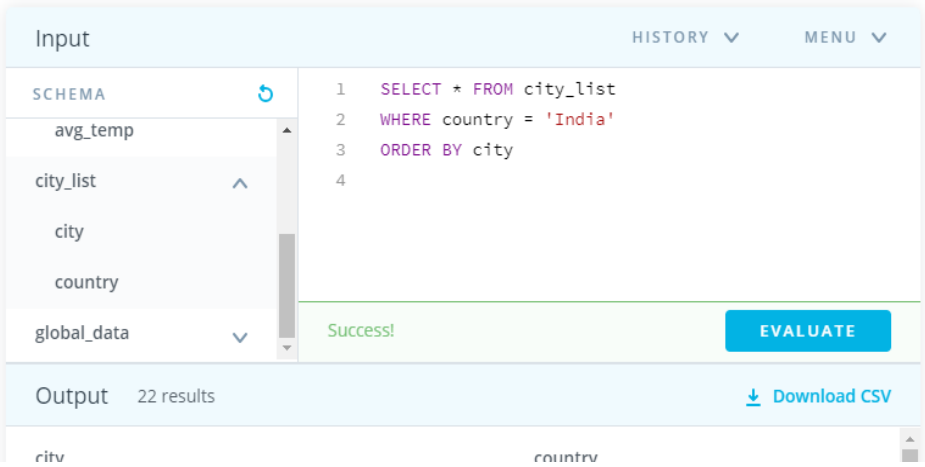
PROJECT 1- EXPLORING WEATHER TRENDS

The following is outline of steps taken to prepare the data and the visualization with a line chart:

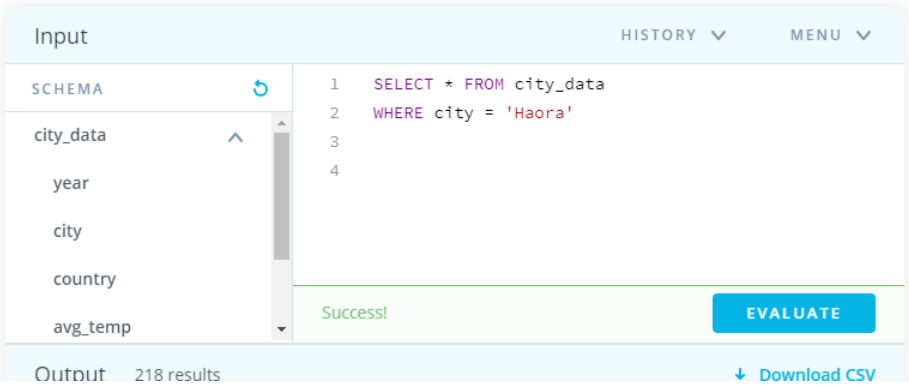
EXTRACTING DATA USING SQL AND EXCEL:

- I used the SQL Workspace provided to extract data from the 3 tables in the temperatures database, and then downloaded the results to csv files.

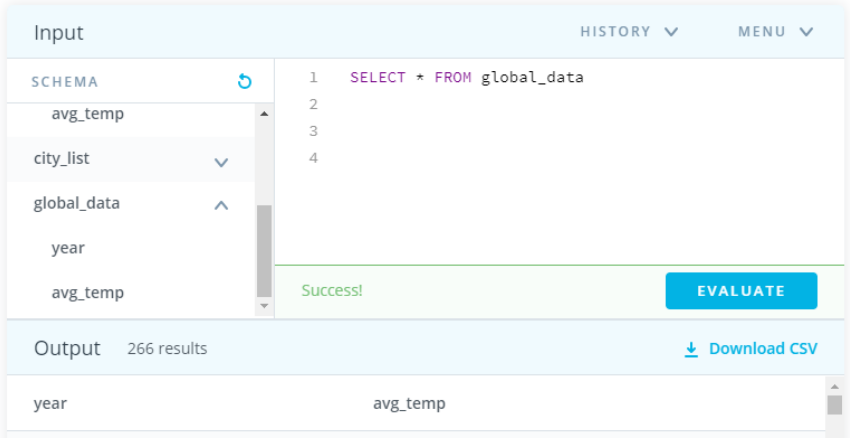
- Firstly, I ran the following SQL query to find the list of all cities, then downloaded the results to a csv file. From the downloaded file I selected the nearest city as 'Haora'.



