Data Analyst Nanodegree: ***Explore weather trends***

The first step will be to extract the data we want through a SQL query, both for Milan (my city) and the rest of the world:

For Milan:

```
SELECT c.year, c.city, c.country, c.avg_temp
FROM city_data c
JOIN global_data g
ON c.year=g.year
WHERE c.city='Milan'
GROUP BY 2,1,3,4
ORDER BY 2,1
```

For the rest of the world:

```
SELECT *
from global_data
```

After that we are going to try to organize the data of both documents in order to get a visualization to be able to compare the temperature tendence in Milan and in the rest of the world working in a Jupyter Notebook.

Regarding the moving averages, they have been calculated both for the “world” data and the “Milan" data using Python. To calculate the moving averages, I used fifteen values in “Milan” and for the “world” file, so the first moving average result is placed in 1764 year.

Then, we will compare the temperatures with a line chart, so the difference in the temperature development is clear and easy to analyse.

In [65]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sb
%matplotlib inline
```

In [66]:
```
milan = pd.read_excel('milanos.xlsx')
```

In [67]:
```
milan['moving_avg1']=milan.iloc[:,4].rolling(window=15).mean()
milan.tail(3)
```

Out[67]:

	year	city	country	avg_temp	avg	moving_avg1
261	2011	Milan	Italy	868	8.68	7.974667
262	2012	Milan	Italy	805	8.05	7.966000
263	2013	Milan	Italy	804	8.04	7.984000

In [68]:
```
world=pd.read_excel('globals.xlsx')
world.head()
```

Out[68]:

	year	avg_temp	avg_temp2
0	1750	872	8.72
1	1751	798	7.98
2	1752	578	5.78
3	1753	839	8.39
4	1754	847	8.47

In [69]:
```
world['moving_avg2']=world.iloc[:,2].rolling(window=15).mean()
```

In [70]:
```
world['city']='World'
world.head(1)
```

Out[70]:

	year	avg_temp	avg_temp2	moving_avg2	city
0	1750	872	8.72	NaN	World

In [71]:
```
temperature=pd.merge(world,milan,on=['year','city'],how='outer')
temperature.head(1)
```

Out[71]:

	year	avg_temp_x	avg_temp2	moving_avg2	city	country	avg_temp_y	avg	moving_avg1
0	1750	872.0	8.72	NaN	World	NaN	NaN	NaN	NaN

In [72]:
```
world['moving_avg2'].mean(), world['moving_avg2'].std()
```

Out[72]: (8.345158730158724, 0.4197224629806731)

In [73]:
```
world['moving_avg2'].min(), world['moving_avg2'].max()
```

