Udacity Data Analyst Nano Degree Project 01: Explore Weather Trends

Swadeep Kumar Singh

Abstract: The aim of the project is to analyze and visualize the local and global temperature data. I have considered "Bangalore" as my local city and compared it with overall global temperature trends. This report focuses on providing five interesting insights/observations about the change in temperature trends. The trends about the weather data (local and global) has been visualized using Excel, where SQL query is used to extract the data from *Udacity* website.

Introduction: The change in weather trend has always been an interesting topic among scientist, politicians, environmentalists, and others. The goal of the project is to compare and analyze the similarities and dissimilarities between the local (where udacity course student lives) and global temperature data. The observations can be drawn by visualizing the temperature data. The report consists of following sections: I. Data extraction from database, II. Data manipulation methods and tools, III. Data visualization, IV. Data interpretation. For this report I have considered "Bangalore" (local city) as I lives here.

Data Extraction: The data is provided by the Udacity as a part of the nanodegree program. The data needs to be extracted through workspace which is connected to the database using SQL query. There are three tables in the database:

city_list - This contains a list of cities and countries in the database. Look through them in order to find the city nearest to you.

city_data - This contains the average temperatures for each city by year (ºC).

global_data - This contains the average global temperatures by year (ºC).

The SQL query used to extract the data: (A) Local city data *(Bangalore, India)* and (B) global data respectively:

A. Write a SQL query to extract the city level data. Export to CSV.

```
SELECT * FROM city_data WHERE
city = 'Bangalore';
```

B. Write a SQL query to extract the global data. Export to CSV.

```
SELECT * FROM global_data;
```

The query runs without error and pulls the intended data.

Data Manipulation: I have used MS-Excel for data manipulation. I have created columns with local_avg_temp and global_avg_temp for storing the moving average data. The calculation for moving average is same as explained in projects *Moving average* section. I have calculated moving average for 6 years separately for global and local data as explained in project instructions.

local_avg_temp	local_mv_avg	global_avg_temp	global_mv_avg
24.49		8.72	
25.18		7.98	
24.65		5.78	
24.81		8.39	
24.85		8.47	
24.49	24.745	8.36	7.95
25.44	24.90333333	8.85	7.971666667
25.22	24.91	9.02	8.145

Data Visualization: To visualize the data I have used the line charts provided within the *Excel*. I have analyzed data by plotting the 3 charts,

(1) Global moving average of temp. vs years
(2) Local moving average of temp. vs years
(3) Comparison between local and global moving average temp. vs years

Observations were drawn and explained in next section where separate color code is assigned to the local moving average data with global moving data. Please refer the graphs below.

