**Name**

**Hina manzoor**

**Roll no**

**17581556-095**

**Assignment**

**Data visualization**

**Submitted to**

**Sir Nauman**



**Data Visualization using python on Jupyter Notebook:**

**Write a program in python using matplotlib**

* First of all, you must have two software’s named python and jupyter installed in your system.



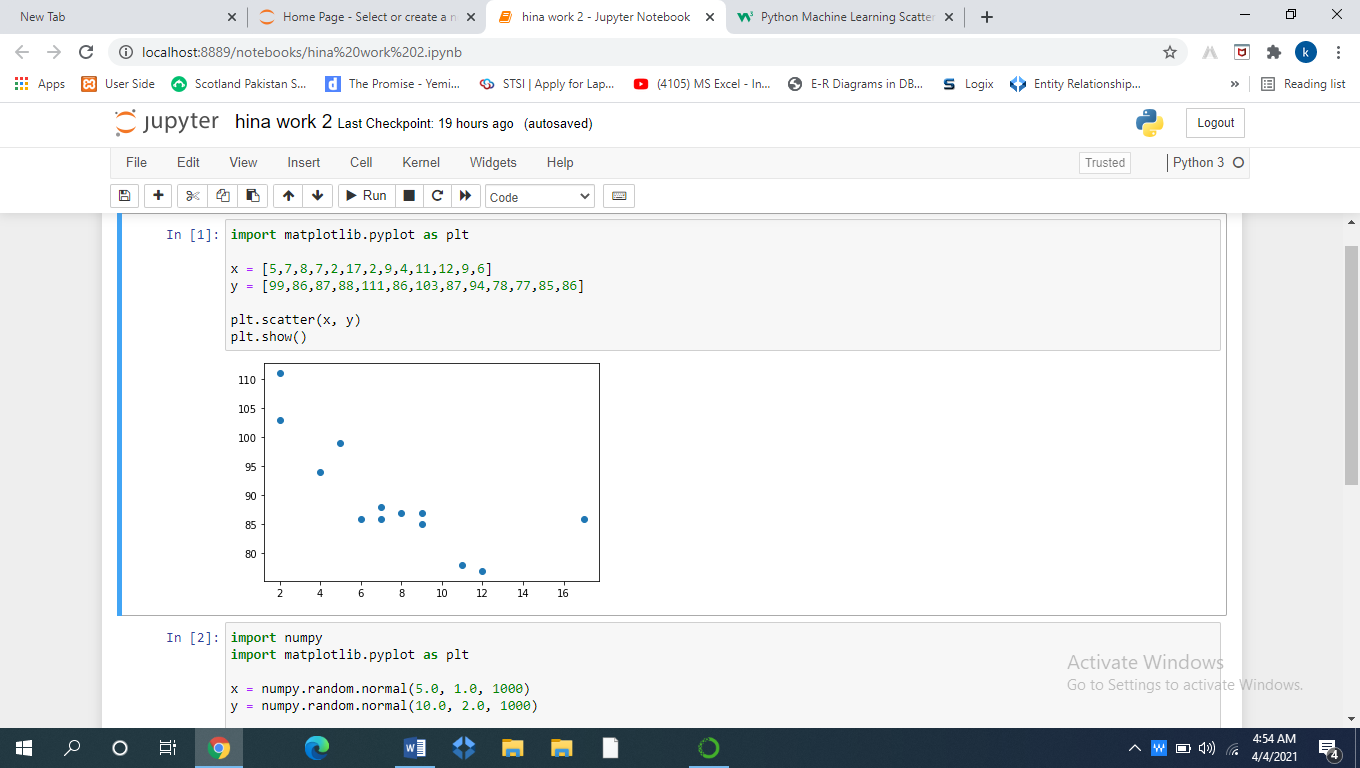


* Download a dataset from kaggle website(www.kaggle.com) in csv format (must be in csv format).
* Then make a workplace.
* Import some libraries of python to make a plot of given csv format dataset:
* **Import pandas as pd**
* **Import seaborn as sns**
* Then we set a color code to make dataset attractive**.**
* **Weather = pd.read\_csv(‘Book1.csv’)**
* **read** is a function of pandas that access the dataset fila nemed **Book1**.
* **Weather.head()** is used to sow dataset with headings.

**Working with so many graphs:**

**Scatter Plot**

A scatter plot is a diagram where each value in the data set is represented by a dot.



The Matplotlib module has a method for drawing scatter plots, it needs two arrays of the same length, one for the values of the x-axis, and one for the values of the y-axis.

## Random Data Distributions

In Machine Learning the data sets can contain thousands-, or even millions, of values.

You might not have real world data when you are testing an algorithm, you might have to use randomly generated values.

As we have learned in the previous chapter, the NumPy module can help us with that!

Let us create two arrays that are both filled with 1000 random numbers from a normal data distribution.

The first array will have the mean set to 5.0 with a standard deviation of 1.0.

The second array will have the mean set to 10.0 with a standard deviation of 2.0:

### **Example**

A scatter plot with 1000 dots:

