DATA ANALYST NANO DEGREE

UDACITY

Project 1: Exploring Weather Trends

Submitted by:

Mohit Sindhani

Senior Analyst – Data Science

Infosys Ltd.

Overview:

In this project, we are analyzing local and global temperature data and compare the temperature trends of the nearest city to the average global temperatures. Out the cities which were present in the Database, Delhi has been taken to represent the local place.

SQL has been used for data pulling and Python for data analysis.

SQL Data Extraction

First, Data was downloaded from Udacity portal using a sequence of SQL queries mentioned below:

1. Retrieving the cities in the city_list for India

```
select * from city_list
  where country='India'
```

2. Retrieving the data for city Delhi in the city_list for India

```
select * from city_data
  where country='India'
    and city='Delhi'
```

3. Retrieved the Temperature data for Global numbers and for the city of Delhi

```
select glob.year as Year, glob.avg_temp as Global_Temp, city.avg_temp as Delhi_Temp
  from global_data as glob
    join city_data as city
    on glob.year=city.year
    where city.country='India'
    and city.city='Delhi'
```

4. Reconfirmed the data as data for Delhi contained some missing values

```
select city, count(year) from city_data
  where country='India'
  and avg_temp is null
  group by city
```

The image shows the output of the number of null values present for each city.

Output 22 results

city	count
Agra	12
Ahmadabad	12
Allahabad	8
Amritsar	22
Bangalore	7
Bhopal	7
Delhi	17

Data Analysis

1. Reading the extracted data in a Jupyter Notebook
   Data shows three columns- year, corresponding global and city temperatures.

In [200]:
```
#Reading Final Data having year,avg temperature globally and for Delhi
df=pd.read_csv('final_data.csv')
df.head()
```

Out[200]:

	year	global_temp	delhi_temp
0	1796	8.27	25.03
1	1797	8.51	26.71
2	1798	8.67	24.29
3	1799	8.51	25.28
4	1800	8.48	25.21

In [201]:
```
df.tail()
```

Out[201]:

	year	global_temp	delhi_temp
213	2009	9.51	26.55
214	2010	9.70	26.52
215	2011	9.52	25.63
216	2012	9.51	25.89
217	2013	9.61	26.71

2. Describing the data to understand major central tendencies

In [130]:
```
df.describe()
# Global temperature - Min 6.86, Max 9.73, Range- 2.87, Percentage Difference- 41.83%
# Delhi temperature - Min 23.7, Max 26.71, Range- 3.01, Percentage Difference- 12.7%
```

Out[130]:

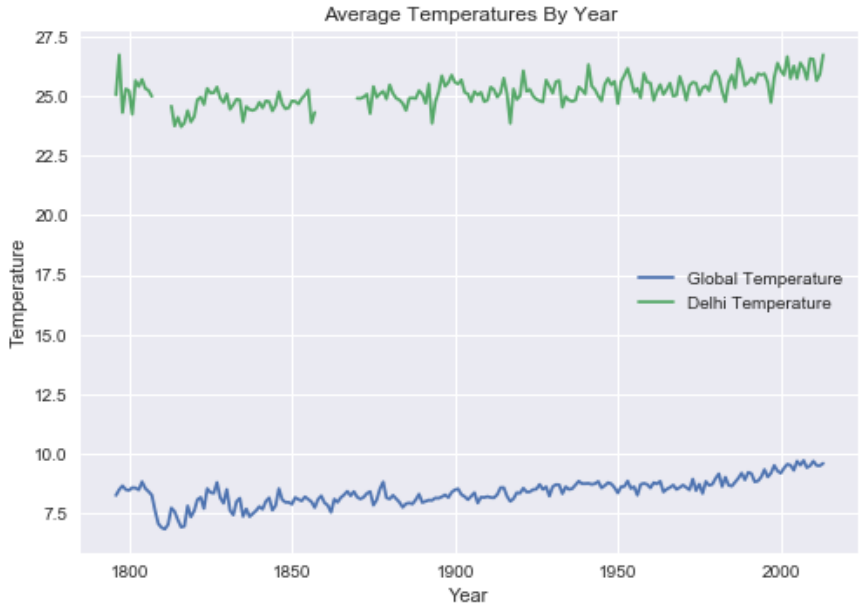
	year	global_temp	delhi_temp
count	218.000000	218.000000	201.000000
mean	1904.500000	8.403532	25.166269
std	63.075352	0.548662	0.594003
min	1796.000000	6.860000	23.700000
25%	1850.250000	8.092500	24.800000
50%	1904.500000	8.415000	25.140000
75%	1958.750000	8.727500	25.550000
max	2013.000000	9.730000	26.710000

