Udacity

Programm:

Data Analyst Nanodegree

PROJECT:

EXPLORE WEATHER TRENDS

Picture 1

Goran Horvat

04.4.2020.

PROJECT INSTRUCTION

Overview

The assignment of the project is to analyze, visualize and make a conclusion about temperature movements around the world based on local and global temperature data.

I live in Graz, Austria and for me the nearest big city included in the database from which I'll extracting data is Vienna. For this city, I'll make exploration and comparison to global temperature.

The analysis will include also a few other cities around the world to make some more conclusions about temperature changes.

In recent years increasing temperature on Earth is in the focus of politicians, economists, and activists for protecting the Earth and this is a hot theme.

Extraction of data

The input data was provided from the Udacity portal database from these three tables:

`city_data, city_list, global_data`

For this purpose SQL query was used to retrieve data:

```
SELECT
      c.year,
      c.country,
      c.city,
      c.avg_temp  AS avg_temp_city,
      g.avg_temp AS avg_temp_global
FROM city_data c
      JOIN global_data g USING(year)
WHERE
      (city IN ('Vienna', 'New York', 'Amsterdam',
              'Delhi', 'Shanghai','Lagos','Sydney'));
```

Retrieved data was saved in CSV file query_7_cities.csv and further used for the process of exploration in Jupiter Notebook using Python.

ANALYSIS

Importing libraries and loading data

The analysis was started with the importing of necessary libraries and loading of .csv file and then selecting the only temperature for the world and Vienna. Required temperature data was retrieved with function .describe().

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
sns.set(style='whitegrid')

data = pd.read_csv('query_7_cities.csv')
df_vienna = data[data['city']=='Vienna'].copy()
df_vienna.reset_index(drop=True, inplace = True)
df_vienna.describe()
```

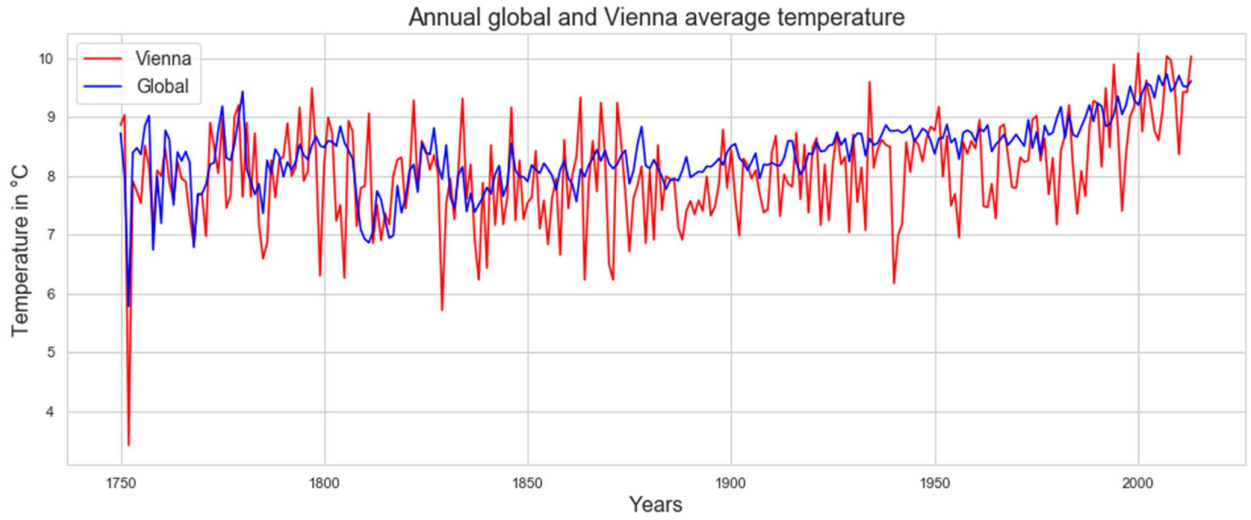
	year	avg_temp_city	avg_temp_global
count	264.000000	264.000000	264.000000
mean	1881.500000	8.042083	8.359394
std	76.354437	0.846756	0.575184
min	1750.000000	3.420000	5.780000
25%	1815.750000	7.507500	8.077500
50%	1881.500000	8.075000	8.365000
75%	1947.250000	8.592500	8.700000
max	2013.000000	10.080000	9.730000

The presented data show that the lowest temperature in Vienna was 3.42°C, and the highest 10.08°C, and on the other hand the global minimum was 5.78°C and maximum 9.73°C, respectively.
The difference between average temperature is 0.32°C and Vienna is on average colder than the world.

Visualization of average temperature

Next view is chart of global and Vienna average temperature.

