**Predicting NHL Playoff Success and Financial Impact**

Helena Mabey

BANA 6760-H01 Spring 2025

Dr. Martin Sabo

This analysis reviewed the current 2024–2025 NHL season using predictive modeling tools and correlation analysis. The intent of this study was to forecast the Stanley Cup Champion along with the estimated financial impacts of playoff performance. Logistic regression, Random Forest, and linear regression models were applied to historical team and season data to attempt to predict playoff success leading to a predicted 2025 Stanley Cup Champion. Z-scores, correlation, and a regression equation were also utilized to better understand patterns in wins, Elo ratings, and revenue impacts. Results highlight the significant economic value tied to playoff success, especially for the teams and home cities directly, which provides insight into the financial impact of professional hockey.

Around the world, people are captivated not just by sports, but by the numbers behind them. This dates back as far back as the beginning of baseball. “The first newspaper box score, which included five categories of statistics for each player – runs, hits, putouts, assists, and errors – was published …in 1859” (Silver, pg.95). Now, with the explosion of sports betting platforms and real-time, data-driven analysis, we see how analytics now shape every play, every game, and every season. As professional sports leagues grow, so does the financial impact, not just for the teams on the ice or on the field, but for the cities that rally behind them. In this analysis, I will use predictive analysis, along with other statistical calculating tools, to predict the National Hockey League (NHL) 2024-2025 Stanley Cup Champion and estimate the potential increased revenue for the team and home city. Because this increased revenue has large potential impacts, it is important for teams and cities to understand how past performance can predict future performance and earnings. It can play a part in city infrastructure and planning, league expansion, and marketing budgeting.

**Pre-Game Prep**

Before puck drop, teams need to prepare for each game. This section outlines the data preparation and transformation used for this analysis. Historical NHL data is available and was sourced for this review through the league’s direct API. Because the NHL did not provide explicit instructions for utilizing their API, a personal GitHub page was utilized as documentation for the Python coding needed to gather this data (Zmalski, pg.15). The API used accessed the team schedules and scoring data for each season. The time period reviewed for this analysis was from the 2020-2021 season through the current 2024-2025 season. Data points obtained included team name and official abbreviation, games played, wins and losses, as well as goals scored and allowed. The NHL data was reported through April 27, 2025. Revenue information for both the individual teams and home cities was AI-generated using estimations of revenue based on known team playoff performance. Revenue is captured in millions of dollars for teams and billions of dollars for home cities. All revenue data provided for cities and teams is hypothetical and was created for the sole purpose of this analysis.

Additionally, Python was used to create an Elo rating system. The Elo rating system was originally created by Arpad Elo to rate chess but it works in any competition where there are two players or two teams (Ryder, pg.1). This rating system gives each team a rating which changes over time based on wins and losses. The base score used in this analysis is 1500, which is common in Elo ratings (Silver, pg.1). Points are added or removed depending on the outcome of each game, adding in factors such as home-ice advantage and a K-Factor. The Silver example utilized home-ice advantage and a playoff adjustment in their Elo formula (Silver, pg.2). The K-Factor is “a fixed parameter that determines how quickly ratings should react to new game results… The higher the K-factor, the more a team’s rating changes based on any individual game’s result” (Silver, pg.2). A key feature I found valuable in their Elo scoring was that it “not only cares if you win, but how you win” (Silver, pg.2). The Ryder article provides a good summary of historical Elo scores for NHL teams and even provides information on the “Greatest Teams Ever” based on this system (Ryder, pg.4). This was interesting to read because it gives some insight into what the scores mean, not just the number itself. It calls out the ‘gold standard’ rating of 1,600 points and a ‘platinum standard’ of 1,650 points (Ryder, pg.4). These insights solidified my thoughts to add an Elo rating into this analysis.

The Elo model used in this analysis is much more basic than what is described in detail in the Silver or Ryder articles. Using AI, an Elo calculation was developed in Python using a base Elo value of 1500, a home-ice advantage value of 35, and a K-Value of 20. Using the historical NHL data obtained, each team was assigned Elo scores. This information was used as an additional data point for the prediction.

