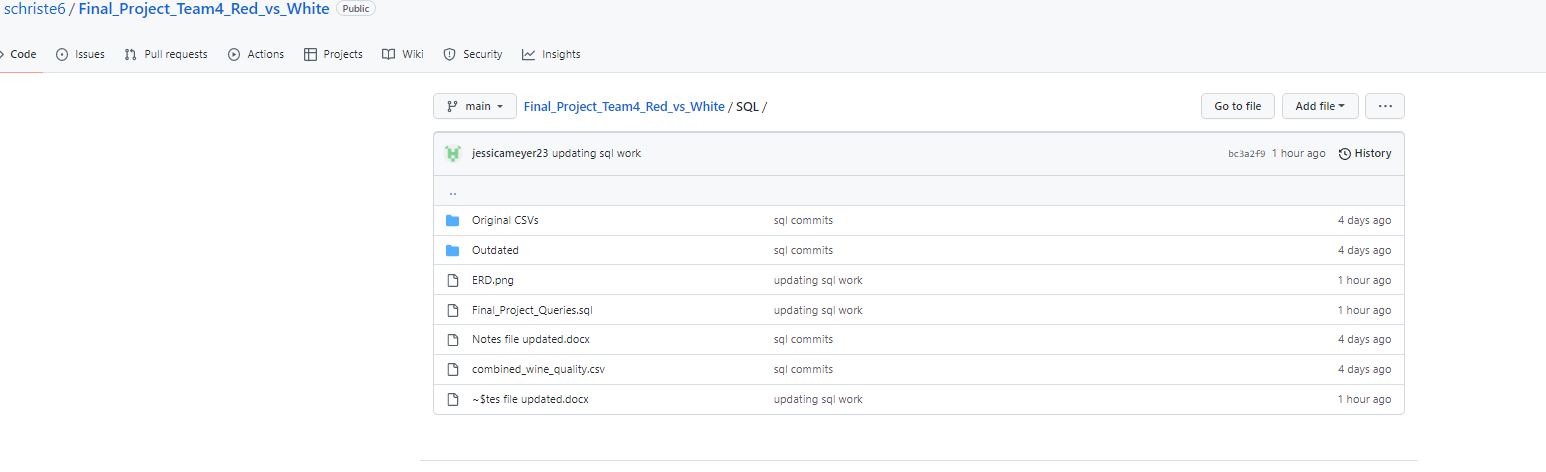
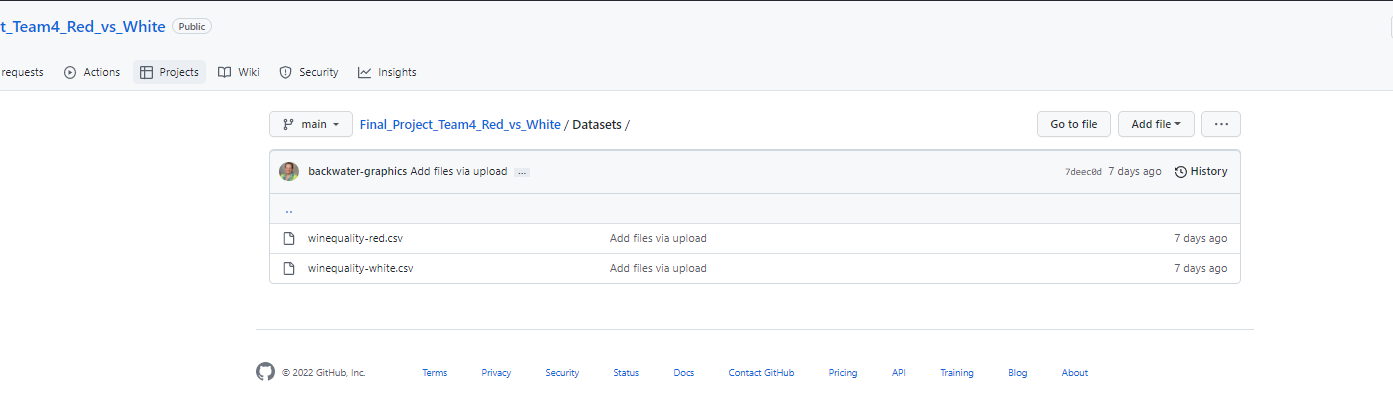
**Author: Jessica Meyer, Database Role for Team 4. “Red\_vs\_White:**

The purpose of this document is to lay out the SQL steps that are also listed in the “Final\_Project\_Queries\_sql” file but this word document will add a little more description and will also list out where the connection string files are located, etc.

