Extracting Data and Downloading CSV

SELECT *

FROM city_list

WHERE country = 'India’

Since, the nearest city to my location is New Delhi, so I extracted the data for the same along with the Global data using SQL statements as follows:

SELECT

c.year "Year",

c.avg_temp "New Delhi",

g.avg_temp "Global"

FROM global_data g

JOIN city_data c **ON** c.year = g.year

WHERE c.city = 'New Delhi'

The data was obtained as a CSV file named results.csv and the same was used for further analysis.

Analyzing the CSV File

I selected Python for data processing and visualisation for this task.

I chose ***pandas*** library for data processing and ***matplotlib*** for visualisation. All operations were done in ***Jupyter Notebooks*** to create this document.

1. Importing the libraries

```
import matplotlib.pyplot as plt

import pandas as pd

%matplotlib inline
```

2. Creating Pandas Dataframe

```
df = pd.read_csv("results.csv")
```

3. Getting insights from the data

```
df.head()
```

	Year	New Delhi	Global
0	1796	25.03	8.27
1	1797	26.71	8.51
2	1798	24.29	8.67
3	1799	25.28	8.51
4	1800	25.21	8.48

df.columns.unique()

Removing Whitespace from Column names:

df.columns = ['Year', 'NewDelhi', 'Global']

df.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 218 entries, 0 to 217
Data columns (total 3 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   Year      218 non-null    int64
 1   NewDelhi  201 non-null    float64
 2   Global    218 non-null    float64
dtypes: float64(2), int64(1)
memory usage: 5.2 KB
```

df.describe()

	Year	NewDelhi	Global
count	218.000000	201.000000	218.000000
mean	1904.500000	25.166269	8.403532
std	63.075352	0.594003	0.548662
min	1796.000000	23.700000	6.860000
25%	1850.250000	24.800000	8.092500
50%	1904.500000	25.140000	8.415000
75%	1958.750000	25.550000	8.727500
max	2013.000000	26.710000	9.730000

Removing NA values using dropna funtion:

df.dropna(axis=0, how='any', thresh=None, subset=None, inplace=True)

4. Calculating Moving Average using Rolling and Mean function

Using Window=10 sets a 10-year window for the moving average. This value was used because it smoothed the result chart optimally for further analysis among a set of other values.

df['NewDelhi'] = df.rolling(window=10)['NewDelhi'].mean()

df['Global'] = df.rolling(window=10)['Global'].mean()

5. Configuring Parameters of the Graph

params = {

'legend.fontsize': 24,

'figure.figsize': (18, 10),

'axes.labelsize': 20,

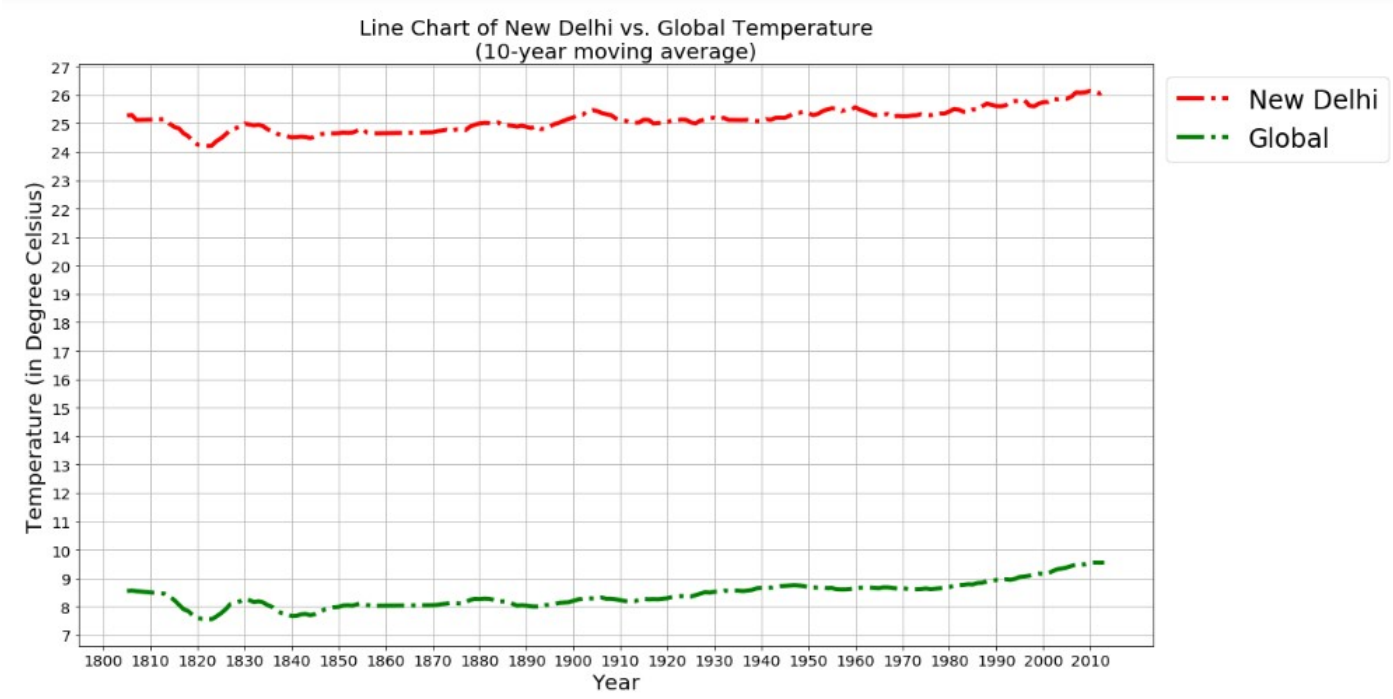
```
    'axes.titlesize': 20,
    'xtick.labelsize': 'x-large',
    'ytick.labelsize': 'x-large'
}
plt.rcParams.update(params)
plt.title('\nLine Chart of New Delhi vs. Global Temperature')
plt.ylabel('Temperature')
plt.xlabel('Year')
plt.xticks(np.arange(1800, 2013, step=10))
plt.plot(df.Year, df.NewDelhi, linewidth=4, color='Red', ls = 'dashdot', label='New Delhi')
plt.plot(df.Year, df.Global, linewidth=4, color='green', ls = 'dashdot', label='Global')
plt.grid(True)
plt.legend(loc='upper left', bbox_to_anchor=(1, 1))
plt.show()
```

6. Plotting the Line Chart

