Exploring Weather Trends

# **Outline**

1. **What tools did you use for each step? (Python, SQL, Excel, etc)?**

I used the following SQL query to extract the dataset from the database:

1. city level data

SELECT year, avg\_temp FROM city\_data WHERE country = 'Japan' and city = 'Tokyo';

1. global data

SELECT \* FROM global\_data;

Then I used Excel & Python to open the dataset. Additionally, I used Google Colab for visualization.

gobal\_data\_df = pd.read\_csv('/content/Global\_Data.csv')

print(gobal\_data\_df)

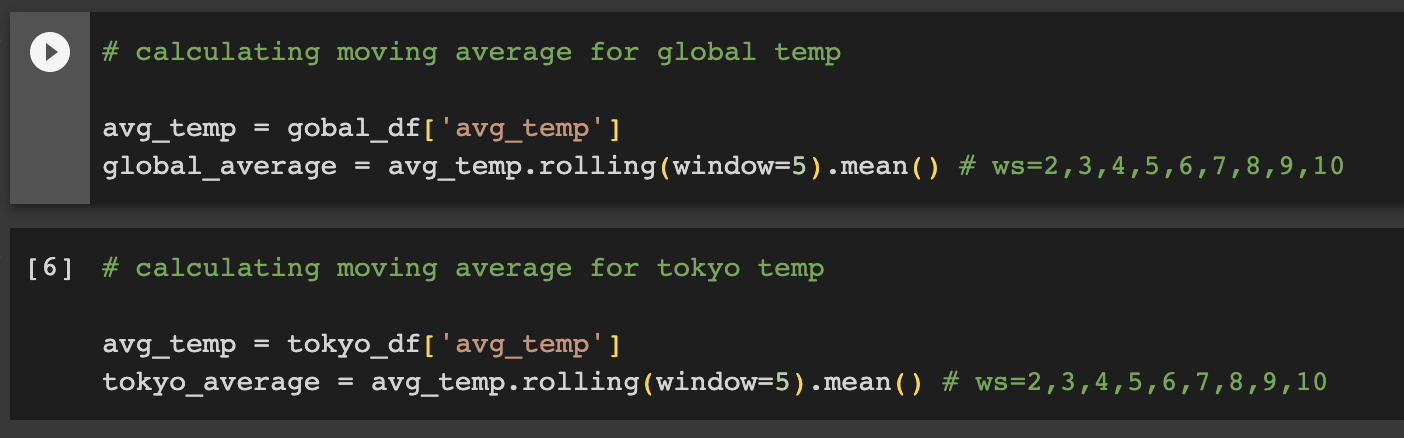
tokyo\_temp\_df = pd.read\_csv('/content/Tokyo\_temp.csv')

print(tokyo\_temp\_df)

1. **How did you calculate the moving average?**

I used Python’s PANDAS library to compute the rolling window.

Figure 1: This code uses the .rolling function from the PANDAS library and sets the window size for computing the moving average.



1. **What were your key considerations when deciding how to visualize the trends?**
   1. No missing data,
   2. all data fields are quantitative,
   3. datafields are (3) standardized,
   4. that year is an independent variable causes an effect and can be manipulated hence plotted on the x-axis,
   5. that avg\_temp is a dependent variable that is an effect and measured thus plotted on the y-axis,
   6. finding the best window size,
   7. and the graph title and legends are labeled.

Figure 2: This code reads the csv file for global data & tokyo data.

