

Project: Explore Weather Trends

by Anupam Shekhar

Preface:

In this project, we will be analyzing the temperature data of Delhi (India) and the Global temperature data and would be drawing a comparison amongst the temperature trends and state our observations.

We would also compare this data against four other cities, namely London (United Kingdom), Moscow (Russia), Munich (Germany) and New York and try to see what we can observe and learn from them.

Tools:

SQL

MS Excel

We would be using SQL for the extraction of the temperature data from the database and would be visualizing the extracted data by restructuring the data and creating visualizations in MS Excel.

Required Steps:

Step 1: Extraction of data using SQL

Upon analyzing the database, we find that there are three tables in it, namely:

city_list - This contains a list of cities and countries in the database. Comprised of columns 'city' and 'country'.

city_data - This contains the average temperatures for each city by year (°C). Comprised of columns 'year', 'city', 'country' and 'avg_temp'.

global_data - This contains the average global temperatures by year (°C). Comprised of columns 'year' and 'avg_temp'.

We begin by looking for the city, Delhi (local city) of the country, India in the city_list. We explore the city_list to see if we can find Delhi using the following command:

```
SELECT *  
  
FROM city_list  
  
WHERE  
  
city like 'D%';
```

Upon successfully finding Delhi, we move towards extracting the required data which are, the year and avg_temp of Delhi from the city_data table; and the avg_temp from the global_data table for our analysis by using the following JOIN command:

```
SELECT city_data.year,  
city_data.avg_temp  
AS delhi_avg_temp,  
global_data.avg_temp  
AS global_avg_temp  
FROM  
city_data  
INNER JOIN  
global_data ON  
city_data.year = global_data.year  
WHERE  
city = 'Delhi'  
AND  
country = 'India';
```

We also extract the temperature data of four additional cities namely, London (United Kingdom), Moscow (Russia), Munich (Germany) and New York along with Delhi and Global temperature data for our further analysis. We do so by using the following command:

```
SELECT city_data.*,  
global_data.avg_temp
```

```
AS global_avg_temp
FROM
city_data
INNER JOIN
global_data ON
city_data.year = global_data.year
WHERE
(city = 'Delhi' OR
city = 'London' OR
city = 'Moscow' OR
city = 'Munich' OR
city = 'New York'
AND
(country = 'India' OR
country = 'United Kingdom' OR
country = 'Russia' OR
country = 'Germany' OR
country = 'United States');
```

Step 2: Visualization of data using MS Excel

The extracted .csv file contains the average temperature data for Delhi against the Global average from the year 1796 to 2013. We find that the temperature data for Delhi is missing from 1808 to 1812 and from 1858 to 1869. Therefore, we would be analyzing the data with the missing values and also after deleting the missing values to see if there is any effect on the overall data.

The creation of the visualization would require us to:

- a) Calculate the Moving Averages of both the average temperature of Delhi as well as the Global average temperature.
- b) Plot a Line Chart using the above calculated Moving Averages to analyze the weather trends and state our observations based on the same.

Calculation Of Moving Averages:

Moving averages are used to smooth out the data by removing the volatility to make it easier to observe long term trends and not get lost in daily fluctuations.

We calculated the Moving Averages for both the average temperature of Delhi and the global average temperature for the periods of 3 years, 5 years, 7 years and 10 years. We settled with the 7 year period Moving Average, which was calculated using the following formula for the columns D8 (7 year moving average of Delhi temperature) and E8 (7 year moving average of Global temperature) respectively:

=AVERAGE(B2:B8)

=AVERAGE(C2:C8)

