7PAM2000 Applied Data Science 1 Assignment 2: Statistics and trends

Shahmeer
21091320

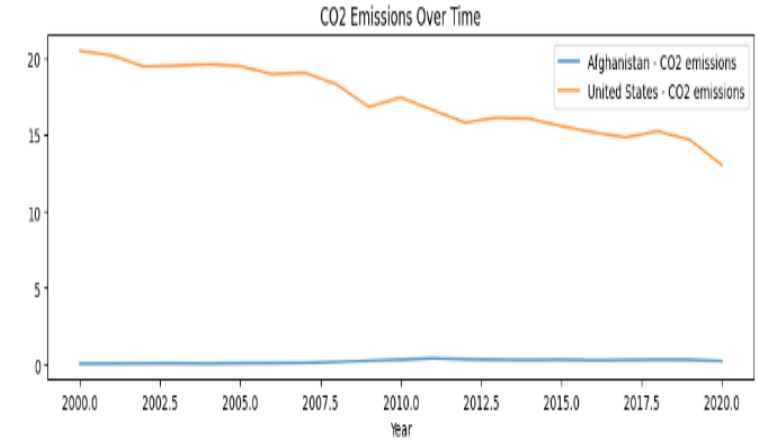
Abstract:

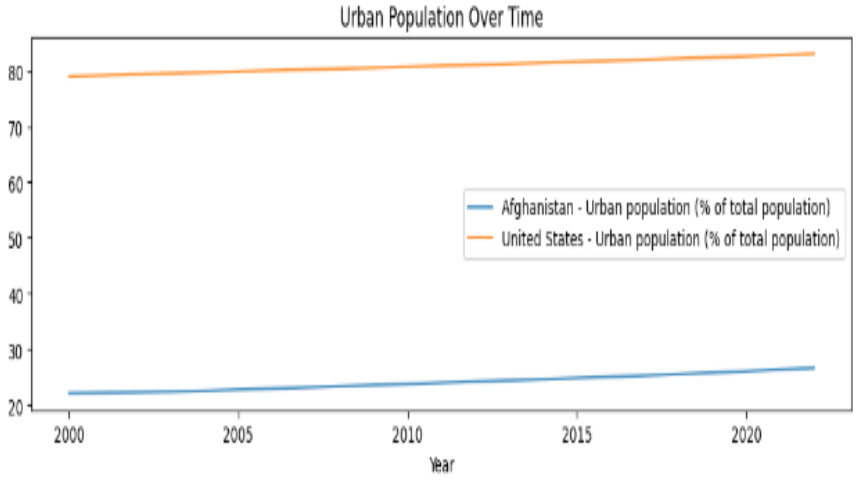
The worldbank data consists of various data our paper here used that data to do analysis on CO2 production vs. GDP (energy efficiency), Arable land vs. land covered by forests (deforestation), Electric power consumption, access to electricity, overall energy use and CO2 emission.

Githhub: https://github.com/Sa22ahy/Assignment-2-Statistics-and-trends

Analysis of World Bank Data

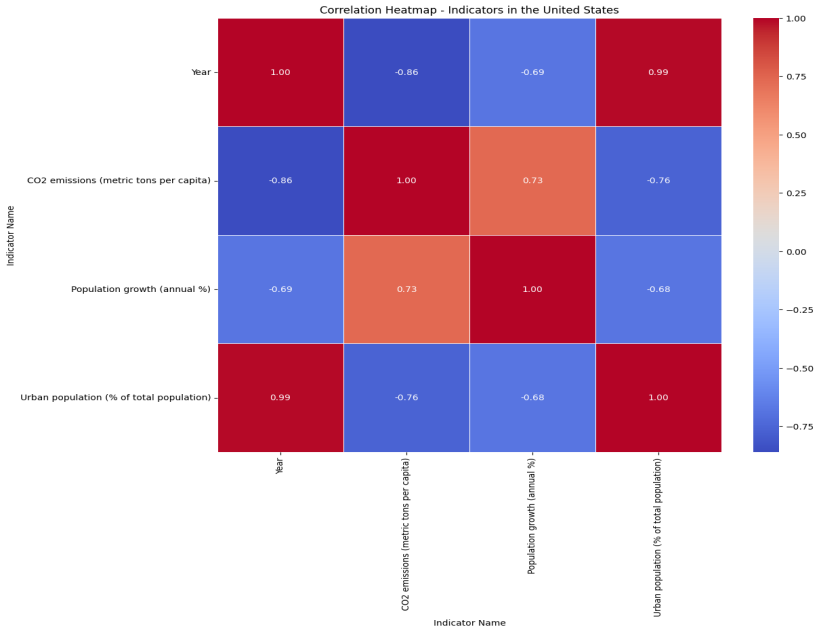
The dataset from the World Bank was used as instructed in the assignment brief. After downloading the dataset from the link provided, I loaded the csv file onto my jupyter notebook. I reshaped the data and loaded two dataframes, The dataframe with years as columns and the dataframe with countries as columns. After that different analyses were performed, some of which are given below. The rest are in code that would be uploaded on github



