

Explore Weather Trends

11/10/2018

Udacity - Data Analyst Nanodegree

Saeed Mohammad yousef Alfalana

Ramallah, Palestine



Note: - I have been search in Udacity database For **Palestine Country** or **Jordan** I cannot find it so I choose **United States country - Alexandria city.**

NOTE: In this project, I have taken help from various resources in order to make use of some libraries, Keywords and commands correctly; I have listed links or resources in the references.

TERM-1, Project-1, Explore Weather Trends

Overview

In this project, I have analysed local temperature of Alexandria City, Alexandria in accordance with the global temperature data and compared.

I had been provided with a database on Udacity portal from where I have to extract, manipulate and visualize the data as in the following goal

Goals

Extraction of data from the database and export to CSV file making a chart

Visualization based on extracted data Observation based on chart.

1-What tools did you use for each step? (Python, SQL, Excel, etc)?

- SQL: To extract the data from the database
- Python: For calculating moving average and plotting line chart
- ANACONDA - Jupyter Notebook: For writing python code and making observations
- Excel Sheets: Having a look at the data and writing project and calculate moving average.

Explore Weather Trends

11/10/2018

Udacity - Data Analyst Nanodegree

Saeed Mohammad yousef Alfalana

Ramallah, Palestine



2-How did you calculate the moving average?

By using Excel, I used this command:

256	2004	9.32	12.79		
257	2005	9.7	12.82		
258	2006	9.53	13.59		
259	2007	9.73	13.13		
260	2008	9.43	12.76		
261	2009	9.51	12.44		
262	2010	9.7	13.28		
263	2011	9.52	13.56		
264	2012	9.51	14		
265	2013	9.61	14.19		

10 year moving average from 2013	
temp_global	temp_city
9.556	13.256
AVERAGE(B256:B265)	AVERAGE(C256:C265)

3-What were your key considerations when deciding how to visualize the trends?

1. Unit of Temperature (°C): Centigrade, on Y-axis
2. Years shown on X-axis
3. Use of Matplotlib library for visualization
4. Applied moving average on City data in order to get relatively smooth line
5. Defined a function for easy code
6. Saved all of the codes in .ipynb files (Jupyter Notebook) for later reference and regenerations or revisions.