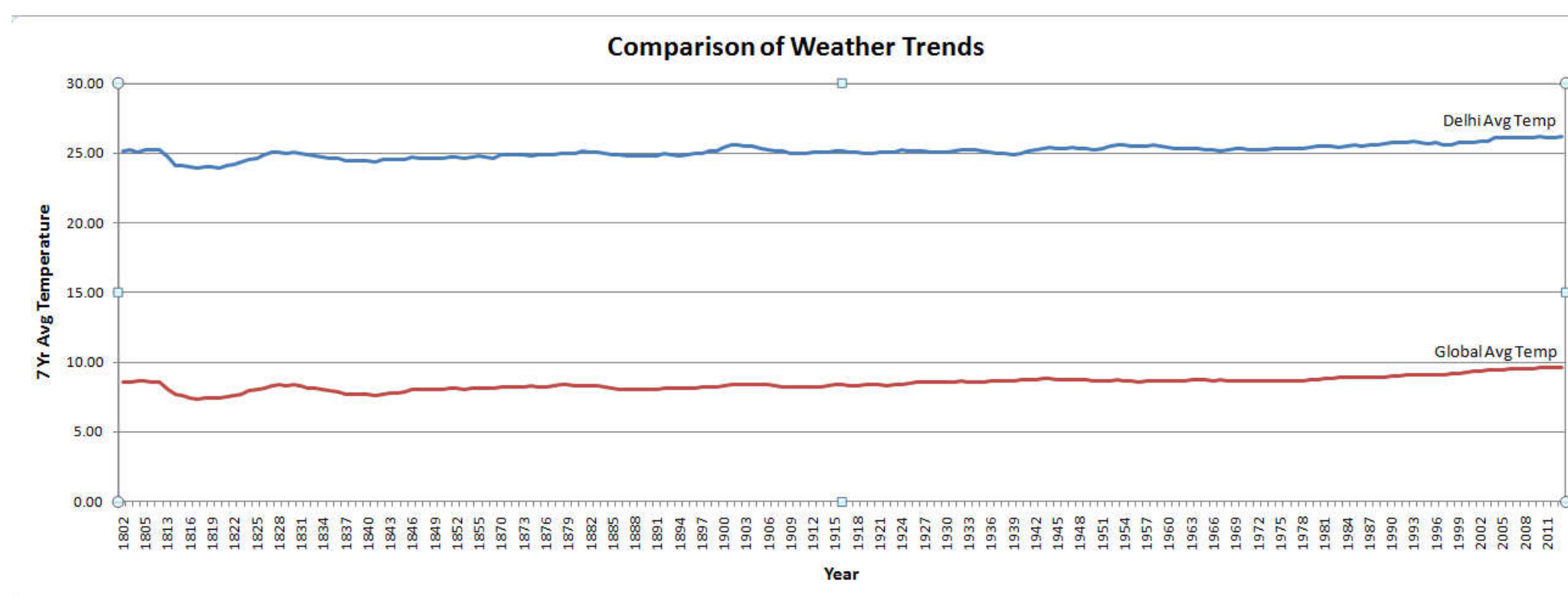
Upon the analysis of the 7 year moving average of Delhi temperature we find that the missing data has caused "#DIV/0!" errors from the period of 1864 to 1869 due to the missing data. Hence, we decided to delete the rows with the missing data to carry out a better clearer analysis of the data as there was no effect of the deleted rows on the data and the line chart looked smoother without them.

We similarly calculated the moving averages of the four additional cities London, Moscow, Munich and New York as above for our further analysis.

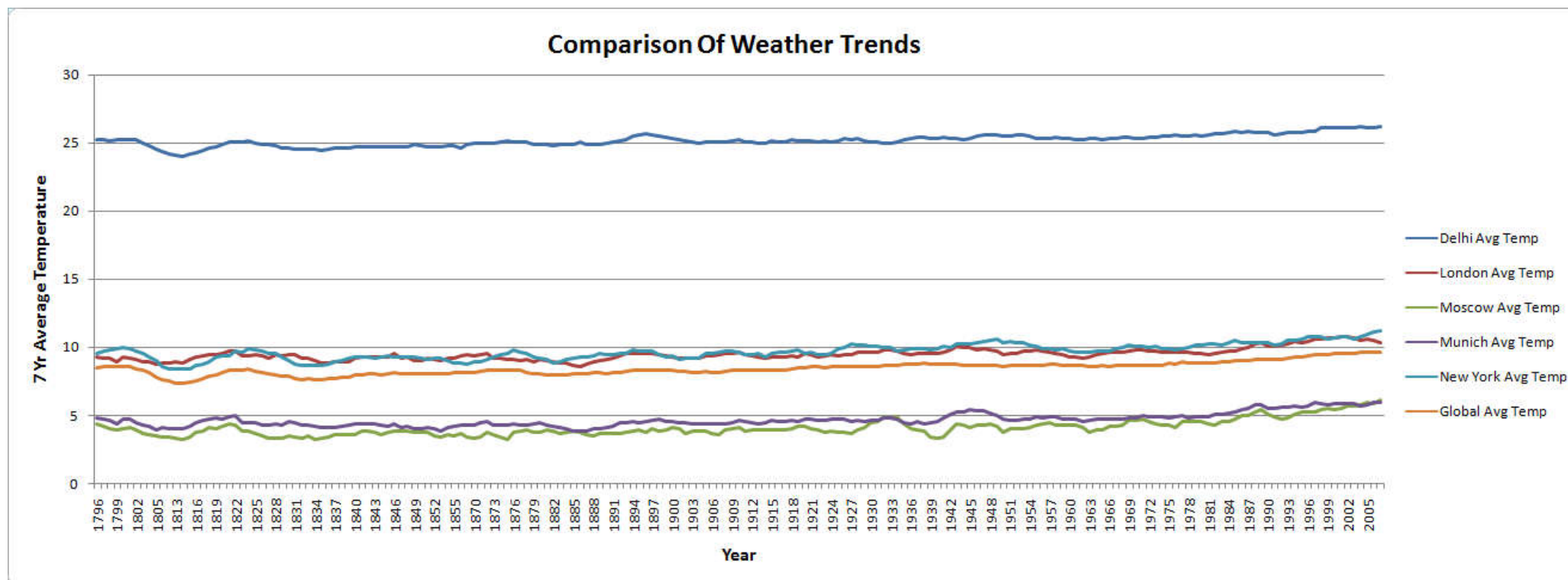
Plotting Of Line Charts:

Line Chart was used to plot the data as it consists of dates (years) and numbers (7 year moving average temperatures). This makes Line Chart an ideal tool to display the weather trends over the time for analyzing and making observations.

We now plot the line chart of the Delhi temperature against the Global temperature data. We plot the years on the X-axis and the 7 year moving average temperatures of Delhi and Global temperature on the Y-axis.



We similarly plotted the line chart of the Delhi, London, Moscow, Munich and New York temperatures against the Global temperature data. We plot the years on the X-axis and the 7 year moving average temperatures of the five cities and Global temperature on the Y-axis.



Calculation Of Coefficient Correlation :

Correlation coefficients are used in statistics to measure how strong a relationship is between two variables. The range of the correlation coefficient is from -1 to 1.

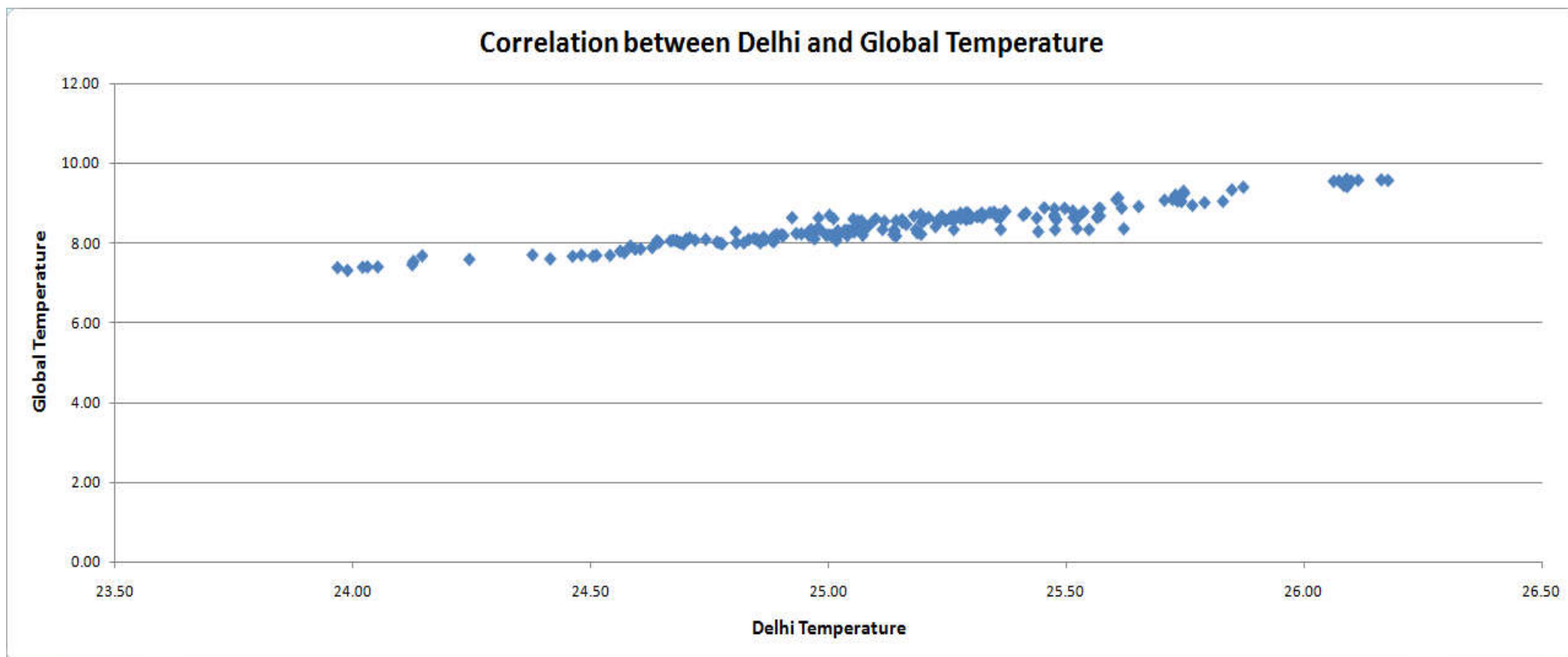
- 1 indicates a strong positive relationship.
- -1 indicates a strong negative relationship.
- A result of zero indicates no relationship at all.

