E X P L O R E W E A T H E R T R E N D S

Data Analyst Nanodegree
Udacity

Sourojyoti Paul

Objective

The objective of this project is to extract data from given tables containing temperature details of all cities and global temperature details. From the extraction it is required to create visualization, understand patterns and derive insights describing the similarities and differences between global temperature trends and temperature trends in the closest big city

Tools Used

SQL
-To query the tabular data and extraction to excel sheet for further analysis.
WPS Spreadsheet
-To clean the extracted data, data preprocessing activities and calculating required moving averages and plotting them to line graphs

Data Extraction

Analysis of data begins at the very first step, which is gathering of data or data extraction. If done properly and efficiently, it can reduce many of the steps down the line. In this project we are required to extract data based on the city we are closer to which is listed in the table 'city_list'.
Here I have considered the city 'Pune' as I am more close to 'Pune,India'.
Here we have been provided data in structured tables, hence SQL has been used to retrieve data and export it to excel for further analysis.

SQL 1 : select c.year as year,c.city as city,c.country as country, c.avg_temp as city_temp_avg, g.avg_temp as global_temp_avg from city_data c inner join global_data g on c.year = g.year and c.city like 'Pune';

SQL 2 : select * from global_data;

Here I have used inner join in order to join two tables city_data and global_data to form a joined table, joining on common column data year from both the tables.

Data available from 1796 to 2013, downloaded as excel for further analysis.

Data Cleanup and Preprocessing

Data cleanup and data preprocessing is a very important step during data analysis. This helps us to identify and remove erroneous or null data which may lead to wrong analysis.

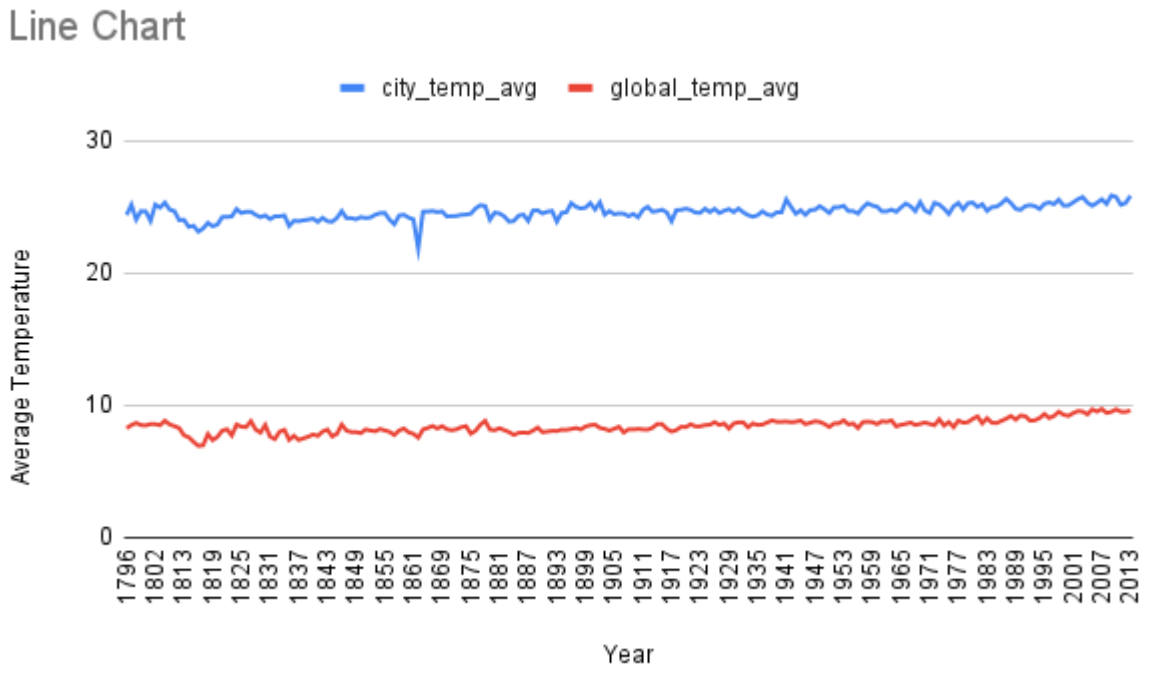
In our dataset we have multiple scenarios where average temperature is not available for a particular year, but in the same year global temperature data is available. To make the data consistent it is necessary to remove the data for both global_data and city_data for that particular year

As part of data preprocessing here I have performed moving averages of the city_temp and global_temp columns respectively, to make line graphs smooth. Three different moving averages are calculated for both the temperature columns, 7-Year-Moving- Average, 14-Year-Moving-Average and 21-Year-Moving-Average. After plotting it as a line graph, it is observed here that the more the interval is for calculating moving average the smoother the graph becomes.

