S&DS 425/625 Capstone Report

Mason Dugas

December 16, 2024

Abstract

This analysis investigates the differences in performance between starting and backup goalies in the NHL, focusing on their ability to prevent goals based on shot location. The motivation stems from evaluating whether starting goalies consistently outperform backups across various spatial clusters on the ice and understanding if opponent offensive strategies significantly impact goalie performance.

The data includes NHL shot-level and team-level statistics from 2008-2024, sourced from MoneyPuck. Observations consist of shot attempts across distinct spatial zones (clusters), game-level outcomes, and goalie-specific performance metrics such as save percentage and goals allowed.

A zero-inflated Poisson (ZIP) model was developed to account for goal-scoring data, which has a high frequency of zero outcomes (no goals). Predictors included offensive shot statistics, goalie-specific metrics, and spatial zones. The model outputs the expected number of goals scored while accounting for cluster-specific performance variations for each goalie.

The analysis reveals that starting goalies generally outperform backups across all teams, as evidenced by 100% starter win rates for most teams in simulated matchups. However, specific cases like the LA Kings and Pittsburgh Penguins highlight outliers where backups performed comparably or better due to situational factors such as poor starter form or unusual roster changes. Notably, in these cases, the same goalie won every simulated game for their team. The simulated data consists of hypothetical matchups where each team's starting and backup goalies faced every other team's offense, using predicted goals scored based on shot-level data and goalie performance metrics derived from the fitted model.

The analysis reveals that starting goalies generally outperform backups across all teams, as evidenced by 100% starter win rates for most teams in simulated matchups. However, specific cases like the LA Kings and Pittsburgh Penguins highlight outliers where backups performed comparably or better due to situational factors such as poor starter form or unusual roster changes.

The results suggest that while starting goalies are often the most reliable choice in the NHL, differences between starters and backups are minimal, likely due to the overall polished skill levels of professional goalies. This type of analysis could yield more pronounced insights when applied to junior hockey leagues (e.g., BCHL, USHL), NCAA hockey, or minor professional leagues, where goalie weaknesses and strengths are more distinct.

Executive Summary

This report evaluates the performance differences between starting goalies and backup goalies in the NHL. The goal is to determine whether starting goalies consistently perform better than backups and to explore whether differences in performance could be linked to specific areas of the ice where shots are taken. This analysis provides insights that can help teams make more informed decisions about which goalie to rely on during games.

The Data

The analysis uses NHL game data from the 2017-2024 seasons, excluding 2020 due to COVID, including details about where shots were taken and wether they resulted in goals. To explore the possibility that goalies may perform better or worse from certain areas of the ice, the ice was divided into 12 zones based on shot locations. Simulated matchups were created where each team's starting and backup goalies faced every other team's offense, and the number of goals each goalie would allow was predicted.

Key Results

The results show that starting goalies consistently outperform backups, as they allowed fewer predicted goals in every simulated matchup. Importantly, the same goalie won every single simulated game for their team, regardless of the offensive strategies or shot locations they faced. This finding suggests that the use of spatial zones did not have a significant impact on the results, as the overall performance of a goalie determined the outcome more than their ability to handle shots from specific areas of the ice.

Conclusions and Future Considerations

This analysis confirms that NHL teams are correct in relying on their starting goalies, as starters consistently outperform backups. The results indicate that the use of shot zones did not reveal any meaningful differences between goalies, as performance was consistent across all zones. This is likely because NHL goalies are highly skilled and well-rounded, leaving little room for significant weaknesses to emerge in specific areas.

While the zones did not add value in this analysis, this approach could be more useful in junior hockey leagues, college hockey, or minor professional leagues, where goalies are still developing their skills. At these levels, weaknesses in specific areas of the ice may be more pronounced, making spatial analysis more impactful for evaluating goalie performance.

Future research could also explore other factors, such as game situations (like power plays or penalty kills) or opponent offensive strategies, to better understand what influences goalie performance. These additional layers of analysis could provide teams with even more actionable insights for decision-making.

