UDACITY- DATA ANALYTICS Nanodegree- Exploring Weather Trends

I first had the preview of the 3 Datasets using **SQL** queries, then downloaded the Datasets as *.csv* files to my local system. The query ran without any error.

```
SELECT * FROM city_data WHERE city IN (SELECT city FROM city_list WHERE city IN
('Bangalore')) AND country IN (SELECT country FROM city_list WHERE country='India')
```

```
SELECT * FROM global_data
```

Also tried my hands-on with **Pandas**, (i.e.) I tried all the above querying using Pandas also, it came out very well again.

	A	B
1	year	avg_temp
2	1750	8.72
3	1751	7.98
4	1752	5.78
5	1753	8.39
6	1754	8.47
7	1755	8.36
8	1756	8.85
9	1757	9.02
10	1758	6.74

global_data.csv

	A	B	C	D
1	year	city	country	avg_temp
2	1849	Abidjan	CÃ´te D'Ivoire	25.58
3	1850	Abidjan	CÃ´te D'Ivoire	25.52
4	1851	Abidjan	CÃ´te D'Ivoire	25.67
5	1852	Abidjan	CÃ´te D'Ivoire	
6	1853	Abidjan	CÃ´te D'Ivoire	
7	1854	Abidjan	CÃ´te D'Ivoire	
8	1855	Abidjan	CÃ´te D'Ivoire	
9	1856	Abidjan	CÃ´te D'Ivoire	26.28
10	1857	Abidjan	CÃ´te D'Ivoire	25.17

city_data.csv

Data Wrangling

I'm making use of Jupyter Notebook for this exercise. All the necessary datasets were imported into my notebook using Pandas *read_csv()* function.

```
city_list=pd.read_csv("D:/SRM books/Online Courses/UDACITY/Data Analyst Nanodegree/MODULE 1/Explore Weather Trends/city_list.csv")
```

```
city_data=pd.read_csv("D:/SRM books/Online Courses/UDACITY/Data Analyst Nanodegree/MODULE 1/Explore Weather Trends/city_data.csv")
global_data=pd.read_csv("D:/SRM books/Online Courses/UDACITY/Data Analyst Nanodegree/MODULE 1/Explore Weather Trends/global_data.csv")
```

I planned to work with the temperature of **Bangalore, India** which is the closest city from my place. There were some *Null/NaN* values in *city_data.csv*. So, I handled these missing values using *fillna()* function available under Pandas library.

```
bangalore2['avg_temp'].fillna(method='pad', inplace=True)
```

Also, I fixed some continuity issues with respect to year.

```
global1= global_data[ (global_data['year']>1795) & (global_data['year']<2014) ]
```

Moving Average Calculation

This is where I carried out a statistical calculation called as Moving Average. Moving Average is a calculation used to analyze data points by creating a series of averages of different subsets of the full data set.

To do this I made use of *.rolling()* function in Pandasc considering the window size to be 10.

```
global1["m_avg"]= global1['avg_temp'].rolling(window=10).mean()
```

