In [1]:

```
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]:

```
#Step-2:Reading the dataset
dt=pd.read_csv(r"C:\Users\91955\Downloads\fiat500_VehicleSelection_Dataset (1).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 [5]:

```
dt=dt[['engine_power','age_in_days']]
#Taking only the selected two attributes from the dataset
dt.columns=['Eng','Age']
#Renaming the columns for easier writing of the code
```

In [7]:

```
dt.head(10)
#Displaying only the 1st 10 rows
```

Out[7]:

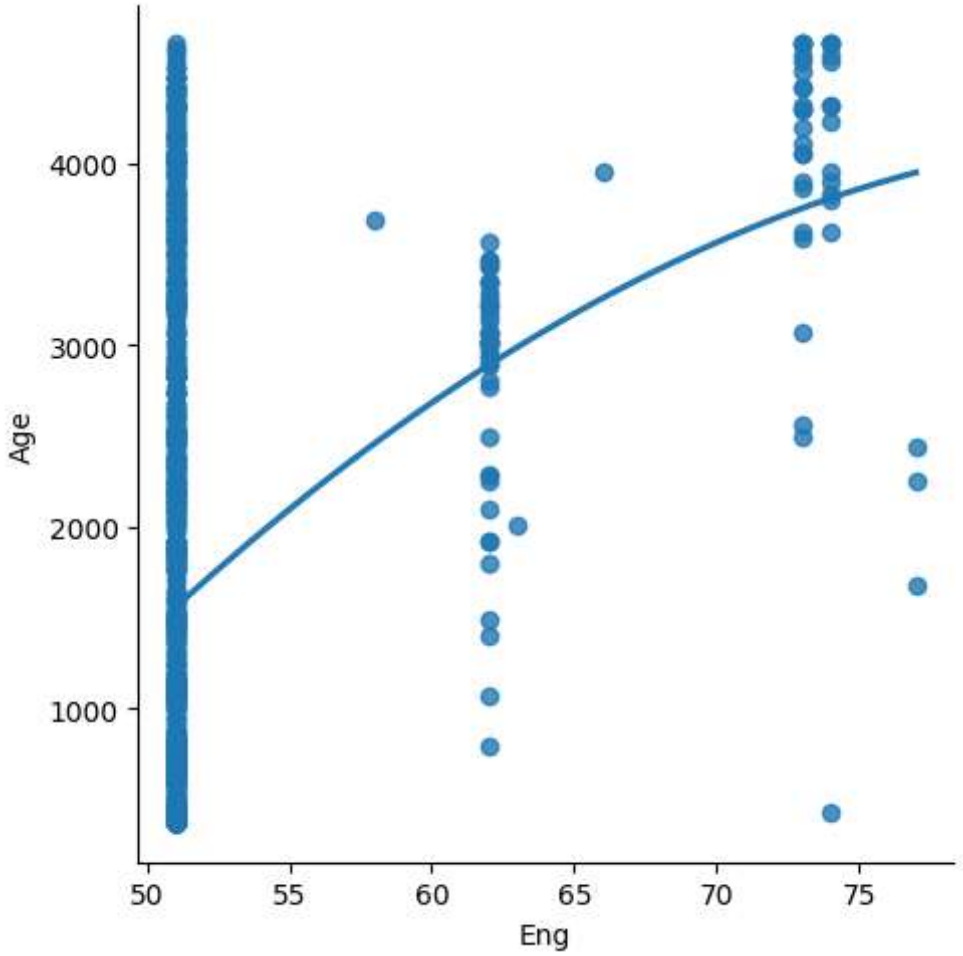
	Eng	Age
0	51	882
1	51	1186
2	74	4658
3	51	2739
4	73	3074
5	74	3623
6	51	731
7	51	1521
8	73	4049
9	51	3653

