**Data Analysis of Human Development Index**

A Python Project

Submitted By:

1. Snehal Kale (BT20CSE170)

Introduction :

This project will analyse the development of various countries on the basis of Human Development Index (HDI) and determine the correlation between development and various factors.

**What is HDI ?**

* It is a tool used to measure a country’s overall achievement in its social and economic dimensions.
* ***Definition*** :

The Human Development Index (HDI) is composite statistic of life expectancy, education and per capita income indicators, which is used to rank countries into four tiers of human development.

* The human development index is a measure of economic development and economic welfare.

**What is correlation ?**

* Correlation refers to the statistical relationship between two entities.
* Positive correlation is a relationship between two variables in which both variables is in the same direction. A positive correlation exists when one variable decreases as the other variable decreases, or one variable increases while the other increases.
* Negative correlation is a relationship between two variables in which one variable increases as the other decreases, and vice versa.

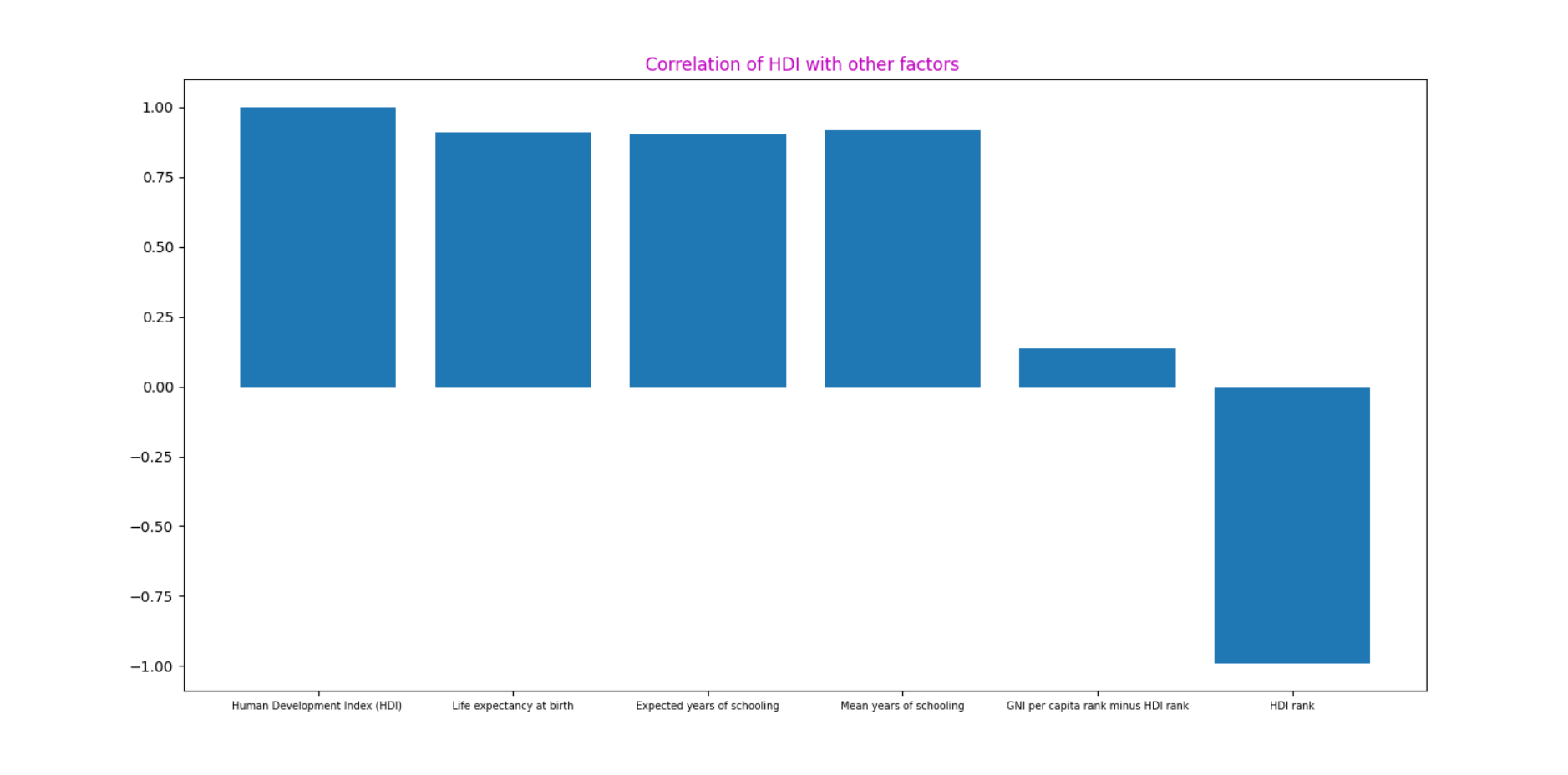
Project overview :

**1.Correlation of HDI to other factors in data set :**

Python libraries used :