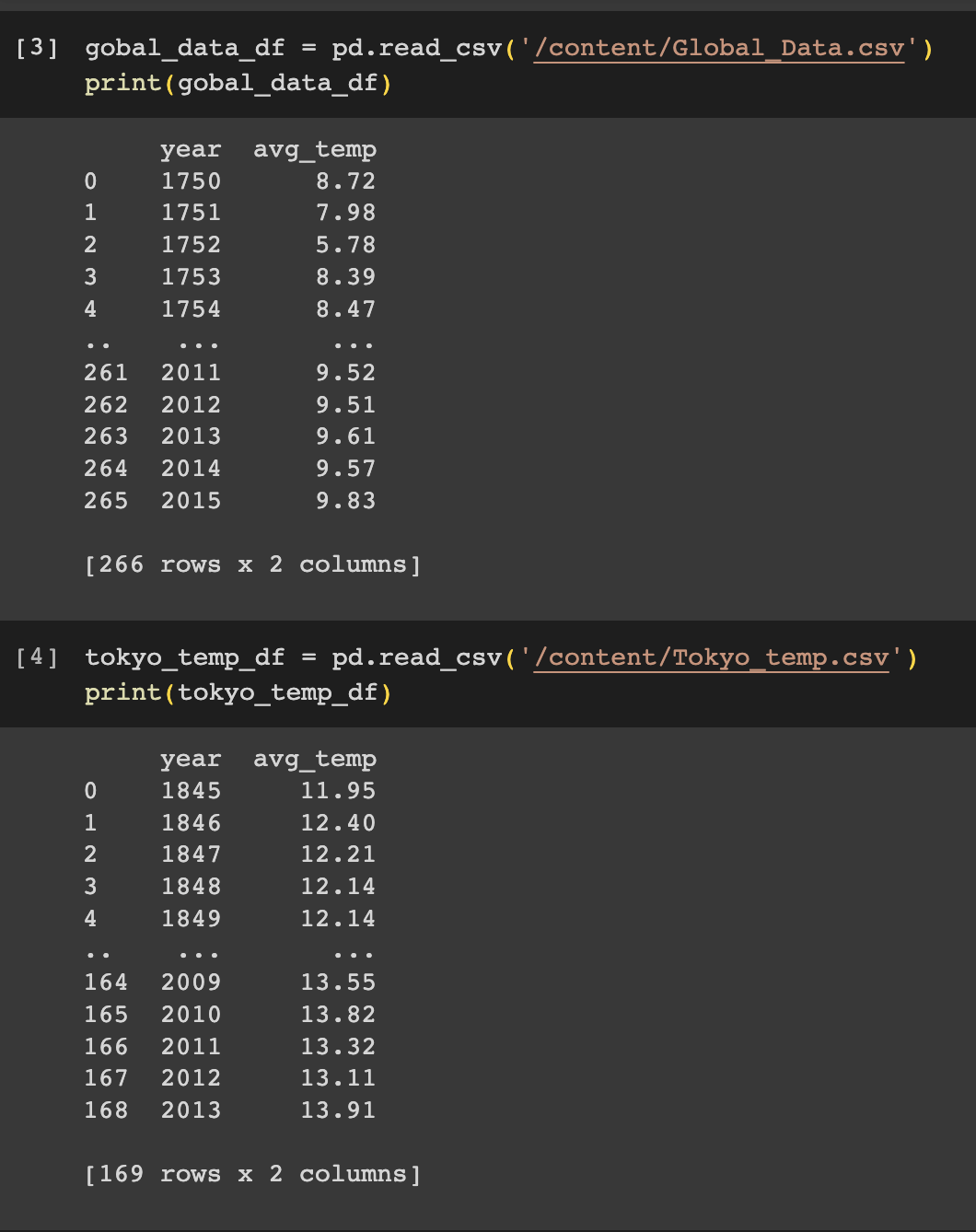
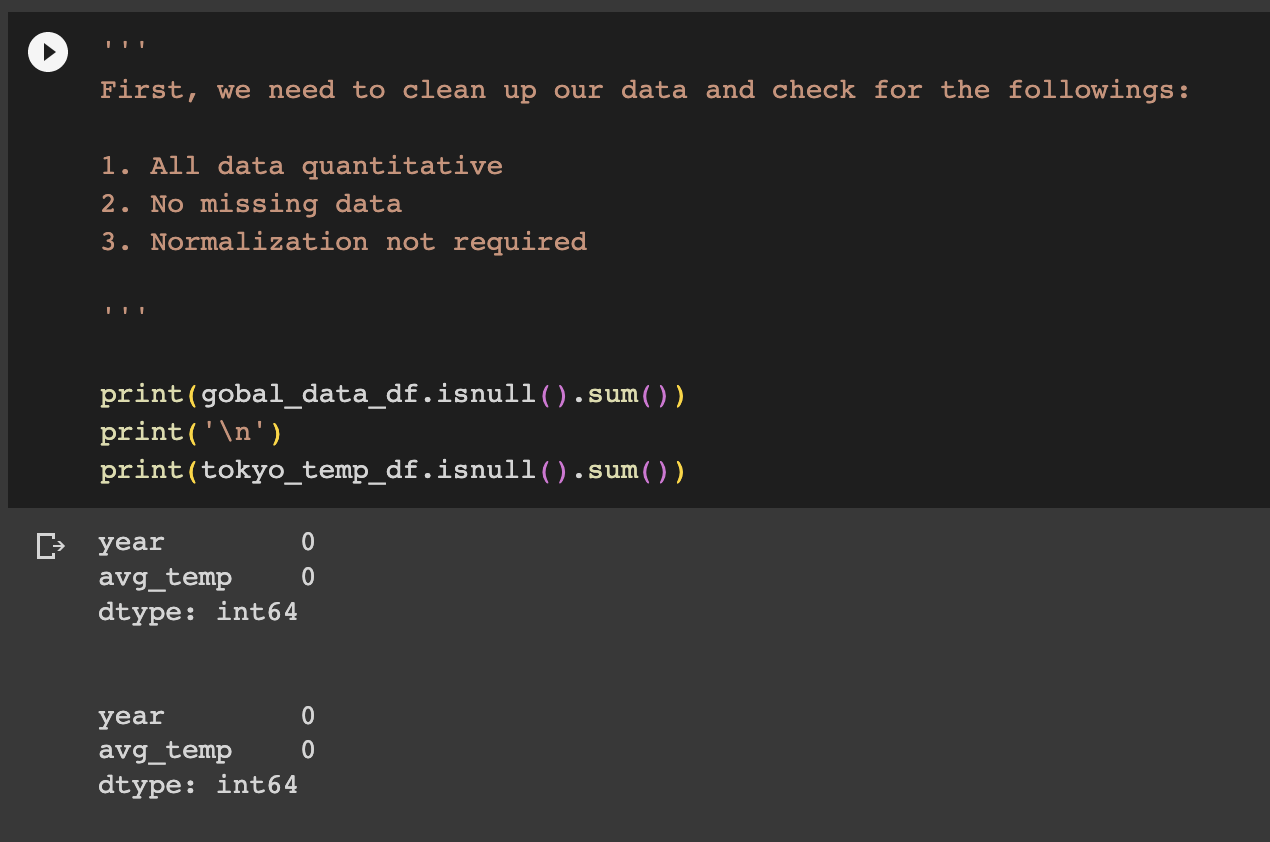


