In [1]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
```

In [2]:
```
df=pd.read_csv(r"C:\Users\chinta pavani\Documents\used_cars_data.csv")
df
```

Out[2]:

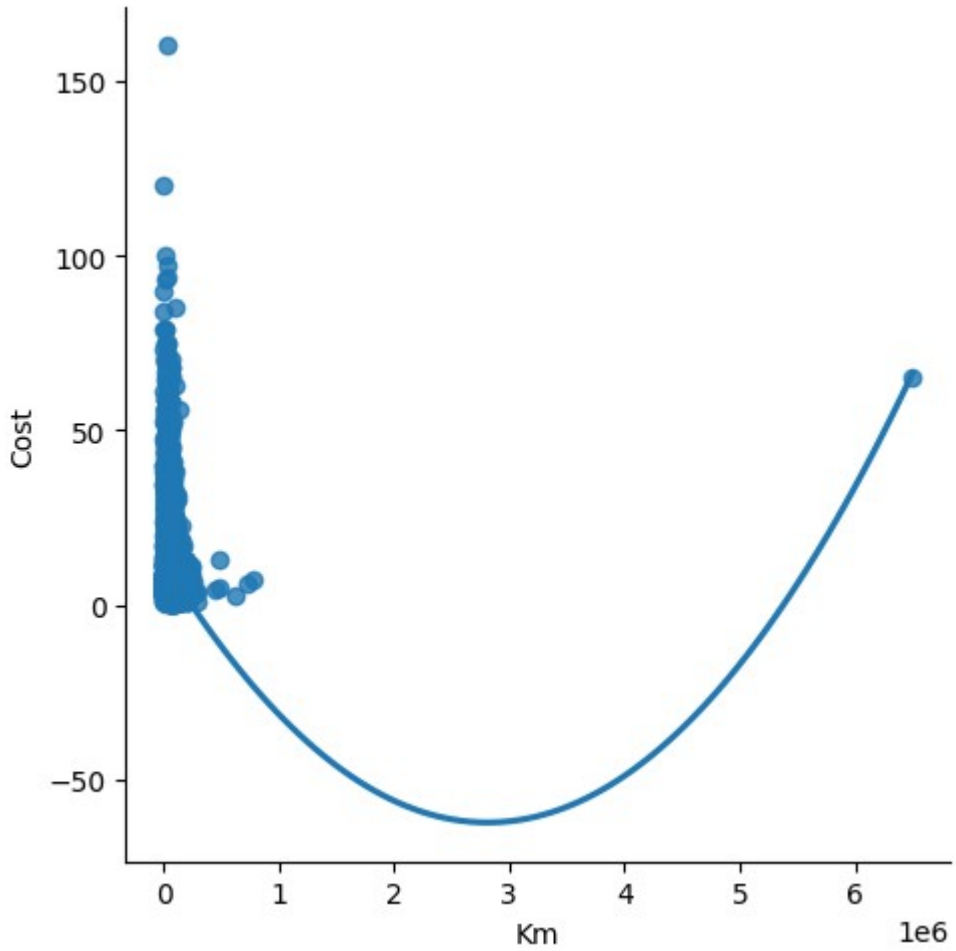
	S.No.	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Ow
0	0	Maruti Wagon R LXI CNG	Mumbai	2010	72000	CNG	Manual	
1	1	Hyundai Creta 1.6 CRDi SX Option	Pune	2015	41000	Diesel	Manual	
2	2	Honda Jazz V	Chennai	2011	46000	Petrol	Manual	
3	3	Maruti Ertiga VDI	Chennai	2012	87000	Diesel	Manual	
4	4	Audi A4 New 2.0 TDI Multitronic	Coimbatore	2013	40670	Diesel	Automatic	
...	...	...	...	...	...	...	...	
