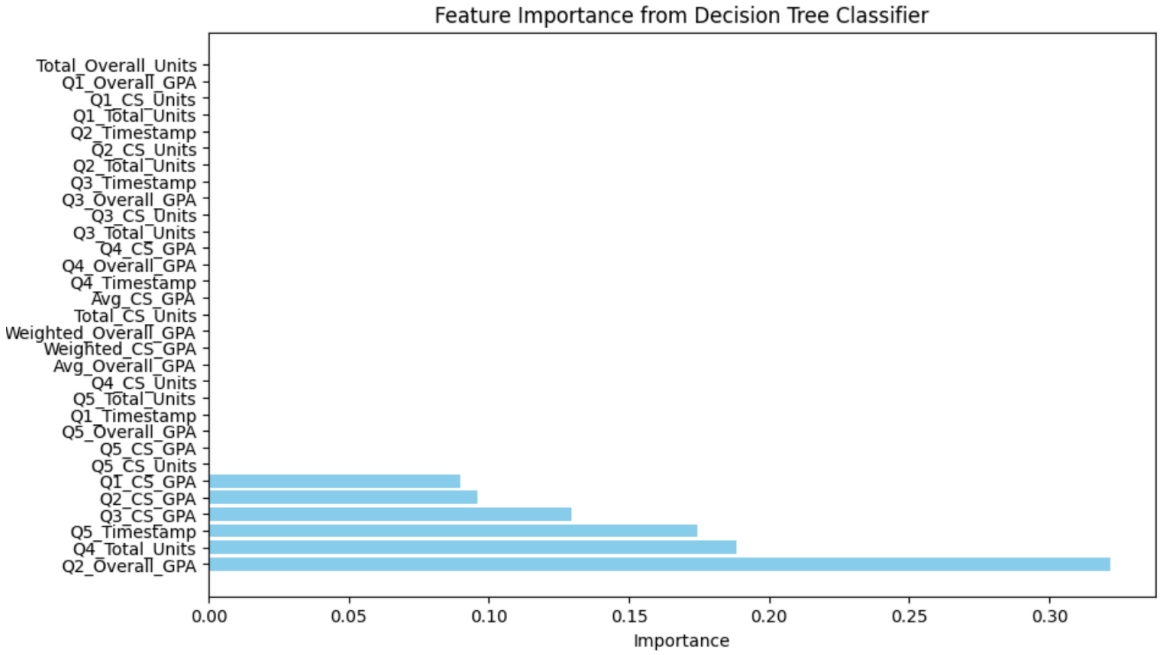
Student Record Analysis

This short analysis will focus on answering the following questions:

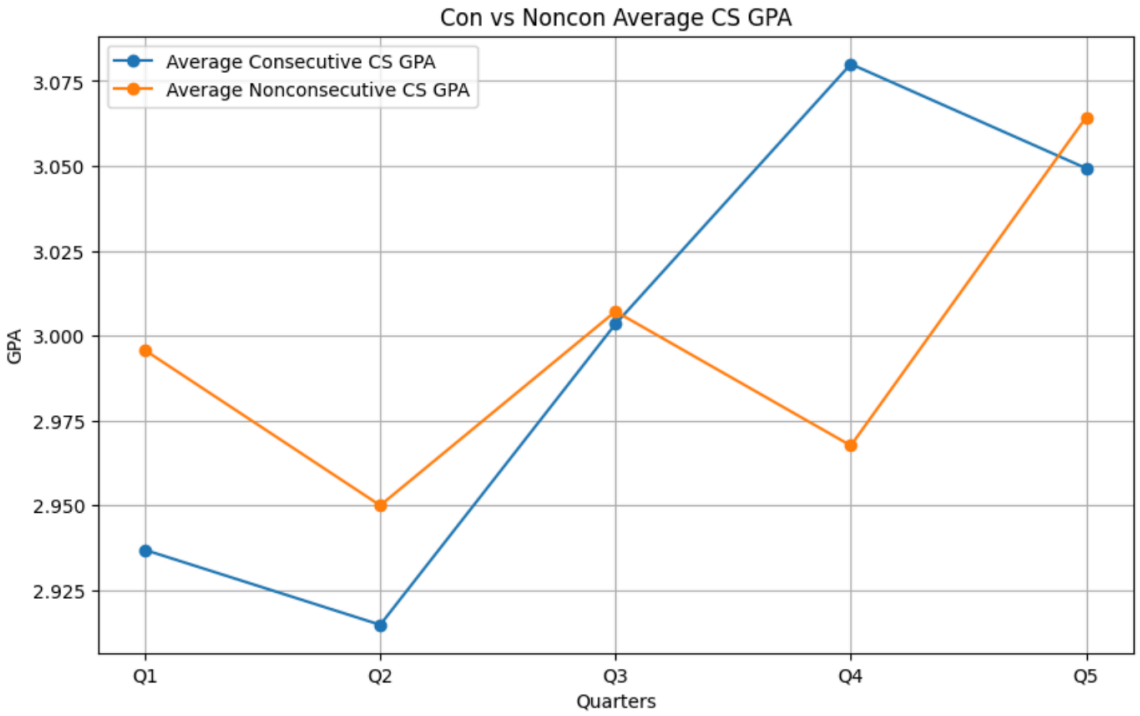
1. What are primary factors students take a gap quarter
2. Does taking a gap quarter affect students' gpa