3. Analyzing correlation in temperature globally and for the city of Delhi

```
df.corr()
# Global temperature and delhi temperature have a positive correlation coefficient of  0.76
# which show strong correlation in both the variables
```

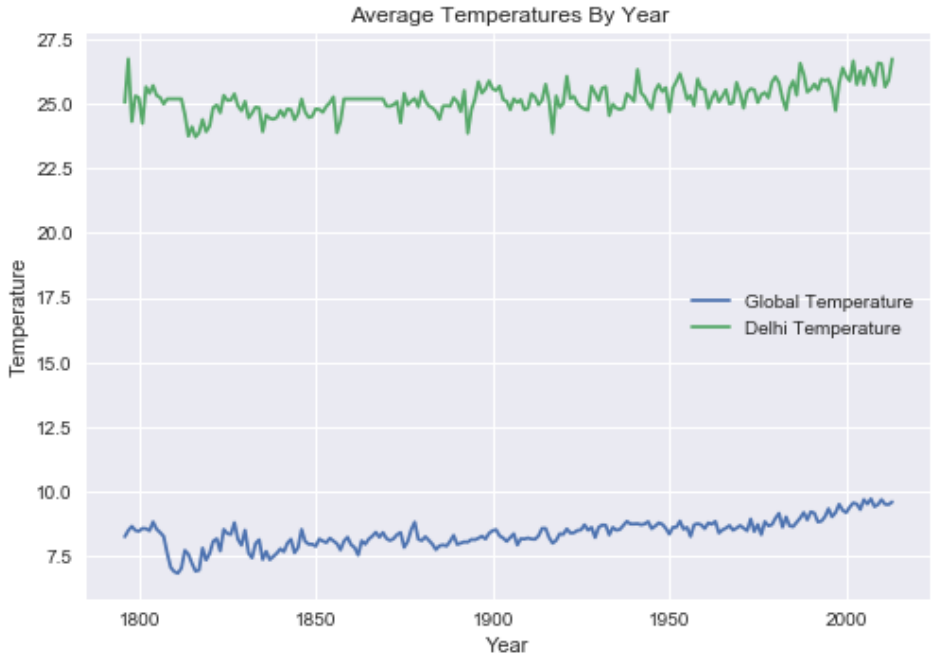
	year	global_temp	delhi_temp
year	1.000000	0.765267	0.602023
global_temp	0.765267	1.000000	0.762654
delhi_temp	0.602023	0.762654	1.000000

4. Plotting the line chart with values of global temperature and Delhi temperature. As there were some null values available for some years for the city of Delhi, we have replaced the values by mean values over the entire period to make the chart continuous.




```
#Replacing missing values with means
mean=df['delhi_temp'].mean()
df['delhi_temp']=df['delhi_temp'].fillna(mean)
```