7248	7248	Volkswagen Vento Diesel Trendline	Hyderabad	2011	89411	Diesel	Manual	
7249	7249	Volkswagen Polo GT TSI	Mumbai	2015	59000	Petrol	Automatic	
7250	7250	Nissan Micra Diesel XV	Kolkata	2012	28000	Diesel	Manual	
7251	7251	Volkswagen Polo GT TSI	Pune	2013	52262	Petrol	Automatic	
7252	7252	Mercedes-Benz E-Class 2009-2013 E 220 CDI Avan...	Kochi	2014	72443	Diesel	Automatic	

7253 rows × 14 columns

In [3]:
```
df = df[['Kilometers_Driven','Price']]
df.columns=['Km','Cost']
```

In [4]:
```
sns.lmplot(x='Km',y='Cost',data=df,order=2,ci=None)
```

Out[4]: <seaborn.axisgrid.FacetGrid at 0x148691330d0>



In [5]:
```
df.describe()
```

Out[5]:

	Km	Cost
count	7.253000e+03	6019.000000
mean	5.869906e+04	9.479468
std	8.442772e+04	11.187917
min	1.710000e+02	0.440000
25%	3.400000e+04	3.500000
50%	5.341600e+04	5.640000
75%	7.300000e+04	9.950000
max	6.500000e+06	160.000000

In [6]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7253 entries, 0 to 7252
Data columns (total 2 columns):
 #   Column  Non-Null Count  Dtype  
---  ------  --------------  -----  
 0   Km      7253 non-null   int64  
 1   Cost    6019 non-null   float64
dtypes: float64(1), int64(1)
memory usage: 113.5 KB
```

In [7]:
```
df.fillna(method='ffill',inplace=True)
```

```
C:\Users\chinta pavani\AppData\Local\Temp\ipykernel_20864\4116506308.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-do
cs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (http
s://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returni
ng-a-view-versus-a-copy)
  df.fillna(method='ffill',inplace=True)
```

In [8]:
```
df.dropna(inplace=True)
