

# NFL Data Analyst Sample Analysis

Michael Dixon

9/3/24

For this report, I wanted to start with some exploratory data analysis on a few of the factors that seem most relevant to a hypothetical NFL organization. The first element of the data I wanted to assess was where current season ticket holders reside. I thus created two graphs, one showing the five most common locations (Figure 1) appearing in the data and a second showing how many NA values exist within the Distance From the Stadium column (Figure 2).

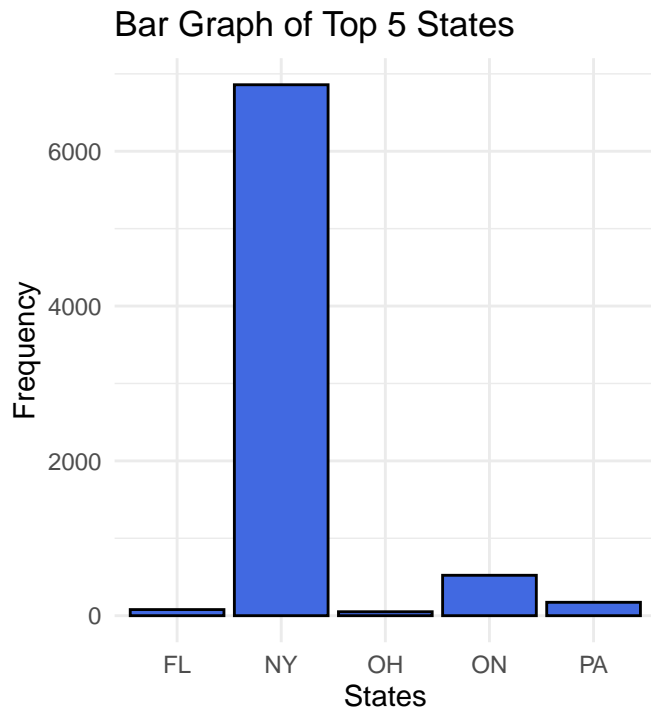


Figure 1: Boxplot Count of Five Most Common States

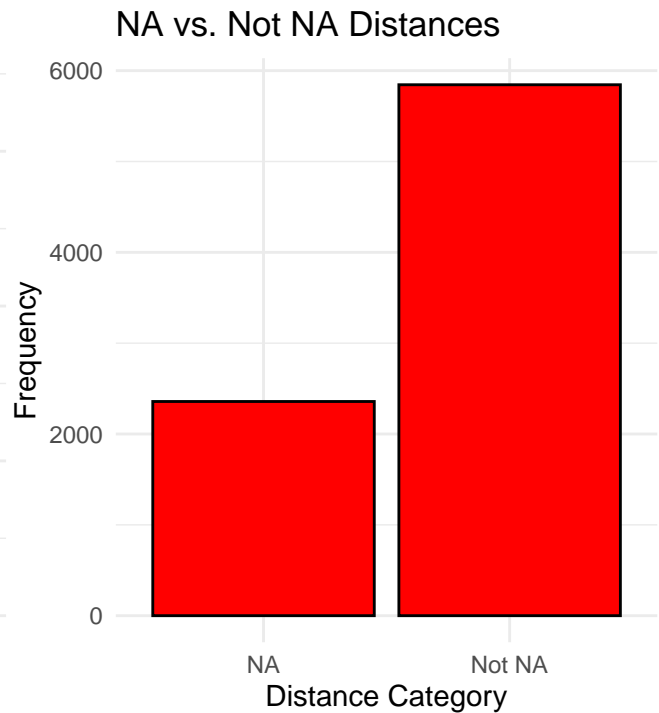


Figure 2: Count of NA Values vs non-NA Values

The most common state where season ticket holders reside is New York. Interestingly, Ontario, Canada is the second most frequent area that fans draw from, which is surprising given it is a province of Canada and the NFL is comprised exclusively of teams in the United States. As for the distance-to-the-stadium values, my initial belief that those values could be useful for some potential modelling turned out to be incorrect given how many values are missing (as reflected by their NA status). With so many missing values, substantial data manipulation would likely be required and even then, the distance-to-the-stadium values might not provide anything useful to the model. For that reason, I have omitted those values from my analysis.

