

# Project - 1

## Explore Weather Trends

May 10, 2020

Udacity - Data Analyst Nanodegree

- *Kuldeep Srivastav*

### Overview

In this project, I will analyze local and global temperature data and compare the temperature trends where I live to overall global temperature trends.

### Instructions

- **Extract the data** from the database and export to CSV file.
- **Open up the CSV** in whatever tool (I will use Google sheets).
- **Create a line chart** that compares your city's temperatures with the global temperatures.
- **Make observations** about the similarities and differences between the world averages and your city's averages, as well as overall trends.

## Goals

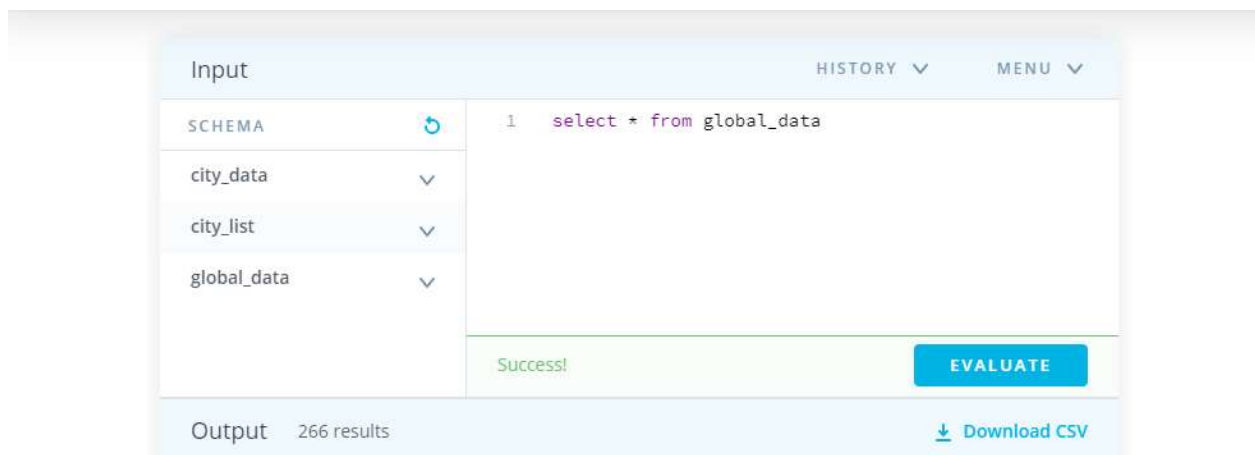
1. Extraction of data from the database using **Sql** and export to **CSV** file.
2. Making a chart visualization based on extracted data using **Google Sheets**.
3. Observation based on chart

## Tools Used:

1. **SQL** : To extract the data from the database
2. **Google Sheets** :
  - a. To calculate *moving average* of city vs global temperature.
  - b. To plot a line chart.

## Extracting data from the database using SQL:

### Step 1: To extract the global data



We have a total of 266 results available

## Step 2 : To see the available cities for country India

The screenshot shows a SQL query editor interface. On the left, under the 'Input' tab, there is a 'SCHEMA' section with a list of tables: 'city\_data', 'city\_list', 'city', and 'country'. The 'city\_list' table is selected. Below the schema, there is a 'global\_data' section. In the center, the SQL query is written as follows:

```
1 select * from city_list
2 where country = 'India'
```

Below the query, there is a green 'Success!' message and a blue 'EVALUATE' button. On the right, there are tabs for 'HISTORY' and 'MENU'. Below the query editor, there is an 'Output' section showing '22 results' and a 'Download CSV' button. The output is displayed in a table with two columns: 'city' and 'country'.

city	country
Agra	India
Ahmadabad	India
Allahabad	India

We have a total of 22 results including AGRA. So as we got our nearest big city.

Now we will Alter the columns "avg\_temp" since both the required data i.e. city\_data

And global\_data contains a column with same name.

4

### Step 3 : Alter the column name (rename avg\_temp)

Input

HISTORY ▼ MENU ▼

SCHEMA

city\_data ▼

city\_list ▼

global\_data ▼

1 `ALTER TABLE city_data RENAME COLUMN avg_temp to City_AvgTemp;`

2 `ALTER TABLE global_data RENAME COLUMN avg_temp to Global_AvgTemp;`

Success!

