Major Occupation in States with Corona Virus Cases

I started with gathering data from BLS on the total employment, by category, of every state. The Covid-19 data was obtained from the daily log of new infections and deaths provided by the World Health Organization. I imported the employment data from the csv I obtained and imported it into a dataframe named emp\_data. I checked the datatypes of the columns to verify whether or not I needed to clean any of the data. Empty data was notated with two asterisks (\*\*) and since nonexistent data on employment wouldn’t be helpful, I removed that data from the dataframe. Next, I had to remove the commas from the numbers to convert the column to the float data type for future calculations if needed. The final cleanup of the emp\_data dataframe consisted of dropping unneeded columns and renaming the columns to prep them for inserting into tables later. After cleaning up the emp\_data dataframe I could divide it into two separate dataframes based on the job categories major and detailed. The major job category provides the industry and the detailed provides the distinct job fields in that industry. Both the detailed and major dataframes were narrowed down to the occupations with the largest amount of people employed.

The covid data was brought in from a csv and read into a dataframe also. The cleanup of this data was more straightforward since I only had to rename the columns for future querying and extracting the data related to US states. I created a world covid dataframe in case I wanted to use that data in the future. The major, detailed, and US covid data were all loaded into tables in postgres where they could be queried.