```
dim=(16,6)
fig, ax = plt.subplots(1,1, figsize=dim)
sns.lineplot('year','avg_temp_city', data = df_vienna,
                                                label = 'Vienna', color = 'red')
sns.lineplot('year','avg_temp_global', data = df_vienna,
                                           label = 'Global', color = 'blue')
plt.title ('Exploring Weather Trends for Vienna and World', fontsize=18)
plt.xlabel ('Years', fontsize=16)
plt.ylabel ('Temperature in °C', fontsize=16)
plt.legend(loc=2, fontsize = 14)
plt.show()
```



It is evident that there are large and frequent oscillations of temperature in Vienna, much more than on a global scale. Keeping track of trends on this graph is difficult, so I will use the 'moving average' technique to get an overall idea of the trends. The moving average will soften the temperature line that the graph becomes more readable, but the question is which time window should be used.
Time windows of 2, 10, 50 and 100 years have been used and the calculation using the .rolling () function was made.

Moving average visualization

```
ma_win =[10,50] # moving average windows

for y in range(len(ma_win)):
    df_vienna['ma_global_' + str(ma_win[y])]= np.nan
    for x in range(len(df_vienna)-ma_win[y]+1):
        df_vienna['ma_global_' + str(ma_win[y])] = df_vienna['avg_temp_global']