Figure 3: This code checks for any missing data.



# **Visualization**

Figure 4: Average temperature trends: Global vs Tokyo

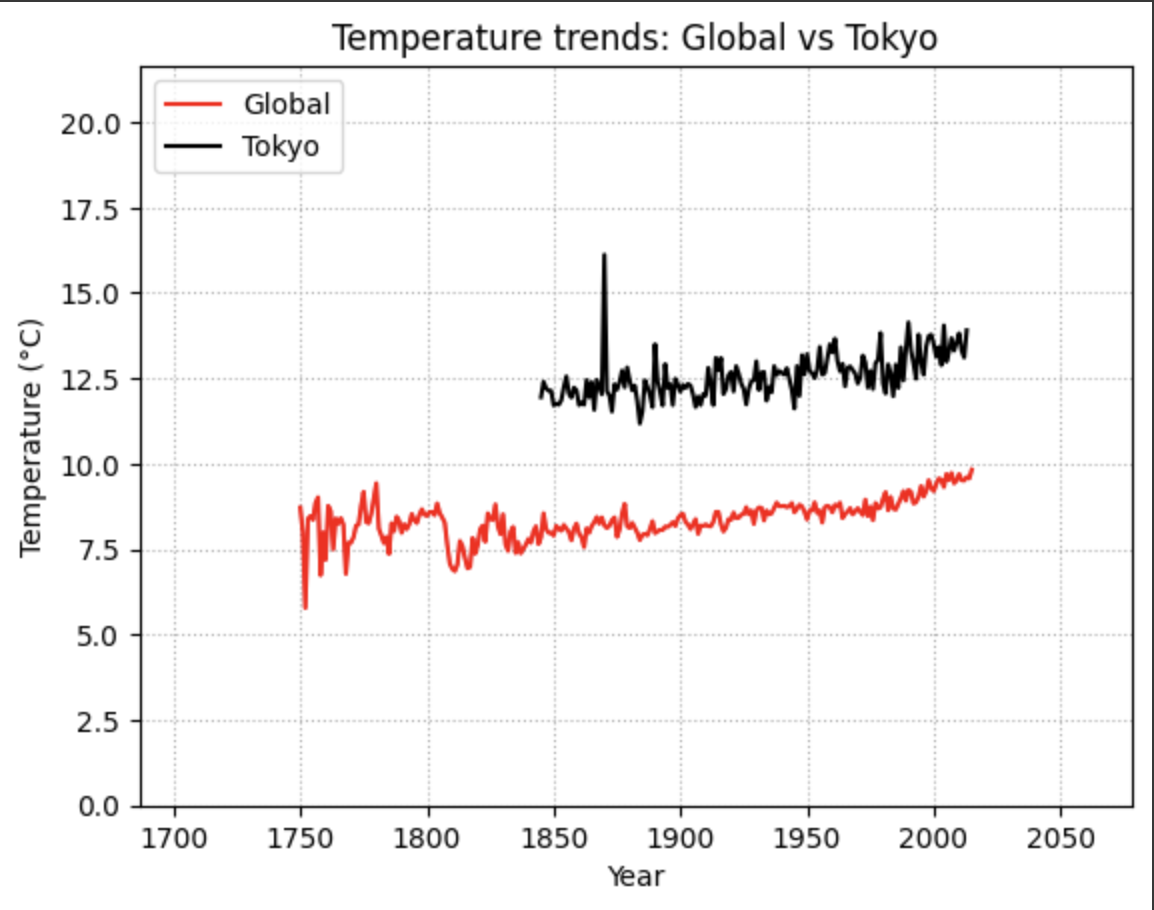


