<u>Project 1: Explore Weather Trends</u>

- **Aim of the Project:** To analyze local and global weather temperatures and derive insights by comparing your city's temperature trend to the global temperature trend.
- **Extracting the Data:**

SQL queries were used to extract the required data. The following are the queries that were used:

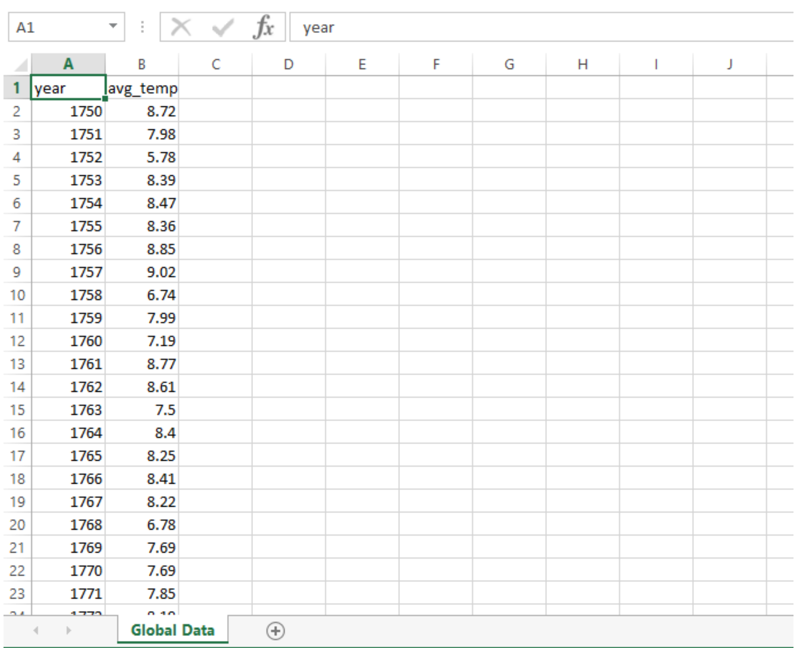
- **Select * from city_list**
  *Used to get a csv file with the list of all the cities available in the database. Since I live in **New Delhi**, I chose it for the purpose of this project*

- **Select * from city_data where city = 'New Delhi'**
  Used the above query to extract a csv file containing the weather data for New Delhi

- **Select * from global_data**
  Used the above quest to extract a csv file containing the Global weather data.

The following are the snapshots of the raw csv data files in excel:

Figure 1: Global Data CSV File



year	avg_temp
1750	8.72
1751	7.98
1752	5.78
1753	8.39
1754	8.47
1755	8.36
1756	8.85
1757	9.02
1758	6.74
1759	7.99
1760	7.19
1761	8.77
1762	8.61
1763	7.5
1764	8.4
1765	8.25
1766	8.41
1767	8.22
1768	6.78
1769	7.69
1770	7.69
1771	7.85

Figure 2: New Delhi Data CSV file

A1 | year

	A	B	C	D	E	F	G	H
1	year	city	country	avg_temp				
2	1796	New Delh	India	25.03				
3	1797	New Delh	India	26.71				
4	1798	New Delh	India	24.29				
5	1799	New Delh	India	25.28				
6	1800	New Delh	India	25.21				
7	1801	New Delh	India	24.22				
8	1802	New Delh	India	25.63				
9	1803	New Delh	India	25.38				
10	1804	New Delh	India	25.68				
11	1805	New Delh	India	25.3				
12	1806	New Delh	India	25.22				
13	1807	New Delh	India	24.97				
14	1808	New Delh	India					
15	1809	New Delh	India					
16	1810	New Delh	India					
17	1811	New Delh	India					
18	1812	New Delh	India					
19	1813	New Delh	India	24.56				
20	1814	New Delh	India	23.73				
21	1815	New Delh	India	24.09				
22	1816	New Delh	India	23.7				
23	1817	New Delh	India	23.86				

New Delhi Data

- **Cleaning the Data:**

I chose to execute this **project using Python** on a Jupyter Notebook. The first steps consisted of cleaning the city data, as it was apparent that some of the years had missing weather data. The mean of the avg_temp was calculated for New Delhi using the following code:

```
fill_data = NewDelhi_Data['avg_temp'].mean()
```

The mean was stored in the variable "fill_data". The missing data was then replaced with this mean using the following code:

```
NewDelhi_Data['avg_temp'].fillna(fill_data, inplace = True)
```

- **Calculating Moving Average:**

The next order of business involved making a new column in the "NewDelhi_Data" Data Frame named "Moving Average" and actually calculating the moving average, taking a period of 7 years. The following code shows the method used to calculate the moving average:

```
NewDelhi_Data['Moving Average'] = NewDelhi_Data.iloc[:,3].rolling(window=7).mean()
```

The built in Pandas function “rolling” was used to calculate the moving average, keeping the window (the period) as 7 years. The data frame looked as given below:

	year	city	country	avg_temp	Moving Average
0	1796	New Delhi	India	25.030000	NaN
1	1797	New Delhi	India	26.710000	NaN
2	1798	New Delhi	India	24.290000	NaN
3	1799	New Delhi	India	25.280000	NaN
4	1800	New Delhi	India	25.210000	NaN
5	1801	New Delhi	India	24.220000	NaN
6	1802	New Delhi	India	25.630000	25.195714
7	1803	New Delhi	India	25.380000	25.245714
8	1804	New Delhi	India	25.680000	25.098571
9	1805	New Delhi	India	25.300000	25.242857
10	1806	New Delhi	India	25.220000	25.234286
11	1807	New Delhi	India	24.970000	25.200000
12	1808	New Delhi	India	25.166269	25.335181
13	1809	New Delhi	India	25.166269	25.268934
14	1810	New Delhi	India	25.166269	25.238401