Explore Weather Trends

11/10/2018

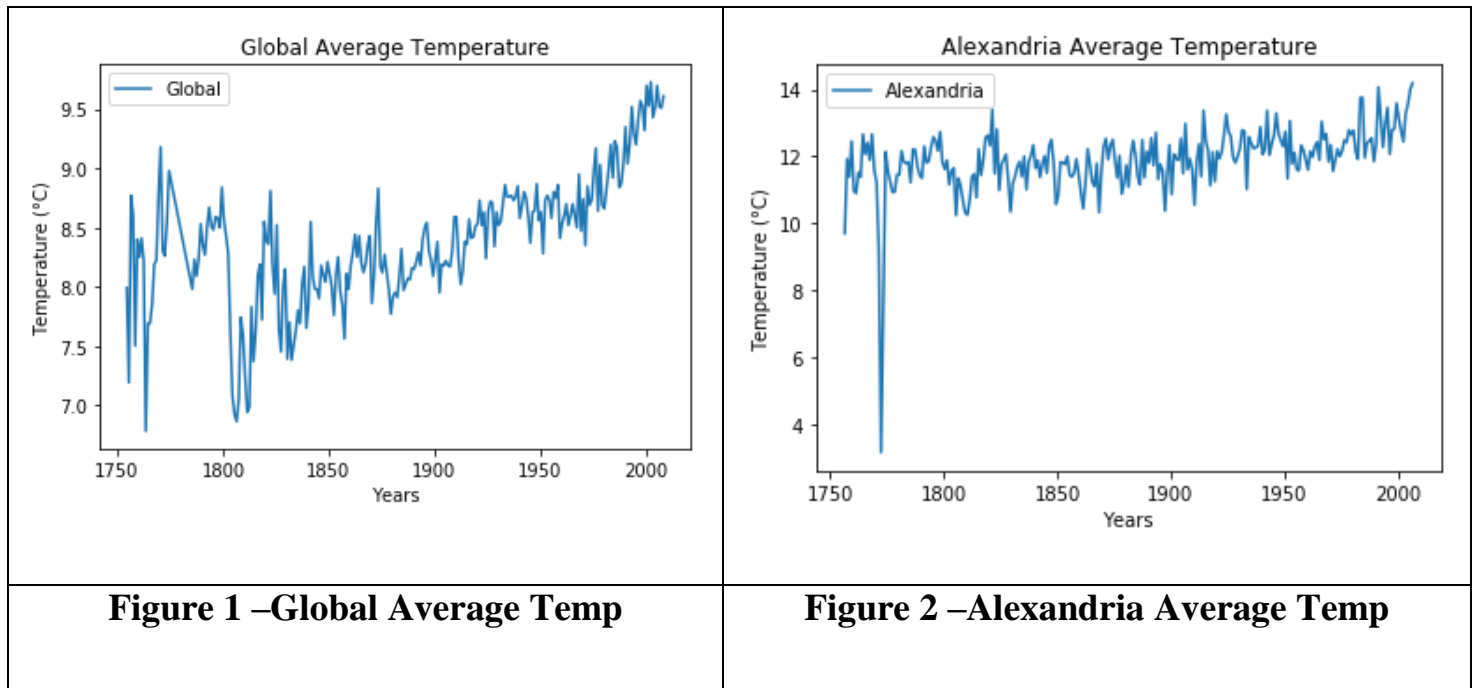
Udacity - Data Analyst Nanodegree

Saeed Mohammad yousef Alfalana

Ramallah, Palestine



4-Line chart with local and global temperature trends



5-RESULT - Observations:

1. I have observed that, if I choose a short range for moving average, say 5, I will get messy line in the graph; also, the range of the “Years” on x-axis becomes short, and if I use a larger moving average, say 180, I will get relatively smooth graph and range of x-axis is longer.
2. if I use a larger moving average, say 180, I will get relatively smooth graph and range of x-axis is longer.

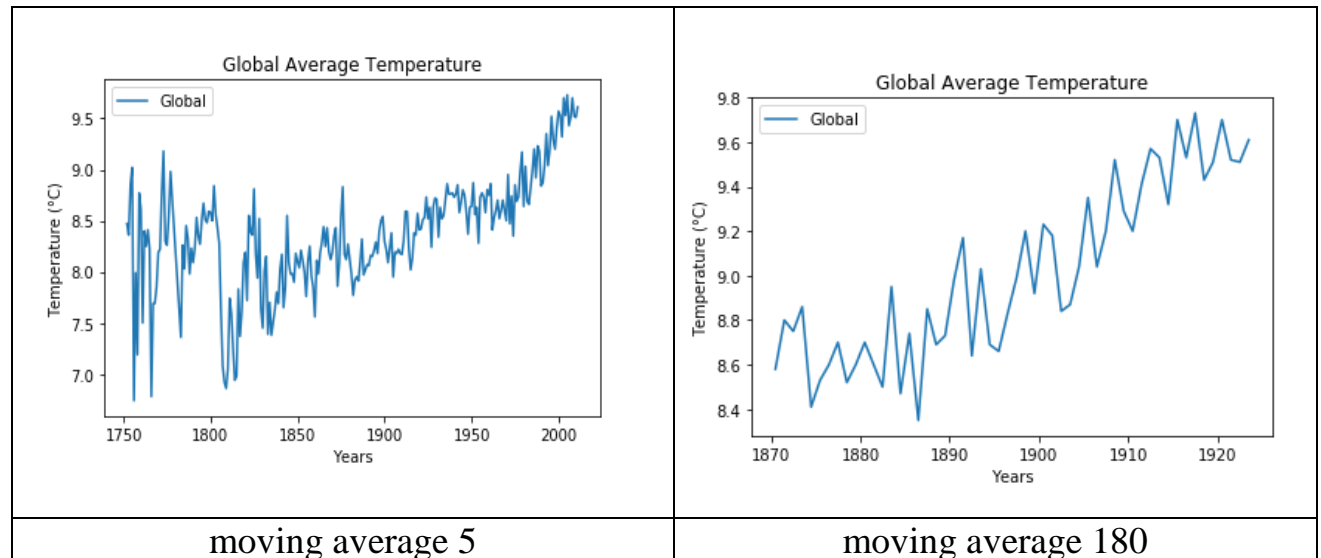
Explore Weather Trends

11/10/2018

Udacity - Data Analyst Nanodegree

Saeed Mohammad yousef Alfalana

Ramallah, Palestine



3. the temperature of the world is on constant rise.
4. The chart of Alexandria Vs Global Temperature: Very big difference between the average temperature of Alexandria and that of Global.
5. From the global graph Figure 1: I noticed that global temperature is increasing quiet constantly with years.

To further verify this I have used the command: `data.tail (264)` which displayed me the 264 rows from the last and I have seen the same insight as in the graph.

After 1750, the global temperature (with a sudden change) started increasing at a higher rate. This may be an effect of worldwide industrialization and newer technologies. Global warming comes to play from the year of 1750.

6. Again coming back to the Figure 2, I observe that Alexandria have temperatures greater than the global average.
7. the local trend is increasing more rapidly than global trend

Explore Weather Trends

11/10/2018

Udacity - Data Analyst Nanodegree

Saeed Mohammad yousef Alfalana

Ramallah, Palestine



STEP 1 - Extraction of Data from provided Database

I have done the following activity in order to make a relevant dataset.