```
plt.plot(df.Year, df.NewDelhi, linewidth=4, color=’Red’, ls = ‘—‘, label=’New Delhi’)
plt.plot(df.Year, df.Global, linewidth=4, label=’Global’)
plt.grid(True)
```

Legend:

```
plt.legend(loc=’upper left’, bbox_to_anchor=(1, 1))
plt.show()
```



Observations:

1. Compared to Global average, my city's temperature is considerably higher i.e. on an average, it is almost 17 degrees hotter than global average!

2. However, after observing the graph, it is clear that the temperature of my city as well as the Global temperature have risen steadily throughout the years. The world is getting hotter every passing year.

3. The temperature of my city has increased by 3 degrees while the global temperature has also increased by an equivalent amount. This is evident from the Min and Max temperature recordings.

4. The ratio of the temperatures of my city and global average is also almost equal to 3 degrees i.e. my city's average temperature is almost 3 times the global average!

5. The Correlation coefficient between my city's and Global temperature is 0.76 so we can conclude it is positive but weak relationship and hence the temperature of my city cannot be estimated based on Global temperature.

Additional observations:

1. The graph clearly shows that there has been a dip in Global temperatures since the 1800s. This can be attributed to the period known as "Little Ice age" where climatologists claim that Europe and Asia have seen dip in temperatures from 1300 to 1800-1850s.

2. Global temperatures saw a spike in 1830 and subsequent steady rise onwards thereafter. A recent article in Carbonbrief.org states that a study published in the journal "Nature" claims that man-made Global warming may have started way before than the 1950s as earlier thought by climatologists. They claimed that they traced Global warming had started specifically in 1830! This could be attributed to the starting of the Industrial Revolution which seemed to have peaked during the 1830s and 1840s.

References:

https://en.wikipedia.org/wiki/Industrial_Revolution

https://www.carbonbrief.org/scientists-clarify-starting-point-for-human-caused-climate-change

https://web.stanford.edu/~moore/Boon_To_Man.html