NFL Game and Arrests Analysis

Rojan Khatri

Introduction

American Football is the most popular spots in the USA. Football is more than just sports; it has become a new religion in America. Before I came to USA, I was not exposed to this sport. If anything, close I had watched was Rugby few times, but it was never the subject of my interest. After I arrived here and got more exposed to the American football, my interest towards NFL grew. I started watching and following NFL games with my friends and cousins. With sports there are fans who are very loyal to their teams. Like any other sports NFL have the fans rooting for their teams. Almost all the time, loyalty stays even when fans move out of city or state. With winning and losing of the game there comes some absurd activities by fan that leads to an arrest. They take it to next level by starting a fight, get in trouble when their team underperforms. After any major NFL games, there will be increased number of news about the brawl, stabbing and shootings relating to the game. It's interesting to see how the outcome of NFL games are related to the arrests made that day.

Outcome:

- Correlation = -0.047
- Spearman Rank Correlation Coefficient -0.009
- Chi-square Statistics 1874.492 p-value 1.000
- P value is 1, which suggests that variables are independent of each other.

There seemed to be strong correlation between the score difference and the number of arrest but while calculating linear regression the score was only 2.2% which would not be efficient model to predict the number of arrests based on the score difference. More analysis needs to be conducted to create model to predict the number of arrests. Possibly combining the categorical variable divisional game and difference might be able to predict the number of arrests.

Were there any variables you felt could have helped in the analysis?

The main objective for this EDA was to see if NFL game contributed to the number of arrests made in the city during or proximity of game day. That variable/category would have helped significantly. Because of the complexity of the NFL schedules it's not always possible to guarantee that same two teams will play against each other every season with exception for

division games, if there was a variable indicating those arrests were NFL related then the host city could come up with better plan to manage security.

Were there any assumptions made you felt were incorrect?
The assumption that all the arrest made were NFL game related seems to be quite incorrect.

What challenges did you face, what did you not fully understand?
The challenges I faced was obviously the lack of data as discussed above, which prevented the analysis to be more accurate and fulfilling. The one thing that I did not understand fully was the PMF for both divisional and non-divisional game were very similar. I was under the impression that divisional game would have more influence in the arrest.