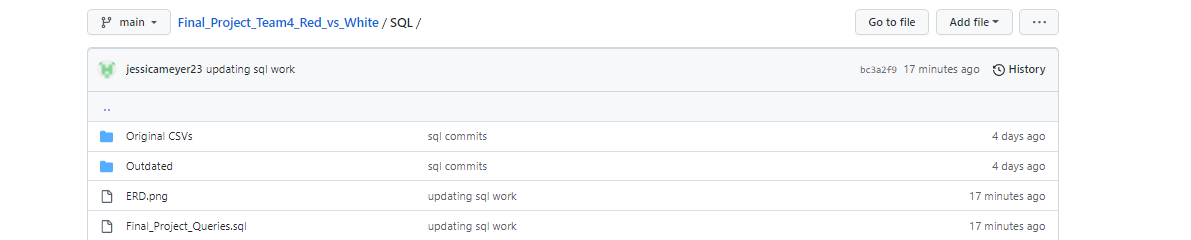
See below for the location of the actual SQL queries—“Final\_Project\_Queries.sql”



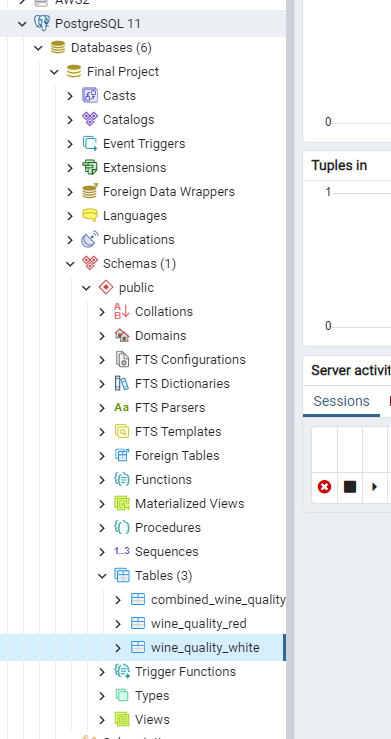
1. First I obtained the original CSV files which are located in the “Datasets” folder in our github location as shown below:



1. I also saved these files to the SQL folder that I created and included them in the “Original CSV” file in the location shown below:

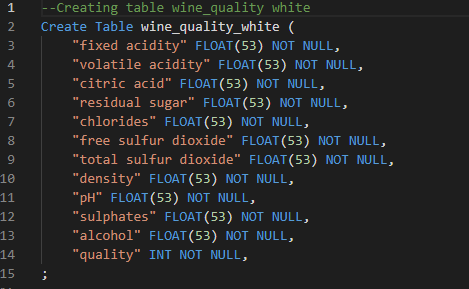


1. Database Creation:
   1. I then created a database in Postgres called Final Project

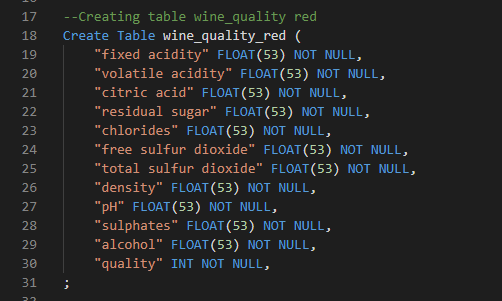


1. Table Creation: The next steps were to create two tables for the original CSVs. The following is the code used and is also located in the “Final\_Project\_Queries.sql” file as noted previously.

4a. Creation of wine\_quality\_white table

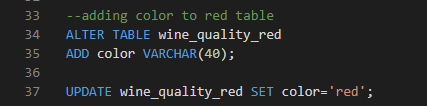


4b. Creation of wine\_quality\_red table

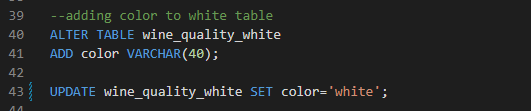


1. The next step was to add a color column to the tables.

5a. These are the steps to add the color column to the red table and then set the color column to “red”:

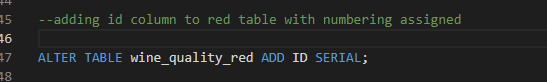


5b. These are the steps to add the color column to the white table and then set the color column to “white”:



1. The next steps were to add an id column to the red and white wine tables.

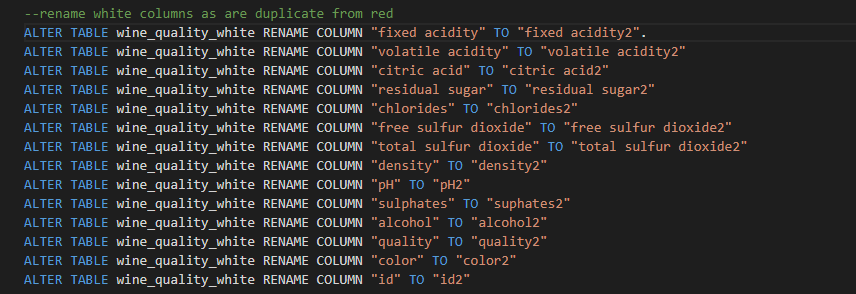
6a. Adding an id column with numbering assigned to the red table:



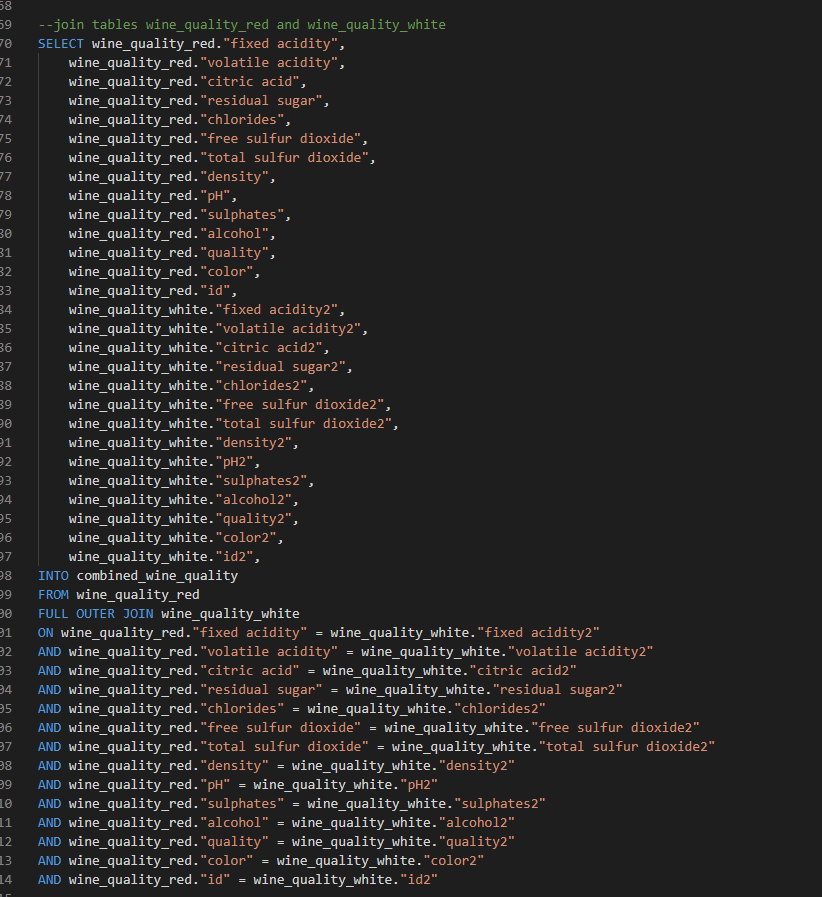
6b. Adding an id column with numbering assigned to the white table:



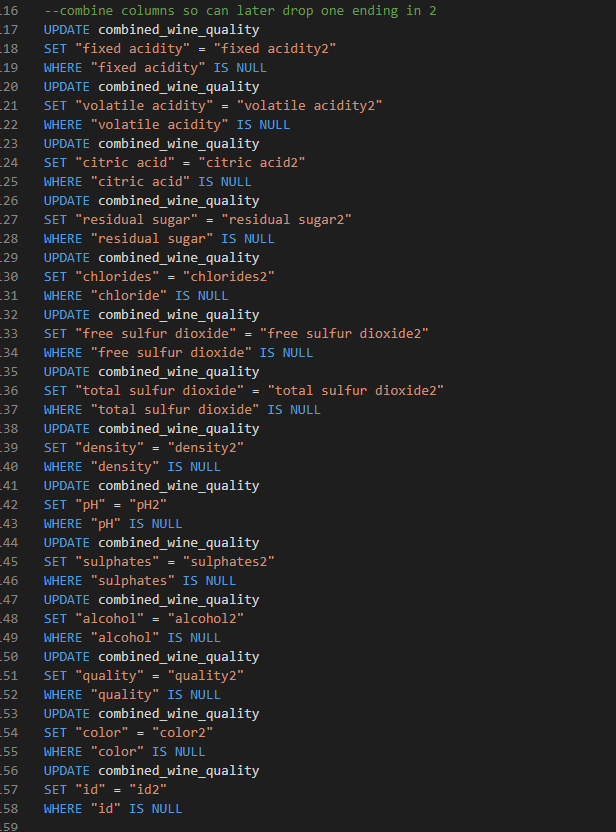
1. The next step was to rename the columns in the white table as they are duplicates from the red (had to do this step to enable us to combine the tables):



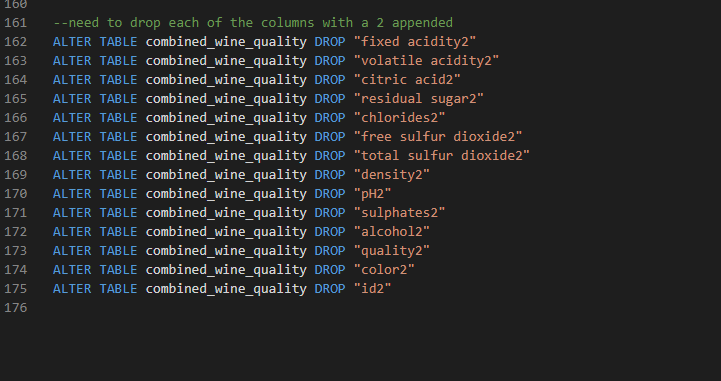
1. The next step was to join the two tables:



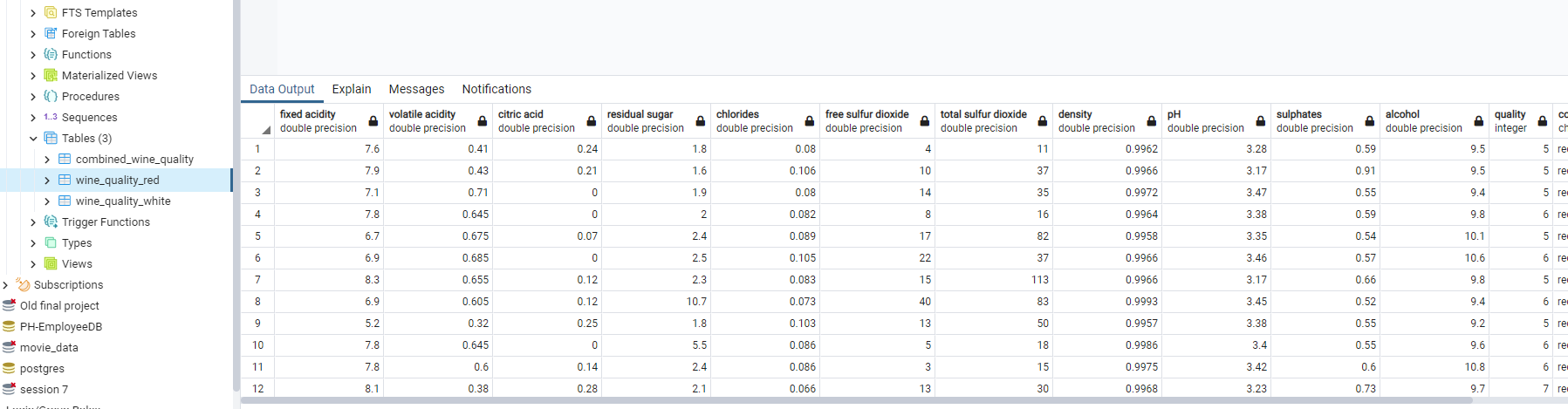
1. The next step was to combine the columns so that we could later drop the one ending in 2:

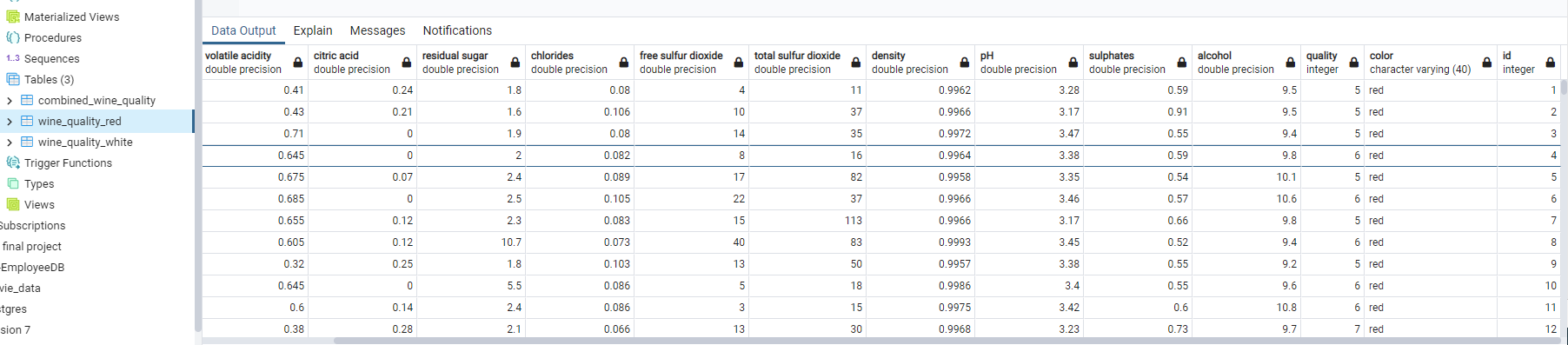


1. The next step was to drop the columns with the 2 appended:

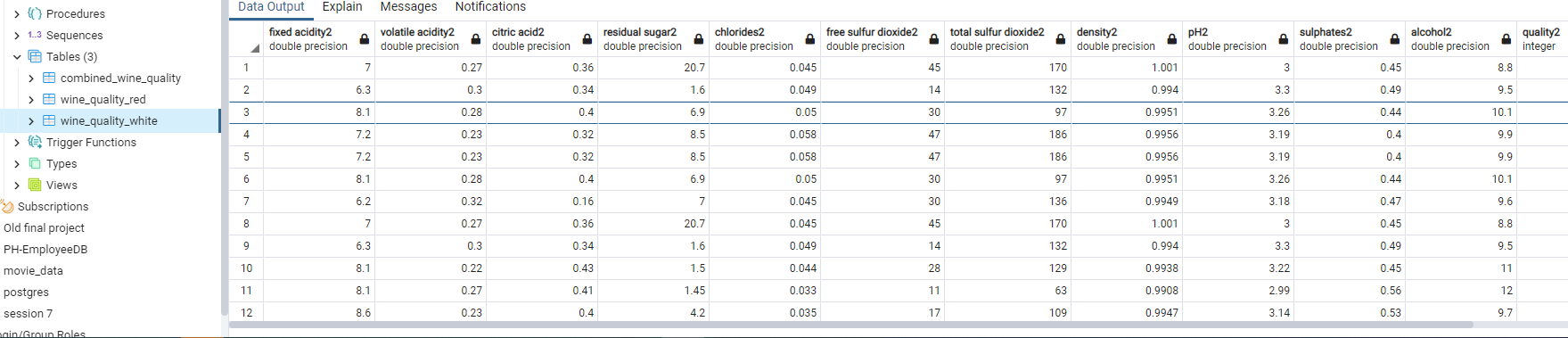


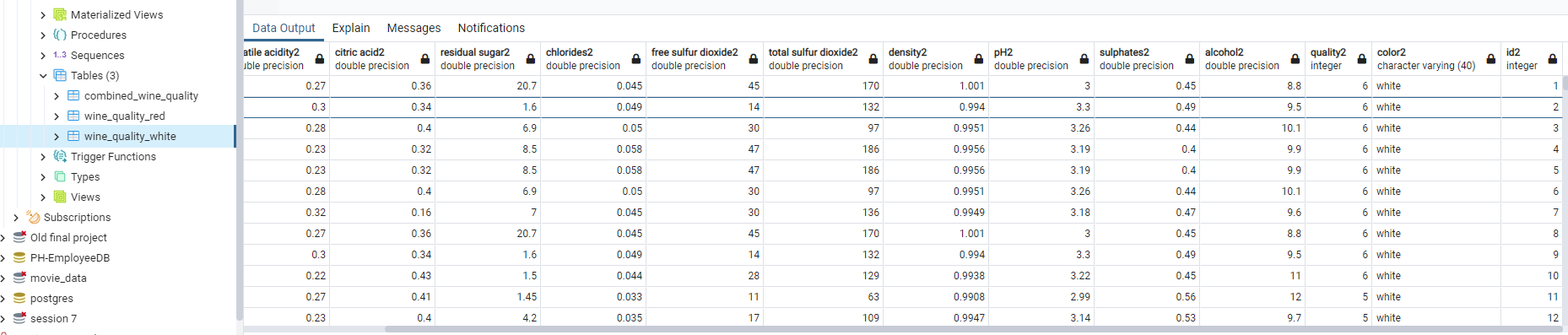
10a. The following screenshot shows the wine\_quality\_red table and columns. Two pictures are included to show all the columns.





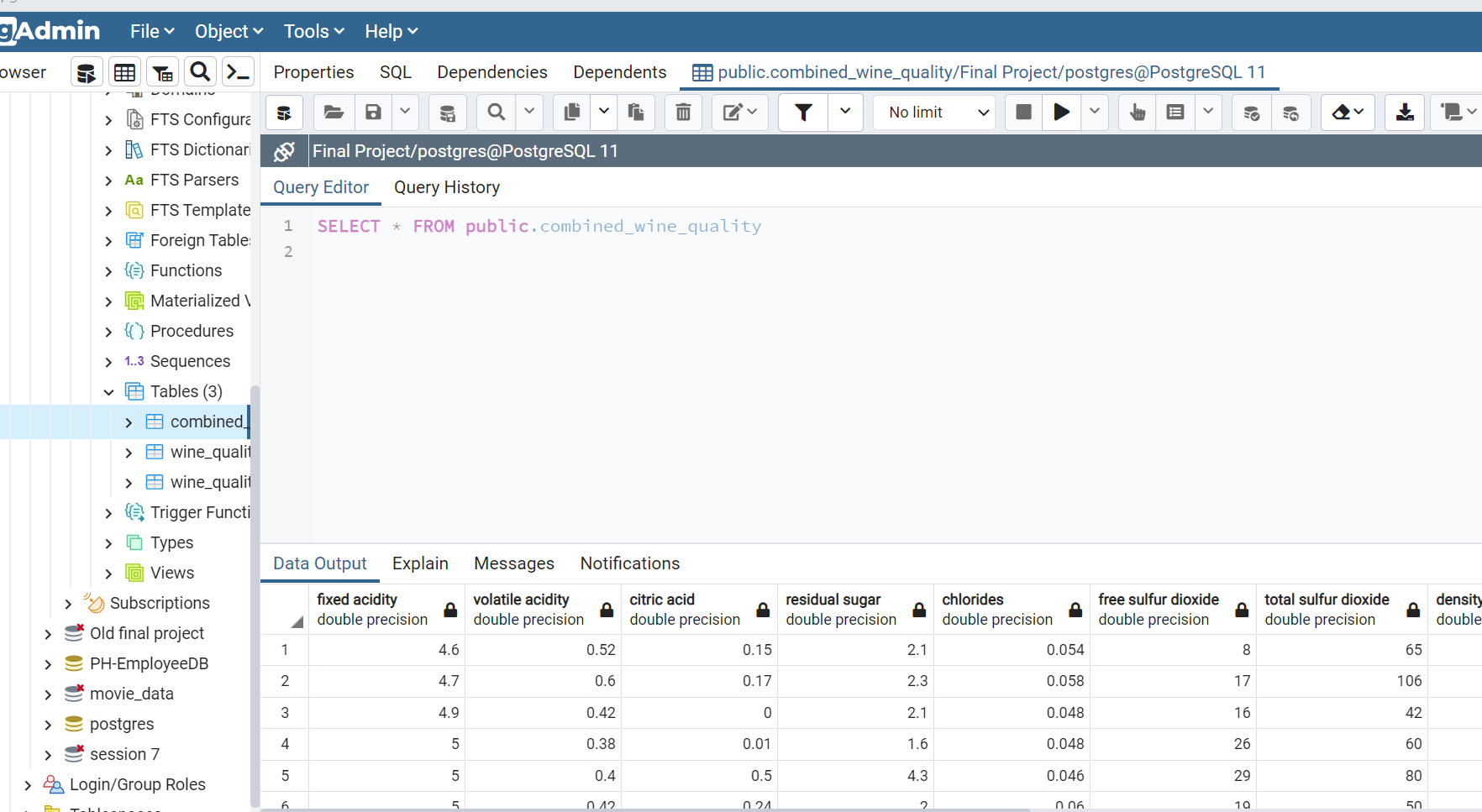
10b. The following screenshot shows the wine\_quality\_white table and columns. Two pictures are included to show all the columns.