**Power Play Predictions**

The section shows the methods used for predictive modeling to forecast playoff success and, ultimately, the Stanley Cup Champion. The first method attempted was a logistic regression model. This model is used for binary comparisons (Brownlee, pg.1). The example used in this reference is determining people’s sex as male or female given their height (Brownlee, pg.4-5). This isn’t related to sports directly uses a unique classification, say games won, to determine if “an input (X) belongs to the default class (Y=1)”, in this case Stanley Cup Champion (Brownlee, pg.4). Unfortunately, this method yielded no potential winner. This is likely because the determining factor, number of prior games won, was not key in determining the very limited defined number of Stanley Cup Champions (four for this data set).Ultimately, it was unable to provide a complete analysis. If there were more examples of the default class, Stanley Cup Champions, this model may have been able to determine if a team would be a future winner. Another method tested was a Random Forest model. I reviewed an existing Random Forest model created by Christian Lee which showed that this method could be useful in prediction using NHL data as he referenced the same dataset used for this analysis (Lee, pg. 3). He utilized “Summary, Faceoff Percentages, Miscellaneous, Penalties and Shot Attempt Percentages” for his analysis (Lee, pg. 3). Unfortunately, the Random Forest model using my dataset had similar struggles to the logistic regression model and the results were disregarded. This model may have been useful if the dataset was considerably larger than the few data points and time period selected for this analysis.

The most reliable method for determining the potential 2024-2025 Stanley Cup Champion was a multiple linear regression model. This model seemed like the best fit because it uses multiple different features to determine a potential champion. Using AI to assist with the Python coding, models were created to attempt to predict the winning team. The independent variables used in the analysis were wins, goal difference, win percentage, Elo as well as three transformed features: goal efficiency (goal difference / wins), offensive power (goal difference / wins\*2), and Elo \* win percentage. Using these features, the most efficient model was able to perform with an R-squared value of 69.6% and an adjusted R-squared value of 68.1% meaning that there is a moderately strong relationship between the included variables and the number of playoff games. The skewness value of 0.428 shows a relatively evenly distributed dataset. The kurtosis value of 5.854 means there may be some pretty extreme outlier residuals, which is expected in sports values. Additionally, a value of 2.069 for Durbin-Watson shows that there is little autocorrelation in the residuals meaning that the errors are relatively independent and random, a strong indicator of reliable regression results.

Figure 1: OLS Regression Results

A screenshot of a computer

AI-generated content may be incorrect.

In addition to the regression model, a correlation matrix was created the show the relationships between variables. The majority of the variables have a strong positive correlation value, nearly 1, showing that many of the variables tend to increase together. For example, Elo increases as win percentage increases which is an expected behavior. This shows that these variables have strong value in this predictive model.

Figure 2: Correlation Matrix

A table with numbers and letters

AI-generated content may be incorrect.

Another interesting analysis was evaluating the Z-scores for the teams based on number or wins and Elo scores for the current season. The scores showed a strong split between the higher and lower performing teams but both scores were not necessarily consistent for each team, meaning that a high wins Z-score did not always correspond to a high Elo Z-score. The two top teams, Toronto (TOR) and Winnipeg (WPG) led the Z-score ratings, WPG by wins and TOR by Elo rating. There was more variability between higher performing teams. Some higher performing teams were more moderate (near the mean) in wins but had an Elo Z-score near or above 1. Lower performing teams were usually below the mean in both wins and Elo. Reviewing these scores added validity to the predicted champion determination and provided insight on how each team measured up against the mean wins and Elo rating values.

Figure 3: Z-Scores

A table of numbers with black text

AI-generated content may be incorrect.

**Winning the Cup**

Once the models were in place, it was time to see how all the numbers played out. The final winner was actually too close to call. Based on the analysis, the predicted Stanley Cup Champion for the 2024-2025 season will be either the Toronto Maple Leafs or the Winnipeg Jets. The linear regression model identified features such as wins, Elo ratings, and goal differential as very strong indicators of playoff success. Both teams scored strongly in these categories. Additionally, Z-scores showed Toronto leading in Elo ratings while Winnipeg led in wins while each was in second position for the other score. Because both teams performed strongly with similar results, it would be difficult to call out one team as a solid predicted winner.

There are also outside variables that may impact this prediction that were not included in this analysis. Two major excluded factors are team injuries, which can happen at any time, and team composition. There are many team trades mid-season which would impact the team’s performance during later games. This could change the statistics drastically but possibly only for a smaller number of games. Additionally, there is an element of luck involved with winning team sports that cannot be accounted for in any analysis. In hockey, a team may get lucky bounces or one-off poor officiating which also could impact the outcome of a game. Even with these potential impacts, this analysis is well-founded on commonly examined statistics.

**The Cup Payoff**

Winning in the playoffs means more for teams and cities than just prestige, it means increased revenue as well. There are many positive financial benefits to teams making the playoffs and winning the championship. A historical analysis showed a clear link between playoff advancement and financial gains. On average, teams that made it further into the playoffs increased their revenues by between $10–20 million annually. The home cities associated with those teams saw additional economic activity of $2–4 billion across the metro areas (hypothetical generated data).

