The first step I took was to import the data by using tabs as the delimiter. Then, I made a dictionary corresponding the column names of the .tsv file that I was interested in with the meaning of the code according to the codebook provided by the Substance Abuse and Mental Health Services Organization. I used the keys of the dictionary to boolean index the DataFrame so I could get a new DataFrame with the columns I was interested in. I then used the pd.rename() function to rename the columns so that I could see inside the DataFrame what each column name really meant. I passed in the dictionary I made earlier. Then, I filtered the age to make sure that I was only getting the observations inside the age range of 18-25. What followed was the most time intensive step thus far: I had to replace the values in the DataFrame based on their meaning according to the codebook. I had to do this for every column, and there were 47. I then realized that I made a mistake with entering in the age that subjects first tried marijuana, and had to go through a series of steps to correct this error. There was an issue with copying the series over, so I had to use a for loop to get the values in the new DataFrame. This created an issue of repeat indices, so I had to use another for loop to put the values in the right indices and then delete the repeat indices. After all this, I was still getting errors and realized that I could simply change the values in the dictionary I entered from before and restart the kernel. Once I finished fixing all the meanings for the columns, I inspected the numerical data for outliers that would definitely be bad data. All of the outliers that I found made sense in the context of the problem, and so there was no need to really eliminate any of the outliers. I did some inspection on the hallucinogen data and found many outliers, but all the outliers made sense in the context of the problem: There are very few people who do hallucinogens more than 100 days per year, but those people do exist and are worth investigating for our analysis. Finally, I wrote the cleaned data to a new .csv file.