Exploring Weather Trends

A Udacity Project Submission by Timothy Quan

Contents

Outline

Tools Used

- Python (pandas and matplotlib)
- SQL – to extract the data from the provided workspace

```sql
-- just city data from bangkok:
-----------------------------------------------------------
SELECT *
FROM city_data cd
WHERE city='Bangkok';
-----------------------------------------------------------
-- just global_data:
SELECT * FROM global_data
-----------------------------------------------------------
```

- Excel – for inspection of the .csv files

Moving Average Calculation Methodology

Using Pandas, this project implements rolling averages by combining the dataframe.rolling() and .mean() functions.

By default, and as implemented in this project, the .rolling() function uses the preceding values, eg., with a 5 year window .rolling(window=5).mean() in 1754 will use data from 1750-1754 and so on:

year	avg_temp	Rolling Average (5 Year)
1750	8.72	
1751	7.98	
1752	5.78	
1753	8.39	
1754	8.47	7.868
1755	8.36	7.796

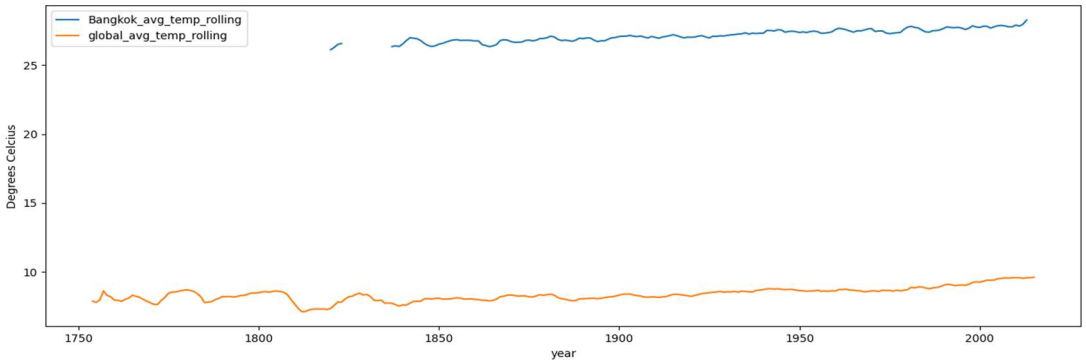
Key Considerations for Visualization

Foremost, the rubric specifies a line chart using rolling averages.

A 5-year rolling average appears to reveal the highest level of detail desirable considering the time span. 3 years shows too much detail to the point that there are no obvious trends; 10 years shows too little.

Python/pandas/matplotlib was chosen with the intent of practicing a somewhat new skill vs formula and chart generation in excel being already known.

Line Chart



Observations

- Bangkok is significantly hotter than the global average, generally around 20 degrees more.
- While Bangkok matches the overall global trend, at this time scale it is difficult to see if there are any other matching trends.
- The overall global and local trend is an increase in temperature at around 2-3 degrees over all time.
- There is no weather data for Bangkok before 1816, and missing from 1826-1833.

External Documentation

This project can be found on github; including explore_weather_trends.py (code to generate the line graph) and extract_all_data.sql (the sql queries to extract the necessary data for csvs).

https://github.com/timothyquan/Explore_Weather_Trends