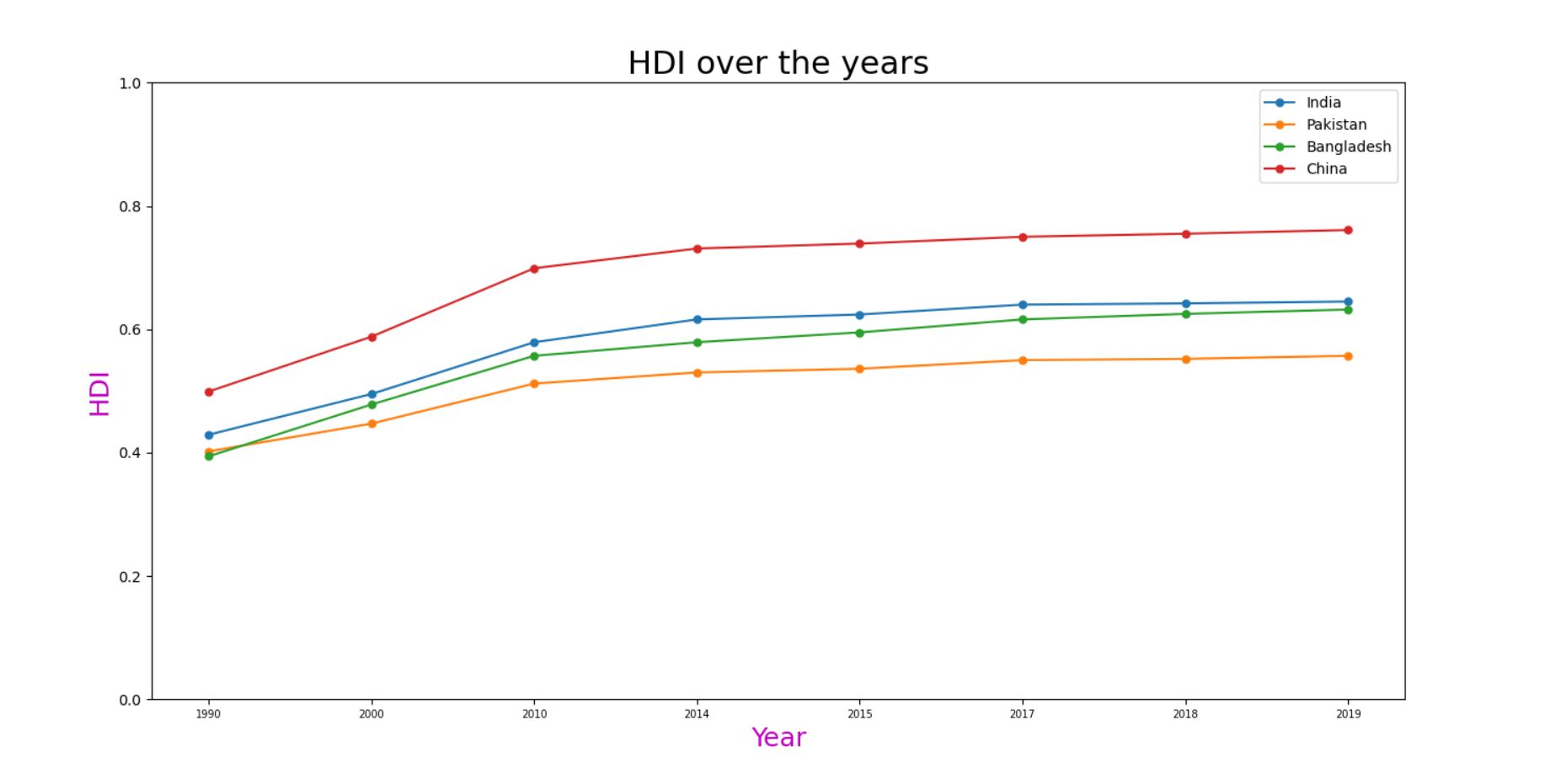
1. Pandas
2. Numpy
3. Matplotlib

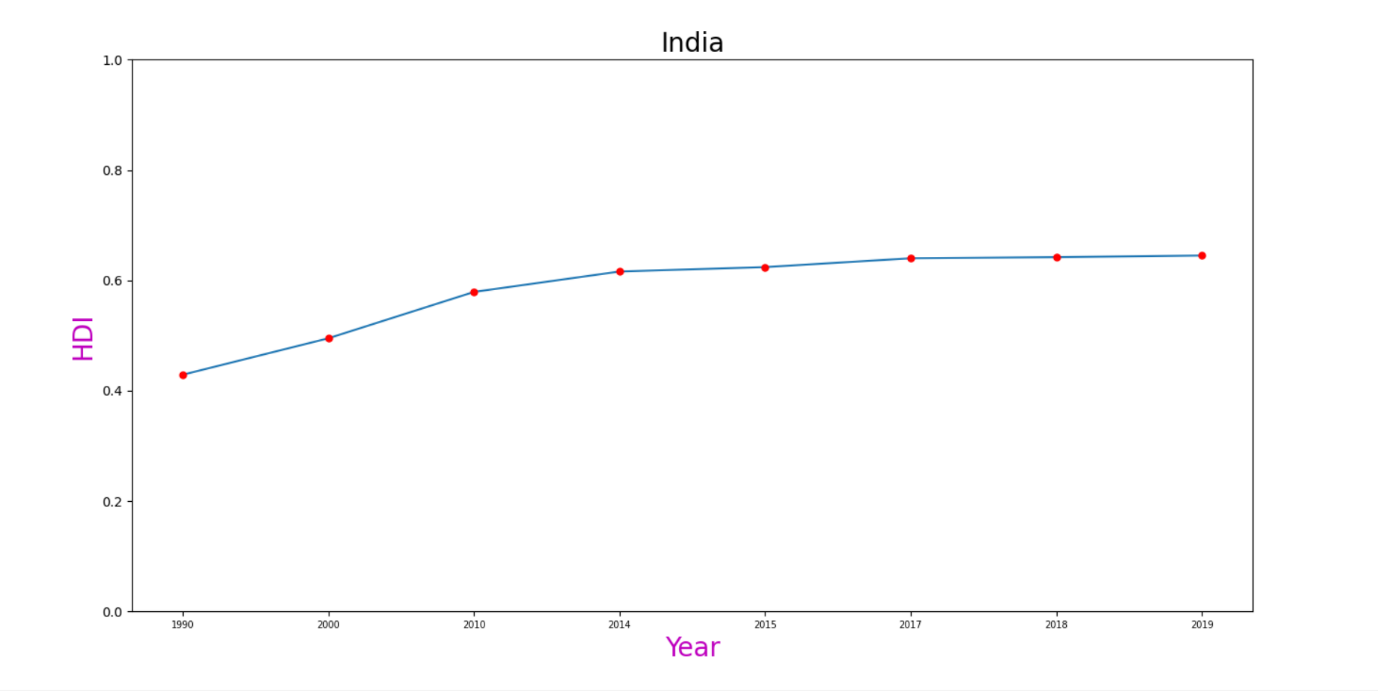
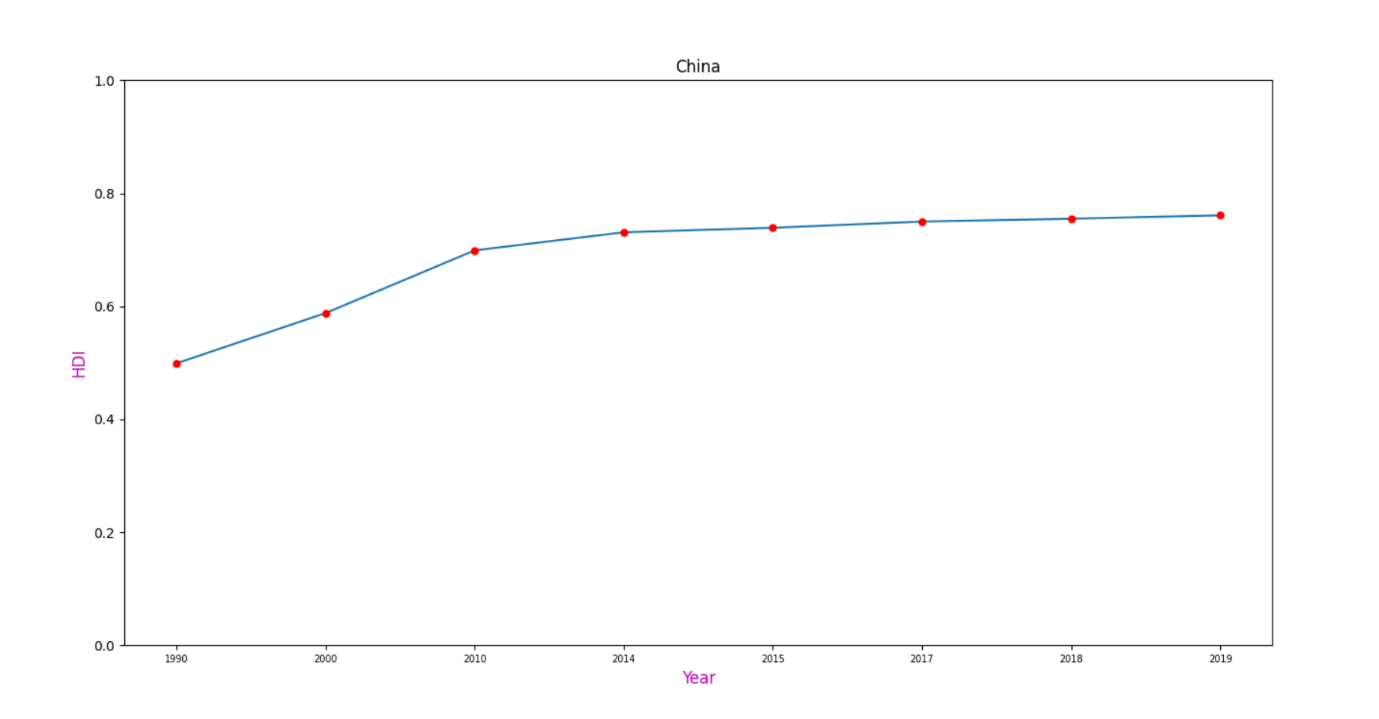
* The correlation of various factors in data set is taken by using **.corr()** function.
* To know the correlation of HDI with other factors graph is plotted using **pyplot** module.
* To plot the graph x co-ordinates and y co-ordinates are taken from **numpy.array()** function.
* X axis have all parameters in data set and Y axis have correlation values correspondingly in a list using **.tolist()** function.