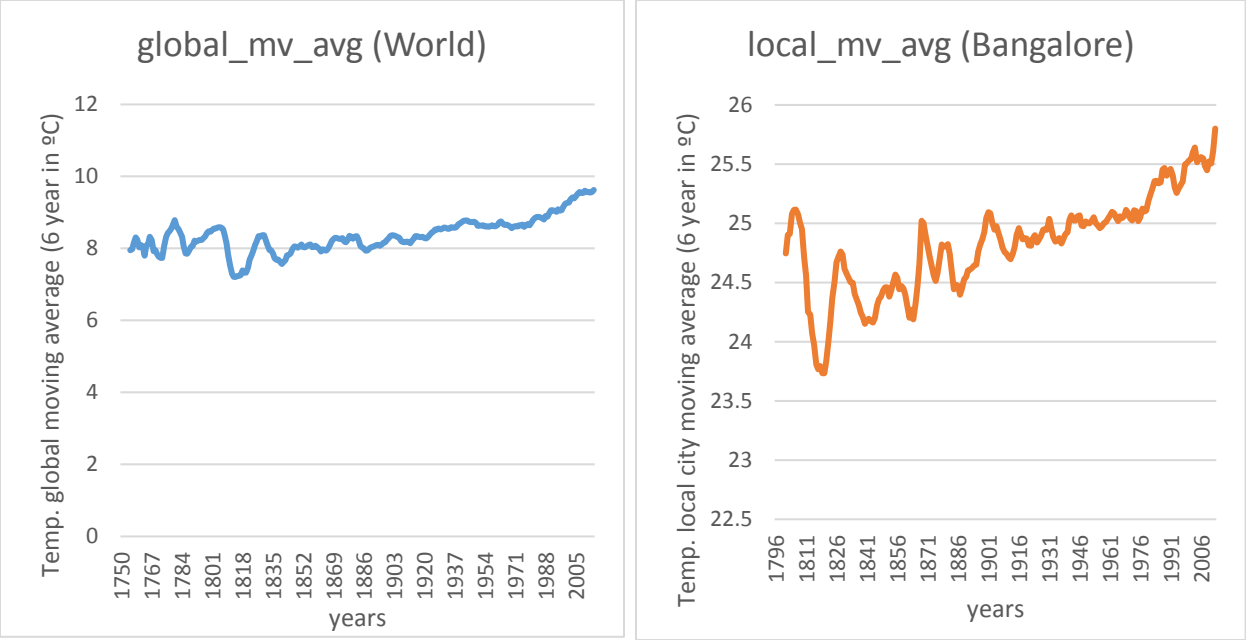

Figure 1. (On left), Global temperature data taken over 6 year moving average, and (on right) Bangalore city temperature data taken over 6 year moving average

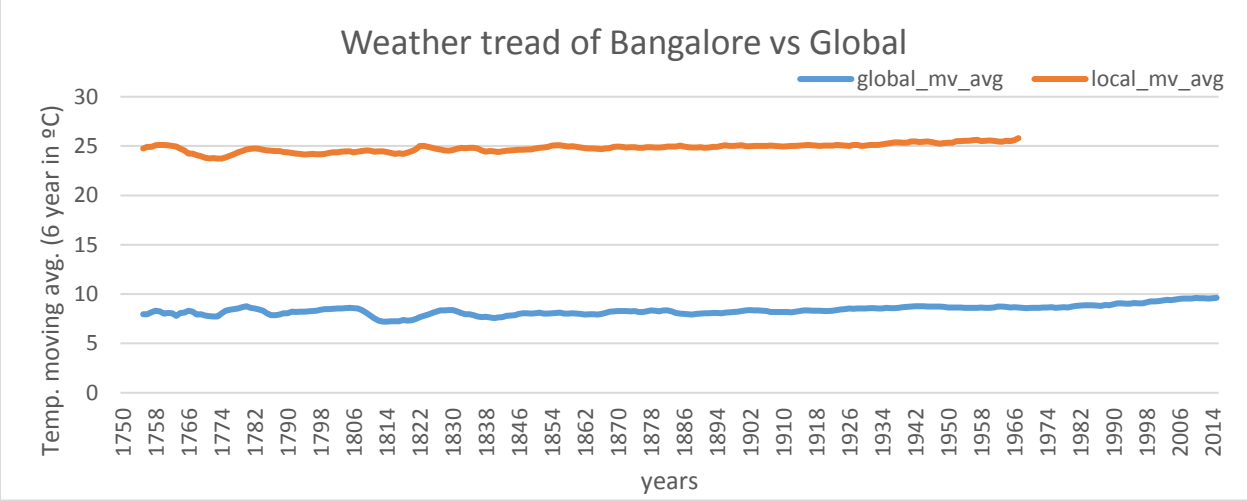

Figure 2: Comparison of Bangalore (local city) with Global temperature trend.

Data Interpretation: There following observations were drawn after analyzing the data from line charts:

(1) The Bangalore is hotter on average compared to global average temperature trend and the difference is consistent over time.
(2) Global and Bangalore average temperature has seen a consistent incremental change around the starting of 1900, before that there was fluctuations in average temperature.
(3) The overall trend shows increase in temperature in both global and local level. This means world is getting hotter and Bangalore as well. This trend is consistent over last hundred years.
(4) The upward trend after analyzing the last century data shows that it could be because of the industrialization or heavy use of fossil fuel.
(5) There is a significant increase in temperature can be observed in global data.