Progress Report

As I am trying to extract the data in python from the BRFFS csv files, I am having trouble extracting the data due to the large data size and column headers, which are grouped and aren't individually coded or differentiated. I didn't realize this when I was looking for datasets online. So, I came to the realization that I am working with an aggregated dataset, which is causing me trouble in extracting the data. I am assuming I will run into the same issue with the other csv file I have obtained since its population based as well. What I am trying to gather from both data sets is to see what behavioral risk factors lead to a chronic disease but I am realizing that it's harder than I initially thought due to the file I am utilizing. I also initially proposed to do apply a logistic regression model for my data but I am realizing that a linear regression model is better suited since I am working with a population dataset. If I can seem to make these files work then I would have to use new datafiles and I would update the professor if I do so but I am trying my best to make the original csv files work for my project objectives and I will also reach out to the professor for guidance on this matter. I am currently extracting the data through the API on a population level of U.S health rankings. It show the start year and the end year that the data was gathered from and I am trying to figure out how to build a linear regression model for my project.