

# Data Analyst Nanodegree Project

## Exploring Weather Trends

### Objective:

The goal of this project is to analyze the local and global temperature data and compare the temperature trends.

### Steps taken:

- Data Acquisition / Data Cleaning
- Data Exploration
- Data Presentation

### Data Acquisition:

The Database Schema:

There are three tables in the database:

- city\_list - This contains a list of cities and countries in the database.
- city\_data - This contains the average temperatures for each city by year (°C).
- global\_data - This contains the average global temperatures by year (°C).

SQL used to extract the data:

```
SELECT
    global_year AS year,
    MAX(global_avg) AS global_avg,
    MAX(bangalore_city_avg) AS bangalore_city_avg,
    MAX(sanjose_city_avg) AS sanjose_city_avg
FROM
(
    SELECT
        global.year as global_year,
        global.avg_temp as global_avg,
        CASE WHEN city.city = 'Bangalore' THEN city.avg_temp END AS bangalore_city_avg,
        CASE WHEN city.city = 'San Jose' THEN city.avg_temp END AS sanjose_city_avg
    FROM
        global_data AS global
    JOIN
    (
        SELECT *
        FROM city_data
        WHERE
            city IN ('Bangalore', 'San Jose')
        ) AS city
    ON global.year = city.year
) AS temp
GROUP BY year
ORDER BY year
```

## Data Analyst Nanodegree Project

I have extracted data of two cities - San Jose and Bangalore - the two cities where I have lived. Case statement and Max function are used to transform the data of the two cities and the global data into columns. The inner join between global and city table ensures comparison of data only for the years for which the global data is available. I exported this data to Excel to analyze and visualize.

### Data cleaning:

Handling of missing data – data for the city Bangalore had a few missing values and this was handled by filling them up with previous year averages.

Moving averages – to get a better visualization of the data points, 5 yr, 10 yr, 20 yr, 30 yr moving averages were calculated. Both 10yr and 20 yr MA seemed to give better visualization and I picked 10 yr MA since a decade seemed to be an intuitive/appropriate unit.

year	global_avg	bangalore_city_avg	sanjose_city_avg	global_10yr_MA	bangalore_10yr_MA	sanjose_10yr_MA	global_20yr_MA	bangalore_20yr_MA	sanjose_20yr_MA
1796	8.27	24.49							
1797	8.51	25.18							
1798	8.67	24.65							
1799	8.51	24.81							
1800	8.48	24.85							
1801	8.59	24.49							
1802	8.58	25.44							
1803	8.5	25.22							
1804	8.84	25.67							
1805	8.56	25.01		8.55	24.98				
1806	8.43	24.87		8.57	25.02				
1807	8.28	24.25		8.54	24.93				
1808	7.63	24.93		8.44	24.95				
1809	7.08	24.95		8.30	24.97				
1810	6.92	24.97		8.14	24.98				
1811	6.86	24.98		7.97	25.03				
1812	7.05	25.03		7.82	24.99				
1813	7.74	24.23		7.74	24.89				
1814	7.59	23.91		7.61	24.71				
1815	7.24	23.79		7.48	24.59		8.02	24.79	
1816	6.94	23.3		7.33	24.43		7.95	24.73	
1817	6.98	23.6		7.20	24.37		7.87	24.65	
1818	7.83	23.94		7.22	24.27		7.83	24.61	
1819	7.37	23.86		7.25	24.16		7.77	24.56	
1820	7.62	23.91		7.32	24.05		7.73	24.52	
1821	8.09	24.4		7.45	24.00		7.71	24.51	
1822	8.19	24.33		7.56	23.93		7.69	24.46	
1823	7.72	24.62		7.56	23.97		7.65	24.43	
1824	8.55	25.1		7.65	24.09		7.63	24.40	
1825	8.39	24.69		7.77	24.18		7.63	24.38	
1826	8.36	24.88		7.91	24.33		7.62	24.38	
1827	8.81	24.67		8.09	24.44		7.65	24.40	
1828	8.17	24.61		8.13	24.51		7.68	24.39	
1829	7.94	24.46		8.18	24.57		7.72	24.36	