Introduction

In hockey, goalies play a pivotal role in determining the outcomes of games, as their ability to stop opposing shots can make or break a team's success. Teams traditionally rely on their starting goalies, who are assumed to outperform backups in most situations. However, questions remain about whether starters consistently justify their status and whether backups can offer similar performance levels. The primary motivation for this study is to examine whether starting goalies consistently perform better than backups across all teams and to explore how performance differences might vary depending on shot location and opponent strategies. Understanding these factors can help teams make better decisions about goalie selection and optimize performance throughout the season.

Previous research on goalie performance has often focused on high-level metrics like save percentage or goals-against average but has not fully explored how these metrics interact with spatial shot data.

The data for this study was sourced from data provided by MoneyPuck, covering shot-level and team-level information from the 2008-2024 seasons. The dataset includes information about shot locations, game outcomes, and goalie performance metrics, such as save percentage and goals allowed. To explore whether certain goalies perform better in specific areas, the ice was divided into 12 spatial clusters, allowing for a detailed examination of performance across different shot zones.

The remainder of this paper is organized as follows: Section 2 presents an exploration of the data and visualizations, highlighting patterns in shot locations and goalie performance. Section 3 details the modeling approach, including the use of a zero-inflated Poisson (ZIP) model to predict goals allowed by starting and backup goalies across simulated matchups. Section 4 discusses the results of the simulations, showing that starting goalies consistently outperform backups, with the same goalie winning every matchup for each team. Finally, Section 5 provides conclusions, recommendations, and ideas for future work, including applying similar methods to junior or collegiate hockey leagues, where differences in goalie performance may be more pronounced.

This study contributes to the understanding of goalie performance by confirming the reliability of starting goalies in the NHL and exploring the potential value of spatial data for evaluating goalies at different levels of competition.

Data exploration and visualization

First, we need to decide what years to include in our analysis. As the game of hockey is constantly changing throughout the years, whether that is due to new rules, evolving coaching schemes, or young players having a new impact, we should not assume that the goal scoring environment in the NHL is consistent throughout all seasons.

This is where a Kruskal-Wallis test can be valuable. It allows us to compare the goal-scoring distributions across seasons without assuming they follow a normal distribution. By identifying whether significant differences exist, we can better decide which seasons to include in our analysis. Here, we are using the goal totals for each team in each game from seasons 2008 - 2024.

```
    Kruskal-Wallis rank sum test

data:  goalsFor by season
Kruskal-Wallis chi-squared = 384.17, df = 16, p-value < 2.2e-16
```

The Kruskal-Wallis test indicates significant differences in the distribution of goals scored across NHL seasons. This suggests that goal-scoring is not consistent across seasons. To explore these differences further, we will a Poisson regression model using 2024 as the reference level for season, allowing us to identify which seasons are not significantly different in goals scored per game. This model assumes a Poisson distribution for the data, which we will confirm later. If the variance of goals scored exceeds the mean, a negative binomial regression would be a better fit but should yield similar results for identifying significant differences between seasons.

```
Call:
glm(formula = goalsFor ~ season, family = poisson, data = all_teams)

Coefficients:
             Estimate Std. Error z value Pr(>|z|)    
(Intercept)  1.111122   0.022299  49.827  < 2e-16 ***
season2023   0.009857   0.024773   0.398 0.690689    
season2022   0.034124   0.024716   1.381 0.167390    
season2021   0.023246   0.024740   0.940 0.347407    
season2020  -0.051968   0.026065  -1.994 0.046173 *  
season2019  -0.029708   0.025242  -1.177 0.239222    
season2018  -0.023120   0.024926  -0.928 0.353635    
season2017  -0.035528   0.024963  -1.423 0.154667    
season2016  -0.112094   0.025240  -4.441 8.95e-06 ***
season2015  -0.130624   0.025283  -5.166 2.39e-07 ***
season2014  -0.135661   0.025302  -5.362 8.24e-08 ***
season2013  -0.125204   0.025264  -4.956 7.20e-07 ***
season2012  -0.140738   0.027062  -5.201 1.99e-07 ***
season2011  -0.138543   0.025316  -5.472 4.44e-08 ***
season2010  -0.104084   0.025216  -4.128 3.66e-05 ***
season2009  -0.088273   0.025173  -3.507 0.000454 ***
season2008  -0.067120   0.025124  -2.672 0.007551 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)
```

```
    Null deviance: 46337  on 41521  degrees of freedom
Residual deviance: 45914  on 41505  degrees of freedom
AIC: 157940

Number of Fisher Scoring iterations: 5
```