Figure 5: 5-year moving average temperature trends: Global vs Tokyo

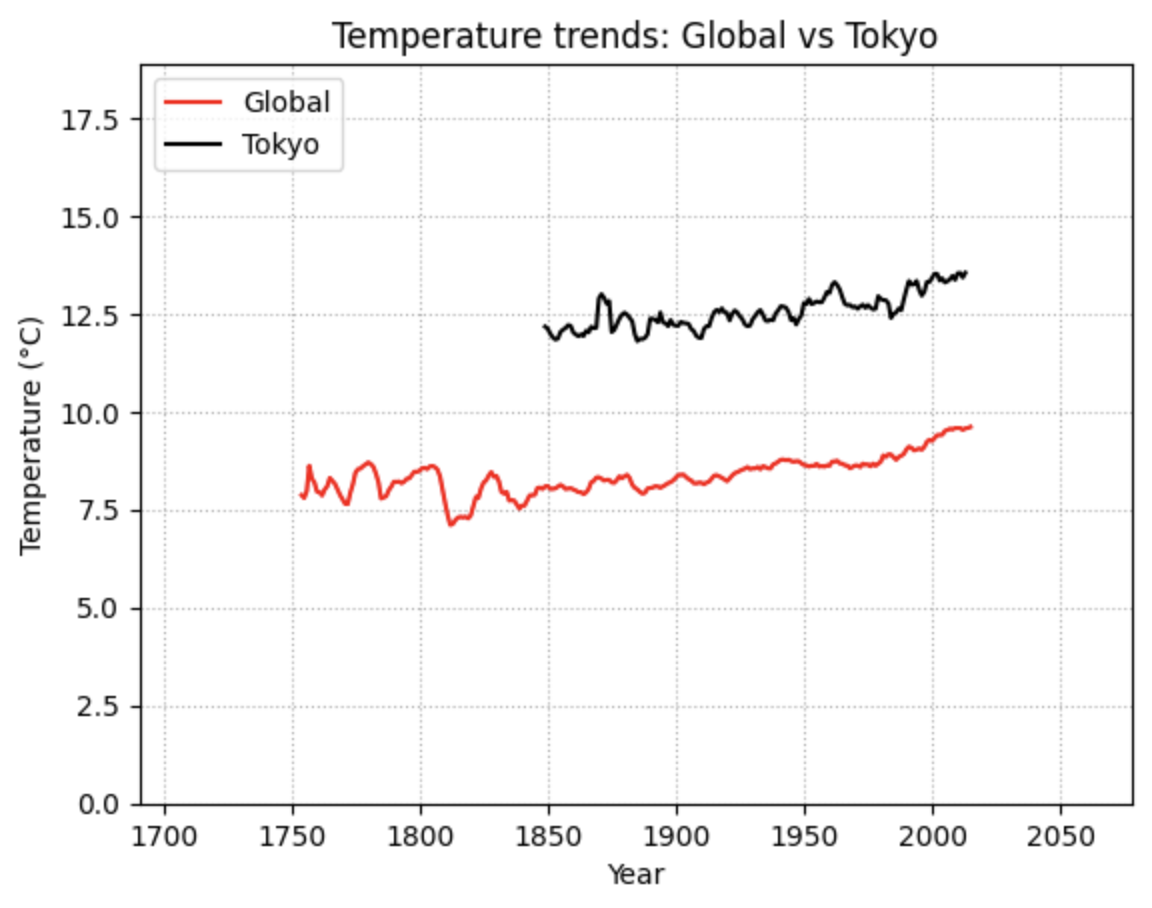


Figure 6: 7-year moving average temperature trends: Global vs Tokyo

