Exploring Weather Trends

By: Md Rahamat Ullah

Extracting datasets from the database

Following **SQL query** were used to get the data. Datasets were downloaded manually by pressing download command.

SELECT * FROM city_data;

SELECT * FROM global_data;

SELECT * FROM city_list;

Data analysis process was done with Python.

In [1]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plot
import seaborn as sb
%matplotlib inline
```

In [2]:

```
#Data Importing from local drive
df_city_data = pd.read_csv("city_data.csv")
df_global_data = pd.read_csv("global_data.csv")
df_city_list = pd.read_csv("city_list.csv")
```

In [3]:

```
#Assessing Data visually for Messy data, Dirty Data
df_city_data.head(2)
```

Out[3]:

	year	city	country	avg_temp
0	1849	Abidjan	Côte D'Ivoire	25.58
1	1850	Abidjan	Côte D'Ivoire	25.52

In [4]:

```
df_global_data.head(2)
```

Out[4]:

	year	avg_temp
0	1750	8.72
1	1751	7.98

In [5]:

```
df_city_list.head(2)
```

Out[5]:

	city	country
0	Abidjan	Côte D'Ivoire
1	Abu Dhabi	United Arab Emirates

In [6]:

```
#Deteccting issue Programmatically
df_city_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 71311 entries, 0 to 71310
Data columns (total 4 columns):
year        71311 non-null int64
city        71311 non-null object
country     71311 non-null object
avg_temp    68764 non-null float64
dtypes: float64(1), int64(1), object(2)
memory usage: 2.2+ MB
```

In [7]:

```
df_global_data.info()
```

```
<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
```

In [8]:

```
df_city_list.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 345 entries, 0 to 344
Data columns (total 2 columns):
city       345 non-null object
country    345 non-null object
dtypes: object(2)
memory usage: 5.5+ KB
```

Data Cleaning:

Issues:

- Missing 'avg_temp' Data in 'df_city_data'

In [9]:
```
df_city_data_null = df_city_data[df_city_data['avg_temp'].isnull()] #Create a dataframe 'df_city_data_null' with null 'avg_temp'
df_city_data_clean = df_city_data.dropna(subset=['avg_temp'])       #Create a dataframe 'df_city_data_clean' with no null 'avg_temp'
```

In [10]:
```
df_city_data_clean[df_city_data_clean['avg_temp'].isnull()]         #Testing 'df_city_data_clean' contains no null value
```

Out[10]:

year	city	country	avg_temp

In [11]:
```
#Calculating mean of 'avg_temp' corresponding to 'city'
nan_data = df_city_data.groupby('city').avg_temp.mean()

#Testing
nan_data.sample(2)
```

Out[11]:
```
city
Alexandria    15.704376
Tijuana       16.126364
Name: avg_temp, dtype: float64
```

In [12]:
```
#Making an empty list 'a' and append value of city data (mean temp) according to the index position of the city in 'df_city_data_null'
a = []
for idx, city_name in enumerate(df_city_data_null.city):
   a.append(nan_data[city_name])
```

In [13]:
```
avg_temp = pd.DataFrame(a)
df_city_data_null  = df_city_data_null .drop('avg_temp', axis =1)
df_city_data_null['avg_temp']  = a
df_city_data_null.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2547 entries, 3 to 71145
Data columns (total 4 columns):
year        2547 non-null int64
city        2547 non-null object
country     2547 non-null object
avg_temp    2547 non-null float64
dtypes: float64(1), int64(1), object(2)
memory usage: 99.5+ KB
```

In [14]:

```
df_city_data = pd.concat([df_city_data_null, df_city_data_clean])
df_city_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 71311 entries, 3 to 71310
Data columns (total 4 columns):
year        71311 non-null int64
city        71311 non-null object
country     71311 non-null object
avg_temp    71311 non-null float64
dtypes: float64(1), int64(1), object(2)
memory usage: 2.7+ MB
```

df_city_data no longer contains missing values.

In [15]:

```
#Taking a subset dataframe called 'My_City' from 'df_city_data' that for cit
y:'New York'
My_city = df_city_data[df_city_data.city == 'New York']
My_city = My_city[['year','avg_temp']]
df_global_data.count()
```

Out[15]:

```
year        266
avg_temp    266
dtype: int64
```

Moving averages

Moving averages calculation for average temperature: This moving average was calculated by using **rolling()** function that was adding average temperatures over a 8 years period and **mean()** function was dividing the sum by the total number of periods.

In [16]:

```
My_city['moving avg_temp_new_york'] = My_city.avg_temp.rolling(8).mean()
My_city = My_city.drop('avg_temp', axis = 1)
My_city.head()
```

Out[16]:

	year	moving avg_temp_new_york
46344	1746	NaN
46345	1747	NaN
46346	1748	NaN
46347	1749	NaN
46378	1780	NaN

In [17]:

```
df_global_data['moving avg_temp_global'] = df_global_data.avg_temp.rolling(8).
mean()
df_global_data = df_global_data.drop('avg_temp', axis = 1)
df_global_data.head(12)
```

Out[17]:

	year	moving avg_temp_global
0	1750	NaN
1	1751	NaN
2	1752	NaN
3	1753	NaN
4	1754	NaN
5	1755	NaN
6	1756	NaN
7	1757	8.19625
8	1758	7.94875
9	1759	7.95000
10	1760	8.12625
11	1761	8.17375

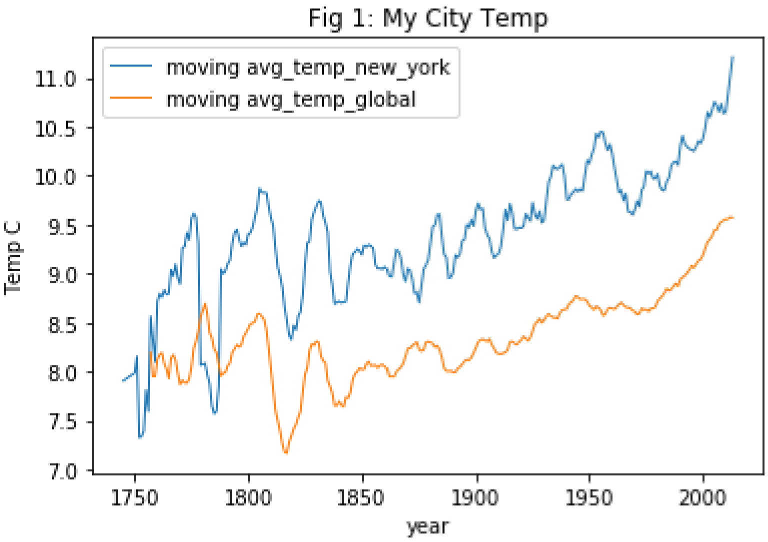
In [18]:

```
#Merging all in a dataframe for easy visual and easy plotting.
Line_chart = pd.merge(My_city,df_global_data, on = ['year'], how = 'left')
Line_chart['moving avg_temp_diff'] = Line_chart['moving avg_temp_new_york'] -
Line_chart['moving avg_temp_global']
Line_chart.head(20)
```

Out[18]:

	year	moving avg_temp_new_york	moving avg_temp_global	moving avg_temp_diff
0	1746	NaN	NaN	NaN
1	1747	NaN	NaN	NaN
2	1748	NaN	NaN	NaN
3	1749	NaN	NaN	NaN
4	1780	NaN	8.71000	NaN
5	1743	NaN	NaN	NaN
6	1744	NaN	NaN	NaN
7	1745	7.906391	NaN	NaN
8	1750	7.985113	NaN	NaN
9	1751	8.153835	NaN	NaN
10	1752	7.325056	NaN	NaN
11	1753	7.335028	NaN	NaN
12	1754	7.390000	NaN	NaN
13	1755	7.808750	NaN	NaN
14	1756	7.593750	NaN	NaN
15	1757	8.563750	8.19625	0.36750
16	1758	8.323750	7.94875	0.37500
17	1759	8.101250	7.95000	0.15125
18	1760	8.716250	8.12625	0.59000
19	1761	8.798750	8.17375	0.62500

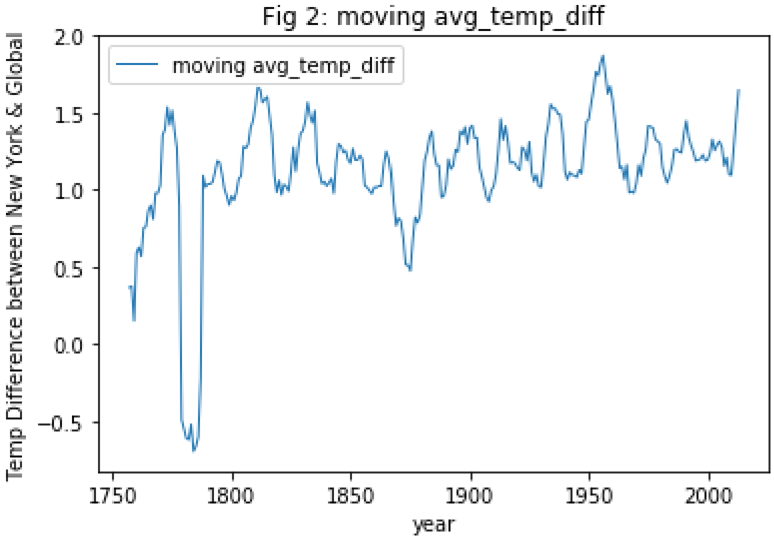
```
In [19]: #Drawing a line chart