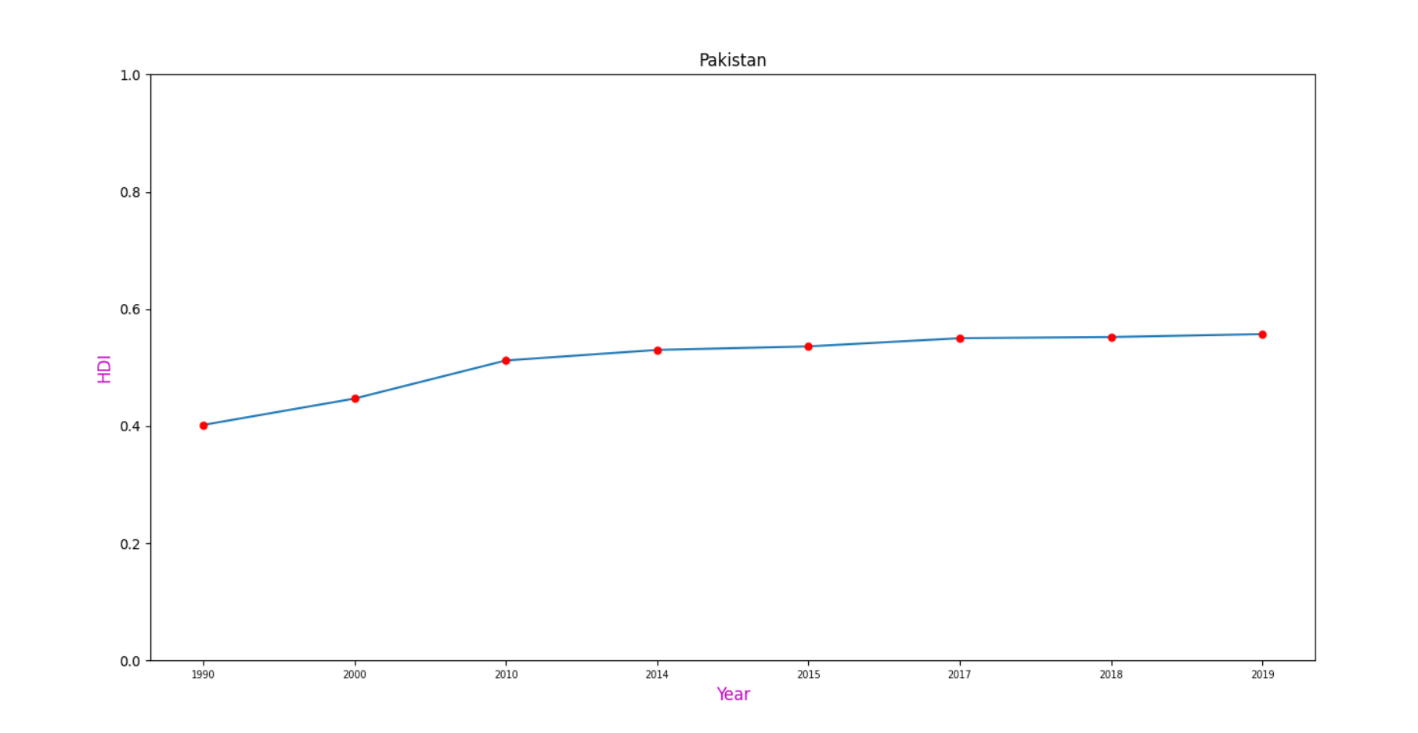
**2.Journey of all countries HDI over years** :

* The graph of various countries HDI over years is plotted using **pyplot** module.
* First to remove the missing values in data set **.dropna ()** function is used.
* From **.T** attribute numpy array is made .T attribute is the transpose of the array, s which can be conveniently separated in x and y parts by sequence unpacking.

