Hyundai_ML1

August 13, 2021

[59]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

[60]:
```
data = pd.read_excel('toyota_data4.xlsx') #read from dataset

data.head() # view first few rows of the data
```

[60]:

	x	r
0	23.0	13500
1	23.0	13750
2	24.0	13950
3	26.0	14950
4	30.0	13750

[61]:
```
data.info()
```

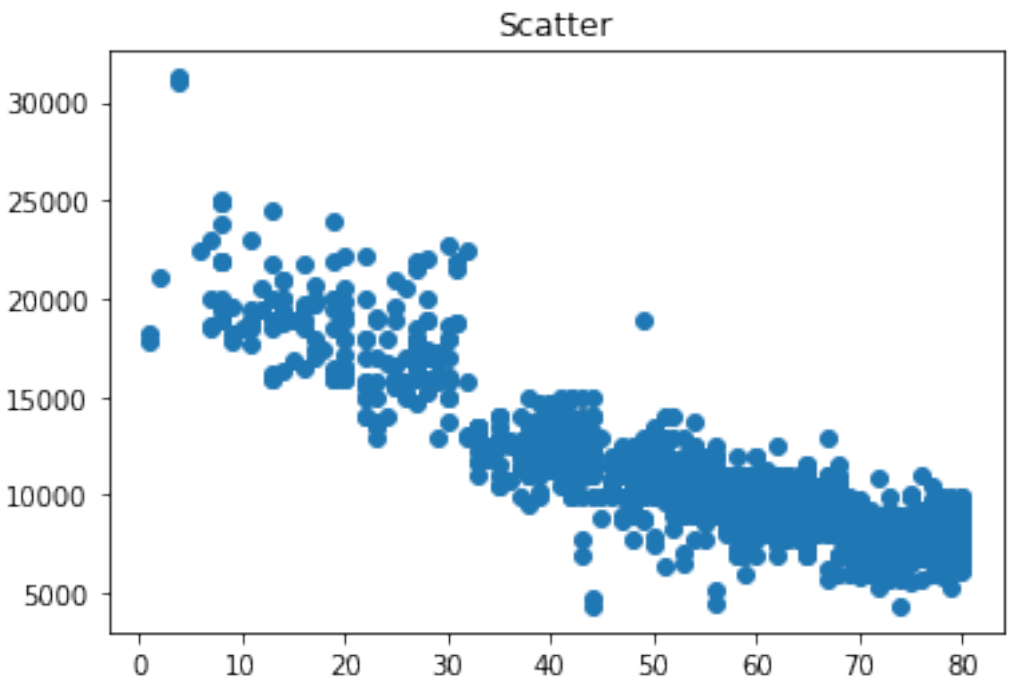
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1436 entries, 0 to 1435
Data columns (total 2 columns):
 #   Column  Non-Null Count  Dtype  
---  ------  --------------  -----  
 0   x       1336 non-null   float64
 1   r       1436 non-null   int64  
dtypes: float64(1), int64(1)
memory usage: 22.6 KB
```

[62]:
```
data = data.dropna()
```

[63]:
```
size = data['x'].size
```

[64]:
```
plt.scatter(data['x'],data['r'])
plt.title('Scatter')
```

[64]:
```
Text(0.5, 1.0, 'Scatter')
```



[65]:
```
data['x_sqr'] = data['x']**2
data
```

[65]:

	x	r	x_sqr
0	23.0	13500	529.0
1	23.0	13750	529.0
2	24.0	13950	576.0
3	26.0	14950	676.0
4	30.0	13750	900.0
…	…	…	…
1429	78.0	8950	6084.0
1430	80.0	8450	6400.0
1432	72.0	10845	5184.0
1434	70.0	7250	4900.0
1435	76.0	6950	5776.0

[1336 rows x 3 columns]

[66]:
```
data['r_sqr'] = data['r']**2
data
```

[66]:

	x	r	x_sqr	r_sqr
0	23.0	13500	529.0	182250000
1	23.0	13750	529.0	189062500

```
2     24.0  13950   576.0  194602500
3     26.0  14950   676.0  223502500
4     30.0  13750   900.0  189062500
…     …     …       …          …
1429  78.0   8950  6084.0   80102500
1430  80.0   8450  6400.0   71402500
1432  72.0  10845  5184.0  117614025
1434  70.0   7250  4900.0   52562500
1435  76.0   6950  5776.0   48302500

