Exploring Weather Trends

-R.Shweta

SQL queries for retrieving data.

1. For global data :
   ```
   select * from global_data;
   ```

2. For city data :
   ```
   select * from city_data
   where city ='Singapore' & country ='Singapore';
   ```

3. Both global data and city data have same column names 'avg_temp'. Hence we rename the column names in both the data sets.

   ```
   alter table city_data
   rename column avg_temp to city_avg_temp;

   alter table global_data
   rename column avg_temp to global_avg_temp;
   ```

4. Joined both the tables and saved it in a new table called 'new_weather'.
   ```
   select city_data.year,city_data.city_avg_temp,global_data.global_avg_temp into
   new_weather
   from global_data
   join city_data on
   global_data.year=city_data.year
   where country='Singapore'
   ```

5. Checked for missing values
   ```
   select * from new_weather
   where country='Singapore' and avg_temp IS NULL;
   ```

6. After the data is fetched from new table, I observed that there were many missing values for avg_temp column, Singapore. Hence, I took the **mean (avg_temp)** of Singapore and substituted it in place of the missing values.

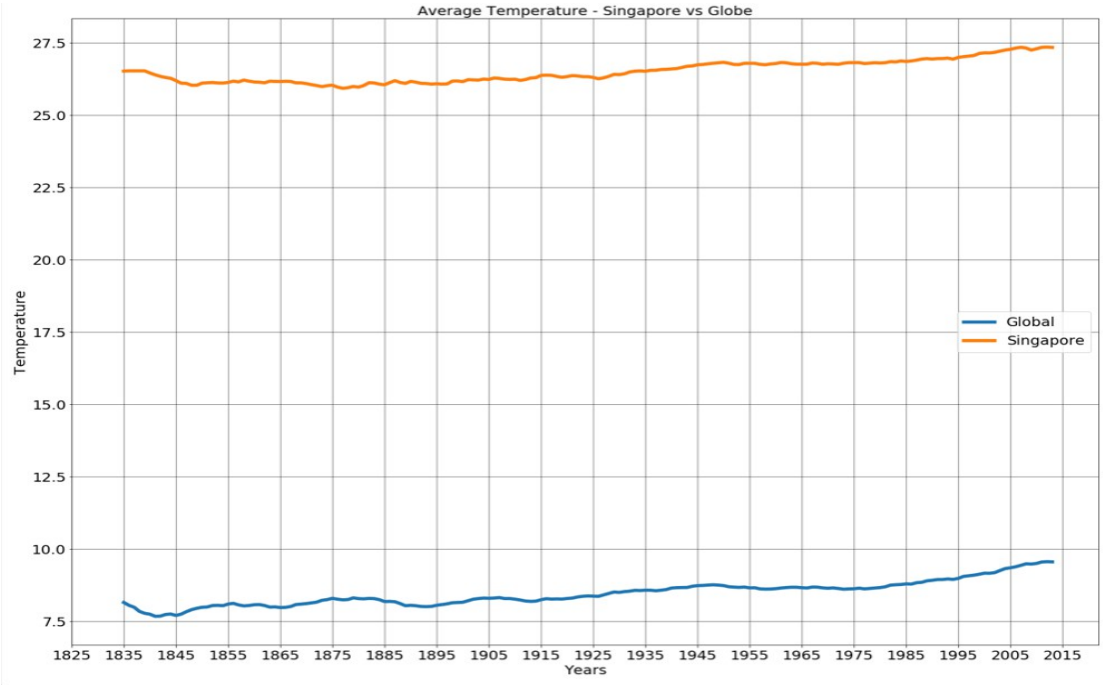
   ```
   select year,coalesce(city_avg_temp,26.52)as city_avg_temp,global_avg_temp from
   new_weather;
   ```

Excel for calculating Moving Average columns:

- ➔ Created two columns **MA_10_city** and **MA_10_global** for smoothing out trends in line chart
- ➔ Calculated the moving average for **10 years**. Command used = **Average(B2:B11)**
- ➔ Calculated correlation coefficients for city_avg_temp and global_avg_temp.Formula used = **correl(B2:B190;C2:C190)**
  correlation coefficient observed = 0.804897

Visualization using Python matplot.lib:

```
import pandas as pd  # for data related queries
import matplotlib.pyplot as plt  # for visualizations
temp_df=pd.read_csv("Results.csv")
# for generating a line plot between  MA_10_global and year
plt.plot(temp_df['year'],temp_df['MA_10_global'],label='Global',linewidth=5.0)
# for generating a line plot between MA_10_city and year
plt.plot(temp_df['year'],temp_df['MA_10_city'],label='Singapore',linewidth=5.0)
plt.legend(loc='center right',fontsize=20)
plt.xlabel("Years", fontsize=20)
plt.ylabel("Temperature", fontsize=20)
plt.title('Average Temperature - Singapore vs Globe', fontsize=20)
plt.rc('xtick', labelsize=20)
plt.rc('ytick', labelsize=20)
# for setting my own interval of  year intervals on x axis (Interval of 10 years)
plt.xticks(np.arange(min(temp_df['year']), max(temp_df['year'])+10,10))
# for displaying x and y grids
plt.grid(which='major', linestyle='-', linewidth='0.5', color='black')
# Setting the desired size of figure
fig_size = plt.rcParams["figure.figsize"]
fig_size[0] = 24
fig_size[1] = 18
plt.rcParams["figure.figsize"] = fig_size
```



Observations

- ➔ Global Moving Average temperature (10 years) between the years 1825 and 2015 varies between 7.6 and 9.5 degrees with the average temperature being 8.5 degrees
- ➔ Singapore on the other hand recorded higher moving average temperatures(10 years) ranging between 26 and 27.3 degrees with average temperature being 26.5 degrees
- ➔ A large difference in temperatures observed between Singapore and the globe.

- ➔ Singapore is hotter in comparison with the average global temperature
- ➔ A positive correlation observed between the global average temperature and Singapore's.Hence, indicating that as global average temperature increases, Singapore's average temperature is also increasing.

Conclusion

Both Singapore and global average temperatures are increasing every year making them hotter

source: google, github