DATA ANALYST NANODEGREE

UDACITY

Explore Weather Trends

07th July 2020

VinothKumar Kolluru

Project - 1, Explore Weather Trends
New York, USA

Project Overview:

There are Three data sets where I can find data related to average temperature of cities and global temperature and my project is to find useful information related to Avg temperature of the city I am currently living with respect to the global temperature with provided data sets.

Goals :

1. Selecting city and country from the database "city_list"
2. Extracting the City level data from the database "city_data" and export to CSV file.
3. Extracting the global temperature from the database "global_data" and export to CSV file.

Tools Used :

1. **SQL** : To extract the data from the database
2. **Google Sheets :**
   > To calculate Moving Averages of global and city temperatures
   > To plot Line Chart

STEP : 1 Extraction of data from the database

1. To check which countries and cities are available in the database.

```
SELECT *
    FROM city_list
    WHERE City ILIKE 'New York'
```

2. To select data from the City database

```
SELECT avg_temp,year,city,country
    FROM city_data
    WHERE city ILIKE 'New York'
```

3. I observed there is a column called avg_temp which is same in both city_data and global_data. I want to change the schema so I joined both the tables and changed the column names in both the databases.

```
ALTER TABLE city_data
    RENAME COLUMN avg_temp to CITY_AVG_TEMP
    ; ALTER TABLE global_data
    RENAME COLUMN avg_temp to GLOBAL_DATA_AVG_TEMP;
```

4. I have joined the two tables using JOIN also called as INNER JOIN as avg_temp is same in both the tables.

```
SELECT city_data. CITY_AVG_TEMP, global_data.
GLOBAL_DATA_AVG_TEMP,global_data.year FROM
global_data
JOIN city_data
ON global_data.year = City_data.year
WHERE city ILIKE 'New York';
```

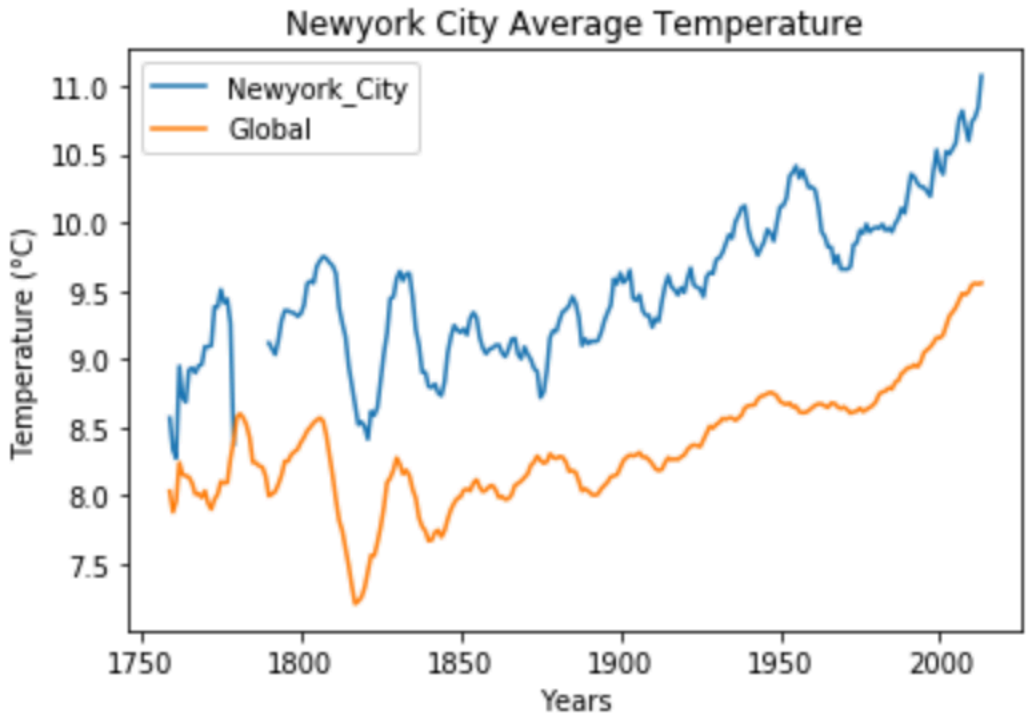
Now, I have got an option to download a CSV file . I downloaded file as “results.csv”

Moving Averages :

1. I’ve used pandas for importing the csv file in python
2. I’ve used .rolling(10).mean() formula for smooth graph
3. Then using matplotlib library I’ve plotted graph between city data and global data
4. I have plotted Line chart for global data separately to observe difference between Global Average Temperature and the city New york.

Considerations:

X-Axis : Years
Y-Axis : Temperature(°C)
Legends are plotted in the charts



Report :

1. Global Average Temperature for 10 yr MA varies between 8.5°C to 9.5°C
2. New York city Average Temperature for 10 yr MA varies between 8.5°C to 11°C

3. The Graph of Newyork Vs Global has some difference in the temperatures.

4. In comparison with Global and New York Average Temperatures, Newyork is hotter than global average temperature .

5. From the first Graph, I observed global temperature is increasing from 8.5 to 9.5. From the second graph I observed the temperature of both global and Newyork average temperatures are ups and downs during the early years, later during 1996 to 2013 both the temperatures increased due to increase in temperature.

6. The final conclusion of this project is Newyork is hotter than global temperature and temperature is increasing day by day due to environmental conditions and global warming.