**REG NO:20B01A1220**

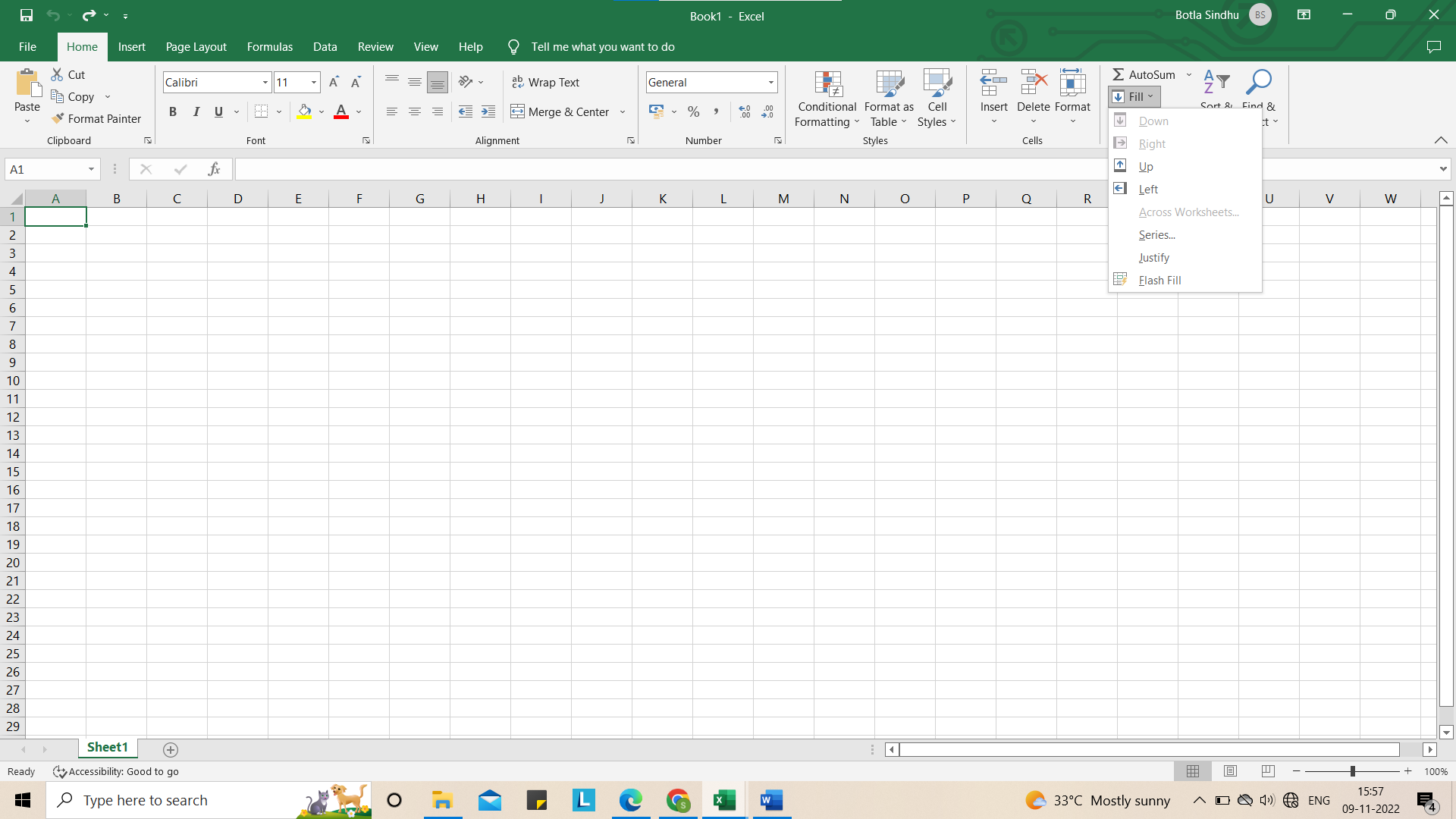
**DATE:21-08-22**

**EXPERIMENT NO: 01**

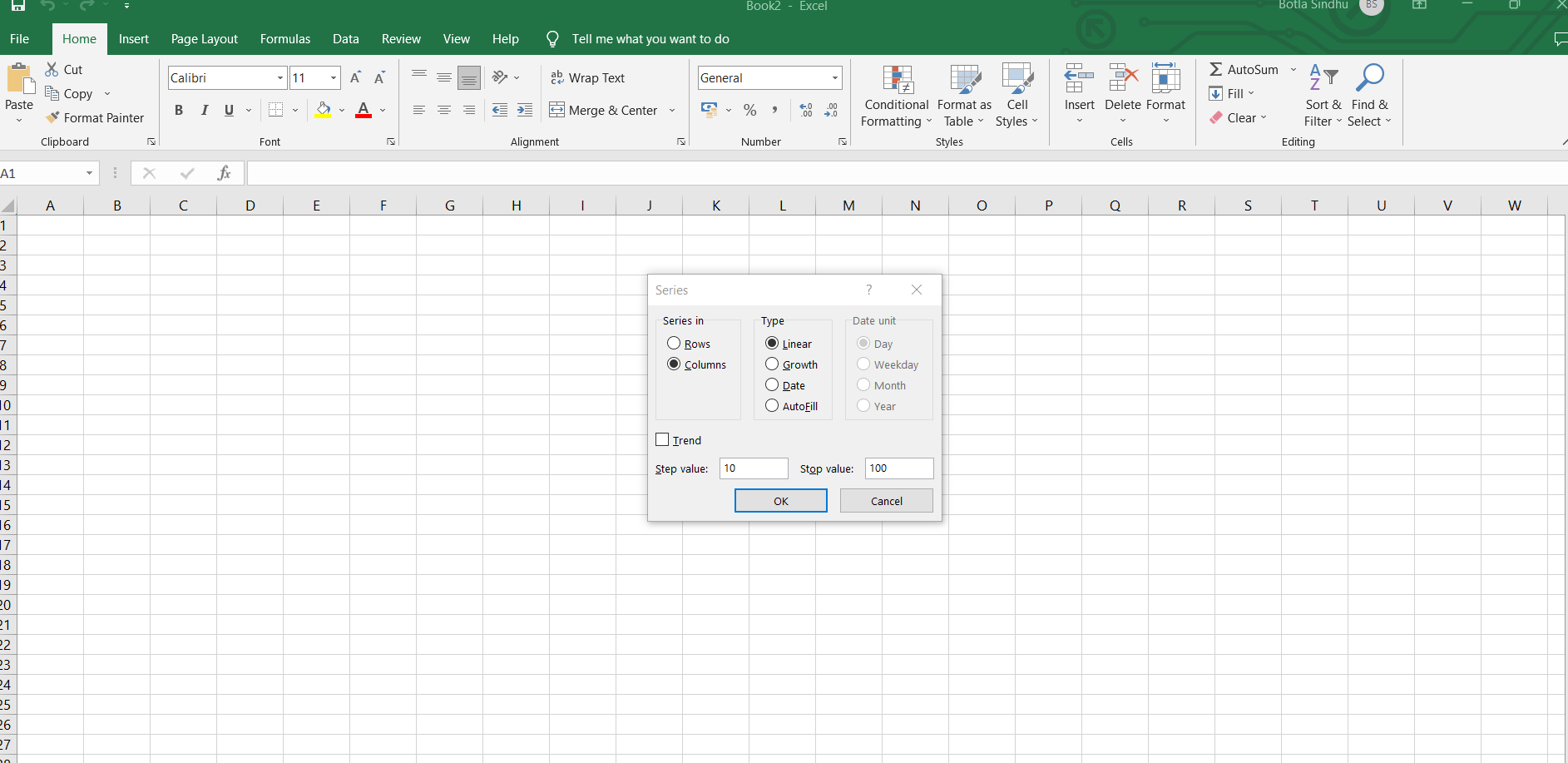
Perform Descriptive Statistics of given dataset using Data Analysis Toolbox of Excel.

PROCEDURE:

Step-1: Fill first cell with number which we want to start and then go to Home Menu Bar select fill option in Editing and then select series.