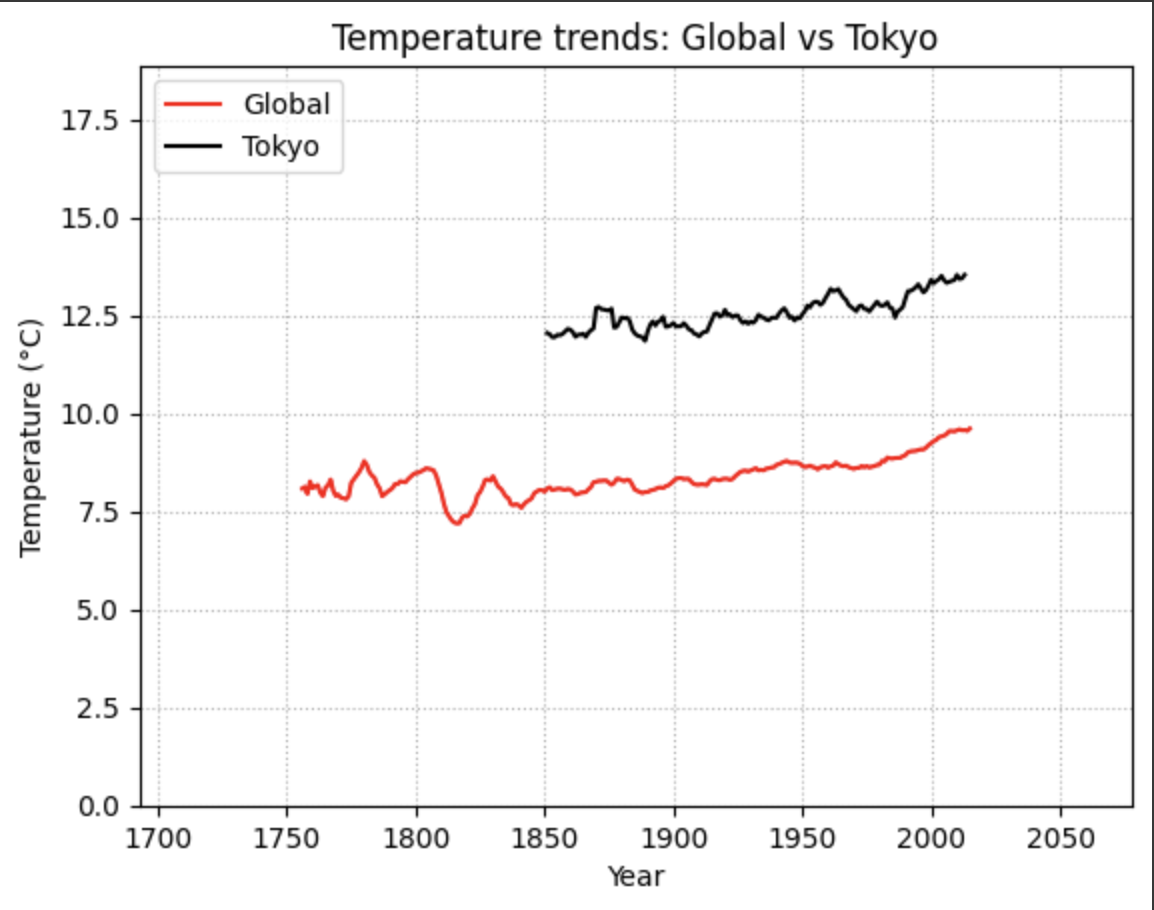


Figure 7: 10-year moving average temperature trends: Global vs Tokyo

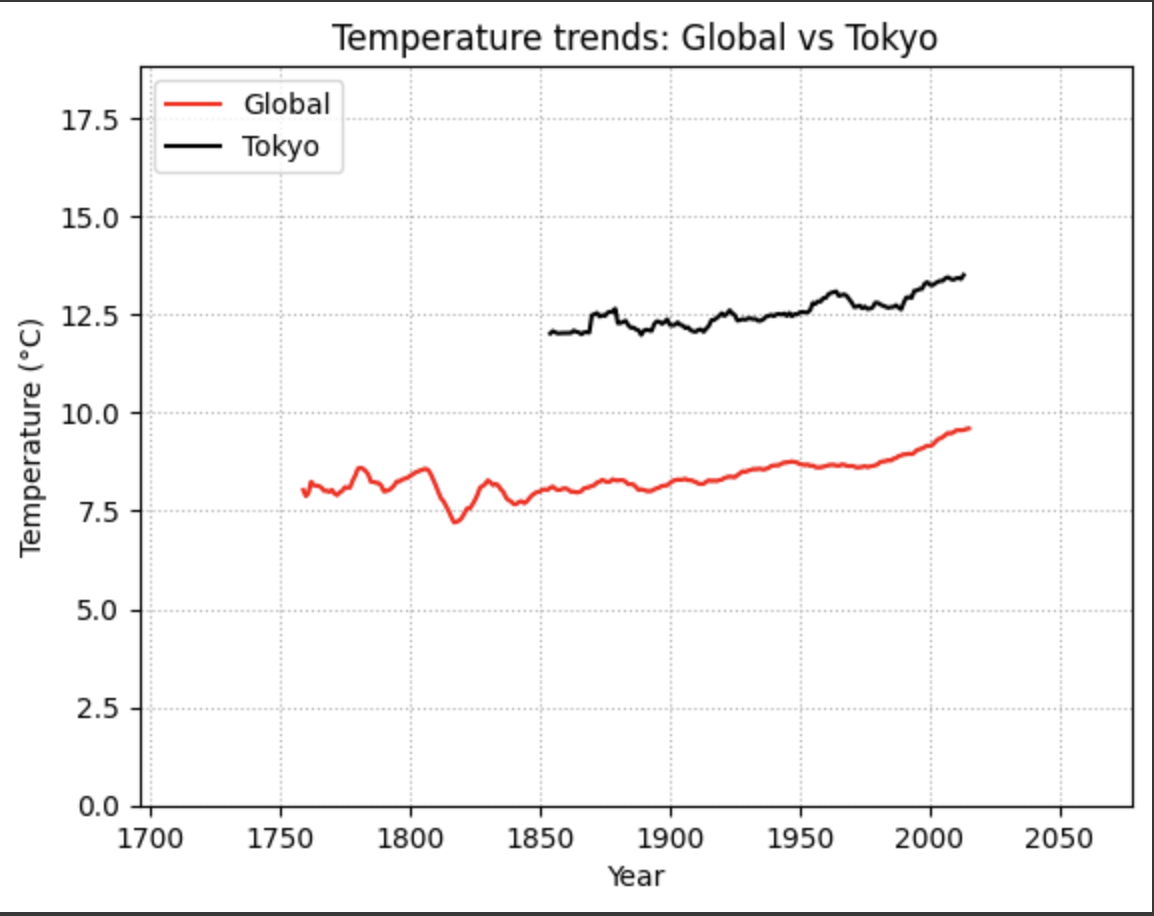


