



WEATHER TRENDS

Kshitiz Tripathi

OVERVIEW-

In the project named explore weather trends, I have analyzed the weather temperature of my city & rest of the world. The data was provided by udacity, and I used different tools for analyzation purpose.

Tools Used: -

1. SQL - for data extraction
2. Python (jupyter notebook, Anaconda) – for analyzation purpose

Project Methods-

Step 1-

To see the available city in the city_list in SQL

```
SELECT *  
FROM city_list
```

Step 2-

To select a data of my city from city_data table

```
SELECT *  
FROM city_data  
WHERE city='New York'  
ORDER BY country;
```

Step 3 –

As observed there is a column named avg_temp available in both the city_data & global_data, So renaming it into different

```
ALTER TABLE city_data  
RENAME COLUMN avg_temp to city_avg_temp;  
ALTER TABLE global_data  
RENAME COLUMN avg_temp to world_avg_temp;
```

Step 4 -

To extract the required data for analysis and visualization.

```
SELECT *  
FROM city_data c  
JOIN global_data d  
ON c.year=d.year  
WHERE c.city='New York'
```

After that I got an option of downloading a csv file.

Step 5 –

Python code for making visualization

➔ Here importing some of the python library/packages

```
import pandas as pd  
import numpy as np  
import matplotlib.pyplot as plt  
%matplotlib inline  
from scipy.stats import pearsonr
```

➔ Here reading the the csv file of weather trends which I had exporeted through SQL

```
df=pd.read_csv('results.csv')  
df.head(2)  
df.info()
```

➔ As there as some missing rows in city_avg_temp, we can see through df.info() method, so filling those rows with overall avg temp-

```
df['city_avg_temp'].fillna(df['city_avg_temp'].mean(),inplace=True)
```

Step 6 –

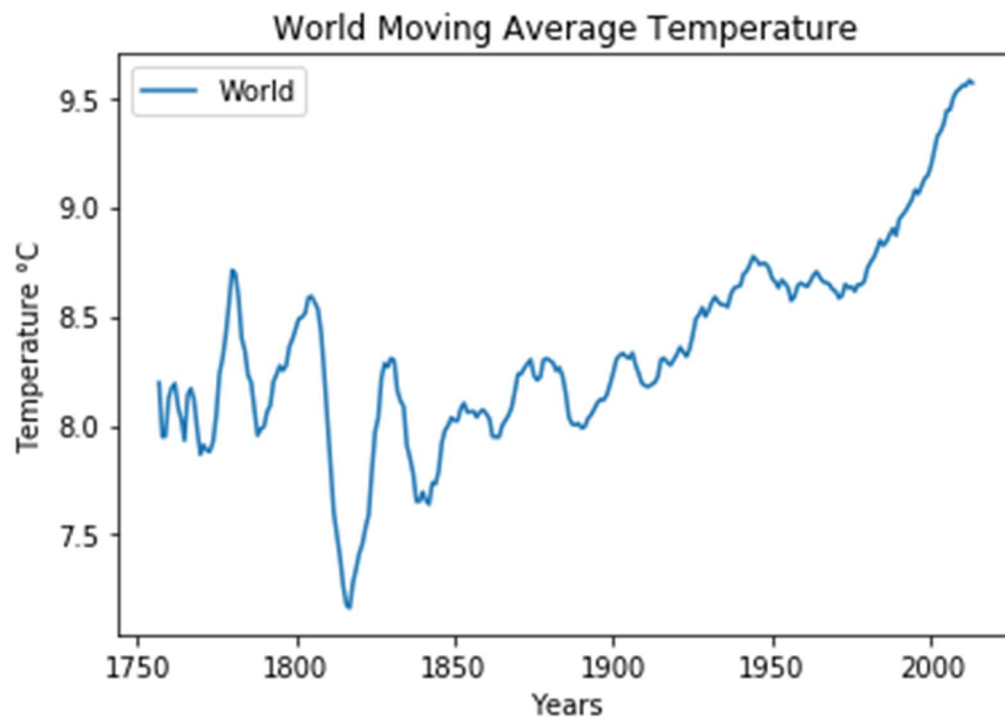
➔ Moving Averages

- Rolling averages has been calculated to to smooth out data and make it easier to observe the trends.
- I have done for 8 years Moving Average for smooth plot
- Python have built-in functions such as ('rolling','mean') used for calculation.

```
world_MV_avg = df.world_avg_temp.rolling(8).mean()  
city_MV_avg = df.city_avg_temp.rolling(8).mean()
```

