Name:**ONDIEK ELIJAH OCHIENG**
Date:**13TH May 2021**

Udacity Data Analysis Nanodegree Project 01
Exploring Weather Trends.

Introduction

Provided with both global and city-level data, students are expected to analyze local and global temperature data and compare the temperatures where we live to the global temperatures.
The data is to be extracted from a database, a workspace connected to a database is also availed.

In this project, I have extracted both the city level and global data, cleaned the data and visualised comparison of the two temperature levels.

Steps Involved
1. Extracting data from the database

The tool used: SQL

Using SQL queries I have been able to interact with the database and downloaded a CSV file containing the city-level data and the global data using a single query as shown underneath. However, I first needed to alter the city level and global level average temperature columns to distinct labels since both contained a similar column label. I altered the column labels as shown underneath.
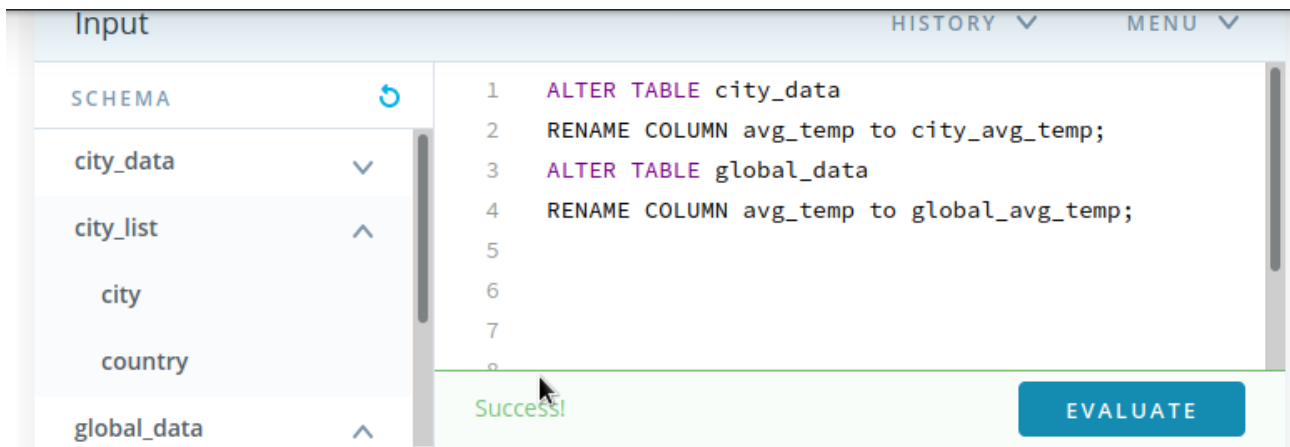
After changing the column labels, proceeded and wrote a query to combine the data as a single unit.
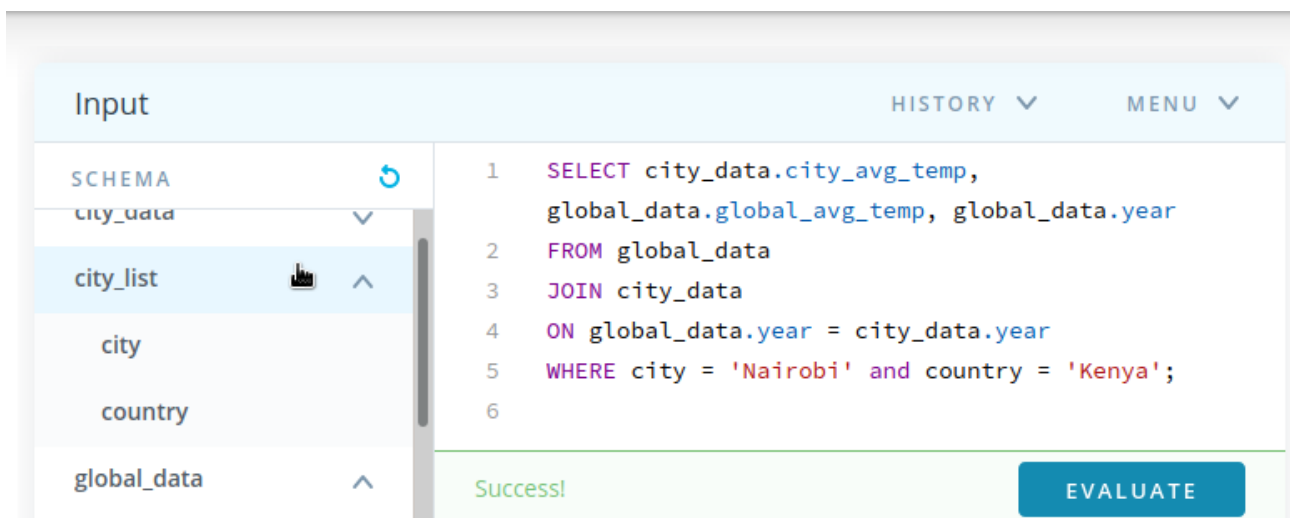
2. Open up the CSV

The tool used : Jupyter notebook, Python Pandas

Using the above query I managed to extract all the required data to a CSV file. In Jupyter notebook, I was able to read and open the CSV as shown underneath.

```python
In [1]: import pandas as pd

df = pd.read_csv('results.csv')
df.head()
```