In [23]:

```
1 sns.lmplot(x='Eng',y='Age',data=dt,order=2,ci=None)
```

Out[23]:

```
<seaborn.axisgrid.FacetGrid at 0x230977c8b20>
```



In [10]:

```
1 dt.describe()
```

Out[10]:

	Eng	Age
count	1538.000000	1538.000000
mean	51.904421	1650.980494
std	3.988023	1289.522278
min	51.000000	366.000000
25%	51.000000	670.000000
50%	51.000000	1035.000000
75%	51.000000	2616.000000
max	77.000000	4658.000000

In [11]:

```
1 dt.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1538 entries, 0 to 1537
Data columns (total 2 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   Eng     1538 non-null   int64
 1   Age     1538 non-null   int64
dtypes: int64(2)
memory usage: 24.2 KB
```

In [12]:

```
1 dt.fillna(method='ffill')
```

Out[12]:

	Eng	Age
0	51	882
1	51	1186
2	74	4658
3	51	2739
4	73	3074
...	...	...
1533	51	3712
1534	74	3835
1535	51	2223
1536	51	2557
1537	51	1766

1538 rows × 2 columns

In [15]:

```
1 x=np.array(dt['Eng']).reshape(-1,1)
2 y=np.array(dt['Age']).reshape(-1,1)
```

In [16]:

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

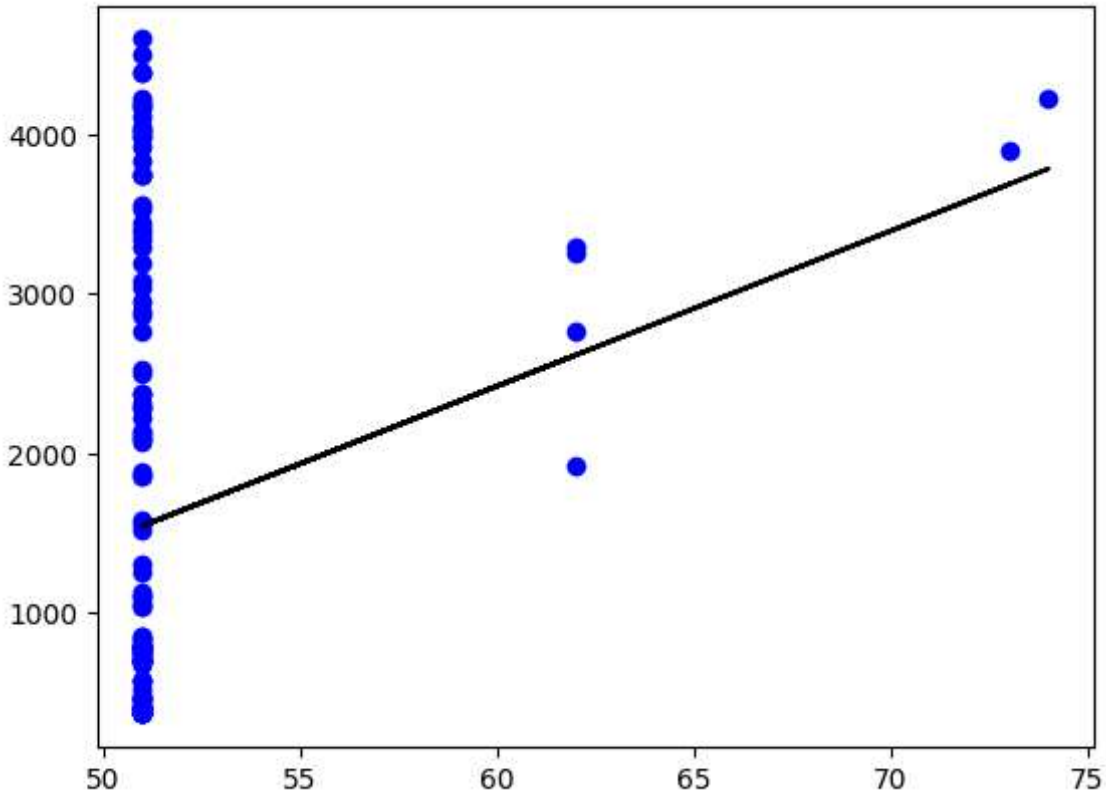
In [17]:

```
1 X_train,X_test,y_train,y_test=train_test_split(x,y,test_size=0.25)
2 reg=LinearRegression()
3 reg.fit(X_train,y_train)
4 print(reg.score(X_test,y_test))
```

```
0.08100123668984383
```

