Project Proposal: R programming (NBA)

In this Capstone Project, I will be analyzing the data obtained that involves late game referee calls that the National Basketball Association has reviewed. Using this data, I want to consider several issues that come with mistake and mistake free referring at the end of the games, in particular, the last two minutes. The main problem I want to resolve is if the referees in the NBA are making mistakes late in games and if those mistakes are possible the deciding factor to whether they win or not. However, there are a lot of subset issues with this that may come afloat. I also want to see if any specific players are getting better treatment late in games compared to others. Sometime the popularity, and prestige of a player could be a factor in whether the referee makes a call against them or for them in the last two minutes. With specific players being treated well, it could also mean that specific teams are being treated better due to their skill level and ranking in the league, this is a possibility and I want to find the underlying cause of it.

My client that will be receiving this information is the National Basketball Association. The NBA cares about this issue the most because it tells them which referees are preforming well in the league, and which are not. It also tells them if there is a possible bias towards players, or teams in the league by referees. We could also go into more detail to wither referees paired with other specific referees tend to make mistakes. Based on my analysis, it possible that the NBA may begin punishing referees for making bad calls, they could remove these specific referees that make the most mistakes, or maybe the NBA could do a better job of pairing specific referees together for a game if it is too difficult to keep hiring new referees, because sometimes even a mediocre experiences referee is better than a brand-new amateur referee who is still learning.

The data I will be using for this project is coming from the official NBA's last two minute reports that they have given to the public on their website. I have found a website that has complied the data all into one excel file. The issue is this data is being updated daily because there are games every day, so even my results will be old regardless of when I complete it, however there is enough data still to make a profound conclusion to my questions.

My approach to solving this problem is to analyze the date by sorting the it into proper rows and columns that I can go to easily when making the code to decipher the data. I want to first look at how many incorrect calls there are compared to correct calls made. After looking at that, I will begin to sort the data by specific players and how many times they were in the advantage and disadvantage based on correct and incorrect calls made. I will then sort by team and do the same thing. Once this is completed, I will look at which referees showed up the most when players were in the disadvantage the incorrect call was made. Just to summarize the whole data analysis in the end, I will see if there are more correct calls made compared to incorrect calls, just to see the balance on how correctly the referees are making calls.