Out[1]:

	city_avg_temp	global_avg_temp	year
0	15.33	7.90	1850
1	NaN	8.18	1851
2	NaN	8.10	1852
3	NaN	8.04	1853
4	NaN	8.21	1854

```python
In [4]: df.columns[df.isnull().any()]
```

Out[4]: Index(['city_avg_temp'], dtype='object')

Also, I checked for the presence of columns that might have had null values. The city average temperature column contained NaN values that I did replace with the column mean as follows;

```python
In [6]: mean = df['city_avg_temp'].mean()
mean
```

Out[6]: 16.079716312056735

```python
In [9]: df['city_avg_temp'].fillna(mean,inplace=True)
```

Confirmed the changes as shown.

```python
In [10]: df.head()
```

Out[10]:

	city_avg_temp	global_avg_temp	year
0	15.330000	7.90	1850
1	16.079716	8.18	1851
2	16.079716	8.10	1852
3	16.079716	8.04	1853
4	16.079716	8.21	1854

```python
In [11]: df.columns[df.isnull().any()]
```

Out[11]: Index([], dtype='object')

Having a good looking dataset I proceeded and saved it to a CSV to be used in the comparison.

```
In [16]: df = df.to_csv('results_edited.csv')

In [ ]:
```

<h1 style="text-align:center">3.Creating a line chart</h1>

The tool used : Google sheets

I used google sheets to calculate a 5-year moving average. To achieve this I added new columns to store the moving average for both the city level and global level. The results were as below.

A	B	C	D	E	F
	city_avg_temp	city_moving_avg	global_avg_temp	global_moving_avg	year
0	15.33		7.9		1850
1	16.07971631		8.18		1851
2	16.07971631		8.1		1852
3	16.07971631		8.04		1853

The next step was to calculate the moving average for the two levels as shown underneath.

fx =AVERAGE(B151,B150,B149,B148,B147)

A	B	C	D	E	F
149	16.7		9.29		1999
150	16.73	16.592	9.2	9.28	2000
151	16.1	16.634	9.41	9.25	2001
152	16.61	16.566	9.57	9.324	2002
153	16.87	16.534	9.53	9.398	2003

Using the moving average columns I inserted a line chart. Below is a visual of the chart.

A Moving Average Comparing City's emperatures with the Global Temperatures For Nairobi City
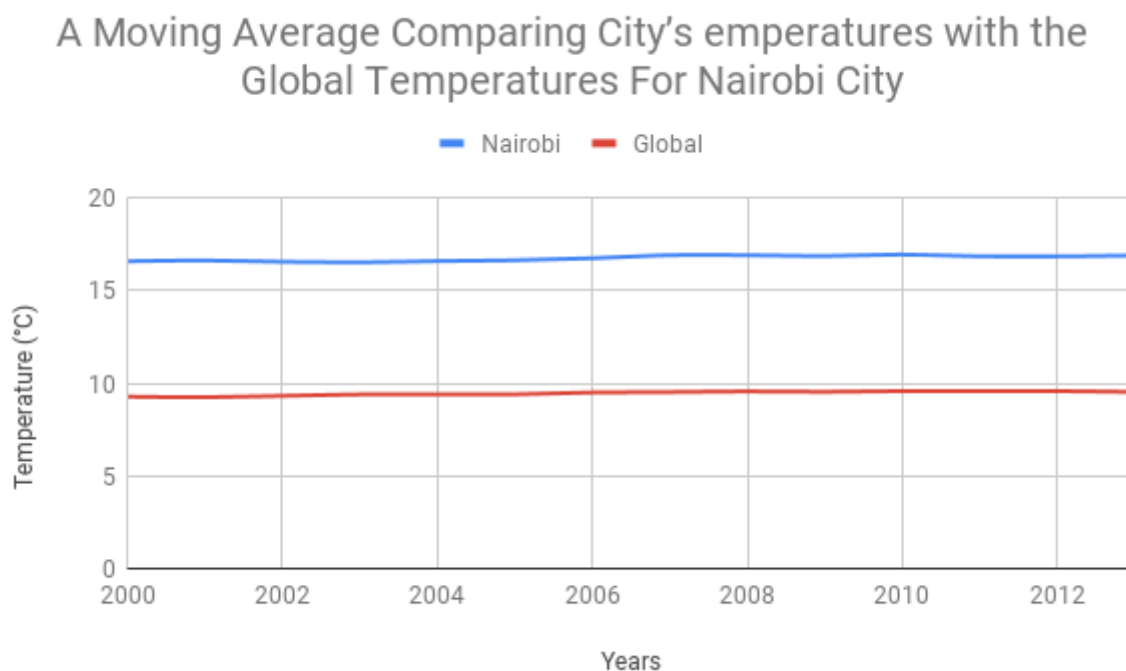
Observations

Nairobi city appears to be hotter on average when compared to the global average. The difference is observed to be consistent over time with a minimal increase and decrease in temperatures.

From the chart, it is quite clear that the temperature change at both levels over the years has been changing gradually however the change is very minimal

The overall trend as visualised indicates that the world is getting hotter over the years

The global temperature averages between 2006 and 2012 have been constant unlike the city temperatures that fluctuate over the years

Conclusion

Temperatures are slightly increasing due to change in climate