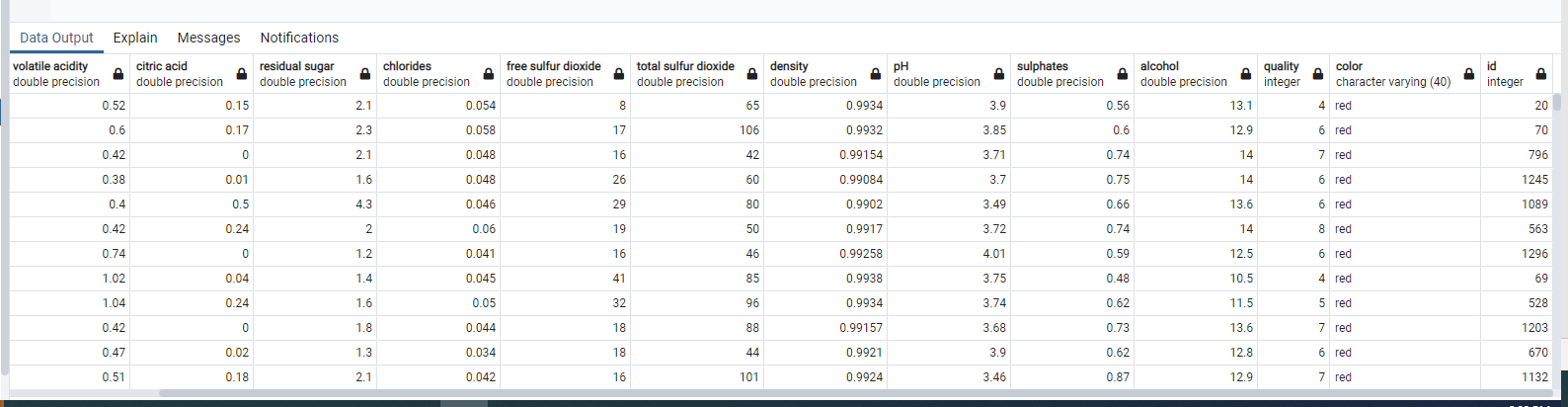


1. The end result is the combined file “combined\_wine\_quality” table with the following columns noted (see two pictures as there were too many columns to fit).

11a. Picture-showing the first columns

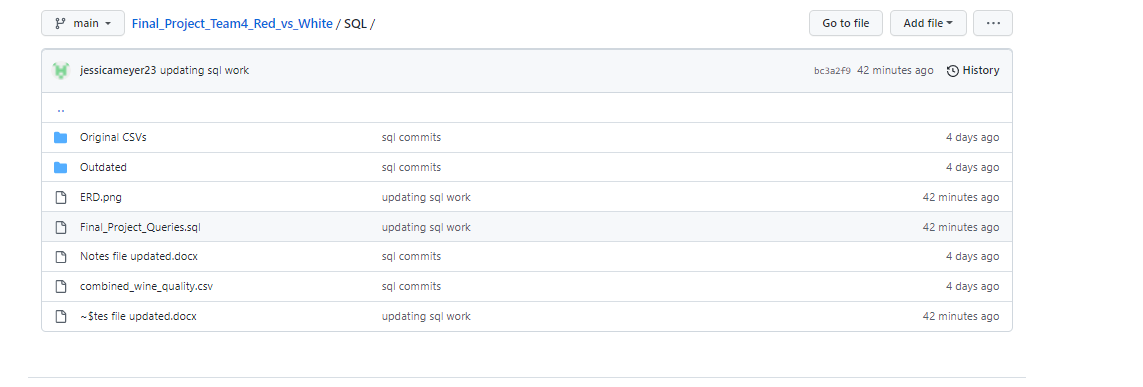


11b. Second picture (showing the last columns, up to ID):