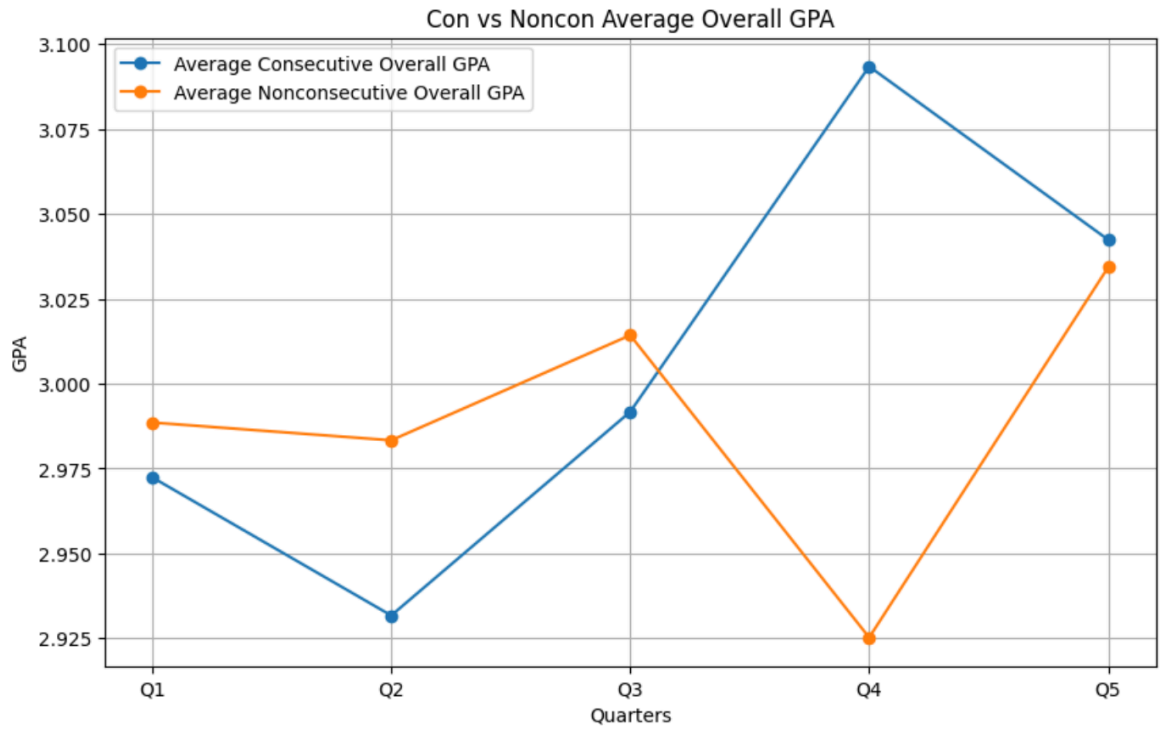
Definition:
Gap quarter: A gap quarter refers to students who did not take consecutive quarters and took some time off (i.e. F21 W21 S21 F25 W25)

What are primary factors students take a gap quarter
To answer this question, a decision tree classifier is trained based on the available data and then the classifier's information gain was obtained to see which of those features made students decide to take a gap quarter.

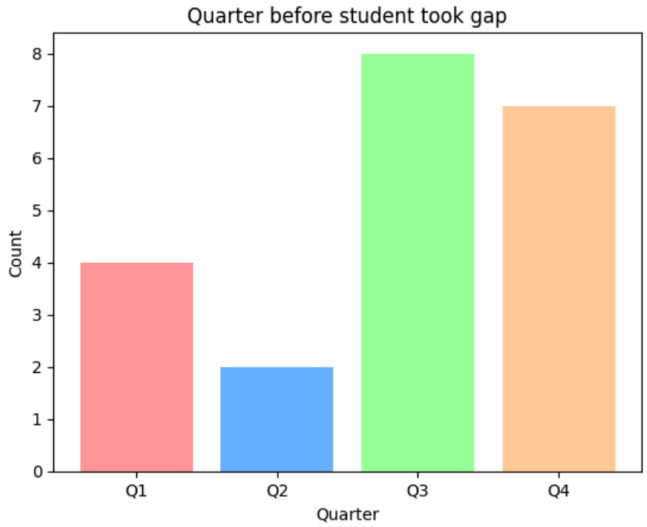


Based on the information gain, GPA is a significant contribution to whether a student decides to take a gap quarter. Below is a closer look at the comparison between students who take consecutive quarters and students who take a gap quarter.