Out[73]: (7.4086666666666625, 9.564666666666659)

In [74]:
```
milan['moving_avg1'].mean(), milan['moving_avg1'].std()
```

Out[74]: (6.7855680000000005, 0.36536911711715775)

In [75]:
```
milan['moving_avg1'].min(), milan['moving_avg1'].max()
```

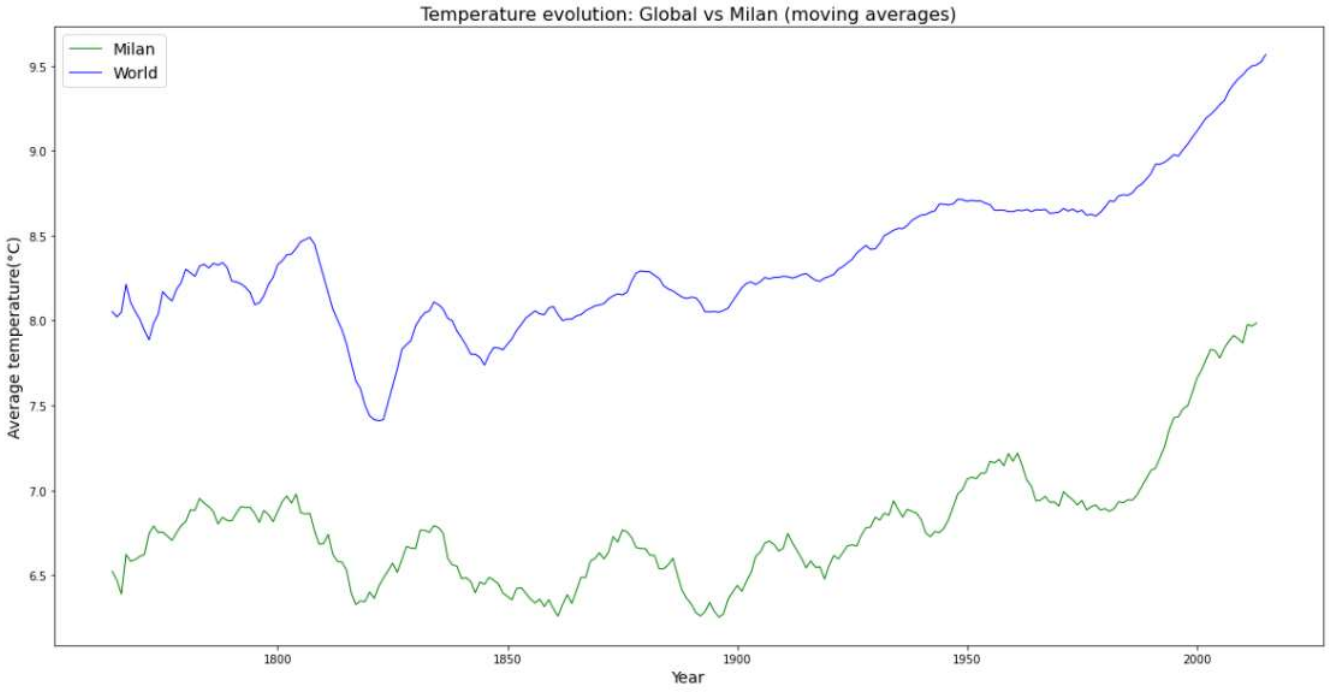
Out[75]: (6.252000000000001, 7.983999999999998)

In [76]:
```
temperature.drop(temperature.columns[[1,2,5,6,7]], axis=1,inplace=True)
temperature.tail(1)
```

Out[76]:

	year	moving_avg2	city	moving_avg1
529	2013	NaN	Milan	7.984

In [83]:
```
plt.figure(figsize=(20,10))
plt.plot( 'year', 'moving_avg1', data=temperature, color='green', linewidth=1, label='Milan')
plt.plot( 'year', 'moving_avg2', data=temperature, color='blue', linewidth=1, label='World')
plt.title('Temperature evolution: Global vs Milan (moving averages)', loc='center', fontsize=16)
plt.xlabel("Year", fontsize=14)
plt.ylabel("Average temperature(°C)", fontsize=14)
plt.legend(loc='upper left', fontsize=14);
```



As we can see on the visualization, the temperature in Milan has always been colder than on the rest of the world, being the average for Milan in the observed time range 6,8°C and 8,3°C the average for the rest of the world. Furthermore, the maximum and minimum temperatures have always been higher in the rest of the world compared to Milan, being 9,5°C the maximum for the world and 8°C for Milan. It happens the same with the minimum temperature, which was 6,2°C for Milan and 7,4°C for the rest of the world. In addition, it is also possible to see how the world temperature is always getting higher, including Milan, especially after 1970, where the tendence is an even stronger temperature increase. If we check the standard deviation values, we can see that the value for Milan is 0,36 and 0,42 for the rest of the world, which means that the temperature values in Milan are closer to the overall average temperature than for the global data.

With this information it is possible to say that the visualization is showing that the world is experiencing a global warming, possibly because of the development of technologies which contribute to this temperature evolution. On the other hand, Milan follows the same development shown by the global blue line but always keeping a difference of 1-1,5°C for the mean, the minimum and maximum values.