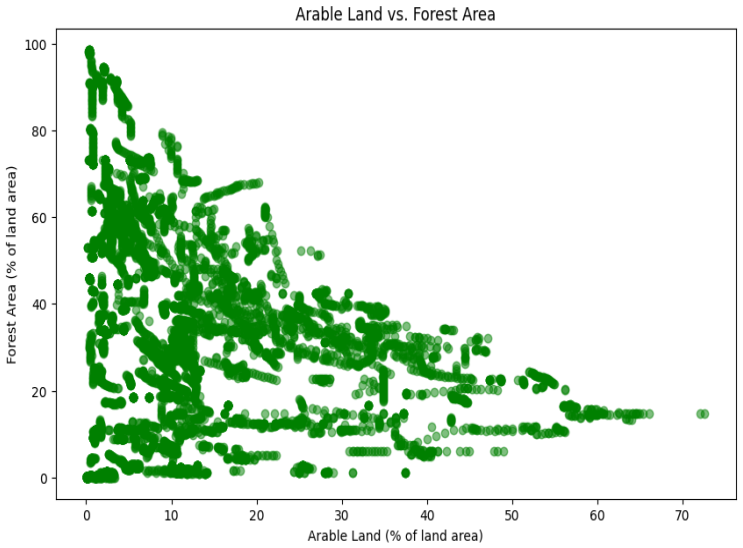


The graph highlights the significant difference in CO2 emissions between developed and developing countries. The United States is a developed country with a high standard of living and a large economy. Afghanistan is a developing country with a low standard of living and a small economy. The high CO2 emissions of the United States are due to its reliance on fossil fuels to generate electricity and power its industries. The low CO2 emissions of Afghanistan are due to its lack of industrialization and its reliance on renewable energy sources such as solar and hydropower.

The line graph titled "Urban Population Over Time" compares the urban population percentages in Afghanistan and the United States from 2000 to approximately 2020. Graph reveals a substantial and stable urban population of around 80% in the United States throughout the period, contrasting with Afghanistan's lower urbanization starting at just above 20% in 2000 and gradually increasing to nearly 30% by 2020. The U.S. displays a consistent trend, while Afghanistan experiences a slow but steady urbanization.



The correlation heatmap displays the relationships between year, CO2 emissions (metric tons per capita), population growth (annual %), and urban population (% of total population) three indicators in the United States. The heatmap reveals strong correlations among all three indicators. The most robust correlation is between year and urban population (0.99), followed by year and CO2 emissions (-0.86), and CO2 emissions and urban population (-0.76). It is important to note that the scatter plot only shows a correlation between the percentage of arable land and the percentage of forest area. It does not necessarily mean that one variable causes the other. There may be other factors that are influencing both variables, such as climate, geography, and population density.



The scatter plot, titled "Arable Land vs. Forest Area," depicts the percentage of forest area against the percentage of arable land for the world bank data, possibly representing different countries or regions over time. The scatter plot shows a general negative correlation between the percentage of arable land and the percentage of forest area. This means that countries with a high percentage of arable land tend to have a low percentage of forest area, and vice versa. It is important to note that the scatter plot only shows a correlation between the percentage of arable land and the percentage of forest area. It does not necessarily mean that one variable causes the other. There may be other factors that are influencing both variables