We will use Coefficient Correlation to find out the correlation between 7 year moving average of Delhi temperature and 7 year moving average of the Global temperature. We can calculate the Correlation using the formula:

=CORREL(D8:D202,E8:E202)

The above formula returns a value of 0.95 which means that the 7 year moving average temperature of Delhi and the Global temperature have a highly positive correlation.

We can also plot the correlation between 7 year moving average of Delhi temperature and 7 year moving average of the Global temperature using a Scatter Plot as follows:



Observations:

The following are the observations that can be made based on the two Line Charts above:

- 1) Both the trends of the local temperature of Delhi as well as that of the Global temperature seemingly follow more or less the same pattern of changes overall. Though the average temperature of Delhi is much higher than the Global average implying that it is hotter than most of the countries comprising the global average temperature.
- 2) The average temperature of Delhi is approximately 2.6 times the Global average temperature. Per 1 degree increase or decrease in the Global average temperature corresponds to approximately 1 degrees (+/- 0.03 to 0.05 degrees) increase or decrease in the average temperature of Delhi.
- 3) While the 80's witnessed a greater rise and fall in the local as well as the average global temperatures across the years, the difference has been significantly lesser in the 90's. From 1890 onwards up to 2013 there has been a gradual rise in the average temperature. This implies that the world seems to be getting consistently hotter with the passing years.
- 4) While the changes per degree in the average temperature of Delhi as compared to the average Global temperature seemed slightly out of sync in the 80's up to the earlier 90's, they appear to get

more in sync later on. This could be due to the advancements in technology with the passing years that provide more accurate data.

- 5) The Coefficient Correlation between the 7 year moving average temperature of Delhi and the Global temperature is 0.95, which means that they have a highly positive correlation (the range of the correlation coefficient is from -1 to 1).
- 6) On observing the average temperature data of colder cities like Moscow (Russia) and Munich (Germany), it can be seen that there is a gradual but steady rise in the average temperatures of these cities since the 90's in compliance with the global temperature. This could imply to the increase in Global Warming.
- 7) On comparing the average temperatures of New York (United States) and London (United Kingdom) against the Global average, it can be observed that both can be seen to be at odds with each other. When there is a rise in temperature of New York, there is a fall in the temperature of London and vice versa. This could be due to the fact that New York is 40 degrees North from the equator while London is 52 degrees North, which makes them 12 degrees apart from each other.
- 8) The average temperatures of Delhi, London, Moscow, Munich and New York all follow the Global average trend and showcase a gradual increase in recent years. Only the average temperature of London showcases a gradual decline.

Sources:

<https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/correlation-coefficient-formula/>

<https://www.statisticshowto.datasciencecentral.com/moving-average/>

<https://www.real-statistics.com/descriptive-statistics/missing-data/>

<https://www.excel-easy.com/examples/moving-average.html>