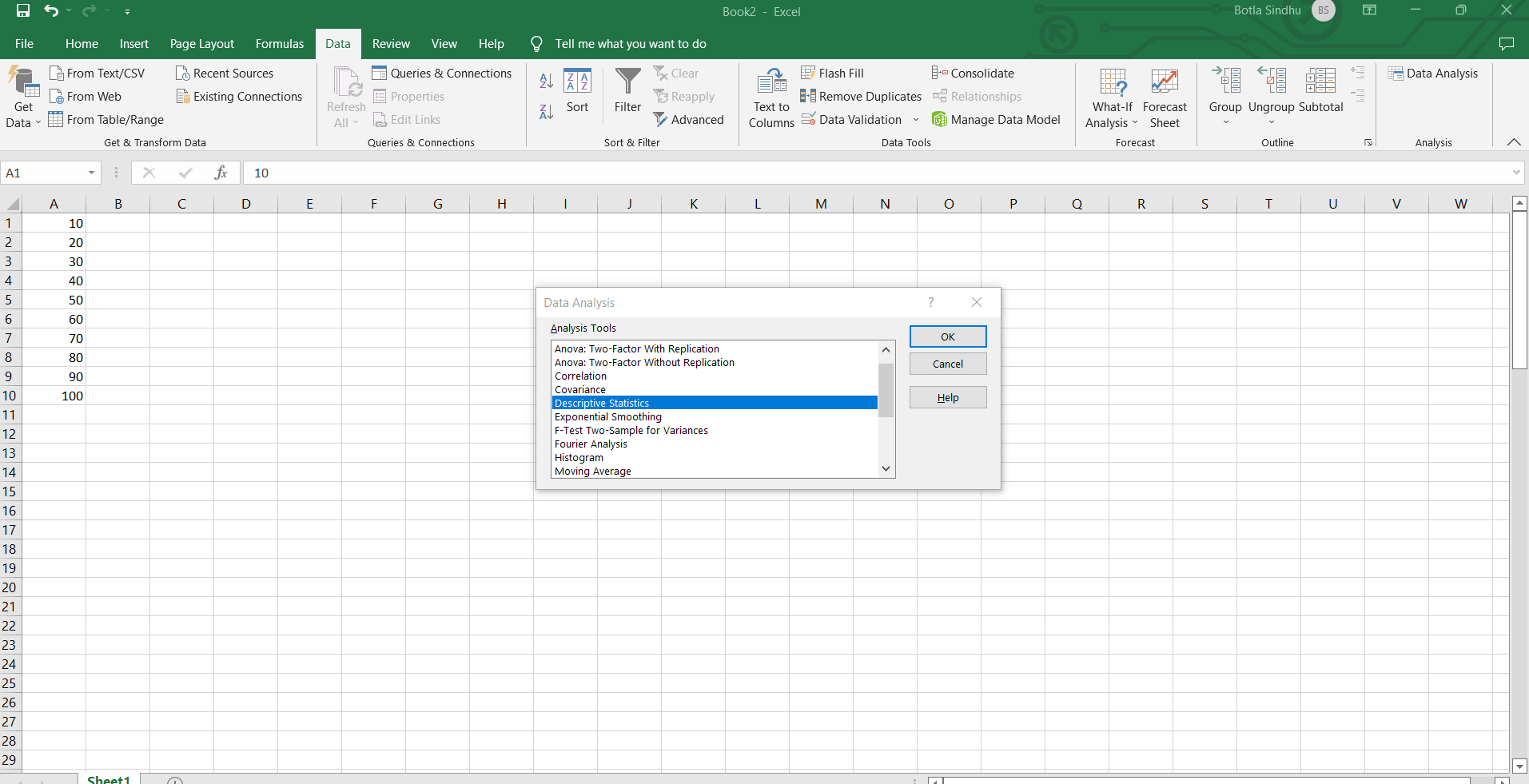


Step-2: Series dialog box will displayed. In that select Columns option and set type as Linear and required set step value and stop value and then click OK.



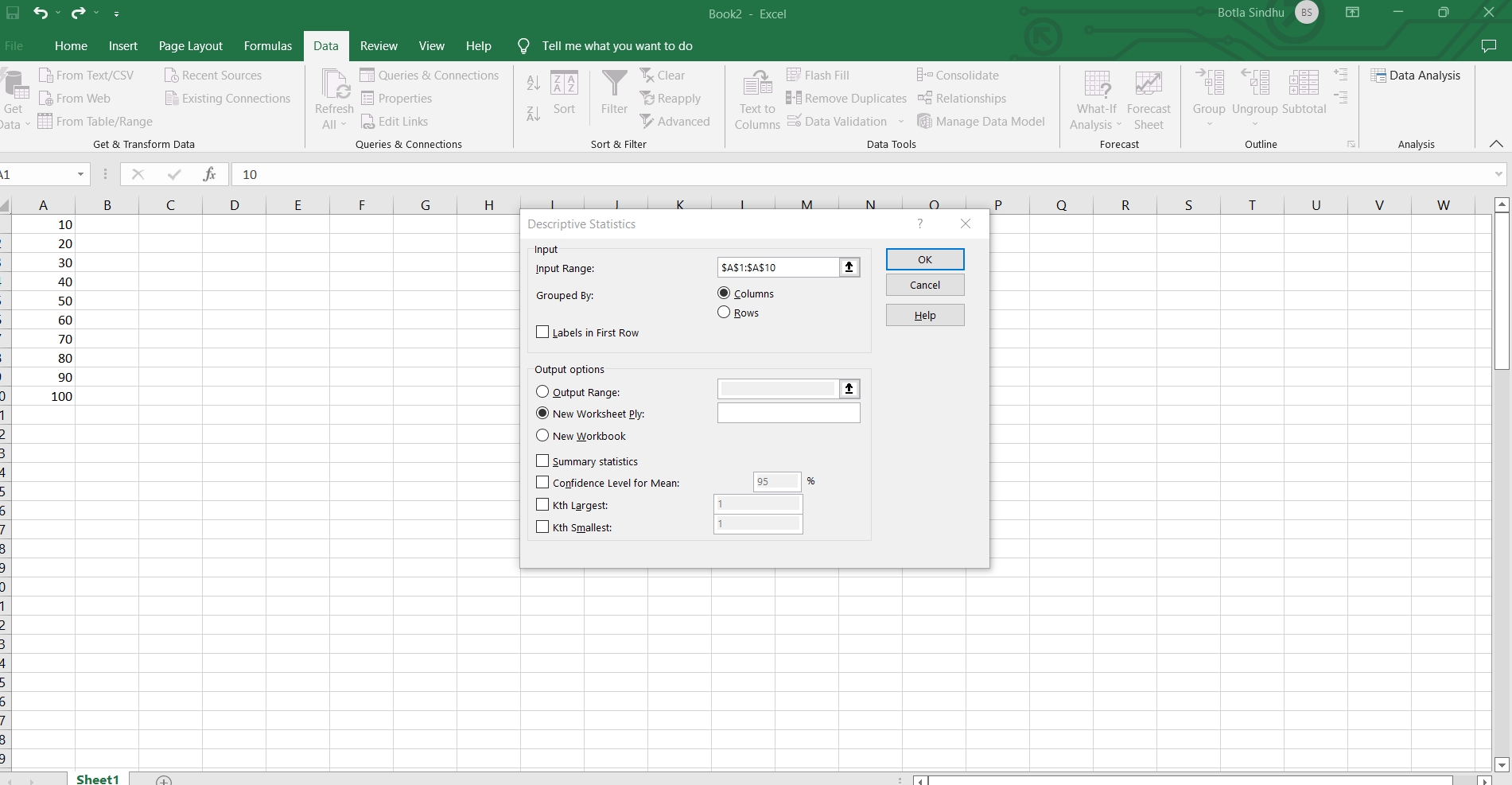
Step-3: Automatically numbers will be generated in column cells.

Then go to Data Header and select Data Analysis then a dialog box will appear in that select Descriptive Statistics and click OK.

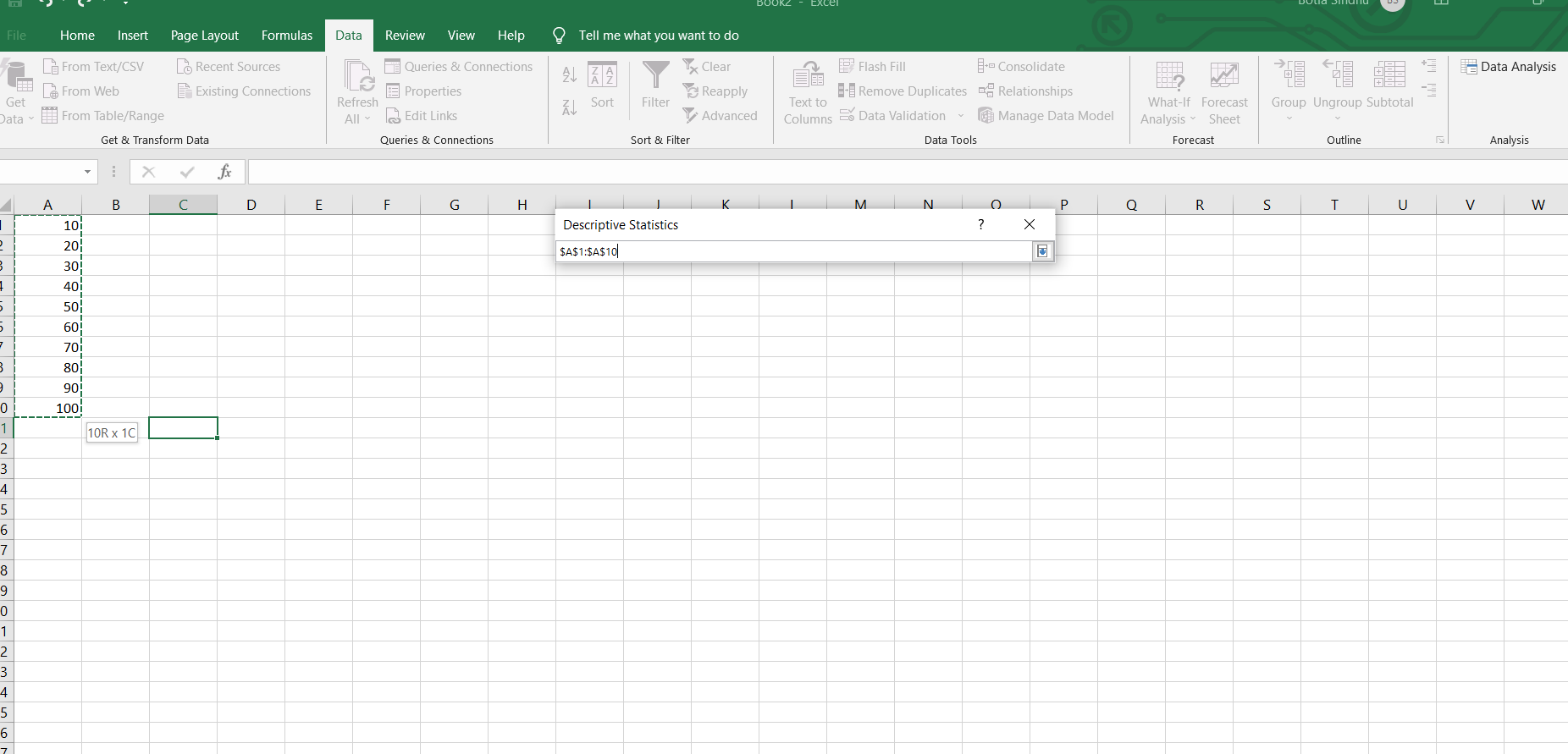


Step-4:

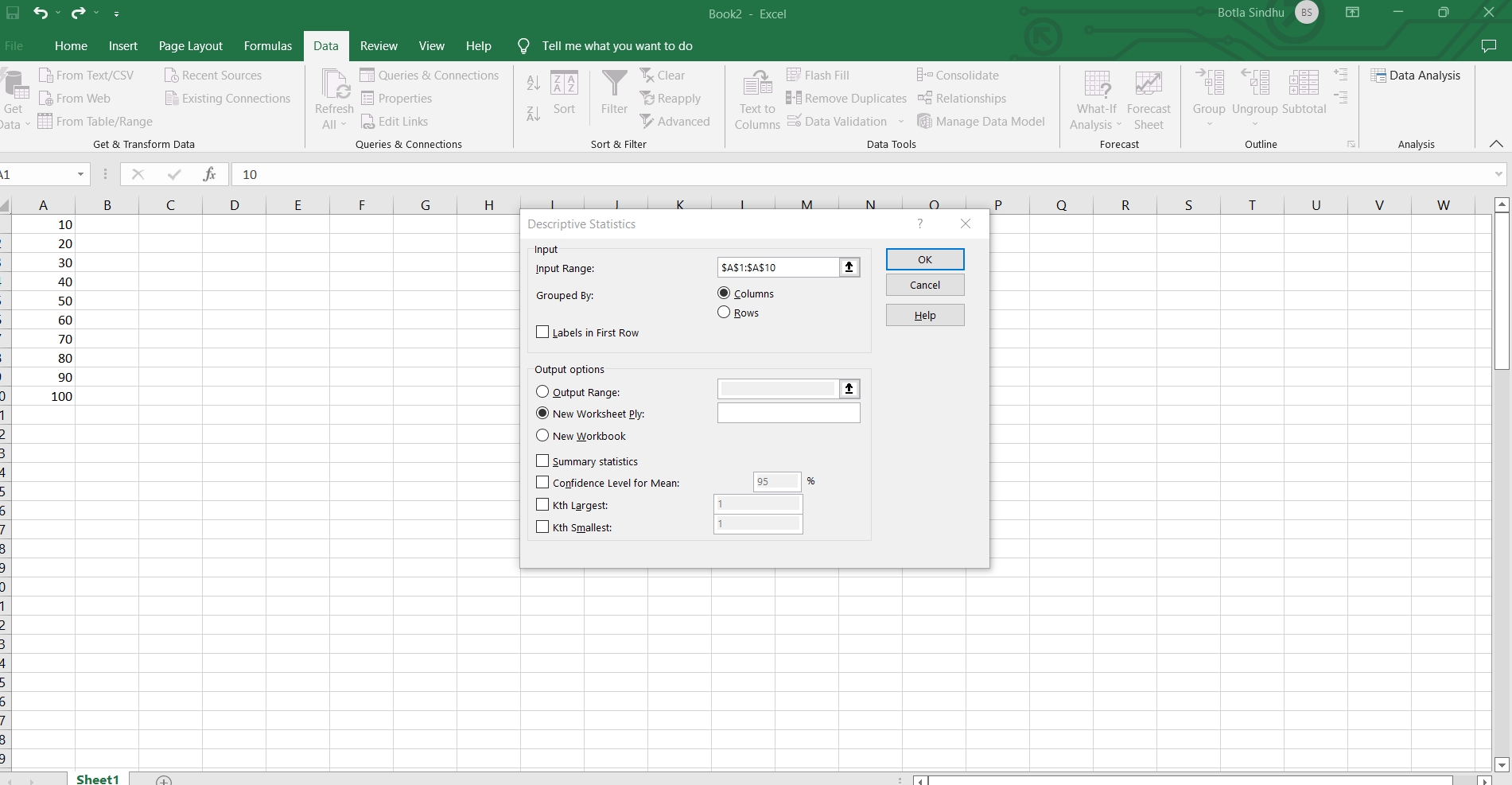
1. Descriptive Statistics dialog box will appear



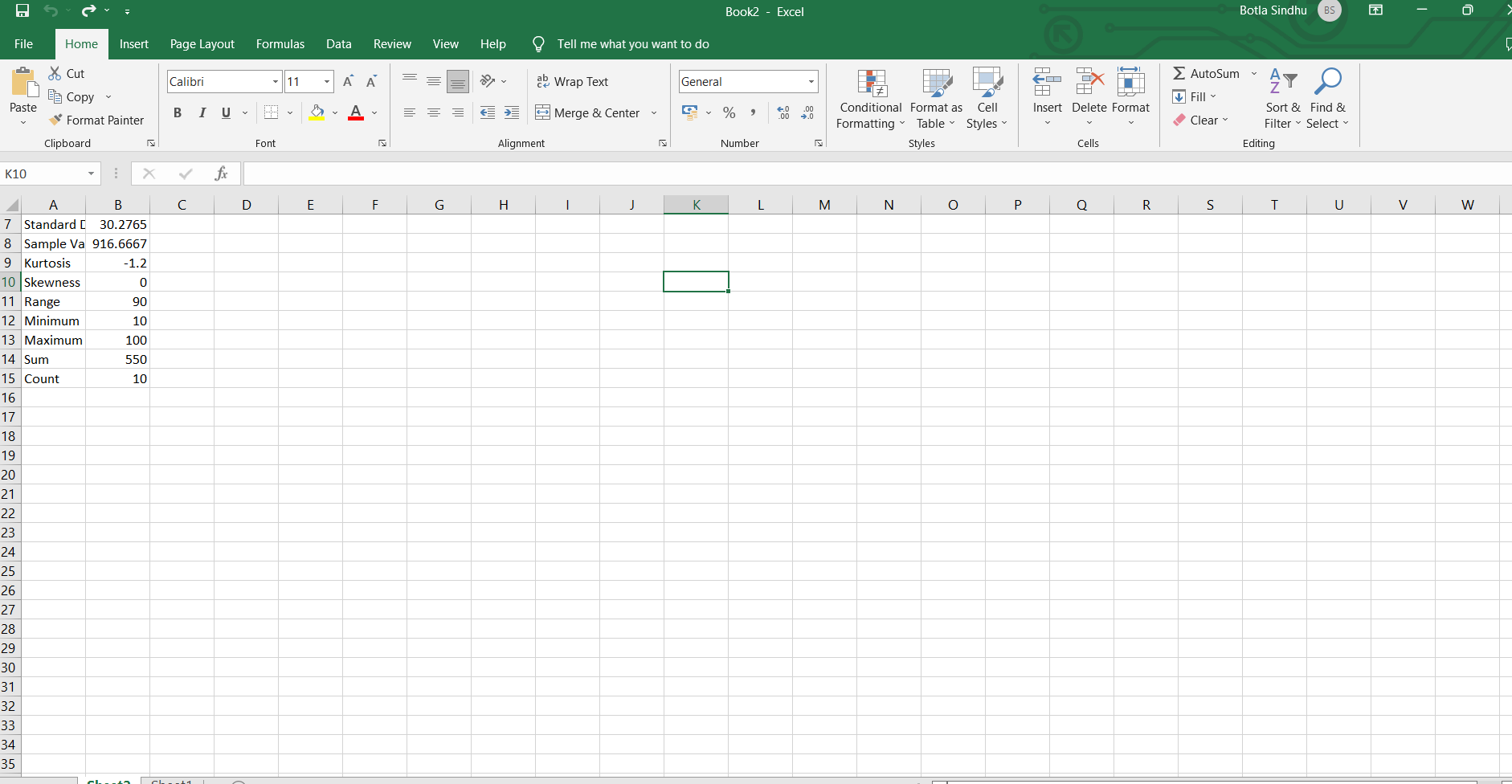
1. Select Input Range. A small dialog box will appear. Then drag the column till last filled cell.



