PROJECT 1
EXPLORE WEATHER TRENDS

Data Analyst Nanodegree Program – Udacity

Abstract

In this project, I will analyze Munich city in Germany and global temperature data to compare the temperature trends in Munich to overall global temperature trends. I extract datasets by using SQL, then analyze and visualize data via Excel with the moving average method. After visualizing data, I find that Munich city is cooler on average compared to global average. However, it is getting hotter both in Munich and globally over more than three centuries. Moreover, I also compare temperature data for two other cities – Berlin and Hamburg – in Germany to Munich and global data. I conclude that the global warming trend is accelerating since the average global temperature is continuously rising over the time.

Phan Le Thanh Truc

12 June, 2021

1. Summary

In this project, I will analyze local and global temperature data and compare the temperature trends of Munich city in Germany to overall global temperature trends.

2. Extract Data from Database

Firstly, I need to find the closest big city to where I live. I am living in Germany, so I wrote an SQL query as below to retrieve the list of cities in Germany.

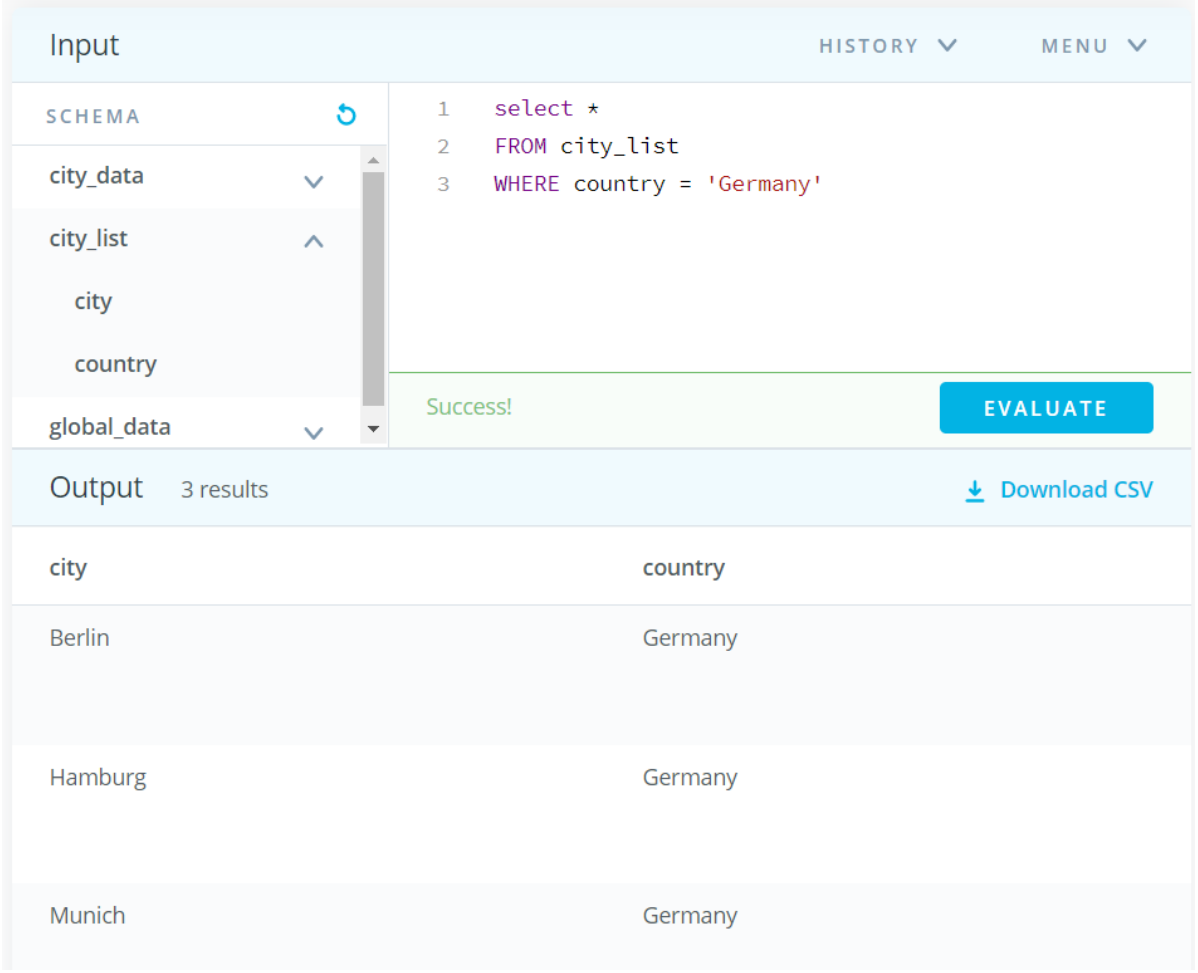

Figure 1. SQL Query to extract local city list

As can be seen from the output, three results appear for Germany, including Berlin, Hamburg and Munich. In this report, Munich is chosen as the local city to compare with global data because it is the closet city to where I currently reside in. However, I will extract data from all three cities for the later comparison between their average temperatures and global average temperature.