Figure 8: 20-year moving average temperature trends: Global vs Tokyo

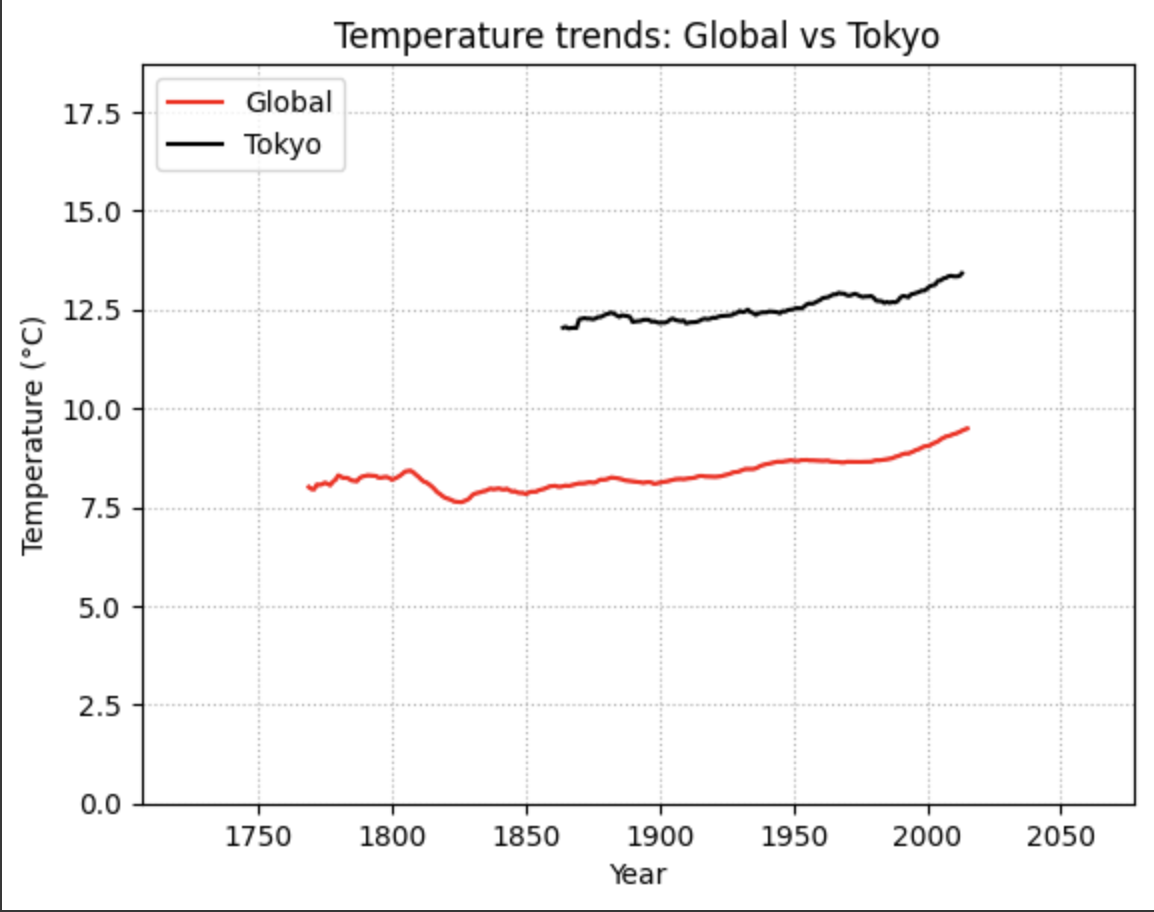


Figure 9: This is the heatmap used from the seaborn library. It