

DS2002 Data Project 1 Reflection Report

For our project, we looked at weather trends for capital cities around the world, comparing the predicted average to the historical average for a given month and day. The dataset we used was from Kaggle and contained weather data from capital cities around the world, along with their longitude, latitude, temperature min/max, date, and other information. For our API, we used Open-Meteo to get weather data based on the city's latitude and longitude.

One of the key challenges we encountered during implementation was integrating the right data from the Open-Meteo API. We wanted to make sure that we were getting the right data and if there was an error, we would throw the correct exception to handle it correctly. Another tricky aspect was working with date formats, extracting components like month and day, and making sure they were in the right format. Another challenge was deciding the best way to present our analysis or visualize our analysis. We ended up going with a bar graph over other alternatives such as a line graph or chart, as it presented the differences between the historical and predicted averages in a clear and intuitive way.

Some aspects that were easier than expected were converting between the different file formats and loading the data. Using the pandas built-in libraries to convert from CSV to JSON using the `.to_json` and `.to_csv` functions made it really simple to return the results in the output the user wanted.

One aspect that was more difficult than expected was handling location consistency. Although retrieving latitude and longitude from the dataset was straightforward, ensuring consistent matches between user input (city name) and the dataset entries required making them case insensitive accounting for error.

A utility like this can be useful for future data projects that involve merging historical data with external APIs. It brings together different pieces—loading data, working with APIs, doing some analysis, and then storing and visualizing the results—in a way that mirrors real-world workflows. Whether it's for climate research, forecasting tools, or just exploring patterns over time, having a reusable pipeline like this is something we could build for more complex projects down the road.