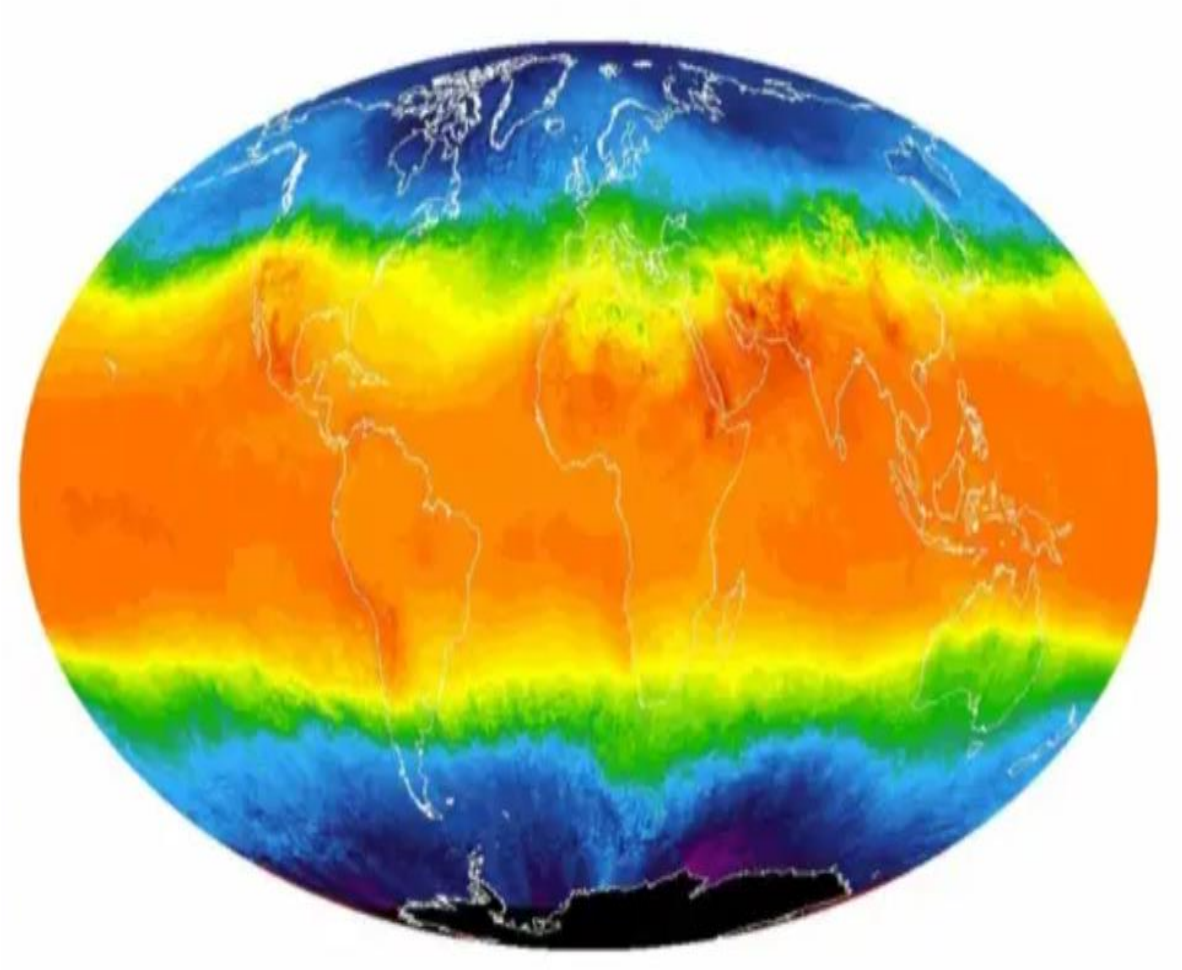


Exploring Weather Trends in Berlin

Ekaterina Kuznetsova | Data Analyst Nanodegree | 06.10.2021

UDACITY

Summary

In this project, we will analyze local and global temperature data and compare the temperature trends of Seattle city to overall global temperature trends.

Content

1. Data Extraction

There are three tables in the given database:

1. city_list - contains a list of cities and countries in the database.
2. city_data - contains the average temperatures for each city by year (ºC).
3. global_data - This contains the average global temperatures by year (ºC).

This SQL commands were used to extract the data from tables.