         Line_chart.plot.line(x = 'year', y =['moving avg_temp_new_york','moving avg_temp_global'] , title="Fig 1: My City Temp",linewidth=1.0);
         plot.ylabel('Temp C')
         plot.show(block=True);
```



Observation_1:

Is your city hotter or cooler on average compared to the global average? Yes, in fig 1, From the line chart, New york city was hotter compared to global trend except around 1750 and 1775-1790.

```
In [20]: Line_chart.plot.line(x = 'year', y =['moving avg_temp_diff'] , title="Fig 2: moving avg_temp_diff",linewidth=1.0);
         plot.ylabel('Temp Difference between New York & Global')
         plot.show(block=True);
```



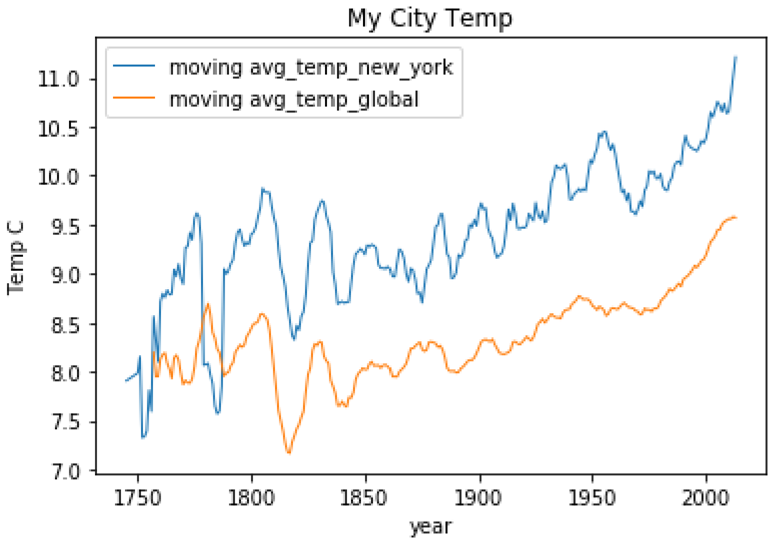
Observation_2:

Has the difference been consistent over time? From fig 2 we can se that, different was always positive most of the time. Difference was fairly consistant between positive 1.0 and 1.5 most of the time except some outlier.

Observation_3:

How do the changes in your city's temperatures over time compare to the changes in the global average? Answer: in fig 1, From the line chart, we can see that both new york city and global average are showing upward trends in temperature.

```
In [21]: #Drawing a line chart
Line_chart.plot.line(x = 'year', y =['moving avg_temp_new_york','moving avg_temp_global'] , title="My City Temp",linewidth=1.0);
plot.ylabel('Temp C')
plot.show(block=True);
```



Observation_4:

What does the overall trend look like? Answer: From the overall trend, it looks like both New York & Global temperature is in upward direction. Its is consistently increasing and thus pose a risk to global warming.

Observation_5:

Is the world getting hotter or cooler? Answer: The world is getting hotter consistently. From the moving average line plot we can see several peaks, but the most alarming observation is in resent years the peak is the highest and there were no downward curve in the last 50 years in global temperature.

Observation_6:

Has the trend been consistent over the last few hundred years? Answer: Temperature was reducing after 1810 but after that Global temperature is rising steadily after 1850. Same applies for New York temperature.