- I then extracted the temperature data corresponding to the city Haora from the database city_data using the SQL query below, and downloaded the data to a csv file.
  The output file contained average temperature data from 1796-2013, with missing data for the period 1808-1812.



- I then extracted the global temperature data from the database global_data using the SQL query below, and downloaded the data to a csv file.
  The output file contained global average temperature data from 1750-2015.



PREPARING THE DATA:

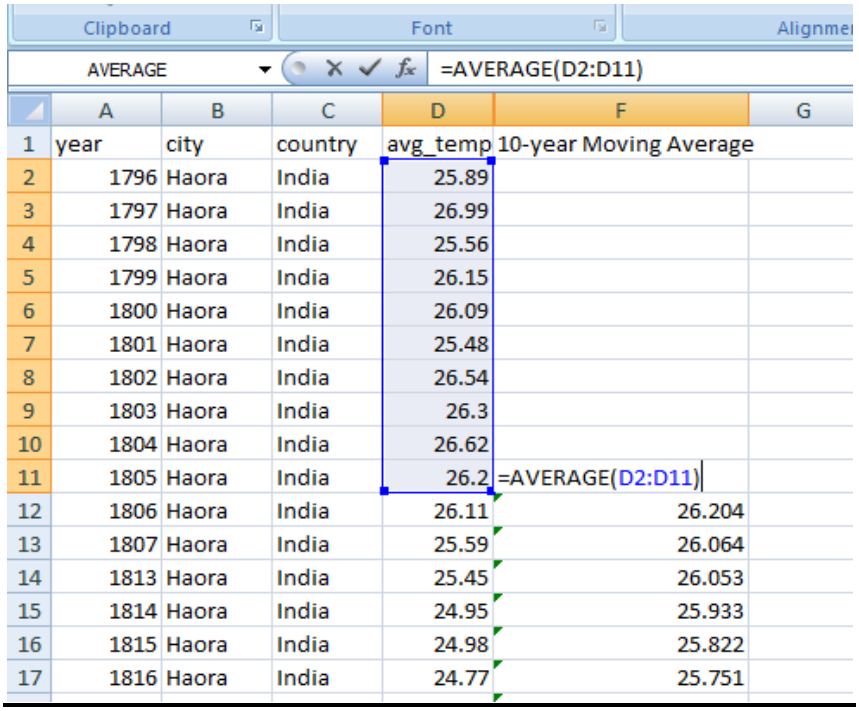
- As the global data was present for 1750-2015, whereas the nearest city data was available only from 1796 onwards, and again had missing records for the years 1808-1812, I therefore decided to include the data only from 1796 onwards for our analysis purpose. The years 1808-1812 for which the nearest city data is not available has been excluded from our analysis, as these years are fairly towards the beginning of our data range, and the available data excluding these rows is large enough for our purpose to analyze trends.

CALCULATING THE MOVING AVERAGE:

- Next, the moving average is calculated in order to smooth out the data and observe the long term trends. I decided to take a 10-year period for the moving average calculation.

- I calculated the 10-year moving average values for the nearest city and global temperature data separately, by opening the data in Excel sheets. A new column called Moving Average was created, and the AVERAGE() function was used to calculate the average temperature for the first ten years. This process was used to calculate the 10-year moving average for all the rows till the end of the dataset, by dragging down the formula in all the cells of this column.

