

Exploring Weather Trends

Submitted by Edward Kyalo

Tools Used

1. Jupyter Notebooks via Google Colab (<https://colab.research.google.com/>)

Outline of steps taken

1. **Extract required data from the datasets provided** using the following SQL commands:

```
SELECT *  
FROM city_data  
WHERE city = 'Kampala'
```

```
SELECT *  
FROM global_data
```

The data returned was stored in CSV files.

2. **Created a new Jupyter notebook using Google Colab.**
Imported the CSV files, named `global_temp.csv` and `kampala_temp.csv`.
3. **Imported required python libraries**

```
import numpy as np  
import pandas as pd  
import matplotlib.pyplot as plt  
%matplotlib inline
```

The inline 'magic command' was used to embed the plotted line graph in the notebook¹.

4. **Loaded the CSV files into pandas DataFrames**

¹ <https://ipython.readthedocs.io/en/stable/interactive/plotting.html>

```
global_data = pd.read_csv('global_temp.csv')
city_data = pd.read_csv('kampala_temp.csv')
```

5. Inspect the global data

```
global_data.head(10)
```

The head of the global DataFrame was as follows:

	year	avg_temp
0	1750	8.72
1	1751	7.98
2	1752	5.78
3	1753	8.39
4	1754	8.47
5	1755	8.36
6	1756	8.85
7	1757	9.02
8	1758	6.74
9	1759	7.99

```
global_data.describe()
```

The DataFrame description was as follows:

	year	avg_temp
count	266.000000	266.000000
mean	1882.500000	8.369474
std	76.931788	0.584747
min	1750.000000	5.780000
25%	1816.250000	8.082500
50%	1882.500000	8.375000
75%	1948.750000	8.707500
max	2015.000000	9.830000

6. Inspect the local data

```
city_data.head(10)
```

The first 10 rows of the local DataFrame were as follows:

	year	city	country	avg_temp
0	1850	Kampala	Uganda	22.47
1	1851	Kampala	Uganda	NaN
2	1852	Kampala	Uganda	NaN
3	1853	Kampala	Uganda	NaN
4	1854	Kampala	Uganda	NaN
5	1855	Kampala	Uganda	NaN
6	1856	Kampala	Uganda	NaN
7	1857	Kampala	Uganda	NaN
8	1858	Kampala	Uganda	NaN
9	1859	Kampala	Uganda	22.88

```
city_data.describe()
```

The city DataFrame description was as follows:

	year	avg_temp
count	164.00000	141.000000
mean	1931.50000	23.231702
std	47.48684	0.429265
min	1850.00000	22.310000
25%	1890.75000	22.930000
50%	1931.50000	23.220000
75%	1972.25000	23.470000
max	2013.00000	24.290000

7. Fill NaN values

The city data was shown to have 23 NaN values in the average temperature column within the first rows of the DataFrame. To fill these NaN values, the pandas interpolate function was used.

```
city_data.interpolate()
```

Interpolation was suitable to fill the NaN values because of the absence of outliers in the city DataFrame, evidenced by the relative consistency in interquartile range shown in the DataFrame description.

8. Calculate Moving Averages

Moving averages were calculated using the pandas rolling function, with a window of 10 being used. The resulting data was saved to new dataframes.

```
ma_global = global_data.avg_temp.rolling(window=10).mean()  
ma_city = city_data.avg_temp.interpolate().rolling(window=10).mean()
```

9. Delete unnecessary data

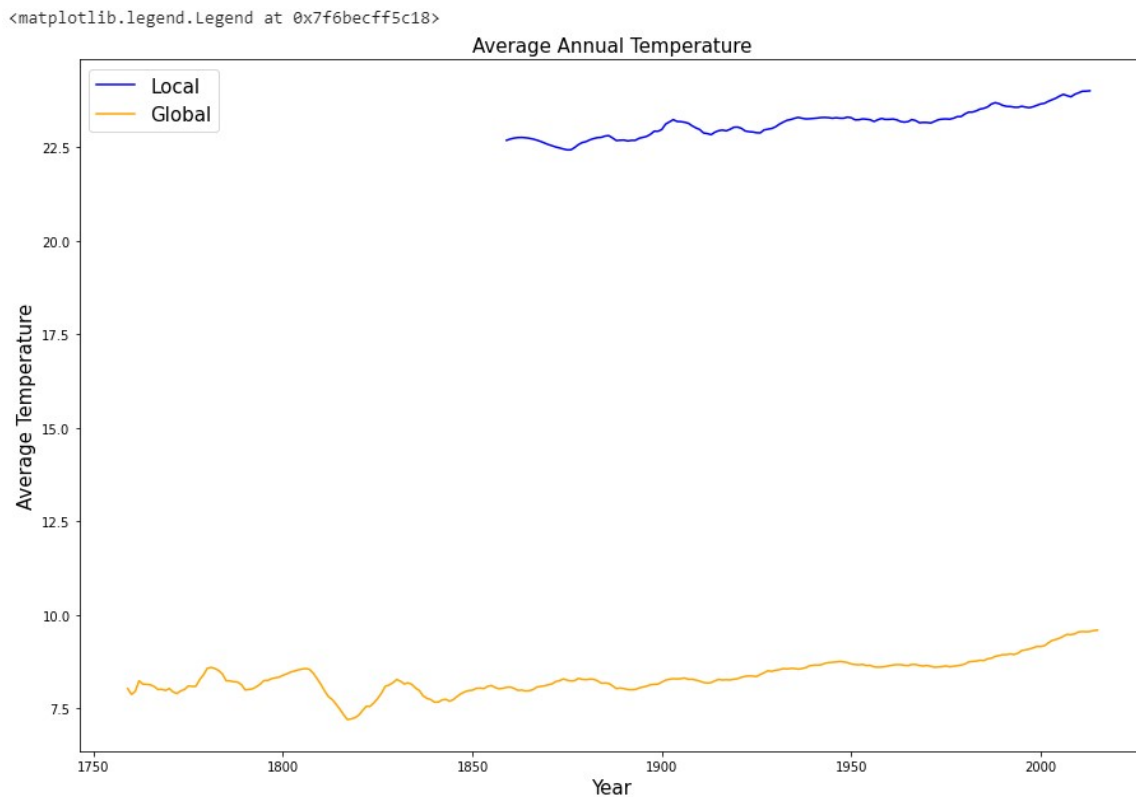
The city DataFrame contained 2 columns ('City' and 'Country') that contained redundant data, seeing as the entire dataframe was obtained from one locale. Therefore, these columns were deleted.

```
del city_data['city']
del city_data['country']
```

10. Plot the global and city dataframes on a line graph.

```
plt.figure(figsize=(15,10))
plt.plot(city_data.year, ma_city, label='Local', color='blue')
plt.plot(global_data.year, ma_global, label='Global', color='orange')
plt.xlabel('Year', fontsize=15);
plt.ylabel('Average Temperature', fontsize=15)
plt.title('Average Annual Temperature', fontsize=15)
plt.legend(loc='upper left', fontsize=15)
```

The following line graph was plotted:



Observations

1. The annual average temperature in Kampala, is on average, hotter than the global average.
2. The graph shows a very gradual increase in both global and local average temperature over time.
3. The rate of change in the average annual temperature is slightly greater in Kampala as opposed to the global average.
4. There are fewer outliers in the local temperature trends as opposed to the global temperature. These outliers were, however, recorded before the local temperature recordings began.