Explore Weather Trends - Udacity Project 1 ND DA

July 9, 2020

1 Implementation with Jupyter Notebooks

1.1 Import necessary modules and data

[1]:
```
# pandas for data handling and calculating the moving average,
# matplotlib for the visualisation of the data
# seaborn for optimizing the plots

import pandas as pd
%matplotlib inline
import matplotlib.pyplot as plt
import seaborn as sns
```

[2]:
```
# Import the data downloaded as csvs from the Udacity page as a pandas␣
 ↪dataframe;
# Munich is my hometown and I wanted to compare its avg temperature to the avg
# temperature of cities in other climate zones, thats why I have choosen New␣
 ↪Orleans and Moscow.

df_munich = pd.read_csv('results_munich.csv')
df_no = pd.read_csv('results_no.csv')
df_moscow = pd.read_csv('results_moscow.csv')
df_global = pd.read_csv('results_global.csv')
```

1.2 SQL Queries to extract the data from the Udacity page

- Query_Munich: Select * From city_data Where city = 'Munich';
- Query_Moscow: Select * From city_data Where city = 'Moscow';
- Query_NewOrleans: Select * From city_data Where city = 'New Orleans';
- Query_global: Select * From global_data

1.3 Aligning time scale and checking the data

[3]:
```
# harmonizing the range of the years of the datasets
# and aligning them to the time scale of the New Orleans
# dataset, the dataframes have the same number of rows
# and start with the year 1758 and end with 2013.
```

```
df_munich = df_munich[15:271]
df_moscow = df_moscow[15:271]
df_global = df_global[8:264]
```

[4]:

```
# checking the raw data and the statistics

print('The first five rows of the "Munich" dataset:\n\n', df_munich.head(),␣
 ↪'\n')
print('The last five rows of the "Munich" dataset:\n\n',df_munich.tail(), '\n')
print('The statistics for the "Munich" dataset:\n\n',df_munich.describe(), '\n')
print('The first five rows of the "New Orleans" dataset:\n\n', df_no.head(),␣
 ↪'\n')
print('The last five rows of the "New Orleans" dataset:\n\n',df_no.tail(), '\n')
print('The statistics for the "New Orleans" dataset:\n\n',df_no.describe(),␣
 ↪'\n')
print('The first five rows of the "Moscow" dataset:\n\n',df_moscow.head(), '\n')
print('The last five rows of the "Moscow" dataset:\n\n',df_moscow.tail(), '\n')
print('The statistics for the "Moscow" dataset:\n\n',df_moscow.describe(), '\n')
print('The first five rows of the "Global" dataset:\n\n',df_global.head(), '\n')
print('The last five rows of the "Global" dataset:\n\n',df_global.tail(), '\n')
print('The statistics for the "Global" dataset:\n\n',df_global.describe(), '\n')
```

```
The first five rows of the "Munich" dataset:

     year    city  country  avg_temp
15  1758  Munich  Germany      3.83
16  1759  Munich  Germany      4.89
17  1760  Munich  Germany      5.02
18  1761  Munich  Germany      4.94
19  1762  Munich  Germany      4.49

The last five rows of the "Munich" dataset:

      year    city  country  avg_temp
266  2009  Munich  Germany      5.89
267  2010  Munich  Germany      4.85
268  2011  Munich  Germany      6.56
269  2012  Munich  Germany      5.88
270  2013  Munich  Germany      6.00

The statistics for the "Munich" dataset:

               year    avg_temp
count   256.000000  256.000000
mean   1885.500000    4.658945
std      74.045031    0.719263
min    1758.000000    2.790000
```

```
25%    1821.750000    4.187500
50%    1885.500000    4.665000
75%    1949.250000    5.132500
max    2013.000000    6.640000

The first five rows of the "New Orleans" dataset:

    year         city        country  avg_temp
0  1758  New Orleans  United States     20.87
1  1759  New Orleans  United States     19.73
2  1760  New Orleans  United States      9.24
3  1761  New Orleans  United States       NaN
4  1762  New Orleans  United States       NaN

The last five rows of the "New Orleans" dataset:

      year         city        country  avg_temp
251  2009  New Orleans  United States     21.23
252  2010  New Orleans  United States     20.33
253  2011  New Orleans  United States     21.17
254  2012  New Orleans  United States     21.81
255  2013  New Orleans  United States     22.00

The statistics for the "New Orleans" dataset:

               year    avg_temp
count   256.000000  211.000000
mean   1885.500000   20.221422
std      74.045031    1.136327
min    1758.000000    9.240000
25%    1821.750000   19.915000
50%    1885.500000   20.290000
75%    1949.250000   20.725000
max    2013.000000   22.000000

The first five rows of the "Moscow" dataset:

     year    city country  avg_temp
15  1758  Moscow  Russia      2.22
16  1759  Moscow  Russia      3.45
17  1760  Moscow  Russia      2.41
18  1761  Moscow  Russia      4.14
19  1762  Moscow  Russia      4.10

The last five rows of the "Moscow" dataset:

      year    city country  avg_temp
266  2009  Moscow  Russia      5.69
```

```
267  2010  Moscow  Russia      5.91
268  2011  Moscow  Russia      6.01
269  2012  Moscow  Russia      5.20
270  2013  Moscow  Russia      6.80

The statistics for the "Moscow" dataset:

               year    avg_temp
count   256.000000  256.000000
mean   1885.500000    4.013477
std      74.045031    0.995745
min    1758.000000    1.500000
25%    1821.750000    3.307500
50%    1885.500000    4.000000
75%    1949.250000    4.607500
max    2013.000000    6.800000

The first five rows of the "Global" dataset:

     year  avg_temp
8   1758      6.74
9   1759      7.99
10  1760      7.19
11  1761      8.77
12  1762      8.61

The last five rows of the "Global" dataset:

      year  avg_temp
259  2009      9.51
260  2010      9.70
261  2011      9.52
262  2012      9.51
263  2013      9.61

The statistics for the "Global" dataset:

               year    avg_temp
count   256.000000  256.000000
mean   1885.500000    8.364492
std      74.045031    0.557958
min    1758.000000    6.740000
25%    1821.750000    8.077500
50%    1885.500000    8.360000
75%    1949.250000    8.700000
max    2013.000000    9.730000
```

