UDACITY

DATA ANALYST NANODEGREE

GABRIEL SALVADOR BARÇANTE BARROS

EXPLORE WEATHER TRENDS

Brazil

2020

Summary

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

Step 1: Extracting the data

First, I needed to find my city or somewhere close to me. So, I wrote the following SQL query:

```
SELECT COUNT(*)
FROM city_list
WHERE country = 'Brazil' AND city = 'Rio De Janeiro';
```

The query outputted the following result, which confirm that there is my city, Rio De Janeiro in Brazil, in the Database.

Output 1 results

count
1

After confirming the city that I had to study, I had to extract the data from my city and global. So, I wrote the following SQL query to output the data:

```
SELECT c.year,
       c.country,
       c.city,
       c.avg_temp city_avg_temp,
       g.avg_temp global_avg_temp
FROM city_data c
JOIN global_data g
ON c.year = g.year
WHERE c.country = 'Brazil' AND c.city = 'Rio De Janeiro';
```

Finally, I could export the results to a CSV file. The following image provides a preview of what the resulting table is (opened in Microsoft Excel):

year	country	city	city_avg_temp	global_avg_temp
1832	Brazil	Rio De Janeiro	23.05	7.45
1833	Brazil	Rio De Janeiro	24.11	8.01
1834	Brazil	Rio De Janeiro	23.27	8.15
1835	Brazil	Rio De Janeiro	22.73	7.39
1836	Brazil	Rio De Janeiro	22.91	7.70
1837	Brazil	Rio De Janeiro	22.29	7.38
1838	Brazil	Rio De Janeiro	22.81	7.51
1839	Brazil	Rio De Janeiro	22.54	7.63
1840	Brazil	Rio De Janeiro	23.32	7.80
1841	Brazil	Rio De Janeiro	22.97	7.69

Step 2: Wrangle the data

After importing the CSV file into Microsoft Excel, I could start work with the data.

While I was assessing the data, I identified some problems in the city temperature column. I have 7 missing values, from 1844 to 1850. So, I needed to clean my data, and I had two choices:

1. I could delete these years from the global temperature column;
2. I could fill the missing values with the mean of the city temperature column.

I chose the second option, because if I had deleted the years from 1844 to 1850, the years before 1844 would have a big gap, which would impact the calculation of the moving averages. So, I filled the missing values with 23.79°C, which is the mean of my city's temperature column.

Step 3: Calculate the Moving Averages

After the data were ready, I calculated the **10 years Moving Averages** for the city temperatures and the global temperatures, in order to smooth out the lines to make it easier to observe long term trends.

I used the "MÉDIA" (Portuguese for Average) formula in Excel to calculate the moving averages, as follows:

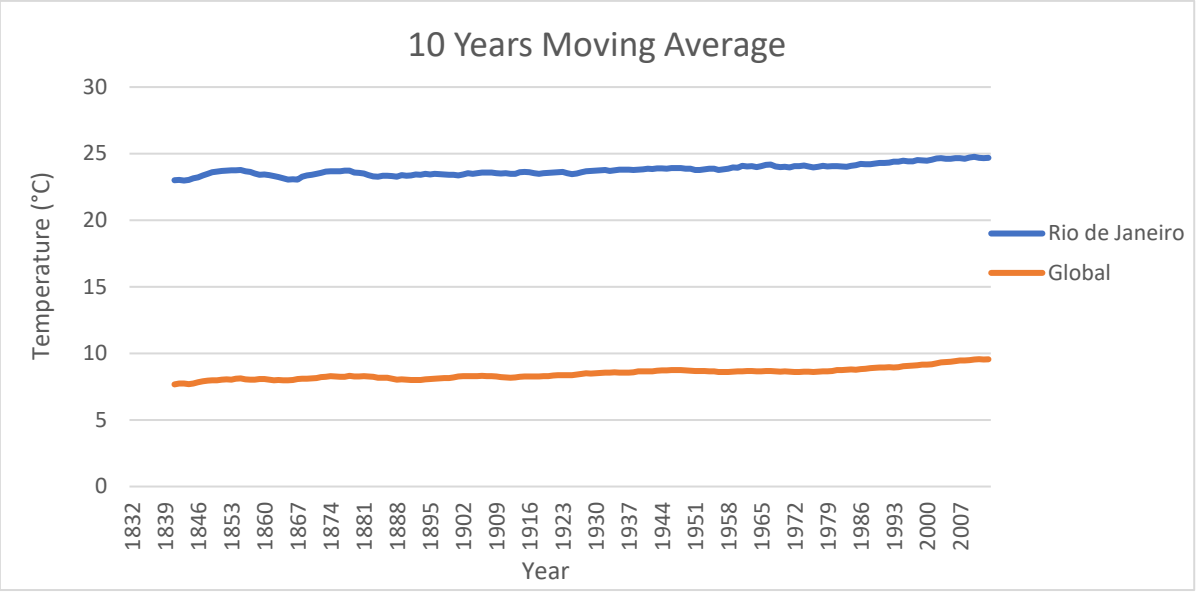
year	country	city	city_avg_temp	10-year MA City
1832	Brazil	Rio De Janeiro	23.05	
1833	Brazil	Rio De Janeiro	24.11	
1834	Brazil	Rio De Janeiro	23.27	
1835	Brazil	Rio De Janeiro	22.73	
1836	Brazil	Rio De Janeiro	22.91	
1837	Brazil	Rio De Janeiro	22.29	
1838	Brazil	Rio De Janeiro	22.81	
1839	Brazil	Rio De Janeiro	22.54	
1840	Brazil	Rio De Janeiro	23.32	
1841	Brazil	Rio De Janeiro	=MÉDIA(D2:D11	
1842	Brazil	Rio De Janeiro	MÉDIA(**núm1**; [núm2]; ...)	23.036
1843	Brazil	Rio De Janeiro	23.55	22.98