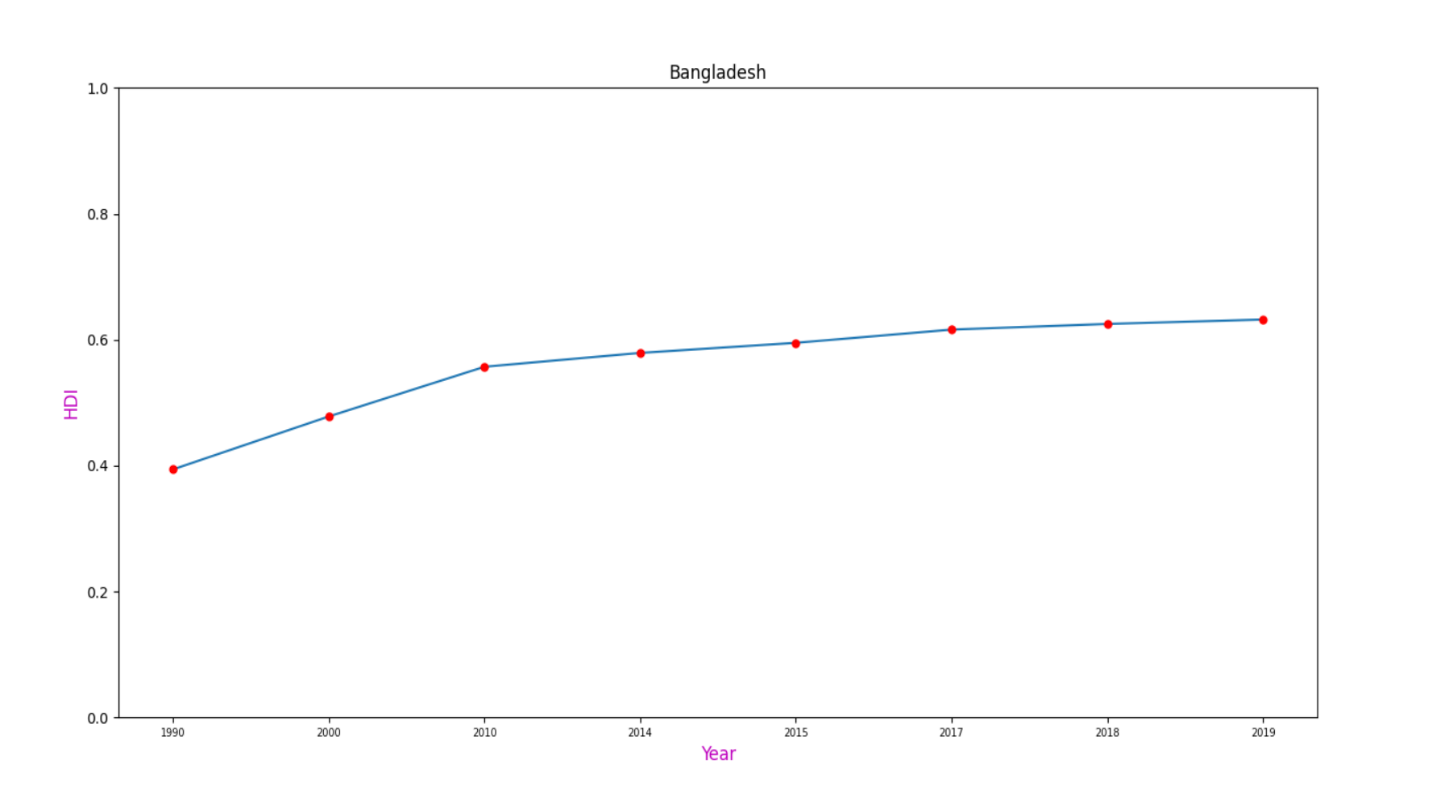


HDI of India over years :

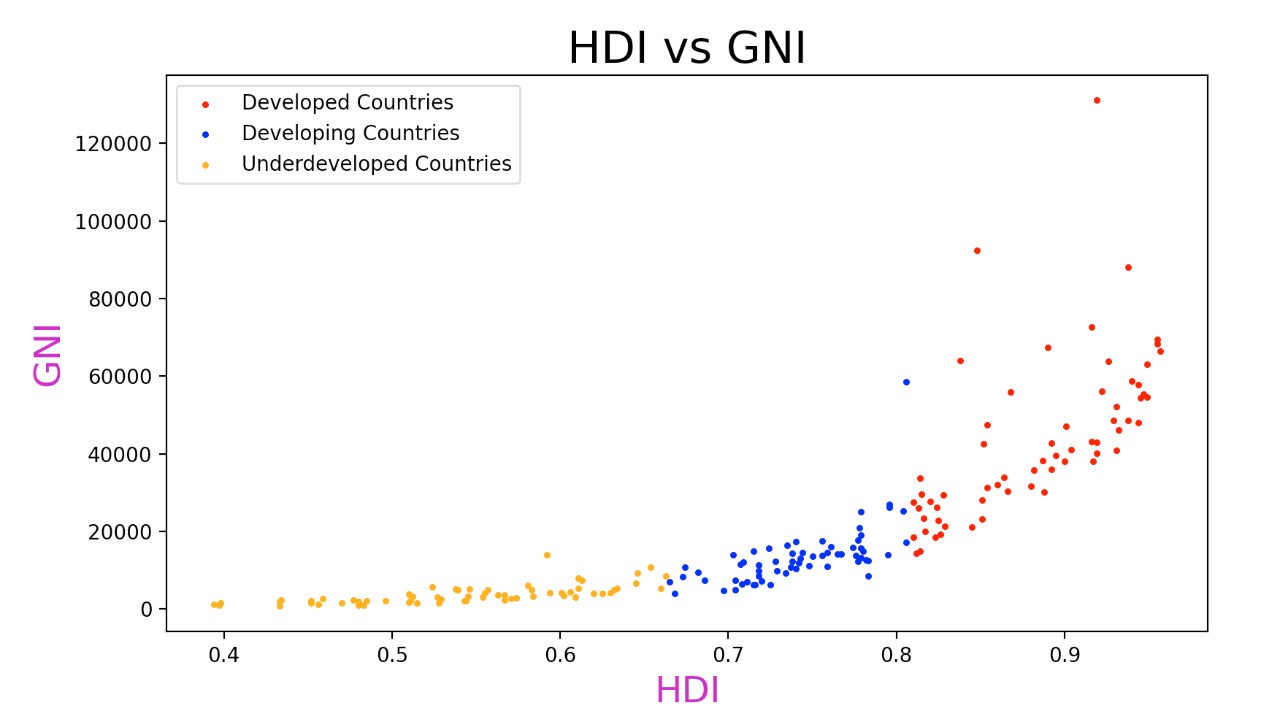
HDI of China over years :

