Udacity training: **Data analyst for enterprise nanodegree program**

Student name: **Delia Amarandei**

Project name: **Project 1: Explore weather trends**

Steps followed with results:

1. Explore and extract the data from the database

For this first step I used SQL to query the data in order to understand it and extract the needed values.

Check the data:

```
SELECT * FROM city_data LIMIT 10;
SELECT * FROM city_list LIMIT 10;
SELECT * FROM global_data LIMIT 10;

SELECT COUNT(*) FROM city_data;
SELECT COUNT(*) FROM city_list;
SELECT COUNT(*) FROM global_data;
```

I found the nearest city to me by checking the available cities in the country I live in, Germany:

```
SELECT  DISTINCT city
  FROM  city_list
 WHERE  country = 'Germany';
```

--Berlin, Hamburg and Munich out of which I will choose Berlin.

2. Extract the data from the database

I exported the temperature data for the world by year:

```
SELECT  year,
        avg_temp
  FROM  global_data
 ORDER  BY year;
```

and for Berlin, also by year:

```
SELECT  year,
        city,
        avg_temp
```

```
  FROM  city_data
 WHERE  city = 'Berlin'
 ORDER  BY year;
```

Another way to download the data together, since we have it already aggregated it by year in both tables:

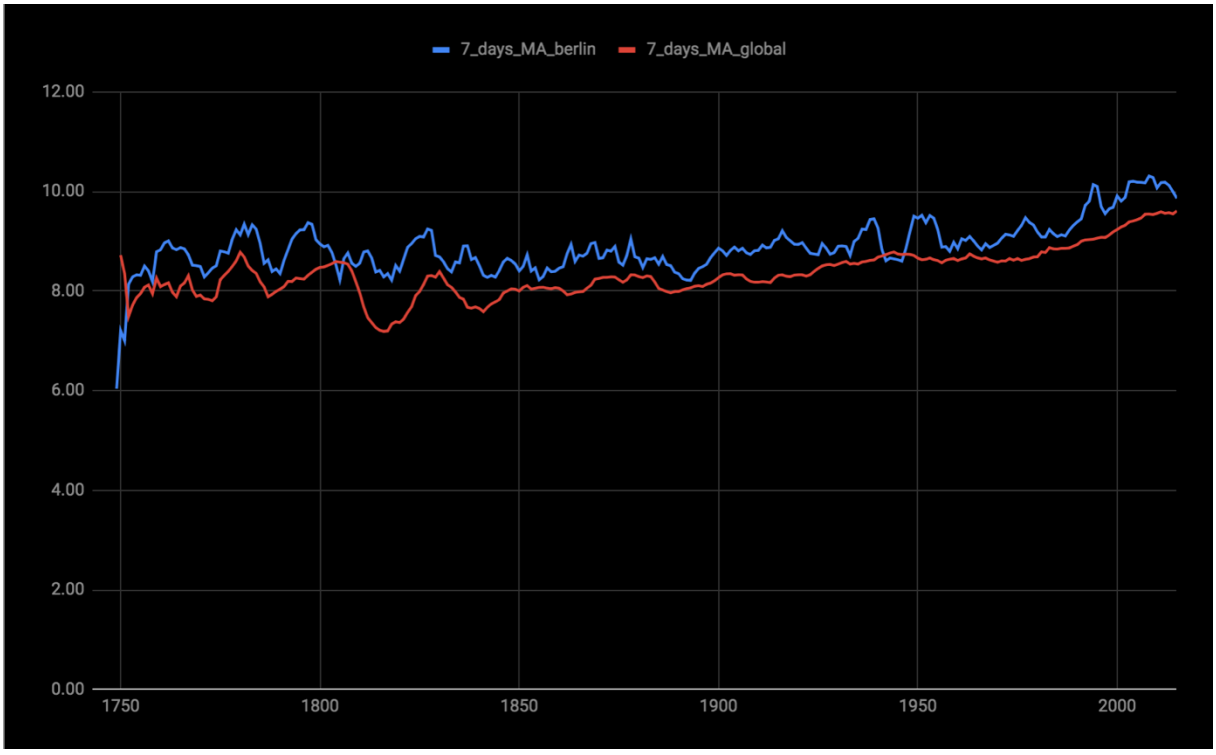
```
SELECT  COALESCE(global.year, city.year) AS year,
        city.avg_temp                    AS berlin_avg_temp,
        global.avg_temp                  AS global_avg_temp
  FROM  global_data global
  FULL  OUTER JOIN (
        SELECT year,
               avg_temp
          FROM city_data
         WHERE city = 'Berlin') city
        ON global.year = city.year
 ORDER  BY COALESCE(global.year, city.year);
```