                        .rolling(window=ma_win[y]).mean()
        df_vienna['ma_city_' + str(ma_win[y])] = df_vienna['avg_temp_city']
                        .rolling(window=ma_win[y]).mean()
```

Chart 1. - Moving average temperature for 10 years period

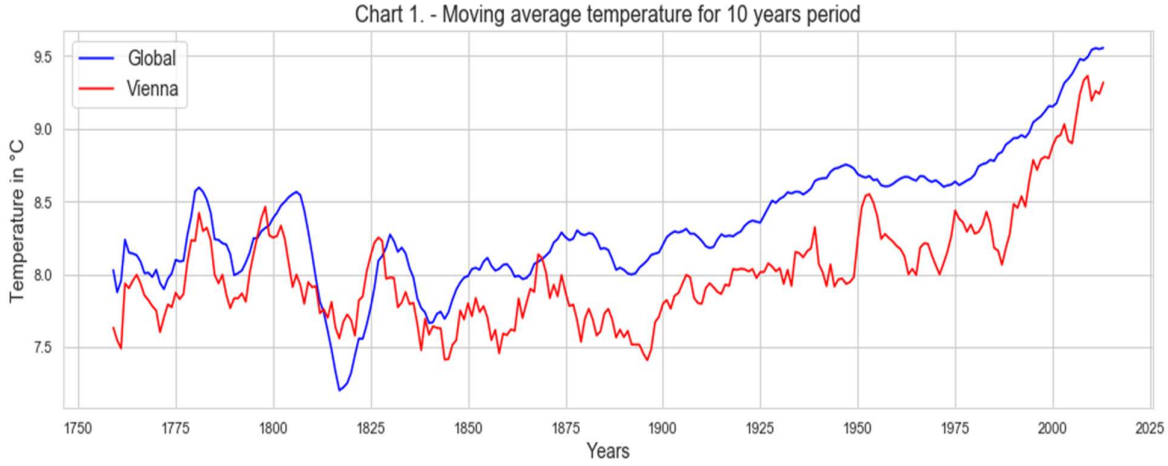
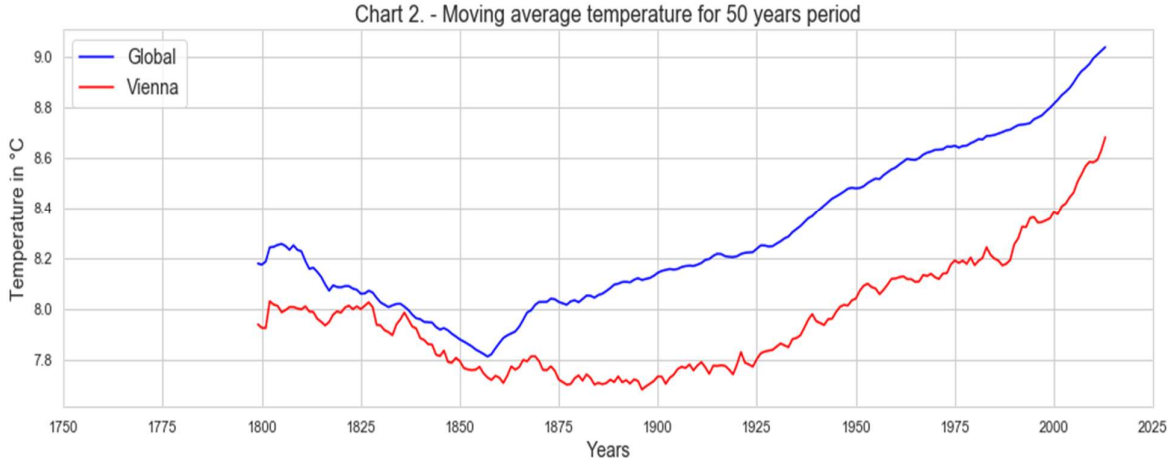
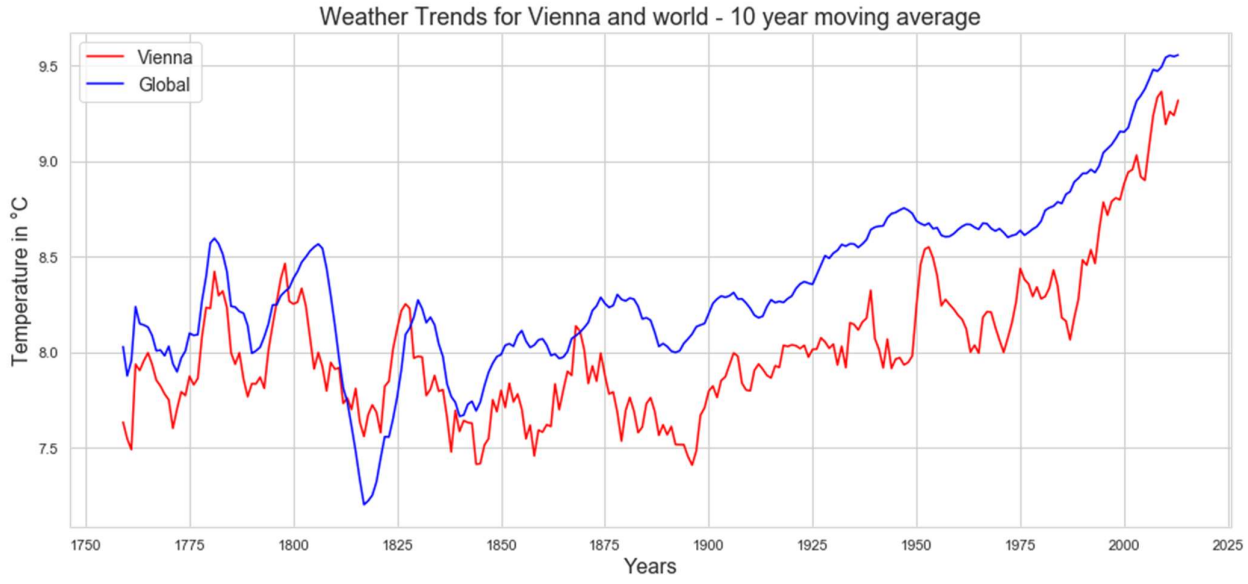


Chart 2. - Moving average temperature for 50 years period



Observations:

The easiest to read are graphs with a time window of 10 and 50 years. 10 years window gives us a clear insight into periods of rising and falling in average temperatures, while with the 50 years time window we can see two main trends in the observed period of 264 years.
Therefore, a 10-years time window will be used, because it is more detailed.



The main observation here is that the average temperature line for Vienna follows the global average temperature line, but we can define three main periods of this timeline:

The first period is from 1750 to 1870, where the global and Vienna temperatures are generally in similar frames. During this time, the global temperature is on average only 0.17 ° C higher than in Vienna with a slight decrease in temperature for both observations, which we will later confirm through the linear regression line.

Second period: the rise of global temperature.

