Exploring Weather Trends

This is my first Udacity project. So, for this project I selected Victoria City which is nearest to me as I live in Calgary (Alberta).

Project starts with SQL queries to extract the data as follows:

Extract Data:

I used these queries to find the dataset and to explore given data.

Select * from global_data

Select * from city_data

Select * from city_data where city = 'Toronto';

I have a python programming background. So, I used some standard libraries to complete this project.

Data preparation:

First step is importing the standard libraries and csv files. Then create data frames with help of pandas.

Data Preparation

In [1]:
```
###import all the standard libraries needed for project
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
```

In [2]:
```
Global_data= pd.read_csv('Global_temp.csv')  ## import csv file and create data frame
Victoria_data=pd.read_csv('Victoria.csv')
```

Data exploration:

Data Exploration

In [3]:
```
Global_data.head(5) ##it will display the first rows
```

Out[3]:

	year	avg_temp
0	1750	8.72
1	1751	7.98
2	1752	5.78
3	1753	8.39
4	1754	8.47

In [4]:
```
Victoria_data.head()
```

Out[4]:

	year	city	country	avg_temp
0	1828	Victoria	Canada	6.83
1	1829	Victoria	Canada	6.58
2	1830	Victoria	Canada	NaN
3	1831	Victoria	Canada	NaN
4	1832	Victoria	Canada	3.25

In [5]:
```
Global_data.info()                                  ## it will display the concise summary of data frame
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 266 entries, 0 to 265
Data columns (total 2 columns):
year        266 non-null int64
avg_temp    266 non-null float64
dtypes: float64(1), int64(1)
memory usage: 4.2 KB
```

To explore both data, I used above methods to know about rows and columns. Info () method gave us concise summary of our both data frames. Global dataset starts from year 1750 to 2015 but local city data starts from year 1828 to 2013.

```
In [7]: Global_data['avg_temp'].describe()       ## it will display the basic statstical detail of selected column from data frame su

Out[7]: count    266.000000
        mean       8.369474
        std        0.584747
        min        5.780000
        25%        8.082500
        50%        8.375000
        75%        8.707500
        max        9.830000
        Name: avg_temp, dtype: float64

In [8]: Victoria_data['avg_temp'].describe()

Out[8]: count    183.000000
        mean       7.259126
        std        0.771946
        min        3.250000
        25%        6.760000
        50%        7.230000
        75%        7.735000
        max        9.850000
        Name: avg_temp, dtype: float64
```

Minimum temperature was 3.25 in Victoria city(Canada) and Maximum was 9.85 over the years.

Describe () method gives us the statistical detail of data frames. Here, I used specific column avg_temp to get the mean, standard deviation etc. The mean temperature of global data is higher than local city (Victoria). The maximum temperature recorded over the years is similar (9.8).

```
In [9]: Global_data.isnull().sum()

Out[9]: year        0
        avg_temp    0
        dtype: int64

In [10]: Victoria_data.isnull().sum().sum()

Out[10]: 3
```

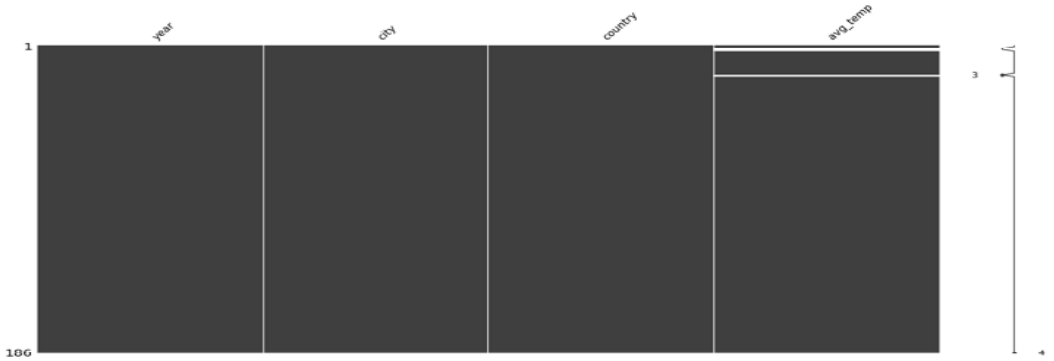
It is better to know about null values in dataset before analysing. There is no missing value in Global data but there are 3 missing values in Victoria data in year 1830,1831, and 1846.We can also plot them as below.

