Udacity Data Analyst Nanodegree

EXPLORING WEATHER TRENDS IN RABAT - MOROCCO

PREPARED BY:

FAHD ARHALAI

fahd.arh@gmail.com

(September 8, 2020)

Table of Content

I. Introduction :

Predicting weather trends and analyzing weather data is the current focus of many organizations and companies, however, the aim changes from one to the other.
Extracting valuable information can help industries such as air companies to keep updated on weather conditions and thus avoid surprising situations. Farmers on the other hand can take advantage of weather data analysis to improve the yield and production. Far away from businesses and industries, scientists are trying to analyze weather data to overcome many challenges and take action to solve climate change and global warming.
The focus of this study is to simply analyze Rabat weather data in Morocco, and compare it to global weather data to find similarities and/or differences using the Moving Average technique.
The analysis is done using Jupyter Notebooks and Python.

II. Extracting Data :

The dataset used in this study is provided by the Data Analyst Nanodegree from Udacity. Using the SQL Workspace, we query the database to extract the "**city_data"** as well as the "**global_data"** tables. The corresponding queries are:

```
SELECT * FROM city_data;
SELECT * FROM global_data;
```

For each query, we download the output in CSV format to be able to read it in a Jupyter Notebook.

Note: The following Python libraries are necessary for the analysis: Numpy, Pandas, Matplotlib, and Seaborn.
We start by reading the two CSV files using Pandas and making the first column - which corresponds to the year - as the index of the DataFrame, then, we extract the local weather data of Rabat city:

```
city_data = pd.read_csv('city_data.csv', index_col=0)
global_data = pd.read_csv('global_data.csv', index_col=0)
rabat_data = city_data[city_data['city'] == 'Rabat'].copy()
```

III. Data Exploration and Preparation :

By looking at the line graph (Figure 1); starting from the year 1900, the average global temperature is increasing smoothly without large oscillations. Before the year 1900, it had many fluctuations that seems random to each other. However, the average temperature of Rabat city has the same fluctuations over the years that seems also random with the exception that it started increasing from the year 1900.

Apart from the shape of the line graphs, the average temperature of Rabat is above the global average. Which is reasonable since the weather of Africa tends to be warmer.

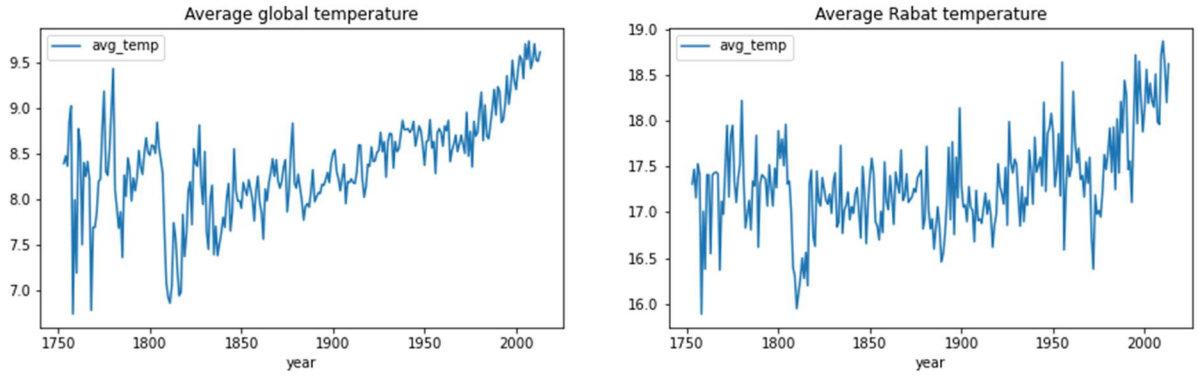

Figure 1. Average global temperature (LHS). Average Rabat temperature (RHS).

Our global temperature data describes the average temperature between the year 1750 and 2015. However, the average temperature of Rabat city is only available between the year 1753 and 2013 (Figure 2).

```
global_data.reset_index().year.agg(['min', 'max', 'count'])
```

```
min      1750
max      2015
count     266
Name: year, dtype: int64
```

```
rabat_data.reset_index().year.agg(['min', 'max', 'count'])
```

```
min      1753
max      2013
count     261
Name: year, dtype: int64
```

Since the data has no missing values (At least in the provided data, since the missing years can be regarded as missing values of average temperature), we only need to deal with the years interval mismatch. For that, we will drop the rows in the global data where the year is missing in the local data. In other words, we will focus our study between the year 1753 and 2013.

```
global_data.drop(index=[1750, 1751, 1752, 2014, 2015], inplace=True)
```

IV. Moving Average :

The moving average is a measure of the average over a period of time. It is used to smooth out the data by removing the changes details and having a more general and clear view on the behavior of the data.