  Snapshot of process for one of the excel sheets is shown below:



Clipboard | Font | Alignme

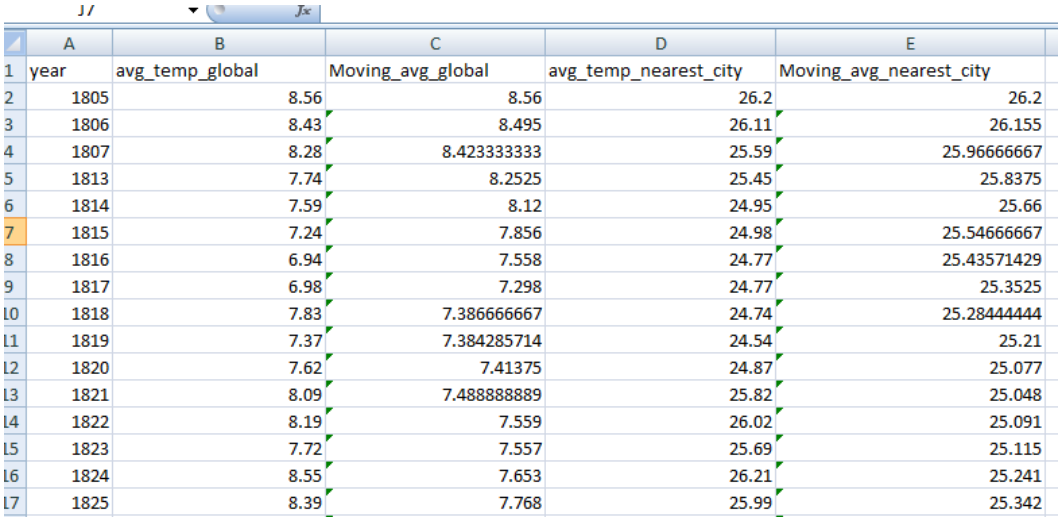
AVERAGE | =AVERAGE(D2:D11)

	A	B	C	D	F	G
1	year	city	country	avg_temp	10-year Moving Average	
2	1796	Haora	India	25.89		
3	1797	Haora	India	26.99		
4	1798	Haora	India	25.56		
5	1799	Haora	India	26.15		
6	1800	Haora	India	26.09		
7	1801	Haora	India	25.48		
8	1802	Haora	India	26.54		
9	1803	Haora	India	26.3		
10	1804	Haora	India	26.62		
11	1805	Haora	India	26.2	=AVERAGE(D2:D11)	
12	1806	Haora	India	26.11	26.204	
13	1807	Haora	India	25.59	26.064	
14	1813	Haora	India	25.45	26.053	
15	1814	Haora	India	24.95	25.933	
16	1815	Haora	India	24.98	25.822	
17	1816	Haora	India	24.77	25.751	

ANALYZING THE DATA:

- The data from both the sheets (nearest city and global) are now combined into a single spreadsheet. The information contained here includes the **year (1805-2015), avg_temp_global, Moving_avg_global, avg_temp_nearest_city and Moving_avg_nearest_city.**

  Snapshot of the final spreadsheet is given below:

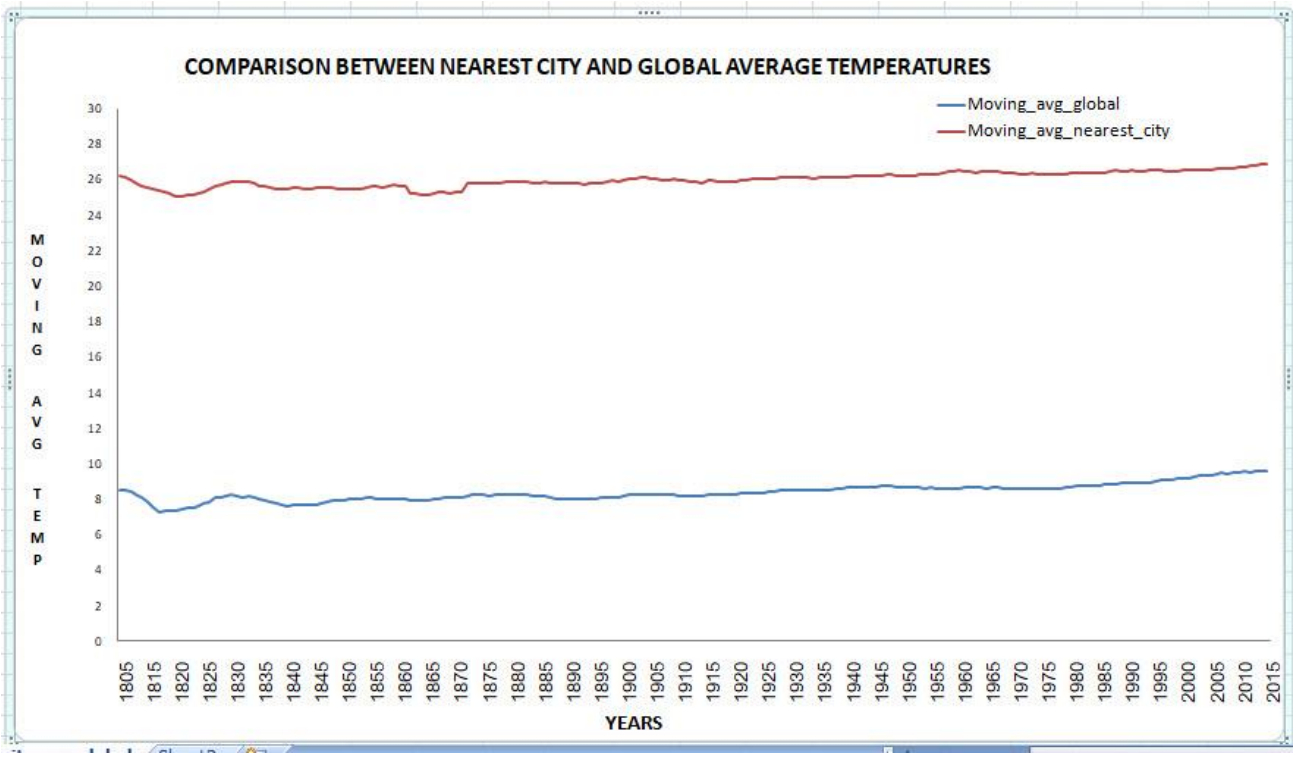


J7

	A	B	C	D	E
1	year	avg_temp_global	Moving_avg_global	avg_temp_nearest_city	Moving_avg_nearest_city
2	1805	8.56	8.56	26.2	26.2
3	1806	8.43	8.495	26.11	26.155
4	1807	8.28	8.423333333	25.59	25.96666667
5	1813	7.74	8.2525	25.45	25.8375
6	1814	7.59	8.12	24.95	25.66
7	1815	7.24	7.856	24.98	25.54666667
8	1816	6.94	7.558	24.77	25.43571429
9	1817	6.98	7.298	24.77	25.3525
10	1818	7.83	7.386666667	24.74	25.28444444
11	1819	7.37	7.384285714	24.54	25.21
12	1820	7.62	7.41375	24.87	25.077
13	1821	8.09	7.488888889	25.82	25.048
14	1822	8.19	7.559	26.02	25.091
15	1823	7.72	7.557	25.69	25.115
16	1824	8.55	7.653	26.21	25.241
17	1825	8.39	7.768	25.99	25.342

VISUALIZING THE DATA USING LINE CHART:

A line chart is then created that compares the nearest city's temperatures with the global temperatures. The moving average temperature is plotted against the year and the following visualization is obtained:



OBSERVATIONS:

- It is seen that the graphs depicting the nearest city moving average temperature and the global moving average temperature follows a similar trajectory, that is, they are roughly parallel in nature.

- There is a temperature difference of around 18 degrees between the nearest city average temperature and the global average temperature (the local temperature being hotter), and this difference has remained consistent over time.

- Both the local and global temperatures show an increasing trend, which indicates that the world is getting hotter over time, and average temperatures are increasing leading to global warming.

- It appears that between the years 1860-1865, the trend shows a decrease of temperature instead of the usual increasing trend for both the local and global temperatures, however this observation is more pronounced for the nearest city.