shows a weak positive correlation between avg temp and year from the global dataset.

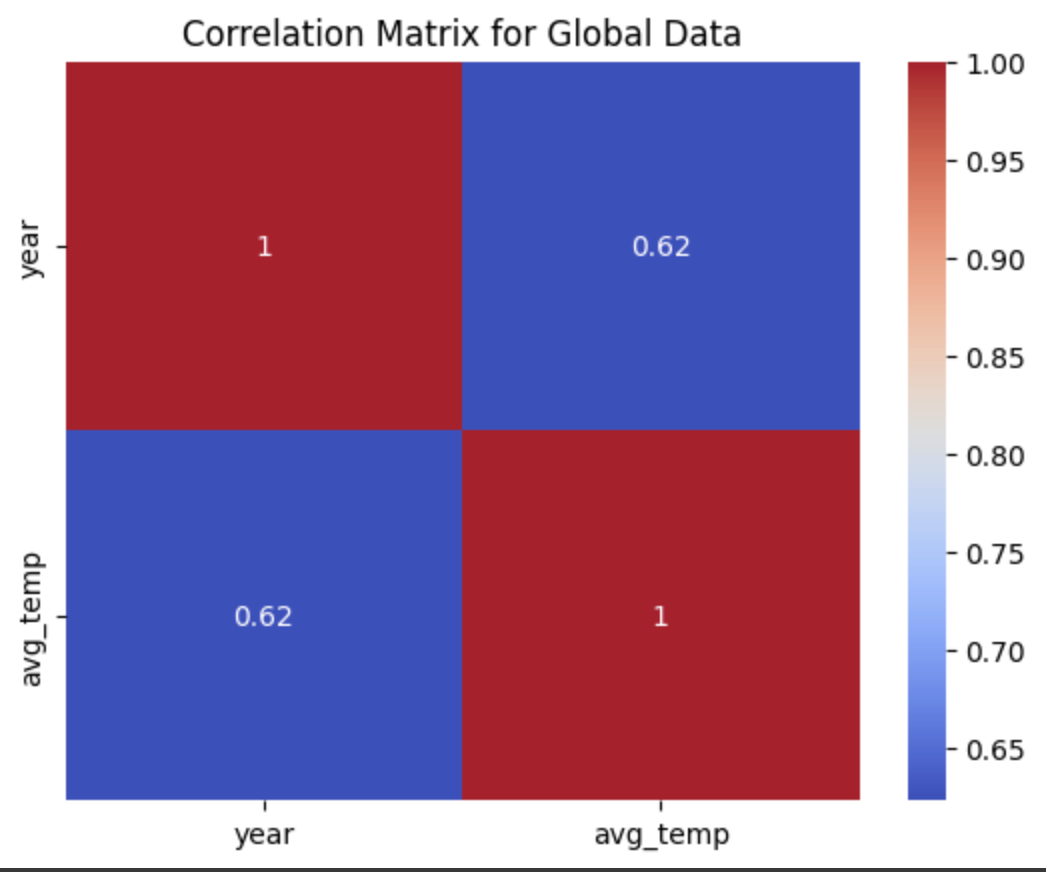
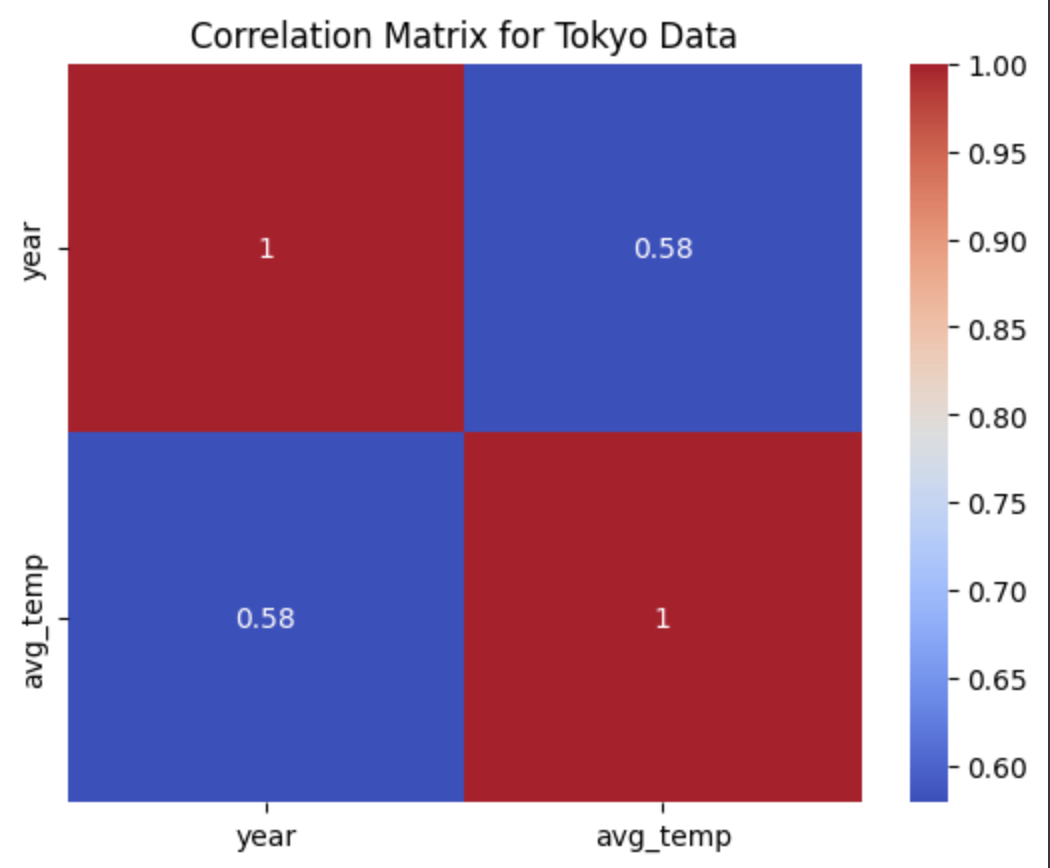