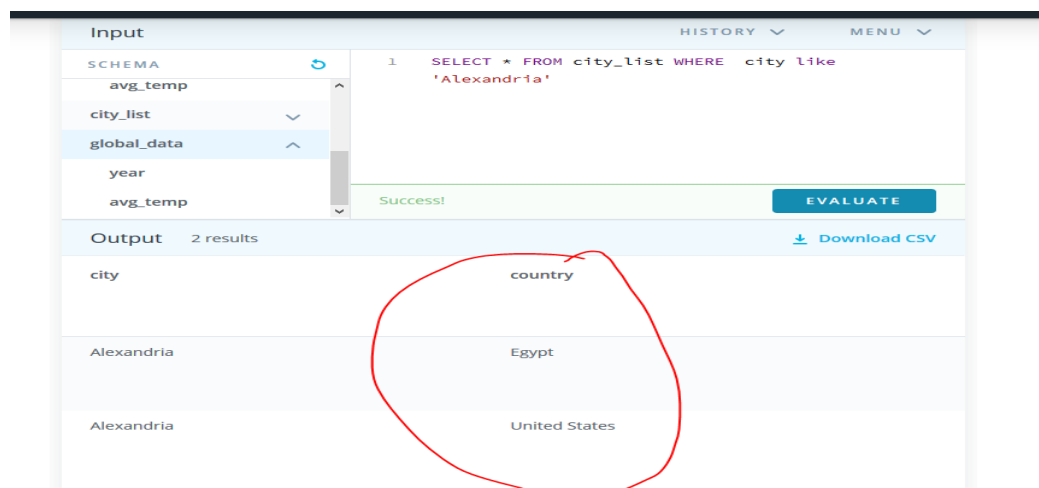
I have learn the SQL basics from lessons provided before this project. I have also done an introductory course on SQL and relational database from which I have used some concepts.²

1. *To see which cities are available for "United States" in the given dataset:*

SELECT * FROM city_list WHERE country LIKE 'United States'

When I try to search 'Alexandria' city to compare temp V/S Global I have Duplicate in year the reason was 'Alexandria' city was found in 'Egypt' And 'united states' I used this SQL command

SELECT * FROM city_list WHERE city like 'Alexandria'

A screenshot of a web-based SQL query editor. The 'Input' section shows a schema with tables 'avg_temp', 'city_list', 'global_data', 'year', and 'avg_temp'. The query entered is 'SELECT * FROM city_list WHERE city like 'Alexandria''. The 'Output' section shows '2 results' with columns 'city' and 'country'. The results are 'Alexandria' from 'Egypt' and 'Alexandria' from 'United States'. A red circle highlights the 'country' column and its values.

city	country
Alexandria	Egypt
Alexandria	United States

So I write this SQL command to Extract correct data

SELECT * FROM city_list WHERE country LIKE 'United States' and city like 'Alexandria'

Explore Weather Trends

11/10/2018

Udacity - Data Analyst Nanodegree

Saeed Mohammad yousef Alfalana

Ramallah, Palestine



2. I know that I can make a relevant dataset by joining the two tables. But, I found from the SCHEMA that both city_data and global_data contains same column named 'avg_temp', so I have changed the names of the columns respectively in order to have distinct columns.

```
ALTER TABLE city_data RENAME COLUMN avg_temp to  
temp_city;
```

```
// temp_city = City Average Temp.
```

```
ALTER TABLE global_data RENAME COLUMN avg_temp to  
temp_global;
```

```
// temp_global = Global Average Temp
```

3. Now I have written following code in order to join the two tables and have the relevant data:

```
SELECT global_data.year, global_data. temp_global, city_data.  
temp_city
```

```
FROM global_data JOIN city_data ON global_data.year =  
city_data.year
```

```
WHERE city LIKE 'Alexandria' and country LIKE 'United States';
```

I have an option of downloading the file as CSV format. Downloaded as "results.csv" file.

