In [ ]:
```
%matplotlib inline
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
plt.rcParams["figure.figsize"] = [6.0, 3.0]

# Reading Data
data= pd.read_csv('realdataOne.csv')
# data= pd.read_csv('realdataTwo.csv')
print(data.shape)
data.head()
```

```
(10, 2)
```

Out[ ]:

	X_Values	Y_Values
0	1.0	1.8
1	2.0	2.4
2	3.3	2.3
3	4.3	3.8
4	5.3	5.3

In [ ]:
```
#Collecting X and Y
X= data['X_Values'].values
Y= data['Y_Values'].values
```

In [ ]:
```
x_values=np.array(X)
print("Sum of all the X_Values: ", round(x_values.sum(),2))

y_values=np.array(Y)
print("Sum of all the Y_Values: ", y_values.sum())

xy = [x_values * y_values for x_values, y_values in zip(X, Y)]
xy_values=np.array(xy)
print("Sum of all the XY_Values: ", round(xy_values.sum(),2))

xx = [x_values * x_values for x_values, x_values in zip(X, X)]
xx_values=np.array(xx)
print("Sum of all the XX_Values: ", xx_values.sum())
```

```
Sum of all the X_Values:  31.8
Sum of all the Y_Values:  32.5
Sum of all the XY_Values:  120.8
Sum of all the XX_Values:  121.34
```

In [ ]:
```
# Using the formula to calculate slope(b)
n= len(X)
b = round(((n* xy_values.sum())-(x_values.sum()*y_values.sum()))/ ((n*xx_values.sum()))
print("slope(b) is: " , b)
```

```
slope(b) is:  0.86
```

In [ ]:
```
# Using the formula to calculate intercept(a)
a = round((y_values.sum() - (b*x_values.sum()))/n ,2)
```

```
print("intercept(a): ", a)
```

```
intercept(a):  0.52
```

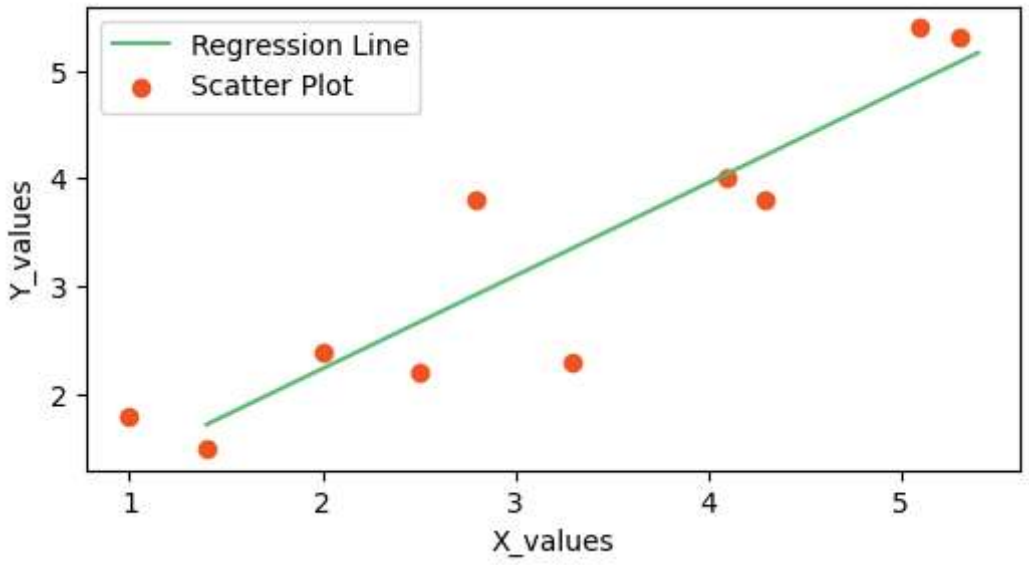
*** Training phase **

In [ ]:
```
# Then substitute Intercept(a) and Slope(b) in regression equation formula
# Regression Equation(y) = a + bx
x = np.array(X)
y=a+b*x
print(y)
```

```
[1.38  2.24  3.358 4.218 5.078 1.724 2.67  2.928 4.046 4.906]
```

In [ ]:
```
plt.plot(x,y, color='#58b970', label='Regression Line')
plt.scatter(X,Y, c='#ef5423', label='Scatter Plot')

plt.xlabel('X_values')
plt.ylabel('Y_values')
plt.legend()
plt.show()
```



** validation phase ***

In [ ]:
```
# After calculating a1, b1, a2, b2 in Training Phase, the values are not changed with
# Only ŷ values are changed with the new Real Data Sets.
# Regression Equation(y) = a + bx
x = np.array([1.5,2.9,3.7,4.7,5.1])
y=a+b*x
print(y)
```

```
[1.81  3.014 3.702 4.562 4.906]
```