**Final Project Report**

**INTRODUCTION**

Ice hockey, typically referred to in North America simply as “hockey,” is a sport of systems. One of the systems most subject to innovation is the offensive system, which often will make or break a team. The historical wisdom on the matter holds that teams should attempt to rack up as many shots on goal as possible. After all, as more pucks enter the vicinity of the goal, the chances increase that one may find its way inside.

Fueled by personal observations and frustration with current coaching paradigms, the research question is a relatively simply one: does the amount of shots on goal really have a meaningful correlation with goals scored? If not, is the amount of high-danger shots a better metric for predicting goals? The hypothesis is that the answers to these questions will be “no” and “yes,” respectively. If this hypothesis is subsequently corroborated by the data, then one of the oldest tenets of hockey coaching is brought into question. This hypothesis also implies a null hypothesis (no metric of shot data correlates with goals scored) and an alternate hypothesis (that the traditional approach is the correct one).

Fortunately, the techniques required to perform this analysis are mercifully simple. Simple linear regression will be used to attempt to draw correlations between the amount of shots on goal and the goals scored by players (adjusting both variables for time on ice). Next, the regression model shall be used to generate predictions. Those predictions will then be compared against the actual data in order to extract an R^2 value as well as perform visual analysis. If they do not significantly correlate or the correlation is tenuous, then the first half of the hypothesis will be considered valid. In this case, the same process will be repeated, this time with power play shots on goal (adjusted for time on ice), in order to verify the second half of the hypothesis.

**BACKGROUND**

Hockey is a sport that lends itself to very team-centric play. Unlike several other sports, one or two powerhouse players are not sufficient to carry a team to a championship, and all forward lines and defensive pairings must be firing on all cylinders to ensure consistent success in the sport. This dynamic of consistent production across the board as a necessity requires heavy involvement from coaching staff, with adept coaches being lauded for their clever ability to squeeze the most out of their players and inept coaches being kicked to the curb when their deficiencies become apparent.

Hockey teams live and die by their systems, both offensive and defensive. The offensive systems that are the subjects of this research can be split into two families: those that aim to maximize the quantity of shots on goal regardless of quality and those that aim to maximize quality of shots at the expense of shot volume. Teams who abide by the first ideology tend to feature dump-and-chase routines where the puck is shot into the zone by defensemen to be recovered by forwards. If this is successful, they aim to pepper the opposing goaltender with shots from anywhere they can find free space to shoot, with the idea being that enough shots on goal will eventually create a favorable rebound. Teams who prefer the second school of thought tend to be more adaptive teams, modifying their systems to fit their opponents. Nevertheless, they retain the core methodology of establishing a passing network in the offensive zone and using their passes to manipulate the positions of opponents until a high-danger chance reveals itself. This naturally reduces the amount of shots that are taken, but the ones that are taken often have a higher chance of getting through.

Recently, teams who have used the first approach have repeatedly underachieved in the playoffs, often failing to produce adequate offensive output despite their objectively skilled rosters. The Dallas Stars, coached by Pete DeBoer, and the Toronto Maple Leafs, coached by Sheldon Keefe (who was fired after their most recent playoff elimination), exemplify this problem. The Dallas Stars were eliminated in six games in the Western Conference Final two years in a row after offensive production ground to a halt, and the Toronto Maple Leafs, despite qualifying for the postseason every year in question, have not won a single playoff series since before 2017, with the exception of the 2023 playoffs, in which they eliminated the Tampa Bay Lightning but were then eliminated in five games in the second round by the Florida Panthers. In the Dallas Stars’ most recent exit, they outshot the victorious Edmonton Oilers more than twofold1 but failed to score more than one goal. Clearly, something else is at play.