Data Visualizations

Line graph on extracted data

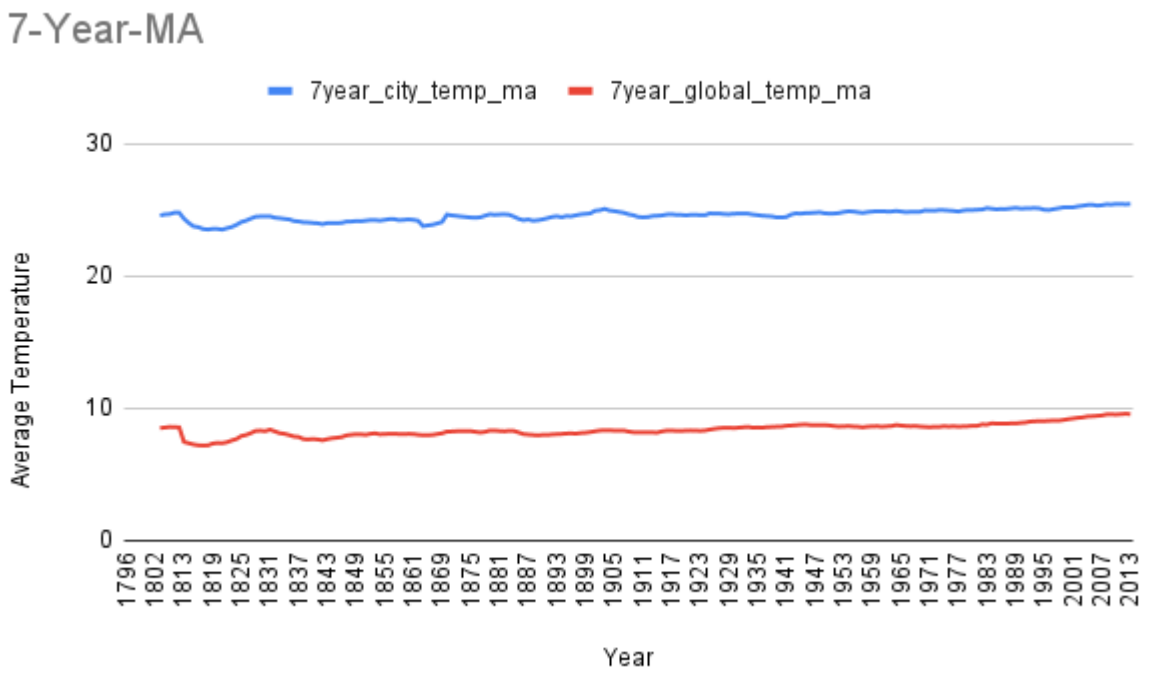
Prepared line graph on extracted data. The data was cleaned for null values before visualization. The graph incorporates city_temperature_avg, global_temperature_avg for every year from 1796 till 2013, except null values.



From the above graph, nothing in particular can be inferred since there are lot of variations in the graph.

Line graph on 7-Year-Moving-Average

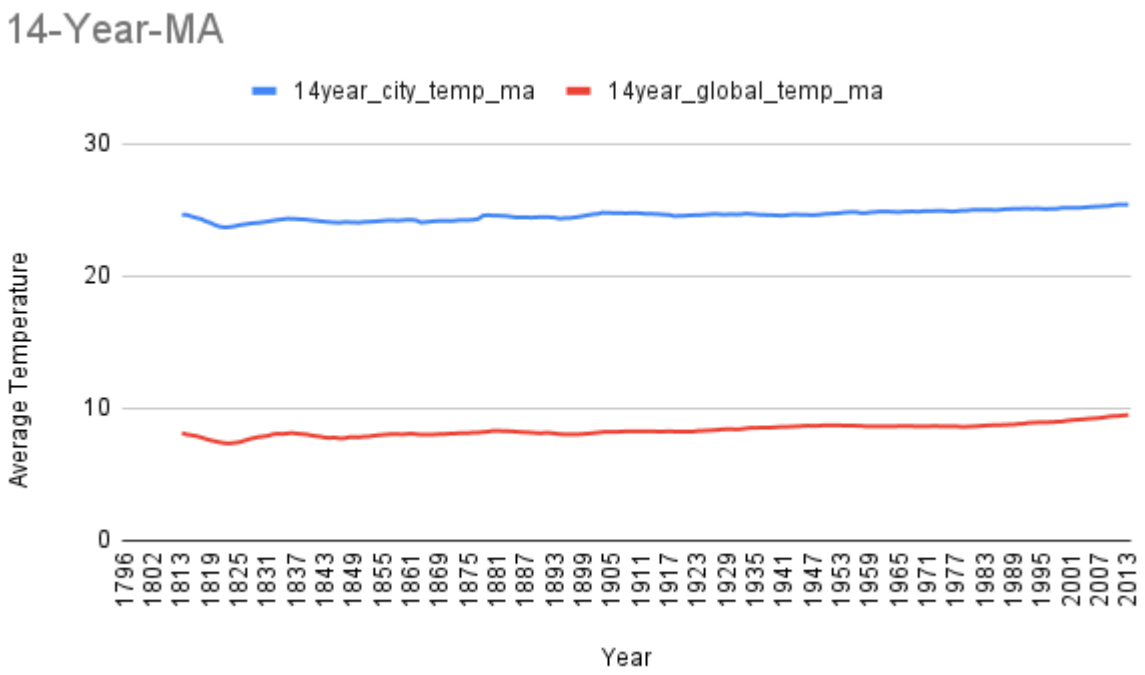
Prepared line graph on 7-Year-Moving-Average data. The data was cleaned for null values before visualization, and 7-Year-moving-Average was calculated for both city_temp data and global_temp data. The graph incorporates city_temperature_avg, global_temperature_avg for every year from 1796 till 2013, except null values.



