Machine Learning

Machine learning is an aspect of data analysis that can definitely be applied to my particular project. Because my dataset deals with different teams over a period of time, we can use clustering to find patterns between variables. By using different machine learning functions, the data can be filtered to show correlations. We can apply machine learning to find relationships between various variables in the dataset such as team, wins, home average attendance, and year. The main dependent variable used from this dataset would be the home average attendance. Several of the other variables such as team, year, and wins can be used as independent variables.

Hierarchical clustering can be used to group similar objects into clusters and we may be able to find some interesting relationships. We can cluster teams with similar amount of wins or similar home average attendance and see if anything significant is found. Hierarchical clustering can be useful as a comparative tool to juxtapose two teams to each other.

Machine learning can also be used as a predictive measure. By analyzing the trends of home attendance for each team, we could apply machine learning to predict the general average of attendance for teams taking into consideration their wins.

Different types of machine learning also include linear and logistic regression. These regression are also capable of taking the dataset and making predictions using analysis. My analyzing the relationships between two categories, a prediction can be made such as how many wins would be needed for a team to reach a certain amount of average home attendance.

I would expect hierarchical clustering to be most practical and useful for this particular dataset and project.