1. Extract the data by SQL query

   ```
   SELECT *
   FROM city_list
   WHERE country = 'Germany';
   ```

   I got three cities in Germany where I lived : Berlin, Munich, Hamburg . I select Hamburg as my target city.

   Then I go to the city_data sheet

   ```
   SELECT *
   FROM city_data
   WHERE country = 'Germany' AND city = 'Hamburg';
   ```

   And the selected data which fulfils the condition country = 'Germany' AND city = 'Hamburg' are saved as csv file. And I also notice there's some missing data in average temperature, which I need to clean the data before plotting.

   ```
   SELECT DISTINCT avg_temp, year,city
   FROM city_data
   WHERE country = 'Germany' AND city = 'Hamburg';
   ```

   I download this csv file as the temperature variation dataset for Hamburg.

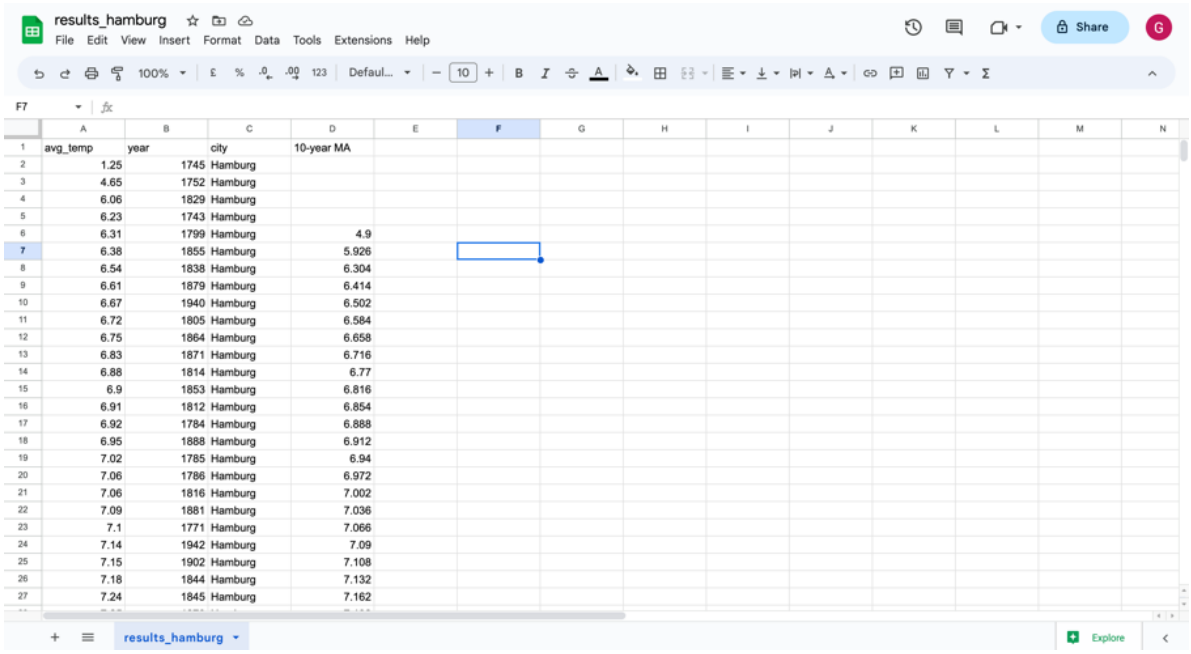
   In the last step, in order to compare the temperature variation in global, I use the SQL query below to fetch the global temperature changing dataset:

   ```
   SELECT DISTINCT avg_temp, year
   FROM global_data;
   ```

2. Data cleaning: this step has been done when extracting data by SQL query.

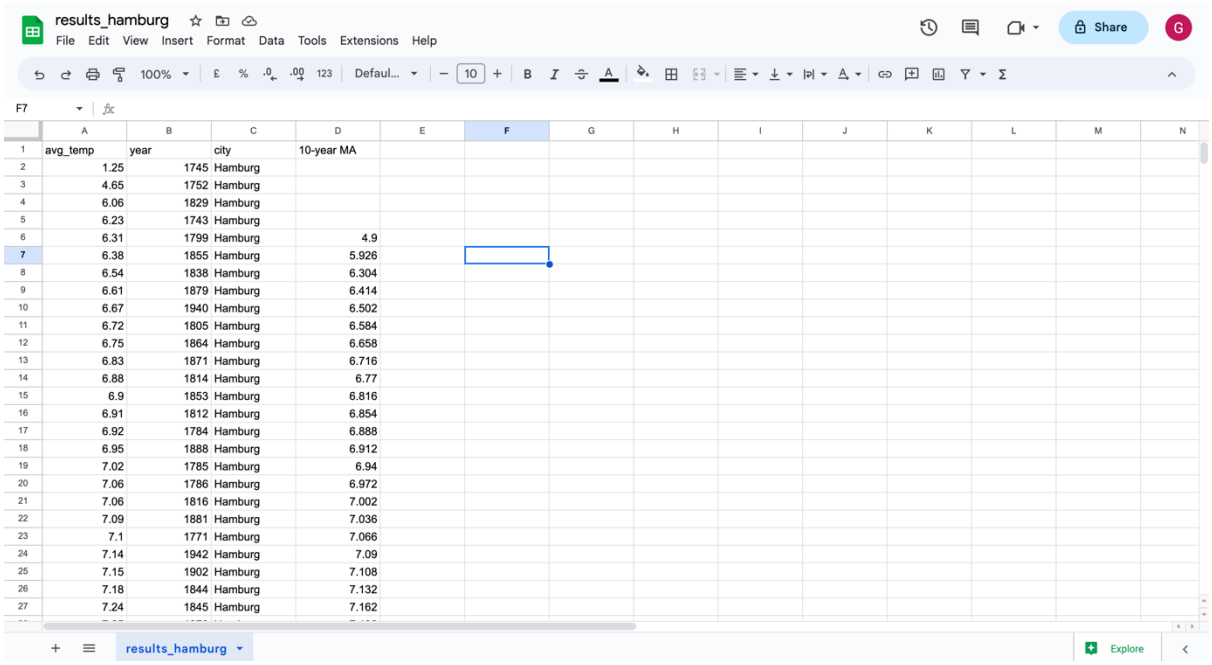
3. Data exploration:
   As the instruction in the project, I calculated averaged temperature in moving average every 5 year for all dataset.

For the Hamburg dataset :



	A	B	C	D
1	avg_temp	year	city	10-year MA
2	1.25	1745	Hamburg	
3	4.65	1752	Hamburg	
4	6.06	1829	Hamburg	
5	6.23	1743	Hamburg	
6	6.31	1799	Hamburg	4.9
7	6.38	1855	Hamburg	5.926
8	6.54	1838	Hamburg	6.304
9	6.61	1879	Hamburg	6.414
10	6.67	1940	Hamburg	6.502
11	6.72	1805	Hamburg	6.584
12	6.75	1864	Hamburg	6.658
13	6.83	1871	Hamburg	6.716
14	6.88	1814	Hamburg	6.77
15	6.9	1853	Hamburg	6.816
16	6.91	1812	Hamburg	6.854
17	6.92	1784	Hamburg	6.888
18	6.95	1888	Hamburg	6.912
19	7.02	1785	Hamburg	6.94
20	7.06	1786	Hamburg	6.972
21	7.06	1816	Hamburg	7.002
22	7.09	1881	Hamburg	7.036
23	7.1	1771	Hamburg	7.066
24	7.14	1942	Hamburg	7.09
25	7.15	1902	Hamburg	7.108
26	7.18	1844	Hamburg	7.132
27	7.24	1845	Hamburg	7.162

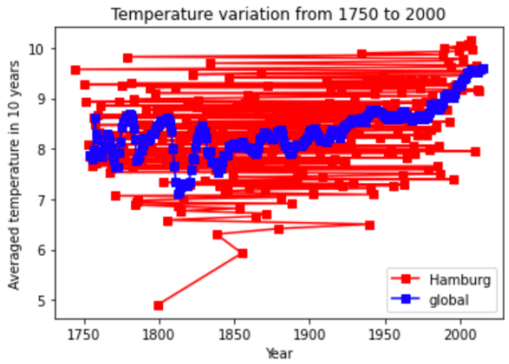
For global dataset:



	A	B	C	D
1	avg_temp	year	city	10-year MA
2	1.25	1745	Hamburg	
3	4.65	1752	Hamburg	
4	6.06	1829	Hamburg	
5	6.23	1743	Hamburg	
6	6.31	1799	Hamburg	4.9
7	6.38	1855	Hamburg	5.926
8	6.54	1838	Hamburg	6.304
9	6.61	1879	Hamburg	6.414
10	6.67	1940	Hamburg	6.502
11	6.72	1805	Hamburg	6.584
12	6.75	1864	Hamburg	6.658
13	6.83	1871	Hamburg	6.716
14	6.88	1814	Hamburg	6.77
15	6.9	1853	Hamburg	6.816
16	6.91	1812	Hamburg	6.854
17	6.92	1784	Hamburg	6.888
18	6.95	1888	Hamburg	6.912
19	7.02	1785	Hamburg	6.94
20	7.06	1786	Hamburg	6.972
21	7.06	1816	Hamburg	7.002
22	7.09	1881	Hamburg	7.036
23	7.1	1771	Hamburg	7.066
24	7.14	1942	Hamburg	7.09
25	7.15	1902	Hamburg	7.108
26	7.18	1844	Hamburg	7.132
27	7.24	1845	Hamburg	7.162

