Submitted By: Deepshikha Patni

UDACITY Data Analysis Nanodegree Program - Project 1 (Exploring Weather Trends)

Introduction:

In this project, I've analyzed local and global temperature data and compared the temperature trends where I live to overall global temperature trends.

Instructions:

Goal is to create a visualization and prepare a write up describing the similarities and differences between global temperature trends and temperature trends in the closest big city to where I live.

- Create a line chart that compares my city's temperatures with the global temperatures
- Make observations about the similarities and differences between the world averages and my city's averages, as well as overall trends.

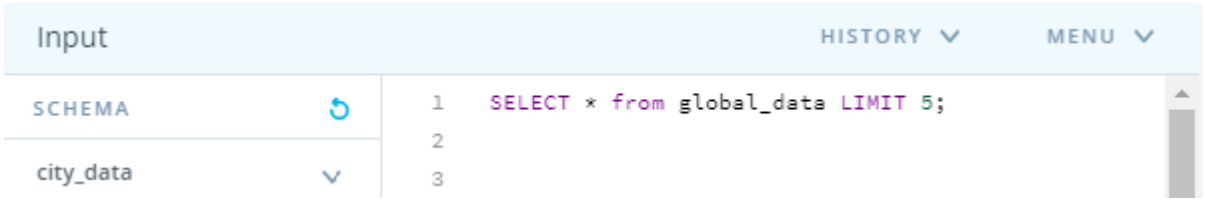
Tools Used:

SQL (For extracting the data), Microsoft Excel (for calculating MA's), Python (Jupyter Notebook for plotting)

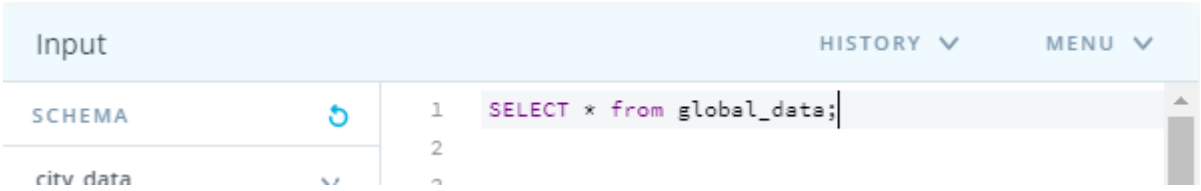
Step 1: Extracting the data:

I've used SQL queries to export the data into csv.

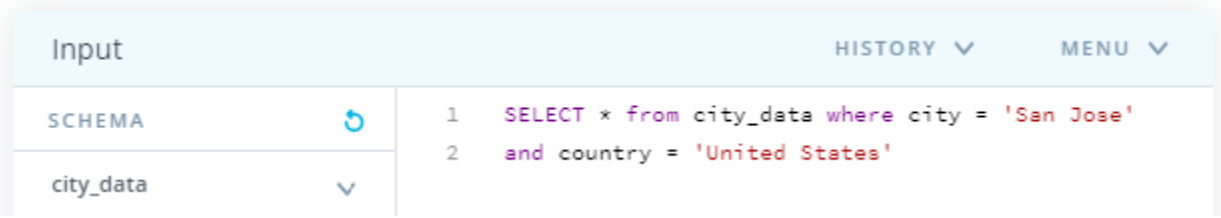
i) To export the global data, I used the below SQL query,
First I queried few records to get an understanding of the data,



Then I took a dump of the entire table,



ii) To export the data for the closest big city to where I live i.e. San Jose, United States, I used the below query,