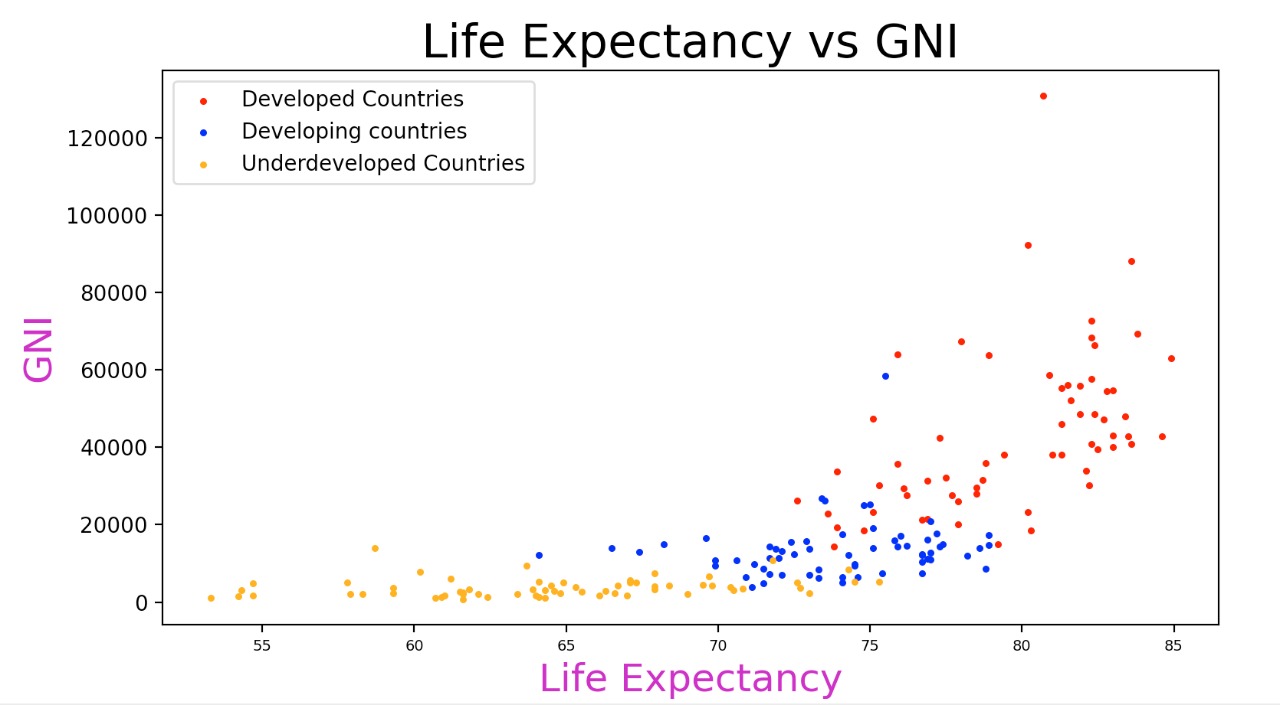
HDI of Pakistan over years :

HDI of Bangladesh over years :

**3.Scatter diagram between HDI and corresponding GNI of all countries :**

* For the scatter diagram HDI list is taken from data set by using **.tolist()** function and the array of this list is created using **numpy .array()** function
* This array is split into developed countries ,developing countries and underdeveloped counties by using **numpy.array\_split()** function.
* For the scatter diagram GNI list is taken from data set by using **.tolist()** fuction and the array of this list is made using **numpy .array()** function
* This array is split into developed countries ,developing countries and underdeveloped counties by using **numpy.array\_split()** function.





**4.Histogram of HDI , GNI per capita and life expectancy :**

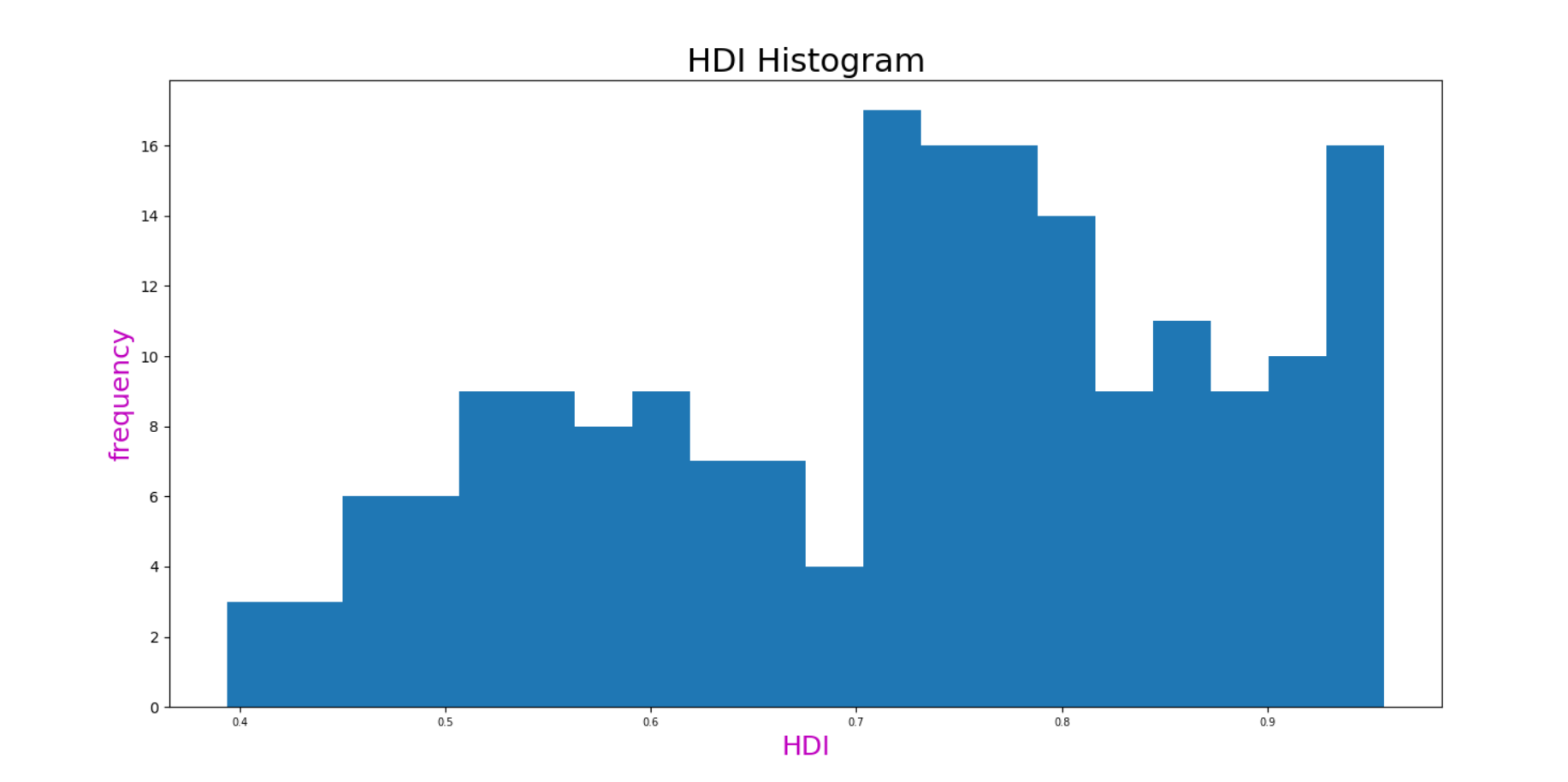
To plot the histogram of HDI , GNI and life expectancy values from data set are taken using **numpy.array()** function.

