Pakistan Aisa Kiun Hai !

Submitted By: Muhammad Ali Raza
Submitted To: Dr. Junaid Akhtar

Abstract

The purpose of writing the report is to elaborate the semester project given to us known as “Pakistan Aisa Kiun hai !” .

In this project we have to explore any country given to us in all the prospectives , socially , economically , politically , socio-culturally etc.

We were given the data in which there were some factors (qualities)

throughout many years . Some of them were of ‘good-qualities’ and some were of bad. These factors were given the value according to given year. By using these indicators(quality-factors), we have to know that “Pakistan Aisa Kiun Hai !”

Table of Contents:

1.Introduction:

There are many ways to check the condition of any country or state. Ranking can be done separately depending upon the years, so that rank of the country can be diagnosed depending upon the year it is called, But it will not give you the Overall rank of that specific country nor will show the ups and downs of that country. Keeping this point in mind , the ranking can also be done by taking the overall position of that specific country or state. This will eliminate the problem of not knowing ups and downs regarding to the state or country. There are indicators (properties) describing the value of that particular indicator based to the country given. Among the indicators , there are some good indicators and bad indicators. For instance , Murder is the indicator in the given data, its low value will show that the country is good in rank and vice versa, on the other hand, the indicator like Imports should be high in order to raise the rank of its corresponding country. So sorting out good and bad indicators is necessary. As data was

given to us in CSV file, file reading and many data sorting techniques are required to sort out the ranks. Many data structures are used while sorting out the ranks. As asked to do this work in Python language, python has been very helpful for doing its purpose. Python has helped a lot in storing the data in lists and sets.

2.Objective:

The main focus of the task is to rank precisely every country in a list or set according to the rank of the indicators (both good and bad) from very big data. Learn new ways to deal with clusters of the data. Get used to Python Language as it is very handy to deal with. List of the objectives is

1- Handling large amount of data.
2- Differentiating good and bad indicators and assigning them values.
3- To make a good algorithm (efficient and reliable) to sort out all the data, keeping all the prospectives in mind.
4- To have a sound grip in the programming language that would be used.

3.Methodology:

The methodology which is used to deal with Data given us in Gapminder.csv file is given below.

3.1. Data Handling:

The data that was given was in CSV file. That file contains data of more than 200 countries, each country was given a value no. of 49 properties (indicators) each, Over the almost 22 years. This data was raw and mixed up. The countries were in column and their indicators were given in there row year by year.

These properties were very much related to each other. For Instance, Co2 emission and yearly co2 emission both should be low in order to show correctness. Similarly many factors were like one should be high the other must be low and vice versa.

Moreover in some year the value of some indicators was not given which was also affecting the rank of that country in that particular year.

So, it was necessary to take the calculate every good and bad indicator, giving it a value according to its category (good or bad).

The sum of bad or good indicators will affect the ranking or value of the country in overall world rank.

3.2. Encountering Problems:

There were many problems faced in ranking the countries. Many indicators of the countries in some years were not given any value. Data was given to us in a '.csv ' file. First of all it has to be converted into any list or set storing all the data. In order to check any given country instead of reading in to the file, that particular set will be used to rank all the countries. Many indicators were bad and many were good, so separating them was necessary to avoid any error in sorting the countries. Besides this, as all the 49 indicators were given a specific value, in order to differentiate and standardize any country by using its indicators value, ranking all the indicators was much more preferable.

While storing data from file we made a dictionary (data structure) having average of all the indicators separately for every country and also made a dictionary of max value of every indicator separately.

Taking the max value is helpful while ranking good and bad indicators.

As some indicators of the countries have very big value while on the other hand same indicators of other country has very less value e.g , -221 . In order to take all the indicators in same page taking average of all the indicators according to the country given and then giving rank to that particular indicator weather it is bad or good was advantageous.

3.3. Ranking The Data:

Ranks given to any particular indicator of the country was then passed to a categorization process which decides if that is a good indicator or bad and then whether that specific indicator is in good, bad, average or worst condition.

In case of country, as all the indicators both bad and good was given a certain value. To rank any country, all that values are summed up and then stored in another dictionary.

In the final step we have to sort out that dictionary where all the countries were stored in the previous process when that dictionary is sorted out all the countries would get their particular special rank in that dictionary.

This method would give us ranks, first according to the indicators (by summing all the values gained differentiating good and bad indicators)

and the other, by sorting out the values of every country in the dictionary where all the values of indicators was summed up. That will be the actual rank of the country, while the former would be the overall rank according to the indicators.

4.Code And Design:

Most of the code written is in the main function. However as besides the 'gapminder.csv' file, some code was also given to check the values of some countries using graph. As graph was not needed, that part of code is removed. Another python file named "Functions.py" file is used to apply some functions in the main file "Project.py" file using import function. The function.py file contains the functions to rank good and bad indicators, rank_sort function which sorts all the countries to rank them in final process also a rank_top10 function which is simply a ranking function(rank_sort) called just for top 10 countries.

The main file ("Project.py") contains many functions some are as same as given to us like 'fetchIndices()' etc. dataTypeConversion() removes all the missing values in the indicators. Beside this there is a ranking function("ranking()") to rank condition of good and bad indicators("good", "bad" , "average") and then printing them according to the country entered.

