In [2]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

In [5]:
```
data = pd.read_csv('studentdataset.csv')
print(data.describe())
```

```
             Age     Gender  Mid term marks  Assignments  Final Term  \
count  69.000000  69.000000       69.000000    69.000000   69.000000
mean   19.159420   1.507246       17.797101     6.456522   29.492754
std     1.779154   0.503610        9.080651     3.005900   13.241671
min    16.000000   1.000000        0.000000     0.000000    0.000000
25%    18.000000   1.000000       12.000000     4.000000   20.000000
50%    19.000000   2.000000       19.000000     7.000000   30.000000
75%    20.000000   2.000000       25.000000     9.000000   40.000000
max    22.000000   2.000000       30.000000    10.000000   50.000000

           Total  Performance Score
count  69.000000          69.000000
mean   53.746377          52.462319
std    17.075676          18.781758
min    16.000000          12.000000
25%    44.800000          42.000000
50%    55.700000          51.000000
75%    65.300000          66.000000
max    86.200000         100.000000
```

In [13]:
```
X = data['Total'] #independent
Y = data['Performance Score'] #dependent
mean_x = np.mean(X)
mean_y = np.mean(Y)
size = len(X)
```

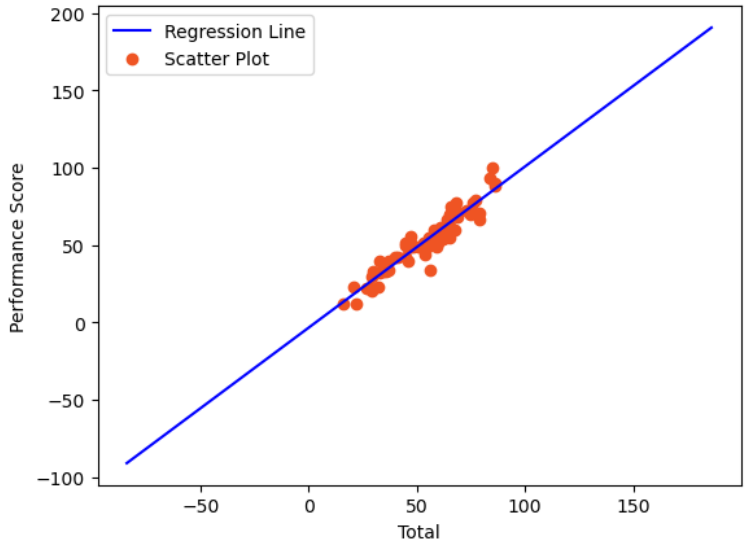

In [14]:
```
numerator = 0
denominator = 0
for i in range(size):
    numerator += (X[i] - mean_x)*(Y[i] - mean_y)
    denominator += (X[i]-mean_x) ** 2
m = numerator/denominator
c = mean_y - (m * mean_x)
c
```

Out[14]: -3.522404250003774

In [24]:
```
max_x = np.max(X) + 100
min_x = np.min(X) - 100

x = np.linspace(min_x, max_x, 100)
y =  m*x + c
```

In [25]:
```
import matplotlib.pyplot as plt

plt.plot(x, y, color='Blue', label='Regression Line')
plt.scatter(X, Y, c='#ef5423', label='Scatter Plot')

plt.xlabel('Total')
plt.ylabel('Performance Score')
plt.legend()
plt.show
```

Out[25]:
```
<function matplotlib.pyplot.show(close=None, block=None)>
```



In [23]:
```
n = 0
d = 0
for i in range(size):
    y_pred = m * X[i] + c
    n += (y_pred-mean_y) **2
    d += (Y[i] - mean_y) **2
r_square = n/d
r_square
```

Out[23]:
```
0.8968586231180552
```

In [ ]:
```

```