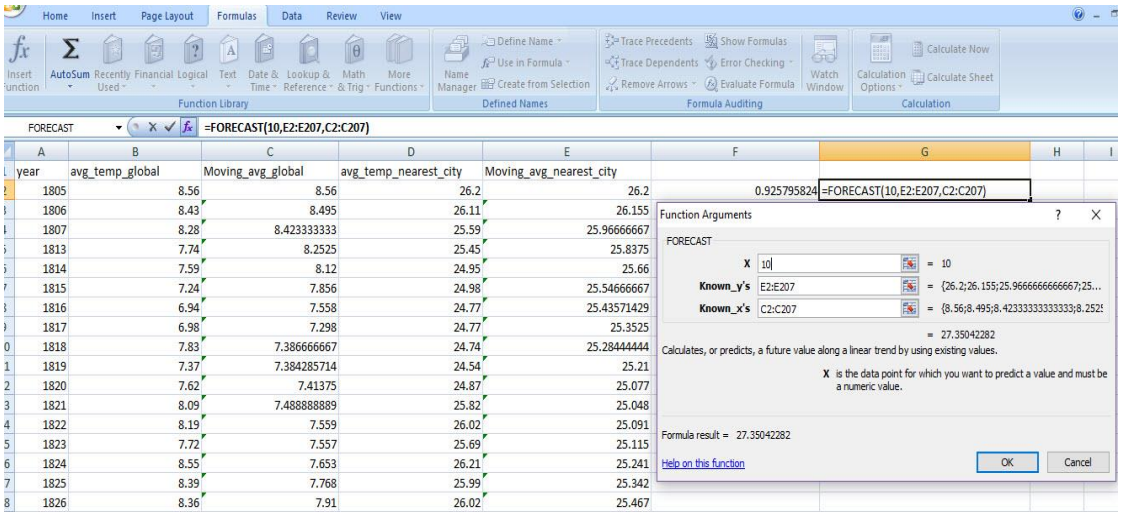
ADDITIONAL INSIGHTS FROM THE DATA:

- The **correlation** coefficient (a value between -1 and +1) tells us how strongly two variables are related to each other. I used the **CORREL** function from the 'Formula' tab in Excel to find the correlation coefficient between two variables - **Moving_avg_global** and **Moving_avg_nearest_city**. The **correlation coefficient** is obtained as **0.926** which indicates that the average global temperature and the average local temperature have a **strong positive correlation** with each other.

- The Microsoft Excel **FORECAST** function returns a prediction of a future value based on existing values provided. I used the built in **FORECAST** function in the 'Formula' tab in Excel to find the linear trend equation between the variables **Moving_avg_nearest_city (Known Ys)** and **Moving_avg_global (Known Xs).**The equation obtained is Y **= 18.92X**. Using this formula, we can predict any future value of Y for any given value of X.

  For example, if the value of X (Moving_avg_global) were to reach 10 degrees, then we can use this formula to predict the value of Y (Moving_avg_nearest_city) to be equal to 27.35 degrees.

  A snapshot of this is given below:



	A	B	C	D	E	F	G	H	I
	year	avg_temp_global	Moving_avg_global	avg_temp_nearest_city	Moving_avg_nearest_city				
2	1805	8.56	8.56	26.2	26.2	0.925795824	=FORECAST(10,E2:E207,C2:C207)		
3	1806	8.43	8.495	26.11	26.155				
4	1807	8.28	8.423333333	25.59	25.96666667				
5	1813	7.74	8.2525	25.45	25.8375				
6	1814	7.59	8.12	24.95	25.66				
7	1815	7.24	7.856	24.98	25.54666667				
8	1816	6.94	7.558	24.77	25.43571429				
9	1817	6.98	7.298	24.77	25.3525				
10	1818	7.83	7.386666667	24.74	25.28444444				
11	1819	7.37	7.384285714	24.54	25.21				
12	1820	7.62	7.41375	24.87	25.077				
13	1821	8.09	7.488888889	25.82	25.048				
14	1822	8.19	7.559	26.02	25.091				
15	1823	7.72	7.557	25.69	25.115				
16	1824	8.55	7.653	26.21	25.241				
17	1825	8.39	7.768	25.99	25.342				
18	1826	8.36	7.91	26.02	25.467				

CONCLUSION:

Thus we have extracted the data from the databases, prepared the data for analysis and used it to create visualizations and make observations and draw inferences from it.