After 1870, a sharp rise in global temperature is visible, while the Vienna temperature just starting to rise slightly.

The period from 1870 to 2013 is where we see the separation of the two lines and it is clear to see increasing world temperature.

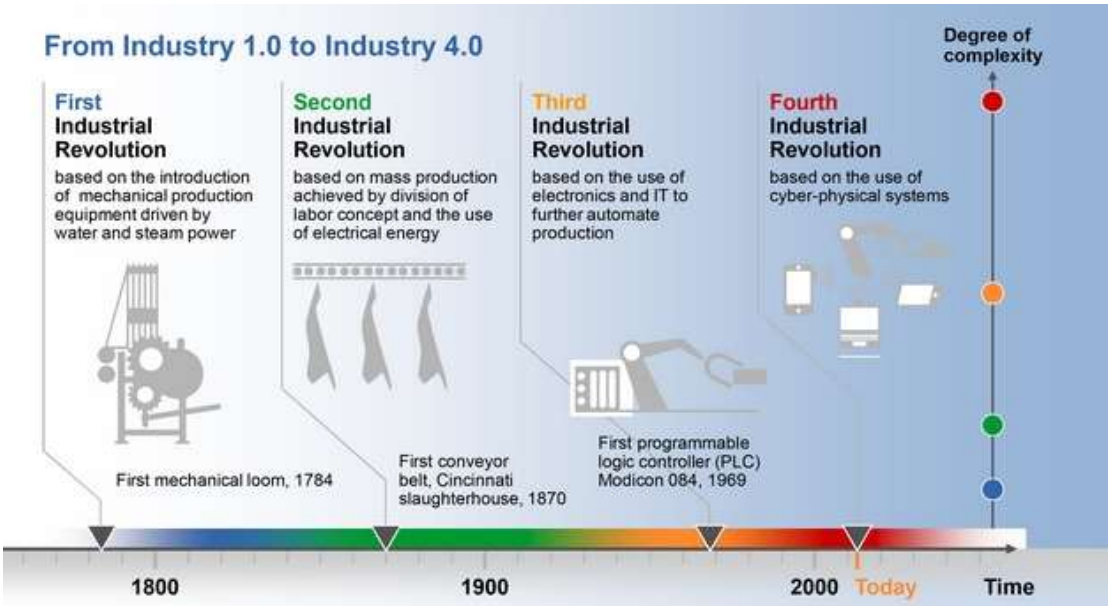
The average global temperature is rising faster than in Vienna and this difference is 0.44° C.

The third observation period is from 1975 to 2013 and in this small period, there is a sharp increase in the average temperature in Vienna and the world. In this period we have an average difference between world and Vienna of 0.38C, and the global average temperature is 9,06.

Average temperature increase is more easily visible from next table.

	Difference	Global average	Vienna average
1750 - 1870	0.171342	8.035631	7.864288
1870 - 2013	0.435329	8.585378	8.150049
1975 - 2013	0.378053	9.064158	8.686105

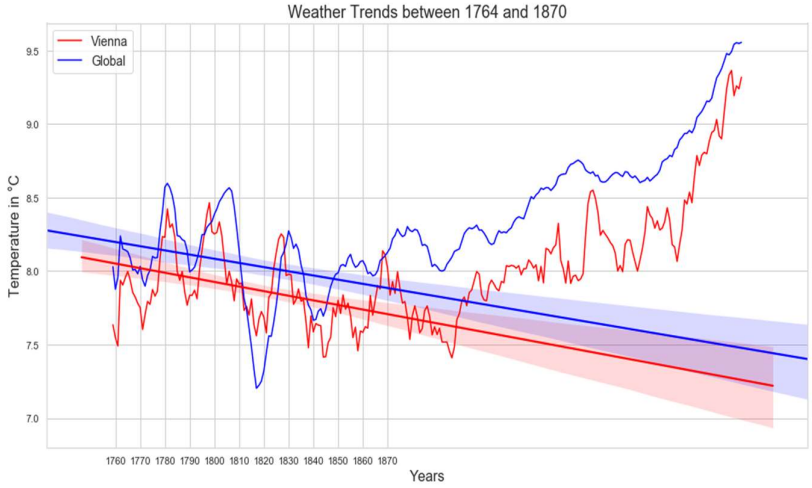
The three observation periods correspond to the 1st, 2nd and 3rd Industrial Revolution, ie. with the introduction of new technologies that are fueled by the high combustion of fossil fuels, especially coal and oil. Please, see the picture.

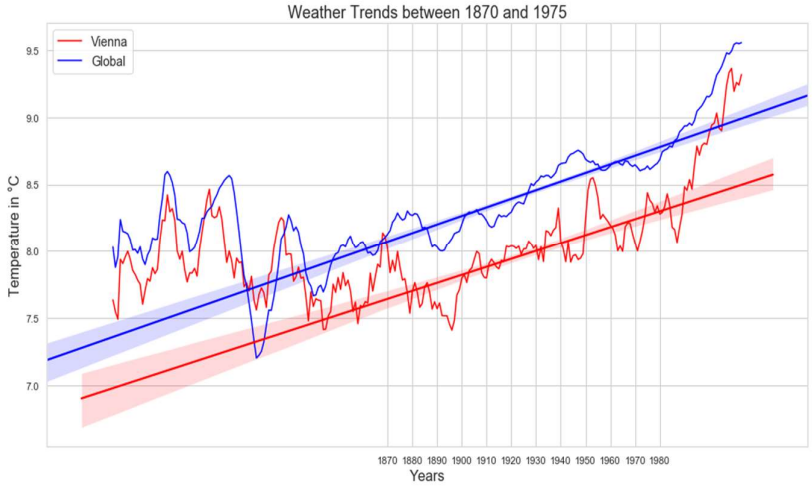


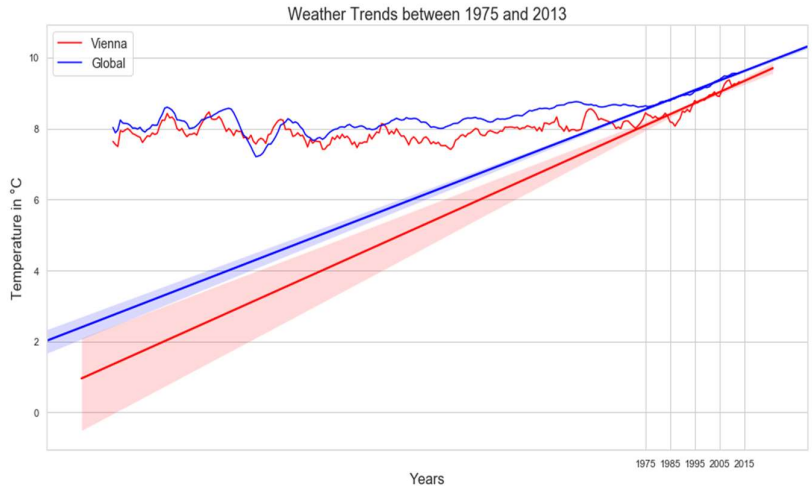
Picture 2

This is certainly not the only reason for the increase in temperature, so the increase in population and, consequently, intensive food production should also be taken into account.