In the initial model, we observe that for the years 2017–2023, there is no significant difference in goals scored per game, with the exception of the shortened 2020 season due to COVID. To validate this finding, we will now fit a model using only these seasons (excluding 2020) to confirm that there is no significant difference in goals scored per game across these years. This step will further reinforce the consistency in goal-scoring during this period.

```
Call:
glm(formula = goalsFor ~ season, family = poisson, data = all_teams %>%
    filter(as.numeric(as.character(season)) > 2016, as.numeric(as.character(season)) !=
        2020))

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.111122   0.022299  49.827   <2e-16 ***
season2023   0.009857   0.024773   0.398    0.691
season2022   0.034124   0.024716   1.381    0.167
season2021   0.023246   0.024740   0.940    0.347
season2019  -0.029708   0.025242  -1.177    0.239
season2018  -0.023120   0.024926  -0.928    0.354
season2017  -0.035528   0.024963  -1.423    0.155
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 18973  on 16913  degrees of freedom
Residual deviance: 18937  on 16907  degrees of freedom
AIC: 65779

Number of Fisher Scoring iterations: 5
```

As we can see, non of the `season` coefficients are significant, indicating a similar goal scoring environment in these seasons.

To assess whether the data fits the assumptions of a Poisson regression model, we will calculate the mean and variance of goalsFor. If these two values are approximately equal, it suggests that the Poisson distribution is a reasonable assumption for our data. However, if the variance significantly exceeds the mean, it indicates overdispersion, and we may need to consider alternative modeling approaches, such as a quasi-Poisson or negative binomial regression. This simple check will guide our choice of model moving forward.

```
[1] "Mean of goals scored per game: 2.85"
```

```
[1] "Variance of goals scored per game: 2.82"
```

Since the mean of `goalsFor` is only 0.03 greater than the variance, the assumption of equal mean and variance required for a Poisson distribution is reasonably satisfied. This indicates that fitting a Poisson regression model is appropriate for this data. We can proceed with this approach, confident that the model aligns well with the underlying distribution of the data.

However, since our hypothesis is that goalies have on-ice zone strengths and weaknesses—areas where they are more or less likely to prevent goals based on shot location compared to other goalies—it is crucial to analyze the data at a more granular level. To do this, we need to split the ice into clusters (zones) and calculate the mean and variance of `goalsFor` within each cluster. This will help us assess whether the Poisson distribution assumption holds within these specific zones and guide our modeling choices for analyzing goalie performance in different areas of the ice.

At first glance, manually defining zones on the ice might seem like a straightforward approach to analyzing goalie performance. However, there are more robust methods, such as clustering algorithms, that can objectively identify meaningful patterns in the data and better capture the nuances of shot location and scoring probability.

This is why a more dynamic and probablistic method is better suited for this analysis. GMM is the best choice for this task because it provides a flexible, probabilistic approach to clustering that can accommodate the inherent variability in shot locations and their associated likelihood of resulting in a goal. Unlike hard clustering methods, GMM assigns probabilities to each observation for belonging to different clusters, which aligns well with the continuous and nuanced nature of shot data in hockey.

