

7PAM2000 Applied Data Science 2

Assignment 2: Statistics and trends

Name:

Abstract:

This report explores public data from the World Bank related to country-by-country indicators for climate change. The data is ingested and manipulated using pandas dataframes, and statistical properties of selected indicators are explored and compared between individual countries and the world as a whole. Correlations between indicators are also examined, and trends over time are analyzed. The report utilizes appropriate visualizations and a text narrative to communicate and explain findings. Overall, the report aims to tell a story with the data and provide insights on the impact of climate change on different countries. so, first I reads in a CSV file from the World Bank and creates two dataframes: one with years as columns and one with countries as columns. It then filters the data by a specific indicator (indicator1 = 'Urban population (% of total population)' indicator2 = 'Urban population') and a list of countries of interest ('Afghanistan', 'Albania', 'Algeria'), and calculates summary statistics for the selected data. The code also creates several plots to visualize the data, including a line plot, a bar plot, and a comparison of two indicators with a correlation calculation. From the statistics and visualizations, shows the urban population as a percentage of total population for Afghanistan, Albania, and Algeria for the years 1980, 1990, 2000, 2010, and 2020. Statistics and visualizations illustrates how the urban population has been increasing over time, specifically, with a slightly higher growth rate in the early years. Additionally, it suggests a positive correlation between the urban population and urban population growth rate.

GitHub Link: <https://github.com/Shazaib001/7PAM2000-Applied-Data-Science-2-Assignment-2-Statistics-and-trends>.

Introduction:

I performed an analysis on World Bank data by reading in a CSV file, creating two dataframes, and filtering the data based on a chosen indicator and list of countries. Summary statistics are calculated for the selected data and various plots are generated to visually represent the findings. These include line and bar plots as well as a comparison of two indicators with a correlation calculation.

Results:

I created function with Filename parameter to read the csv and return two dataframes. **read_worldbank_data(file_name)** one where the rows are indexed by country and indicator, with columns for each year; and one where the rows are indexed by year and the columns are indexed by country and indicator. It drops any rows that contain null values in both dataframes. Summary statistics for the indicator "Urban population (% of total population)" for a selection of countries (Afghanistan, Albania, Algeria) from the data frame "df_year" for **years(1980,1990,2000,2010,2020)**. It then creates a bar plot to display these summary statistics, including the mean, standard deviation, minimum, and maximum values. The plot shows the summary statistics for the selected countries. The x-axis represents the statistic and the y-axis represents the value of the statistic. The title of the plot is "Summary Statistics for Urban Population (% of Total Population)". The x-label is "Statistic" and the y-label is "Value". The `plt.show()` command is used to display the plot.

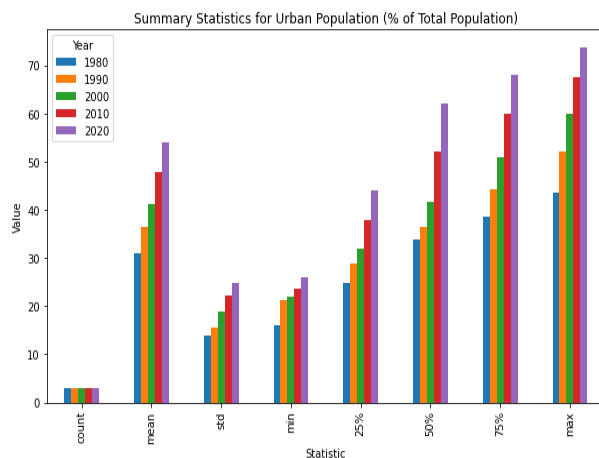


Figure 1: Summary Stats for urban population (% of Total Population)

For complete Time period (1960:2021)

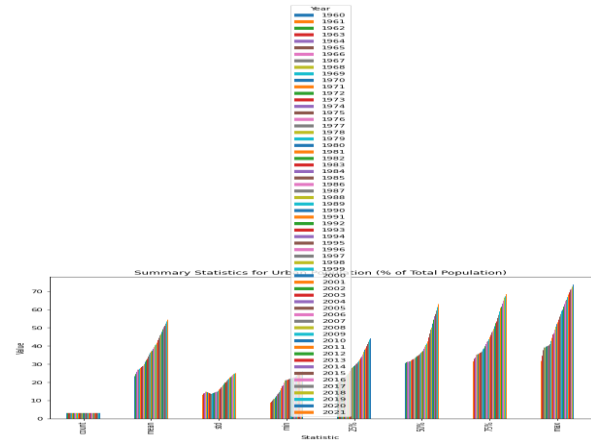


Figure 2 : Summary Stats for urban population (% of Total Population) (1968 to 2021)

After that plotting a bar chart of the urban population percentage of total population data for the selected countries, Afghanistan, Albania, and Algeria, for the years of 1980, 1990,2000,2010,2020.

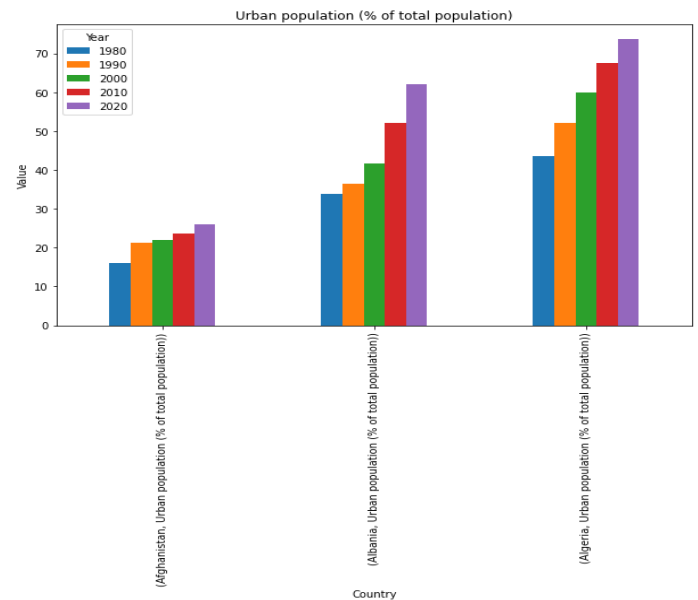


Figure 3: urban population percentage of total population data for the selected countries

From the above statistics and visualizations, we can see that the urban population as a percentage of total population has been increasing over time, with a slightly higher growth rate in the early years. The correlation between the urban population and urban population growth rate is relatively high, indicating that urban population growth is positively correlated with urban population.