Report on Weather Trends

By Vijay Raj Saravanan Radhakrishnan

Tools Utilized

Python
Pandas framework
Matplotlib framework

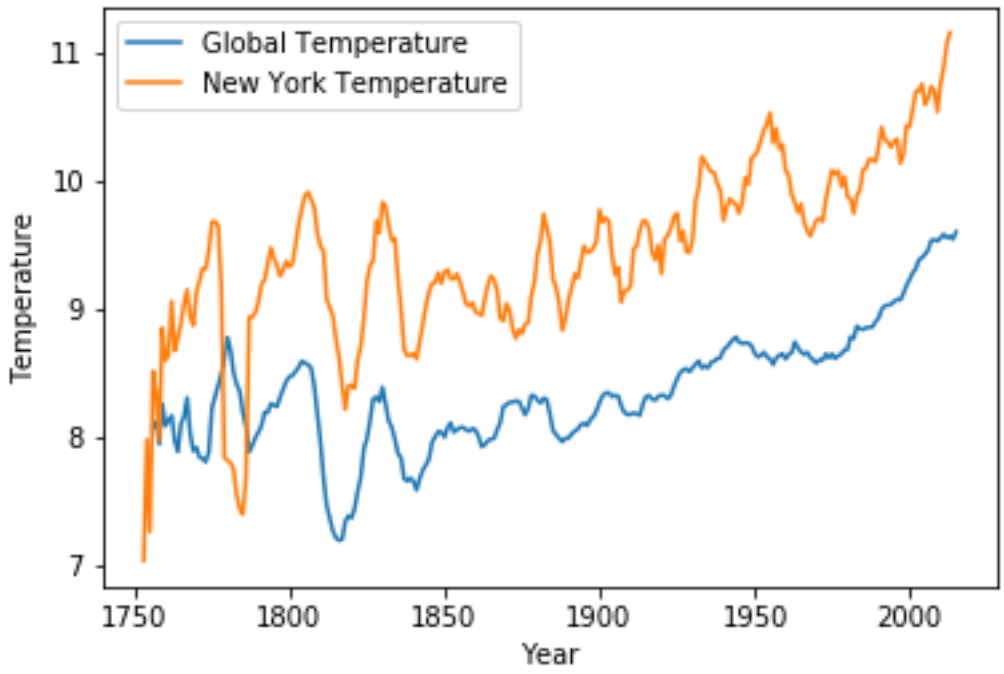
SQL Queries used

SELECT * FROM city_data;
SELECT * FROM global_data;
SELECT * FROM city_list;

Steps used to arrive to conclusion

- I imported the CSV files into pandas data frames and analyzed their columns and data in it.
- I wanted to compare global weather trends to New York's weather over the past years and hence I checked if New York exists as a record in city_list table and found it to be present.
- I found that there were a very small number of records in city_data table that had Nan values for average temperatures of New York for certain years. Since they were a very small amount comparatively, I removed them from my data frame.
- I proceeded to calculate moving average of 7 years for New York's average temperatures and global average temperatures using pandas rolling().mean() function and plotted them on a line graph using Matplotlib framework.

Year VS Average Temperature Line Graph



Observations

- On average, New York's temperature seems to be higher(hotter) than the global temperature except in the 1750s and 1780s where it seems to have taken sudden falls and spikes which might be due to missing data.
- The lower extremes of average temperature seem to have increased significantly from the 1750s to the 2000s indicating a rise in global temperature which is evident from the corre lation coefficient. The correlation coefficient was calculated to be 0.7310703144258599 which states that a strong positive linear relationship exists between years and temperatur e signifying that temperatures have been increasing along the years.
- Over the recent years, i.e., after 2000, New York's temperature seems to be increasing significantly compared to previous years.

- There seems to be a significant difference between lows and highs in New York's temperature whereas global temperature does not vary too much comparatively.