Data Analyst Nanodegree

Explore Weather Trends

Project #1

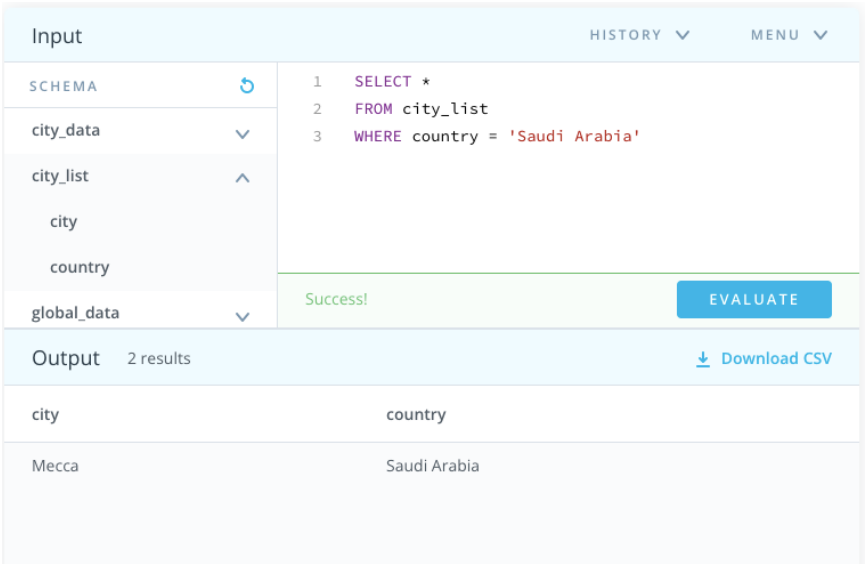
By Sultan Alharbi

1- Tools used in this project:

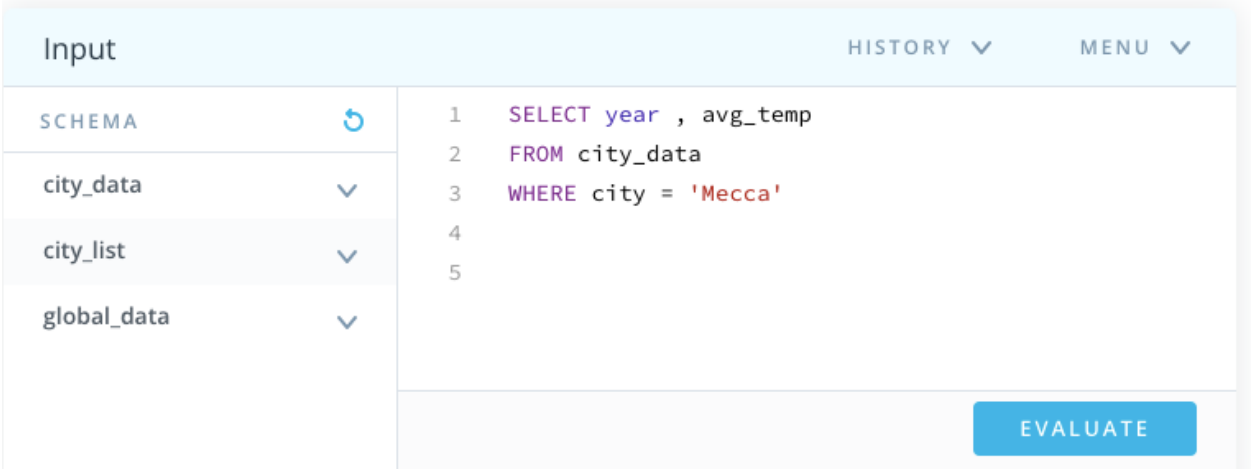
The tools I have used in this project are SQL , Python , Pandas, Seaborn and Matplotlib libraries.

