Antonis Daniil
Report for Code

Part 1:

I started getting the data from the API url from three different locations for 7 days. I did a for loop for 7 days for the three different locations and I added range, year, month as variables for the user to choose which days and months they want. Due to the url address, I had to add an “IF” to get the 7 days from a particular month, I could do an extra IF ‘statement’ if the days were in a different month and/or year. We added an extra column to each table to have the three different locations and then we Concat the three tables. After that, we applied some data cleaning to get the number of decimals that we need and update the “created” column to be readable as datetime from MySQL. I decided to keep the null values in the 'visibility' column because it showcases that we don't have data for it one week ahead.

Part 2:

After creating the table by setting the correct variables in MySQL, I imported the csv file into the table. I had an issue with importing the data due to MySQL security where I find it hard to bypass it, after a long google search I found a solution which was to add two ‘\’ in the path so the path ended up looking like this ‘C:\\ProgramData\\MySQL\\MySQL Server 8.0\\Uploads\\all_data.csv’. Furthermore, to check my results for the 2nd part of the 3rd task, I run a query in SQL which you can find in the same SQL file as the table.

Part 3:

After loading the downloaded data from the “all_data.csv” file in Jupyter notebook I made the API for the first task and with groupby and max, I found the requested data.

Afterward, I moved to task two (2) where at the start I had to create a new column to take the day out of the ‘created’ column in a date format. Then I made a new column ‘ranked’ where I saved the results out of my dense_rank function so I would get the top 3 results for each day (because we had multiple temperatures at the same time on some days). After that with a mean function and a group by I took the average of the top 3 counted results. An issue that I had was passing the data that I calculated in a dictionary because it can’t take datetime values so I had to change them to a ‘string’.

Finally, I had to make an API that took the n parameter from a user. I had an issue because the ‘n’ value was read as a string so after researching I found my solution which was to make ‘n’ an integer.

Finally, please note that any further guidelines can be shown as comments (#) in the code.