2.1. Data Exploration

Before deciding which data to be retrieved for a final dataset, I first explored the database. I extracted data of all three cities in Germany from the table city_data and global data from the table global_data to see how many observations each has as well as whether there are any null values. Table 1 below shows SQL query and output description for each data extract.

Table	Data	SQL Query	Output Description
city_data	to extract the average temperatures of Munich city	`SELECT year, avg_temp FROM city_data WHERE country = 'Germany' and city = 'Munich'`	▪ 271 results from 1743 – 2013. ▪ 4 null values for 1746 – 1749
	to extract the average temperatures of Belin city	`SELECT year, avg_temp FROM city_data WHERE country = 'Germany' and city = 'Berlin'`	
	to extract the average temperatures of Hamburg city	`SELECT year, avg_temp FROM city_data WHERE country = 'Germany' and city = 'Hamburg'`	
global_data	to extract the average global temperature	`SELECT * FROM global_data`	▪ 266 results from 1750 – 2015 ▪ No null values

Table 1. SQL query and output description for each data extract

According to Table 1, there are 271 observations for the cities compared to 266 for the global as the cities' data is available for a larger range of years. Moreover, in each city dataset, there are four null values for the years of 1746 – 1749 in 'avg_temp' column. Thus, four rows of those years will be excluded.

In order to provide a tidier and more accurate dataset for comparison, I proceed with choosing the range of common years only (**1750 – 2013**), which also contains **all non-null values**, for cities and global data.

Hence, the combined dataset will contain the followings.

- Year (1750 – 2013)
- Munich average temperature
- Berlin average temperature
- Hamburg average temperature
- Global average temperature

2.2. Extract the combined data

Below is a SQL query to extract the combined temperature data of Munich, Berlin, Hamburg cities and Global, then export to CSV file called 'weather_trends.csv'.

```
SELECT munich.year,
        munich.avg_temp AS munich_avg_temp,
        berlin.avg_temp AS berlin_avg_temp,
        hamburg.avg_temp AS hamburg_avg_temp,
        global.avg_temp AS global_avg_temp
FROM city_data munich
    LEFT JOIN city_data berlin ON munich.year = berlin.year
    LEFT JOIN city_data hamburg ON munich.year = hamburg.year
    LEFT JOIN global_data global ON munich.year = global.year
WHERE munich.country = 'Germany' and munich.city = 'Munich'
        and berlin.country = 'Germany' and berlin.city = 'Berlin'
        and hamburg.country = 'Germany' and hamburg.city = 'Hamburg'
        and global.avg_temp IS NOT NULL
```

After running the SQL Query above, I get an output with 264 results and no missing values as shown in Figure 2 below.

Output 264 results Download CSV

year	munich_avg_temp	berlin_avg_temp	hamburg_avg_temp	global_avg_temp
1750	5.40	9.83	9.31	8.72
1751	5.54	9.75	8.94	7.98
1752	0.53	4.84	4.65	5.78
1753	4.61	8.72	8.12	8.39

Figure 2. SQL query output of cities' and global average temperature data (1750 – 2013)

3. Data analysis and manipulation

3.1. Data analysis

The descriptive statistics of the dataset that I have collected is now provided in Table 2 below.

	year	*munich_avg_temp*	*berlin_avg_temp*	*hamburg_avg_temp*	*global_avg_temp*
count	264	264	264	264	264
mean		4.64	8.92	8.33	8.36
std		0.76	0.88	0.85	0.58
min	1750	0.53	4.84	4.65	5.78
max	2013	6.64	10.96	10.37	9.73

Table 2. Descriptive statistics

The dataset contains 264 observations from 1750 to 2013 with a yearly interval. It contains data for the average temperature in Munich, Berlin, Hamburg and globally. In Munich, the mean average temperature is 4.64°C, its standard deviation is 0.76, the lowest average yearly temperature is 0.53°C and the highest yearly average temperature is 6.64°C. Meanwhile mean average temperatures in Berlin and Hamburg are higher than in Munich, with 8.92 and 8.33 respectively. The lowest yearly average temperature in Berlin is 4.84°C, while the highest is 10.96°C, with the standard deviation of 0.88. Similarly, in Hamburg, the lowest yearly average temperature is 4.65°C, and the highest is 10.37°C, with the standard deviation of 0.85. Globally, the mean average temperature is 8.36°C with the lowest average temperature of 5.8°C and the highest of 9.73.

3.2. Data manipulating

As there are fluctuations in yearly average temperature (°C), a moving average method is considered to make lines smoother and trends clearer during data visualization. To do this, I import the CSV file into Excel.

