Exploring Weather Trends

UDACITY - DATA ANALYST NANODEGREE

Mohamed Balam |Project-1|Benghazi, Libya |6th-Feb-2020

Overview:

In this report I compared the global average temperature to the average temperature of my city Benghazi by analyzing the data available in the database, which is provided by Udacity, and then drawing the graph to clarify, and finally the observation and summary.

Goals:

1. An overview of all the contents of the attached database to bring in the most important elements that aid analysis and comparison.
2. Fetch data from the database using SQL commands.
3. Save the data as a .csv file.
4. Analyze and show results.

Tools Used:

1. SQL statements: Their purpose is to retrieve the data I want from the database.
2. Python: To write the code that allows me to read data, analyze it, and draw charts using libraries like pandas, matplotlib.
3. ANACONDA and Jupyter Lab: To install packages, write and modify the Python code and print the results.
4. Microsoft excel: To look at the data before and after the analysis process.

Executing Steps:

I followed certain steps in this project, I learned from the lessons presented in this course, and I also used some help resources, which I will mention in the references section.

Step 1- Fetch data from database that given from Udacity

I have good experience in using SQL statements so I did not have any problems except for renaming columns, which I took the idea from one of the references that I will mention later, so I used these steps To see the available cities of my country "Libya" in this database:

1. See what Libyan cities are in the database using this statement:

SELECT * FROM city_list WHERE country LIKE 'Libya'

2. Due to the similarity of the column names in the "city_data" and "global_data" tables in the "avg_temp" column, I renamed the columns to help me in the process of merging the two tables together later so I used these statements:

ALTER TABLE city_data RENAME COLUMN avg_temp to BEN_AT;

ALTER TABLE global_data RENAME COLUMN avg_temp to GLB_AT;

3. Here I wrote this statement to combine the two tables and retrieve important data from them:

SELECT city_data.BEN_AT, global_data.GLB_AT ,global_data.year

FROM city_data JOIN global_data

ON city_data.year = global_data.year

WHERE city LIKE 'Benghazi';

After that the button to save the file as csv has been activated and I have saved it as BENvsGLB.csv

Step 2- Write code using Python

You will see here that I used some of the libraries that I learned how to use, after watching the course videos of the data analysis lesson, which I had little experience with, then after that I opened the jupyter-lab and wrote this code:

```python
import pandas as pd
import matplotlib.pyplot as plt


df = pd.read_csv( " BENvsGLB.csv" )

df.head()

#Import libraries and read data file and test if readed by df.head()
```


After that I wrote a function that calculates the moving average, I used some references here to see some of the auxiliary functions, here is the code:

```
def cal_mov_avg (movingAVGR, dataFile):
output = dataFile.rolling(window = movingAVGR, on = "ben_at" ).mean().dropna()
return output
```

Then I called the function to execute:

chart = cal_mov_avg(130, df)

