COMP3590
Programming for Artificial Intelligence

Class 9

In this class we will practice manipulating Pandas. To get started, create a Jupyter workbook and call it Class 9. The solutions to the following questions are based on Pandas concepts that we have seen in the lectures. If you find that your solution is not based on what was discussed then you might want to reconsider your approach.

Question 3, marked with a *, forms part of the assessment. Please show your work to the supervisor for credit.

We will be using http://raptor.kent.ac.uk/~ds756/Data/heart.csv.

Exercises:

1. Using the dataframe idiomatically, compute the sum of each feature. Note, this only makes sense for some features. Skip over the ones for which sum is not defined; use the Pandas mechanism for exclusion. The solution should be 1 line.
2. Compute the sum of the sums of the features. Your solution should be 1 line.
3. * The mode is a measure of centre of a set (as are mean and median). It is the element with the highest frequency.

   mode ([4, 4, 1, 5, 5, 5, 7]) = 5 # There are more instances of 5

   Implement a function that computes the mode of a column. Name it ComputeMode. It will be invoked as ComputeMode (your_data_frame["column name"]). You will have to do something to the column before you call your function to ensure that it runs in linear time.

Demonstrate correctness by using it to compute the mode of “Age”.

4. Compute the rate of heart disease by age.

5. Create a new column, called “AgeNorm,” that is the age column divided by the max age.