```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt


import os
from google.colab import drive


drive.mount('/content/drive')
```

```
Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.m
```

```
!ls
```

```
 data.csv     Salaries.csv                Untitled0.ipynb
 data..xlsx  'salery of the year.ipynb'
```

```
os.chdir('/content/drive/My Drive/Colab Notebooks/')


dataset = pd.read_excel('data..xlsx')


pd.read_excel('data..xlsx')
```

	S. No.	Constituency	Winning Margin	Air India Top 20 flights	Maximum Delay in Minutes
0	1	Alappuzha	19407	Bangalore-Mumbai	182
1	2	Alathur	37312	Ahmedabad-Mumbai	203
2	3	Attingal	69378	Hyderabad-Mumbai	240
3	4	Chalakudy	13884	Mumbai-Goa	164

```
dataset
```

	S. No.	Constituency	Winning Margin	Air India Top 20 flights	Maximum Delay in Minutes
0	1	Alappuzha	19407	Bangalore-Mumbai	182
1	2	Alathur	37312	Ahmedabad-Mumbai	203
2	3	Attingal	69378	Hyderabad-Mumbai	240
3	4	Chalakudy	13884	Mumbai-Goa	164
4	5	Ernakulum	87047	Delhi-Kolkata	265
5	6	Idukki	50542	Chennai-Delhi	226
6	7	Kannur	6566	Delhi-Bangalore	156
7	8	Kasaragod	6921	Mumbai-Chennai	161
8	9	Kollam	37649	Kolkata-Delhi	219
9	10	Kottayam	120599	Mumbai–Delhi	328
10	11	Kozhikode	16883	Hyderabad-Delhi	181
11	12	Malappuram	194740	Delhi–Mumbai	340
12	13	Mavelikkara	32737	Mumbai-Ahmedabad	202
13	14	Palakkad	105300	Mumbai-Hyderabad	284
14	15	Pathanamthitta	56191	Chennai-Mumbai	234
15	16	Ponnani	25410	Bangalore-Delhi	199
16	17	Thiruvananthapuram	15470	Goa-Mumbai	178
17	18	Thrissur	38228	Delhi-Chennai	225
18	19	Vadakara	3306	Delhi-Hyderabad	146
19	20	Wayanad	20870	Mumbai-Bangalore	197

```
x = dataset.iloc[:,-1].values.reshape(-1,1)
```

```
y = dataset.iloc[:,-3].values.reshape(-1,1)
```

```
x
```

```
array([[182],
       [203],
       [240],
       [164],
       [265],
       [226],
       [156],
       [161],
       [219],
       [328],
       [181],
       [340],
       [202],
       [284],
       [234],
       [199],
       [178],
       [225],
       [146],
       [197]])
```

```
y
```

```
array([[ 19407],
       [ 37312],
       [ 69378],
       [ 13884],
       [ 87047],
       [ 50542],
       [  6566],
       [  6921],
       [ 37649],
       [120599],
       [ 16883],
       [194740],
       [ 32737],
       [105300],
       [ 56191],
       [ 25410],
       [ 15470],
       [ 38228],
       [  3306],
       [ 20870]])
```

```
x.shape
y.shape
```

```
(20, 1)
```

```
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x,y,test_size=1/3,random_state=0)
```

```
from sklearn.linear_model import LinearRegression
regressor = LinearRegression()
```

```
regressor.fit(x_train,y_train)
```

```
LinearRegression()
```