From the data, we observe that on average students with higher CS GPA and overall GPA are more likely to take a gap quarter compared to students who did not take a gap quarter. A closer look at when the student took a gap also reveals that many students decide to take a gap after 3rd or 4th quarter.

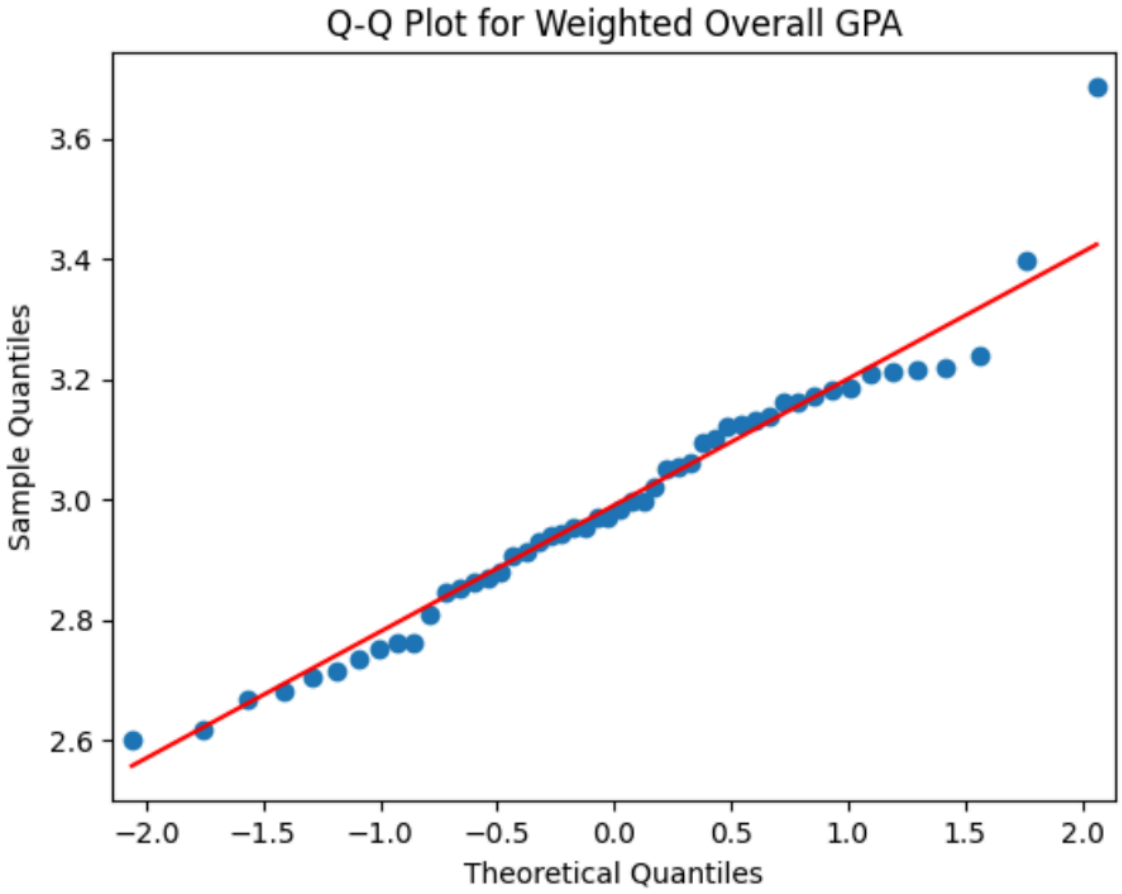


A reasonable assumption is that students with higher CS GPA are more likely to secure internships and thus are more likely to take a gap quarter. Many students decide to take the gap after 3rd or 4th quarter where they accumulated a reasonable amount of knowledge/gpa in CS to enable them to find and secure an internship.

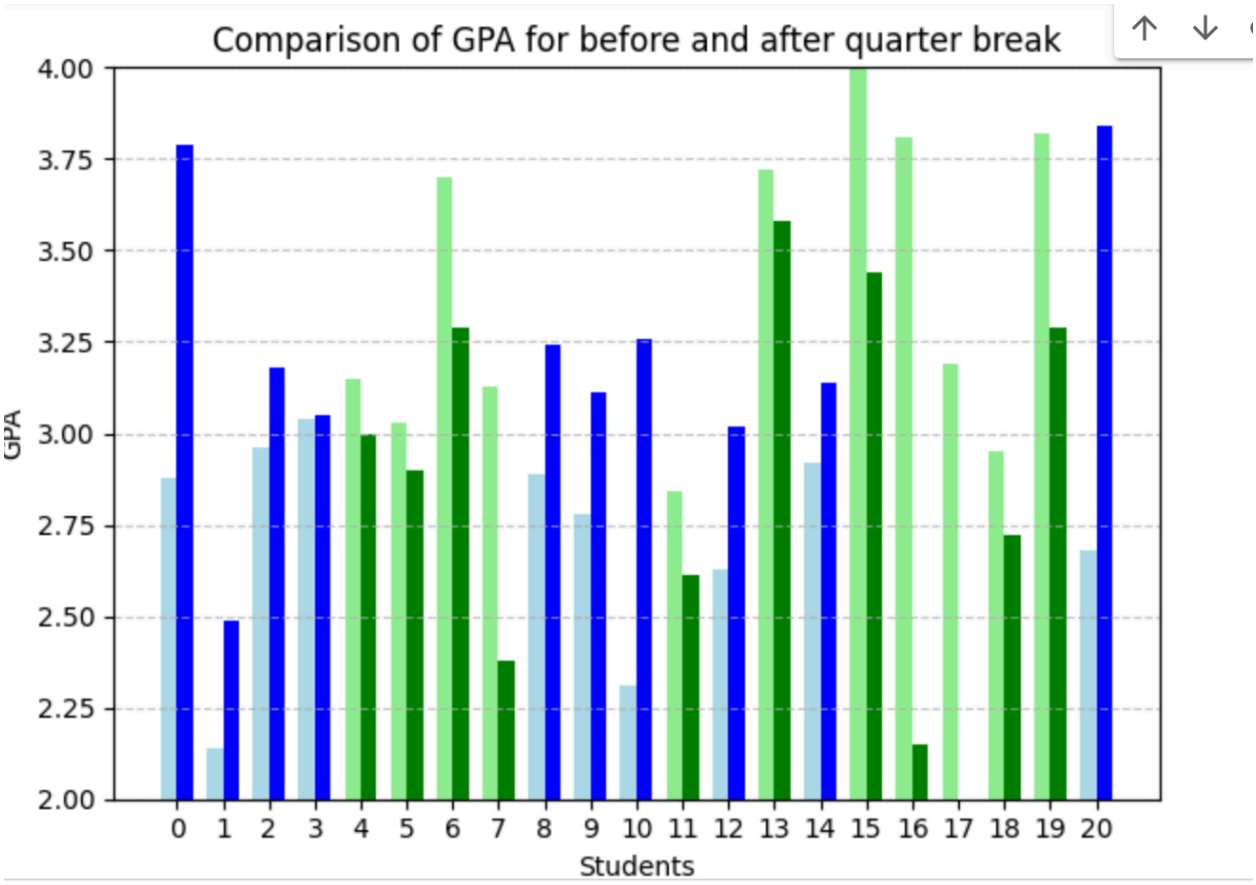
Does taking a gap quarter affect students' gpa

One reason to ask this question is to see if after a gap quarter, student's become "rusty" at CS and may see a drop in GPA. My hypothesis is that we will observe a drop in students' CS GPA after they take a quarter off.

To answer this question, significance tests such as the t-test should be used to check statistical significance. Before running the t-test, we must check if the data is normally distributed.



Based on the QQ Plot, the data seems to be reasonably normally distributed. This gives us the opportunity to use t-test to test statistical significance. The choice of t-test is paired t-test that calculates 2 related sample scores. In this case, would be the students' GPA before and after they take their gap quarter with the expectation that the student's GPA is higher before they take their gap quarter. A series of t-tests performed on CS GPA reveals no statistical significance ($p > 0.05$) on student's who took a gap quarter. This is further supported by the example bar chart below



Blue indicates students performed better after break, green indicates students' performance dropped after break. They are roughly equal and thus no statistical significance is revealed. Thus, the answer to whether or not a gap quarter affected a student's cs performance is likely to be no. More tests needed to be performed to be rigorous. Some test could be

1. Check academic performance of students over the span of 5 quarters rather than immediately before and after break quarters
2. Check units taken and how they relate to academic performance of students who took a break
3. Take in consideration the time frame comparing students of the same term only (before and after break). The reasoning for this could be that some quarters are generally harder than others quarters and looking at the mean overall does not provide enough information. This could be a problem when there are only 50 data points available.