2- Data extraction:

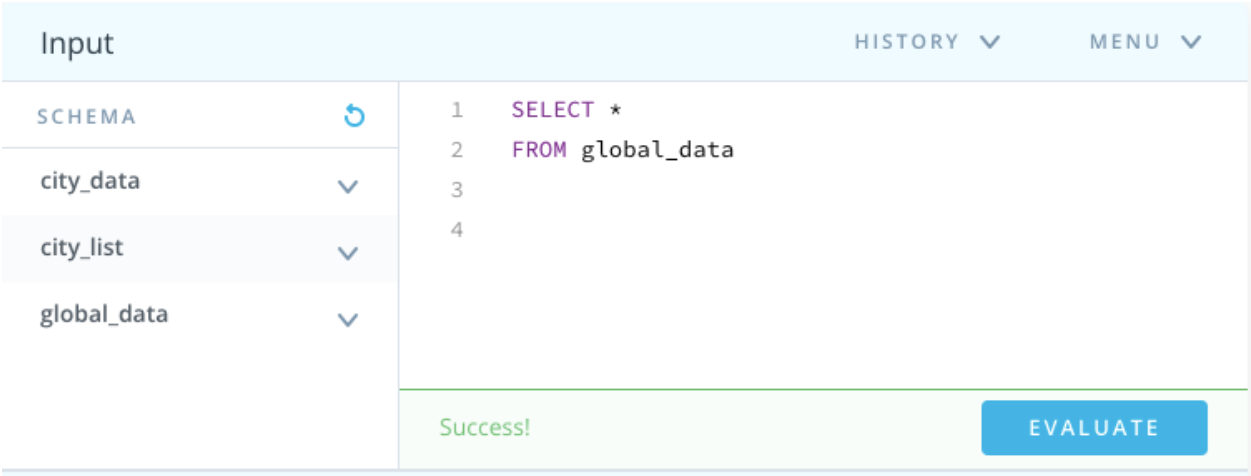
I have extracted the data by typing queries in SQL workspace provided by Udacity classroom, at first, I had to check what are the cities from my country that were in the data, so I did the following query:



As you can see there were two results in the output section, first one was Mecca and the second one was Riyadh, and since I'm from Jeddah and the closest city to Jeddah is Mecca, I have chosen to work on Mecca's data rather than Riyadh's.



Before doing this query I did (SELECT *) , but the output was 4 columns , (year , city , country , avg_temp) , and since the columns city and country will always returns Mecca and Saudi Arabia , I have decided to not include it in the data, that's why I did (SELECT year , avg_temp) .



For the global data I didn't do much , I just select everything from (global_data) and download it.

3- Data importing and preprocessing:

I did import Pandas and Matplotlib which I'll use in the project, after that I have imported the two .CSV files that I have download it from SQL workspace, I have used pandas function .head() to check the first five rows in City dataframe , and I have found out that there was NaN values, so I had to deal with it by dropping the rows that contains NaN values.

```
In [10]: import pandas as pd
         import seaborn as sns
         import matplotlib.pyplot as plt
```

```
In [6]: city = pd.read_csv('/Users/SulMac/Desktop/B.S in AI/Data Analyst/ExploreWeatherTrends/Data/City.csv')
        glo = pd.read_csv('/Users/SulMac/Desktop/B.S in AI/Data Analyst/ExploreWeatherTrends/Data/global.csv')
```

```
In [7]: city.head()
```

Out[7]:

	year	avg_temp
0	1843	25.16
1	1844	19.05
2	1845	22.46
3	1846	NaN
4	1847	NaN

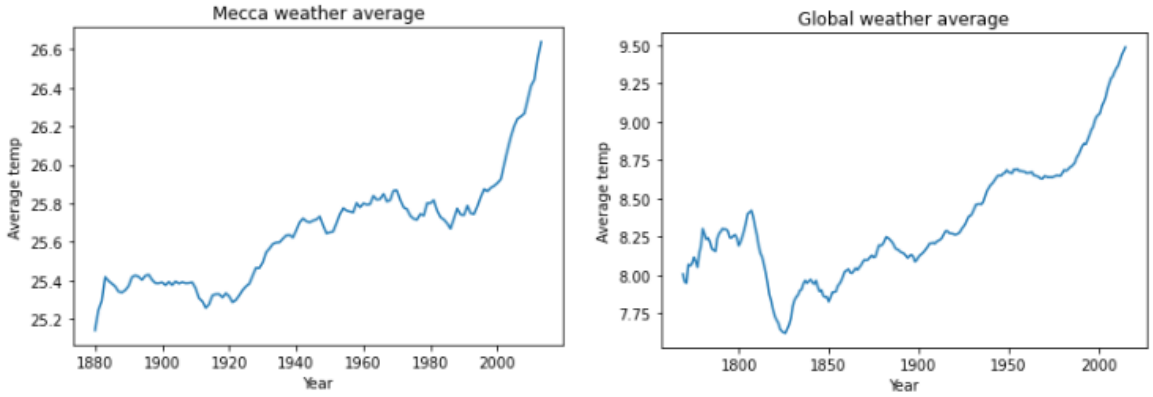
```
In [17]: (city.dropna()).reset_index(drop=True)
```

After that I added a column in both dataframes , which contains the values of Moving average (also knowns as rolling), the columns names were (ma_mecca and ma_glo) and I have calculated the value of moving average using Pandas library function (rolling) .

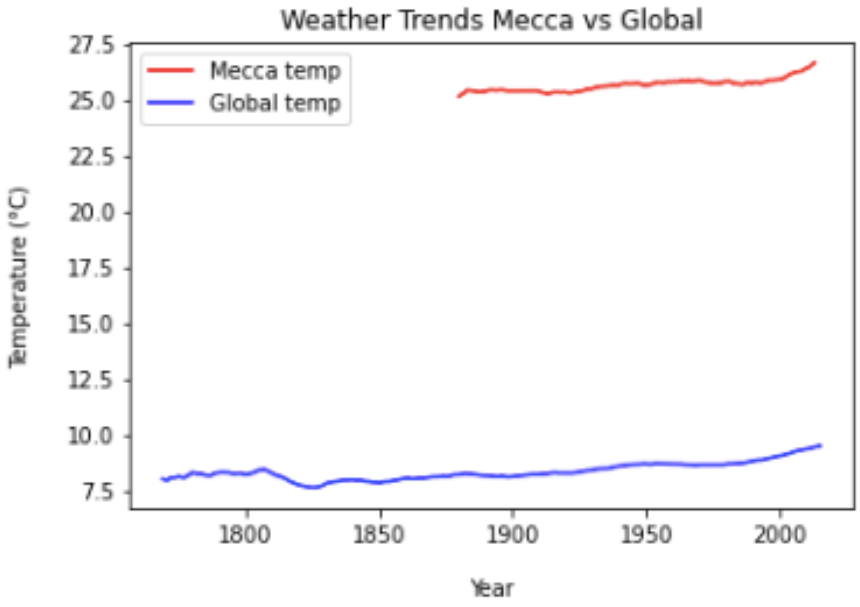
```
city['ma_mecca'] = city['avg_temp'].rolling(window=20).mean()

glo['ma_glo'] = glo['avg_temp'].rolling(window=20).mean()
```

4- Creating Line Chart:

I have made a line chart that shows the weather average for Mecca through the years, and so on for the global weather.



Then I have made a line chart to compare between Mecca average and the global average, by plotting the years on the x-axis and the temperature moving average on the y-axis.



5- Observations:

- Mecca has a higher average temperature than the global average temperature.
- The average temperature in Mecca is 25.6° C and the average global temperature is 8.3° C, based on the moving average .
- The temperature in Mecca has increased through the years, and so on for the global temperature.
- The average temperature will increase in both of Mecca and the world, based on the data that we had.
- The minimum temperature average for Mecca was 25.1° C , and the maximum was 26.6° C .
- The minimum global temperature average was 8.3° C, and the maximum was 9.5° C .