5. Continuous Line chart showing Global and Delhi Temperature corresponding to various years



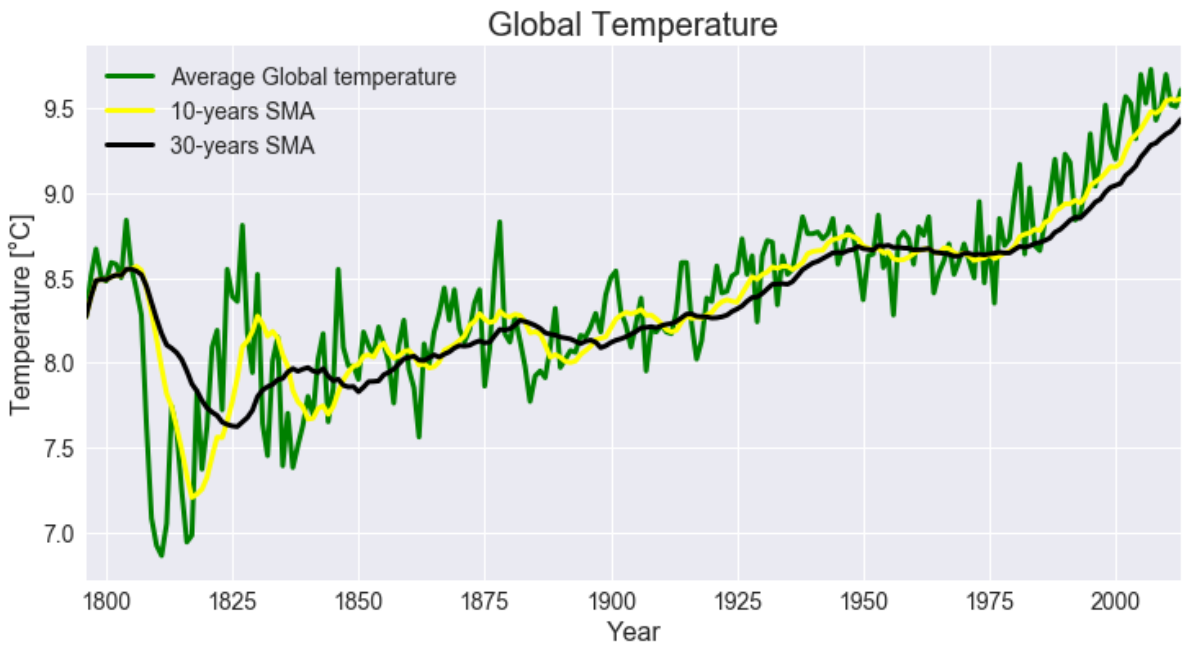
6. Calculating Moving Average for 10 year and 30 year period to see which looks best, also it would help smoothen out the variations which occur when yearly values as taken.

```
# the simple moving average over a period of 10 year
df['Global_MA_10'] = df.global_temp.rolling(10, min_periods=1).mean()
# the simple moving average over a period of 30 years
df['Global_MA_30'] = df.global_temp.rolling(30, min_periods=1).mean()

# the simple moving average over a period of 10 year
df['Delhi_MA_10'] = df.delhi_temp.rolling(10, min_periods=1).mean()


# the simple moving average over a period of 30 year
df['Delhi_MA_30'] = df.delhi_temp.rolling(30, min_periods=1).mean()
```

