Daniel Stenberg
2022-06-28

Project: Explore Weather Trends

Udacity Nanodegree – Data Analyst

Background

Dataset Description

The Gapminder Foundation collects and provides data on global trends relating to population growth, economics, energy, education, health and more. It was founded by Professor Hans Rosling, one of my favorite scientists, to provide insights into what is really happening in the world backed up by data.

In this project, the following data sets have been selected:

- Yearly population by country
- Yearly CO2e emissions by country
- Yearly change in GDP (economic growth)

The Gapminder Foundation collects and provides data on global trends relating to population growth, economics, energy, health and more. It was founded by Professor Hans Rosling, one of my favorite scientists, to provide insights in to what is really happening in the world backed up by data.

Questions for Analysis

Some questions that may be asked from these data sets are:

- At what rate is the global population growing in different regions?
- What is the relation between population and CO2e emissions?
- Are increased CO2 emissions linked to economic growth (GDP)?
- How does the global numbers on this compare to regional numbers like population growth related to energy production in Africa as compared to Europe?

Data Extraction

The data is extracted from CSV files obtained from Gapminder's github page "Gapminder Systema Globalis" (https://github.com/open-numbers/ddf--gapminder--systema_globalis).

They are loaded into Jupyter notebook using the `pandas` package `read_csv` function call. An example would be:

```
df_pop = pd.read_csv('ddf--datapoints--population_total--by--geo--
time.csv')
```

All data was provided per country. To simplify the analysis, country data was aggregated into region data for:

- Africa
- Americas
- Asia
- Europe

This was done using a combined `groupby()` and `sum()` operation.

Daniel Stenberg
2022-06-28

Data Cleaning

Each of the datasets are thoroughly analyzed and cleaned using pandas function calls such as:

- `info()`
- `head()`
- `tail()`
- `isna()`
- `drop()`

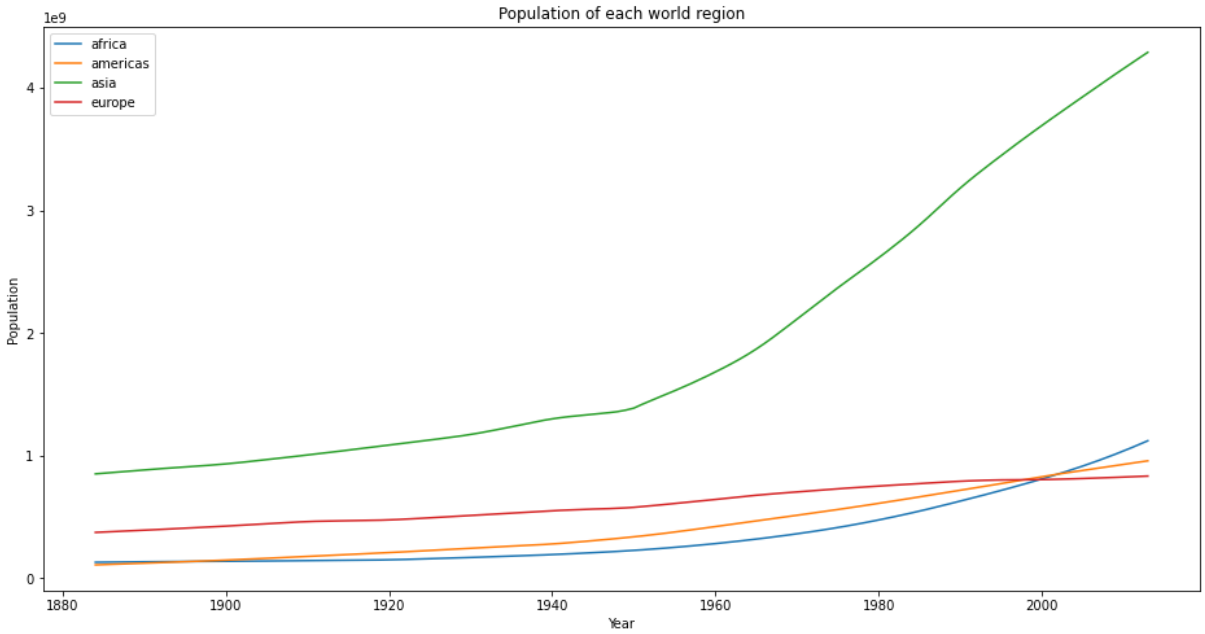
When this was done for each of the datasets, a “master” dataset was constructed by combining the columns using the `join()` and `groupby()` function calls in pandas.

This dataset required additional cleaning to align relevant periods of data.

