DATA ANALYST – PROJECT 1 SUMMARY

1. **Step to prepare the data**

   First, I do the query: SELECT * FROM global_data
   This will get the data temperatures of global
   Second, I do SELECT * FROM city_data WHERE city='Hanoi'
   to get the temperatures of Hanoi. I'm living in Ha Noi, Viet Nam.

   - I'm choosing Excel to be visualized data in the chart.
   - Before visualizing this data. I used python pandas first to merge the data of two .csv files into one.

	A	B	C	D
1	year	Global_temp	Hanoi_temp	
2	1840	7.8	21	
3	1841	7.69	21.3	
4	1842	8.02	21.16	
5	1843	8.17	21.26	
6	1844	7.65	20.78	
7	1845	7.85	20.68	
8	1846	8.55	21.25	
9	1847	8.09	20.83	
10	1848	7.98	20.73	
11	1849	7.98	20.94	
12	1850	7.9	21.11	
13	1851	8.18	21.26	
14	1852	8.1	21.12	
15	1853	8.04	21.51	
16	1854	8.21	21.45	
17	1855	8.11	21.1	
18	1856	8	20.69	
19	1857	7.76	20.89	
20	1858	8.1	21.1	
21	1859	8.25	21.24	

Figure 1: The data results after the merge

   - I'm using the AVERAGE function in Excel to calculate the average of temperatures.

   - I consider that Should I calculate all the years or not? And besides that, using a line chart or bar chart, because I saw that also good. Finally, I decision choose a line chart.

2. Line chart with local and global temperature trends

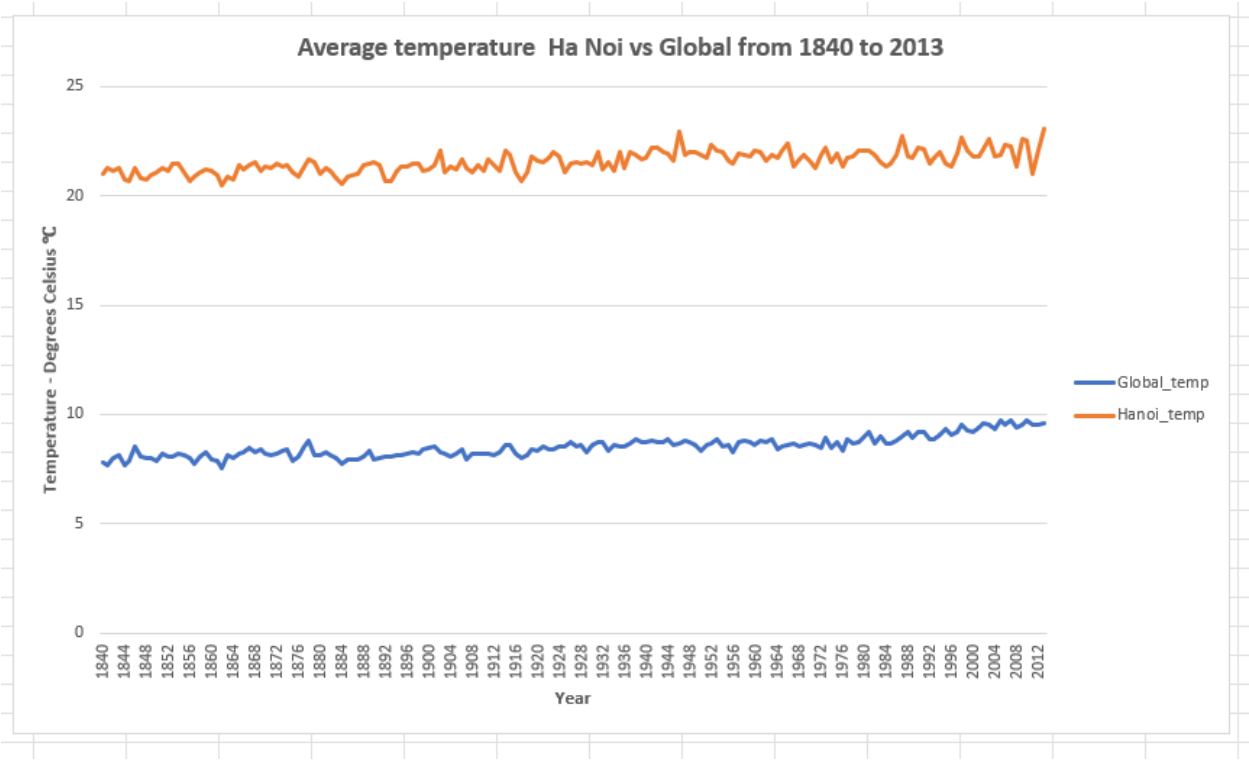

Figure 2: Average temperature HaNoi vs Global

I'm using Excel to create a visualized after merge data by using python pandas.

3. Observations.
   - Firstly, I see the volatility in my local more than global temperatures.
   - Secondly, I saw that the temperature has trended up.
   - Thirdly, The gap between the highest and lowest in the local is quite big (> 3 degrees C), while the global one is insignificant (~2 degrees C)
   - Finally, The alteration in the local is unstable, while the global is in the opposite.