

FUNDAMENTALS OF MACHINE LEARNING

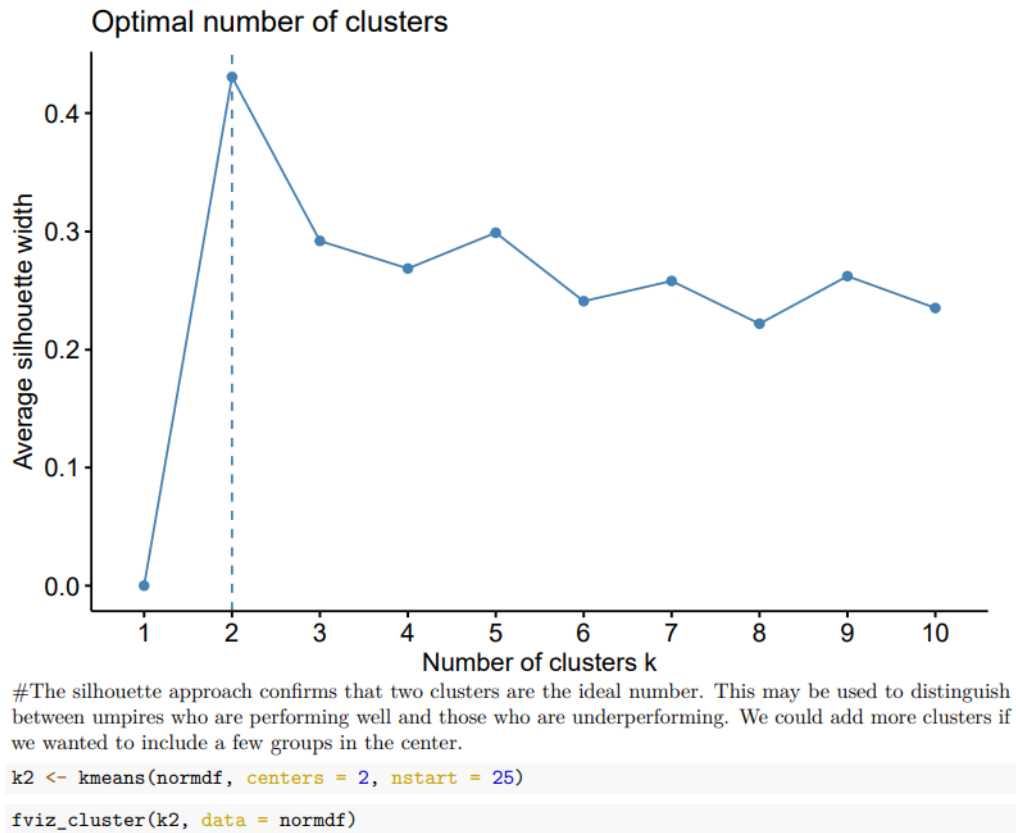
FINAL PROJECT

ROHITH DESAMSHETTY

17-12-2022

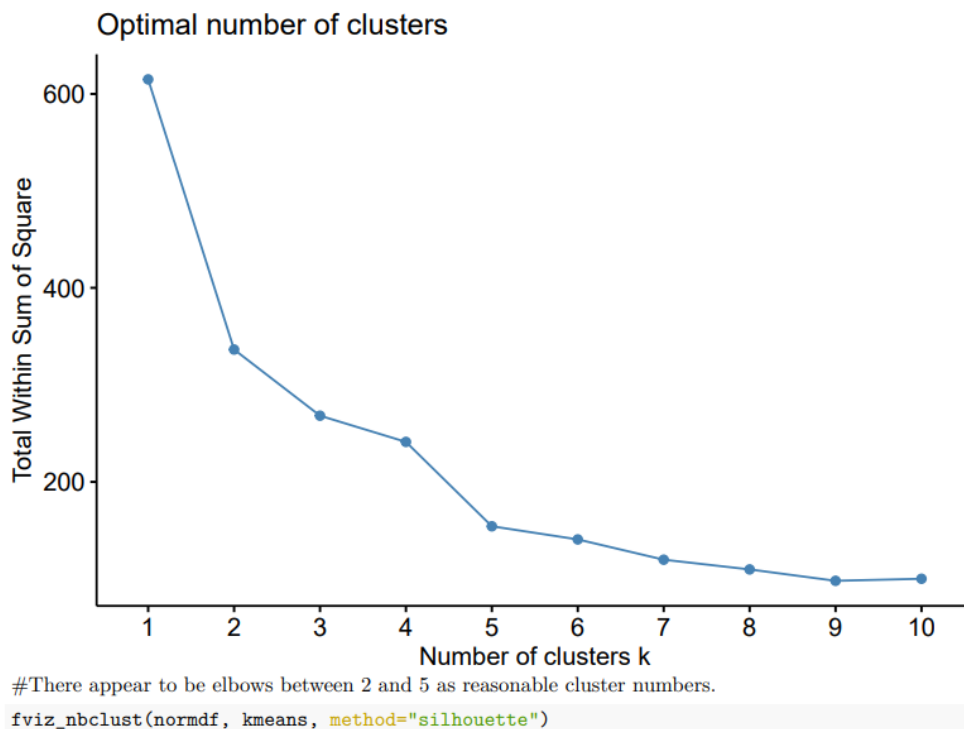
INTRODUCTION: This dataset contains a list of MLB regular season games from the last seven years. Umpire statistics include umpire name, ID, home and away team runs, number of pitches called, accurate versus predicted calls, and overall umpire accuracy.

ABSTRACT: The analysis' purpose is to give valuable insight to MLB executives and commissioner Rob Manfred on the performance and overall efficacy of MLB umpires, as well as to provide suggestions for groupings depending on how they perform in key categories.



K Means Clustering was proven to be the most informative and relevant strategy for this investigation. This approach will group umpires based on critical performance characteristics. Using this data, we can determine the overall accuracy and efficiency of umpires in comparison to expectations. By displaying essential performance indicators, the clusters effectively summarize top performing umpires. Higher-performing umpires consistently outperform expectations in terms of right calls, accuracy, and consistency, all while avoiding considerable home bias. There are who, on average, fall short of accuracy, consistency, and call expectations. The clustered data may be utilized to determine which umpires should supervise high-stakes games in the playoffs and World Series, as well as to identify specific flaws or defects. It is easy to identify areas of opportunity for poorer performing umpires and offer particular training to enhance performance using cluster analysis.

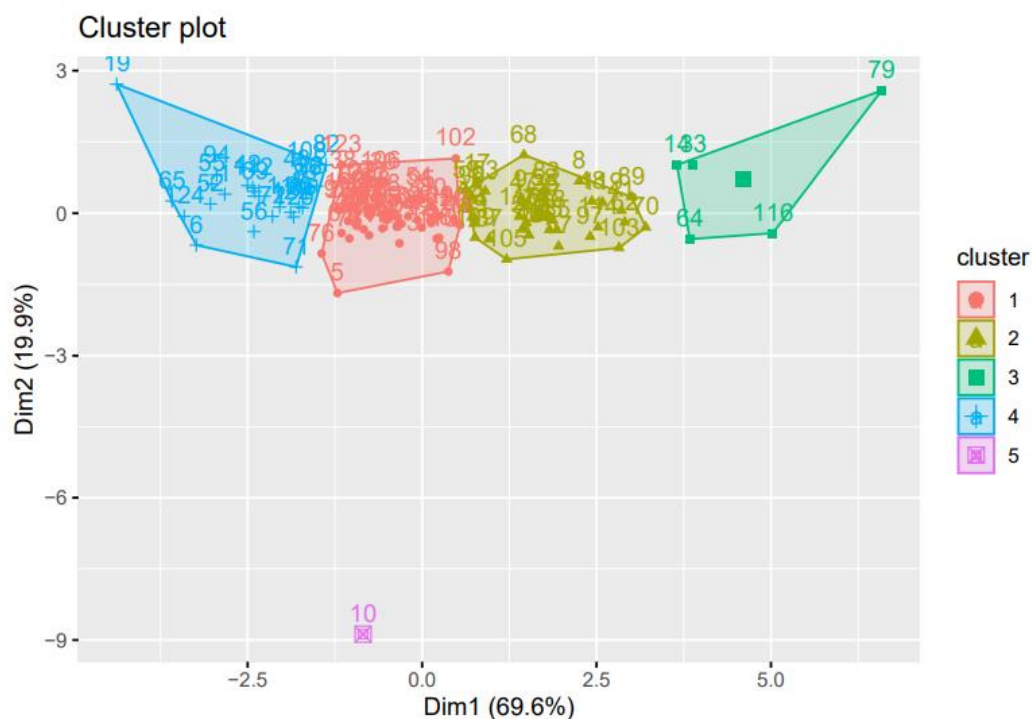
ANALYSIS: Using K Means Clustering analysis on this dataset, we detect five distinct clusters. Four of the clusters are large, while the fifth comprises one umpire who scores particularly well on the "favor home" criterion.



Cluster 4 (blue) represents the best performing umpires. Statistics like as accuracy, consistency, and accurate calls all suggest above-average performance for these umpires while having a below-average influence on total runs scored, indicating that they are not impacting game outcomes. Cluster 1 (red) represents the average achievers. All data indicators, such as accuracy and consistency, are reaching or slightly exceeding expectations. Cluster 2 (yellow) implies below-average accuracy, correct calls, and consistency while having above-average effect, indicating that which indicates They had a stronger influence on game outcomes. The lowest

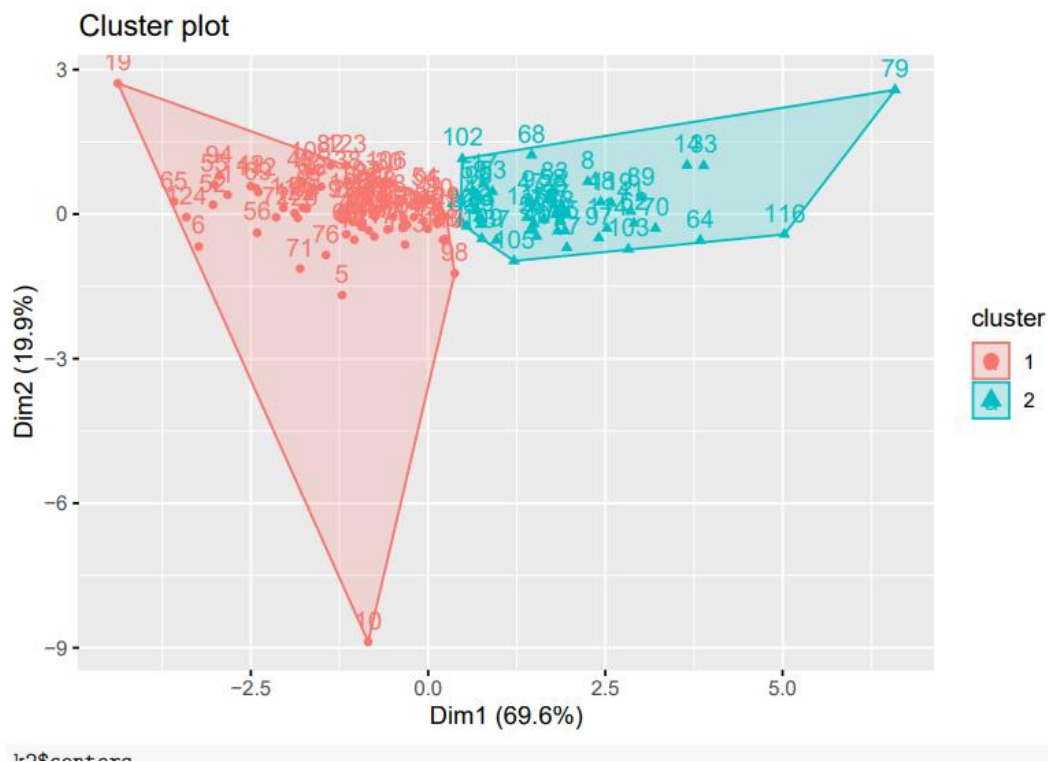
performing umpires are displayed in Cluster 3 (green). This cluster suggests that these umpires' accuracy, accurate calls, and consistency are far below average. Furthermore, they have a substantially greater influence on total runs scored, implying that the performance of the umpires may be influencing the outcome of the game rather than the performance of the players. Cluster 5 (purple) has one umpire. His performance is somewhat below average in terms of correct calls and accuracy, but he is consistent. Although the data implies an excessive home bias, the prefer home measure for this cluster is exceptionally high, indicating that this umpire might be performing pretty well.

Statistics like as accuracy, consistency, and accurate calls all suggest above-average performance for these umpires while having a below-average influence on total runs scored, indicating that they are not impacting game outcomes. Cluster 1 (red) represents the average achievers. All data indicators, such as accuracy and consistency, are reaching or slightly exceeding expectations.



Cluster 2 (yellow) shows below-average accuracy, correct calls, and consistency while having above-average effect, indicating that they had a higher impact on game outcomes. The lowest performing umpires are displayed in Cluster 3 (green). This cluster suggests that these umpires'

accuracy, accurate calls, and consistency are far below average. Furthermore, they have a substantially greater influence on total runs scored, showing that umpire performance, rather than player skill, may be influencing game outcome.



Cluster 5 (purple) has one umpire. His performance is somewhat below average in terms of correct calls and accuracy, but he is consistent. Although the data implies an excessive home bias, the prefer home measure for this cluster is exceptionally high, indicating that this umpire might be performing pretty well.

CONCLUSION:

Based on my results, I would advocate sending the umpires in Cluster 4 to the World Series and including them in the formulation of off-season training programs for the umpire clusters who are performing below average, notably Cluster 3 and 2. Cluster 1 is operating as planned, thus they may be employed with Cluster 4 throughout the regular season and playoffs without altering game outcomes. Cluster 2 should be forced to conduct more offseason training in

order to perform at their projected level. Umpires in Cluster 3 will require considerable training or may be terminated or demoted to the minor leagues if offseason training does not drastically improve. Their performance. Anthony Johnson is the lone outlier in Cluster 5. Some offseason training in addition to resistance train the influence of home crowds is recommended since the data suggests a strong bias for hometeams and a slightly below average performance for correct calls and accuracy. Although the strong home bias is suspicious, Anthony's consistency rating is the highest of the five clusters, suggesting his potential to perform well if he can correct the home bias issue.

