Assignment 5

1. Explanation of the source code

For the code to run properly, I needed to call all my imports and configure the spark code so it could run with the spark command. Afterwards I began to read each of the json files and displaying them with use of the show() function. Following this, I joined the tables and show that table. Lastly, I used the last table to count the number of tweets from the states and display that table.

2)

I had trouble figuring out how to join the two sets but I was able to use some online resources to figure out how to fix it. I also had some issues displaying the tables until I realized that I needed to use show().

2.1) Since we plan to use Dataframe in Spark, it is easy to type in “DF.show()” to visualize the table in the terminal. Please do so and take a screenshot of the output in the terminal. screenshot “

2.2) Got the correct answer.

