Explore Weather Trends

Udacity - Data Analyst Nanodegree

Amdjed Bensalah
05th May 2020

Summary:

I will examine local (Algiers city) and global temperature data in this project, and compare weather trends with global average temperature trends overall.

Goals:

1. Get the data to analyze by writing SQL query, and then extract it from the Udacity portal in .csv format.
2. Manipulating CSV files with Google Sheets and drawing charts with it.
3. To draw smooth graphs we measure the moving average. I used a moving average of 10 years starting from 1913.

Questions to be asked:

1. Is Algiers city on average hotter or cooler compared to the global average? Has the difference remained persistent over time?

2. How do variations in city temperatures in Algiers correlate with changes in the global average over time?

3. What's the overall trend looking like? Will the world get warmer or cooler? Was the pattern over the last couple of hundred years consistent?

Used tools:

- SQL: To extract the data from the database.
- Python: To calculate moving average and plot the charts.
- Jupyter Notebook: To write code and observe.
- MS Excel/ Google Sheets: To have a look at the dataset.

Methodology:

Step 1: **Extraction of data from given database**

1. Get a list of Cities in India from the database.

```sql
SELECT *
    FROM city_list
    WHERE country LIKE 'Algiers'
```

2. I found from the SCHEMA that both city_data and global_data contains same named column 'avg_temp', so I have changed the names respectively

```sql
ALTER TABLE city_data RENAME COLUMN avg_temp to CAT;
    -- CAT = City Average Temp.
ALTER TABLE global_data RENAME COLUMN avg_temp to GAT;
    -- GAT = Global Average Temp.
```

3. Join the tables:

```sql
SELECT global_data.year, global_data.GAT, city_data.CAT
    FROM global_data JOIN city_data -- Joining tables
    ON global_data.year = city_data.year -- reference for join
    WHERE city LIKE 'Algiers';
```

4. Download the CSV file 'results.csv'

Step 2: Python Code for Making Line Chart

So I have used some python libraries here such as: NumPy, Pandas, Matplolib
I have written these codes on Jupyter Notebook.

```python
# Importing the important Libraries
import numpy as np
import pandas as pd                          # For loading data into the notebook
from matplotlib import pyplot as plt         # For making line chart

# Importing the extracted Data Set
data = pd.read_csv("results.csv")
```

I defined a function to calculate moving average:

```python
# function that calculates the MOVING AVERAGE
def moving_avg(mA_range, data_input):
        output = data_input.rolling(window = mA_range, on = "cat").mean().dropna()
        return output
# Function Calling with the range of Moving Average
mA_value = 150
chart_moving_avg = moving_avg(mA_value, data)
# Drawing the graph: Global Temperature
plt.plot(chart_moving_avg ['year'], chart_moving_avg ['gat'], label = 'Global')
plt.legend()
plt.xlabel ("Years")
plt.ylabel ("Temperature (°C)")
plt.title ("Global Average Temperature")
plt.show ()
```

Now we have the following chart:
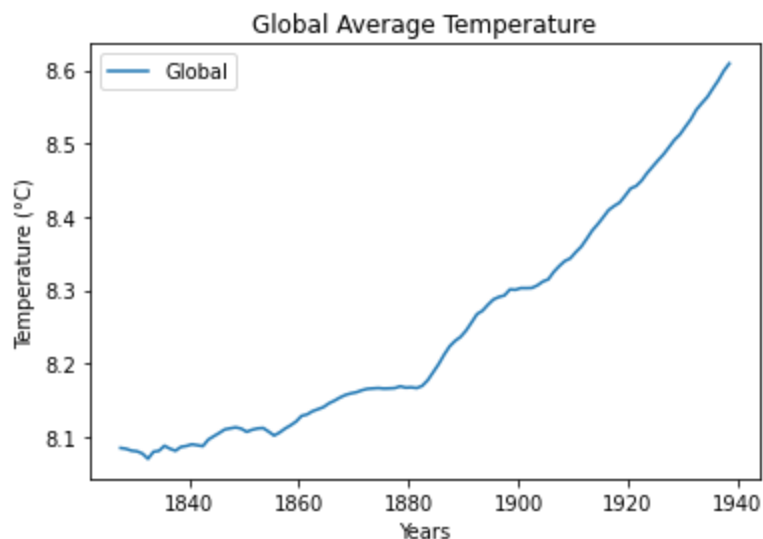
I have separately analysed the global data in order to check and distinguish it from

combined data of Algiers and Global Average temperatures.
Drawing the graph: Algiers and Global Temperature:

# Introducing this line in the previous code above the "global_data plot command"
plt.plot(chart_moving_avg ['year'], chart_moving_avg ['cat'], label = 'Algiers')

Which gives us:
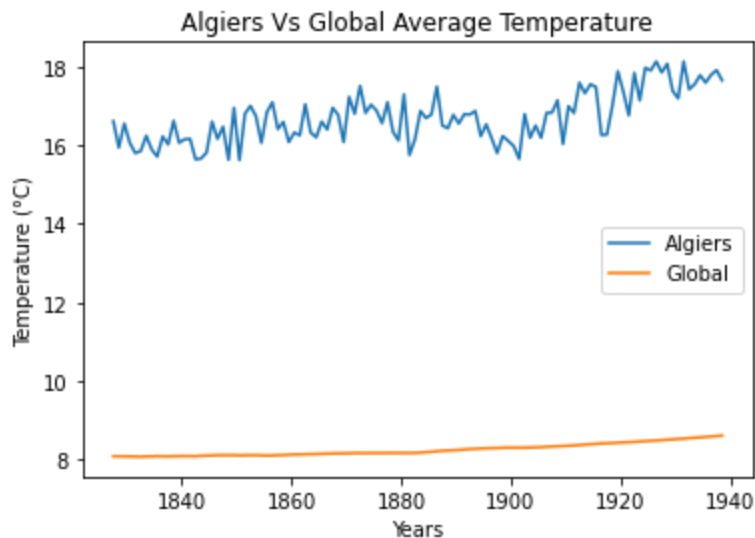
I have also analysed further by using:
data.head (10)
data.tail (10)

Conclusion:

- Algiers: is much warmer than the global average.
- Algiers is the capital of Algeria situated in north Africa where temperature is usually high.
- The global temperature appears to fall after 1950 and then we notice a sudden increase.
- Considering the first chart, we have a clear understanding that due to industrialization the global temperature rises at a very high rate.