The above graph is much smoother than the original one.
But another graph is plotted with 14-Year-Moving-Average, for better inference of and insights.

Line graph on 14-Year-Moving-Average

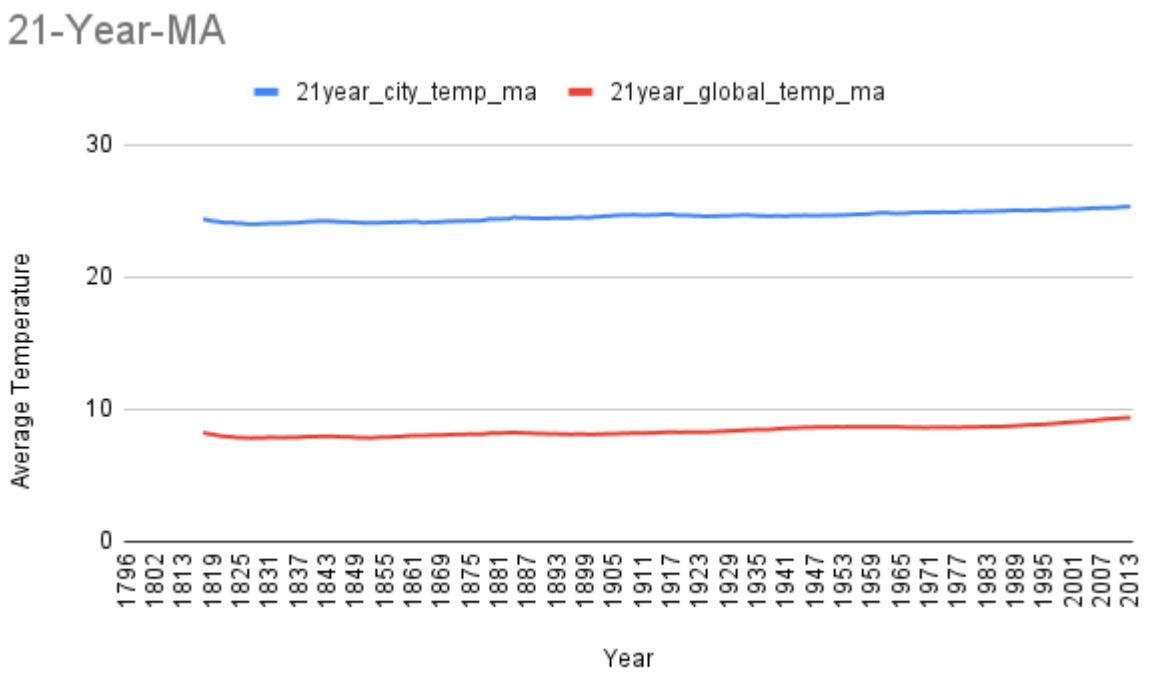
Prepared line graph on 14-Year-Moving-Average data. The data was cleaned for null values before visualization, and 14-Year-moving-Average was calculated for both city_temp data and global_temp data. The graph incorporates city_temperature_avg, global_temperature_avg for every year from 1796 till 2013, except null values, similar to 7-Year-Moving-Average but with increased interval from 7 to 14



The above graph is much smoother than the previous two graphs and much more suitable to infer some insights. It is evident that for all these years, the temperature of 'Pune' was much more than the global temperature. There is an average difference of 15 units in temperature scale.