Explore Weather Trends

11/10/2018

Udacity - Data Analyst Nanodegree

Saeed Mohammad yousef Alfalana

Ramallah, Palestine



STEP 2 - Python Code for Making Line Chart

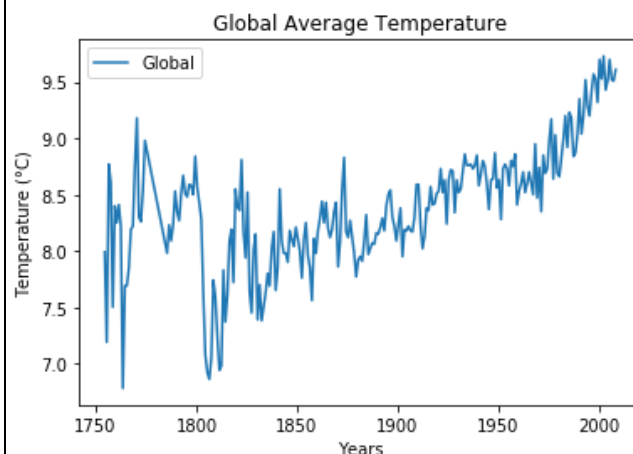
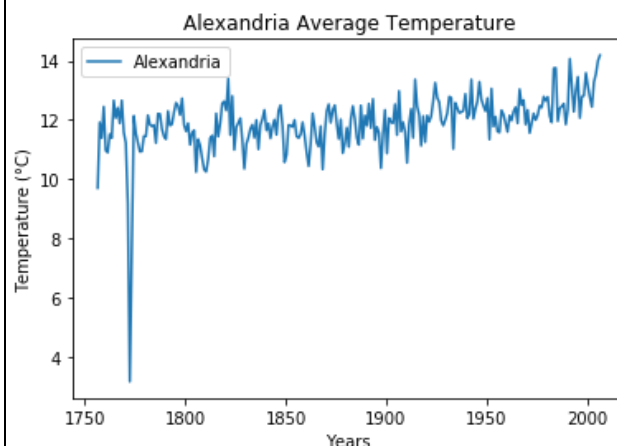
So I have used some python libraries here, which I have learn when I was researching a lot about data science I have written these codes on Jupyter Notebook.

I have separately analysed the global data in order to check and distinguish it from combined data of Alexandria and Global Average temperatures.

```
# 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")
# function that calculates the
MOVING AVERAGE (global)
def moving_avg(mA_range,
data_input):
    output = data_input.rolling(window
= mA_range, on =
"temp_global").mean().dropna()
    return output
```

```
# Drawing the graph: Global
Temperature
mA_value = 5 #global 9.556
chart_moving_avg =
moving_avg(mA_value, data)
plt.plot(chart_moving_avg ['year'],
chart_moving_avg ['temp_global'],
label = 'Global')
plt.legend()
plt.xlabel ("Years")
plt.ylabel ("Temperature (°C)")
plt.title ("Global Average
Temperature")
```

So I have got the following output:



Explore Weather Trends

11/10/2018

Udacity - Data Analyst Nanodegree

Saeed Mohammad yousef Alfalana

Ramallah, Palestine



UDACITY

```
plt.show ()
data.tail(487)

# function that calculates the
MOVING AVERAGE (local)
def moving_avg_local(mlocal_value,
data_input):
    output = data_input.rolling(window
= mlocal_value, on =
"temp_city").mean().dropna()
    return output
# Drawing the graph: alexandria
Temperature
mlocal_value =14 #local 13.256
chart_moving_avg_local =
moving_avg_local(mlocal_value,
data)
plt.plot(chart_moving_avg_local
['year'], chart_moving_avg_local
['temp_city'], label = 'Alexandria')
plt.legend()
plt.xlabel ("Years")
plt.ylabel ("Temperature (°C)")
plt.title ("Alexandria Average
Temperature")
plt.show()
data.tail(10)
```

	year	temp_global	temp_city
254	2004	9.32	12.79
255	2005	9.70	12.82
256	2006	9.53	13.59
257	2007	9.73	13.13
258	2008	9.43	12.76
259	2009	9.51	12.44
260	2010	9.70	13.28
261	2011	9.52	13.56
262	2012	9.51	14.00
263	2013	9.61	14.19

Explore Weather Trends

11/10/2018

Udacity - Data Analyst Nanodegree

Saeed Mohammad yousef Alfalana

Ramallah, Palestine



REFERENCES:

Sql Command and Install it (localy)

<https://docs.microsoft.com/en-us/sql/ssms/download-sql-server-management-studio-ssms?view=sql-server-2017>

<https://docs.microsoft.com/en-us/sql/advanced-analytics/r/sql-server-r-services?view=sql-server-2017>

<https://www.microsoft.com/en-us/sql-server/sql-server-downloads>

<https://sqlchoice.azurewebsites.net/en-us/sql-server/developer-get-started/python/windows/step/2.html>

learn pandas numpy matplotlib

<https://pandas.pydata.org/pandas-docs/stable/10min.html>

installing Jupyter Notebook

<https://jupyter.readthedocs.io/en/latest/install.html>

ANACONDA

<https://www.anaconda.com/download/>

<https://www.youtube.com/watch?v=T8wK5loXkXg>

https://www.youtube.com/watch?v=GqdgxA_Bgz8

<https://www.lynda.com/-tutorials/Windows-command-line-basics/497312/513424-4.html>

Explore Weather Trends

11/10/2018

Udacity - Data Analyst Nanodegree

Saeed Mohammad yousef Alfalana

Ramallah, Palestine



other REFERENCES: sure from udacity classroom