Special Teams and the Success of Scoring a Goal

Rishika Cherivirala, Lauren Mok, Yinglin Zhao

1. Introduction

There are three types of penalties that can occur in a hockey game: a minor (two minutes), a double-minor (four minutes), and a major penalty (five minutes). Whenever a penalty occurs in ice hockey, the player that takes the penalty serves in the penalty box while their team plays shorthanded. Both of these teams play their special teams, a power play or a penalty killer. The team that takes the penalty becomes the penalty killer while the opposing team becomes the power play. A power play team has the advantage and typically plays their strongest offensive players in an attempt to maximize their scoring chances which are increased due to a greater number of players. The penalty killers play their best defensive play in an attempt to prevent the other team from scoring during the penalty. For a minor penalty, the special team ends once the power play team scores a goal. However, for major penalties, the special teams do not terminate once a player scores, it only ends when the five minutes have elapsed or the game ends.

In this paper, we attempt to analyze the impact of special teams and whether the number of power plays for one team increases their success in scoring a goal.

To investigate this, we first isolated the variables that directly correlated with the success of scoring a goal and modeled them against whether a shot was successful or not in an attempt to reduce the amount of noise in our multivariate regression model. We wanted to take into account all the other factors happening in the game instead of just power play to help decrease bias in our model, so we don't attribute goal success to only power plays. We later trained a multivariate regression model against having a power play advantage and the chance of scoring a goal based on other factors such as one-timer, traffic, and distance of the shot.

2. Methodology

To streamline our analysis and develop a multivariate regression model, we first conducted several simple regressions to identify the most influential variables related to goal success. Our focus was on factors such as shot type, presence of traffic, one-timer attempts, Euclidean distance to the goal, and shot angle. Subsequently, we developed these models to examine the influence of power plays on goal success.

To investigate the potential impact of shot type on goal success, we specifically queried the dataset for events categorized as shots or goals. Given that our variables were categorical in nature, we transformed shot types into dummy variables to effectively represent this categorical data. Additionally, boolean variables were converted to integers to facilitate analysis.

Subsequently, we constructed an Ordinary Least Squares (OLS) regression model with shot type as the independent variable and goal outcome as the dependent variable to examine the relationship between these two factors. OLS regression was chosen due to its suitability for analyzing the linear relationship between a discrete dependent variable (goal outcome) and one or more independent variables (shot types). By employing OLS regression, we were able to quantify the influence of different shot types on the likelihood of goal success, providing valuable insights into the dynamics of goal scoring in our analysis. This regression analysis revealed the following insights about the p-values associated with the shot types. These p-values later helped us filter out the significant shot types that are not correlated with goal outcomes. This model is dubbed the **Shot Type Model**.

We also investigated whether or not a goal was successful to the details which are one-timer and traffic. The variables of interest were "Traffic" and "One-timer," representing whether traffic was present and whether the shot was a one-timer, respectively. We divided the dataset into training and testing sets using a 70/30 split. We then fitted a logistic regression model using the training data, with goal outcome ("Goal") as the dependent variable and traffic and one-timer shots as the independent variables. This logistic regression provided insights into the summary statistics, including coefficients and p-values. This model is dubbed the **One-Timer and Traffic Model**.

For our final simple model, we examined the relationship between goal success and spatial factors, specifically the Euclidean distance and angle from the center of the goal to the location of the goal attempt. These variables act as proxies for spatial considerations in goal prediction, enabling the model to capture both directional and proximity aspects of shot attempts. To prepare for this analysis, we calculated the Euclidean distance and angle using the x and y coordinates of the shot attempt and the reference point (189.0, 42.5), which we considered to be the center point of the goal. The shot angle was determined using the arctangent function, which computes the angle in radians based on the differences in Y and X coordinates between the shot location and a reference point. With these two covariates, we constructed a logistic regression model to explore the relationship between shot characteristics (Euclidean distance and shot angle) and goal outcomes through performance metrics and p-values. This model is called the **Distance and Angle Model**.