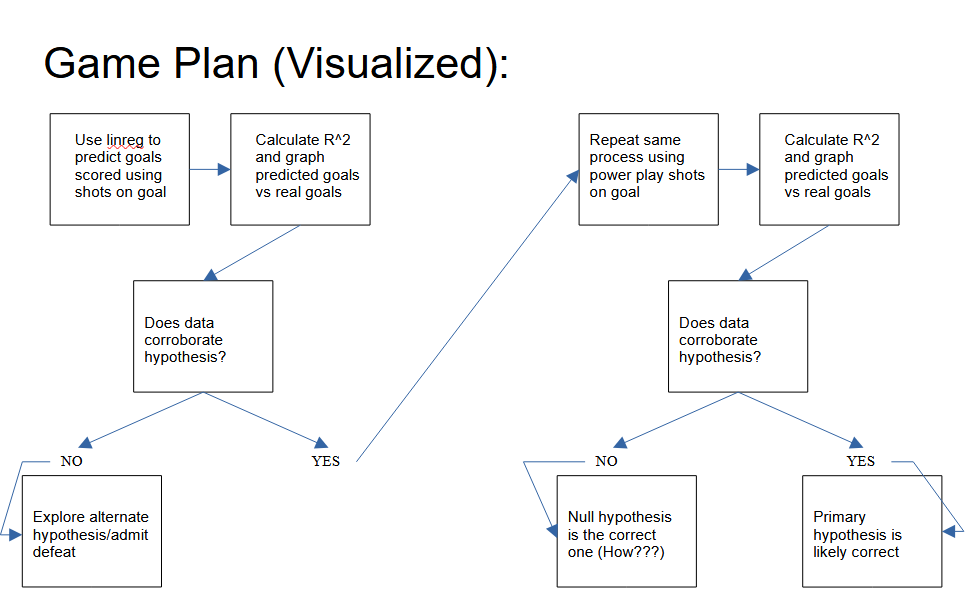
**DATA AND METHODS**

Martin Ellis’s “NHL Game Data”2 dataset was used for the analysis. This extremely detailed dataset provides numerous tables including game records, goaltender statistics, skater (for clarity: a skater means any player who is not a goaltender) statistics, team statistics, and goals scored. The primary table of interest was the table of skater statistics, which contains numerous entries with the statistics of how skaters performed in different games. This is extremely useful, as it provides records of performance in games with a variety of outcomes.

Prior to processing, the data required cleaning. While the dataset is robust and does not lack any entries, entries that contained zero values for either time on ice, shots, goals, or power play time on ice were sought out and removed, as it was discovered early on that such entries muddied the analysis and made the data difficult to visually represent. Once these entries were removed, the rows for time on ice (“timeOnIce”), shots, goals, and power play time on ice (“powerPlayTimeOnice”) were extracted.

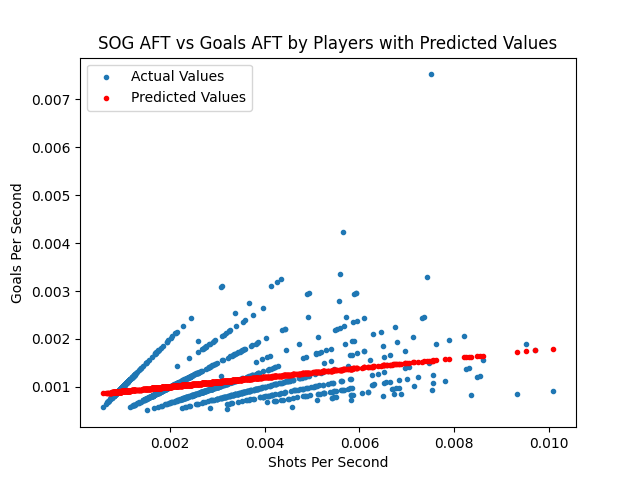
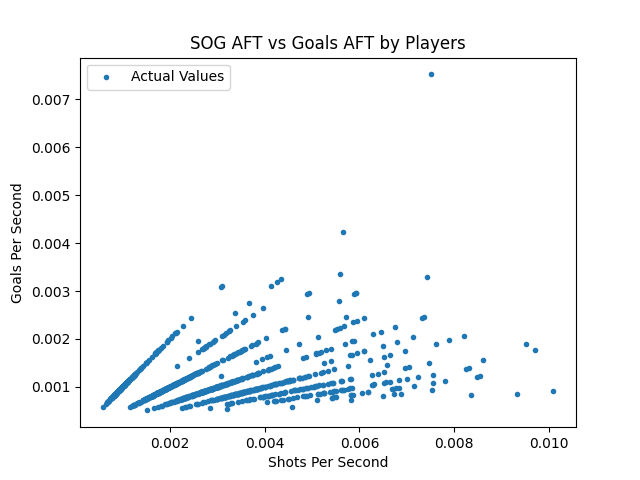
Preparation for the analysis required the variables shots and goals to be adjusted for ice time. Since no official statistic for high-danger shots exists, one had to be fashioned by adjusting shots for power play time. The power play often favors a style of play in which players delay their shots for high-danger opportunities, and on this account it is extremely useful for determining which shots were high-danger shots. Shots and goals were divided by time on ice and power play time on ice, the results of which were stored for analysis. To avoid overly long and repetitive descriptions of data, “PP” will be used to mean “power play,” “SOG” will be used to mean “shots on goal,” and “AFT” will be used to mean “adjusted for time (on ice).”