In [24]:

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

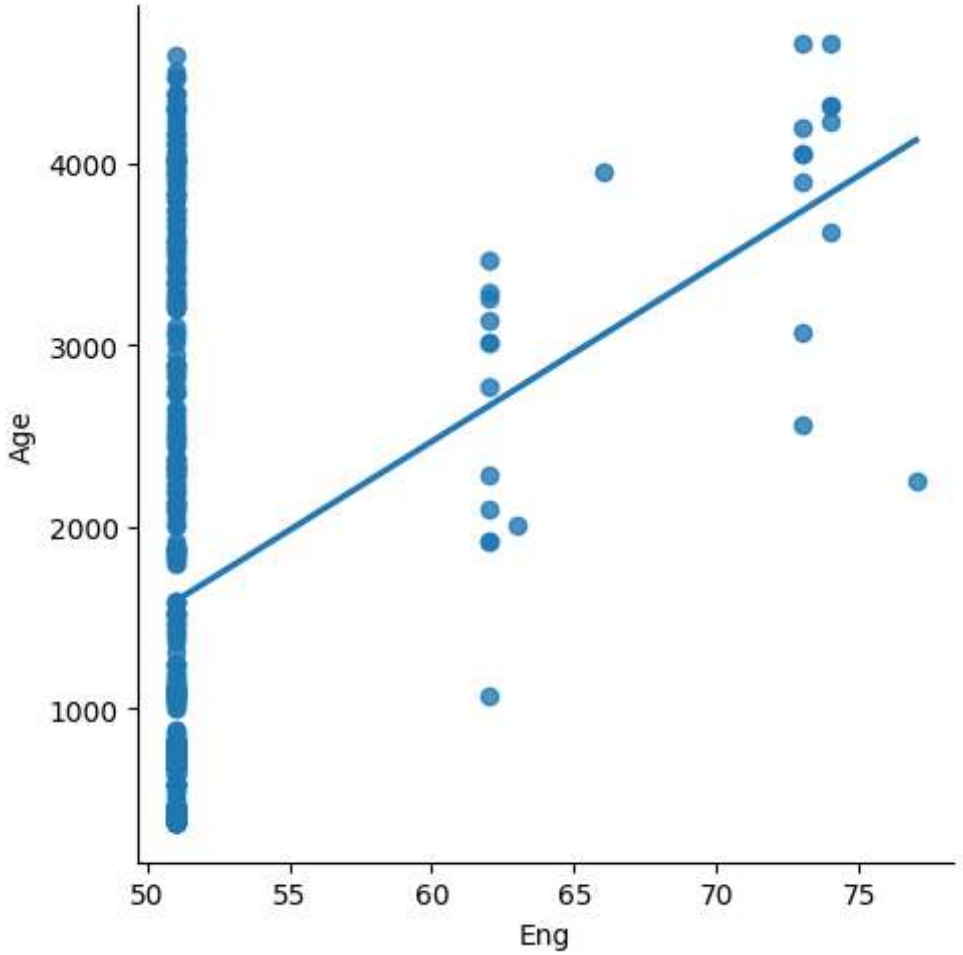


In [19]:

```
dt500=dt[:][:500]
sns.lmplot(x='Eng',y='Age',data=dt500,order=1,ci=None)
```