Summary (raw datasets and statistics):

- Munichs yearly temperature average starts with 3,83 in 1758, and ends with 6.00 in 2013. The yearly temperature average increased 2,17 degrees over time. The mean is 4,65 over time, the standard deviation 0,71. The min value is 2,79 and the maximum value is 6,64.
- New Orleans yearly temperature average starts with 20,87 in 1758, and ends with 22.00 in 2013. The yearly temperature average increased 1,13 degrees over time. The mean is 20,22 over time, the standard deviation 1,13. The min value is 9,24 and the maximum value is 22,00. The max value is also the most recent value.
- Moscows yearly temperature average starts with 2,22 in 1758, and ends with 6.80 in 2013. The yearly temperature average increased 4,58 degrees over time. The mean is 4,01 over time, the standard deviation 1,00. The min value is 1,5 and the maximum value is 6,80. The max value is also the most recent value.
- The global yearly temperture average starts with 8,61 in 1758, and ends with 9,61 in 2013. The yearly temperature average increased 1 degree over time. The mean is 8,36 over time, the standard deviation 0,56. The min value is 6,74 and the maximum value is 9,73.
- All datasets have the same time scale, so they are comparable
- The New Orleans dataset has a lot of NAN Values replaced by interpolated values for the calculation of the moving average
- The New Orleans dataset has the highest standard deviation of and the global dataset the lowest

1.4 Plotting the raw data

[5]:

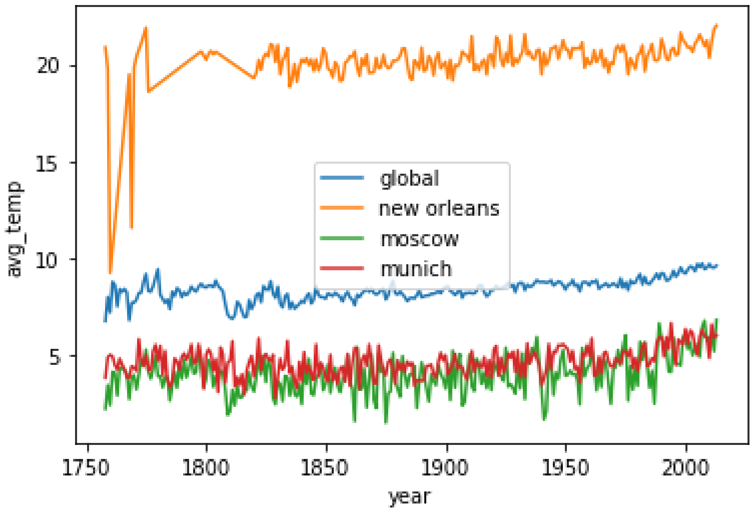
```
# plot all lines in one diagram

plt.title("Global vs. Munichs/Moscows/New Orleans yearly average temperature␣
 ↪from 1750 to 2013 \n")

sns.lineplot(df_global['year'], df_global['avg_temp'], label = 'global')
sns.lineplot(df_no['year'], df_no['avg_temp'].interpolate(), label = 'new␣
 ↪orleans')
sns.lineplot(df_moscow['year'], df_moscow['avg_temp'], label = 'moscow')
sns.lineplot(df_munich['year'], df_munich['avg_temp'], label = 'munich')

plt.show()
```