Next, I wanted to examine how well the team retains customers by looking at the tenure of the team's season ticket holders. I plotted the histogram below (Figure 3) to examine this phenomenon.

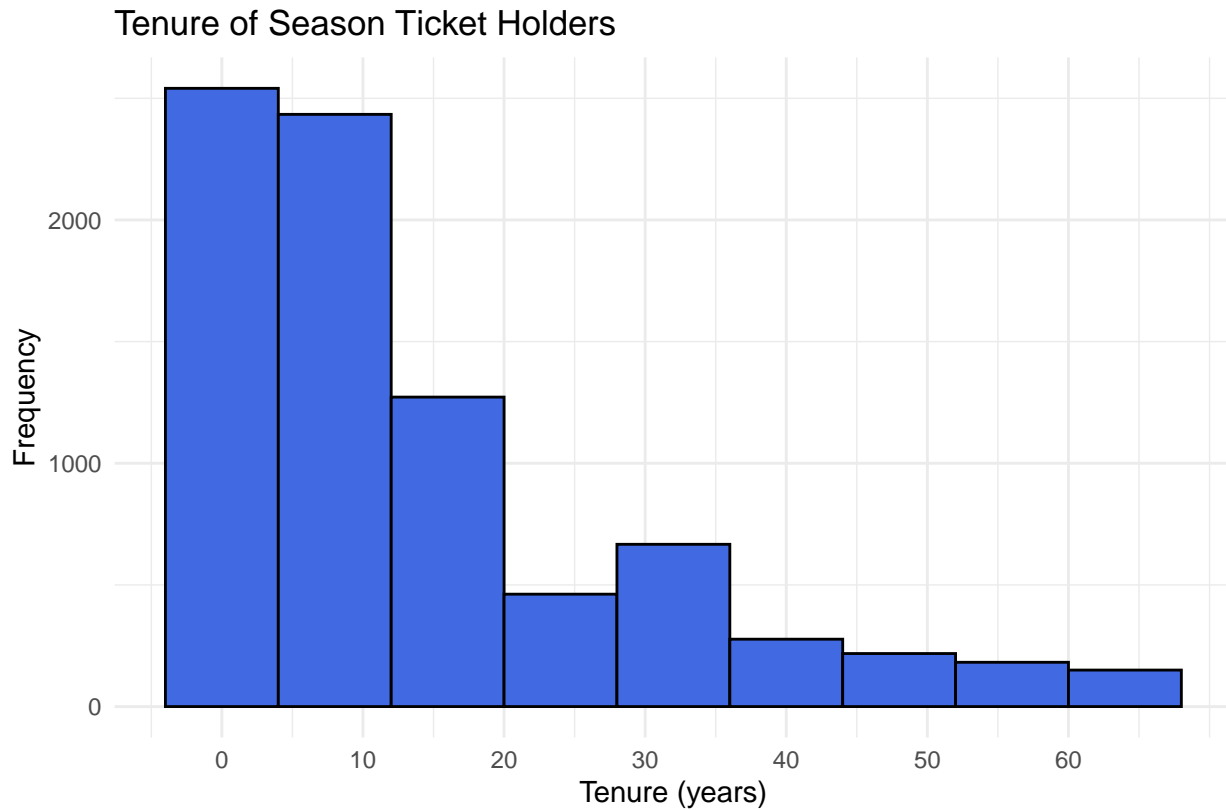


Figure 3: Histogram of Season Ticket Holder Tenure

This chart is particularly valuable because it has been estimated that acquiring new customers can be five times more expensive than retaining old ones. In this plot, we can see that while there are certainly many long-standing customers (exceeding 50 years as season ticket holders), there are far more relatively new customers (less than five years). Information like this would be crucial for the business analytics team to know.