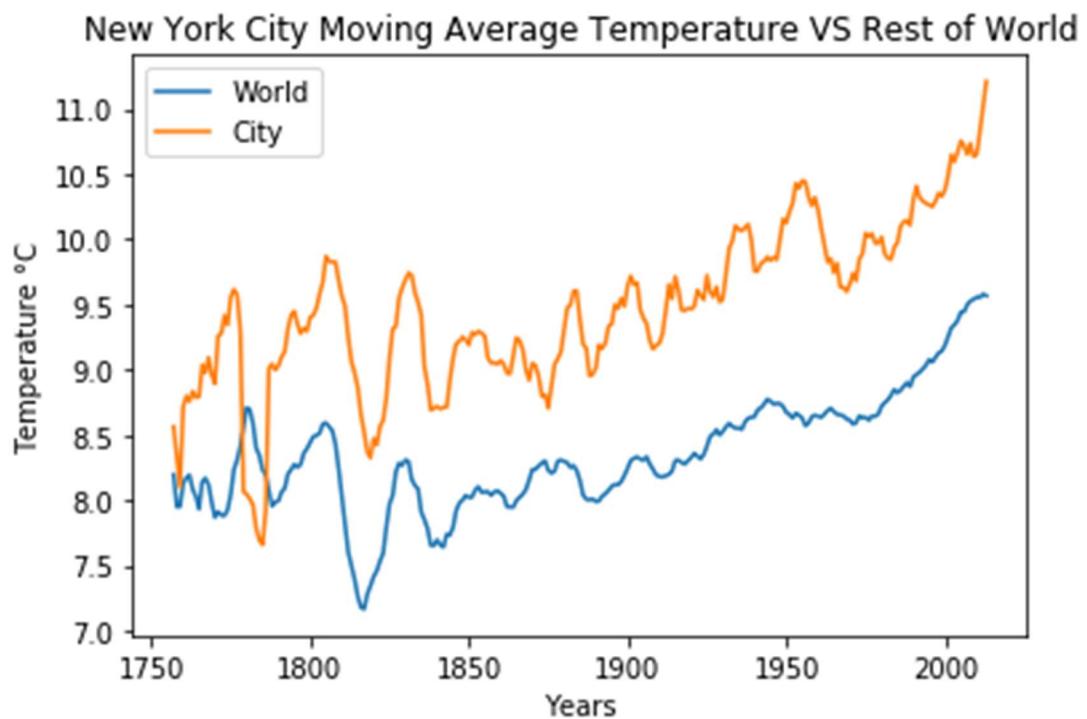
➔ Analysis of global temperature

```
plt.plot(df['year'],world_MV_avg,label='World')  
plt.legend()  
plt.xlabel("Years")  
plt.ylabel("Temperature °C")  
plt.title("World Average Temperature")  
plt.show()
```



➔ Analysis of city VS World temperature.

```
world_MV_avg = df.world_avg_temp.rolling(4).mean()
city_MV_avg = df.city_avg_temp.rolling(4).mean()
plt.plot(df['year'],world_MV_avg,label='World')
plt.plot(df['year'],city_MV_avg,label='City')
plt.legend()
plt.xlabel("Years")
plt.ylabel("Temperature °C")
plt.title("New York City Average Temperature VS Rest of World")
plt.show()
```



Observations: -

1. By noticing the chart of global moving avg. temperature it is sure that the temperature has been rising which is an alarming situation for the world. But when we see the graph, there is sharp change in between the year 1820 to 1840, where temperature drop below 7.5 degree Celsius. But after that the temperature was in between the range of 8 to 8.5 from 1900 to 1950. But after that there is sharp change which leads to temperature get high upto 9.5 at the end of 2000. And there is no drop down after 1950.
2. As city temperature has been also raised with a spike, which is also alarming situation rather having situated beside the North Atlantic Ocean for New York where temperature rise upto 11 degree at the end of 2000.
3. For the New York, if we talk about variations in temperature, from the graph we can see there is a lot of changes. As we can see there is sharp drop down in between 1750 & 1800 to approx. 7.5, whereas the rest of the world temp. got a raise in that particular period.
4. But when start comparing New York temperature with world average temperature, we can see a direct relation between those two. As with increase in time, the temperature change with drastic curve is noticed in New York city graph after 1950. But when compare with rest of world, there are more up and down in New York temperature with respect to rest of world. The temperature of New York ends at 11 degrees whereas world temperature ends at 9 degrees Celsius with a difference of 2 degrees.

Conclusion-

To establish a good relation between those two, we need more different types of data. Otherwise with this data we can only conclude with there high rise in temperature worldwide & no city is left from its grip, which is very alarming situation.

If we want to see the correlation between New York city & rest of World temperature, we can use below method to find correlation coefficient.

Calculated by below method-

```
# prepare data
data1 = df['city_avg_temp']
data2 = df['world_avg_temp']

# calculate Pearson's correlation
corr, _ = pearsonr(data1, data2)

print('Pearsons correlation: %.3f % corr)
```

Consideration-

X-Axis for years

Y-Axis for temperature

Different labels for charts

Source:- [Click here](#) for correlation coefficient calculation method.