Kheireddine Douli

Explore Weather Trends Project

March 31th, 2020

Data Analyst Nanodegree

Project Overview:

In this project, we will analyze local and global temperature data, and compare weather trends in Algiers with global temperature trends. We will analyze the data, create a data visualization and make some patterns observations

Project Instructions:

1. Extract the data from the database using SQL
2. Calculate the moving average temperature for local and global data using Google spreadsheet
3. Create a line chart to compare similarities and differences between temperature trends in the world and in the city of Algiers
4. Make observations about the trends in the chart
5. Make the project stand out!

Project solution outline:

Extract the data:

To pull the data from the database, three tables are provided: city_list, city_data and global_data, one way is to pull the local data from the city_data table in csv file, and extract the global data from the global_data table in another csv file, then either MS Excel or Google spreadsheet can be used to merge the columns in a single sheet, the SQL query to pull local data look like this:

```
SELECT cd.year,cd.avg_temp
FROM city_data cd
WHERE city = 'Algiers'
AND country = 'Algeria'
```

the database contain only the city Algiers for Algeria so we can remove the' city' condition but we preferred to add it to make the query more readable.
And this query is for the global data :

```
SELECT gd.year,gd.avg_temp global_data_avg_temp
FROM global_data gd
```

Another way to do it which more optimized and it can save us more work on Google spreadsheet is to join the city_data and global_data in one query using the attribute' year'.
We had used this one to extract the data:

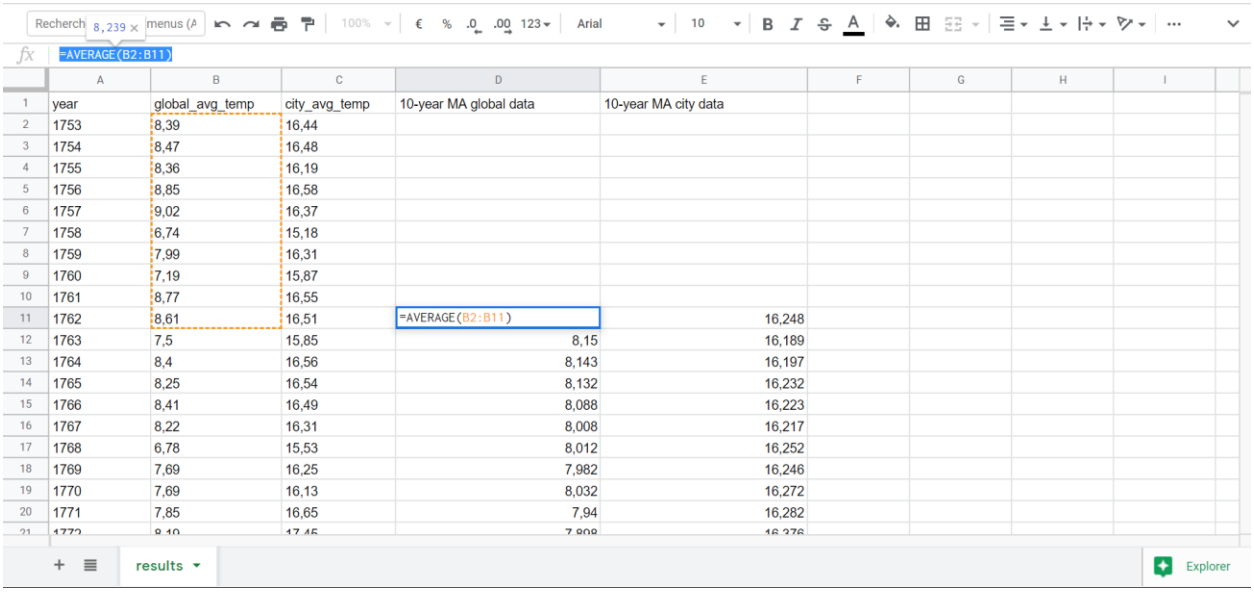
```
SELECT cd.year,gd.avg_temp global_avg_temp,cd.avg_temp city_avg_temp
FROM city_data cd
JOIN global_data gd
ON cd.year = gd.year
WHERE cd.city = 'Algiers'
ORDER BY year
```