Global vs. Munichs/Moscows/New Orleans yearly average temperature from 1750 to 2013



1.5 Correlation coefficient

[6]:
```
# converting the avg_temp columns in pd Series

s_munich = df_munich['avg_temp']
s_no = df_no['avg_temp']
s_moscow = df_moscow['avg_temp']
s_global = df_global['avg_temp']
```

[7]:
```
# calculating the standard correlation coefficient

print('Correlation of Munich dataset and others\n')
print('The Correlation between the average temperature of Munich and Moscow:',␣
 ↪s_munich.corr(s_moscow))
print('The Correlation between the average temperature of Munich and New␣
 ↪Orleans:', s_munich.corr(s_no))
print('The Correlation between the average temperature of Munich and the global␣
 ↪average:', s_munich.corr(s_global))
print('\nCorrelation of Moscow dataset and others\n')
print('The Correlation between the average temperature of Moscow and New␣
 ↪Orleans:', s_moscow.corr(s_no))
print('The Correlation between the average temperature of Moscow and the global␣
 ↪average:', s_moscow.corr(s_global))
print('\nCorrelation of New Orleans dataset and global\n')
print('The Correlation between the average temperature of New Orleans and the␣
 ↪global average:', s_no.corr(s_global))
```

```
Correlation of Munich dataset and others

The Correlation between the average temperature of Munich and Moscow:
0.5759752632787498
The Correlation between the average temperature of Munich and New Orleans:
0.2743534125855344
The Correlation between the average temperature of Munich and the global
average: 0.36023517386163845

Correlation of Moscow dataset and others

The Correlation between the average temperature of Moscow and New Orleans:
0.12805995267127246
The Correlation between the average temperature of Moscow and the global
average: 0.36294906044122505

Correlation of New Orleans dataset and global

The Correlation between the average temperature of New Orleans and the global
average: 0.28436510651171515
```

Summary (correlation of raw data):

- All correlation coefficients are positive
- The datasets of Moscow and Munich have the highest correlation coefficient 0.58
- The datasets of Moscow and New Orleans have the lowest correlation coefficient 0.13
- The correlation coefficient of the global dataset and the city datasets is between 0.28 (New Orleans) and 0.36 (Munich/Moscow)

1.6 Moving Average

[8]:

```python
# calculate the moving average with a window of seven years and adjusting the timescale starting with 1764

s_munich_MA7d = s_munich.rolling(window=7).mean()
s_no_MA7d = s_no.interpolate().rolling(window=7).mean()
s_moscow_MA7d = s_moscow.rolling(window=7).mean()
s_global_MA7d = s_global.rolling(window=7).mean()
s_year = df_munich['year'][6:]
```

