

Explore Weather Trends

This a project geared towards analysing local and global temperature data.

Extracting data with SQL

Looking at the dataset to understand what it contains

The database contains 3 tables, **city_list**, **city_data** and **global_data**.

- global_data: Contains the average temperature for each year. Has **2 columns**; **year** and **avg_temp** with about **266 rows**. This data is ordered by **year**.
- city_list: List each city in the database along with their country. Has **2 columns**; **city** and **country** and has **342 cities** in the table.
- city_data: List the average temperature of each city in the database from **1743 to 2013**. Has 4 columns; **year**, **city**, **country** and **avg_temp** and about **70792 records**. This data is ordered by **year**.
- Exploring**: Figuring out hottest cities in the year 2013

```
SELECT *, RANK() OVER (PARTITION BY year ORDER BY avg_temp DESC) temp_rank
FROM city_data
WHERE year = 2013
LIMIT 5;
```

- Result: The cities Khartoum (Sudan), Niamey (Niger), Maiduguri (Nigeria) and Ouagadougou (Burkina Faso) has been consistent on the top 5 over the last 5 years.

year	city	country	avg_temp	temp_rank
2013	Khartoum	Sudan	30.65	1
2013	Niamey	Niger	30.42	2
2013	Maiduguri	Nigeria	29.71	3
2013	Ouagadougou	Burkina Faso	29.41	4
2013	Bangkok	Thailand	28.98	5

- Exploring**: Finding closest big city to where I live from the city_list table. (I live in Osun state, Nigeria, so I would looking to find Ibadan or Lagos)
- ```
SELECT *
FROM city_list
WHERE country = 'Nigeria';
```
- Result: There are 7 Nigerain cities in the database; **Abuja, Ibadan, Kaduna, Kano, Lagos, Maiduguri** and **Port Harcourt**. Of the 7, **Ibadan** is close to **Osun state** where I live so going with that.

- Exploring**: Looking at global temperature table in database. *(266 result)*

```
SELECT *
FROM global_data;
```

- Result: The global table has data in the table going way back to **1750** up until **2015**.

- Exploring**: Looking at Ibadan temperature data in database. *(158 results)*

```
SELECT year, avg_temp
FROM city_data
WHERE country = 'Nigeria' AND city = 'Ibadan';
```

- Result: Resulting table has temperature data going back to **1856** up until **2013**, the city of Ibadan also has missing data from **1863 to 1872**

- Exploring**: Looking at Lagos temperature data in database (out of curiosity, I'd like to see how lagos compares). *(165 result)*

```
SELECT year, avg_temp
FROM city_data
WHERE country = 'Nigeria' AND city = 'Lagos';
```

- Result: Resulting table has temperature data going back to **1849** up until **2013**, this city of lagos also has missing data from **1852 to 1855** and from **1863 to 1872**.

- Extracting**: Getting require data from database and exporting as csv.

```
SELECT g.year, g.avg_temp global_avg_temp, i.avg_temp ibadan_avg_temp, l.avg_temp lagos_avg_temp
FROM global_data g
JOIN city_data i
ON g.year = i.year AND i.city = 'Ibadan'
JOIN city_data l
ON g.year = l.year AND l.city = 'Lagos'
ORDER BY i;
```

- Result: **158 record** containing the **year**, **global** average temperature and the average temperatures of **Lagos and Ibadan city**. The data starts from **1856 to 2013**, I could have used **LEFT JOIN** to get more data global data, but since I trying to make comparison, I reasoned it would be best getting portions of the data that are relevant for comparison.

There are some missing data in Lagos and Ibadan, but would use interpolation to fix that.

## Moving to Google sheets

- Importing Data**: I imported the `avg_temp_data.csv` file I downloaded into google sheets.
- Interpolation**: As stated above, I observed some missing data in Ibadan from 1863 to 1872 (likewise Lagos, about 12 missing data) and a way I thought to fix this is to fill the missing data with the moving average of 14 rows from the known points of the database. My formular `=round(average(E112:E125), 2)`. *Link to sheet below*
- Descriptive statistics**: I derived the descriptive statistics of the 3 datasets to gain better insight on the data.

|                        | Global  | Ibadan  | Lagos   |
|------------------------|---------|---------|---------|
| Mean                   | 8.58    | 26.34   | 26.53   |
| Standard Error         | 0.036   | 0.043   | 0.043   |
| Median                 | 8.56    | 26.43   | 26.65   |
| Mode                   | 8.73    | 26.7    | 26.66   |
| Standard Deviation     | 0.46    | 0.54    | 0.54    |
| Sample Variance        | 0.21    | 0.29    | 0.29    |
| Kurtosis               | -0.055  | 0.23    | 0.11    |
| Skewness               | 0.56    | -0.46   | -0.43   |
| Range                  | 2.17    | 2.93    | 2.93    |
| Minimum                | 7.56    | 24.63   | 24.86   |
| Maximum                | 9.73    | 27.56   | 27.79   |
| Sum                    | 1354.98 | 4162.02 | 4192.52 |
| Count                  | 158     | 158     | 158     |
| Correlation with Years | 0.86    | 0.72    | 0.77    |
| Slop with Years        | 0.0086  | 0.0085  | 0.0091  |

