CMSC 320 - Introduction to Data Science

Final project

I Cheng Chan

Xinyao Zhang

Data Science

This course would involve data analyzing from cvs file and sometimes data from a website. The data could have over thousands of entities (row) and multiples of attributes (column). The users should pay close attention on useful analysis such as standard distribution, mean, center and other types of analysis which would introduce later in this document.

Before analyzing data, manipulating data is often needed so that unnecessary data could be eliminated from analyzing progress.

There are multiple ways to manipulate data before analysis such as handling missing data, tidying data, etc.

Visualizing data is also an important feature of data analyzing. Different kinds of graph could provide different information to the read, as well as avoiding unimportant information regarding what reader is looking for.

DATA RESOURCES:

https://www.kaggle.com/uciml/student-alcohol-consumption#student-mat.csv

This is a data set that is related to a survey conducted between 2 school's mathematics course students. The entity of the data would contain some personal background information. Our aim is to answer the question regarding whether a student's habits and personal background would affect their performance in their studies.