Data Exploration

Population

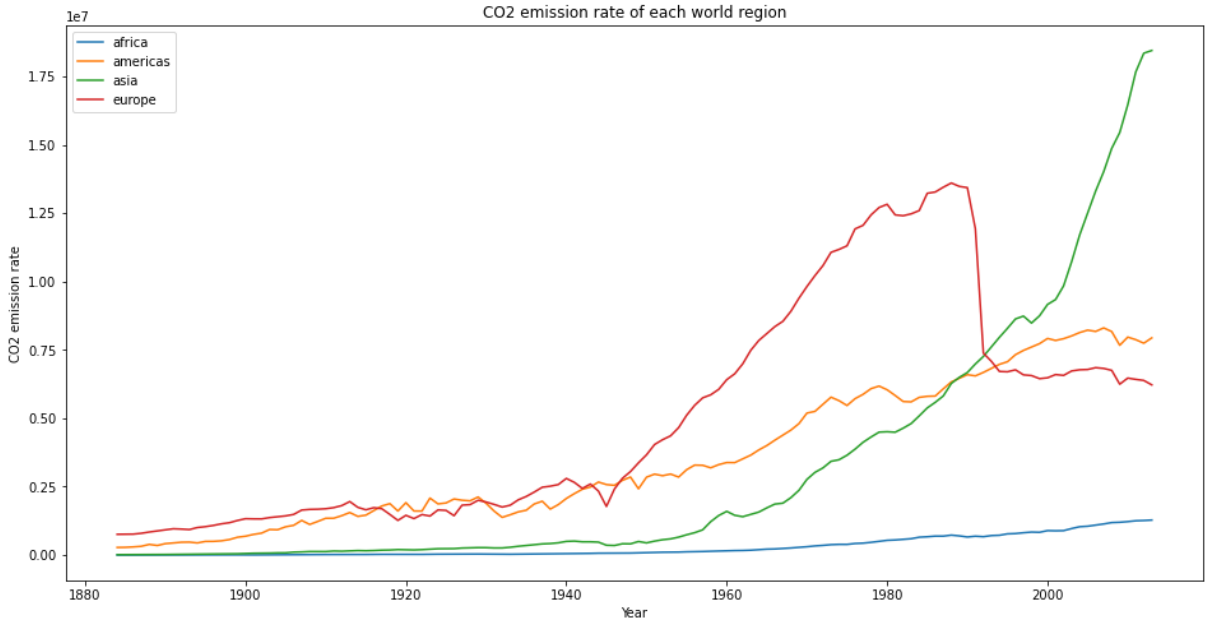
The following graph shows how populations are growing in each world region.