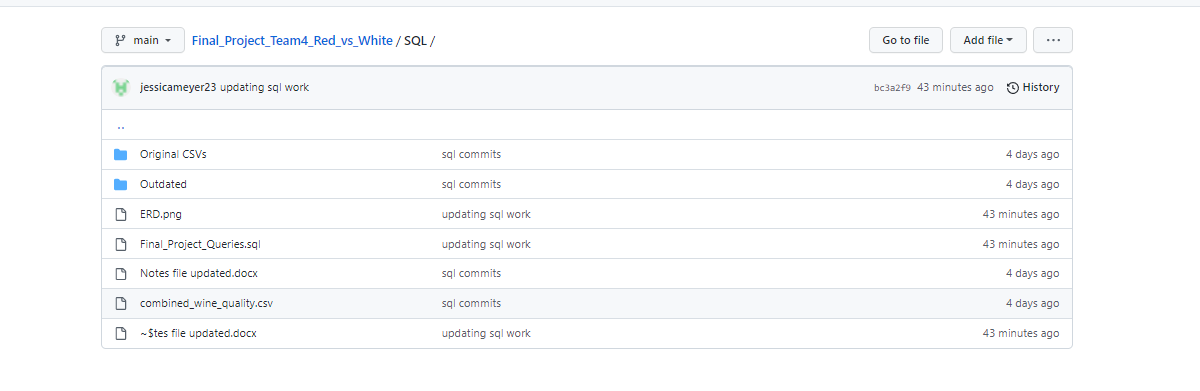
1. I then exported the final table to a combined file, which is called: “combined\_wine\_quality.csv”

It is located in the SQL folder, see below for the location:

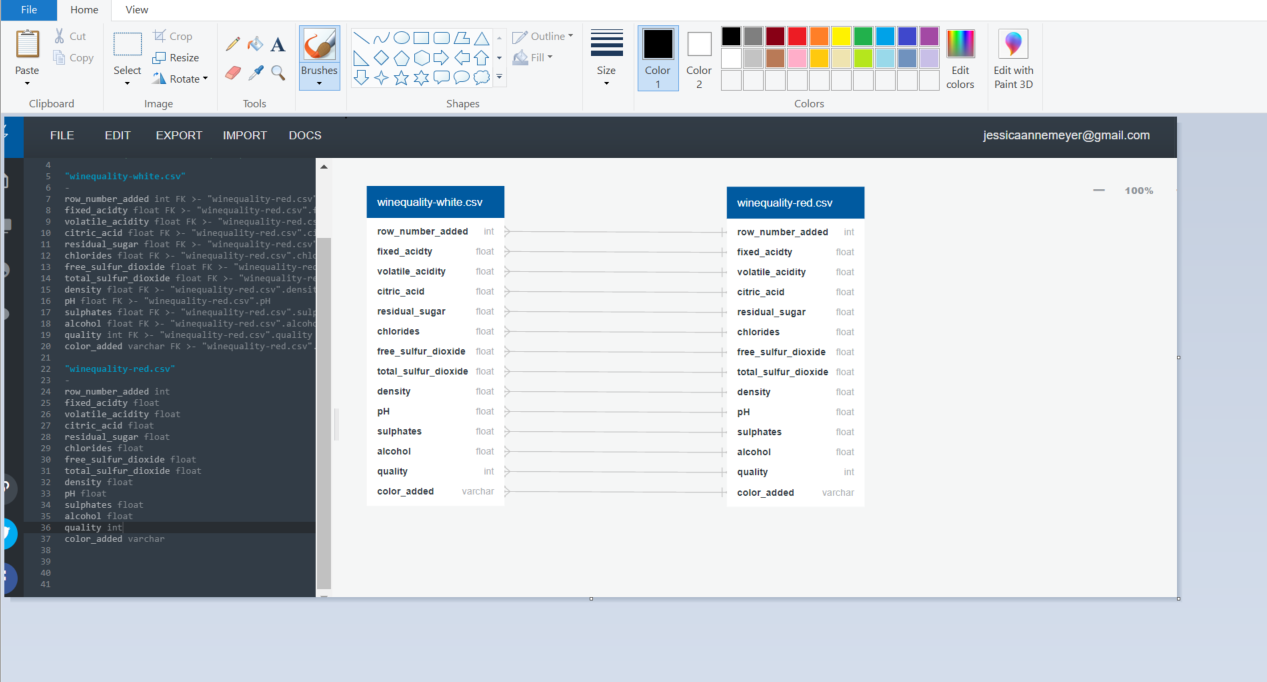


1. I also updated the ERD, which is located in the following location. I have also included a picture of the final ERD below:

13a. Picture of location of ERD file

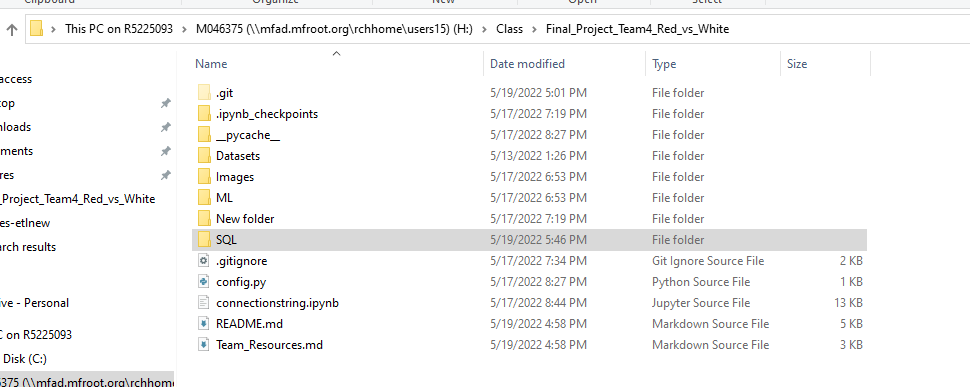


13b. Picture of actual final ERD:

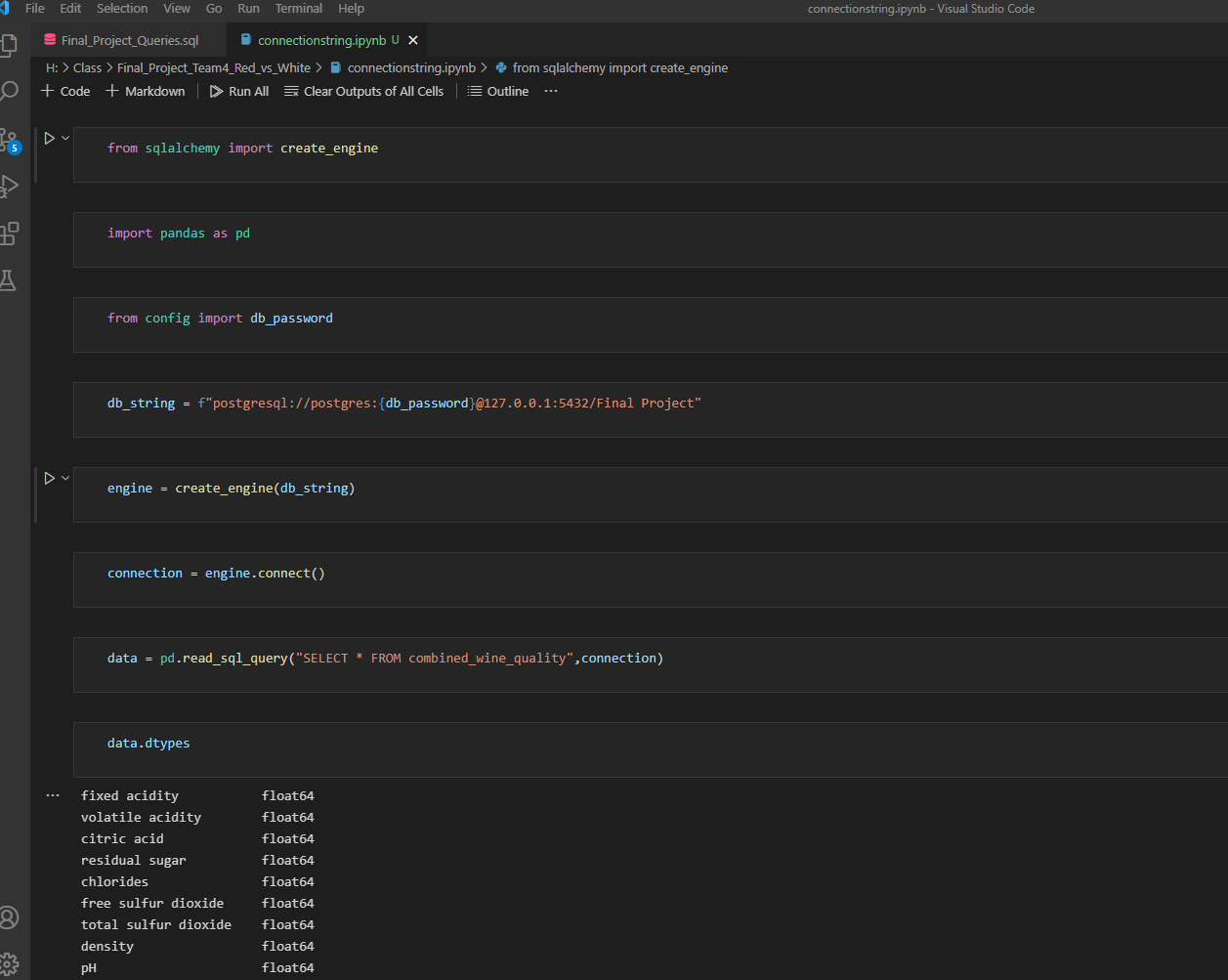


1. The next steps were to create a connection string. For these steps, I need to creat a git ignore and config.py file in order to hold my password so that it would not be updated to the team github folder.

See picture of my class folder, where the config.py file is located (as it is not in the team github folder):



1. The next steps were to create the connection string and then to test it. The proof source file is the “connectionstring.ipynb” file that I created in Jupyter notebook. Here is a picture of the code and also the proof of the connection string.



This second picture validates that the connection string worked, as it is pulling the data in from the database, and you can see it has 6,497 rows and 14 columns.