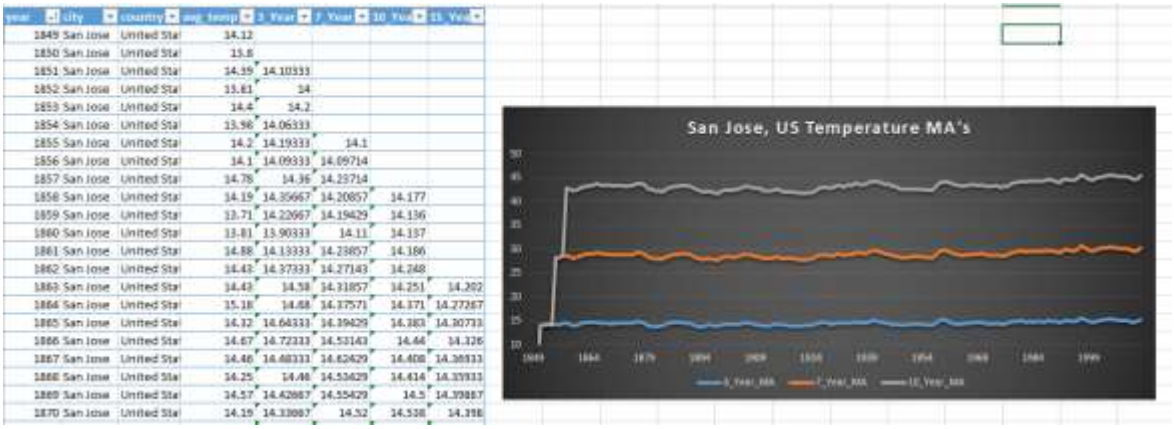


Step 2: Calculating and Determining Moving Average for Plotting:

- Before calculating the MA's, I sorted the data in increasing order of Year.
- I calculated 3 year MA, 7 year MA, 10 year MA and 15 Year MA for both city and global temperature data.

Calculating Moving Average for city_data

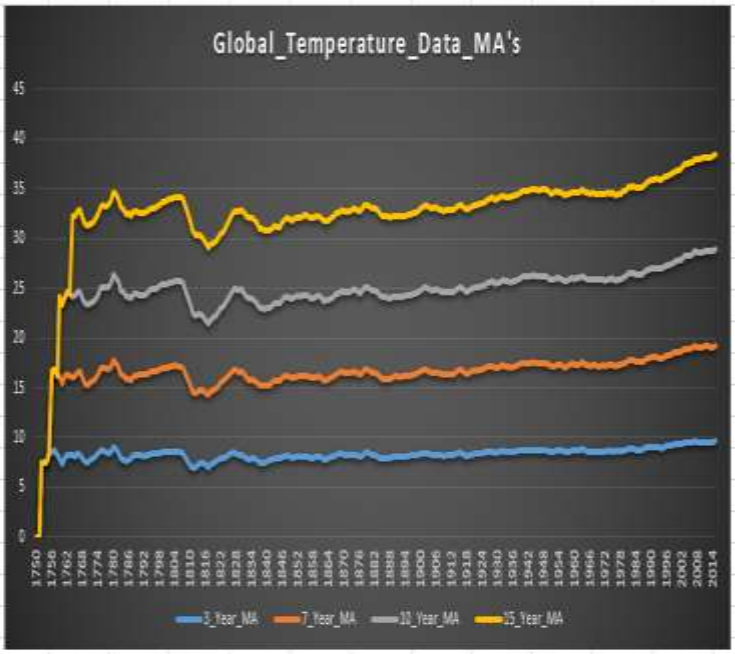
For 3 Year MA, used *=AVERAGE(D2:D4)*
For 7 Year MA, used *=AVERAGE(D2:D8)*
For 10 Year MA, used *=AVERAGE(D2:D11)*



Calculating Moving Average for global_data:

For 3 Year MA, used *=AVERAGE(B2:B4)*
For 7 Year MA, used *=AVERAGE(B2:B8)*
For 10 Year MA, used *=AVERAGE(B2:B11)*
For 15 Year MA, used *=AVERAGE(B2:B16)*

year	avg_temp	3_Year	7_Year	10_Year	15_Year
1750	8.72				
1751	7.98				
1752	5.78	7.493333			
1753	8.39	7.383333			
1754	8.47	7.546667			
1755	8.36	8.406667			
1756	8.85	8.56	8.078571		
1757	9.02	8.743333	8.121429		
1758	6.74	8.203333	7.944286		
1759	7.99	7.916667	8.26	8.03	
1760	7.19	7.306667	8.088571	7.877	
1761	8.77	7.983333	8.131429	7.956	
1762	8.61	8.19	8.167143	8.239	
1763	7.5	8.293333	7.974286	8.15	
1764	8.4	8.17	7.885714	8.143	8.051333
1765	8.25	8.05	8.101429	8.132	8.02
1766	8.41	8.353333	8.161429	8.088	8.048667
1767	8.22	8.293333	8.308571	8.008	8.211333
1768	6.78	7.803333	8.024286	8.012	8.104
1769	7.69	7.563333	7.892857	7.982	8.052
1770	7.69	7.386667	7.92	8.032	8.007333
1771	7.85	7.743333	7.841429	7.94	7.940667



Determining the ideal period for calculating Moving Average: Based on my plots, I decided to go with 10 year MA so that I can smooth the curve.