Here the output dataset

Figure 1 output dataset

Calculate the moving Average:

At this point we need to import our dataset into the Google spreadsheet, then use the `AVERAGE` formula to measure the moving average for local and global data, but this time we need to set a period of years to measure it, a good reason for this is to minimize the noise but not make the curves too flat (Y, n.d.).
For the formula we include 10 cells like this `=AVERAGE(B2:B11)`, then we can drag all to the buttom until the last cell to apply the formula to other cells, we do the same for columns in the sheet, Moreover, the final dataset sheet look like this:



	A	B	C	D	E
1	year	global_avg_temp	city_avg_temp	10-year MA global data	10-year MA city data
2	1753	8,39	16,44		
3	1754	8,47	16,48		
4	1755	8,36	16,19		
5	1756	8,85	16,58		
6	1757	9,02	16,37		
7	1758	6,74	15,18		
8	1759	7,99	16,31		
9	1760	7,19	15,87		
10	1761	8,77	16,55		
11	1762	8,61	16,51	=AVERAGE(B2:B11)	16,248
12	1763	7,5	15,85	8,15	16,189
13	1764	8,4	16,56	8,143	16,197
14	1765	8,25	16,54	8,132	16,232
15	1766	8,41	16,49	8,088	16,223
16	1767	8,22	16,31	8,008	16,217
17	1768	6,78	15,53	8,012	16,252
18	1769	7,69	16,25	7,982	16,246
19	1770	7,69	16,13	8,032	16,272
20	1771	7,85	16,65	7,94	16,282

Figure 2 moving average calculation on spreadsheet

Create a Line Chart:

Once we have the dataset ready, we can visualize the trends by creating a simple line chart that shows the changing in temperature in Algiers city and the global temperature in Celsius.
To do so we had followed a YouTube tutorial (Boulmetis, s.d.). And we get this awesome chart:

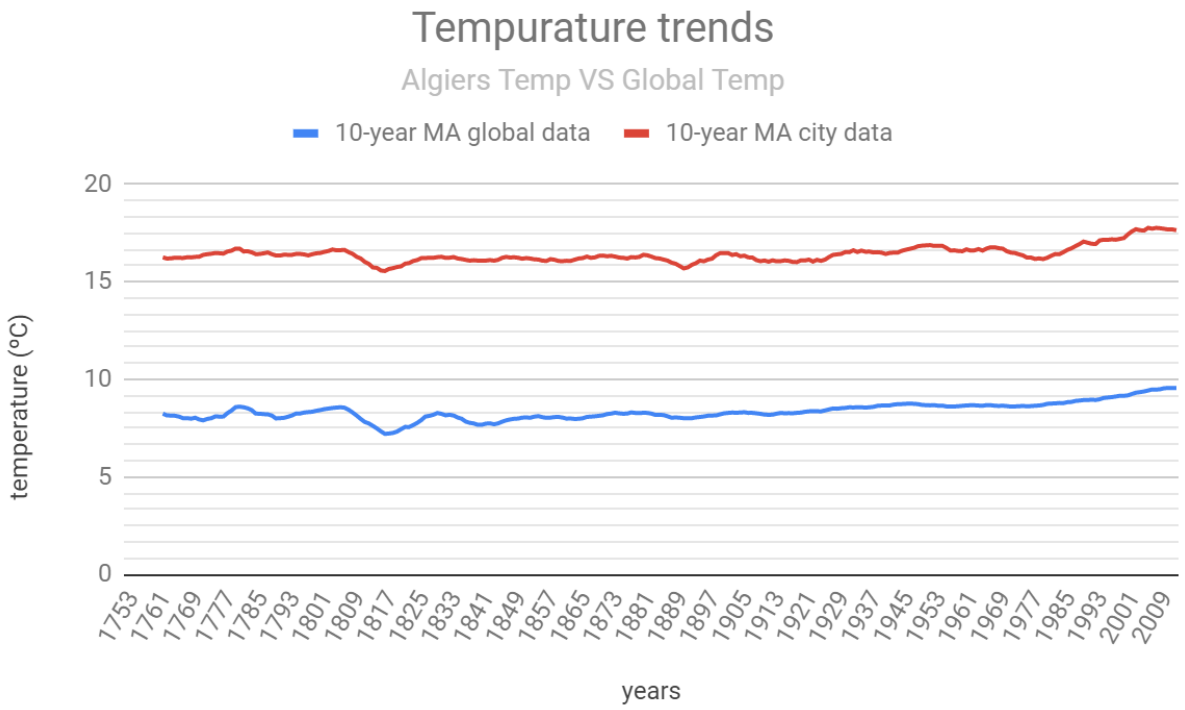


Figure 3 Temperature Trends Algiers vs global

Observation from the trends chart

- Algiers city is hotter compared to the global temperature
- Algiers Temperature goes up and down but without any significant peaks and most the of time it climb slowly (see correlation coefficient below
- From 1841 The global temperature starts increasing slowly
- The overall trend seems to be as the world is getting hotter over time.

Making the Project Stand Out!

What Is the Correlation Coefficient?

The correlation coefficient is a statistical measure of the strength of the relationship between the relative movements of two variables. The values range between -1.0 and 1.0.
There are several types of correlation coefficients, but the one that is most common is the **Pearson correlation** (r), this measures the strength and direction of the linear relationship between two variables.(GANTI, s.d.)

In our case the correlation coefficient could be calculated to determine the level of correlation between the temperature in Celsius which is the dependent data and the time by years which is the independent data, and we ca easily see from the chart that both the local and global temperature had a correlation that is positive.

In the image below, we show how to calculate the correlation coefficient for both local and global data:

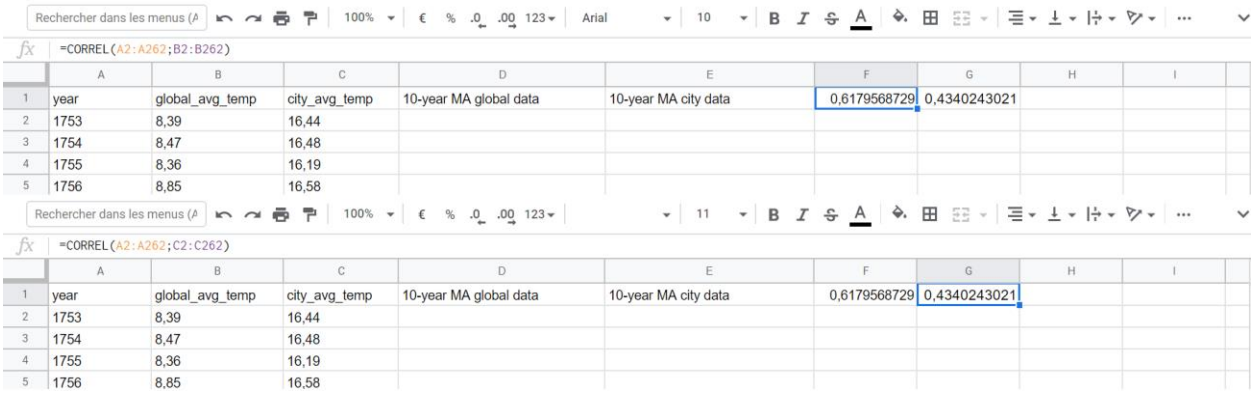


Figure 4 Correlation Coefficient Formula

And we can add a linear line that describe the correlation coefficient as it showed in the chart below:

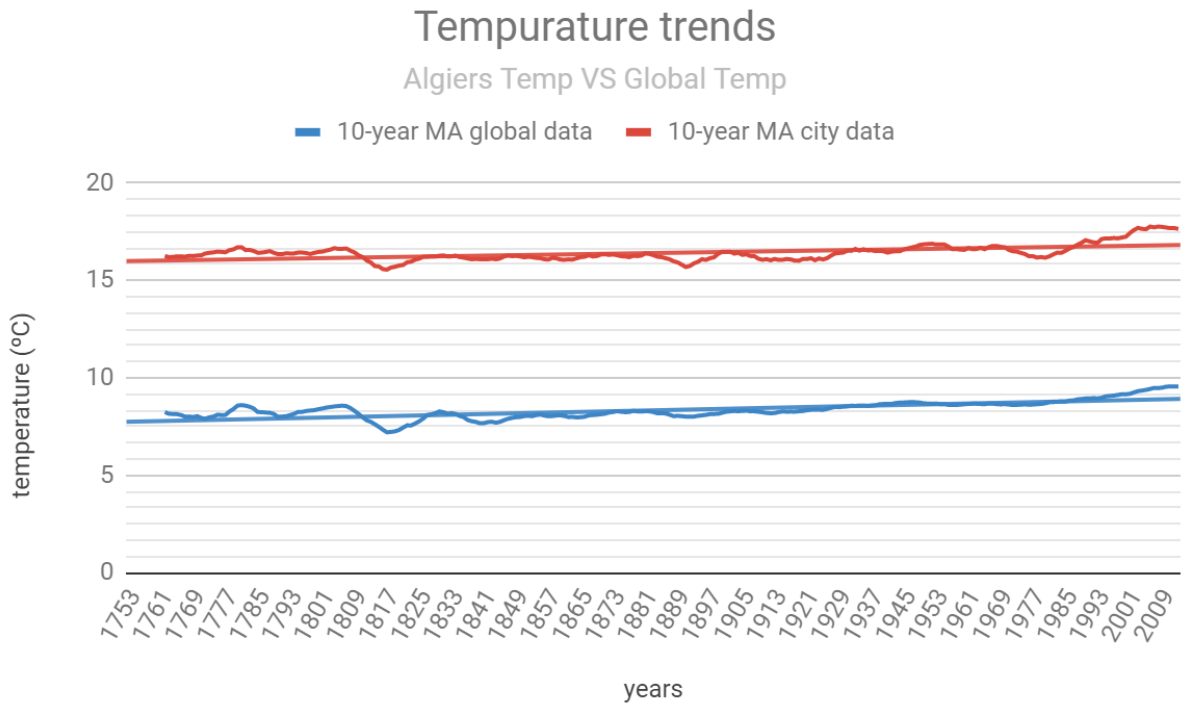


Figure 5 Linear Correlation of Temperature Trends

Estimation of average temperature city based on the average global temperature

We assume that we can estimate the average temperature of any city in the world based on the gl obal average temperature, but this calculation cannot be reliable because each city has its own ge ographical location and there are many factors influencing the temperature in that area, such as m anufacturing facilities, the population and many others.(team, 2019), (Newsbeat, s.d.)

Bibliography

Boulmetis, M. (n.d.). Retrieved from youtube: https://www.youtube.com/watch?v=lxkPWeCXm5I

GANTI, A. (n.d.). Retrieved from https://www.investopedia.com/terms/c/correlationcoefficient.asp

Newsbeat, B. N. (n.d.). *Newsbeat*. Retrieved from https://www.bbc.com: https://www.bbc.com/news/newsbeat-48947573

team, B. V. (2019, July 31). Retrieved from https://www.bbc.co.uk/news/resources/idt-985b9374-596e-4ae6-aa04-7fbcae4cb7ee

Y, H.-y. (n.d.). Retrieved from https://knowledge.udacity.com/questions/86248