[9]:

```python
# checking the data with the moving average

print('The moving average over seven years of Munichs average temperature starts with:\n', s_munich_MA7d.head(10))
print('The moving average over seven years of Munichs average temperature ends with:\n', s_munich_MA7d.tail())
print('The statistics of the moving average over seven years of Munichs average temperature:\n', s_munich_MA7d.describe())
print('The moving average over seven years of New Orleans average temperature starts with:\n', s_no_MA7d.head(10))
print('The moving average over seven years of New Orleans average temperature ends with:\n', s_no_MA7d.tail())
print('The statistics of the moving average over seven years of New Orleans average temperature:\n', s_no_MA7d.describe())
print('The moving average over seven years of Moscows average temperature starts with:\n', s_moscow_MA7d.head(10))
print('The moving average over seven years of Moscows average temperature ends with:\n', s_moscow_MA7d.tail())
print('The statistics of the moving average over seven years of Moscows average temperature:\n', s_moscow_MA7d.describe())
print('The moving average over seven years of the global average temperature starts with:\n', s_global_MA7d.head(10))
print('The moving average over seven years of the global average temperature ends with:\n', s_global_MA7d.tail())
print('The statistics of the moving average over seven years of the global average temperature:\n', s_global_MA7d.describe())
print('The time scale starts with:\n', s_year.head(10))
print('The time scale ends with:\n', s_year.tail())
print('The statistics of the time scale:\n', s_year.describe())
```

```
The moving average over seven years of Munichs average temperature starts with:
 15          NaN
16          NaN
17          NaN
18          NaN
19          NaN
20          NaN
21     4.605714
22     4.704286
23     4.617143
24     4.450000
Name: avg_temp, dtype: float64
The moving average over seven years of Munichs average temperature ends with:
 266    5.801429
267    5.620000
268    5.781429
269    5.905714
270    5.914286
Name: avg_temp, dtype: float64
The statistics of the moving average over seven years of Munichs average
temperature:
 count    250.000000
mean       4.645349
std        0.434946
min        3.794286
```

```
25%        4.370000
50%        4.587143
75%        4.810000
max        5.914286
Name: avg_temp, dtype: float64
The moving average over seven years of New Orleans average temperature starts
with:
 0          NaN
1          NaN
2          NaN
3          NaN
4          NaN
5          NaN
6    14.226786
7    13.478750
8    13.076250
9    14.355000
Name: avg_temp, dtype: float64
The moving average over seven years of New Orleans average temperature ends
with:
 251    21.125714
252    21.082857
253    21.098571
254    21.178571
255    21.245714
Name: avg_temp, dtype: float64
The statistics of the moving average over seven years of New Orleans average
temperature:
 count    250.000000
mean      20.037441
std        1.041198
min       13.076250
25%       19.915000
50%       20.190000
75%       20.480357
max       21.245714
Name: avg_temp, dtype: float64
The moving average over seven years of Moscows average temperature starts with:
 15         NaN
16         NaN
17         NaN
18         NaN
19         NaN
20         NaN
21    3.368571
22    3.665714
23    3.777143
24    4.011429
```

```
Name: avg_temp, dtype: float64
The moving average over seven years of Moscows average temperature ends with:
 266    5.662857
267    5.794286
268    5.892857
269    5.844286
270    6.114286
Name: avg_temp, dtype: float64
The statistics of the moving average over seven years of Moscows average
temperature:
 count    250.000000
mean       4.000971
std        0.573754
min        2.665714
25%        3.667500
50%        3.892143
75%        4.268214
max        6.114286
Name: avg_temp, dtype: float64
The moving average over seven years of the global average temperature starts
with:
 8          NaN
9          NaN
10         NaN
11         NaN
12         NaN
13         NaN
14    7.885714
15    8.101429
16    8.161429
17    8.308571
Name: avg_temp, dtype: float64
The moving average over seven years of the global average temperature ends with:
 259    9.535714
260    9.560000
261    9.588571
262    9.561429
263    9.572857
Name: avg_temp, dtype: float64
The statistics of the moving average over seven years of the global average
temperature:
 count    250.000000
mean       8.358943
std        0.466867
min        7.191429
25%        8.056071
50%        8.315000
75%        8.635357
```

```
max        9.588571
Name: avg_temp, dtype: float64
The time scale starts with:
 21    1764
22    1765
23    1766
24    1767
25    1768
26    1769
27    1770
28    1771
29    1772
30    1773
Name: year, dtype: int64
The time scale ends with:
 266    2009
267    2010
268    2011
269    2012
270    2013
Name: year, dtype: int64
The statistics of the time scale:
 count     250.000000
mean     1888.500000
std        72.312977
min      1764.000000
25%      1826.250000
50%      1888.500000
75%      1950.750000
max      2013.000000
Name: year, dtype: float64
```

Summary (checking data with moving average):

- The Munich dataset with the moving average starts with 4,61 and ends with 5,91. This is a difference of 1,3 degrees over time. The mean is 4,65 over time, the standard deviation 0,43. The min value is 3,79 and the maximum value is 5,91. The max value is also the most recent value.
- The New Orleans dataset with the moving average starts with 14,23 and ends with 21,25. This is a difference of 7,02 degrees over time. The mean is 20,04 over time, the standard deviation 1,04. The min value is 13,08 and the maximum value is 21,25. The max value is also the most recent value.
- The Moscow dataset with the moving average starts with 3,37 and ends with 6,11. This is a difference of 2,74 degrees over time. The mean is 4,00 over time, the standard deviation 0,57. The min value is 2,67 and the maximum value is 6,11. The max value is also the most recent value.
- The global dataset with the moving average starts with 7,88 and ends with 9,57. This is a difference of 1,69 degrees over time. The mean is 8,36 over time, the standard deviation 0,47. The min value is 7,19 and the maximum value is 9,59.

