SELECT project_submission
FROM Jonathan_Wright
WHERE
        project_focus = "Database Design";

Aim:

The aim of this project was to use MySQL to locally host a custom database. Using data scraping and cleaning techniques to prepare a set of data from the NASA website. The end goal of the project was to be able to write complex queries for this database.

Method:

I downloaded data for the years 2018, 2019 and 2020 regarding the aerosol concentration in the air as measured by the ATLAS16 satellite. The data is split into 2°x2° squares. Each of these "granules" then has the following measurements (followed by column title in database) recorded on a month-by month basis:

- Latitude - LAT
- Longitude - LON
- global_asr_obs – ASR_OBS
- global_aerosol_fraction – AERO_FRAC
- global_asr - ASR
- global_cloud_aerosol_obs – CLOUD_AERO_OBS
- global_cloud_fraction – CLOUD_FRAC
- global_column_od – COLUMN_OD
- Global_ground_detection - GRND_DETECTION FLOAT
- Tcod_obs - TCOD_OBS

Explanations for all of these variables can be found in the PDF in the linked GitHub Repo. The database also has two additional columns: Granule_ID and MON which comprise the composite primary key for each of the tables. The data are then grouped by year and the resulting database consists of roughly 1e5 entries. Included in this entry are several screenshots of the SQL code run to create the database, the tables and return certain entries. The entirety of the code can also be found on the repository.

Conclusion

This project was a success. I created a functional database containing interesting and related data.

Extensions

To enrich my database I would add the ancillary data from each granule this is data that details the method of measurement and other constants for each granule.