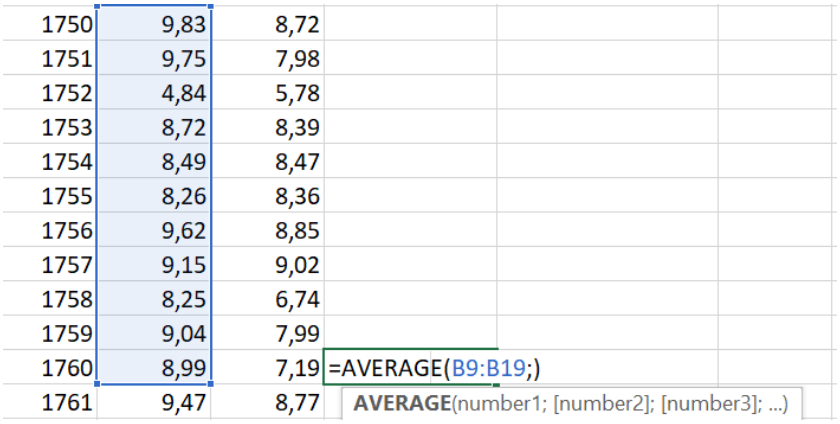
Extracting the average global temperature from the table global_data

```
SELECT * FROM global_data
```

Extracting the average temperatures of the city from the table city_data

```
SELECT year,avg_temp FROM city_data

WHERE city = 'Berlin'
```



2. Open the CSV

I opened CSVs in Jupiter Notebook and below just a part of code. Full code you can find in *Exploring_Weather_Trends_in_Berlin.ipynb*

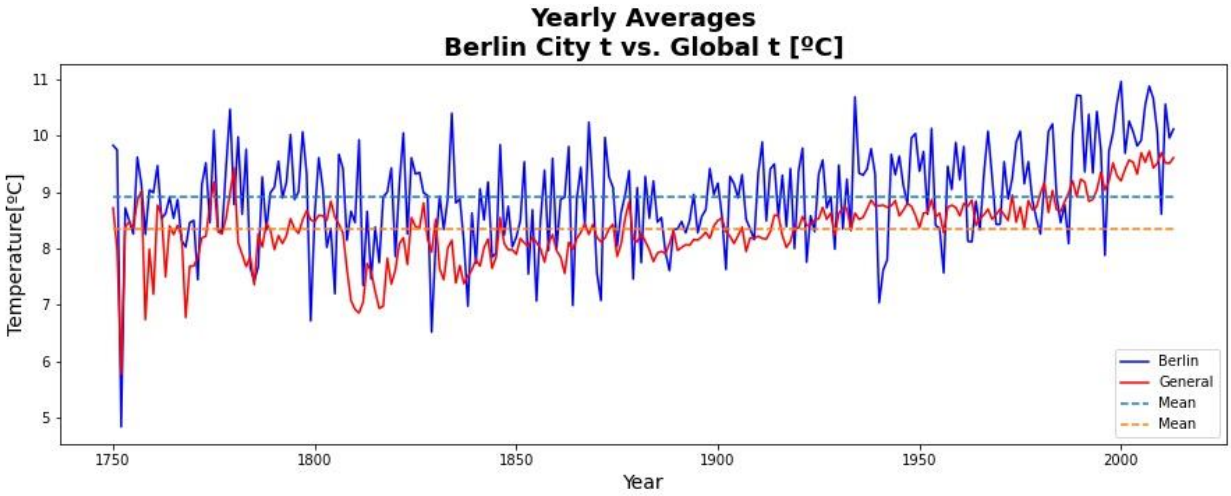
```
# open csv
berlin_data = os.path.abspath('berlin_data.csv')
general_data = os.path.abspath('general_data.csv')
# read csv
berlin = pd.read_csv(berlin_data)
general = pd.read_csv(general_data)
```

For convenience in further work, I merged csv-Files and renam columns in Jupiter Notebook.

3. Create a Line Chart

Line chart that compares Berlin temperatures with the global temperatures.

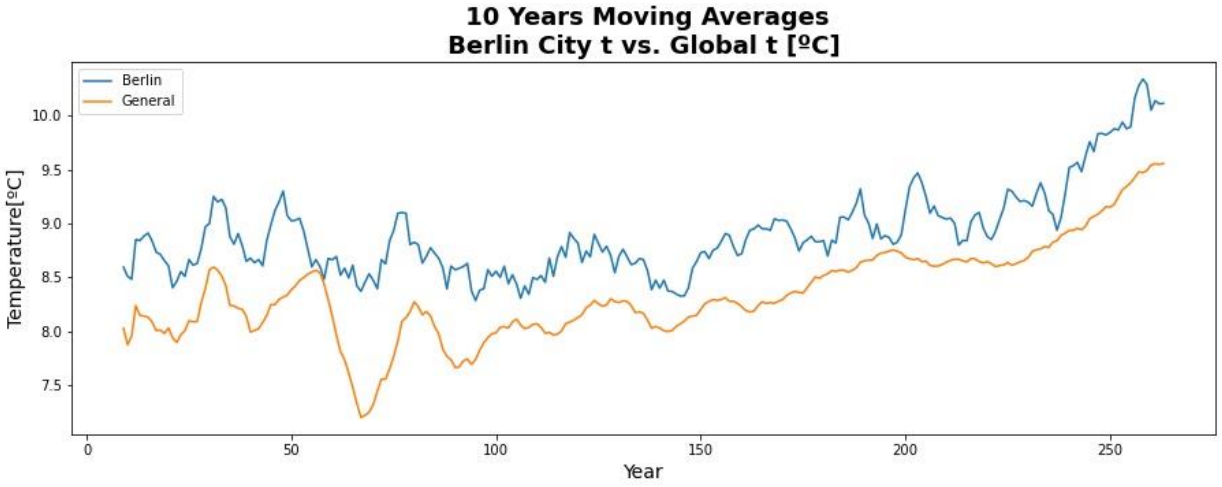
In this case I will plot the moving average and the yearly averages in order to smooth out the lines, making trends more observable (the last concept in the previous lesson goes over how to do this in a spreadsheet). I make it specially for comparing yearly averages with moving averages.