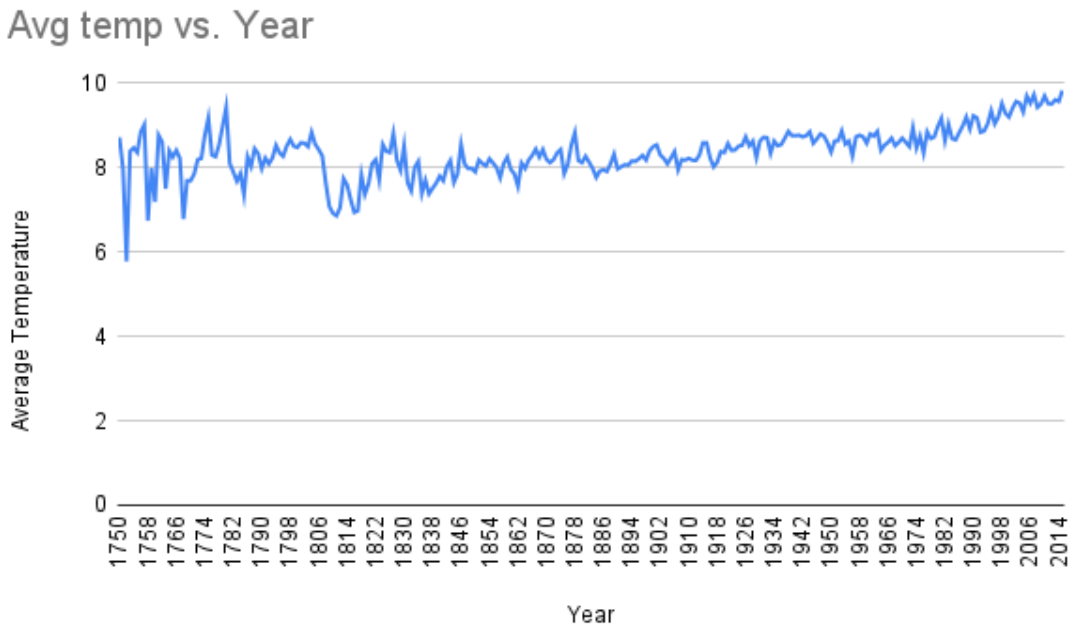
Line graph on 21-Year-Moving-Average

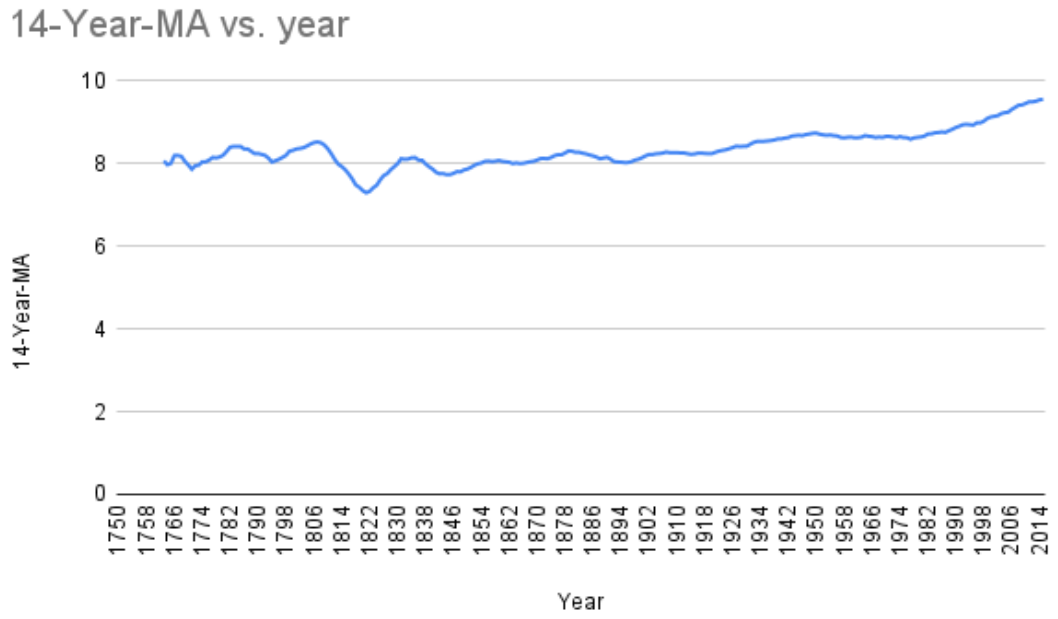
Similar to the previous two graphs, a graph consisting 21-Year-Moving-Average is plotted



The above graph is smoothest than the previous graphs and much more suitable to infer some insights. It is evident that more the interval is considered for Moving-Average, the smoother the graph becomes.

Additionally plotted line graph for global temperature





It is evident that the global temperature has increased over the centuries.

Observations

1. The weather of 'Pune' is consistently much warmer than the average global temperature.
2. There is an average difference of 15 units in temperature scale comparing each year of city_temperature data and global temperature data
3. For both the cases we could see a gradual increase in the temperature which could be due to global warming across these years, more data required to support the observation.
4. The more the interval is considered for Moving-Average, the smoother the graph becomes.