Finally, I wanted to examine if “perks” might influence a customer’s decision to renew season tickets. In the dataset, the ‘club-flag’ identifies premium tickets that offer prime sections of the stadium and added benefits. The bar chart below (Figure 4) analyzes if there is a stark difference between the club seat retention and the non-club seat retention.

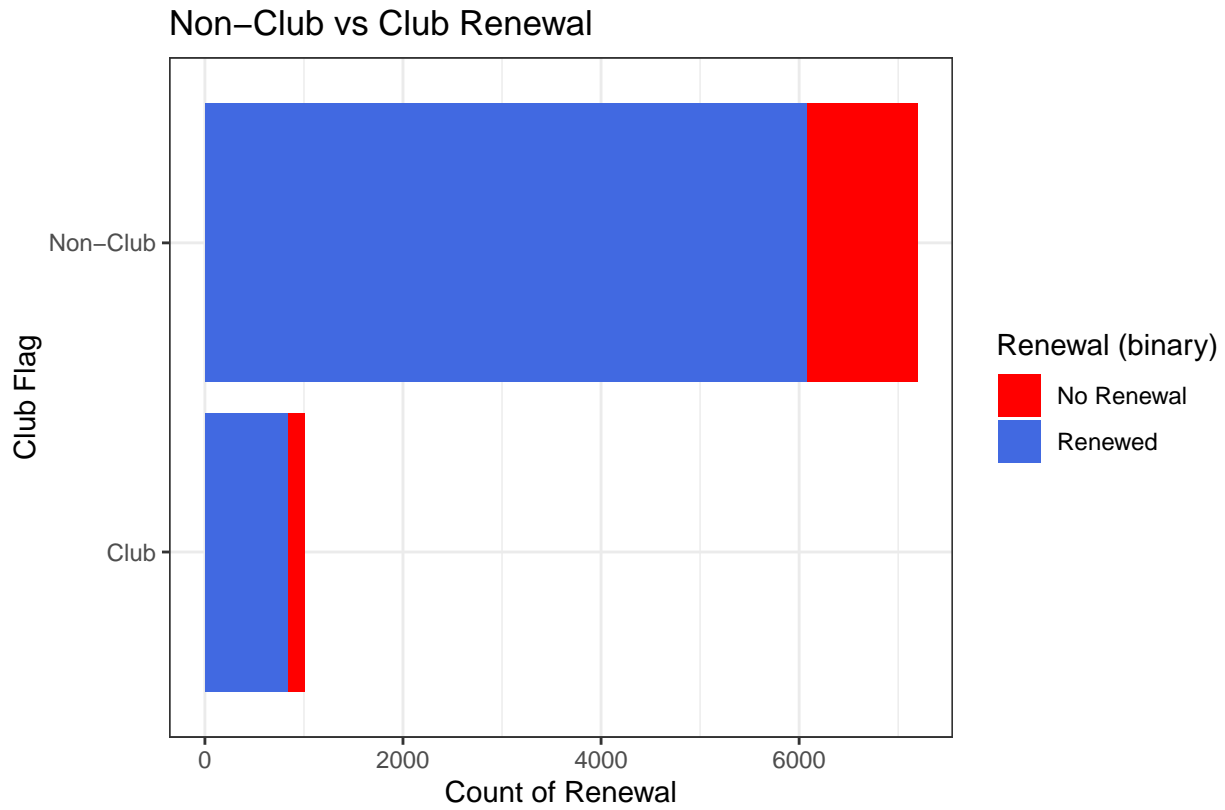


Figure 4: Boxplot of Renewal, by Club Status

In this result, we can see that while a much smaller percentage of the season tickets sold had club access, the team had a higher number of them renew their tickets compared to the average. Other factors could be at play here (for example, perhaps more club tickets are owned by corporations and are more likely to be renewed as a corporate expense). Regardless, these benefits seem to hold weight with the clients.