

#### Analysis:

1. The HCUP provides us with the count of the population in each of the regions participating in the HCUP program. I have decided to utilize this data to find the percentage of population suffering from a disease in each region. This can be achieved by using features such as gender, age, area/ division against a disease (e.g. Heart ailments). From this we will be able to analyze which community suffers the most because of this disease.
2. A linear regression based on the diagnoses types, age, gender, income and to predict the LOS (Length of stay) at the hospital. This would help us to analyze the correlation of these features with LOS.
3. A binary classification model to predict the addition to alcohol, diabetes, depression and obesity based on features such as diagnoses type, age, gender and region. This would be helpful as these diseases are interconnected and effect a lot of people in today's world. I would like to come forward with a successful model to predict whether a person suffers from any of these issues.

#### Limitation:

- 1 years data and hence will have an adverse effect on the accuracy of my model.