1. Again Descriptive Statistics will appear in that select Summary statistics and then click OK.



1. Required Output will be generated.



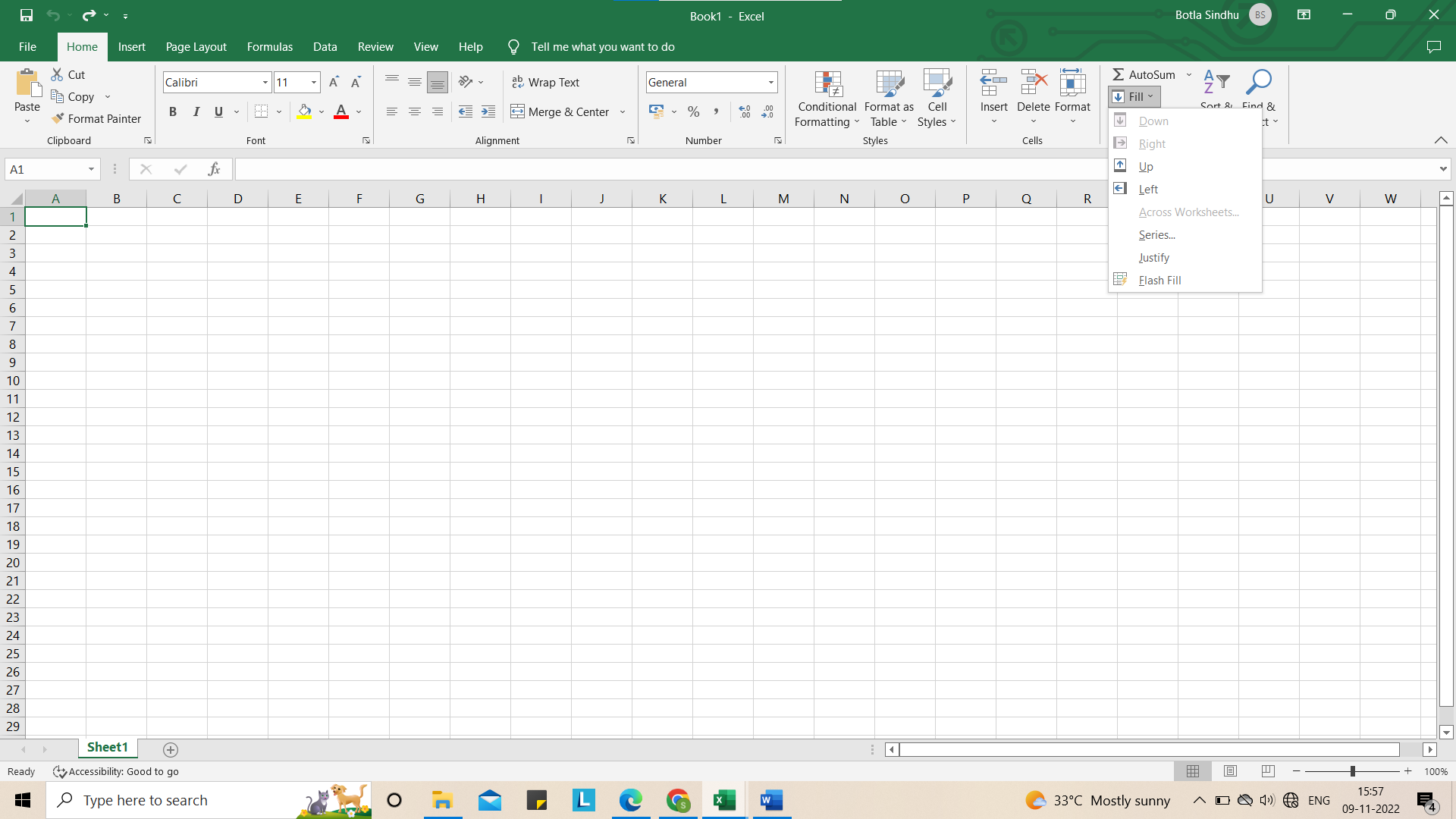
**REG NO:20B01A1220**

**DATE:23-08-22**

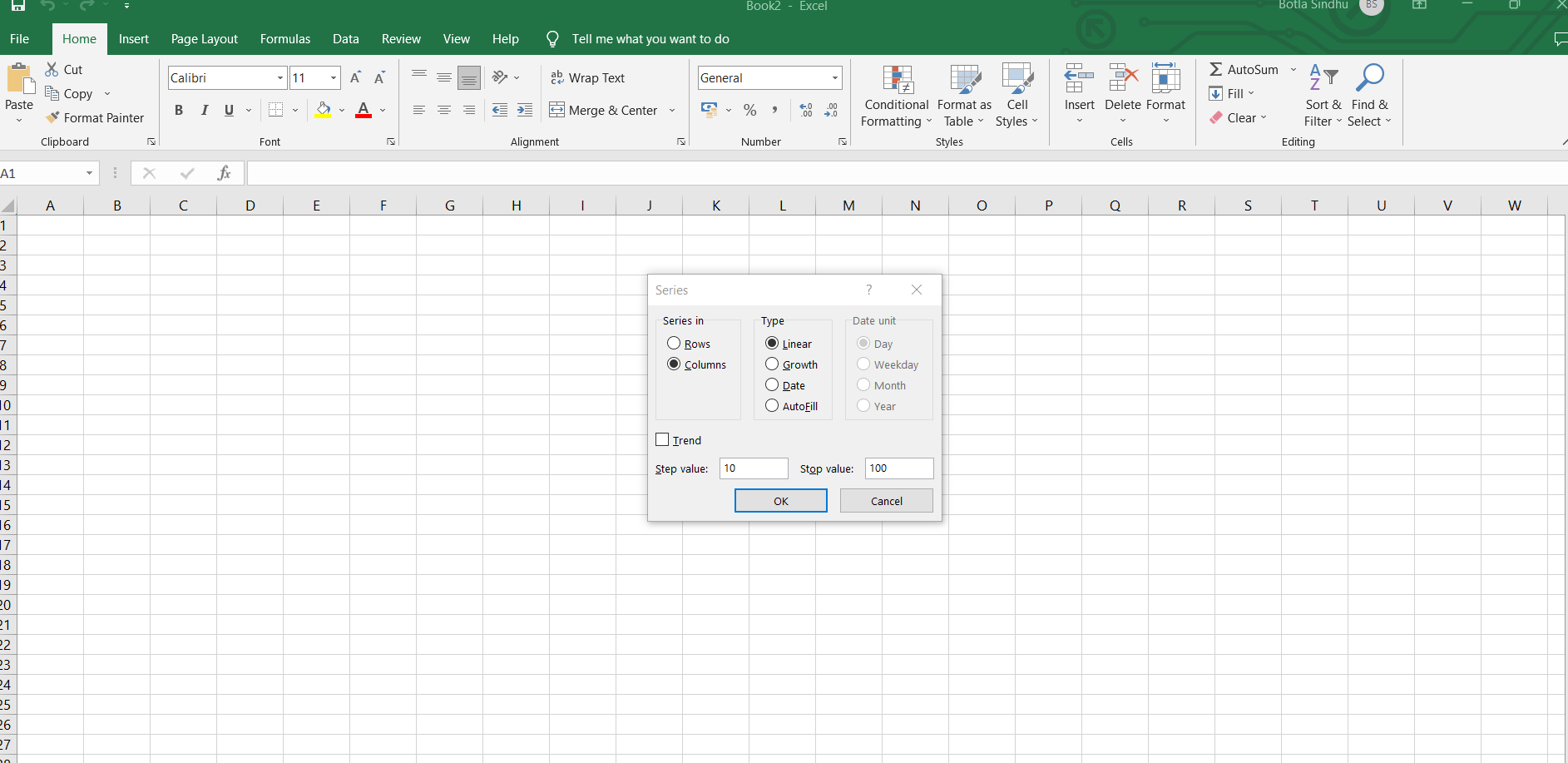
**EXPERIMENT NO: 02**

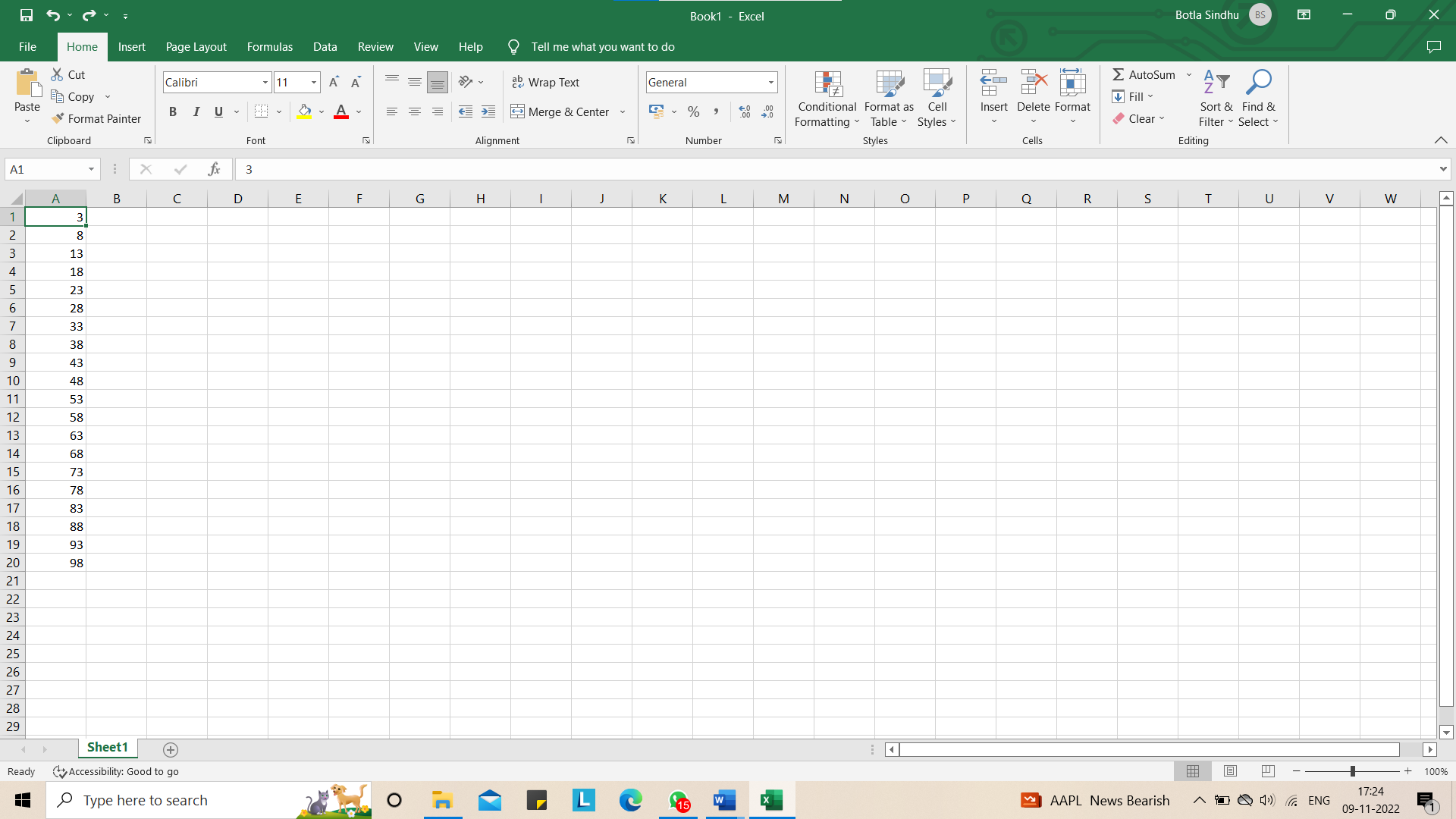
Perform the Histogram Analysis of given dataset using Data Analysis Toolbox of Excel.

PROCEDURE:

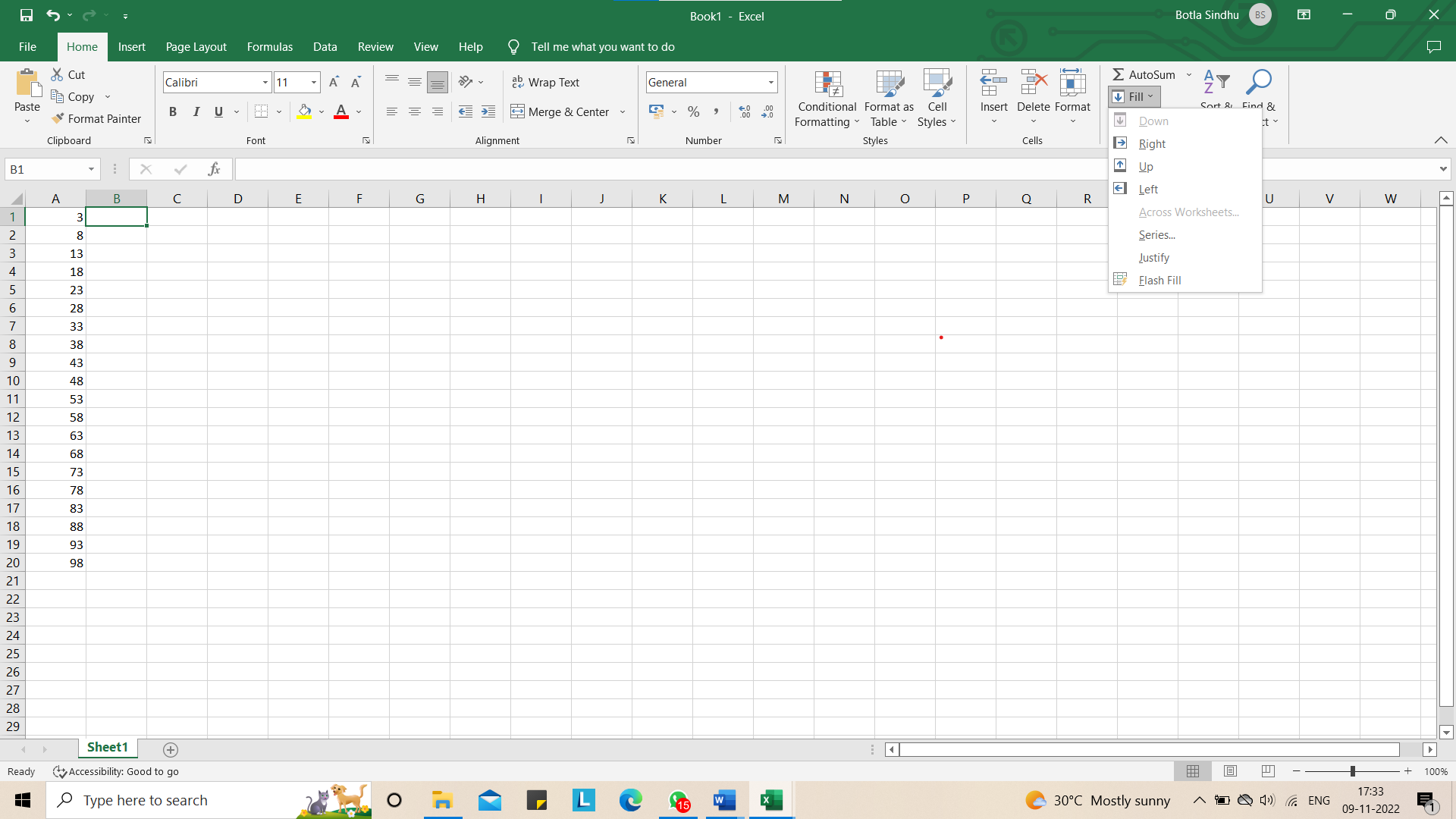