Step 3: Plotting the line charts:

```
### Importing necessary libraries:
```

```
In [2]: import pandas as pd
        import matplotlib.pyplot as plt
        import numpy as np
```

```
### Importing the data:
```

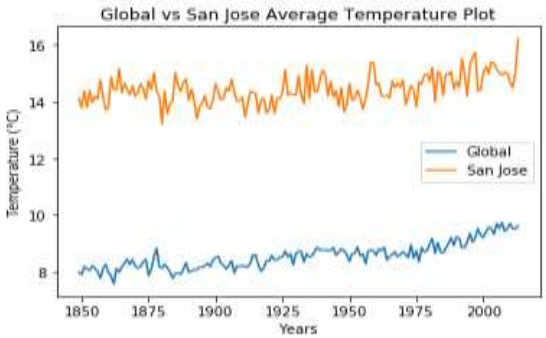
```
In [9]: df_global=pd.read_csv('GLOBAL_DATA.csv')
        df_city=pd.read_csv('san_jose_data.csv')
        df_global.head(10)
        df_city.head(10)
```

```
### Merging data frames (df_global and df_city) using year column
```

```
: df_merged=pd.merge(df_global, df_city, on="year")
  df_merged.head(15)
```

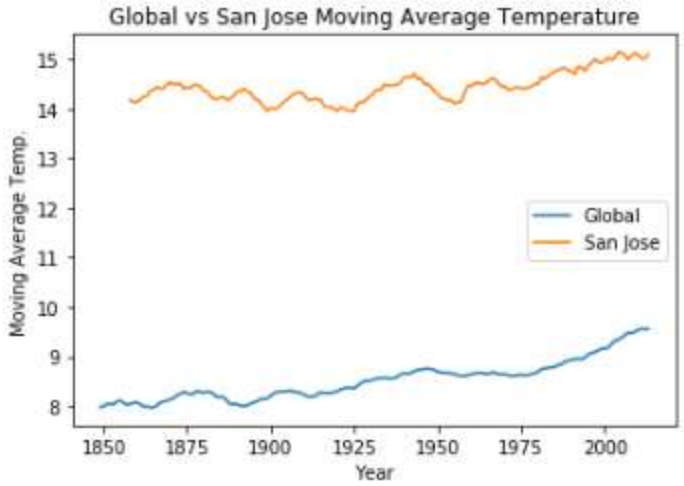
```
### Plotting df_global and df_city with year on X Axis and avg_temp_x, avg_temp_y on Y axis
```

```
: plt.plot(df_merged['year'],df_merged['avg_temp_x'],label='Global')
  plt.plot(df_merged['year'],df_merged['avg_temp_y'],label='San Jose')
  plt.legend()
  plt.xlabel("Years")
  plt.ylabel("Temperature (°C)")
  plt.title("Global vs San Jose Average Temperature Plot")
  plt.show()
```



There were lot of fluctuations in this graph, so to make it smooth I decided to use 10 year moving averages to plot this graph.

```
plt.plot(df_merged['year'],df_merged['10_Year_MA_x'], label = 'Global')
plt.plot(df_merged['year'],df_merged['10_Year_MA_y'], label = 'San Jose')
plt.legend()
plt.title("Global vs San Jose Moving Average Temperature")
plt.xlabel('Year')
plt.ylabel('Moving Average Temp.')
plt.show()
```



Step 4: Observations:

- Temperature of San Jose city is on an average hotter than compared to global temperature.
- This trend is consistent since 1970 onwards.
- While there was a dip in global temperature between 1880 -1890, but San Jose observed a spike in temperature during that time
- Temperature is rising steadily for both global and San Jose.
- While the difference in lowest and highest temperature during 1850-2010 in San Jose is between ~1.7 degree centigrade but globally this difference is close to 2 degree.