Along with the functions in the main file there are operations like which first convert data from csv file to dictionary the "CountriesDict" , to store the max value of every indicator against every particular country or state in a dictionary called "max_dict" , to rank every good and bad indicator of every country by first taking average of that indicator then taking its substitute value after that checking that whether it is good or bad and then passing the substitute value to the good or bad functions,

which will give the indicators a rank according to their value (average) for that particular country or state.

Moreover there is also an ranking operation which calculates the sum of all the ranks given to every indicator of any specific country and then passes the value to "total_rank_dict" dictionary.

For the interface there are just two while loops asking for ranking the country and the other for viewing the condition of the indicators of any country.

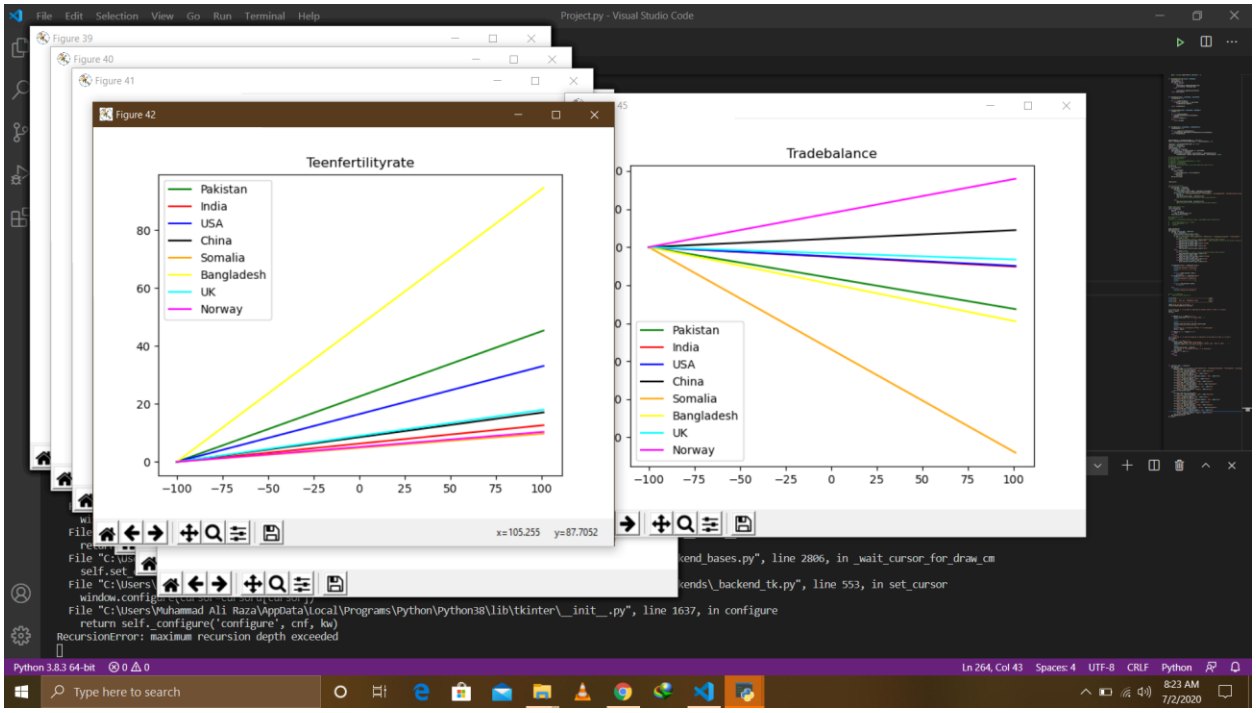
All the functions are named according to their functionality, variable, dictionaries etc, are also named according to their purpose, to avoid any confusion and misunderstanding , the reason separating the functions.py file is also this to avoid intermingling of functions and operations.

5. Graphical Representation:

In the end of the program to illustrate the difference in the indicators over the ranks that were concluded of each indicator individually for every particular country. Some countries where short listed like ‘Pakistan’ , ‘India’ , ‘Norway’ , ‘USA’, ‘UK’, ‘Somalia’, ‘Bangladesh’ and ‘China’.

Every indicator of above mentioned countries would be plotted at the end of the program to show the change in the ranks (values) of each indicator for that specific country.

The elaboration is given below.



In the x-axis there are values that were given to each indicator separately for their value taken as average to rank the country, so to show the change of the values limits from -100 to 101 as a lot of indicators were also given negative values.

In the y-axis there is simply the range of every indicator.

6. Summary:

By using the data given one can't tell that whether the country is progressing, developing or not as one has to compare it with the progress (value) of same country in another year that is very absurd.

Taking average is good in this prospective that it sums up all the good and bad years for any specific indicator and it is easy to rank any particular country basing this method. For instance any indicator of a country is good for last decade and then it suddenly falls. If one see progress by depending upon the year one can't summarize the rank of that country.

Ranking of any country needs basically reading its indicators, taking average of that , giving the indictor a precise value and then passing it to rank in good or bad indictors and finally summing up all value in a sorted list.

7. Conclusion:

The final result of the code is good every country has its own specific ranking over all the countries given in the data. Every country has sorted out good and bad indicators. Moreover every country now has the condition of its good and bad indicators that whether any particular is in good condition or bad or average etc.

Coding implementation logic is up to scratch. We learnt new ways to deal with large amount of data. Python was very handy and useful for storing that much data in different data stuctures.