ML hw 1

Michael Fosco

1)

ID Age GPA Days\_missed

count 1000.000000 771.000000 779.000000 808.000000

mean 500.500000 16.996109 2.988447 18.011139

std 288.819436 1.458067 0.818249 9.629371

min 1.000000 15.000000 2.000000 2.000000

25% 250.750000 16.000000 2.000000 9.000000

50% 500.500000 17.000000 3.000000 18.000000

75% 750.250000 18.000000 4.000000 27.000000

max 1000.000000 19.000000 4.000000 34.000000

Note the median is the 50%

ID 0

First\_name 0

Last\_name 0

State 116

Gender 226

Age 229

GPA 221

Days\_missed 192

Graduated 0

dtype: int64

First\_name mode: Amy

Last\_name mode: Ross

State mode: Texas

Gender mode: Female

Age mode: 15.0

GPA mode: 2.0

Days\_missed mode: 6.0

Graduated mode: Yes

For part C, instead of just taking the means of the values (conditional or nonconditional) it should be better to condition on even more so that when we take means we have a more homogenous group. To improve the prediction I conditioned on whether or not someone graduated, their gender, and their state.

2)