Linear Regression line







Regression Line

On every chart for the regression line, it's clear that the average temperature in Vienna more or less follows the World average temperature.

When we observe timeline for complete data we can say that the trend is growing But, we don't have data before the year 1752 and we can't see what was trends for previous years or centuries

From 1750 until 1840 both temperature trends are decreasing in values

At the end of the 19th century (around the year 1870) started the second industrial revolution and huge use of steam engines and mass production.

From there until the year 1975 is a growing trend of temperature in Vienna and the world, and after 1975 trend is even more increasing.

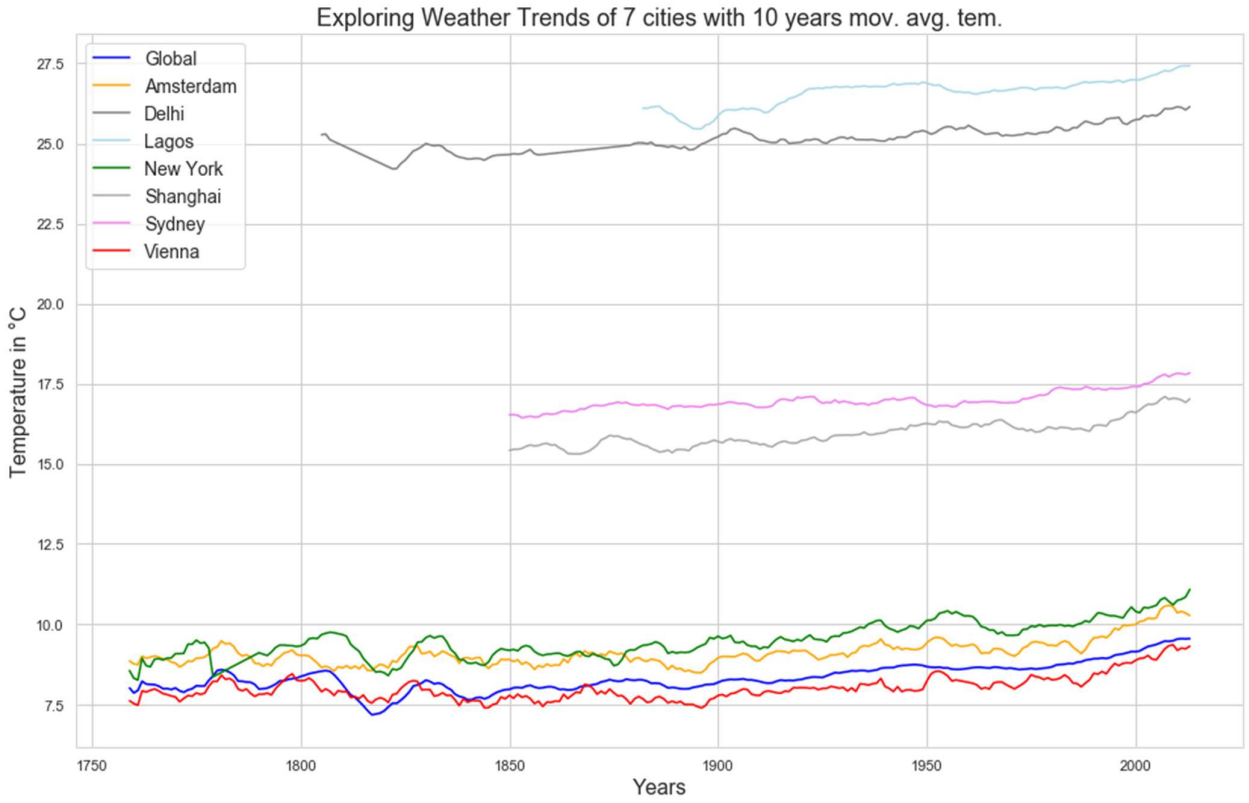
Cities around the World

As we have been seen the linear regression is positive in the last 130 years. If this trend continues, the global temperature and the temperature in Vienna will continue to increase. And what is about other cities in the world?
In this weather trends analysis, 7 cities were included from different parts of the world. The relationship between cities and the global temperature was identified, through comparison of temperatures and trends using seaborn.lineplot() and the .corr () function.