The moving average is performed using Excel command on a 15-year basis for both the local city data and global data to obtain two new moving average data columns on the same Excel spreadsheet which are used for data visualization.

Figure 3 below shows the calculation of the average temperature for the first 15 years (1750 – 1764) for Munich city. The same is then repeated from years 1751-1765 and so forth. Then, I do the same for Berlin, Hamburg and global data.

yea	munich	berlin_a	hambur	global_	Munich_m.avg	Berlin_m.avg	Hamburg_m.a	Global_m.avg
1750	5.4	9.83	9.31	8.72				
1751	5.54	9.75	8.94	7.98				
1752	0.53	4.84	4.65	5.78				
1753	4.61	8.72	8.12	8.39				
1754	4.33	8.49	7.88	8.47				
1755	4.05	8.26	7.69	8.36				
1756	4.64	9.62	8.64	8.85				
1757	4.3	9.15	8.48	9.02				
1758	3.83	8.25	7.76	6.74				
1759	4.89	9.04	8.69	7.99				
1760	5.02	8.99	8.36	7.19				
1761	4.94	9.47	8.86	8.77				
1762	4.49	8.53	7.81	8.61				
1763	4.25	8.62	7.8	7.5				
1764	4.82	8.91	8.25	8.4	=AVERAGE(B2:B16)	8.70	8.08	8.05
1765	4.52	8.54	7.91	8.25	4.32	8.61	7.99	8.02
1766	4.28	8.87	8.27	8.41	4.23	8.55	7.94	8.05

Figure 3. Excel spreadsheet screenshot of the 15-year moving average calculation.

4. Data visualization

4.1. Line chart

As mentioned, Munich is chosen as the main local city to compare with global data. Therefore, in this section, I will only use the 15-year moving average temperature data of Munich city and global.

A line chart is now created to compare the Munich city's average temperatures with the global average temperatures. I use Excel to plot the moving average temperatures on the y-axis and the year range on the x-axis. The line chart is then generated and shown as follows.

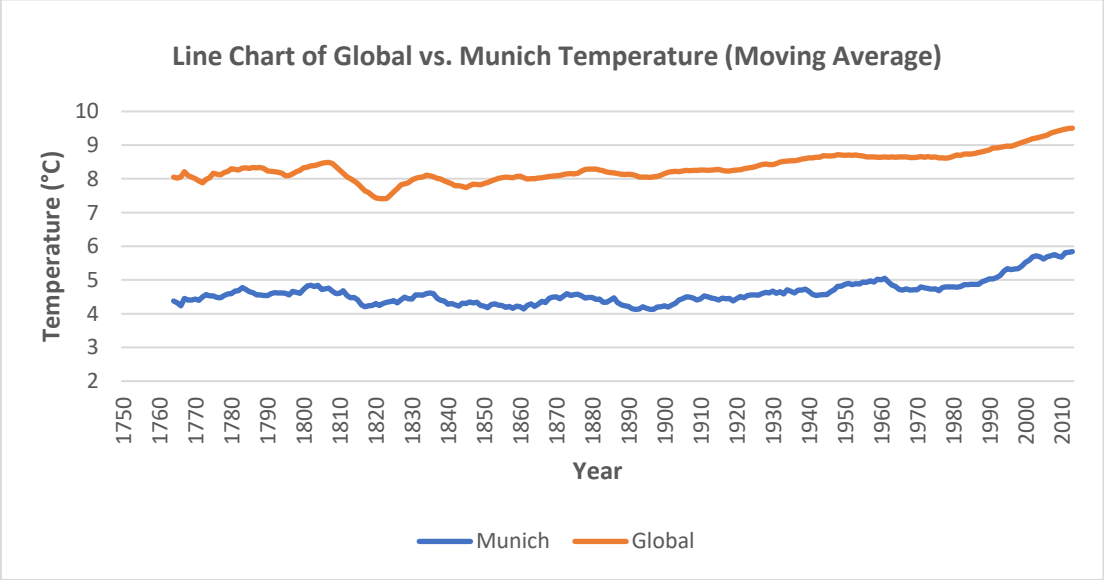

Figure 4. Line Chart of Global vs. Munich Temperature (Moving Average)

The line chart in Figure 4 shows the 15-year moving average temperature (°C) for Munich city and global from 1964 to 2013.

4.2. Observations

According to the line chart of moving average temperature (Figure 4), the following observations might be inferred.