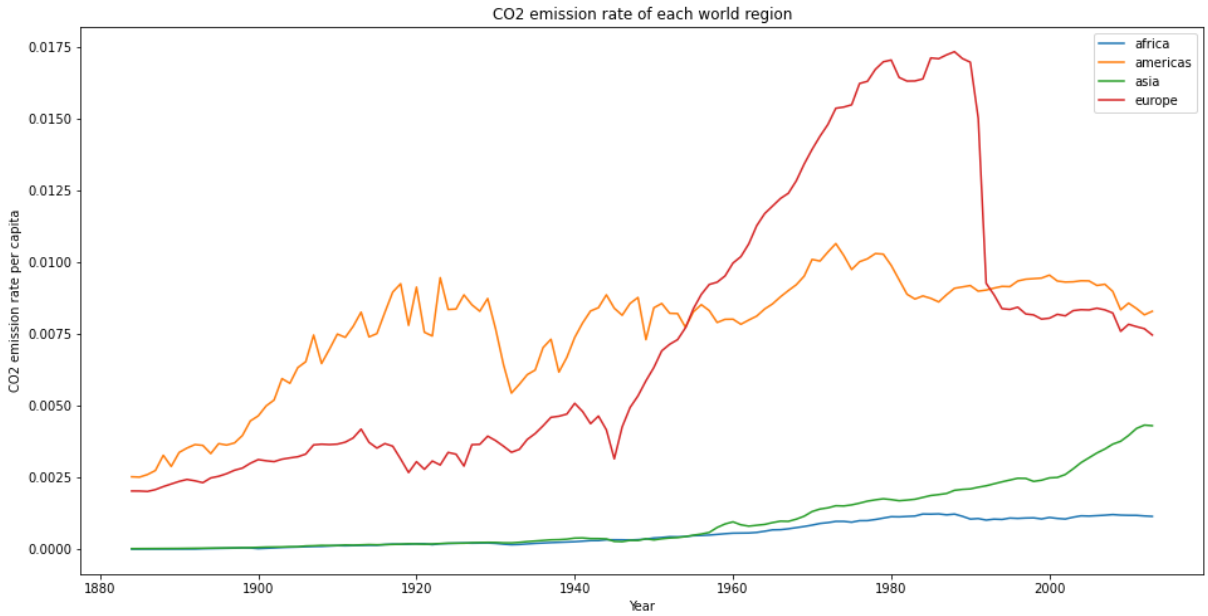


CO2 emissions and populations

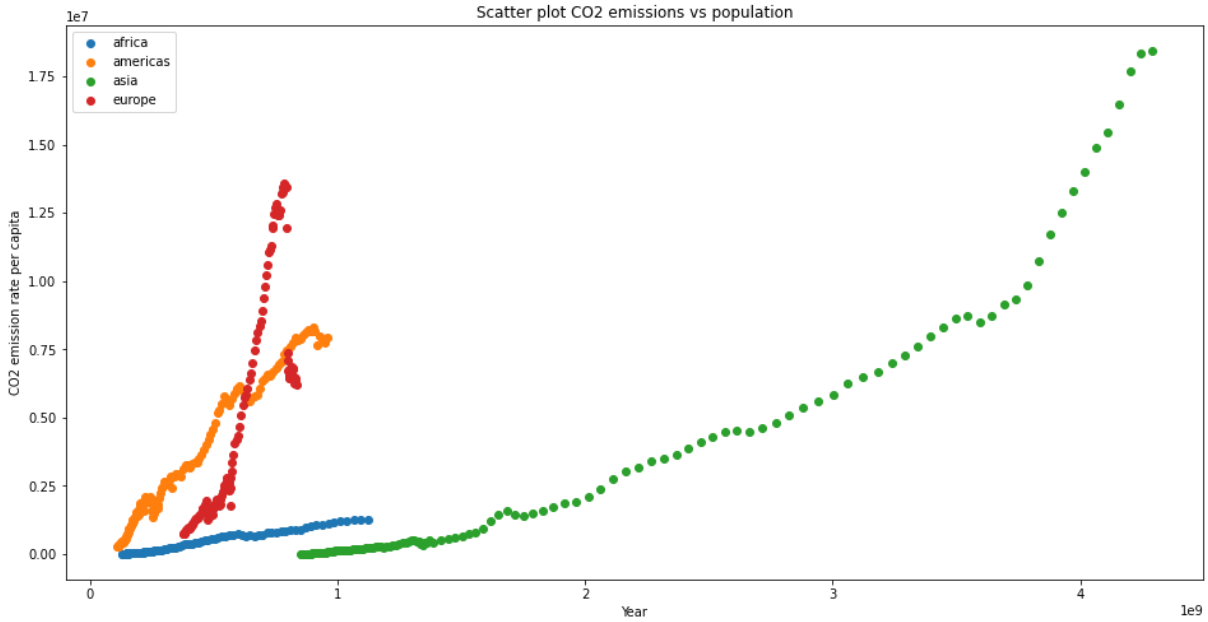
The following graphs show how CO2 emissions have been developing in each world region and how population growth relates to CO2 emissions.