Step 1: Fill first cell with number which we want to start and then go to Home Menu Bar select fill option in Editing and then select series. This column is known as Input Range.

Step 2: Series dialog box will displayed. In that select Columns option and set type as Linear and required set step value and stop value and then click OK.

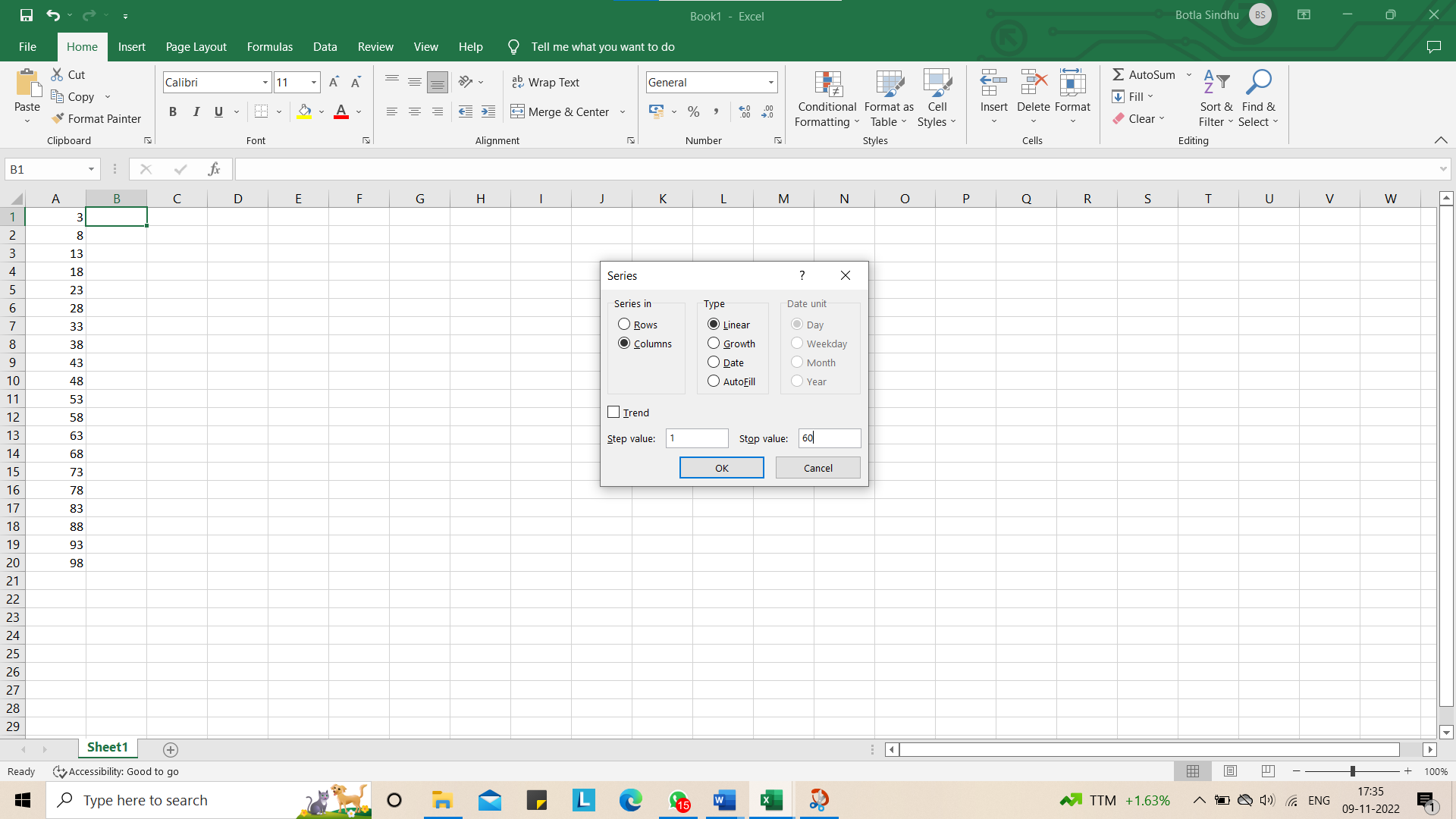
 Step 3: Automatically numbers will be generated in Input Range column cells