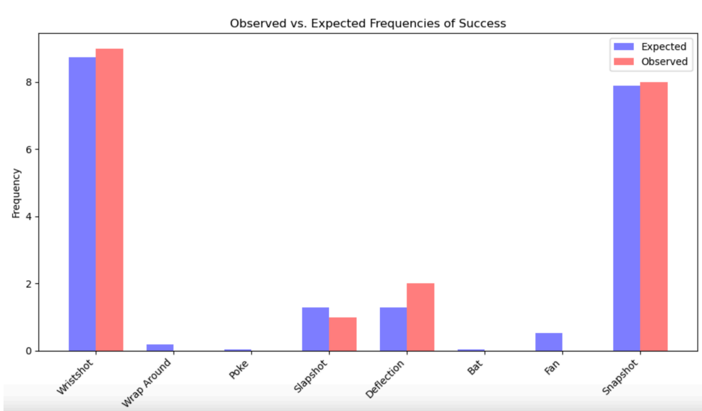
To create our final model to assess the impact of power plays on goal success we defined a binary variable, "Power Play," to indicate whether the team of the player attempting the goal was currently in a power play situation. The number 1 indicates that the team was in a power play while the number 0 indicates that the team was not in a power play. Using this, we calculated the success rate and the total number of power plays against the total number of goals in the dataset to determine the efficiency of power plays to make goals. We also included the three other significant covariates, euclidean distance, traffic, and one-timer, to create our final model. A custom function, "special_coder," was developed to encode each goal or shot entry with a binary value representing the team's power play status, taking the home team, home team skaters, away team skaters, and team affiliation of the player into consideration. We then assembled our predictor variables, including traffic presence, one-timer shots, shot distance, and power play status, into a feature matrix denoted as X_power. This matrix, along with the binary goal outcomes, served as input for our logistic regression model. After splitting the dataset into training and testing sets, we fitted a logistic regression model using the training data. The model was trained to predict goal outcomes based on the specified covariates. This model is dubbed as the **Power Play Model**.

Utilizing the Power Play Model, we wanted to identify the optimal combination of predictor variables for predicting goal outcomes. We devised a custom function, best_models, which systematically evaluates all possible combinations of predictor variables. This function iteratively removes one variable at a time and selects the model with the lowest AIC (Akaike Information Criterion) value. We used the AIC as a basis to select the better model because the AIC penalizes overly complex models, favoring simpler yet effective models that adequately explain the data without overfitting. By selecting the model with the lowest AIC value, we aim to identify the most suitable model that captures the essential patterns in the data while avoiding unnecessary complexity.

3. Discussion

3.1 Shot Type Model Results and Evaluation

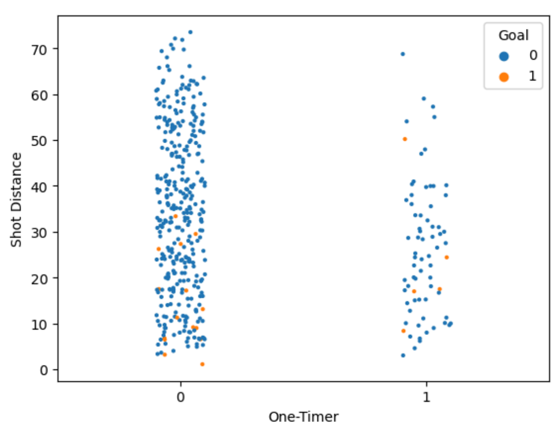
For this model, the p-values for all shot types are significantly higher than 0.05, indicating that these predictors are not statistically significant at the 95% confidence level. We did notice that snapshot and wrist shot types lead to more goals, but these two shot types accounted for more than 80% of the shot types. It's important to note that the lack of statistical significance doesn’t necessarily mean that these shot types have no impact on goal outcomes in reality. The lack of observations for the other categories likely led to the high p-values for each respective category. However, for the sake of our analysis, we are only considering p-values that are less than 0.05 to include in our final model, so for this reason, the shot type was not included as a predictor in our final model.

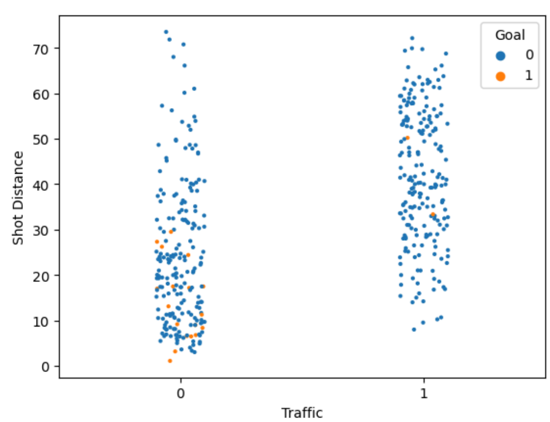


3.2 One-Timer and Traffic Model Results and Evaluation

From the summary of the logistic regression, we have that both of these variables are significant. If there is traffic, assuming all else equals, the change of log odds is approximately -4.8497. In other words, we can conclude that if there is traffic, the model predicts an approximately 0.78% increase in the odds of a goal being made. If the shot is a one-timer, assuming all else equals, the player shooting shot the goal with a slapshot right after receiving a pass, the change of log odds is approximately -1.6932 which means the model predicts an approximate 18.39% increase in the odds of a goal being made. The presence of traffic significantly reduces the likelihood of scoring a goal. Coaches and players could use this insight to strategize ways to create space and minimize traffic in front of the goal to increase goal-scoring opportunities. While one-timer shots are often associated with quick and unexpected shots, the model suggests that they are less likely to result in goals compared to other shot types.