Now it is the turn of the great library Matplotlib where I wrote the following code to draw the following two charts:

```
plt.figure(figsize=[15,10])
plt.plot(chart [ 'year' ],chart [ 'glb_at' ], label = 'Global' )
plt.legend()
plt.grid(True)
plt.xlabel ( "Years" )plt.ylabel ( "Temperature (°C)" )
plt.title ( "Global Average Temperature" )
plt.show ()
```
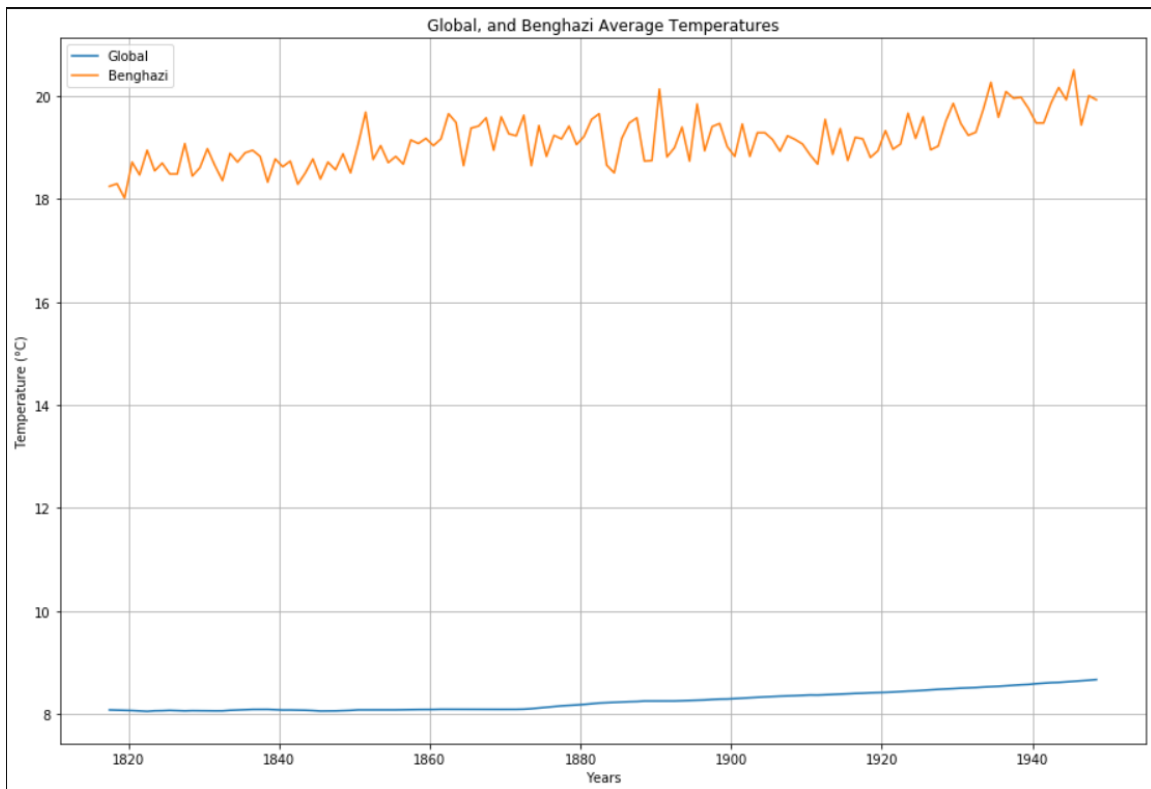
So, I got the following output:
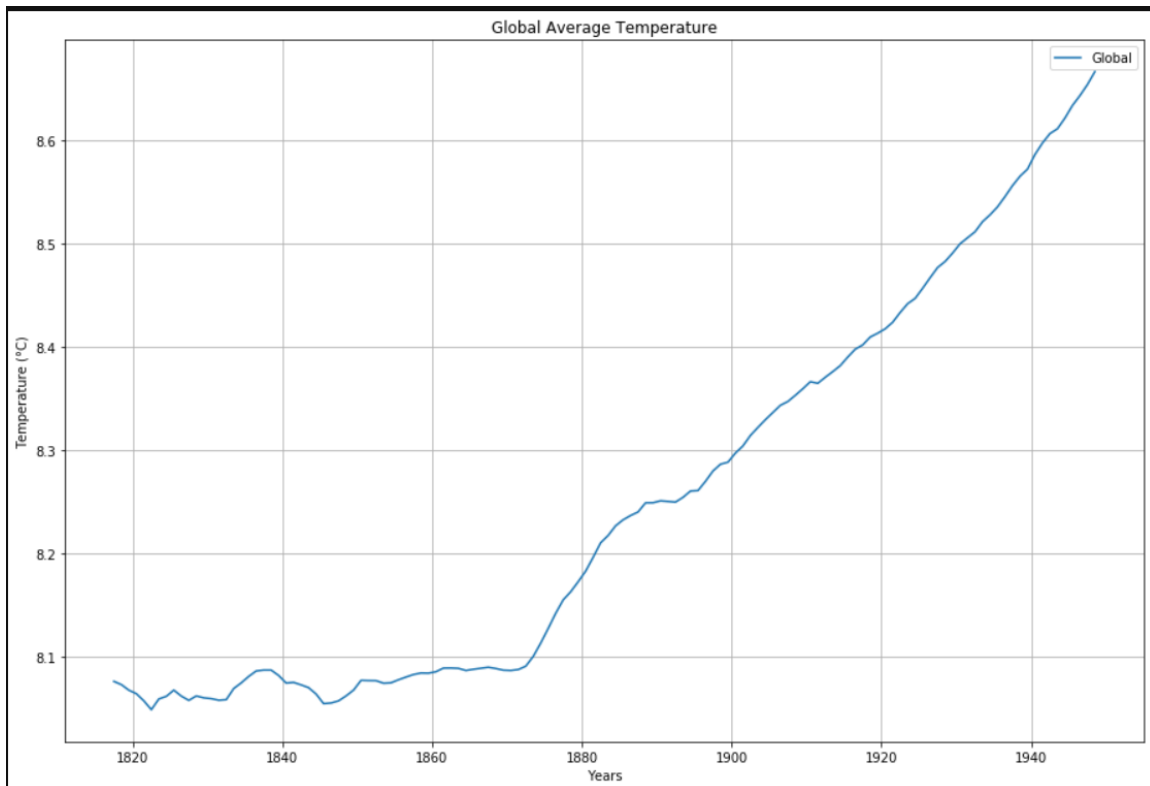
```
plt.figure(figsize=[15,10])
plt.plot(chart [ 'year' ],chart [ 'glb_at' ], label = 'Global' )
plt.plot(chart [ 'year' ],chart [ 'ben_at' ], label = 'Benghazi' )
plt.legend()
plt.grid(True)
plt.xlabel ( "Years" )plt.ylabel ( "Temperature (°C)" )
plt.title ( " Global, and Benghazi Average Temperatures)
plt.show ()
```