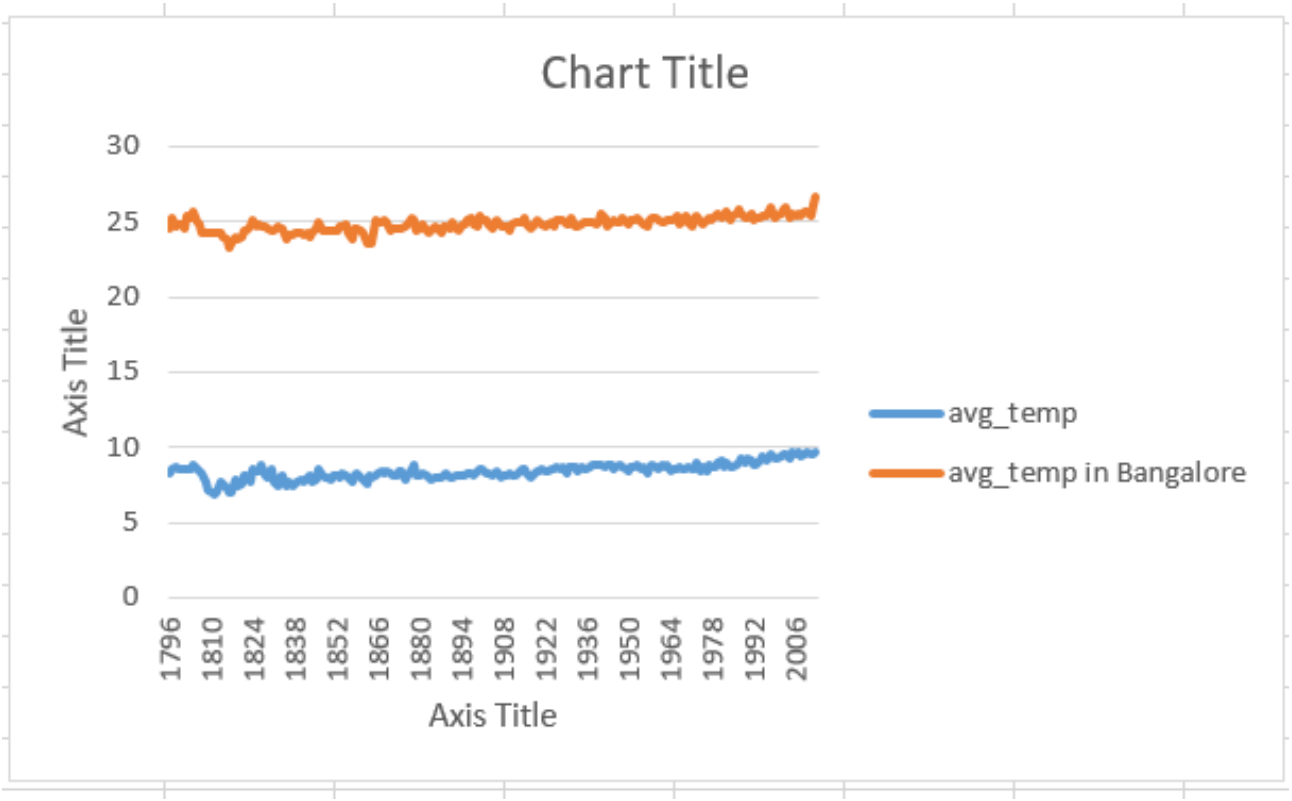
	A	B	C
1		year	avg_temp
2	0	1796	8.27
3	1	1797	8.51
4	2	1798	8.67
5	3	1799	8.51
6	4	1800	8.48
7	5	1801	8.59
8	6	1802	8.58
9	7	1803	8.5
10	8	1804	8.84

```
bangalore2["m_avg"]= bangalore2['avg_temp'].rolling(window=10).mean()
```

	A	B	C	D
1		year	city	avg_temp
2	0	1796	Bangalore	24.49
3	1	1797	Bangalore	25.18
4	2	1798	Bangalore	24.65
5	3	1799	Bangalore	24.81
6	4	1800	Bangalore	24.85
7	5	1801	Bangalore	24.49
8	6	1802	Bangalore	25.44
9	7	1803	Bangalore	25.22
10	8	1804	Bangalore	25.67
11	9	1805	Bangalore	25.01
12	10	1806	Bangalore	24.87

Line Chart

I have plotted the values of moving average temperatures from 1796 to 2013 of Bangalore, India and the Global Moving average temperature using MS Excel. This is a simple Line graph and it solves the purpose.



Observations

1) The value of moving averages show that Bangalore, India is hotter when compared to the moving average of the global temperatures.

2) In the time frame between 1808 and 1816 there was a dip the moving average temperatures in the global temperatures.

3) The average temperature of Bangalore, India is pretty much centred on 25°C from 1796 to 2013.

4) The temperatures for both the categories have increased marginally in the last 200+ years.

5) The global temperatures remained consistent until the year 1957, but after 1957 it increased, there seems to be lot of change in temperatures.

6) With every 50 years there is a marginal increase in temperature in both the categories.