```

```
C:\Users\chinta pavani\AppData\Local\Temp\ipykernel_20864\1379821321.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-do
cs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (http
s://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returni
ng-a-view-versus-a-copy)
  df.dropna(inplace=True)
```

In [9]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7253 entries, 0 to 7252
Data columns (total 2 columns):
 #   Column  Non-Null Count  Dtype  
---  ------  --------------  -----  
 0   Km      7253 non-null   int64  
 1   Cost    7253 non-null   float64
dtypes: float64(1), int64(1)
memory usage: 113.5 KB
```

In [10]:
```
df.isnull().sum()
```

Out[10]:
```
Km      0
Cost    0
dtype: int64
```

In [11]:
```
df.head(10)
```

Out[11]:

	Km	Cost
0	72000	1.75
1	41000	12.50
2	46000	4.50
3	87000	6.00
4	40670	17.74
5	75000	2.35
6	86999	3.50
7	36000	17.50
8	64430	5.20
9	65932	1.95

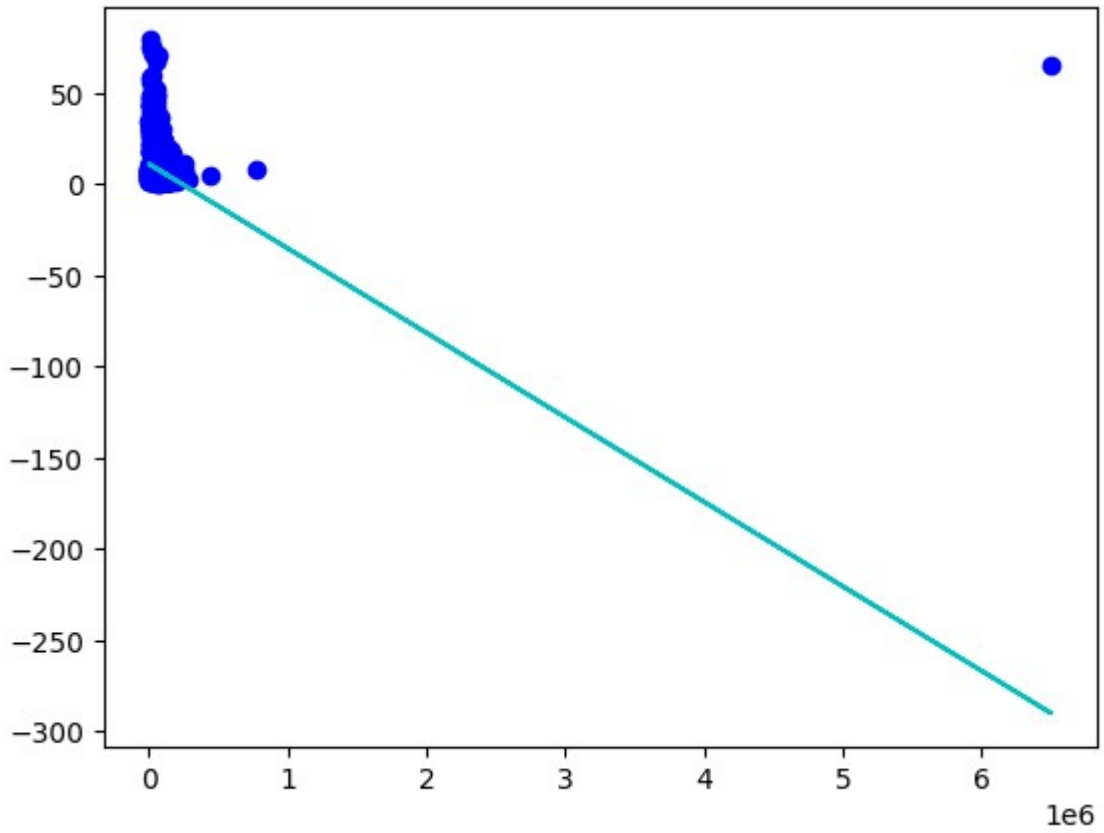
In [13]:
```
x=np.array(df['Km']).reshape(-1,1)
y=np.array(df['Cost']).reshape(-1,1)
```

In [14]:
```
x_train, x_test, y_train, y_test = train_test_split(x,y,test_size=0.25)
regr = LinearRegression()
regr.fit(x_train,y_train)
print(regr.score(x_test,y_test))
```

```
-0.7037051332330637
```

In [15]:
```
y_pred = regr.predict(x_test)
plt.scatter(x_test,y_test,color='b')
plt.plot(x_test,y_pred,color='c')
```

Out[15]:
```
[<matplotlib.lines.Line2D at 0x1486b599f90>]
```



In [16]:
```
df200 = df[:][:200]
df200
```

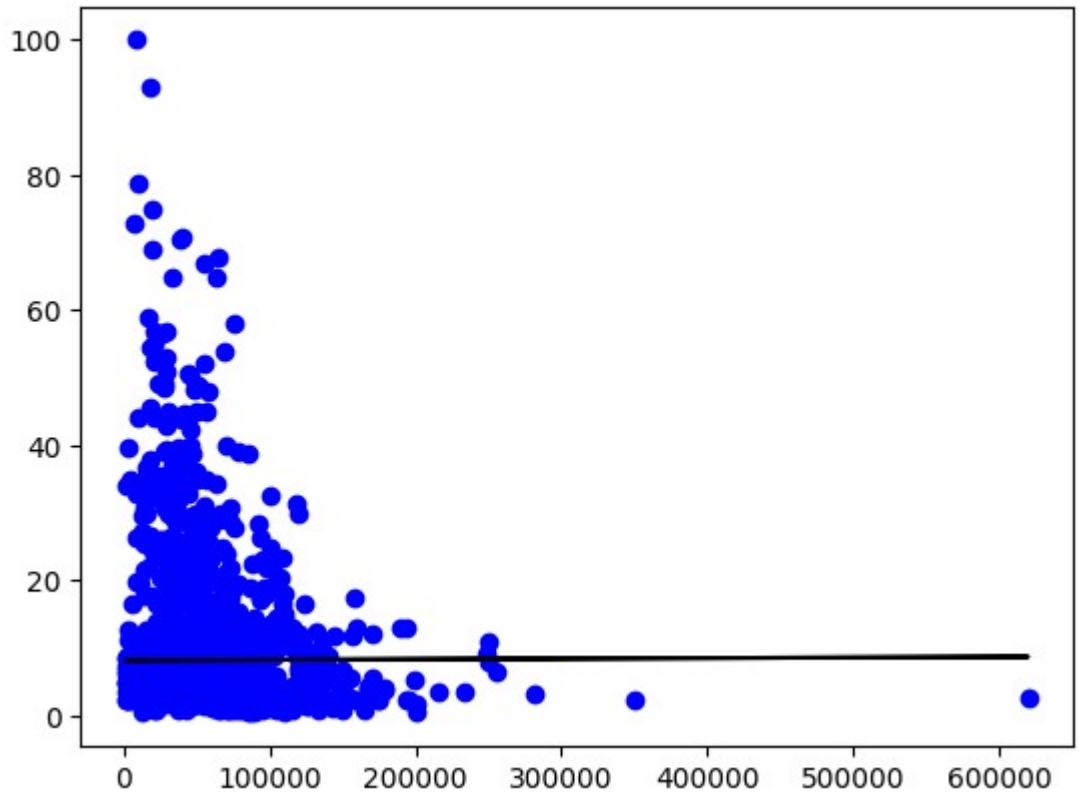
Out[16]:

	Km	Cost
0	72000	1.75
1	41000	12.50
2	46000	4.50
3	87000	6.00
4	40670	17.74
...	...	...
