**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.

Table

Description automatically generated

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.

1. Line chart with local and global temperature trends

Graphical user interface, chart

Description automatically generated

Figure 2: Average temperature HaNoi vs Global

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

1. 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.