Mathematically speaking, the simple moving average denoted as $\mathbf{P_{SM}}$ of a numerical variable **P** starting from the $\mathbf{(M-(n-1))^{th}}$ element, over a period of length **n,** can be calculated using the following formula:

$$\begin{aligned}\overline{p}_{\mathrm{SM}} &= \frac{p_M + p_{M-1} + \cdots + p_{M-(n-1)}}{n} \\ &= \frac{1}{n}\sum_{i=0}^{n-1} p_{M-i}.\end{aligned}$$

We picked three different intervals (5 years, 10 years, and 20 years) to calculate the moving averages and plot the corresponding line graphs for local and global data.

Pandas library provides an easy way to calculate the moving average using a rolling window of a specific size that moves sequentially through the entire dataset. We use the rolling window to calculate the mean.

```
rabat_data['mva_temp_5'] = rabat_data.iloc[:,2].rolling(window=5).mean()
rabat_data['mva_temp_10'] = rabat_data.iloc[:,2].rolling(window=10).mean()
rabat_data['mva_temp_20'] = rabat_data.iloc[:,2].rolling(window=20).mean()
```

```
global_data['mva_temp_5'] = global_data.iloc[:,0].rolling(window=5).mean()
global_data['mva_temp_10'] = global_data.iloc[:,0].rolling(window=10).mean()
global_data['mva_temp_20'] = global_data.iloc[:,0].rolling(window=20).mean()
```

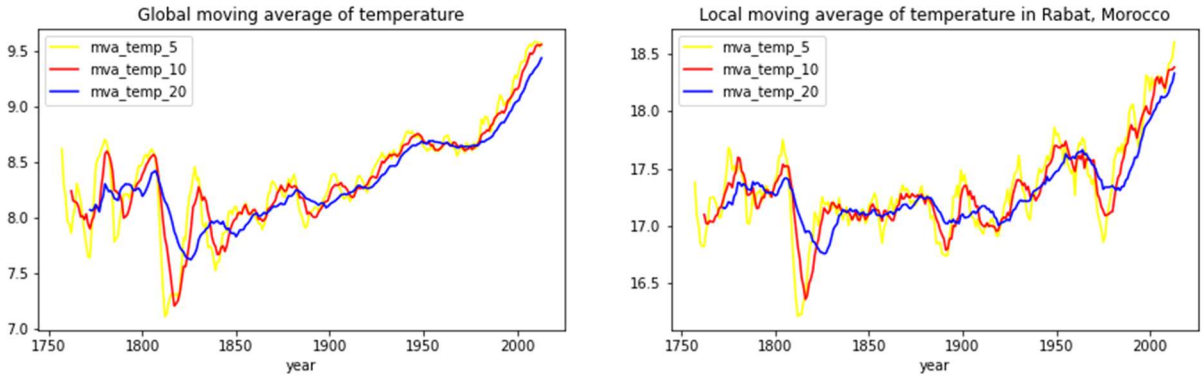

Figure 2. Global (LHS) and Local (RHS) moving average.

The plots for the moving average (Figure 2) shows that a 5 years interval (labeled **mva_temp_5**) is showing too many details while the 20 years interval (labeled **mva_temp_20**) hides some of the details that might be useful to look at when comparing the two datasets. Therefore, a 10 years interval (labeled **mva_temp_10**) is the one we selected for this analysis.

The next step is to join the two datasets to form one dataset containing only the moving average (over 10 years) of the local and global data, as well as the year columns.

```
mva_rabat_data = rabat_data.loc[:,['mva_temp_10']]
mva_global_data = global_data.loc[:,['mva_temp_10']]
joined_data = mva_rabat_data.join(mva_global_data, how='inner',
                                  lsuffix='_rabat', rsuffix='_global')
```

The joined data (Figure 3) contains null values for the first 9 years because of the calculation of the moving average. We drop those rows to keep only the data with non-null values.

year	mva_temp_10_rabat	mva_temp_10_global
1753	NaN	NaN
1754	NaN	NaN
1755	NaN	NaN
1756	NaN	NaN
1757	NaN	NaN
1758	NaN	NaN
1759	NaN	NaN
1760	NaN	NaN
1761	NaN	NaN
1762	17.099	8.239

Figure 3. Joined data.

V. Analysis and Results :

Using the joined data that contains only the moving average over 10 years for local (Rabat, Morocco) and global data, we plot a line graph for the moving average of Rabat city (labeled **mva_temp_10_rabat**) and global moving average (labeled **mba_temp_10_global**).

