REGRESSION PROJECT MVP

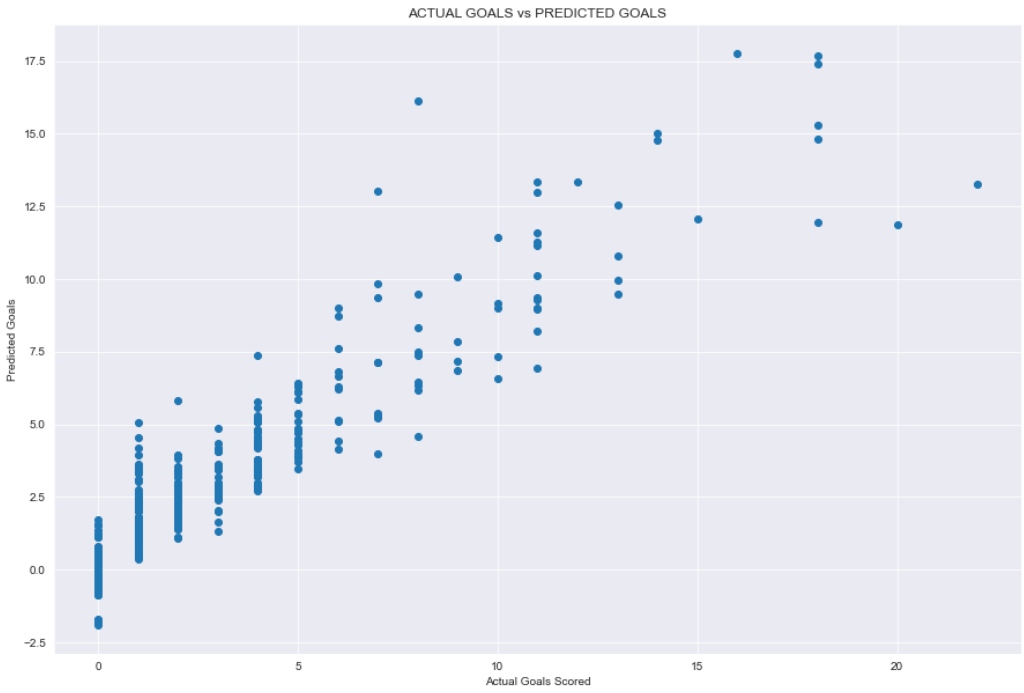
Project Summary:

The project seeks to predict a player's goal outlay a given season based on their performance stats in a previous season or more.
The project is employing data scraped from fbref.com and includes features like age, playing position, shots taken (general), shots-taken-on-target, goals scored and using these to train a model to accomplish the project objective.

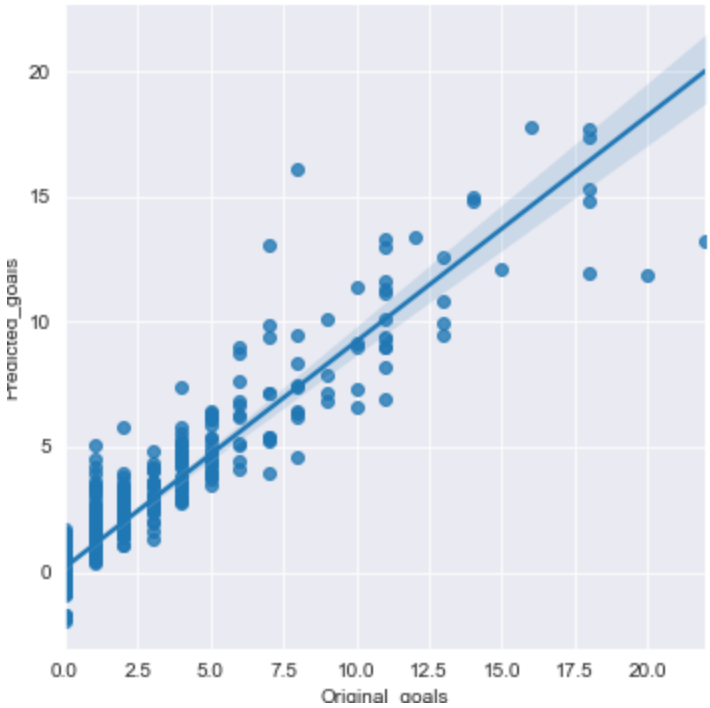
Milestones:

Data Acquisition had have been concluded, and it has. Data was acquired from www.fbref.com.

MVP: A working Model with an 'as near to normal' a residual-plot as possible:
Below is a scatter-plot taken from a rough but working model trained on a train-set of 70% and a test-set of 30% of the dataset respectively.
As you can see, there is a lot of congestion between 0 - 5 goals because that's what the majority of players will return. It thins out on the higher goal outlays, which is quite reflective of the numbers in the dataset (and in real-life football).



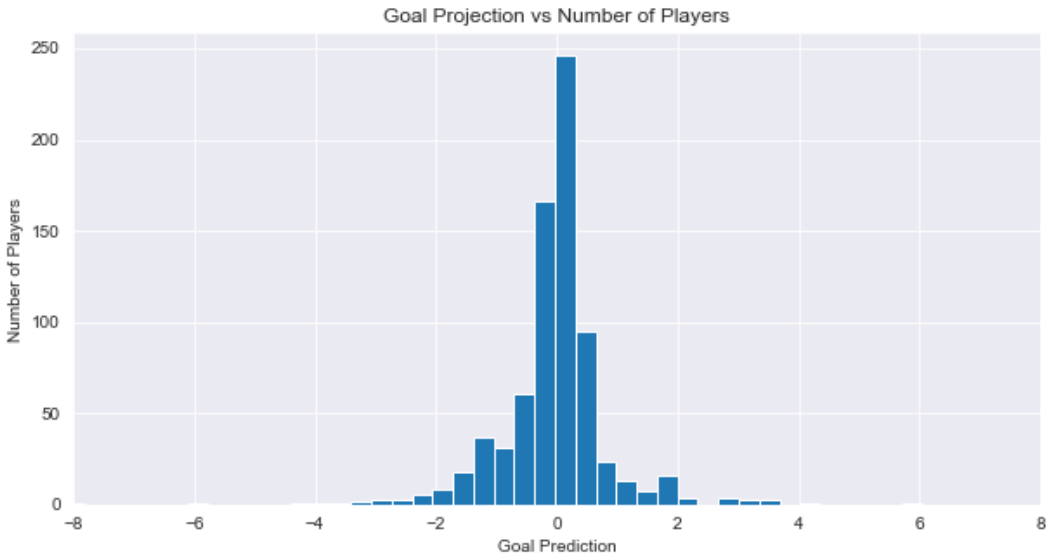
The line is fitted on the same plot below:



It is linear because the model expects players who score lots of goals to keep them up.

Residuals

Below is histogram visualizing the model's prediction relative to the reality in the dataset. In reality though,it shows the distribution of the goal-predictions across the players in the dataset.



As expected, most players will not score any goals.

Negative projections however are a concern which hopefully should be addressed before project submission.

Dennis Ssekamaanya