195	52000	3.50
196	43571	3.55
197	50000	3.25
198	113000	4.50
199	90000	5.35

200 rows × 2 columns

In [17]:
```
df200.fillna(method='ffill',inplace=True)
X=np.array(df['Km']).reshape(-1,1)
y=np.array(df['Cost']).reshape(-1,1)
df200.dropna(inplace=True)
X_train,x_test,y_train,y_test = train_test_split(X,y,test_size=0.25)
regr=LinearRegression()
regr.fit(X_train,y_train)
print("Regressin: ",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()
```

```
Regressin:  -0.0010269867051837522
```



In [18]:
```
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:  -0.0010269867051837522
```

In [19]:
```
from sklearn.linear_model import LinearRegression
from sklearn import metrics
model = LinearRegression()
model.fit(X_train,y_train)
y_pred = model.predict(x_test)
r2=metrics.mean_squared_error(y_test,y_pred)
print('MSE: ',r2)
```

```
MSE:  114.70068714601699
```

In [20]:
```
import pandas as pd
import numpy as np
from sklearn.linear_model import LogisticRegression
from sklearn.preprocessing import StandardScaler
```

In [21]:
```
a=pd.read_csv(r"C:\Users\chinta pavani\Documents\used_cars_data.csv")
a
```

Out[21]:

	S.No.	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Ow
0	0	Maruti Wagon R LXI CNG	Mumbai	2010	72000	CNG	Manual	
1	1	Hyundai Creta 1.6 CRDi SX Option	Pune	2015	41000	Diesel	Manual	
2	2	Honda Jazz V	Chennai	2011	46000	Petrol	Manual	
3	3	Maruti Ertiga VDI	Chennai	2012	87000	Diesel	Manual	
4	4	Audi A4 New 2.0 TDI Multitronic	Coimbatore	2013	40670	Diesel	Automatic	
...	...	...	...	...	...	...	...	
