# Data science assignment 2.

**By HAMZA RAQEEB. 21031287**

## INTRODUCTION.

This assignment is based on data visualization tools and some statistical properties for data analysis. Here, we are using the library like NumPy pandas and mat plots for data visualization, and the data sets are downloaded from the official websites of the World Bank. Here we use the indicators like population growth, agricultural land, and the Urban population in the selected countries. Are different it will show in the code visualization tool are pie chart frequency graph and histogram here, we have two data frames one is for the years record and the second one is for the countries different docstrings are used to describe the written code.

## Statistical properties.

Some statistical properties are used in the code to describe the data set. These are a built-in function of the panda's library. These are shape, types and description. There are many different statistical properties for analyzing the dataset, but we used these basic properties for this assignment.

```
shape of dataset is  (20216, 65)
```

```
description of dataset
               1960          1961
count  3.140000e+03  4.546000e+03
mean   2.015940e+07  1.476890e+07
std    1.766754e+08  1.579775e+08
min   -3.667000e+00 -3.667000e+00
25%    1.658243e+00  2.727489e+00
50%    4.214050e+01  3.036830e+01
75%    6.640735e+03  1.764865e+03
max    3.040000e+09  3.428000e+09
```

```
data types of each feature

Country Name        object
Country Code        object
Indicator Name      object
Indicator Code      object
1960               float64
```

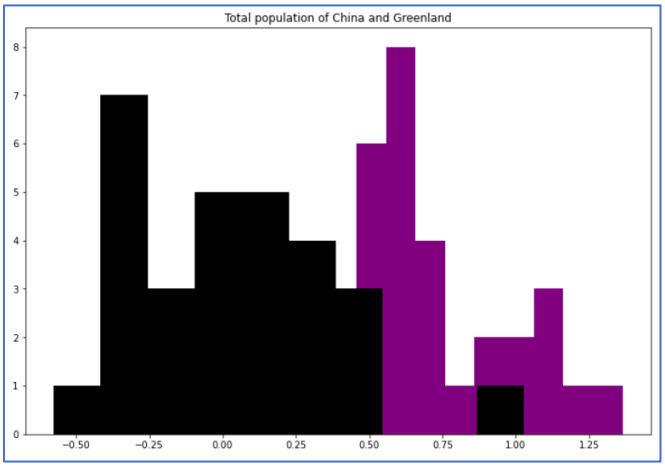
## Indicators.

Here we selected three indicators. One is population growth, and here we used the two countries to make the relationship between the two countries that is China and Greenland. The cycle indicator is the agricultural land selected country is Pakistan, and the third indicator is the urban population, which is only 2017. The selected countries are Armenia, Aruba, China, Spain, Switzerland and Bangladesh. Now we use these indicators for different types of visualization.

## Data visualization 1.

Indicator name: population growth.

Visualization tool: histogram.



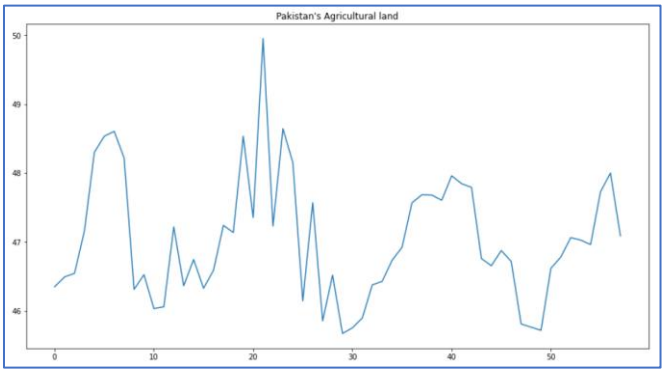

## Results.

The first data visualization technique is a histogram, one of the two types of histograms. One is for the population growth of China and the second one is for Greenland, so the blue describes the China and orange describes the Greenland population growth. At the start, the population growth of Greenland was more than China. After 2001 or 2002, the population of China suddenly increased more than Greenland because it's a big country, and it increased to a higher level if the population growth is more than Greenland after increasing population China. China started working to control the population, and now it is in a controlled situation that is about 1 % almost.

## Data visualization 2.

Indicator name: agricultural land.

Visualization tool: line graph.



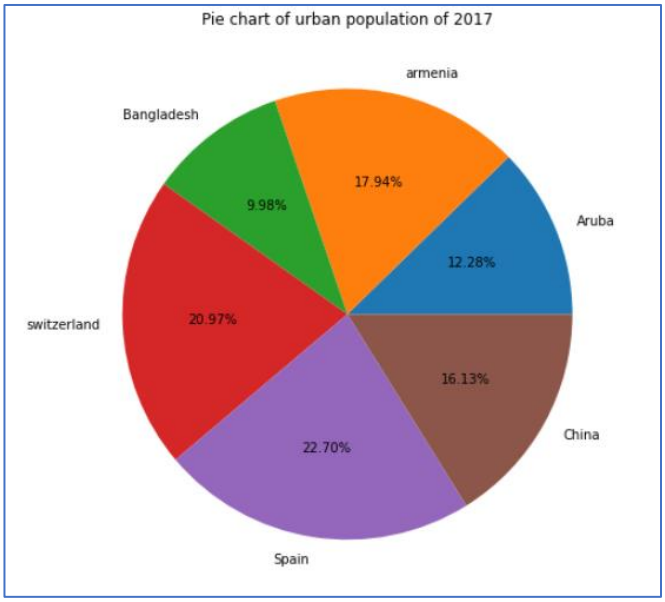
**Results.**

The country is Pakistan, and the selected indicator is agricultural land. Pakistan is situated in Asia, so this is good according to the given results at the start after independence. Pakistan's hard work increased agricultural land and made more fertile land for agricultural growth. After that, due to some political reasons, agricultural land is decreasing. From 1990 to 2003 the agricultural land is on the peak after 2005 it suddenly decreased to very low level and increasing step by step still 2013 from 2013 to 2019 it again increasing due to covid-19 now its decreased.

## Data visualization 3.

Indicator name: urban population.

Visualization tool: pie chart.




**Results.**

Have we selected the different countries for plotting the Urban population of 2017 urban population of Armenia is 17%, is Aruba is 12%, 16% for China, 22% for Spain, 20% for Switzerland, and 9% for Bangladesh, the top most urban country in Spain in 2017 where is a 22% increase in the Urban population after that the second urban populated country is Switzerland third one is China force one is Armenia then Aruba. Then Bangladesh, so this is the record of 2017 and here using the pie chart to distribute all the values.

## Conclusion.

In this assignment, we use a PANDAS library for the data framing, the data visualization that is matplotlib, and some function of the NumPy for deleting the null values in the records. We also learn some statistical properties of the data set and how we describe the data set. Who Learn Data visualization in a very clean way. And one more interesting thing is how to use a world bank data set according to the countries and different indicators.

## Git hub link.

https://github.com/HamzaRaqeeb/data-science-assignment-2