World Urban Population Year by Year Analysis

Student Name :Muhammad Adeel
Student ID : 22021887
Assignment No.2

Github link:
https://github.com/muhammadadeel22/ADS2_assognment2.git

Abstract

The world urban population is increasing year by year as is obvious from the dataset which comprises data from the years 1960 to 2021. The dataset used for this study is taken from the world bank website. Statistical parameters like a five-number summary were used to take an insight into the dataset. Furthermore after summing up two six years of urban population data. I have found a positive correlation of 0.88.

I have chosen one of the datasets from world bank data available online. I then wrote a func in python and read that CSV file into it. Since I was getting errors in reading that file. So I opened the file and remove unbalanced rows in the beginning of the file. After reading the csv in that function it will return 3 data frames d1 with county name and indicator other d2 with the country name and country code. The third data frame named d3 was of my interest and to take further insight into this dataset.

Since in the original data frame the years are used as columns from 1960 to 2021. So I use the melt function to convert the dataset columns to rows values.

Then I further extracted data from 1980 to 1985 six-year data into data frame d4. Similarly in data frame d5 I have extracted the data from years 1985 to 1991.

Since in this data frame before extraction year-wise data I have to change the year data which was string so I changed the year column as an int using astype.

I then dealt with missing values of both the data frames using the function fillna(0).

Then to get an insight in these two dataframes I used describe function which will give me five number summary of every five year data.

Then I further summed up year-wise data of both these data frames.

I also used another function of statistics named pearsonr from scipy.stats library to see whether ther is any correlation between each five years i.e. 1980 to 1985 and 1986 to 1991.

What I come to know is that there is a positive correlation of 0.88 in both of the datasets.

I have made graphs of both the datasets to visualize urban population year wise.