EDA Summary

By Carlos Palomo

The outcome of my EDA analysis was that there was no relationship between the amount attempts to score and the amount goals during FIFA 2018 tournament. This went against my intuition, that the more attempts the more likely a team would score and therefore the amount goals would increase. I feel I did not include enough variables. If I included amount of block and the percentage a team was in possession of the ball, if feel would yield better results. As that would give a better picture of a team’s actual performance. I had attempted to create such a model, but I was not able to create a model that did not error when I attempted to have multiple variables. Another variable that would have like to include was Man of the Match. The introduction a categorical variable would cause difficulties that I was not prepared to resolve, so it was omitted.

One thing that I did not fully understand was the mixed results that I had received. From the correlation it would appear that the attempts and amounts was at least weakly correlated. This was supported by the correlation tests that was performed. But based of the actual model I could not say that attempts did have a statistically significant role in the amounts.

Lastly one assumption I felt that I made that was incorrect was in the distribution of the data. I used a Normal distribution but upon reflection I do not feel that correct. I believe difficulty in understand which distribution stems from the fact that goals are discreet variable as oppose a continuous variable. Therefore, getting smooth is more difficult.