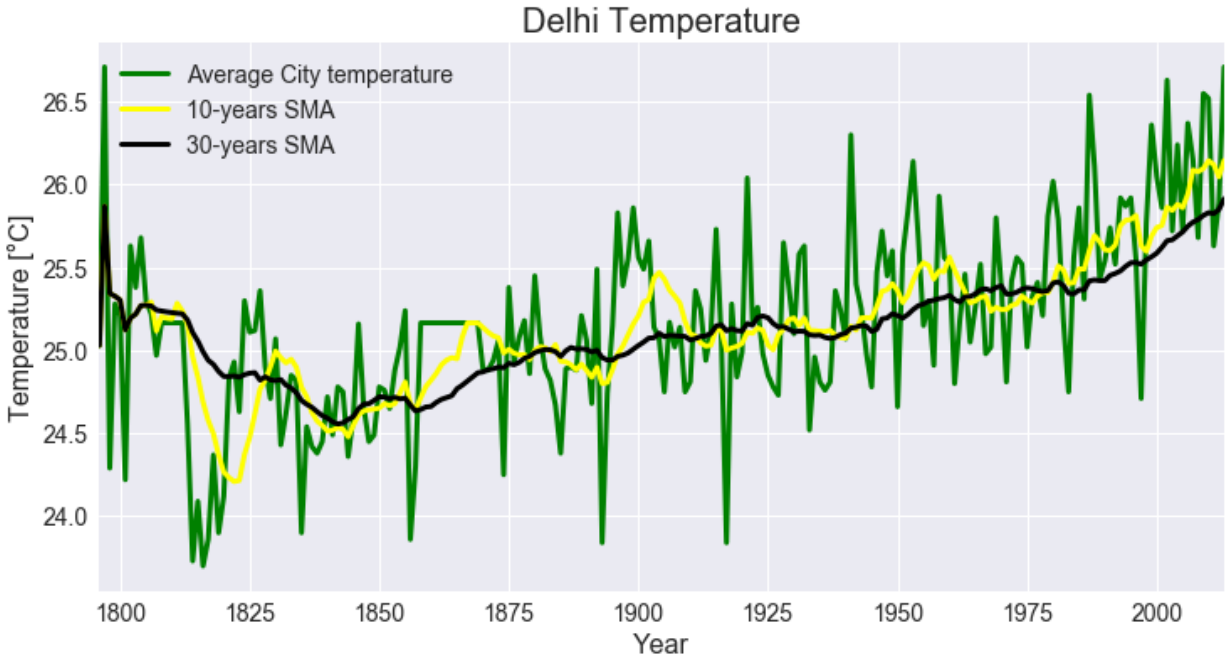
```
df.head()
```

	year	global_temp	delhi_temp	Global_MA_10	Global_MA_30	Delhi_MA_10	Delhi_MA_30
0	1796	8.27	25.03	8.270000	8.270000	25.030000	25.030000
1	1797	8.51	26.71	8.390000	8.390000	25.870000	25.870000
2	1798	8.67	24.29	8.483333	8.483333	25.343333	25.343333
3	1799	8.51	25.28	8.490000	8.490000	25.327500	25.327500
4	1800	8.48	25.21	8.488000	8.488000	25.304000	25.304000

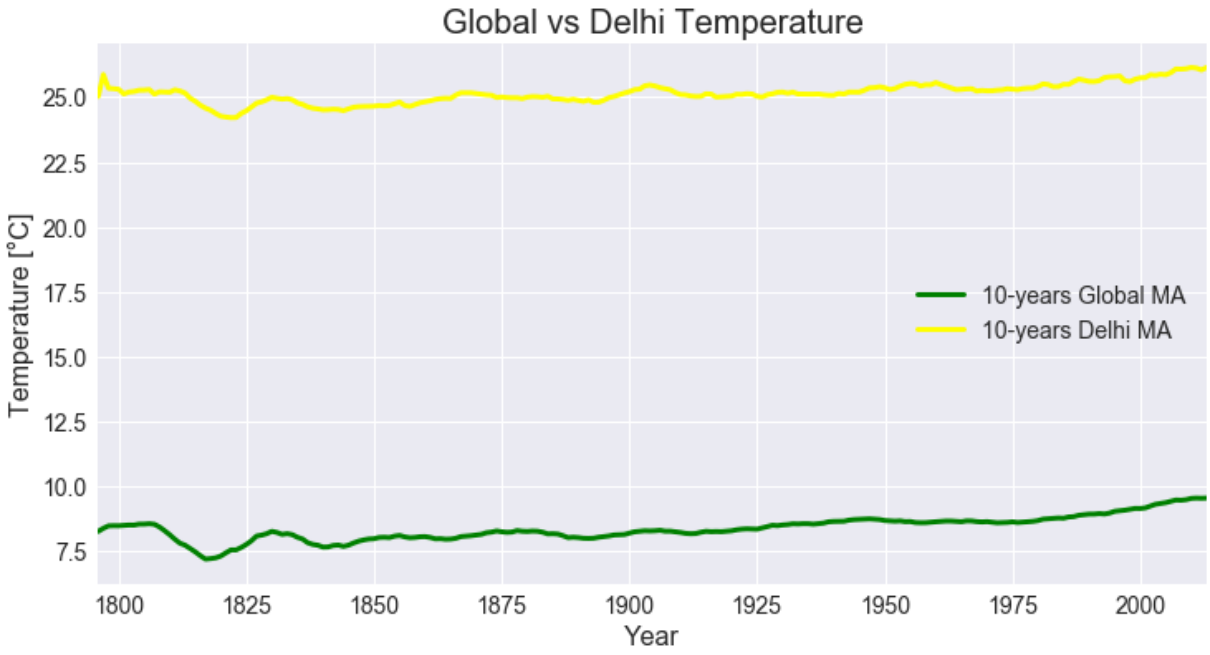
7. Line chart showing Global Temperatures for yearly values, 10 year and 30-year SMA



