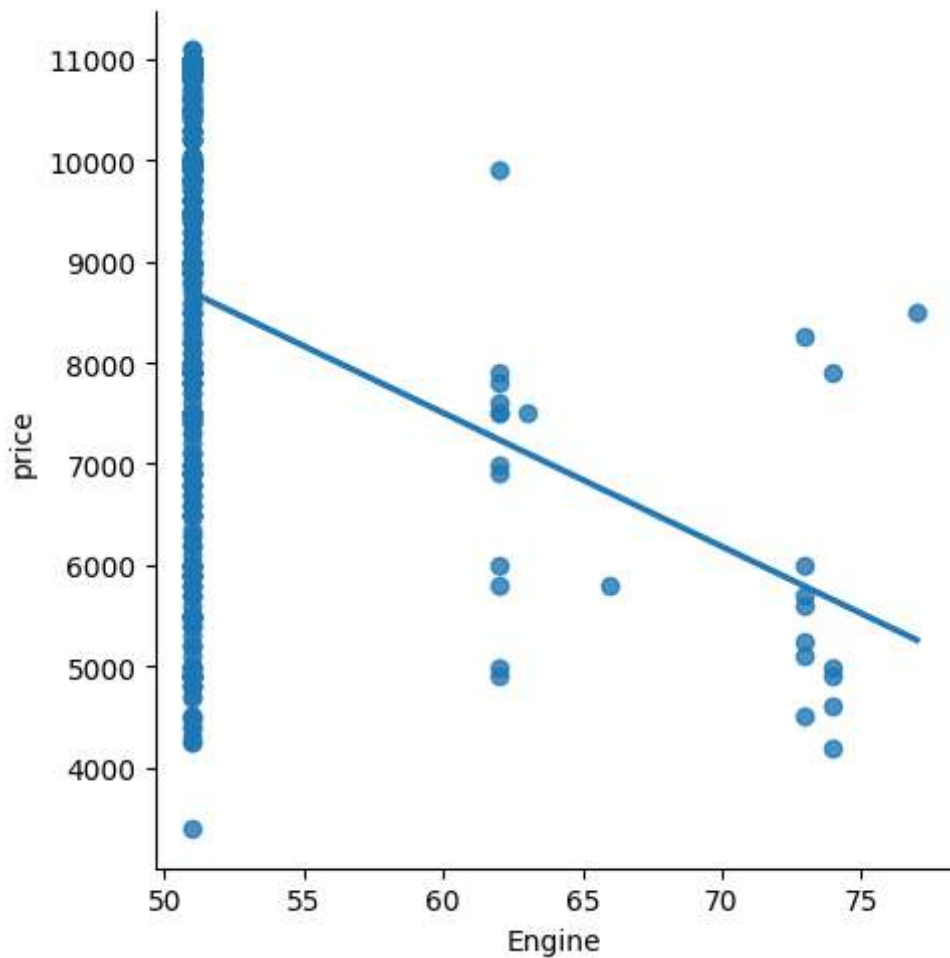
```
dt500=dt[:,[:500]]
```

In [16]:

```
sns.lmplot(x="Engine",y="price",data=dt500,order=1,ci=None)
```

Out[16]:

<seaborn.axisgrid.FacetGrid at 0x220bf6e6e10>



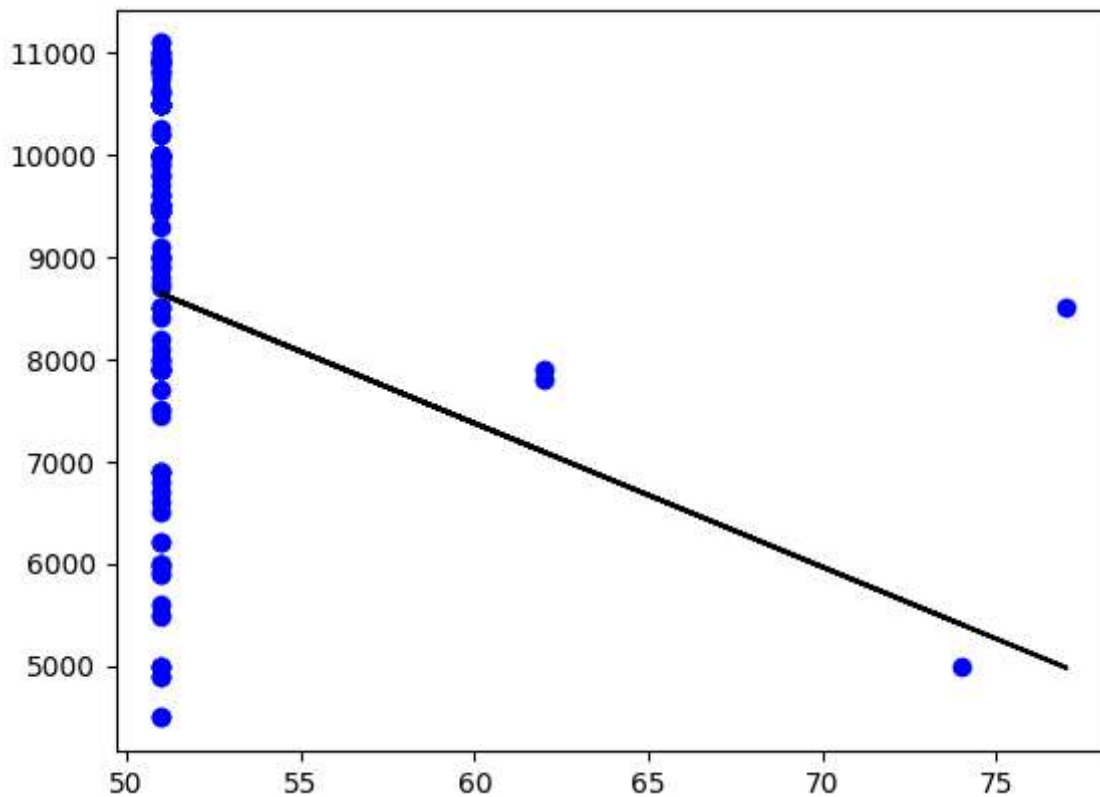
In [17]:

```
dt500.fillna(method='ffill',inplace=True)  
x=np.array(dt500['Engine']).reshape(-1,1)  
y=np.array(dt500['price']).reshape(-1,1)  
dt500.dropna(inplace=True)
```

In [18]:

```
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.25)
regr.fit(x_train,y_train)
print("Regression:",regr.score(x_test,y_test))
y_pred=regr.predict(x_test)
plt.scatter(x_test,y_test,color='b')
plt.plot(x_test,y_pred,color='k')
plt.show()
```

Regression: -0.006392486607532222



In [19]:

```
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score
model=LinearRegression()
model.fit(x_train,y_train)
y_pred=model.predict(x_test)
r2=r2_score(y_test,y_pred)
print("R2 score:",r2)
```

R2 score: -0.006392486607532222