EVALUATE

Output

No data to download

### Step 4 : Joining the two tables to obtain the relevant data

Input

HISTORY ▼ MENU ▼

SCHEMA

city\_data ▼

city\_list ▼

global\_data ▼

1 `SELECT global_data.year, global_data.Global_AvgTemp, city_data.City_AvgTemp`

2 `FROM global_data`

3 `JOIN city_data`

4 `ON global_data.year = city_data.year`

5 `WHERE city LIKE 'Agra';`

Success!

EVALUATE

Output 218 results [Download CSV](#)

year	global_avgtemp	city_avgtemp
1796	8.27	25.05
1797	8.51	26.71
1798	8.67	24.19

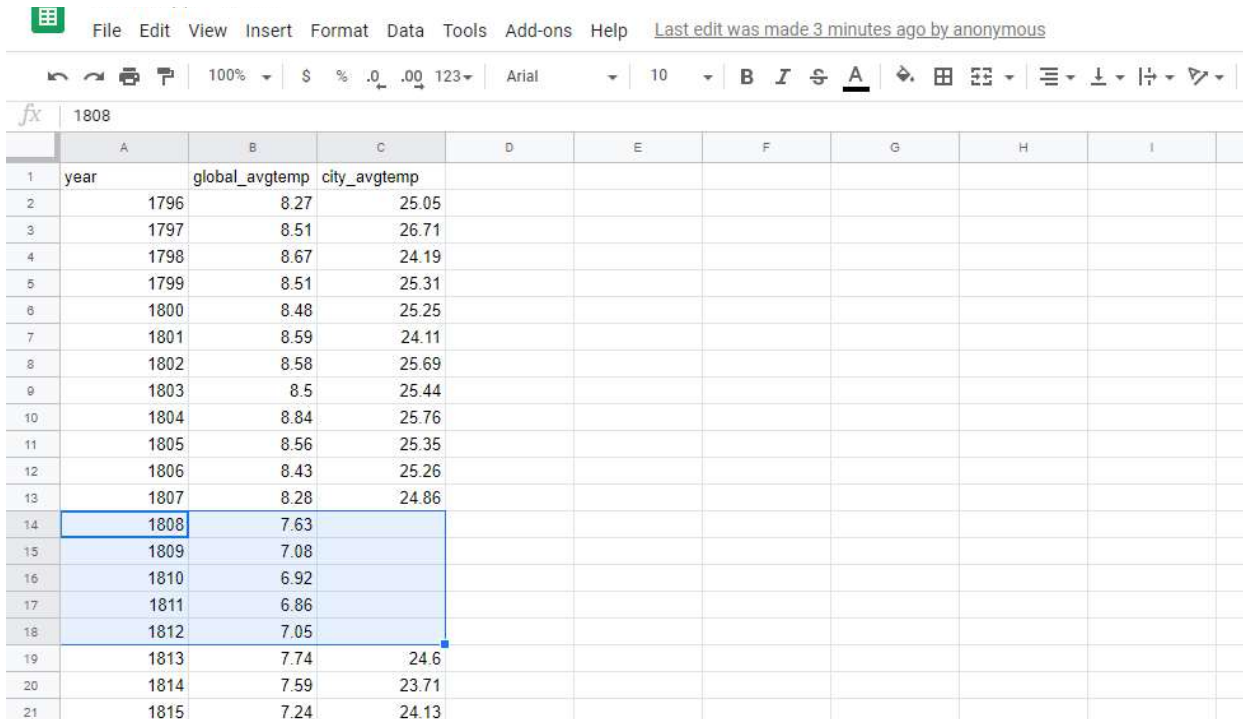
So we finally got the required data with 218 results starting from year 1796 (same for both).

Now I have got an option of downloading the file as CSV format. Downloaded as "results.csv".

& hence the extraction part is complete.

## Moving Average & Plotting The Graph:

Step 1 : First of all open the CSV file using Google Sheets  
And delete the missing year data by simply selecting it.



