**DATASET INFORMATION FOR STAT 7020 PROJECT**

1. Name of dataset:

Students

1. Description of dataset:

One hundred Five observations were randomly sampled from the High School to help the teachers to make a prediction of what **makes a brilliant student** whether he can apply for a graduate school, whether abroad or not and to help those who fails to apply for a graduate school with advice in job searching.

Id=Student ID.

Class = Students’s class with levels A and B

Gender=Student's gender, with levels female and male

Race=Student's race, with levels African American, Asian, Hispanic, Indian, and white.

GPA = Students’s final percentage

Algebra = Algebra score,

Calculus = Calculus Score

Statistics = Statistics score

Probability =Probability score

Y = final decision with levels one fails to apply for the graduate school, who may apply again or search jobs in the future (0) and inland (1) or abroad (2).

1. A link to where you downloaded the dataset(s), it must be publicly available (or speak with me): *\*You may choose to merge multiple datasets. If you choose to do this, please link all datasets.*

https://www.kaggle.com/datasets/erqizhou/students-data-analysis

1. Number of observations:

105

1. List of quantitative variables you plan to use including range of possible values for each variable:

GPA, Algebra, Calculus, Statistics, Probability are the quantitative variables I plan to use with the range 50 to 100.

1. List of categorical variables you plan to use including possible values for each variable:

Class – A and B

Gender-female

Male

Race-Indian-1

African American-2

Asian-3

Hispanic-4

White-5

Y – 0 =one fails to apply for the graduate school, who may apply again or search jobs in the future

1 = success, inland

2 =success, abroad