Visualizations:

Plots and Visualizations

```
In [11]: import missingno as msno
         # Visualize missing values as a matrix
         msno.matrix(Global_data)

Out[11]: <matplotlib.axes._subplots.AxesSubplot at 0x2714e5270f0>
```

```
In [12]: msno.matrix(Victoria_data)
Out[12]: <matplotlib.axes._subplots.AxesSubplot at 0x2714e593358>
```



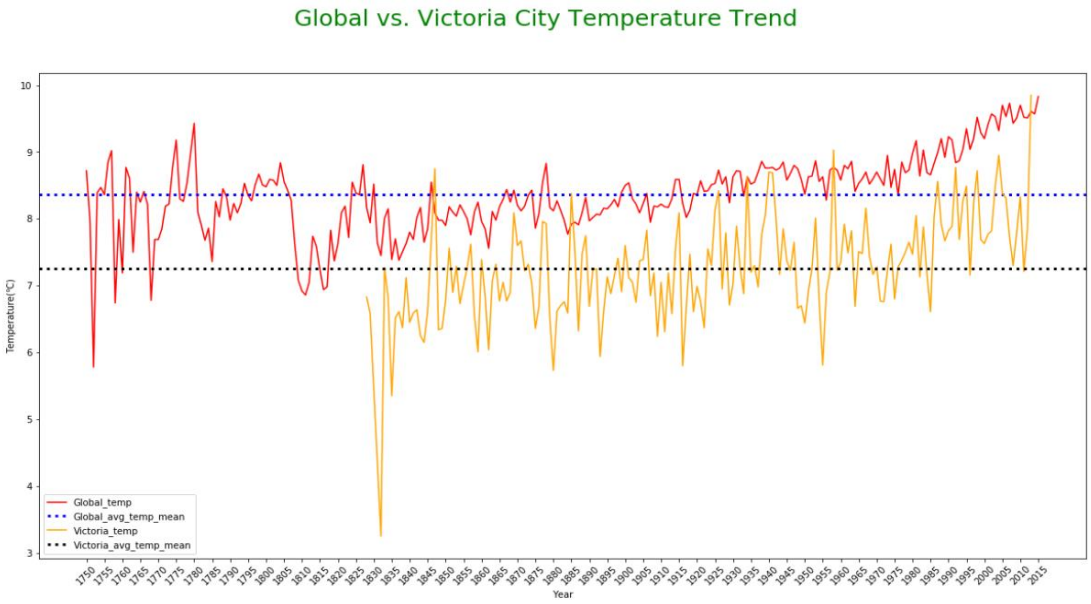

```
In [34]: Victoria_data2=Victoria_data.fillna(Victoria_data.mean())  # we can fill NaN values with Mean value(mean temp)
         Victoria_data2
```

Out[34]:

	year	city	country	avg_temp
0	1828	Victoria	Canada	6.830000
1	1829	Victoria	Canada	6.580000
2	1830	Victoria	Canada	7.259126
3	1831	Victoria	Canada	7.259126
4	1832	Victoria	Canada	3.250000
5	1833	Victoria	Canada	7.270000
6	1834	Victoria	Canada	6.810000
7	1835	Victoria	Canada	5.350000
8	1836	Victoria	Canada	6.520000
9	1837	Victoria	Canada	6.610000
10	1838	Victoria	Canada	6.370000

There are many methods by which you can clean the data (missing values) but I used to fill the values with mean.

Temperature comparison:



Both temperatures have similarities in rising year by year. There are rapidly rising in temperature in 1850,1880 and 2010. Both are rising by 2-3 degree every 5 years. The average mean of global temperature is higher than Victoria City's average mean which means Globally temperature is more affected by changing climate as compared to local city.