7248	7248	Volkswagen Vento Diesel Trendline	Hyderabad	2011	89411	Diesel	Manual	
7249	7249	Volkswagen Polo GT TSI	Mumbai	2015	59000	Petrol	Automatic	
7250	7250	Nissan Micra Diesel XV	Kolkata	2012	28000	Diesel	Manual	
7251	7251	Volkswagen Polo GT TSI	Pune	2013	52262	Petrol	Automatic	
7252	7252	Mercedes-Benz E-Class 2009-2013 E 220 CDI Avan...	Kochi	2014	72443	Diesel	Automatic	

7253 rows × 14 columns

In [22]:
```
a.describe()
```

Out[22]:

	S.No.	Year	Kilometers_Driven	Seats	Price
count	7253.000000	7253.000000	7.253000e+03	7200.000000	6019.000000
mean	3626.000000	2013.365366	5.869906e+04	5.279722	9.479468
std	2093.905084	3.254421	8.442772e+04	0.811660	11.187917
min	0.000000	1996.000000	1.710000e+02	0.000000	0.440000
25%	1813.000000	2011.000000	3.400000e+04	5.000000	3.500000
50%	3626.000000	2014.000000	5.341600e+04	5.000000	5.640000
75%	5439.000000	2016.000000	7.300000e+04	5.000000	9.950000
max	7252.000000	2019.000000	6.500000e+06	10.000000	160.000000

In [23]:
```
a.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7253 entries, 0 to 7252
Data columns (total 14 columns):
 #   Column             Non-Null Count  Dtype  
---  ------             --------------  -----  
 0   S.No.              7253 non-null   int64  
 1   Name               7253 non-null   object 
 2   Location           7253 non-null   object 
 3   Year               7253 non-null   int64  
 4   Kilometers_Driven  7253 non-null   int64  
 5   Fuel_Type          7253 non-null   object 
 6   Transmission       7253 non-null   object 
 7   Owner_Type         7253 non-null   object 
 8   Mileage            7251 non-null   object 
 9   Engine             7207 non-null   object 
 10  Power              7207 non-null   object 
 11  Seats              7200 non-null   float64
 12  New_Price          1006 non-null   object 
 13  Price              6019 non-null   float64
dtypes: float64(2), int64(3), object(9)
memory usage: 793.4+ KB
```

In [24]:
```
a.isnull().sum()
```

Out[24]:
```
S.No.                0
Name                 0
Location             0
Year                 0
Kilometers_Driven    0
Fuel_Type            0
Transmission         0
Owner_Type           0
Mileage              2
Engine              46
Power               46
Seats               53
New_Price         6247
Price             1234
dtype: int64
```

In [25]:
```
a.fillna(method='ffill',inplace=True)
```

In [26]:
```
a.dropna(inplace=True)
```

In [27]:
```
a.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 7251 entries, 2 to 7252
Data columns (total 14 columns):
 #   Column             Non-Null Count  Dtype  
---  ------             --------------  -----  
 0   S.No.              7251 non-null   int64  
 1   Name               7251 non-null   object 
 2   Location           7251 non-null   object 
 3   Year               7251 non-null   int64  
 4   Kilometers_Driven  7251 non-null   int64  
 5   Fuel_Type          7251 non-null   object 
 6   Transmission       7251 non-null   object 
 7   Owner_Type         7251 non-null   object 
 8   Mileage            7251 non-null   object 
 9   Engine             7251 non-null   object 
 10  Power              7251 non-null   object 
 11  Seats              7251 non-null   float64
 12  New_Price          7251 non-null   object 
 13  Price              7251 non-null   float64
dtypes: float64(2), int64(3), object(9)
memory usage: 849.7+ KB
```

In [28]:
```
a.isnull().sum()
```

Out[28]:
```
S.No.                0
Name                 0
Location             0
Year                 0
Kilometers_Driven    0
Fuel_Type            0
Transmission         0
Owner_Type           0
Mileage              0
Engine               0
Power                0
Seats                0
New_Price            0
Price                0
dtype: int64
```

In [29]:
```
print("This DataFrame has %d rows and %d columns"%(a.shape))
```

```
This DataFrame has 7251 rows and 14 columns
```

In [30]:
```
a.head()
```

Out[30]:

	S.No.	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_T
2	2	Honda Jazz V	Chennai	2011	46000	Petrol	Manual	
3	3	Maruti Ertiga VDI	Chennai	2012	87000	Diesel	Manual	
4	4	Audi A4 New 2.0 TDI Multitronic	Coimbatore	2013	40670	Diesel	Automatic	Se
5	5	Hyundai EON LPG Era Plus Option	Hyderabad	2012	75000	LPG	Manual	
6	6	Nissan Micra Diesel XV	Jaipur	2013	86999	Diesel	Manual	

In [31]:
```
feature_matrix = a.iloc[:,0:13]
target_vector = a.iloc[:,-1]
```

In [32]:
```
print("The feature_matrix has %d rows and %d columns"%(feature_matrix.shap
```

```
The feature_matrix has 7251 rows and 13 columns
```

In [ ]:
```

```