Similarly, moving average for the Global Data was calculated:

	year	avg_temp	Moving Average
0	1750	8.72	NaN
1	1751	7.98	NaN
2	1752	5.78	NaN
3	1753	8.39	NaN
4	1754	8.47	NaN
5	1755	8.36	NaN
6	1756	8.85	8.078571
7	1757	9.02	8.121429
8	1758	6.74	7.944286
9	1759	7.99	8.260000
10	1760	7.19	8.088571
11	1761	8.77	8.131429
12	1762	8.61	8.167143
13	1763	7.50	7.974286

Now, since New Delhi only had moving average for years after and including 1802, the row above the year 1802 in the Global Data and New Delhi Data were dropped so as to ensure Data Frames of equal size. The following code was used:

```
NewDelhi_Data_Cleaned = NewDelhi_Data[NewDelhi_Data['year']>1801]
```

```
Global_Data_Cleaned = Global_Data_Cleaned[Global_Data_Cleaned['year']>1801]
```

- **Plotting the data and drawing conclusions:**

The following code was used to plot the line graphs of the two data sets:

```
x = NewDelhi_Data_Cleaned['year']
y1 = NewDelhi_Data_Cleaned['Moving Average']
y2 = Global_Data_Cleaned['Moving Average']
```

```
fig = plt.plot(x, y1, lw=3, color ='blue' , label = 'New Delhi')
fig = plt.plot(x,y2, lw=3, color = 'orange' , label = 'World')
plt.xlabel('Year ', fontsize = 10)
plt.ylabel("Temp ("+chr(176)+'C)', fontsize = 10)
plt.legend(loc = 'best')
plt.savefig('Comparison.png')
```

Covariance Coefficient between the two data sets was also calculated using the following code:

```
std_NewDelhi = np.std(NewDelhi_Data_Cleaned['avg_temp'])
std_Global = np.std(Global_Data_Cleaned['avg_temp'])
```

```
cov1 = np.cov(NewDelhi_Data_Cleaned['avg_temp'],Global_Data_Cleaned['avg_temp'])
```

```
cov1
```

```
array([[0.31510803, 0.2249239 ],
       [0.2249239 , 0.30886117]])
```

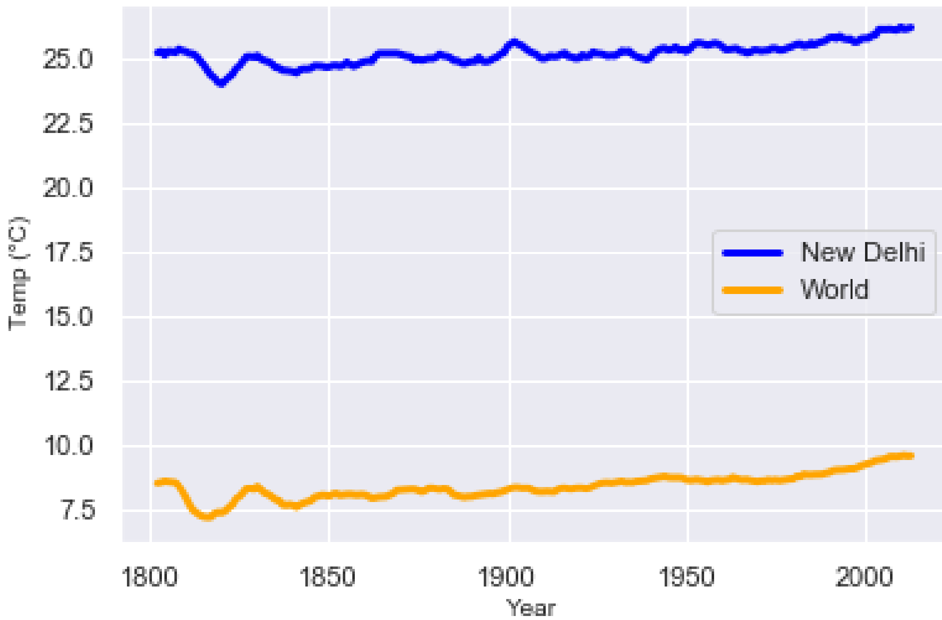
```
cov_coeff
```

```
array([[1.01484913, 0.72439863],
       [0.72439863, 0.99473026]])
```

As shown above, the Covariance Coefficient comes out to be **0.72.**

The final plot was as show below:

Figure 3 : Comparison between Global and New Delhi Weather Trends



Conclusions:

- New Delhi's average temperature has been significantly higher than that of the world. New Delhi's temperature over the past two centuries has hovered around 25° C while Global temperate has stay between 8.5-10° C. The primary reason for the difference is the proximity of New Delhi to the equator, which is characterized by extreme heat and low humidity.

- Both New Delhi and Global Data shows an upward trend in the average temperatures over the years, increasing rapidly after the 1900s. This can be attributed to the increasing industrialization and the resulting global warming from the exhausted green house gases.

- Both the trend lines show a sharp drop in temperatures during the 1820s, indicating severe winters during those years world-wide.

- The covariance coefficient of 0.72 and the trend lines show a strong positive relationship between New Delhi's temperatures and the temperatures globally.

- Global temperatures started rising significantly after the 1900s whereas New Delhi's temperature stayed relatively flat till the 1960s, indicating a slower rate of industrialization in India, which started picking up late in the 20th century.