Heavy Alcohol Consumption Amongst Student Population using Machine Learning

This topic was selected because we needed a dataset that we could use to make possible predictions to an outcome using machine learning models. Meanwhile, the dataset selected is also considered simple and easy to read and majorly contains numerical values. However, this dataset was chosen because of its size and ability to generate potential results which allows our model to predict outcomes on alcohol consumption amongst student population.

Description of Datasource

This dataset was culled from the UCI machine learning website which shows 2 different schools in portugal and shows students drinking consumption. This group will use different machine learning algorithms to make predictions pertaining to whether a student may be prone to alcoholism in the future.

Questions to Answer

Using the sum of the Workday Alcohol Consumption (Dalc) and Weekend Alcohol Consumption (Walc), we will generate a Alcohol Consumption column, where if values fall below a certain number (to be determined), they will be deemed low risk alcohol consumers while if they fall above that number, they will be deemed high risk alcohol consumer, based on several factors, such as sex, age, parent education, parent job, number of class failures, etc