8. Line chart showing Delhi Temperatures for yearly values, 10 year and 30 year SMA



9. Finally, Line chart showing Global and Delhi Temperatures using 10 year SMA seemed to represent the trend, neither over smoothening it, nor having too much variations.
   Title: Global vs Delhi Temperature
   Axes: Y- Temperature, X – Year
   Legend- Green- 10 year Global SMA and Yellow – 10 year Delhi SMA



Observations:

1. During the late 1790s, temperature was on the higher side, at similar ranges as was in 1970s.

There was a major drop in temperatures in the period 1810-1825.

This drop was very significant as global average fell by 1.5 Deg Celsius which is about 20% drop.

2. The drop was even more terms of magnitude where the temperatures fell by about 3 Deg Cel, about 13% drop. May be this is the reason it was referred to as Little Ice Age (LIA) period by some climate experts.

3. Post 1825, there was a gradual increase in the temperatures and till 1975, it can be seen that global temperatures regained the values which used to be exist in late 1790s. So it took about 150 years to recover from the gradual dip which occurred in early 1800s.

4. Average global temperature increase over the 218 years considered in the dataset is 0.0132 Deg Celsius.

5. Average temperature increase for Delhi is 0.0138, which although slightly higher but is very close to global average

6. Considering the temperature rise in last 40 years globally, the average increase globally is 0.0165 which is 25% more than the global average increase over last 218 years

7. Considering the temperature rise in last 40 years for Delhi, the average increase in Delhi is 0.02875 which is about 100% more than the average increase for the city over last 218 years

This shows the rate of increase of temperature increase is considerably high in the last 40 years.

8. If we decrease our range to consider the last 30 years, for the city of Delhi this average rate increases and comes out to be 0.065 Deg Celsius which is 364% more than average rate over the last 218 years. This figure is worth noticing.

Similarly, global average also increases although less rapidly and average increase for last 30 years is 0.019, an increase of about 45% from average of over 218 years.

9. In terms of magnitude, there is significant difference in Global temperatures and Temperature of Delhi city. The mean Global temperature is 8.403532 whereas for the city of Delhi the value is 25.166269, a difference of about 16.8 Deg Cel, making the temperature of Delhi almost 3 times the Global average.

10. Global temperature and temperature of Delhi have a positive correlation coefficient of 0.76 which shows strong correlation in both the variables.

References

https://www.w3resource.com/graphics/matplotlib/basic/matplotlib-basic-exercise-5.php

https://www.kite.com/python/answers/how-to-find-the-correlation-between-two-pandas-dataframe-columns-in-python

https://www.datacamp.com/community/tutorials/moving-averages-in-pandas

https://www.w3resource.com/graphics/matplotlib/basic/matplotlib-basic-exercise-5.php

https://towardsdatascience.com/moving-averages-in-python-16170e20f6c

Appendix

Attaching the python jupyter notebook.

Exploring Weather Trends.ipynb