For the 2024–25 projections, Toronto was forecast to experience the highest team revenue increase at approximately $75.2 million and the highest city economic impact at $15.5 billion if they reached predicted playoff milestones.

Winnipeg, as the other top contender, was projected to generate an estimated $73.3 million in additional team revenue if they achieved similar playoff success. The economic impact for Winnipeg was estimated around $15.1 billion. Winnipeg has a considerably smaller population, 834,678 metro-area population as of 2021 where Toronto’s metro area population was over 6 million in the same timeframe. This explains why Winnipeg’s revenue gains would be lower; however, it still highlights the value in post-season play on team and city economies.

Regression analysis provided a link between playoff games to team revenues. The regression equation is: Predicted Revenue = 4.70 × Playoff Games. Additionally, correlation coefficients between playoff games and revenue increases were near 1.0, further validating the strength of these relationships.

Outside of the team and city increases, local businesses and the general population see benefits of successful teams. Hotel occupancy rises and restaurants are full of patrons supporting their teams. Staff at these restaurants in turn can potentially see increased tips and larger order sizes as people tend to go out more during seasons when their teams play well.

**Postgame Analysis**

Every game ends with a review of what worked, what didn’t, and what was learned. This analysis has shown that predictive analysis, along with other statistical calculating tools, can potentially predict the NHL 2024-2025 Stanley Cup Champion and estimate the potential increased revenue for the team and home city. Some methods proved less effective with the datasets. After trial and error, a method was found to get the answers we were looking for. While the final team analysis was too close to call, strong evidence was presented that either Toronto or Winnipeg will be raising the Stanley Cup at the end of the season. Using the speculative revenue information, it can be gathered that winning the championship, or even making it further into the playoffs, offers significant financial benefits to both the teams and the cities and fans who support them.

Sports, regardless of type, offer both measurable and intangible benefits, for teams, cities, and the fans who support them. Financial gains can be seen through revenue and economic impact, but another major value of sports also lies in community pride, shared experiences, and the cultural identity that teams shape. The dedication these athletes show season after season is matched only by the loyalty of fans and the continued investments of cities working to grow in professional sports.

As professional leagues grow and the potential financial impacts increase, continued research and predictions become even more valuable. By combining analysis with an understanding of human nature, cities and teams can better maximize both economic growth and fan loyalty.

**Overtime Thoughts**

This analysis hit the ice strong but there are some areas for further review. The NHL API has data on each player season to season. It would be interesting to see how player statistics change year over year and even when they change teams. Additionally, while this analysis touched on the retail impact outside of team and city revenue, it would be interesting to use real-world data to not only see how the teams and cities benefit from playoff runs, but how local businesses experience this impact as well.

Hockey and other sports are part of world culture. Fans share the joy and pain of their favorite teams year after year. Using predictive data analysis allows for teams and cities to maximize their revenue while growing the culture and support of the loyal fans.

**Bibliography**

* Brownlee, Jason. Logistic Regression for Machine Learning. Machine Learning Mastery, 2016. [https://machinelearningmastery.com/logistic-regression-for-machine-learning/], pg.1, pgs.4-5
* Lee, Christian. Identifying the Best Predictors of NHL Game Outcomes Using Random Forest, 2021 [https://medium.com/hockey-stats/identifying-the-best-predictors-of-nhl-game-outcomes-using-random-forest-b4c11f46bc97], pg.3
* Ryder, Alan. Elo Rating for the NHL. HockeyAnalytics.com, 2016. [https://hockeyanalytics.com/2016/07/elo-ratings-for-the-nhl/#:~:text=And%20that%20has%20led%20to,team%2Don%2Dteam%20games.], pg.1, pg.4
* Silver, Nate. The Signal and the Noise: Why So Many Predictions Fail — But Some Don't. Penguin Press, 2012. pg. 95
* Silver, Nate. How Our NHL Predictions Work. FiveThirtyEight, 2023. [https://fivethirtyeight.com/methodology/how-our-nhl-predictions-work/], pg.2
* Zmalski, NHL API Reference. GitHub, 2024. [https://github.com/Zmalski/NHL-API-Reference], pg.15

**AI Statement**

I utilized ChatGPT and Claude to assist in creating the Python code to complete API requests, financial datasets, regression modeling, correlation determination, Z-score analysis, and Elo team ratings. AI was also used to assist in visualizations for the presentation using Python and Tableau.

Supporting documents are located here: [Predicting NHL Success](https://github.com/helenamabey/Predicting_NHL_Success)