- All datasets have 250
- The timescale starts with 1764 and ends with 2013
- The use of the moving average over 7 years evens the data

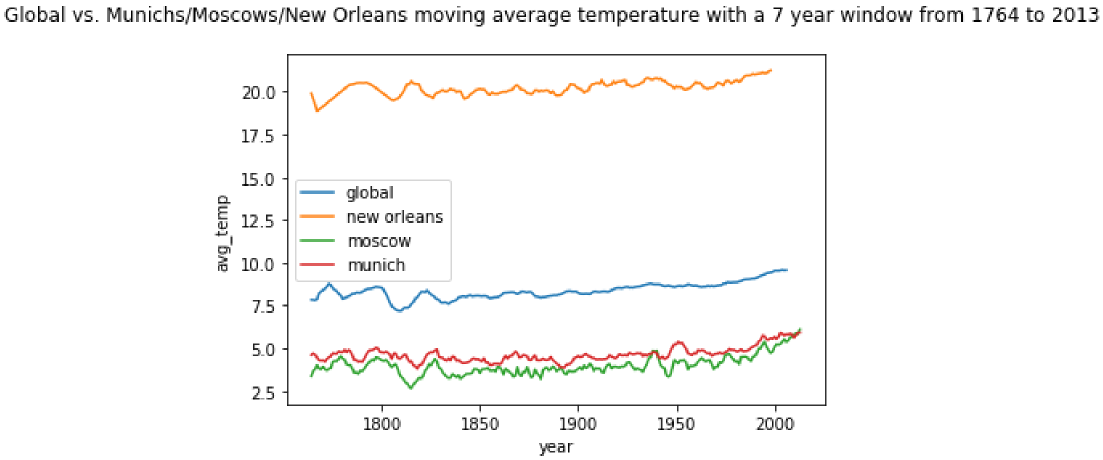
1.7 Plotting the data with the moving Average

[10]:
```
# plot all lines in one diagram

plt.title("Global vs. Munichs/Moscows/New Orleans moving average temperature␣
 ↪with a 7 year window from 1764 to 2013 \n")

sns.lineplot(s_year, s_global_MA7d, label = 'global')
sns.lineplot(s_year, s_no_MA7d, label = 'new orleans')
sns.lineplot(s_year, s_moscow_MA7d, label = 'moscow')
sns.lineplot(s_year, s_munich_MA7d, label = 'munich')

plt.show()
```



2 Observations about Similarities/Differences in the Data Trends

- Munichs/Moscows average temparature is lower than the global average, New Orleans temperature is higher
- All curves are rising. Some of the curves seem to rise faster, Moscows average temperature seems to increase faster than Munichs. Munichs and Moscows lines cross, the 7 year moving average of Moscows temperature became higher than Munichs in recent years.
- The peaks and lows are similar, but have a different intensity, f.e. between 1800 and 1850 the impact in Moscow was heavier and and a little bit earlier than in Munich.
- Checking the data with, I saw that the moving average temperature of the cities in 2013 is also the max value in the dataset, only the global dataset has the max value in 2011. Taking the raw data this is also the case for New Orleans and Moscow. My conclusion is that the we all over the world live now in the hottest times ever since we are measuring the temperature

- The correlation coefficients are all positive, Munich and Moscow have the highest correlation regarding their average temperature. New Orleans and Moscows average temperature correlation coefficient is the lowest. My interpretation is that the average temperatures of Munich and Moscow are quite similar in comparison to the other curves, as the diagram shows. The correlation coefficient of the global dataset and the city datasets is between 0.28 (New Orleans) and 0.36 (Munich/Moscow).
- The highest standard deviation has the "New Orleans" dataset, the lowest standard deviation has the "Munich" dataset. From my point of view a high standard deviation could indicate relativly inconsistent temperatures.