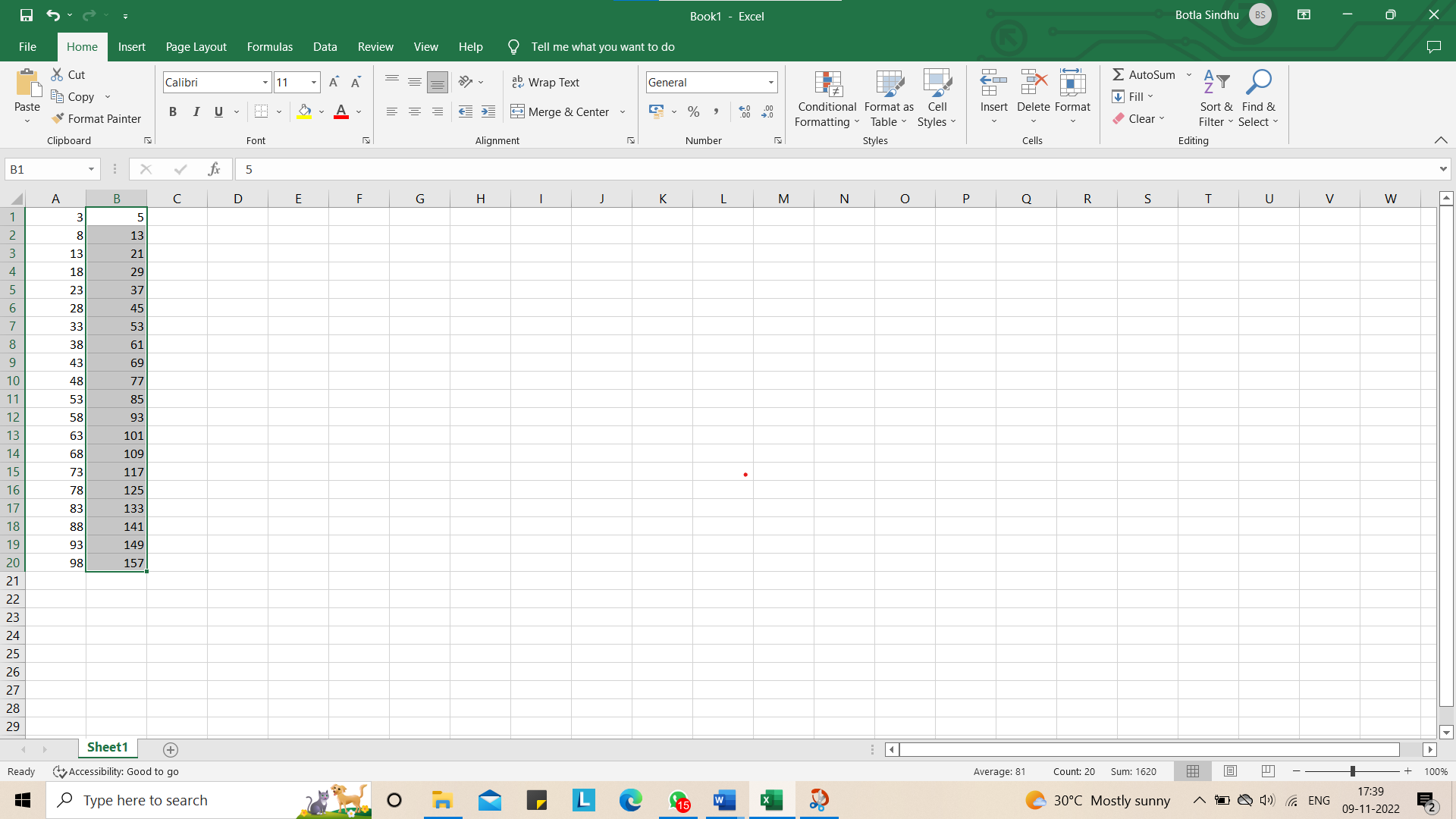
Step 4: Fill another column cell with number which we want to start and then go to Home Menu Bar select fill option in Editing and then select series. This column is known as Bin Range



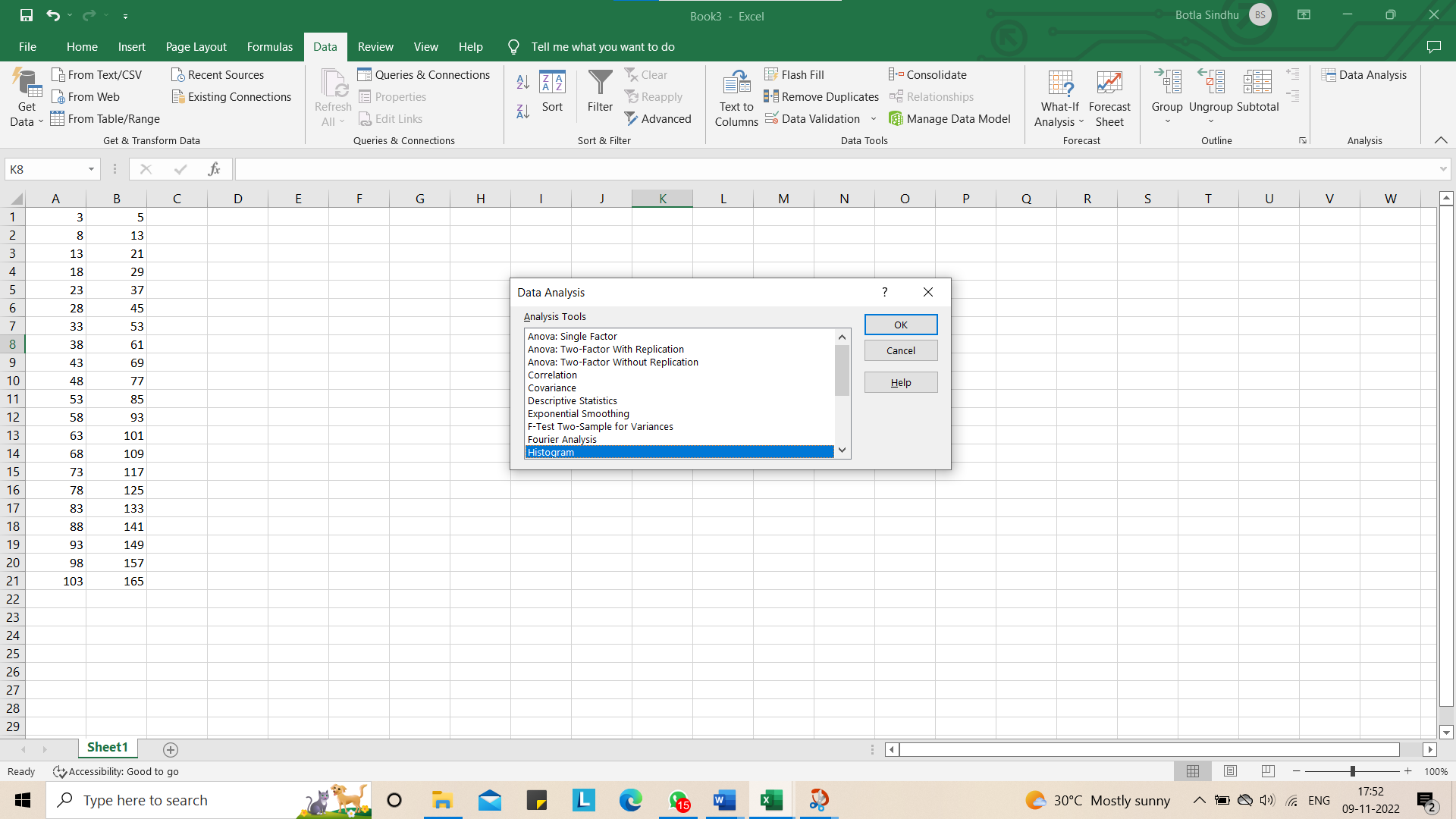
Step 5: Series dialog box will displayed. In that select Columns option and set type as Linear and required set step value and stop value and then click OK.