Then, I continued this process all the way to the end of the dataset for the city values and the global values. The following image provides a preview of what the resulting table is after calculating the moving averages:

year	country	city	city_avg_temp	global_avg_temp	10-year MA City	10-year MA Global
1832	Brazil	Rio De Janeiro	23.05	7.45		
1833	Brazil	Rio De Janeiro	24.11	8.01		
1834	Brazil	Rio De Janeiro	23.27	8.15		
1835	Brazil	Rio De Janeiro	22.73	7.39		
1836	Brazil	Rio De Janeiro	22.91	7.7		
1837	Brazil	Rio De Janeiro	22.29	7.38		
1838	Brazil	Rio De Janeiro	22.81	7.51		
1839	Brazil	Rio De Janeiro	22.54	7.63		
1840	Brazil	Rio De Janeiro	23.32	7.8		
1841	Brazil	Rio De Janeiro	22.97	7.69	23	7.671
1842	Brazil	Rio De Janeiro	23.41	8.02	23.036	7.728
1843	Brazil	Rio De Janeiro	23.55	8.17	22.98	7.744

Step 3: Data Visualization

Now that I have the moving averages, I can create a line chart, in Excel, that compares my city's temperatures with the global temperatures, using the 10 years moving averages for each. The resulting visualization can be seen bellow.



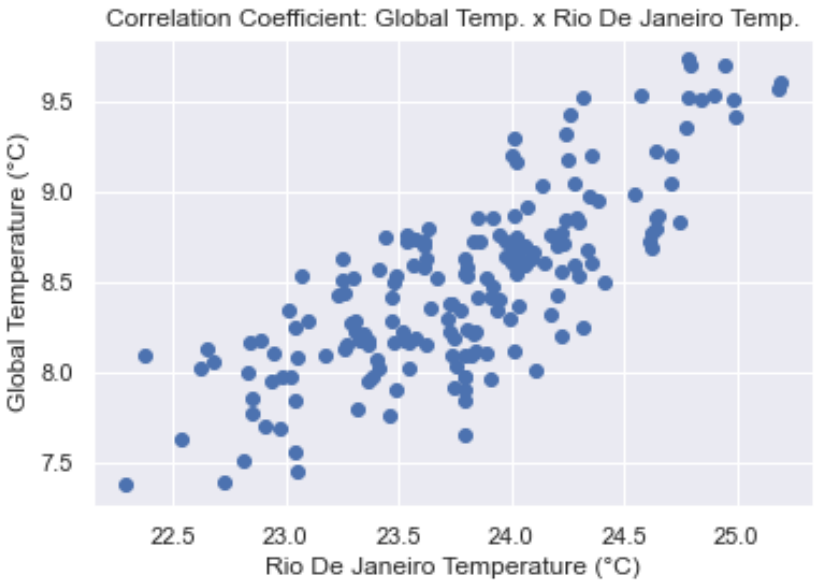
Step 4: Observations

1. The city of Rio De Janeiro, Brazil, has been hotter than the global temperature in the time period 1832-2013. This trend is consistent as the global line stays in the range of 7 °C and 10 °C, whereas the Rio de Janeiro line stays in the range of 23 °C and 25 °C, well above the global line.

2. The average difference between the global temperature and my city temperature is, approximately, 15 °C consistently over the time period 1832-2013. In which, both have been getting warmer, approximately, 2.5 °C on that period of time.

3. The moving averages are a great way to see the trend in the data. However, when we use that, we lost the fluctuation of it. So, in order to see how much each point varies from the mean of the points, we can calculate the standard deviation (std).
   The std for the Rio De Janeiro data is, approximately, 0.585, whereas the std for the global temperature is, approximately, 0.496. Both values are small, what demonstrate that the temperature values are not very spread out.

4. In order to see the strength of the relationship between the Rio De Janeiro's temperature and the global temperature, we can calculate the correlation coefficient. The Pearson's coefficient between the city temperature and the global temperature is 0.78, which means they have a positive correlation, that is, the global temperature tends to increase as the city temperature increases, as seen below in the graph.


Correlation Coefficient: Global Temp. x Rio De Janeiro Temp.
Global Temperature (°C)
9.5
9.0
8.5
8.0
7.5
22.5
23.0
23.5
24.0
24.5
25.0
Rio De Janeiro Temperature (°C)