Daniel Stenberg
2022-06-28





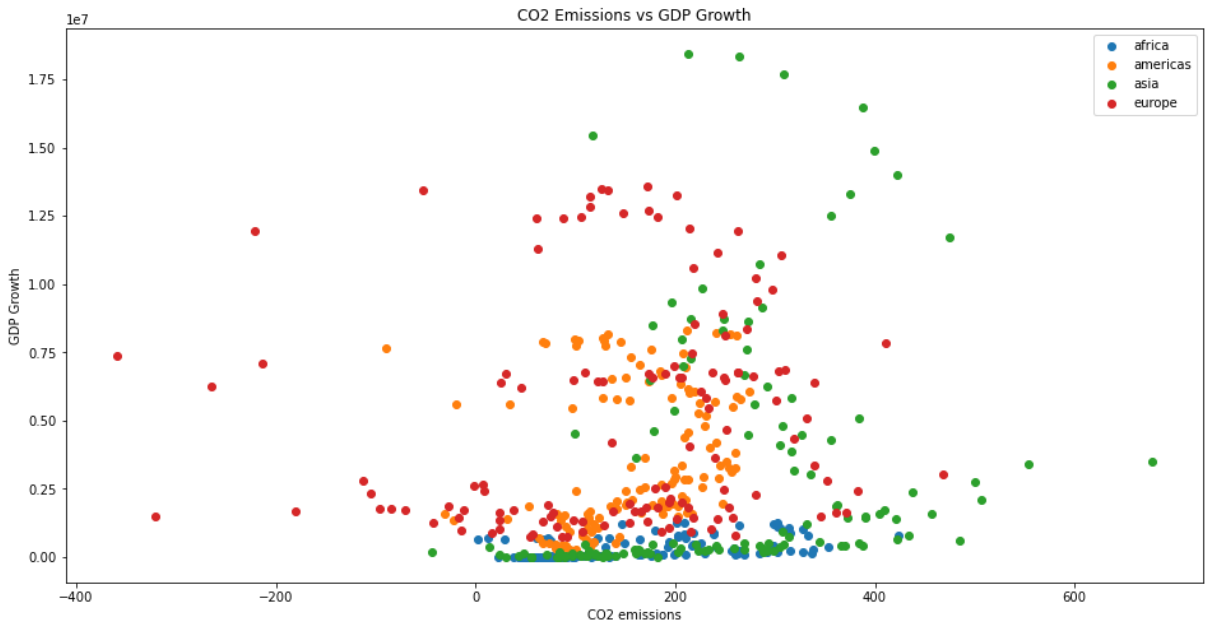
Daniel Stenberg
2022-06-28

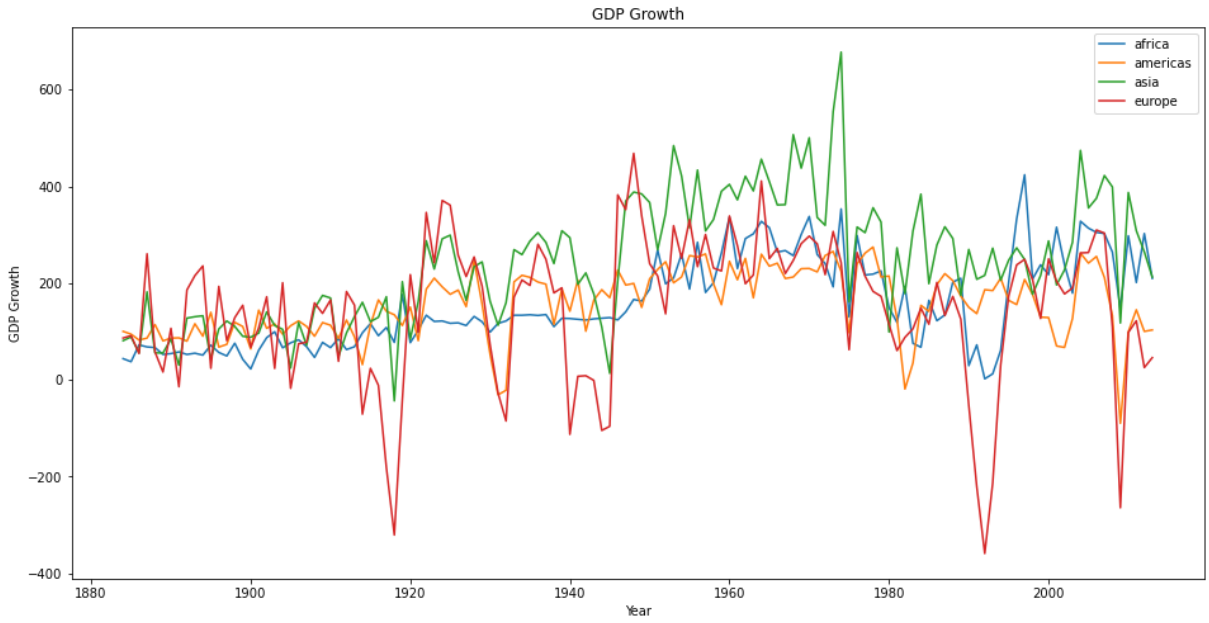


GDP Growth and CO2 emissions

Below follows graph showing GDP growth for each world region as well as its relation to CO2 emission rates.

Daniel Stenberg
2022-06-28





Data Analysis

Conclusions

The analysis shows that there is a clear correlation between population growth and CO2 emissions in all world regions. However, the positive correlation factor between regions can vary in magnitude (Africa lowest correlation, Europe high correlation).

World population is also steadily growing, most rapidly in Asia and Africa.

The analysis also shows that there is no clear correlation between GDP growth and CO2 emissions. No pattern can be discerned frow the data.

Limitations

As the analysis originates from 3 separate datasets, it is possible that some countries data has been missed or not included in all datasets.

Daniel Stenberg
2022-06-28

Also, a possible anomaly is the data is noted in the European CO2 emissions data at around year 1990. The background of this anomaly is unknown.