Explore Weather Trends

By

Saud Qadir Khan

Collection of Data

The most important part of any data analysis project is acquiring the data. For this project of weather trend analysis, the data was extracted using SQL query. The example of SQL query is given bellow. The query was run 5 times. In first run the data was extracted from global_data schema and then 4 times data was extracted the city_data schema

```
SELECT year, avg_temp
FROM global_data
WHERE avg_temp IS NOT NULL;
```

```
SELECT year, avg_temp
FROM city_data
WHERE city ='Multan' AND avg_temp IS NOT NULL;
```

```
SELECT year, avg_temp
FROM city_data
WHERE city ='Lahore' AND avg_temp IS NOT NULL;
```

```
SELECT year, avg_temp
FROM city_data
WHERE city ='Islamabad' AND avg_temp IS NOT NULL;
```

```
SELECT year, avg_temp
FROM city_data
WHERE city ='Karachi' AND avg_temp IS NOT NULL;
```

Data Preprocessing

The next step was to get your data ready for the analysis. I ran the query in such a way that missing values did not appear in the csv file. The data was loading in Jupyter Notebook and converted in to pandas DataFrame.

Moving average was calculated using window of 7 year. For this purpose method of pandas library name *.rolling()* was used.

Load the Data

```
df_global=pd.read_csv('global.csv')
df_global['global_avg_mean'] = df_global.avg_temp.rolling(window=7).mean()
df_global.drop(['avg_temp'], axis=1, inplace = True)
```

```
df_multan=pd.read_csv('multan.csv')
df_multan['multan_avg_mean'] = df_multan.avg_temp.rolling(window=7).mean()
df_multan.drop(['avg_temp'], axis=1, inplace = True)
```

```
df_lahore=pd.read_csv('lahore.csv')
df_lahore['lahore_avg_mean'] = df_lahore.avg_temp.rolling(window=7).mean()
df_lahore.drop(['avg_temp'], axis=1, inplace = True)
```

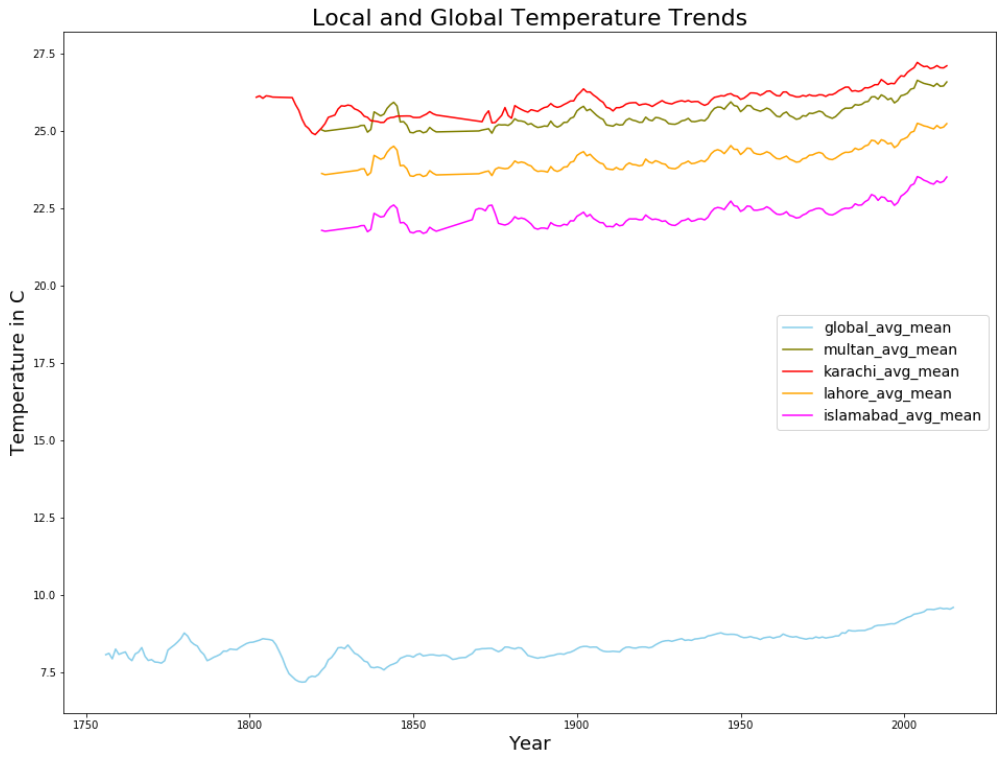
```
df_karachi=pd.read_csv('karachi.csv')
df_karachi['karachi_avg_mean'] = df_karachi.avg_temp.rolling(window=7).mean()
df_karachi.drop(['avg_temp'], axis=1, inplace = True)
```

```
df_islamabad=pd.read_csv('islamabad.csv')
df_islamabad['islamabad_avg_mean'] = df_islamabad.avg_temp.rolling(window=7).mean()
df_islamabad.drop(['avg_temp'], axis=1, inplace = True)
```

Plotting of Data

The data was plotted using matplotlib library of python. It is easy to use. The code used in plotting is given below

```
plt.figure(figsize=(16,12))
plt.plot( 'year', 'global_avg_mean', data=df_global, color='skyblue')
plt.plot( 'year', 'multan_avg_mean', data=df_multan, color='olive')
plt.plot( 'year', 'karachi_avg_mean', data=df_karachi, color='red')
plt.plot( 'year', 'lahore_avg_mean', data=df_lahore, color='orange')
plt.plot( 'year', 'islamabad_avg_mean', data=df_islamabad, color='magenta')
plt.xlabel('Year',fontsize=18)
plt.ylabel('Temperature in C',fontsize=18)
plt.title("Local and Global Temperature Trends",fontsize=22)
plt.legend(fontsize=14)
plt.savefig("temp.png")
```



Observations

1- My country is relatively hot as compared to global temperature
2- The Karachi is hotter than Multan despite of Multan temperature goes up to 48 degree centigrade in, Summer But overall yearly average temperature is less than that of Karachi's Temperature
3- Yearly average temperature is increased upto 2 degree Centigrade in past 2 centuries
4- The rate of increase in temperature was constating till 1950s and there is relative sharp increase in temperature in 21st century.