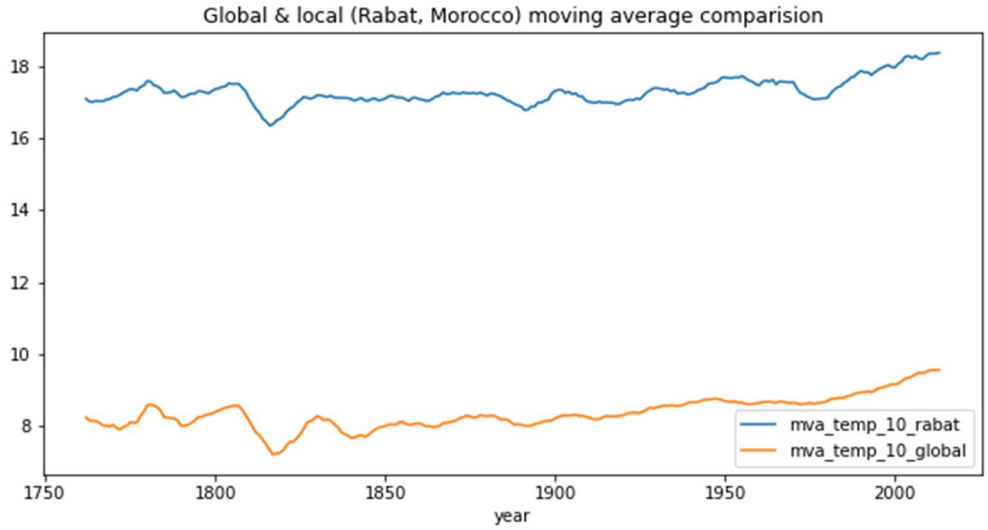

Figure 4. Global & Local moving average over a period of 10 years.

Figure 4 shows a comparison between the two line plots. A first impression leads to the following remarks:

- The moving average of the local temperature sits above the global moving average, which means the local average temperature is warmer than the global average temperature.
- Before the year 1900, the change in the local moving average is relatively the same as the change in the global moving average.
- After the year 1900, the change in the global moving average tends to be smoother, while the change in the local moving average keeps oscillating up and down. This confirms our pre-observation from Figure 1.
- Generally, the average temperature is increasing over time.

To have a more clear and detailed understanding on the difference between local and global temperature, we create a scatter plot to view the correlation between the two variables.

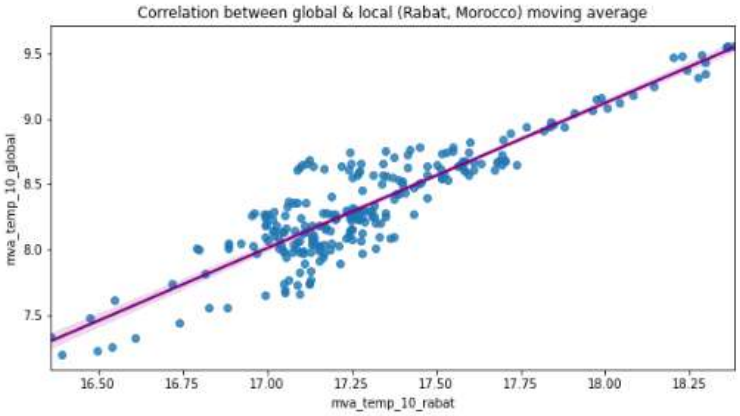

Figure 5. Correlation between global and local moving average.

	value
min	8.43800
mean	8.95529
std	0.19524
max	9.43000

Figure 6. Difference between local and global moving average.

Figure 5 shows the correlation as well as the regression line. Generally, this explains that the local temperature has followed the global temperature trends between the year 1753 and 2013.

Finally, we found it fun to calculate the difference between the local moving average and the global moving average, and find the min, max, mean and standard deviation:

```
mva_diff = (joined_data['mva_temp_10_rabat'] - joined_data['mva_temp_10_global'])
mva_diff.agg(['min', 'mean', 'std', 'max']).to_frame().rename(columns={0: 'value'})
```

The raison is to find out in more details how the local average temperature is behaving compared to the global average temperature. The result (Figure 6) shows that the change in the values is contained in a small interval [8.438, 9.43]. In short, the difference is relatively stable since the standard deviation is small and we can say that approximately:

Rabat Moving Average ≈ Global Moving Average + 8.9

VI. Conclusion :

Rabat temperature tends to be on average 8.9°C warmer than the average global temperature between the year 1753 and 2013. In other words, we found that there is a significant correlation between the local and global temperature and they follow generally the same trends. However, it is clear that the average temperature keeps increasing, which leads to ask many question on what could be the cause of it; Is it the effect of the industrial revolution and the emerging industries? Or an unknown factor that leads to a failure in the equilibrium of nature? It seems like we need more data to explore and analyze, to find any trends and patterns that could explain this phenomena, and maybe to find the next most relevant and necessary action to take as humanity.

REFERENCES

- Data Analyst Nanodegree – Udacity: https://www.udacity.com/course/data-analyst-nanodegree--nd002
- Moving Average – Wikipedia: https://en.wikipedia.org/wiki/Moving_average
- Implementing Moving Average in Python – TowardsDataScience: https://towardsdatascience.com/implementing-moving-averages-in-python-1ad28e636f9d