```
df_cities.groupby('city')['avg_temp_global', 'avg_temp_city'].corr()
```

		avg_temp_global	avg_temp_city
city			
Amsterdam	**avg_temp_global**	1.000000	0.516804
	avg_temp_city	0.516804	1.000000
Delhi	**avg_temp_global**	1.000000	0.762654
	avg_temp_city	0.762654	1.000000
Lagos	**avg_temp_global**	1.000000	0.814152
	avg_temp_city	0.814152	1.000000
New York	**avg_temp_global**	1.000000	0.563413
	avg_temp_city	0.563413	1.000000
Shanghai	**avg_temp_global**	1.000000	0.784995
	avg_temp_city	0.784995	1.000000
Sydney	**avg_temp_global**	1.000000	0.754343
	avg_temp_city	0.754343	1.000000
Vienna	**avg_temp_global**	1.000000	0.534342
	avg_temp_city	0.534342	1.000000

When we look at the coefficient of correlation for Vienna it is 0.53, which is a moderate positive correlation.
Other cities with higher average temperatures are more strongly correlated with global temperature, while cities around the global average temperature have a moderate correlation coefficient.



The cities of Lagos and Delhi have very high average temperatures of around 26 ° C, while Sydney and Shangai around 17 ° C. All other cities in my research work are moving around the average world temperature.
It is significant for all cities to have an increase in temperature after 1870.

Final conclusion

The temperature in Vienna and other cities in this research are definitely in an upward trend, what certainly rises a global temperature.

To get a better picture of the movement of temperature on earth, we need to have much more data, to see the earth's natural temperature fluctuations as well.

Probably hundreds or thousands of years before we began with our observation from 1750.

References:

Picture 1: Kornati https://www.croatiaitinerary.com/kornati/
Picture 2: Industrie 4.0 https://www.tritoninnovation.com/industry40