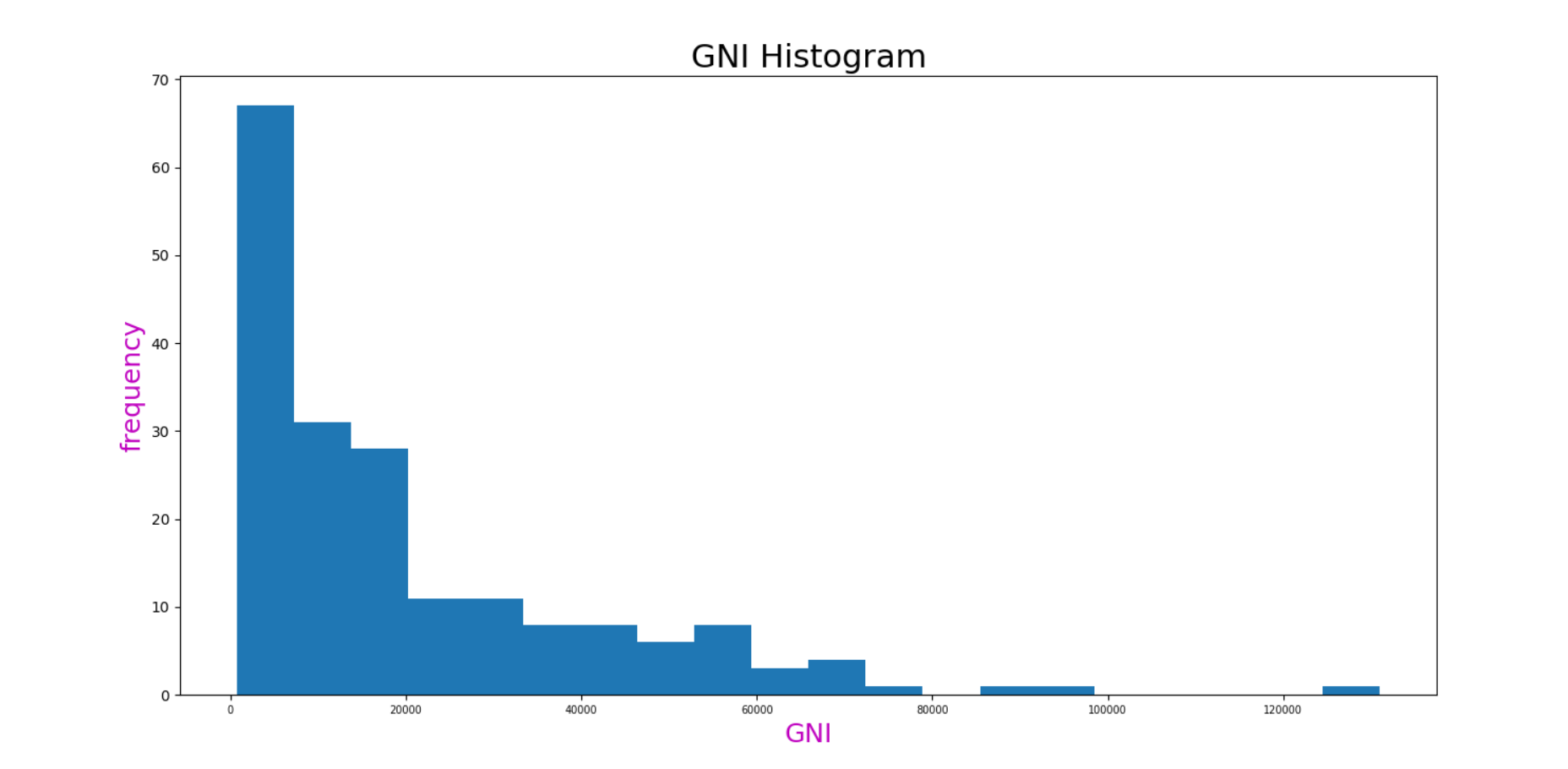
To calculate standard deviation of corresponding parameters **numpy.std()** function is used.