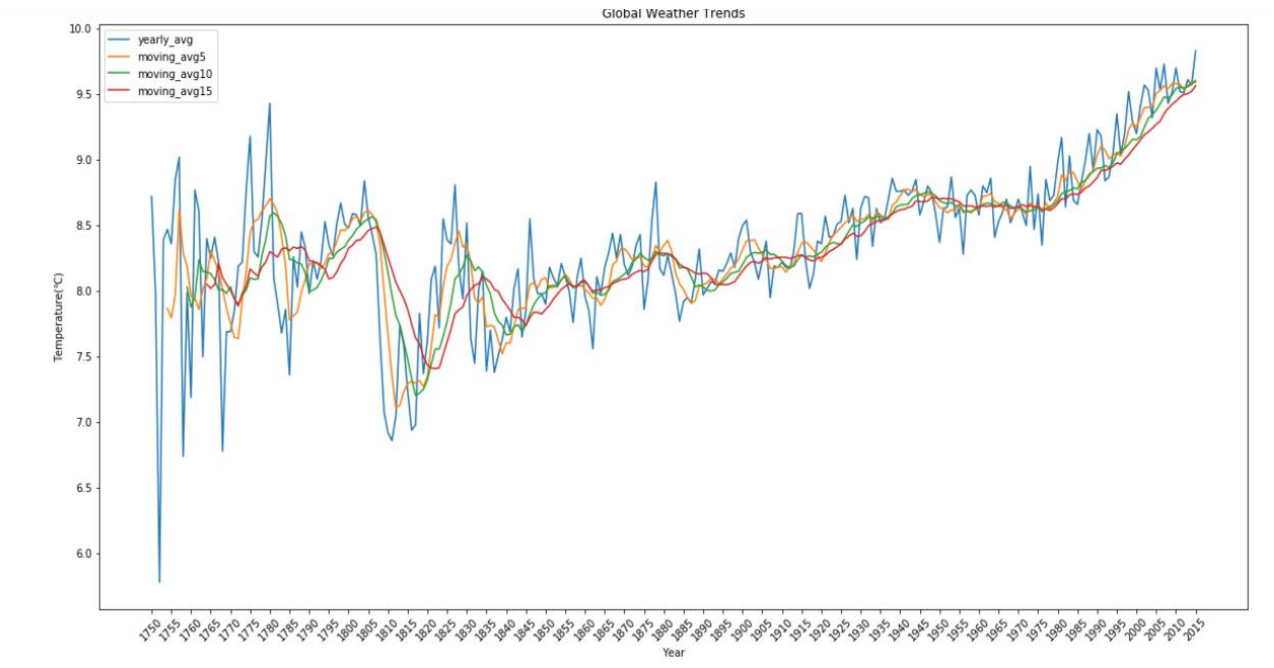
Moving averages and weather trends:

The idea behind the moving averages is to smoother out the values and reduce the noise. It makes easier to watch out the trends more clearly with moving averages. I calculated moving averages for 5, 10 and 15 years because it makes visualization clearer to analyze.

