Exploring Weather Trends

Project report

1 Data extraction

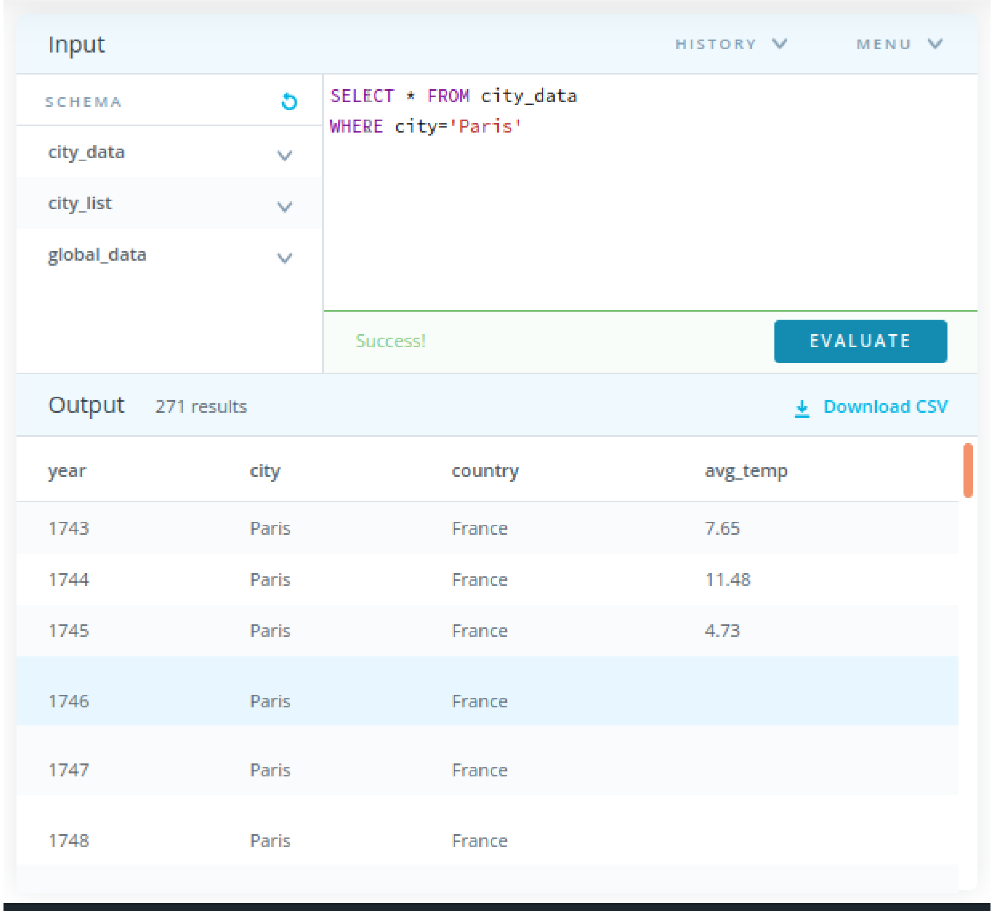
To extract data from the Udacity database I used different SQL queries in the platform workplace.

First, I had to find a closest city to where I live (Toulouse, France), so I ran the following query:

```
SELECT * FROM city_list
   WHERE country='France'
```

The only French city returned by this query was Paris, so I will further contrast global weather trends with weather trends in the capital of France.

Next, the following query was used to extract temperatures for Paris:



```
SELECT * FROM city_data
WHERE city='Paris'
```

Success!

271 results

year	city	country	avg_temp
1743	Paris	France	7.65
1744	Paris	France	11.48
1745	Paris	France	4.73
1746	Paris	France	
1747	Paris	France	
1748	Paris	France	

The previsualisation shows that some data is missing.

Finally, to extract global temperatures, I ran the following query:

```
SELECT * FROM global_data
```

Both data sets (global data and Paris data) were exported as .csv via Udacity platform.

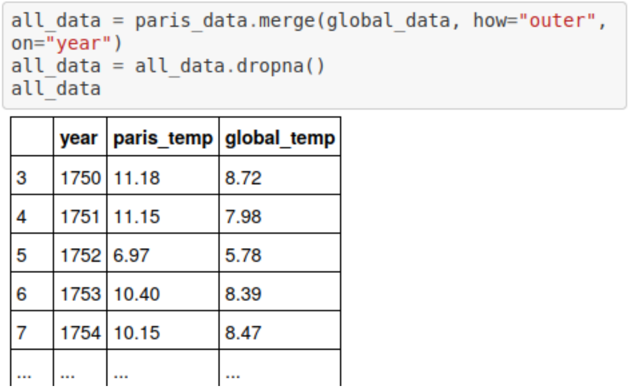
2 Working with CSV

I used python (pandas and seaborn libraries) to work with .csv file.

The whole project can be found on GitHub (https://github.com/elis-wind/Udacity/blob/main/Data-analyst/Weather-Trends/Project.ipynb). It describes all the necessary steps of data wrangling.

3 Line chart

The final data set which contains yearly temperatures for both Paris and global data tables looks loke following:



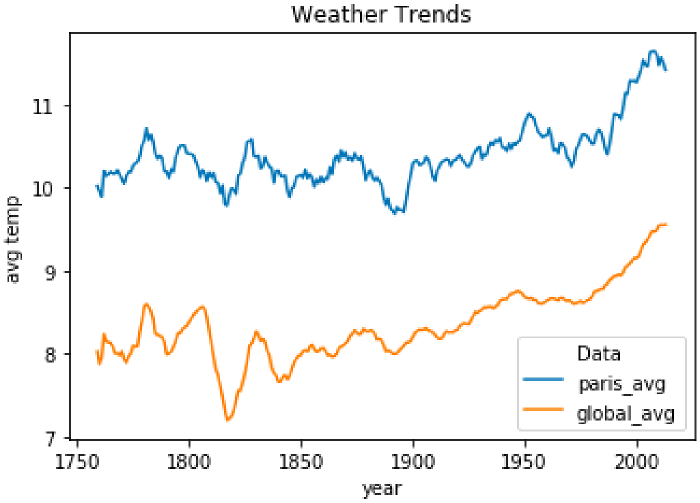
```
all_data = paris_data.merge(global_data, how="outer",
on="year")
all_data = all_data.dropna()
all_data
```

	year	paris_temp	global_temp
3	1750	11.18	8.72
4	1751	11.15	7.98
5	1752	6.97	5.78
6	1753	10.40	8.39
7	1754	10.15	8.47
...	...	...	...

I used the 10 years window to calculate the average temperature for both columns. The example of code line for Paris data is shown below:

```
all_data["paris_avg"] = all_data["paris_temp"].rolling(10).mean()
```

Based on new columns (`paris_avg` and `global_avg`) I plotted a line plot with seaborn:



4 Observations

Using the Weather Trends visualization, I can make the following observations:

- Overall Paris temperatures are about 2 degrees higher than global temperatures;
- Both Paris and global temperatures have increased over time, especially over the past 50 years;
- The coldest period can be found at the beginning of the 19th century in global temperatures;
- The temperature trends are quite similar: the same patterns can be found in both weather lines. Two local minima are presented in Paris data; one of them corresponds to the only local minimum in global data;
- Since patterns are similar, I can hypothesize that knowing global temperatures one can predict the temperatures in Paris for a given period. Further investigations (with correlation coefficients and linear regression model) are needed to prove this hypothesis.