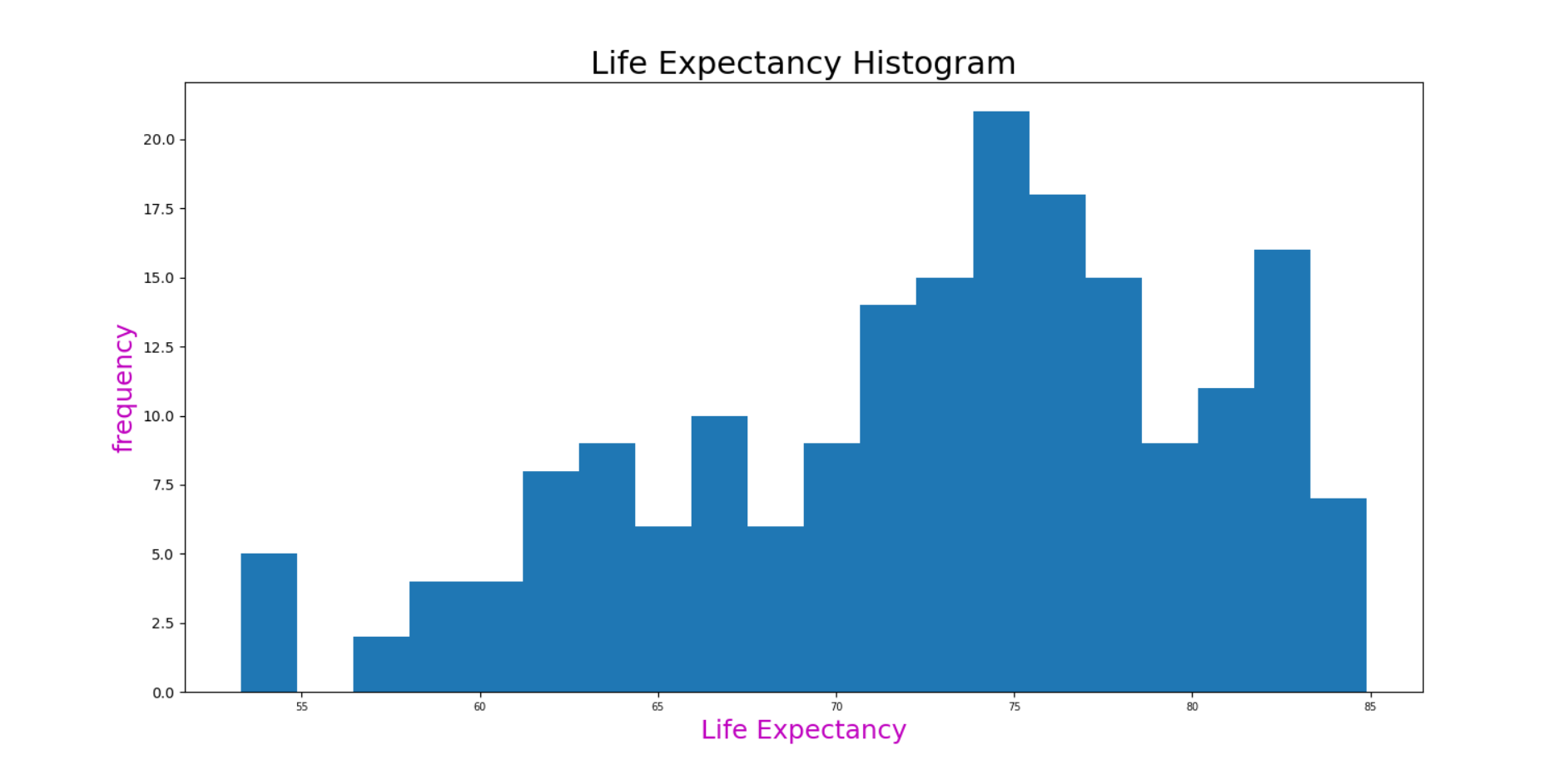
To calculate mean of corresponding parameters **numpy.mean()** function is used.

**.astype()** function is used to take specified int ,float values of respective parameters.

Histogram is plotted using pyplot modules **py.hist()** function.







**5.Other python modules and functions used in project :**

A) File handling and exception:

* **User defined exception : To take valid country name as input from user to display its HDI parameters .valid name of country in data set does not contain any numeric digit and the first letter of name is always capital letter. If user enter a name containing any digit or first letter as lowercase then invalid\_name exception is raised .islower() and .isdigit() functions are used to check first letter is capital or not and any numeric digit is present or not.**
* **Name error : If any variable is not defined in code then to handle this error name error exception is defined.**
* **Import error : while including modules in program import statement is used if the specified module is not written properly or imported module is not found in python library then interpreter throws import error and the remaining program won’t run to avoid this import error exception is used .**

B ) Object Oriented Programming :

4 classes have been created to display all types graphs shown above and a function is also created which display the hdi parameter details of the inputed country.

**$**

Conclusion :

* From the graph of correlation of HDI with other factors we can conclude that life expectancy rate , mean years of schooling , expected years of schooling are correlated with HDI
* From the graph of journey of different countries HDI over years we can say that Bangladesh and China has a better HDI growth than India.
* From the scatter diagram between HDI and corresponding GNI of all countries in data set we can say that HDI and GNI are less correlated.

**DATA SETS USED:**

**Book1.csv & Book2.csv (attached in zip folder)**