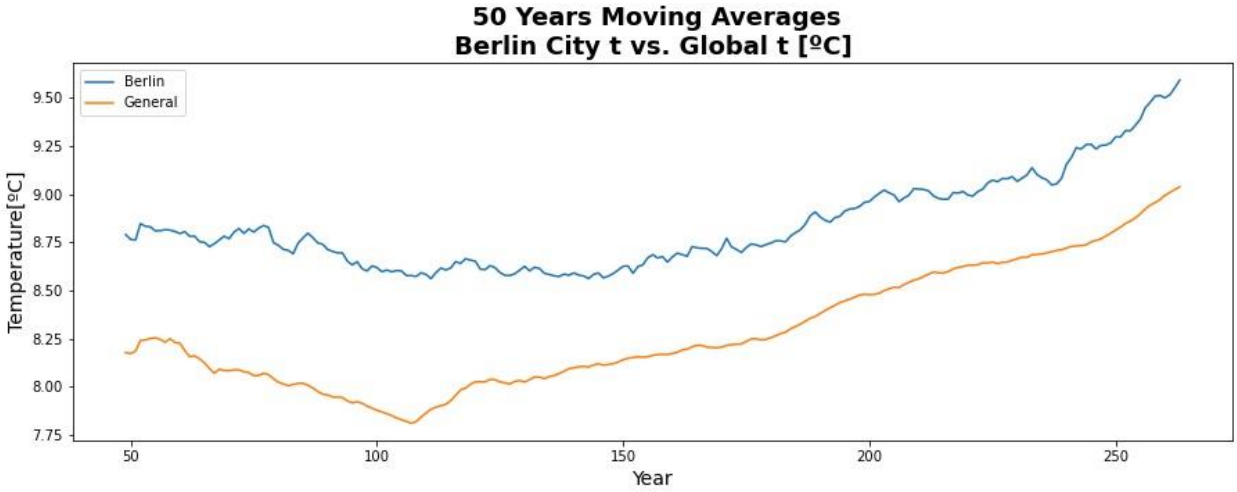
Plot 1: Yearly Averages t in World and in Berlin

10 year moving average is used to smooth out data.

This term makes trends clearer and shows less fluctuations in data then yearly averages.



Plot 2: 10-years moving Average t in World and in Berlin



Plot 3: 50-years moving Average t in World and in Berlin

50-years moving averages make trend even more obvious and linear.

4. Conclusion

Visual analysis let us make 4 clear conclusions:

1. Berlin City Temperature is generally higher than Global Temperature [in ºC].

   Berlin is hotter than world averages. And this difference has been consistent over time, what surprised me. Temperature changes over Time (Based on 10 years moving average) where line chart showing Different City Temperatures against Global Temperature

2. Global Trend - Temperature growing.

   Maximum yearly temperature in Berlin was in year 2010 and made 10,96ºC as we can forecast from moving averages plot that is not a record nowadays. General temperature maximum was in year 2015 and made 9,83ºC. Important to notice that the last entry in the Dataset was made in year 2015 and we can also guess that this temperature record nowadays is new.

3. Berlin and World temperature moderately correlated with each other.

   According to the Formula *of correlation coefficient*:

   result1['Berlin'].corr(result1['General']) we get result = 0.51595

   That means that variables considered **moderately correlated** and is on the boarder a low correlation, because we have 0.5 < |r| < 0.7.

4. Moving Averages in 10 and 50 years let us clearer to see temperature distribution than yearly average.

   50-year running mean (moving average) time series plot of global mean is presented in Plot 3 with the base period 1750–2013 and shows trend even more obvious and linear.

   Temperature series show a steady increase of temperature, what more accused in the long-term (50-years).