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

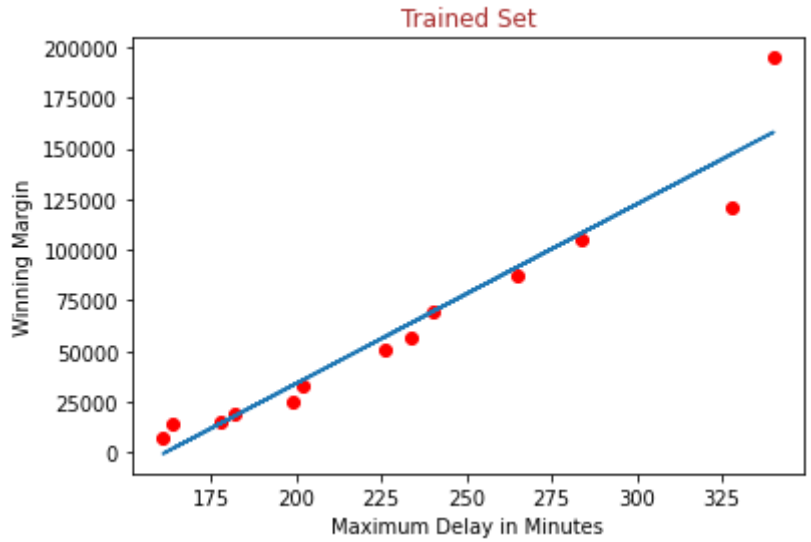
```
array([[-13939.03289191],
       [ 36552.82717407],
       [ 31237.89453555],
       [ 50725.9808768 ],
       [ 17064.74083282],
       [ 56040.91351532],
       [ -5080.8118277 ]])
```

```
y_test
```

```
array([[ 3306],
       [37312],
       [20870],
       [37649],
       [16883],
       [38228],
       [ 6566]])
```

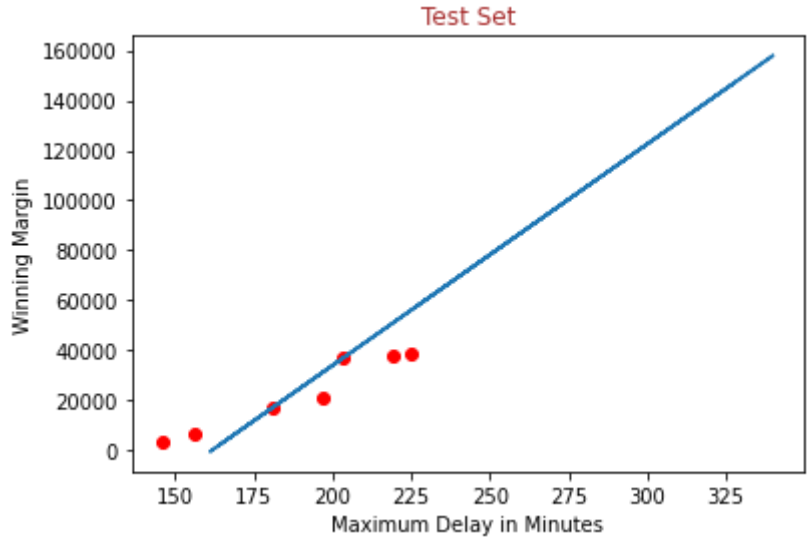
```
#train_set
plt.title('Trained Set', color='brown')
plt.xlabel("Maximum Delay in Minutes")
plt.ylabel("Winning Margin")
plt.scatter(x_train,y_train, color='red')
plt.plot(x_train, regressor.predict(x_train))
```

```
[<matplotlib.lines.Line2D at 0x7faeb8b1aa10>]
```




```
plt.title('Test Set', color='brown')
plt.xlabel("Maximum Delay in Minutes")
plt.ylabel("Winning Margin")
plt.plot(x_train, regressor.predict(x_train))
plt.scatter(x_test,y_test, color='red')
```

```
<matplotlib.collections.PathCollection at 0x7faeb8ab0250>
```




```
x_train
```

```
array([[284],
       [265],
       [240],
       [226],
       [234],
       [328],
       [161],
       [178],
       [340],
       [164],
       [182],
       [199],
       [202]])
```

```
r2_score = regressor.score(x_test,y_test)
```

```
r2_score
```

```
0.23812114554531594
```

```
r2_score*100,'%'
```

```
(23.812114554531593, '%')
```