SDS 323 Final Project

Ashwin Shankar

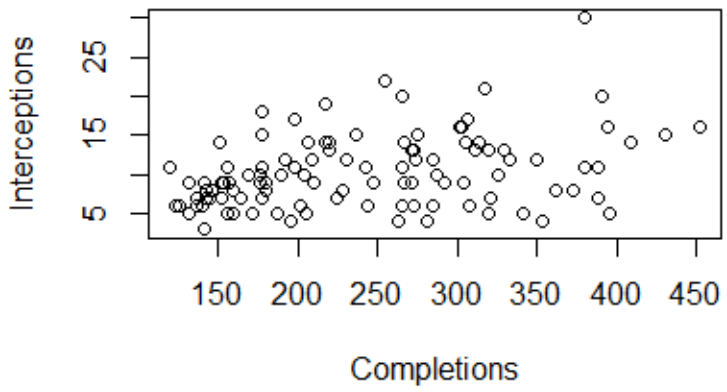
My project will be to utilize tree-based methods to create a model that shows which variables increase the chances of a quarterback throwing an interception.

The dataset I will be using for my project will be the quarterback (QB) passing statistics from the 2008-2019 National Football League (NFL) seasons. I obtained the data from pro-football-reference.com.

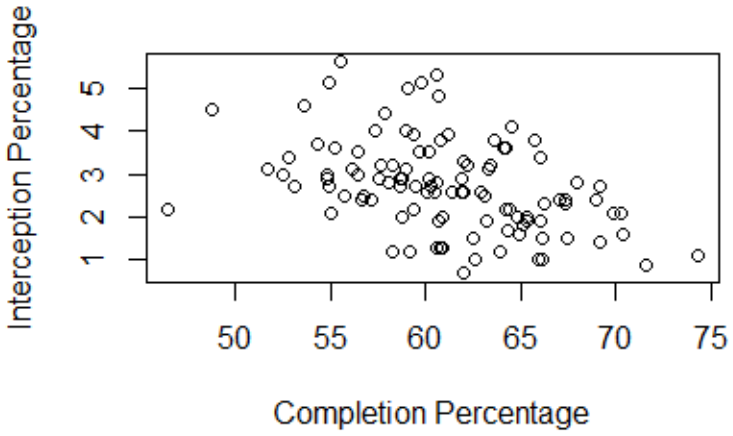
Processing and Descriptives

For each season, the data originally had over 100 observations, so I first removed all the players who were not QBs (e.g. Wide Receiver throwing a pass on a trick play). I then used the NFL minimum pass attempts to qualify for season statistics (224) to eliminate other QBs. The final set resulted in 106 QBs who had at least 224 pass attempts throughout their respective seasons.

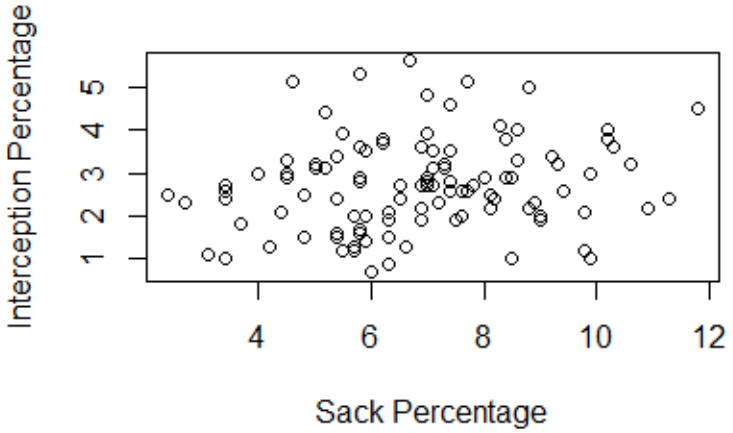
In a single season, the average NFL QB was 28.7 years old, attempted about 393 passes and completed about 243 of them, resulting in a completion percentage of about 61%. The average QB also threw 10 interceptions, 16 touchdowns, got sacked about 28 times and had a quarterback rating (QBR) of 60.37.



QBs with higher amounts of completions tend to throw more interceptions, which makes sense since more completions means more pass attempts, which increases the chances of an interception.



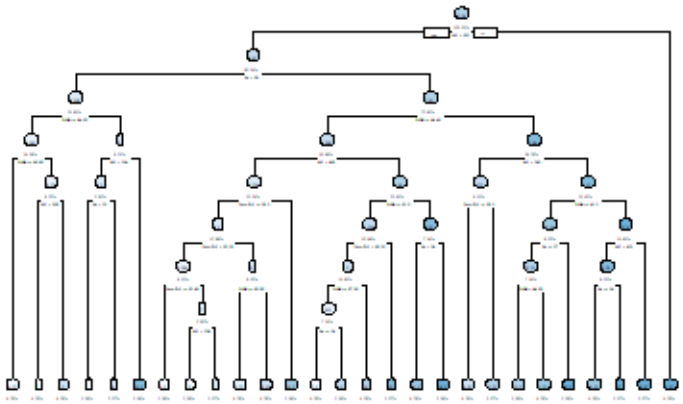
Looking at completion percentage vs interception percentage shows a negative correlation. This makes sense because QBs that complete a higher percent of their passes tend to throw less passes to the other team.



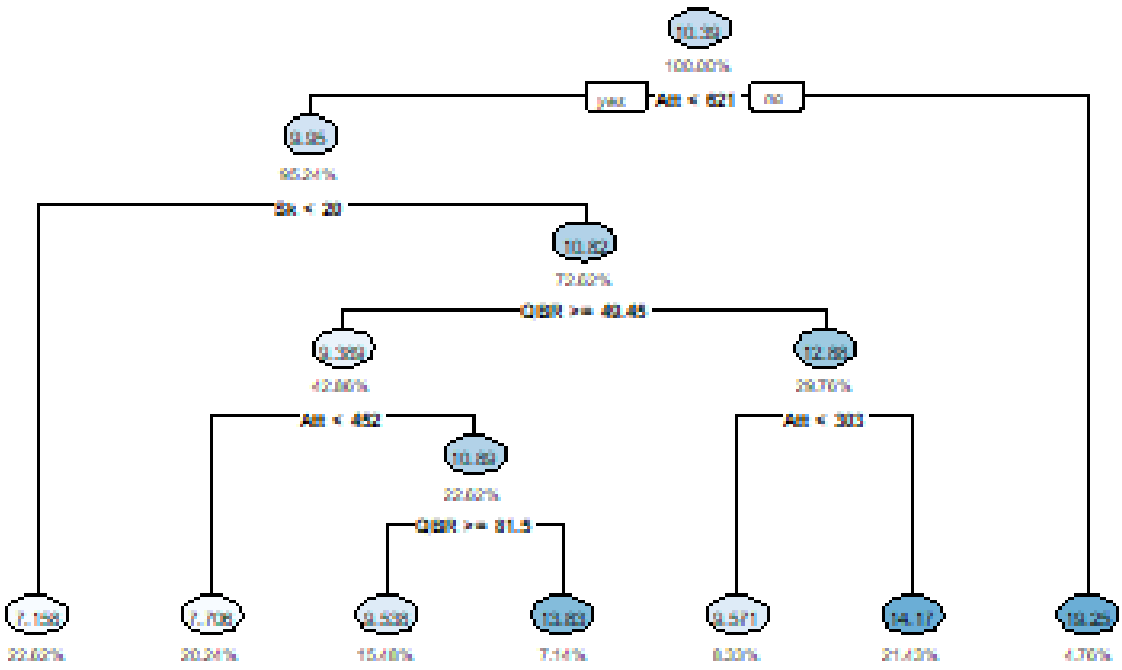
Interestingly, there seems to be a small positive correlation between sack percentage and interception percentage. This could be because QBs that have a higher sack percentage tend to spend more time in the pocket, which increases the chances of throwing an interception. However, sacks can also be the result of not enough time in the pocket due to poor offensive line play, which could rush the QB on future pass attempts and lead to errant throws, increasing the chances of an interception.

Analysis

To analyze the factors that increased the chances of a QB throwing an interception, I used a tree model. I first created a deep tree with the variables attempts, sacks, QBR, touchdowns, completion percentage, first downs, and yards per attempt. The deep tree had a size of 27.



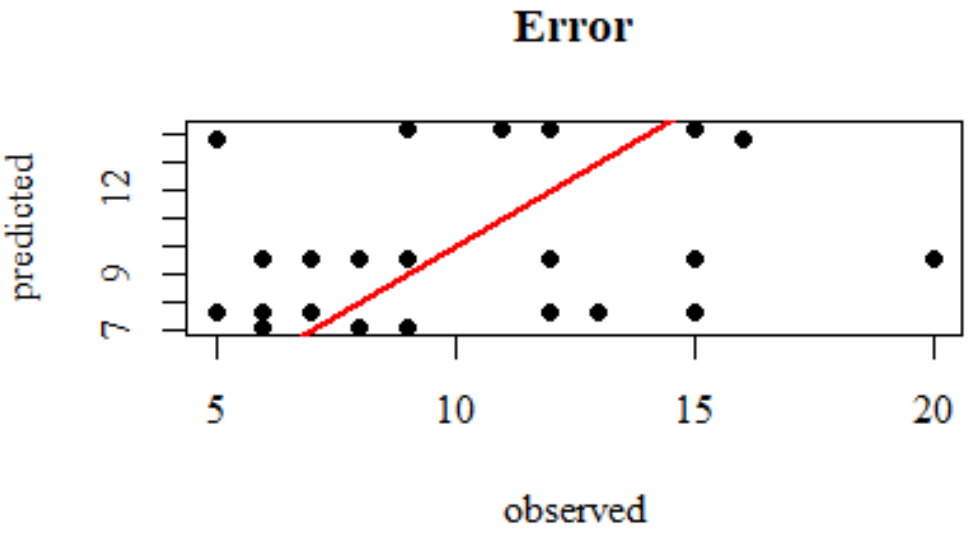
The best alpha value from the deep tree was .0344, which was used to prune the tree, resulting in a tree of size 7.



The final model included the variables sacks, attempts, and QBR. Although the model says that throwing over 621 passes decreases the chances of an interception to about 4.7%, this makes sense since very few QBs throw that many passes and that ones that do tend to be the best QBs, who tend to throw fewer interceptions.

Another branch to note is the sack branch. Having less than 20 sacks decreases the chances of an interception to about 23%. This makes sense since low sack numbers are due to a combination of good offensive line play and quick release times when throwing. Either the QB has more time to make an accurate throw or he is throwing quickly to a crossing or screen route, which have a higher chance of being completed.

The lowest QBR branch is also an interesting branch. Having a QBR below 81.5 actually decreases the chances of throwing an interception to 7.5%. This is probably because these QBs are not the worst, but they are also not an elite QB. These kinds of QBs are called "game managers" because they have a great supporting cast (Receivers, Running Backs, Defense, etc.) and due to that they are asked to do just enough to keep the offense going, resulting in fewer passes thrown, thus lowering the chances of an interception.



After predicting the model on the test set, the mean error was about 4.30, which is pretty good. The model did a good job of analyzing which variables in the dataset were relevant to calculating the chances of a QB throwing an interception.