Figure 10: This is the heatmap used from the seaborn library. It shows a weak positive correlation between avg\_temp and year from the Tokyo dataset.



# **Observations**

1. **Is your city hotter or cooler on average compared to the global average? Has the difference been consistent over time?**

Tokyo is hotter on average compared to the global average. The difference has been shown to be consistent over time.

1. **“How do the changes in your city’s temperatures over time compare to the changes in the global average?”**

Tokyo’s average temperature has been increasing similar to the increase of the global average. Also, Tokyo’s temperature tends to fluctuate more often than the global average.

1. **What does the overall trend look like? Is the world getting hotter or cooler? Has the trend been consistent over the last few hundred years?**

The overall trend appears to be that the world is getting slightly hotter. This trend has been consistent since the 1900s.

1. **The 20-year moving average seems to have the smoothest line in contrast to plotting the average temperature trends. Picking the appropriate sliding window was the trickiest part so I experimented with different numbers. The smoothest was 20 years and beyond it didn’t change much. However, 20 years might also obscure some necessary details, but for this task specifically, I think the 20 years visually reflects more clearly that increase in global and local trending temperatures.**
2. **What's the correlation coefficient?**

As shown in the heatmap, the correlation coefficient is 0.62 between avg\_temp and year in the global dataset while a 0.58 between avg\_temp and year in the tokyo dataset. Both results indicate a weak positive correlation between the two variables. Although the results do not indicate an ideal linear regression solution, it could be the most accurate.

1. **Can you estimate the average temperature in your city based on the average global temperature?**

Yes. I would perform linear regression on the global temperature to find the best fit line. Then average the difference between global and local. Then add the difference to the predicted linear regression output.

# **References**

1. Helmenstine, A. (2022, December 5). *Difference Between Independent and Dependent Variables*. Science Notes and Projects. <https://sciencenotes.org/difference-between-independent-and-dependent-variables/>
2. Effect of window size in moving average. (2020, June 29). <https://mins.space/blog/2020-06-29-moving-average-effect-window-size/>