[1336 rows x 4 columns]
```

[67]:
```
data['xr'] = data['x']*data['r']
```

[68]:
```
data
```

[68]:
```
          x      r   x_sqr      r_sqr        xr
0     23.0  13500   529.0  182250000  310500.0
1     23.0  13750   529.0  189062500  316250.0
2     24.0  13950   576.0  194602500  334800.0
3     26.0  14950   676.0  223502500  388700.0
4     30.0  13750   900.0  189062500  412500.0
…     …     …       …          …          …
1429  78.0   8950  6084.0   80102500  698100.0
1430  80.0   8450  6400.0   71402500  676000.0
1432  72.0  10845  5184.0  117614025  780840.0
1434  70.0   7250  4900.0   52562500  507500.0
1435  76.0   6950  5776.0   48302500  528200.0

[1336 rows x 5 columns]
```

[69]:
```
(sum(data['x_sqr'])-(sum(data['x'])**2/size))
```

[69]:
```
461350.4041916169
```

[70]:
```
w1 = (sum(data['xr']) - sum(data['x'])*sum(data['r']/size)) /␣
↪(sum(data['x_sqr'])-(sum(data['x'])**2/size))
w1
```

[70]:
```
-170.94936688237877
```

[71]:
```
w0 = sum(data['r'])/size - w1*sum(data['x'])/size
w0
```

[71]:
```
20267.598061360455
```

[72]:
```
predicted = w0+w1*data['x']
```

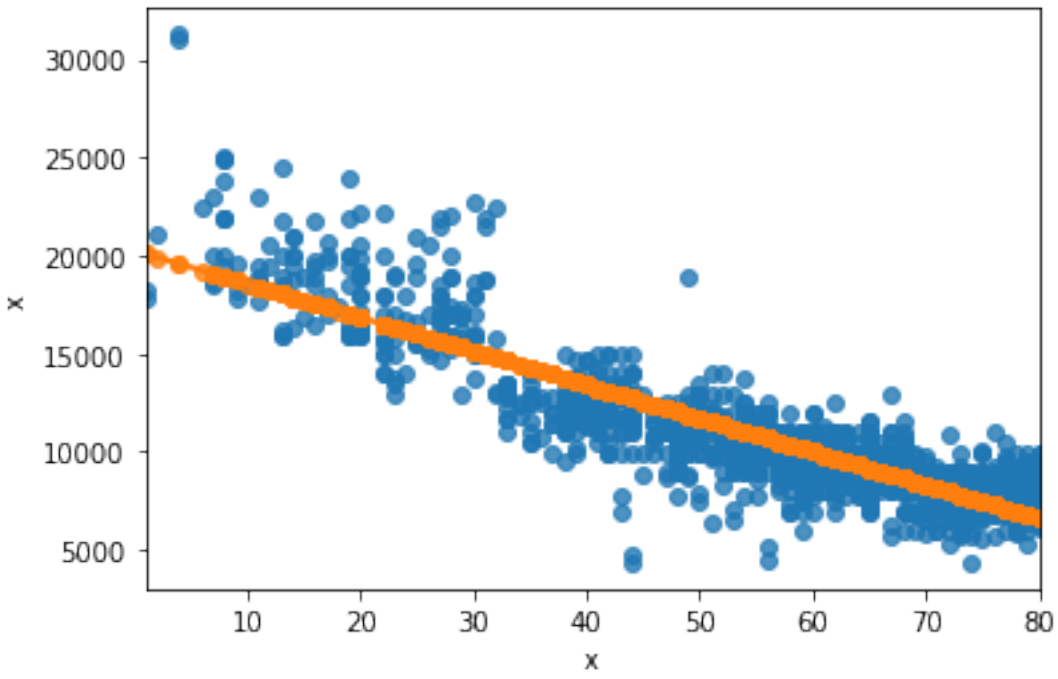
[73]:
```
predicted
```

[73]:
```
0       16335.762623
1       16335.762623
2       16164.813256
3       15822.914522
4       15139.117055
              …
1429     6933.547445
1430     6591.648711
1432     7959.243646
1434     8301.142380
1435     7275.446178
Name: x, Length: 1336, dtype: float64
```

[81]:
```
sns.regplot(x = 'x',y = 'r',data = data,fit_reg=False)#.
 ↪set(xlim=(0,100),ylim=(0,500))

sns.regplot(x = 'x',y = predicted,data = data)
```

[81]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7770a46fd0>
```



[75]:
```
error = np.sum((data['r']-predicted)**2)/size
```

[76]:
```
error
```

[76]: 2987200.437131624

[76]: