Lee Johnston

Prof. Iranitalab

Bellevue University

DSC530-T303

3/2/2024

Term Project Summary/Final Essay

During the EDA for my project my goal was to find the overall relationship between the statistical importance of possession in European Football/Soccer relative to a team scoring goals or creating chances to score goals. The null hypothesis/question that I wanted to have answered was "Is possession an overvalued or unnecessary statistic".

I pre-empted the experiment by saying that to a level possession is a required statistic but could possibly be an overblown concept. I sought out to find just how heavily possession and goals were linked though a series of CDF, PMF, Analytical Distribution, hypothesis testing and regression analysis models/methods to come up with my findings. My outcome on the EDA is that with a 0.36 R-Squared value that Possession has less of an outcome on the variables than originally thought and that with relatively low p-values and f-statistics the findings I have produced are very accurate. What I feel was missed during this analysis was the ability to relate possession to wins. I feel as though this would be a far more definitive indicator as to the importance of possession as this current experiment only relates it to goal actions. While goal actions are important they are seemingly meaningless if a team loses a high scoring game due to things such as quick counterattacks that require very little percentages of possession.

The variables I would have liked added are 'wins' as previously stated but also 'losses' and 'Goals Scored Against'. I feel that if I was able to cross reference the data I collected in my EDA with these variables that my findings could have been even more accurate. I did assume that I did not need to calculate age into this equation regardless of some teams having a higher average age than others. In hindsight I understand that things like experience can influence how well a team is able to maintain possession and how having an average age that is on the higher end could lead to a team choosing to play without the ball and look to play in moments on the counterattack. A challenge/issue that I faced during this EDA is that the dataset is limited to a single tournament that lasted for roughly six weeks. Some teams only played 4 games in this tournament so the sample size is limited in certain instances. Also, European football played at the international level instead of club level is played very differently. If I could re-do this EDA in the future I would like to obtain a dataset on a particular league for their entire season and obtain the previously desired variables as well.

Overall, I vastly enjoyed this EDA as the topic is something I enjoy in my spare time. The models/analysis done was also a great way to utilize the skills learned throughout the course.