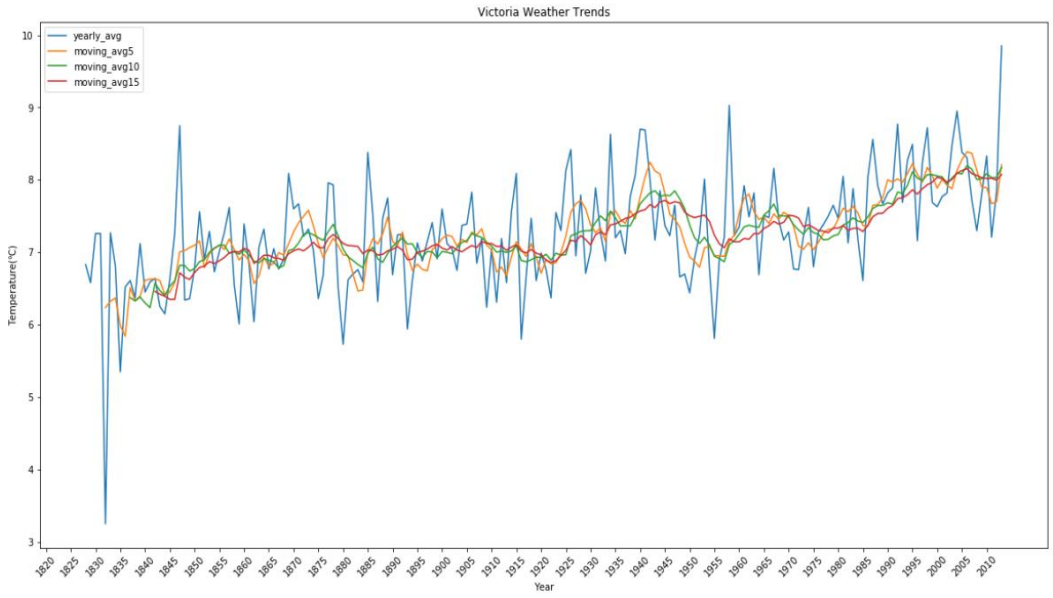
Moving Averages

In [39]:

```
n = 5

Global_data['moving_avg5'] = Global_data['avg_temp'].rolling(window=n).mean()
Victoria_data2['moving_avg5'] = Victoria_data2['avg_temp'].rolling(window=n).mean()

n = 10

Global_data['moving_avg10'] = Global_data['avg_temp'].rolling(window=n).mean()
Victoria_data2['moving_avg10'] = Victoria_data2['avg_temp'].rolling(window=n).mean()

n = 15

Global_data['moving_avg15'] = Global_data['avg_temp'].rolling(window=n).mean()
Victoria_data2['moving_avg15'] = Victoria_data2['avg_temp'].rolling(window=n).mean()

plt.figure(figsize=(20,10))
plt.plot(Global_data.year, Global_data.avg_temp, label='yearly_avg')
plt.plot(Global_data.year, Global_data.moving_avg5, label='moving_avg5')
plt.plot(Global_data.year, Global_data.moving_avg10, label='moving_avg10')
plt.plot(Global_data.year, Global_data.moving_avg15, label='moving_avg15')
plt.title('Global Weather Trends')
plt.xlabel('Year')
plt.xticks(np.arange(1750, 2020, step=5), rotation=45)
plt.ylabel('Temperature(°C)')
plt.legend()
plt.show()
```



```
In [37]:
plt.figure(figsize=(20,10))
plt.plot(Victoria_data2.year, Victoria_data2.avg_temp, label='yearly_avg')
plt.plot(Victoria_data2.year, Victoria_data2.moving_avg5, label='moving_avg5')
plt.plot(Victoria_data2.year, Victoria_data2.moving_avg10, label='moving_avg10')
plt.plot(Victoria_data2.year, Victoria_data2.moving_avg15, label='moving_avg15')
plt.legend()
plt.title('Victoria Weather Trends')
plt.xlabel('Year')
plt.xticks(np.arange(1820, 2015, step=5), rotation=45)
plt.ylabel('Temperature(℃)')
plt.show()
```



We can see the moving average 15 (red line) makes more smoother line than others in both cases. Both follows upward trends smoothly.

1.Is your city hotter or cooler on average compared to the global average? Has the difference been consistent over time?

Victoria City is capital of British Columbia and it sits on the craggy southern end of Vancouver Island. It has most moderate weather in all of Canada**.** Rainfall is ample all over the year but there is more rainfall in winters. With each passing year both global and local city (Victoria) temperatures are increasing due to climate issues. By analysing both data's it is clear in every 30-40 years temperature is doubled.

2."How do the changes in your city's temperatures over time compare to the changes in the global average?"

Despite global temperature's rise, Victoria City's temperature is also rising every year 2-3 degree. Plot shows there are two sharp peaks in both global and local temperature. One peak is around 1850's and the other is around 1950-60. This is not a end but more yet to come..

3.What does the overall trend look like? Is the world getting hotter or cooler? Has the trend been consistent over the last few hundred years?

From visualization, the overall temperature trend is gone upward, which means the world is getting hotter day by day. Despite ups and downs from year to year, global average surface temperature and our local (Victoria) temperature are rising over the last hundred years.

References

https://matplotlib.org/3.1.0/gallery/subplots_axes_and_figures/subplots_demo.html

https://towardsdatascience.com/implementing-moving-averages-in-python-1ad28e636f9d

https://stackoverflow.com/questions/13295735/how-to-replace-nan-values-by-zeroes-in-a-column-of-a-pandas-dataframe

https://en.climate-data.org/north-america/canada/british-columbia/victoria-631/