3. Open up the CSV & create a line chart

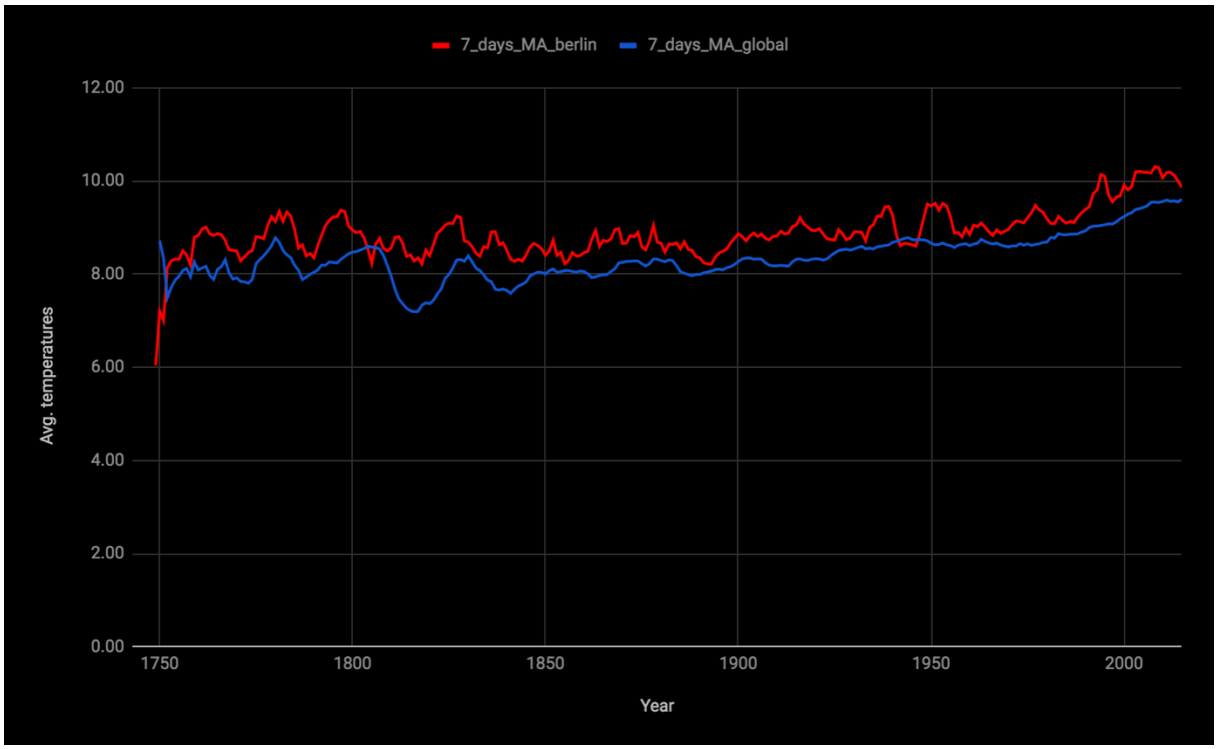
I created a new Google sheet where I loaded my csv file by clicking on File – Import – csv file.

I calculated the moving averages for both columns, berlin_avg_temp and global_avg_temp and named them 7_days_MA_berlin and 7_days_MA_global respectively. I used the average of the last 7 days, including the day I was calculating it for.

Further I selected all data and inserted a chart which I edited to only display the moving averages values.



I chose a black background in order to better focus on the trend lines. As the default colors were blue for 7_days_MA_berlin and red for 7_days_MA_global and as I observed directly that Berlin averages are on average warmer than the global values, I decided to change the colors so that whoever checks this graph immediately associates red with higher temperatures and understands the graph quickly.



4. Observations

- On average, Berlin is a hotter city than the global average, the difference has been somehow consistent over time given Berlin's fluctuation from year to year.
- For both averages we observe a steady growing trend towards the most recent years with temperatures getting higher and higher consistently over the last hundred years (starting somewhere around 1900).
- We observe that the changes in the 7-days moving averages affected both averages in the same way, for example we see a peak between the years 1780 and 1781 in both.
- A visible outlier in the data happened between the years 1942 and 1946 where there is a significant drop in the Berlin 7-days moving averages, when checking the data we will see that in 1838 the average for Berlin was 6.98, a lower temperature than the rest, possibly explained by some meteorological factors.