# Healthcare Data analysis

The following steps are carried out before doing the healthcare data analytical tasks

1. Data Acqusition

2. Data preprocessing and Exploratory Data analysis

3. Missing value imputation

4. Outlier Analysis and handling outliers

5. Univariate analysis

6. Bivariate analysis

Questions:-

1. Question-1 :-age category of people who frequent the hospital and has the maximum expenditure

The age group 0-1 has the maximum expenditure and maximum hospital visits

1. Question-2 diagnosis-related group that has maximum hospitalization and expenditure

The diagnosis group, 640 has the maximum hospitalization and expenditure

1. Question-3 needs to analyze if the race of the patient is related to the hospitalization costs

Race-1 has the maximum patients and therefore maximum cost

1. Question-4 analyze the severity of the hosp costs by age and gender for the proper allocation of resources.

The age group , 0-1 and gender male has the most hospital cost

1. Question-5 Check if the length of stay can be predicted from age, gender, and race.

Ho - LOS is independent of age, gender and race

H1 - LOS is dependent on age, gender and race

Linear regression model , with predictor variables as age, gender and race can only explain 0.01 variance in regression coefficient for predicting LOS. Also p-Value of model is higher than 0.05.

The cor() gives Very minimal correlation between LOS and AGE , FEMALE & RACE. This shows that although some correlation between predictors and response variable is present, but cannot be used for accurately predicting LOS

1. Question-6 agency wants to find the variable that mainly affects hospital costs.

AGE, FEMALE, LOS, APRDRG are the variables which mainly affects hospital costs. Linear regression model can predict hosp costs. No mulitcollinearity is present in the model, variance of regression coefficient is not inflated due the correlation between predictor variables.