	A	B	C	D	E	F	G	H	I
1	year	global_avgtemp	city_avgtemp						
2	1796	8.27	25.05						
3	1797	8.51	26.71						
4	1798	8.67	24.19						
5	1799	8.51	25.31						
6	1800	8.48	25.25						
7	1801	8.59	24.11						
8	1802	8.58	25.69						
9	1803	8.5	25.44						
10	1804	8.84	25.76						
11	1805	8.56	25.35						
12	1806	8.43	25.26						
13	1807	8.28	24.86						
14	1808	7.63							
15	1809	7.08							
16	1810	6.92							
17	1811	6.86							
18	1812	7.05							
19	1813	7.74	24.6						
20	1814	7.59	23.71						
21	1815	7.24	24.13						

Step 2 : Now find the average using the command

**= AVERAGE(B2:C11)** , to see the moving average for the first 10 years ( to get smooth line chart)

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

100% \$ % .0 .00 123

**B I S A**

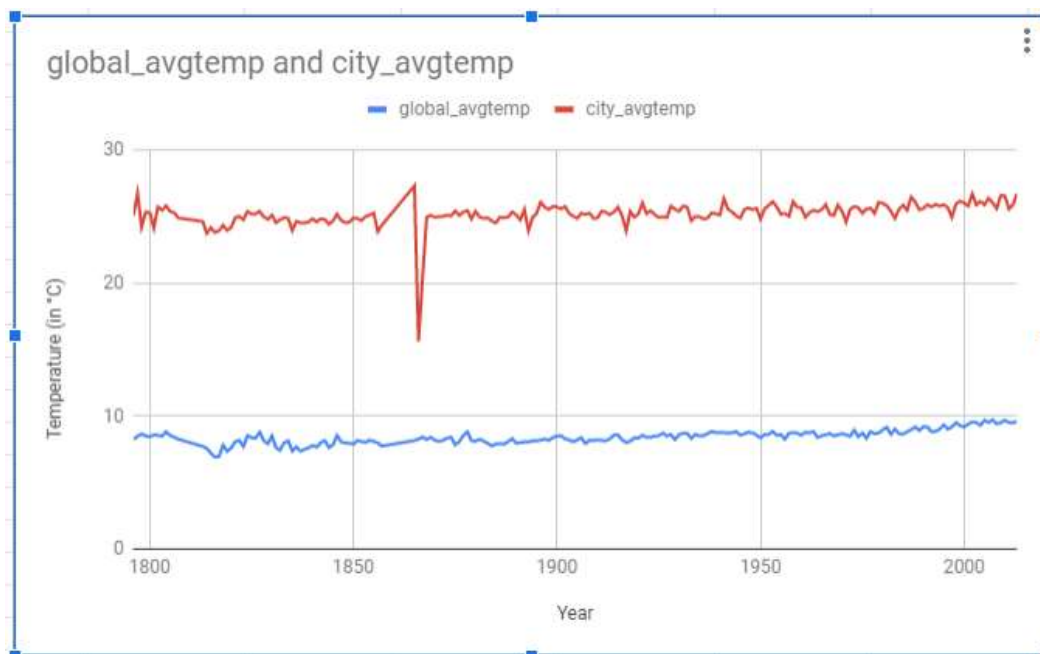
**fx** =AVERAGE(B2:C11)

	A	B	C	D	E	F	G
1	year	global_avgtemp	city_avgtemp				
2	1796	8.27	25.05				
3	1797	8.51	26.71				
4	1798	8.67	24.19				
5	1799	8.51	25.31				
6	1800	8.48	25.25				
7	1801	8.59	24.11				
8	1802	8.58	25.69				
9	1803	8.5	25.44				
10	1804	8.84	25.76				
11	1805	8.56	25.35				
12	1806	8.43	25.26				
13	1807	8.28	24.86				

D11 16.9185 x

=AVERAGE(B2:C11)

Step 3 : Now plot a line chart for Global Average Temperature vs City Average Temperature.



## Observations :

**According to the graph, following observations may be concluded.**

1. The city Agra's temperature is much hotter than the overall global temperature.
2. There is a big difference between the Agra's temperature to that of the global temperature throughout the time frame.
3. Even though it is not increasing rapidly but the city's as well as the global temperature is increasing in the entire time frame which can be due to increase in Global Warming.
4. From the above, we can clearly see the temperature in the year 1865 (27°C approx) has a peak while suddenly it falls to 15°C in the next year i.e. 1866.
5. Looking to the graph, it is clear that the temperature of the world is on constant rise.
6. I have also observed the data for the 50 years and found that the chart is smoother as compared to that of the 10 years data.

This conclude my observations for the project and data provided.

*Thank You...*