Step 6: Automatically numbers will be generated in Bin Range column cells

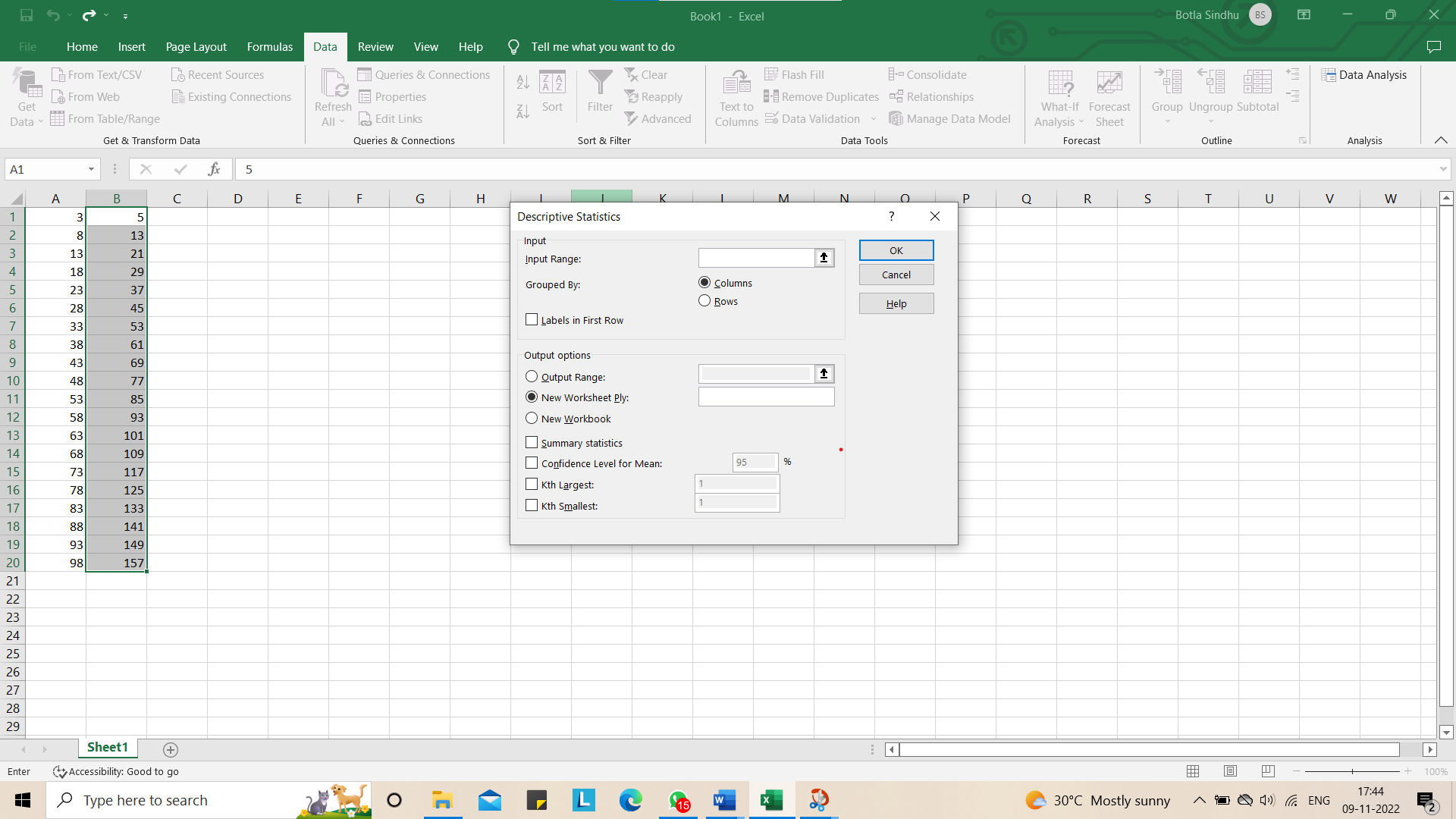


Then go to Data Header and select Data Analysis then Data Analysis dialog box will appear in that select Histogram and click OK.

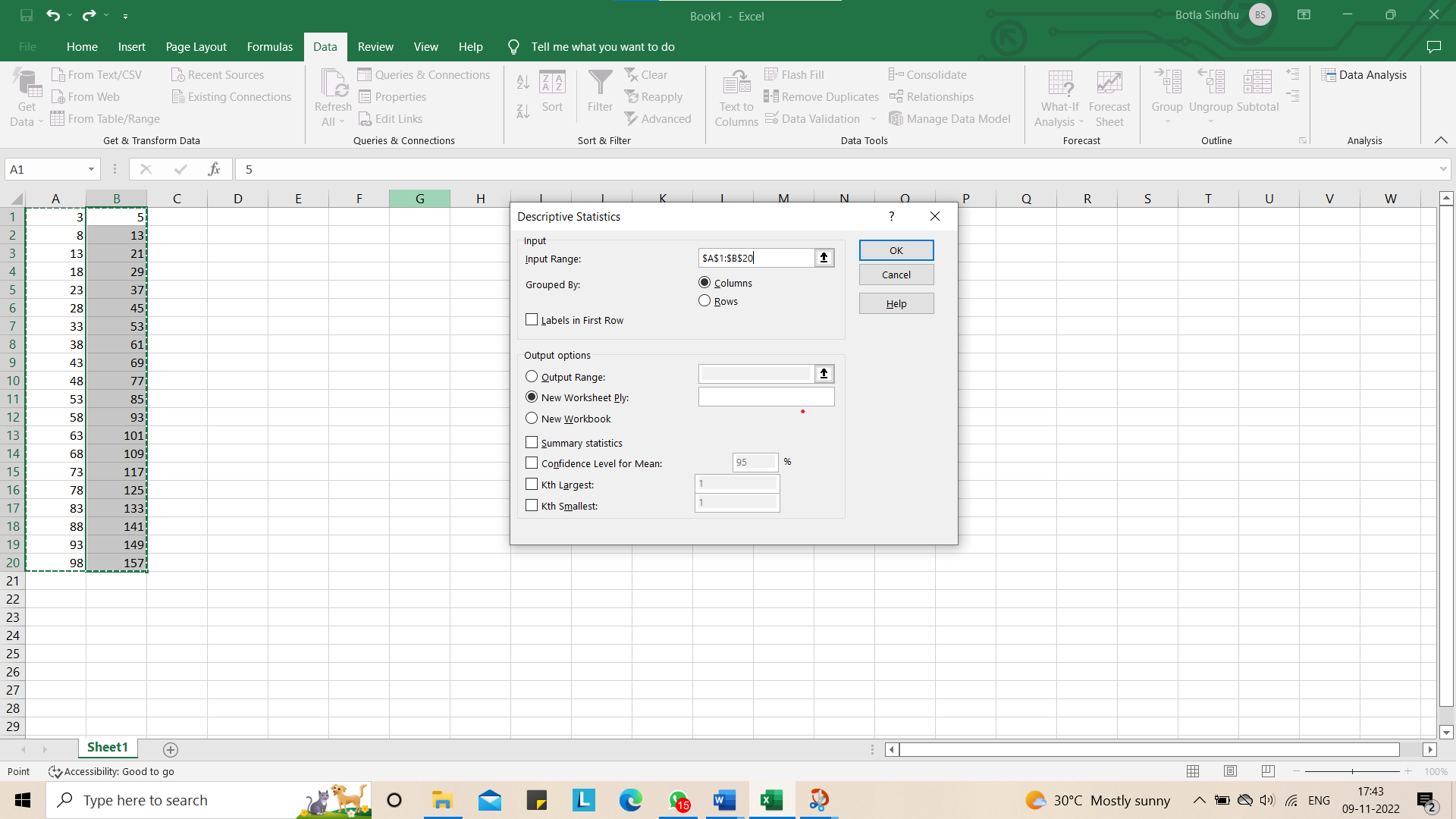


Step 7:

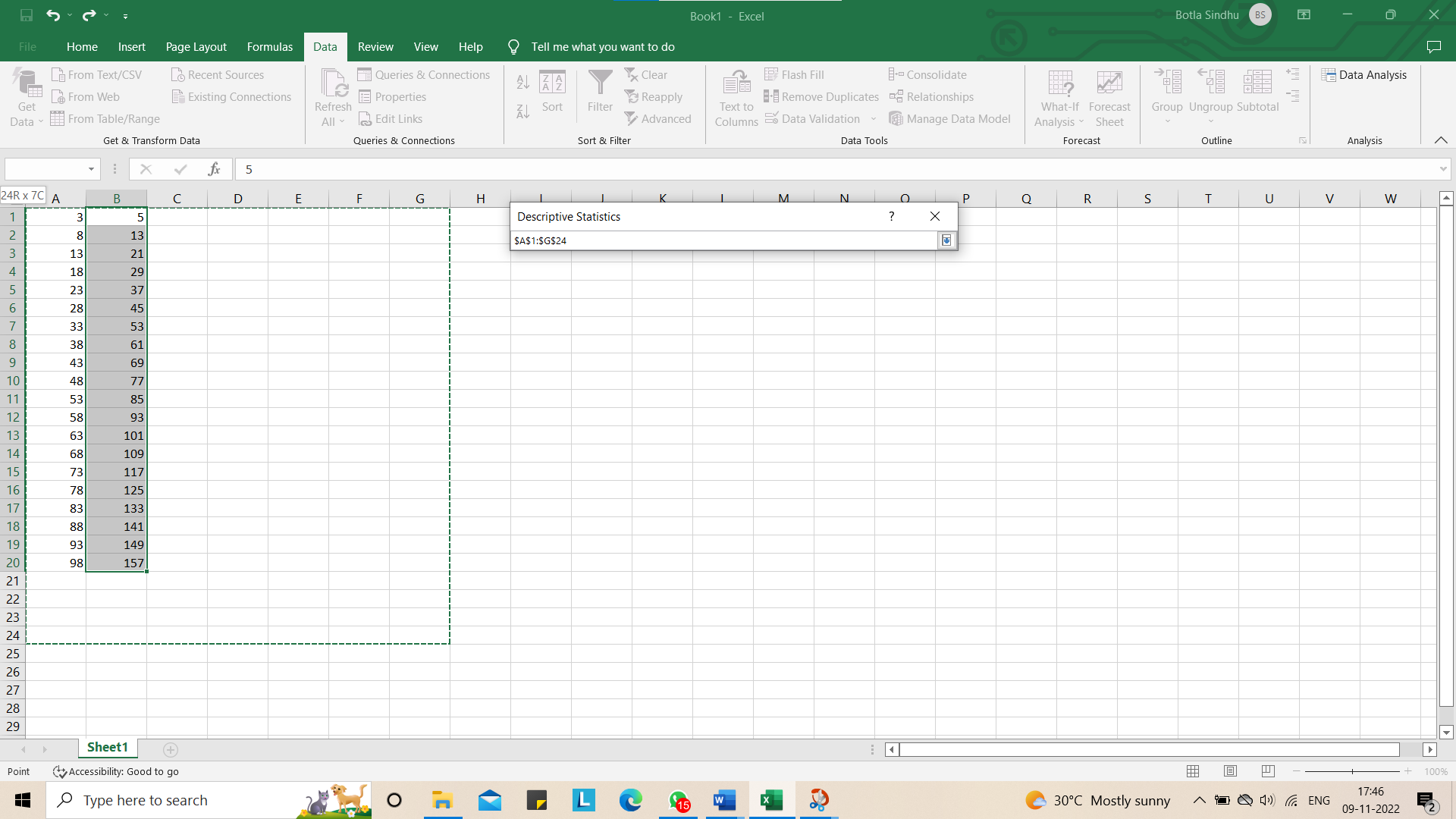
1. Histogram dialog box will appear



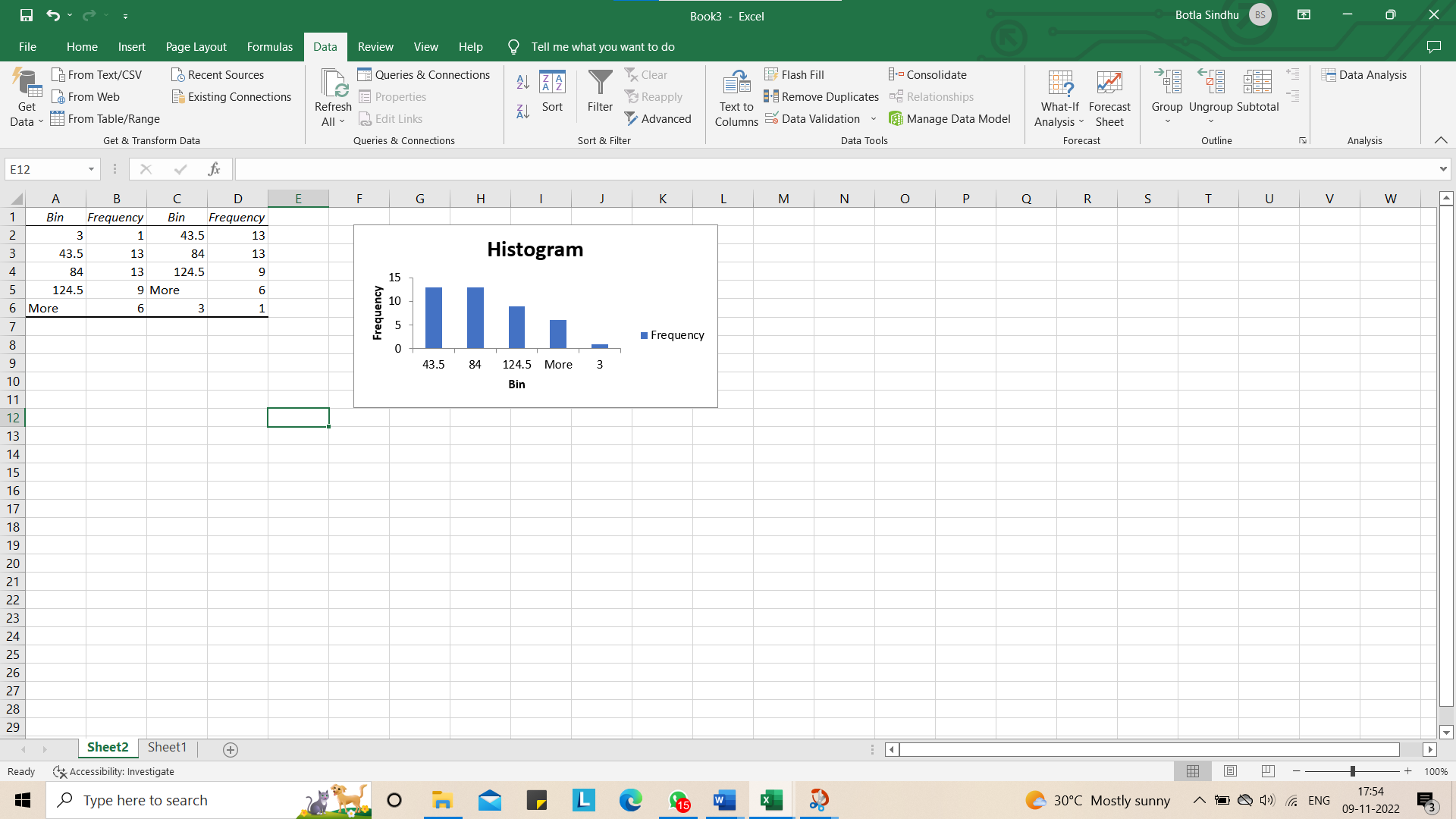
1. Select Input Range. A small dialog box will appear. Then drag the column till last filled cell.



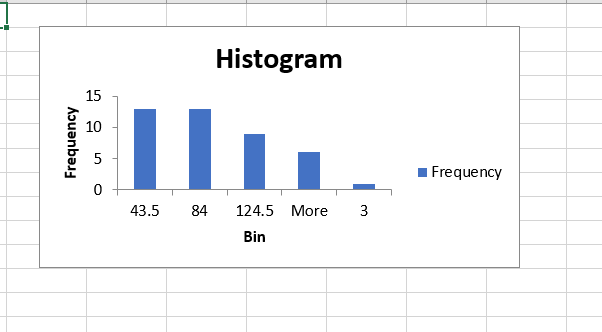
1. Select Bin Range. A small dialog box will appear. Then drag the column till last filled cell.



1. Again Histogram dialog box will appear in that select Pareto and Chart Output and then click OK.



1. Required Output will be generated.



In [1]:

*# REG NO:20B01A1220*

*# Botla Sindhu*

*# DATE:20/10/22*

*# EXPERIMENT NO: 08*

Write an application to apply regression model on the given data set

In [3]

**import** numpy **as** np

**import** pandas **as** pd

**import** matplotlib.pyplot **as** plt

dataset **=** pd.read\_csv('GPA\_SAT\_data.csv')

X **=** dataset.iloc[:, :**-**1].values y **=** dataset.iloc[:, 1].values print(X)

print(y)

[[2.714]

[3.418]

[4.325]

[2.552]

[2.634]

[3.352]

[2.758]

[4.231]

[2.975]

[3.839]

[3.788]

[2.133]

[1.136]

[3.053]

[4.074]

[3.515]

[1.98 ]

[2.788]

[3.681]

[4.141]

[3.13 ]

[2.311]

[2.421]

[1.631]

[3.472]

[1.37 ]

[2.498]

[3.89 ]

[3.952]

[2.254]

[3.193]

[2.835]

[3.287]

[3.587]]

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| [1480 | 1620 | 2070 | 1200 | 1300 | 1310 | 880 | 2230 1180 | 1900 1480 720 760 1490 |
| 1940 | 1400 | 1170 | 1410 | 1290 | 2390 | 1410 | 1090 1120 | 940 1460 980 1460 1880 |
| 2250 | 1480 | 1400 | 1390 | 1350 | 1850] |  |  |  |

In [7]:

**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

In [8]:

**from** sklearn.linear\_model **import** LinearRegression regressor **=** LinearRegression()

regressor.fit(X\_train, y\_train)

Out[8]:

LinearRegression(copy\_X=True, fit\_intercept=True, n\_jobs=1, normalize=False)

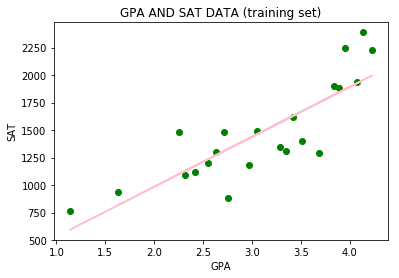
y\_pred **=** regressor.predict(X\_test)

In [10]:

plt.scatter(X\_train, y\_train, color **=** 'green')

plt.plot(X\_train, regressor.predict(X\_train), color **=** 'pink') plt.title('GPA AND SAT DATA (training set)')

plt.xlabel('GPA') plt.ylabel('SAT') plt.show()



In [11]:

plt.scatter(X\_test, y\_test, color **=** 'green')

plt.plot(X\_train, regressor.predict(X\_train), color **=** 'pink') plt.title('GPA AND SAT DATA (test set)')

plt.xlabel('GPA') plt.ylabel('SAT') plt.show()

