Exploring Weather Trends

- Suraj Singh Shahi

Extracting the data

I have extracted data from the given database, using SQL queries.

2 datasets extracted

- Global Temperature ('global_data.csv') USING:

```
SELECT *
FROM global_data
```

- City Temerature ('delhi_temp.csv') As I live in New Delhi, India.

```
SELECT *
FROM city_data
WHERE Country = 'India' AND City = 'Delhi'
```

In [1]:

```
# Impoting the python libraries and the dataset.
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sb
%matplotlib inline

df_global = pd.read_csv('global_data.csv')
df_delhi = pd.read_csv('delhi_temp.csv')
```

Exploring the datasets

In [2]:

```
df_global.sample()
```

Out[2]:

	year	avg_temp
73	1823	7.72

In [3]:

```
df_delhi.sample()
```

Out[3]:

	year	city	country	avg_temp
76	1872	Delhi	India	24.94

In [4]:

```
df_global.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 [5]:

```
df_delhi.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 218 entries, 0 to 217
Data columns (total 4 columns):
year        218 non-null int64
city        218 non-null object
country     218 non-null object
avg_temp    201 non-null float64
dtypes: float64(1), int64(1), object(2)
memory usage: 6.9+ KB
```

In [6]:

```
pd.isna(df_delhi.avg_temp).sum()
```

Out[6]:

```
17
```

In [7]:

```
df_delhi[df_delhi['avg_temp'].isnull()]
```

Out[7]:

	year	city	country	avg_temp
12	1808	Delhi	India	NaN
13	1809	Delhi	India	NaN
14	1810	Delhi	India	NaN
15	1811	Delhi	India	NaN
16	1812	Delhi	India	NaN
62	1858	Delhi	India	NaN
63	1859	Delhi	India	NaN
64	1860	Delhi	India	NaN
65	1861	Delhi	India	NaN
66	1862	Delhi	India	NaN
67	1863	Delhi	India	NaN
68	1864	Delhi	India	NaN
69	1865	Delhi	India	NaN
70	1866	Delhi	India	NaN
71	1867	Delhi	India	NaN
72	1868	Delhi	India	NaN
73	1869	Delhi	India	NaN

Assessment :

- 17 null values in delhi dataset.
- Betweeen years **1808** and **1869** average temperature values are missing at variarous places.

Solution :

- We know that null values will create problems later.
- We will consider data in both the datasets after **1869** to prevent future problems.

In [8]:

```
df_delhi = df_delhi.query('year > 1869')
df_delhi.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 144 entries, 74 to 217
Data columns (total 4 columns):
year        144 non-null int64
city        144 non-null object
country     144 non-null object
avg_temp    144 non-null float64
dtypes: float64(1), int64(1), object(2)
memory usage: 5.6+ KB
```

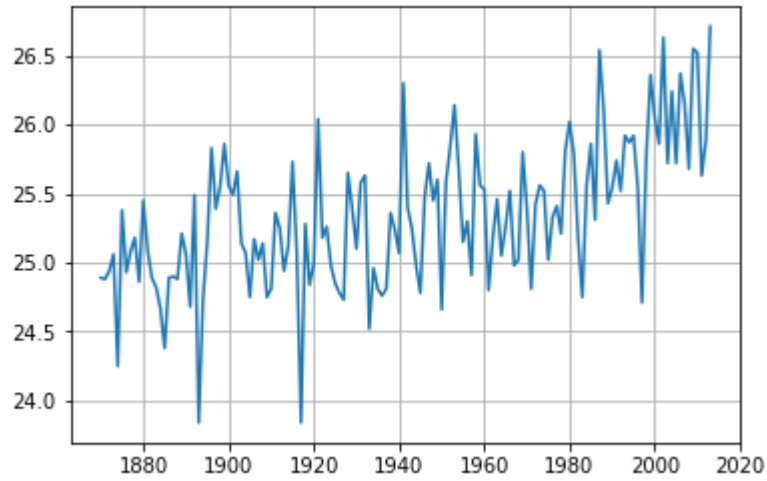
In [9]:

```
df_global = df_global.query('year > 1869')
df_global.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 146 entries, 120 to 265
Data columns (total 2 columns):
year        146 non-null int64
avg_temp    146 non-null float64
dtypes: float64(1), int64(1)
memory usage: 3.4 KB
```

In [10]:

```
plt.plot(df_delhi['year'], df_delhi['avg_temp'], label = "Delhi")
plt.grid(True)
plt.show()
```



Assessment :

- As we plot average annual temperature over time, it creates graphs that is not easy to read.

Solution :

- To create `moving average` of average temperature to smooth out the graph lines.

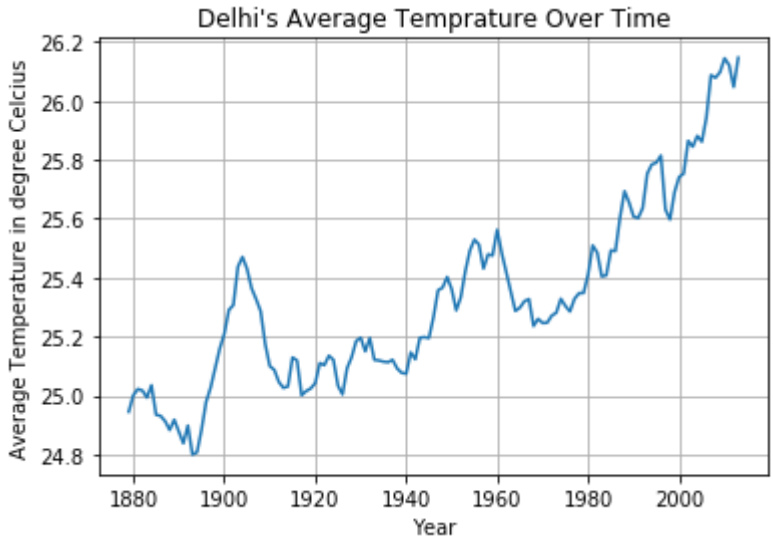
In [11]:

```
global_mavg = df_global['avg_temp'].rolling(10).mean()
delhi_mavg = df_delhi['avg_temp'].rolling(10).mean()
```

Visualization

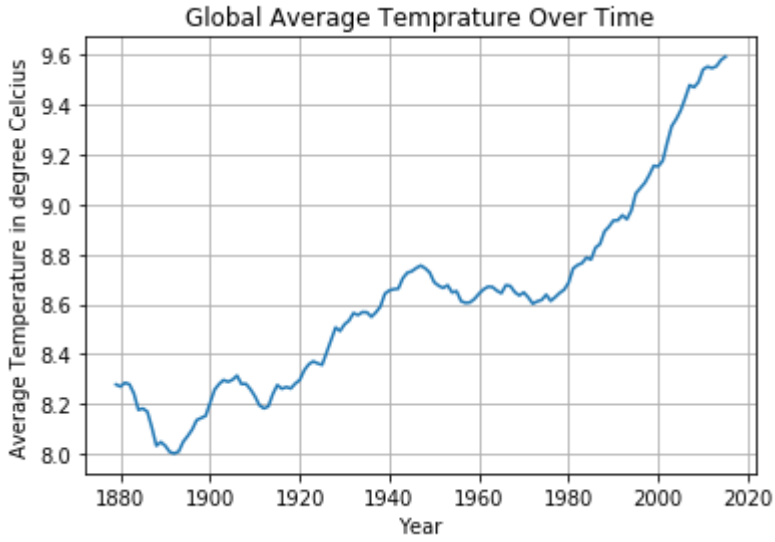
In [12]:

```
plt.plot(df_delhi['year'], delhi_mavg, label = "Delhi")
plt.title("Delhi's Average Temprature Over Time")
plt.xlabel('Year')
plt.ylabel('Average Temperature in degree Celcius')
plt.grid(True)
plt.show()
```



In [13]:

```
plt.plot(df_global['year'], global_mavg, label = "Global Temperatures")
plt.title('Global Average Temprature Over Time')
plt.xlabel('Year')
plt.ylabel('Average Temperature in degree Celcius')
plt.grid(True)
plt.show()
```



OBSERVATION

- Both Delhi's and Global temperature has increased overtime.
- In the last few decades the rate of increase of temperature is considerable larger.

In [14]:

```
df_delhi.avg_temp.mean() - df_global.avg_temp.mean()
```

Out[14]:

16.700742960426176

In [15]:

```
df_delhi.avg_temp.mean() , df_global.avg_temp.mean()
```

Out[15]:

```
(25.339236111111106, 8.63849315068493)
```

In [16]:

```
float(df_delhi.query('year == "2013"').avg_temp.values) - float(df_delhi.query('year == "1870"').avg_temp.values)
```

Out[16]:

```
1.8200000000000003
```

In [17]:

```
# temperature difference in Delhi in the first and last decade.
del_last_decade = df_delhi.query('year >=2004 and year <=2013')['avg_temp'].mean()
del_first_decade = df_delhi.query('year >= 1870 and year < 1880')['avg_temp'].mean()
print("Delhi's average temperature between 1870 to 1880: ",del_first_decade,"\nDelhi's
 average temperature between 2004 to 2013: ",del_last_decade)
print('Difference :',del_last_decade - del_first_decade)
```

```
Delhi's average temperature between 1870 to 1880:  24.945000000000004
Delhi's average temperature between 2004 to 2013:  26.145999999999997
Difference : 1.2009999999999934
```

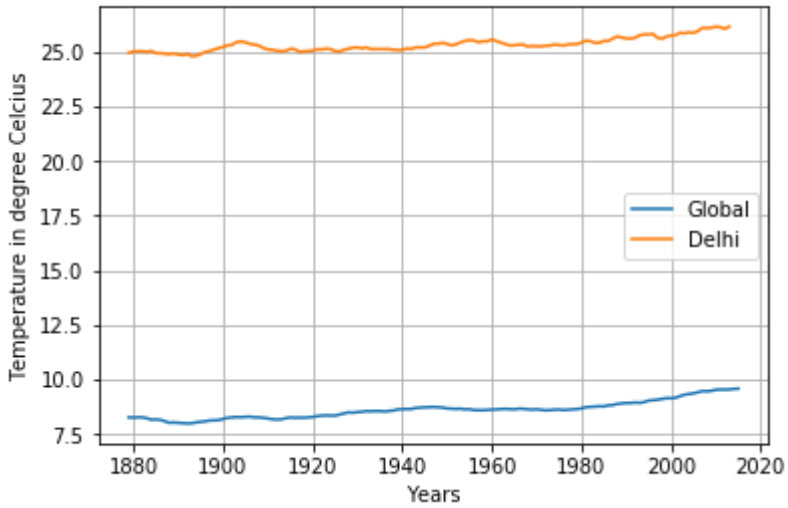
In [18]:

```
# temperature difference in the World in the first and the last decade.
glb_last_decade = df_global.query('year >=2006 and year <=2015')['avg_temp'].mean()
glb_first_decade = df_global.query('year >= 1870 and year <1880')['avg_temp'].mean()
print("World's average temperature between 1870 to 1880: ",glb_first_decade,"\nWorld's
 average temperature between 2006 to 2015: " ,glb_last_decade)
print("Difference :",glb_last_decade - glb_first_decade)
```

```
World's average temperature between 1870 to 1880:  8.277
World's average temperature between 2006 to 2015:  9.593999999999998
Difference : 1.3169999999999984
```

In [19]:

```
# Comparing the two
plt.plot(df_global['year'], global_mavg, label='Global')
plt.plot(df_delhi['year'], delhi_mavg, label = 'Delhi')
plt.grid(True)
plt.legend()
plt.xlabel('Years')
plt.ylabel('Temperature in degree Celcius')
plt.show()
```



FINAL OBSERVATIONS :

- Both Delhi's and Global temperature has `increased` overtime .
- In the `last few decades` the rate of `increase` of temperature is `accelerated`.
- Since `1975` Global average temperature is `raising` without any stops.
- The difference between year **1870** and **2015** in temperature is more than `1.32` °C in the Global average chart.
- The difference between year **1870** and **2015** in temperature is more than `1.2` °C in the National Capital of India, Delhi.
- From the data it looks like world has gotten much hotter than Delhi from **1870** to **2015**.
- Delhi is on average 16.7 °C hotter than the world.

CONCLUSION:

There are `statistically significant` evidences suggesting that the global temperature is raising over the years which support the case of climate change.

TOOLS USED:

- SQL for extraction of data from given database.
- Python libraries to convert the raw data into meaningful, understandable pieces of knowledge by the help of numeric data points, graphs and conclusions. (Pandas, Numpy, Matplotlib).