1) Munich city is much cooler on average than the global average as Munich average temperature has a mean of 4.64°C while the mean of global average temperature is 8.36°C.
2) The mean difference between global temperature and Munich temperature is around 4 Celsius degree with the standard deviation of 0.18, indicating that there is not much variation in the difference of the temperature.
3) From the line chart, we can say that the difference is approximately consistent over time.
4) Both Munich and global average temperatures are gradually rising throughout the entire timeframe as the lowest average temperatures are in earlier years and the highest values are in later year. In details, for Munich, the average temperature has the min value of 0.53°C in 1952 and the max of 6.64°C in 1994, while global average temperature has the min value of 5.78°C also in 1952 and the max of 9.73°C in 2007.
5) Last but not least, global temperature upward trend indicates that the world is getting hotter and this global warming trend has been continuously happening since the early-mid 19th century, following a significant drop in the average temperature from the period of 1810- 1820.

Correlation Coefficient

By using Excel formulae, I calculated the coefficient of correlation for Munich city and global temperature of approximately 0.87. This indicates a strong positive relationship between the two variables.

4.3. Conclusion

Although Munich city average temperature is lower than global temperature, it is getting hotter both in the city and globally over the time.

5. Considering More Cities in Germany

In this part, I will explore and compare the weather trends across Germany from North to South by visualizing datasets from three different cities including Berlin (Northern Germany), Hamburg (North western Germany) and Munich (Southern Germany).

A line chart below shows 15-year moving average temperatures in different three cities in Germany and globally.

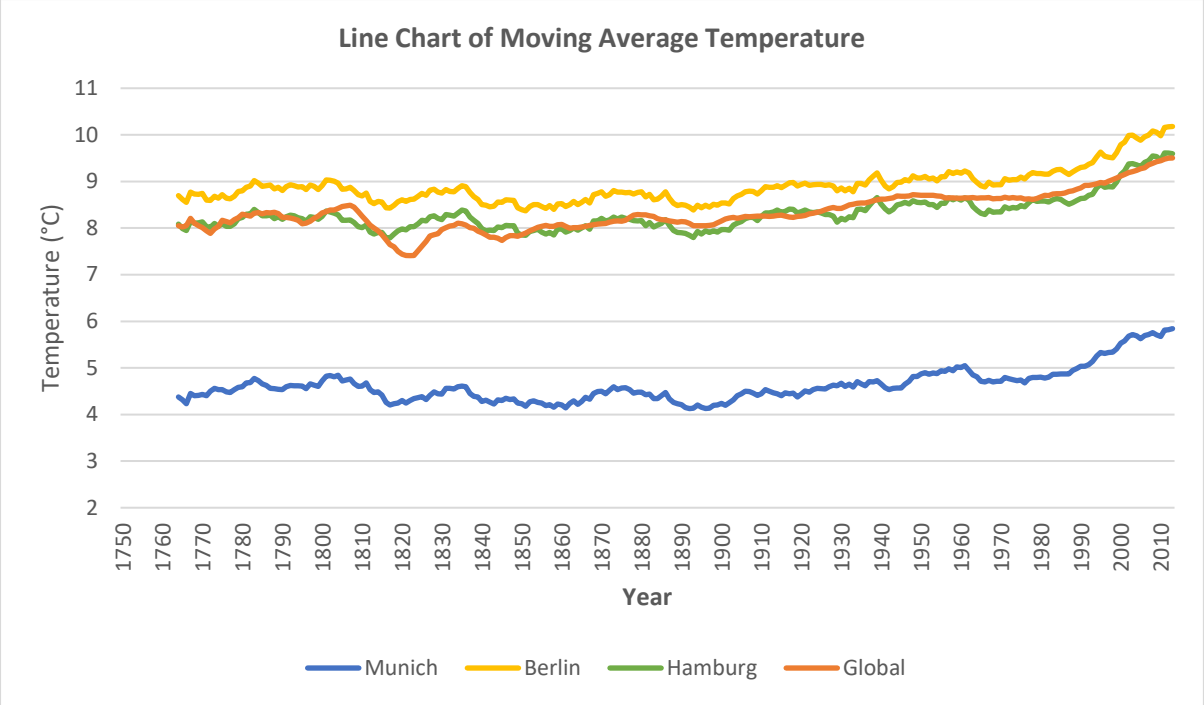

Figure 5. Line Chart of City Temperatures vs Global (10-year MA)

5.1. Observations

According to Figure 5, the following observations might be inferred.

- Surprisingly, even though Munich is in southern Germany, it seems cooler on average (with the mean average temperature of 4.6°C) than two other cities in the North of Germany - Berlin and Hamburg (with the mean of around 9°C and 8.3°C).
- Berlin city in northwestern Germany is on the slightly warmer side on average when compared to global average temperature (8.92°C vs 8.36°C respectively).
- The average temperature of Hamburg city (Northern Germany) is approximately as same as global average temperature (with 8.33 and 8.36), even though there are more fluctuates in the Hamburg average temperature.
- There is an increase of one or two degree Celsius in both Germany and globally during the observed period of time.

5.2. Conclusion

The global warming trend is accelerating since the average global temperature is continuously rising over the time.