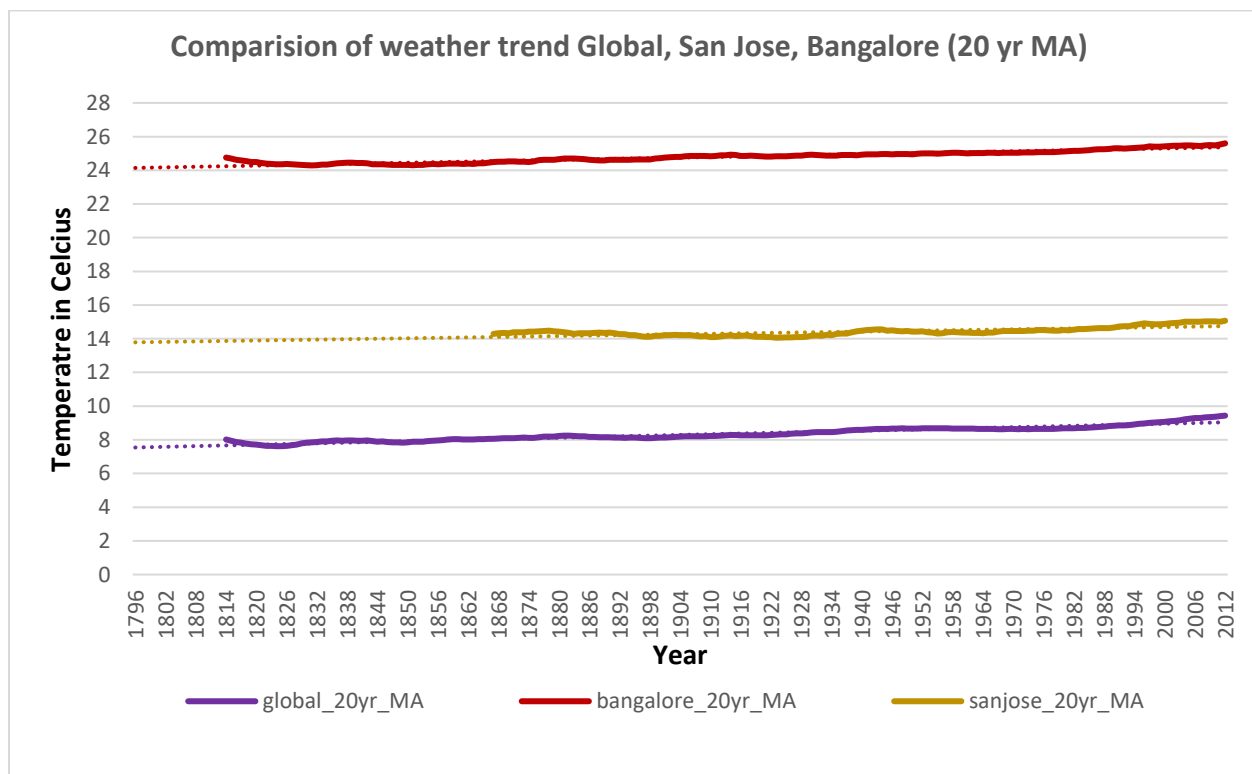
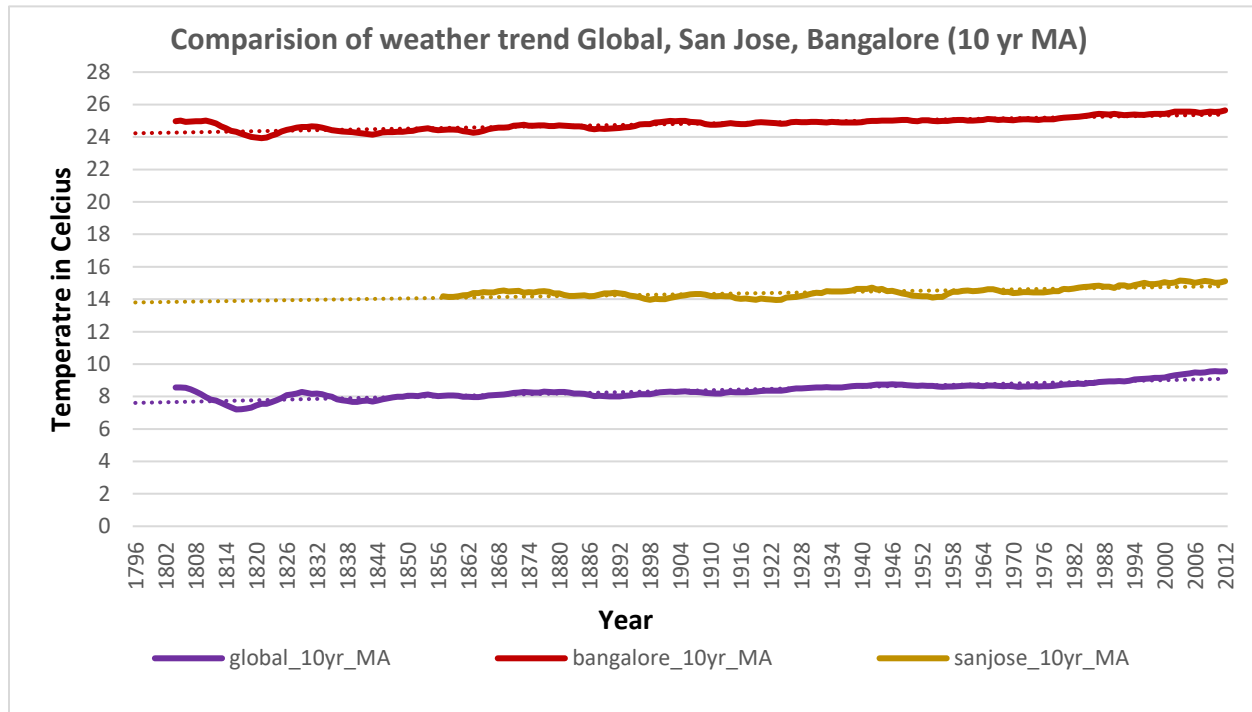
### Data Exploration:

I plotted a line graph with the moving average temperature values on Y axis and the year on X axis.

I also added a trend line to make the trend easily visible.

The correlation coefficients calculated for Global, Bangalore and Global, San Jose is also presented for both 10 yr and 20 yr MA values.

## Data Analyst Nanodegree Project



## Data Analyst Nanodegree Project

### CORRELATION COEFFICIENT

Global, Bangalore	0.924977	10 Year MA
Global, San Jose	0.816018	
Global, Bangalore	0.956953	20 Year MA
Global, San Jose	0.861431	

### Observations:

- All three temperatures – global, San Jose and Bangalore seems to be on an upward trend.
- San Jose seems to be hotter than the overall global temperature and Bangalore even hotter than San Jose which is understandable as Bangalore is located near Tropical region.
- The upward trend starts at the turn of 19<sup>th</sup> century which coincides with the industrial revolution. A steep increase can be seen in all three lines at the beginning of 21<sup>st</sup> century which shows that the climate is changing rapidly now.
- There is a dip in both global temperature and Bangalore temperature around 1816 which needs further research to analyze the reason.
- There is a high correlation between global and Bangalore data with a coefficient value of 0.92, also between global and San Jose data with a coefficient value of 0.8. However, we cannot conclude that increase in temperature of one caused the increase in temperature of the other from this data alone.