4. Data visualization:

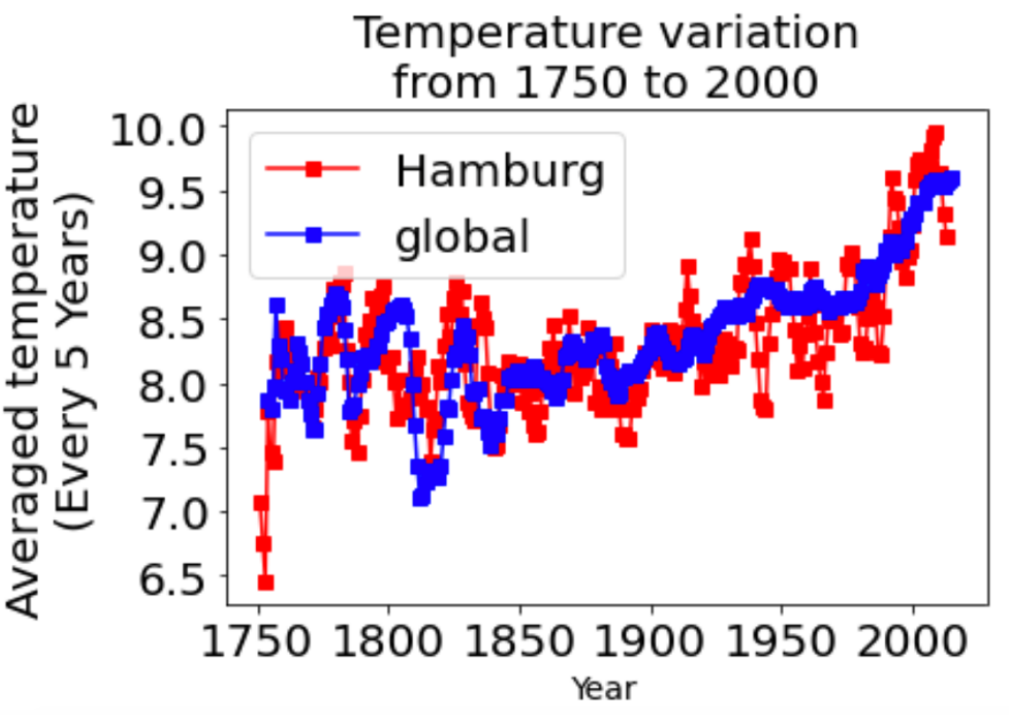
I utilized Jupyter Notebook to access the data, remove any null values, and generate the corresponding plots. The outcome is illustrated below:



To ensure the temperature data for Hamburg is properly organized, I performed another extraction specifically for Hamburg. The SQL query used for this extraction is as follows:

```
SELECT DISTINCT avg_temp, year, city
from city_data
WHERE city = 'Hamburg' AND avg_temp IS NOT NULL
ORDER BY year;
```

I re-extracted the Hamburg temperature data, applied average moving in Excel, removed null values, and plotted the figure below:



5. Observations:

Question 1: Does your city tend to be hotter or cooler than the global average, and has this difference remained consistent over time?

The temperature in my city generally follows the global average trend, with some variations. Initially, around 1750, my city was cooler than the global average. However, after 1800, the temperature in my city began to align with the global trend.

Question 2: How do the temperature changes in your city compare to the global average over time?

In my city, the average temperature experienced a significant increase between 1750 and 1800, followed by fluctuating but overall increasing temperatures until 2000. This trend is similar to the global average temperature trend. However, the temperature fluctuations in my city were more pronounced compared to the global changes. In contrast, the global average temperature remained relatively stable increase after experiencing significant fluctuations between 1750 and 1850.

Question 3: What is the overall trend regarding global temperatures? Is the world getting hotter or cooler, and has this trend been consistent over the last few hundred years?

Both my city and the whole world have been experiencing a consistent trend of increasing temperatures over the years. Looking at the graph, the blue curve representing global average temperatures shows that the temperatures from 1900 to 2000 (the most recent 100 years) are higher than those from 1800 to 1900 (the previous 100 years). This indicates a continuing upward trend. However, the main difference is that the last 100 years have witnessed significant temperature fluctuations, whereas the trend in the most recent 100 years has shown a more stable increase.

Question :
I'd like to join the two table together, but result only show a avg_temp column. How should I fix it ?
The code I used:

```
SELECT gd.avg_temp, cd.avg_temp
FROM global_data as gd
INNER JOIN city_data as cd
ON gd.year = cd.year
```