Out[19]:

```
<seaborn.axisgrid.FacetGrid at 0x230841ae350>
```



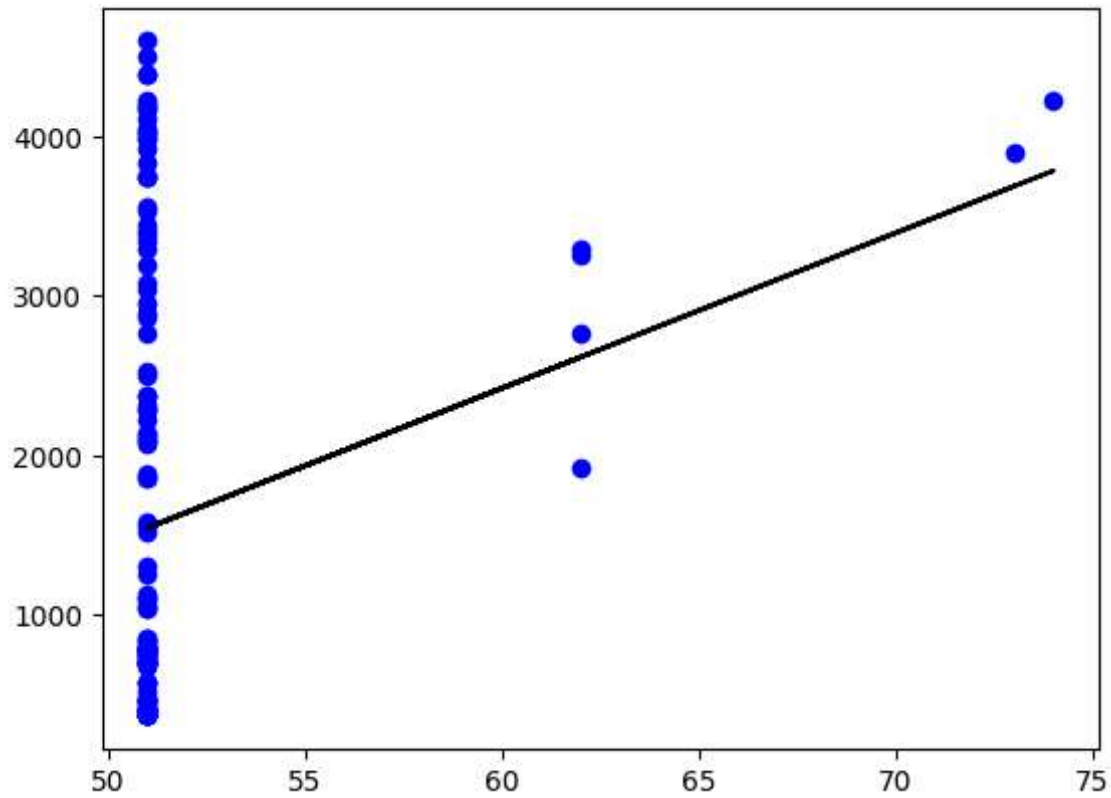
In [20]:

```
dt500.fillna(method='ffill',inplace=True)
X=np.array(dt500['Eng']).reshape(-1,1)
y=np.array(dt500['Age']).reshape(-1,1)
dt500.dropna(inplace=True)
X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.25)
reg=LinearRegression()
reg.fit(X_train,y_train)
print("Regression:",reg.score(X_test,y_test))
y_pred=reg.predict(X_test)
plt.scatter(X_test,y_test,color='b')
plt.plot(X_test,y_pred,color='k')
plt.show
```

```
Regression: 0.044422623802540806
```

Out[20]:

```
<function matplotlib.pyplot.show(close=None, block=None)>
```



In [21]:

```
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score
mode1=LinearRegression()
mode1.fit(X_train,y_train)
y_pred=mode1.predict(X_test)
r2=r2_score(y_test,y_pred)
print("R2 score: ",r2)
```

```
R2 score:  0.044422623802540806
```

In [ ]:

```
#conclusion:Linear regression is not fit for the model
```