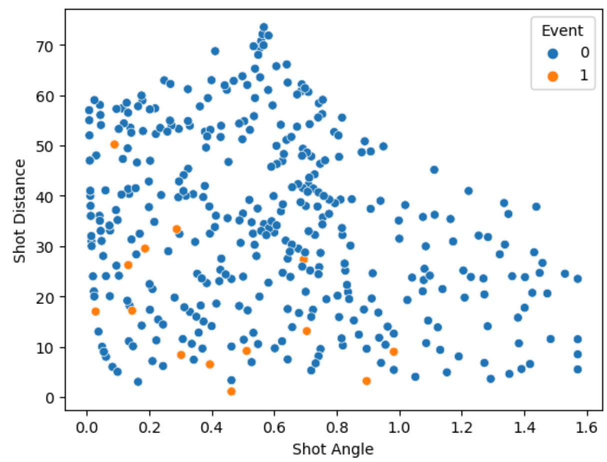
Initially, we hypothesized that in most cases, players do one-timer shots due to the presence of traffic and a lack of players open to receive a pass. However, the correlation between traffic and one-timers is approximately -0.0699, which indicates the relationship between the two variables is very weak and negative. We also found a negative weak correlation between distance to the goal and one-timers, which suggests that it is not due to the close distance that players decide to do one-timers. The left graph below further supports that; most of the goals made for a one-timer shot are when the Euclidean distance is below 30. In this case, if traffic does not exist and the distance is not close to the goal, teams may want to reconsider relying solely on one-timers and explore alternative shooting strategies to maximize goal-scoring potential.
In addition, we can observe from the graph on the right that for traffic, the rate of success is very low and most goals are made when there does not exist any traffic and the distance is below thirty. Of the twenty goals made and the hundreds of attempted shots, only two goals were made given traffic was present. Therefore, the defense team could consider having defense members closer to the proximity of the goal, and the offense team could work on spreading out and passing the puck instead of making a shot attempt.





3.3 Distance and Angle Model Results and Evaluation

From the summary of the logistic regression, we have that shot distance is a statistically significant predictor with a p-value of 0.006, however, shot angle is not statistically significant at a 0.05 significance level with a p-value of 0.094. The coefficient of shot distance to goal success is -1.0950, meaning that as Euclidean distance increases by 1 unit, the probability of scoring a goal increases by approximately 33.45%. Knowing this, coaches and players can strategically position themselves during gameplay and prioritize getting closer to the goal before taking a shot. Additionally, while shot angle did not show statistical significance in this model, players may still benefit from considering optimal shooting angles to increase their chances of scoring. The figure below also shows this, the goals– marked by the orange dots– are clustered at around lower values of shot distance and angle.



3.4 Power Play Model

While the significance of the other variables can be found through the other models, we will focus on the importance of power play on the success of goals in this analysis. The coefficient for power play situations is approximately -0.9267, however, it is not statistically significant (p-value = 0.388). This implies that being in a power play situation may not significantly affect the likelihood of scoring a goal, according to this model. Furthermore, to further analyze this we will try to provide a quantitative measure of the effectiveness of power play opportunities in scoring goals. The success rate of power plays is approximately 23.81% and it takes approximately 4.2 power plays to score one goal. This means that approximately 23 out of 100 power plays result in a goal being scored and the average number of power play opportunities a team needs to convert into a goal is 4.2 opportunities. This statistic provides insight into the efficiency and effectiveness of a special team's power play unit, meaning that these hockey teams might have trouble taking advantage of these opportunities. This suggests that the penalty killers who defend against power plays are relatively effective in preventing goals during these situations. They can disrupt the offensive plays of the opposing team and limit their scoring opportunities despite being at a numerical disadvantage.

The AIC analysis reveals that shot distance consistently emerges as a statistically significant predictor across various models. Specifically, it consistently exhibits p-values below the conventional significance threshold of 0.05, indicating its strong association with goal outcomes. Moreover, among all the models considered, the simplest regression model containing only shot distance achieves the lowest AIC value of 145.3855. This highlights the importance of considering shot distance in predicting goal outcomes.

4. Conclusion

A power play is a pivotal opportunity eagerly sought by both teams, often serving as a game-changing factor in determining victory. Using the player tracking data produced by the 2024 Big Data Cup, we developed a logistic regression to see the impact of special teams and whether the number of power plays for one team actually does increase a team's success in scoring a goal. To do this we took into account other factors that contribute to scoring a goal like shot distance, shot type, one-time, traffic, and shot angle. Based on our results, we saw that other variables except power plays play more of a role in increasing goal success. Shot distance, the distance from the player to the center of the goal, plays a significant role in determining the success of a goal. Players without possession of the puck should look to spread out closer to the goal to assist their team in getting closer to the goal, which increases the chance of shot success. In particular, if a team is in a power-play state, the players should take numerical advantage of their current situation and move further apart from each other to better advance their distance to the goal. By integrating these insights into their training and strategic planning, hockey teams can enhance their goal-scoring efficiency and increase their chances of success on the ice.

5. Appendix - Code:

All our code for the models can be found:
https://github.com/yzhao2433/big_red_cup24/tree/main