First, a linear regression model was fitted to explain goals scored AFT with shots on goal AFT. Then, goals scored AFT was graphed against SOG AFT, once alone and once with the predictions of the model for each SOG AFT entry graphed in a different color above the original data. The R^2 value of the model was calculated using the predicted goals scored AFT and the actual data. This process was then repeated with PP goals AFT and PP SOG AFT.

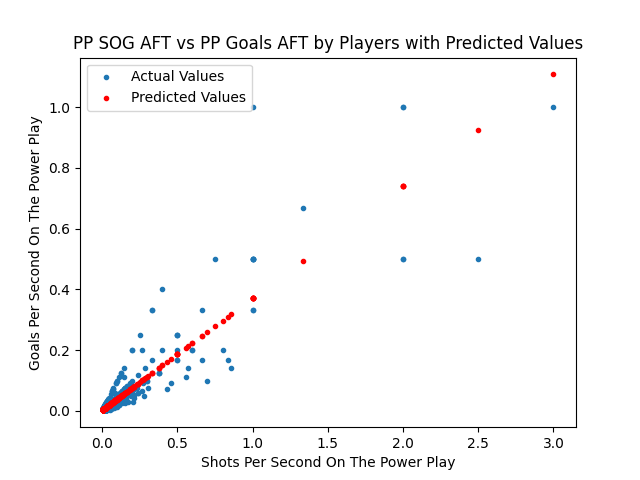
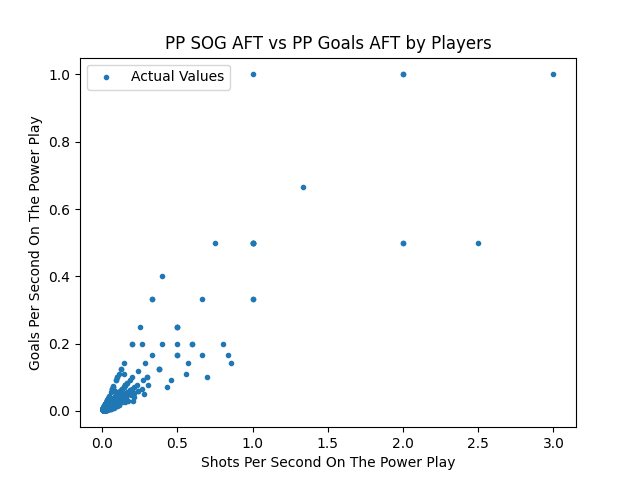


**RESULTS**

SOG AFT was determined to have a very weak correlation with goals scored AFT. The model had an R^2 value of approximately 0.11558077036955272, and the visual representation of the data and the predictions matches expected results for such a score.

****

On the other hand, PP SOG AFT was determined to have a very strong correlation with PP goals scored AFT. Impressively, that model had an R^2 value of approximately 0.8135615979554637, which was reflected in the graphs.



**CONCLUSION**

As is plainly visible through the R^2 value and through simple visual interpretation of the data, the amount of shots taken when prioritizing high-danger chances (in this case, when on the power play) is a much predictor of goals scored for any given player than simply the amount of shots on goal. It is important to note that, while power plays do by nature encourage this sort of play and likely increase the number of high-danger chances available, the adjustment makes these findings still viable for general play. These findings more or less refute the shot-heavy approach that some teams choose to follow and make a good case for playing more positionally, waiting for the right time to strike. As such, a team looking to maximize its performance should focus on intelligent positioning and passing in order to generate high-danger chances rather than mindlessly shoveling pucks toward the opposing team’s net.

**REFERENCES**

1. Dallas Stars vs. Edmonton Oilers, Western Conference Final game 6 “team stats”: <https://www.espn.com/nhl/matchup/_/gameId/401672988>

2. “NHL Game Data” dataset by Martin Ellis: <https://www.kaggle.com/datasets/martinellis/nhl-game-data>