I got the following graph:

Result- Observations

You notice here that the value of "movigAVGR" which I gave **130** because I found small values such as 50 or 30 do not give the best results in the graph.

Observation from the Line Charts:

1. From the first chart, we note that the average temperature level was approximately balanced between 8.1 to 8.2, then after that it increased significantly, especially after the year 1870, to rise by about 0.1 every year.
2. From the second chart, we note the vast difference between the two lines, I think that this is due to the region in which the city of Benghazi is located, as the Mediterranean region is warm, and it is hot in summer.
3. The global average temperature contains many cold countries, and we know that the sum and number of values affect the mean value.
4. Also, from the two charts, we note that temperatures, as the last years recorded in the database have increased significantly from other years, I think it is because of the global warming due to the increase in number of factories and the expansion of the ozone depletion, "**df.tail (20)**" this function came in accordance with my observations when it showed me the last 20 recorded values.

According to the simple experience that I have, and after learning some basics from this program until this moment, so I can say that these are all my observations of this data after analyzing it.

Key consideration:

1. Writing functions makes your code working well an easy to debug.
2. Moving average was applied on Benghazi data "City data" to get a good visualize when drawing a chart.
3. Matblotlib: used for making a visualization.
4. BEN_AT and GLB_AT: it means Benghazi Average Temperatures, and Global Average Temperatures.

References:

1. Cover image from the amazing photographer Mohamed Badi : https://twitter.com/Mohameed_Badi/status/1197657958525612038/photo/1.
2. Rename columns names: https://www.1keydata.com/sql/alter-table-rename-column.html.
3. Auxiliary functions for calculating the moving average, and drawing charts using Pandas and Matplotlib: https://www.datacamp.com/community/tutorials/moving-averages-in-pandas.
4. Udacity: Nanodegree course, section 2, Introduction to Data Analysis, Data analysis process lesson.