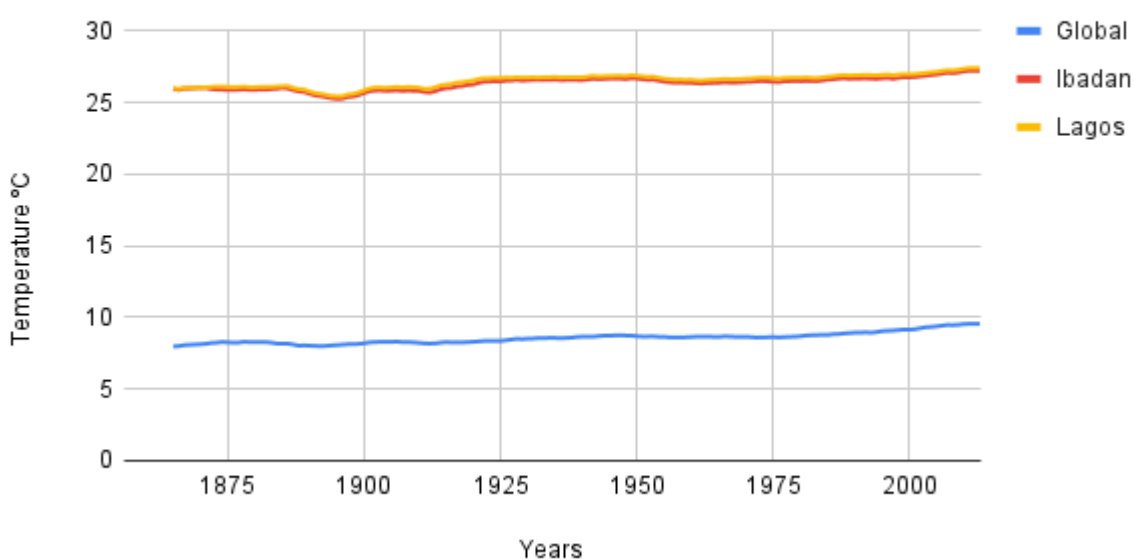
Descriptive statistics table

- Moving Average**: Created 3 columns **Global**, **Ibadan** and **Lagos** where the **10-years** moving average of average temperature values of the 3 data would be stored. Afterwards, I calculated the 10-years moving averages using the fomulars `=round(average(B2:B11), 2)` for **Global**, `=round(average(F2:F11), 2)` for **Ibadan** and `=round(average(G2:G11), 2)` for **Lagos**. I copy and pasted this to cover all the rows needed. *Link to sheet below*

I initially tried using 5 years but the final chart was a bit noisy and a bit difficult to notice trends and So I went with 10 years.

- Plotting Line Chart**: I then plotted a **Line Chart of Global vs. Ibadan vs Lagos Temperature (Moving Average)** in a combined chat.

### Line Chart of Global vs. Ibadan vs Lagos Temperature (Moving Average)



## Observations

- The city of Ibadan Nigeria, is way hotter then global average temperature with a mean of **26.34** and a standard deviation of **0.54** as compared to the global average temperature which has a mean of **8.58** and a standard deviation of **0.46**. From the data, the city of Ibadan tends to be hotter than the global average by a factore of about **17.76 °C** and hence the formular

$$ib = g + 17.76$$

can be used to estimate the average temperature of Ibadan city based on the average temperature of the globe. Where;

**ib -> Ibadan city average temperature for a given year**

**g -> Global average temperature for a given year**

A reason I could say is the course for this, going with the post on [CarbonBrief](#) about how scientists measure global temperatures, is that the global temperature data includes regions above sea levels and region in the Arctic and Antarctic region, hence this values pull down the global average to low values.

Looking at the city data once more, I found **503 cities** with temperatures below zeros;

```
SELECT *
FROM city_data
WHERE avg_temp < 0
ORDER BY year DESC;
```

which would constitute part of the global\_avg\_temp pulling this average down.

- It can be observed that the globe average temperature values has less fluctuations in it's data as compared to Lagos and Ibadan. Lagos and Ibadan has a standard deviation of **0.54** each and a correlation coefficient, r, of **0.72 and 0.77** respectively The globe has a standard deviation of **0.46** and correlation coefficient of **0.86**. *Insight from the descriptive statistics table*
- A careful observation of the **Line Chart of Global vs. Ibadan vs Lagos Temperature (Moving Average)** above and the correlation coefficients of average temperature values of both the globe, Ibadan and Lagos city (**greater than 0**) shows a rising trend in temperature. It can also be observed from the slop value that the rising trend is almost the same for the 3 zones with Lagos rising faster than the globe and Ibadan.
- This rising trend has been consistent over the past few centuries, however, the data has been more correlated and consistent over the last century as can be observed from the 1900s to the 2000s.
- Lagos city seems to have similar trends with Ibadan city mostly likely because they aren't very far apart, about **135.5 km**.

## Links:

[Google sheet link](#)

[Notebook file.ipynb](#))