The probability density function for a GMM is defined as:

$$p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

Where:

- $\mathbf{x}$ is the data point (in this case, (`xCordAdjusted`, `yCordAdjusted`)).
- K is the total number of clusters (here, $K = 12$).
- π_k is the mixing coefficient for cluster k, where $\sum_{k=1}^{K} \pi_k = 1$.
- $\mathrm{N}(\mathbf{x} \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ is the Gaussian distribution for cluster k, with:
  - $\boldsymbol{\mu}_k$: Mean vector of the cluster (centroid in $(xCordAdjusted, yCordAdjusted)$ space).
  - $\boldsymbol{\Sigma}_k$: Covariance matrix describing the spread and orientation of the cluster.

The Gaussian distribution is defined as:

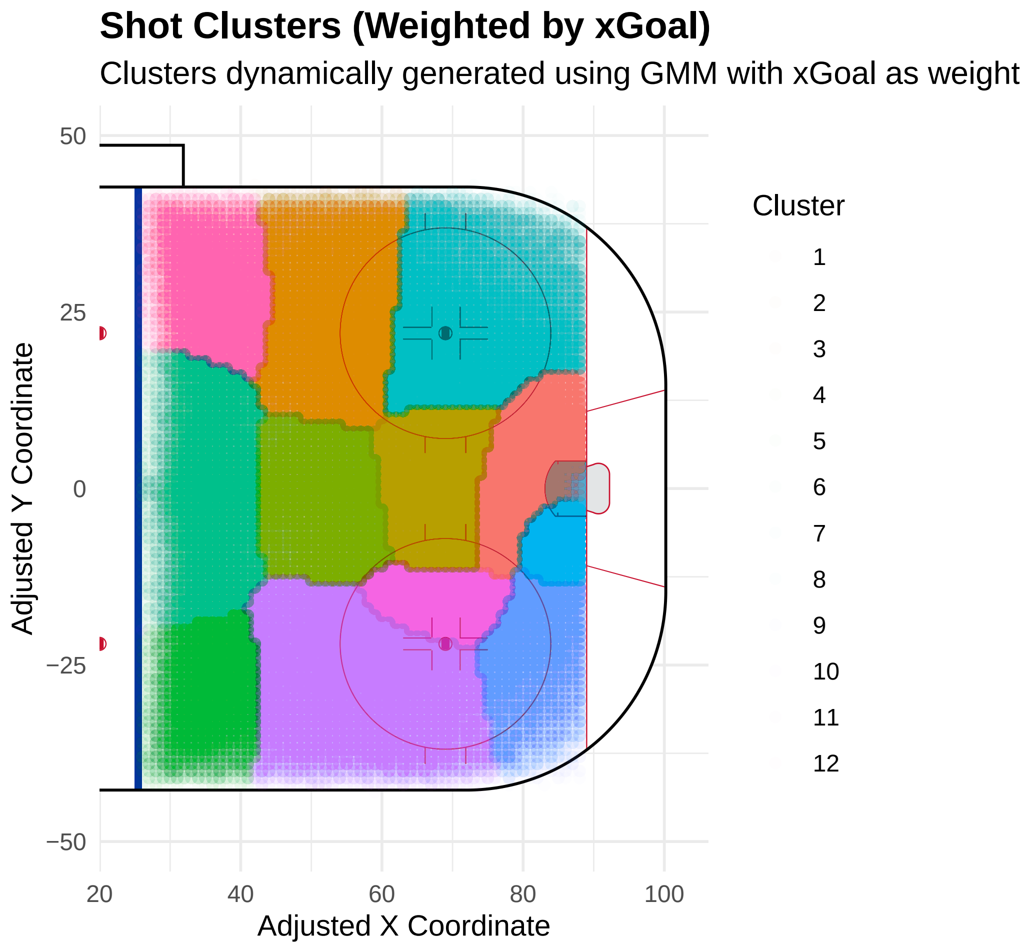
$$\mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \frac{1}{(2\pi)^{d/2} |\boldsymbol{\Sigma}_k|^{1/2}} \exp\left(-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_k)^\top \boldsymbol{\Sigma}_k^{-1} (\mathbf{x} - \boldsymbol{\mu}_k) \right)$$

Where:

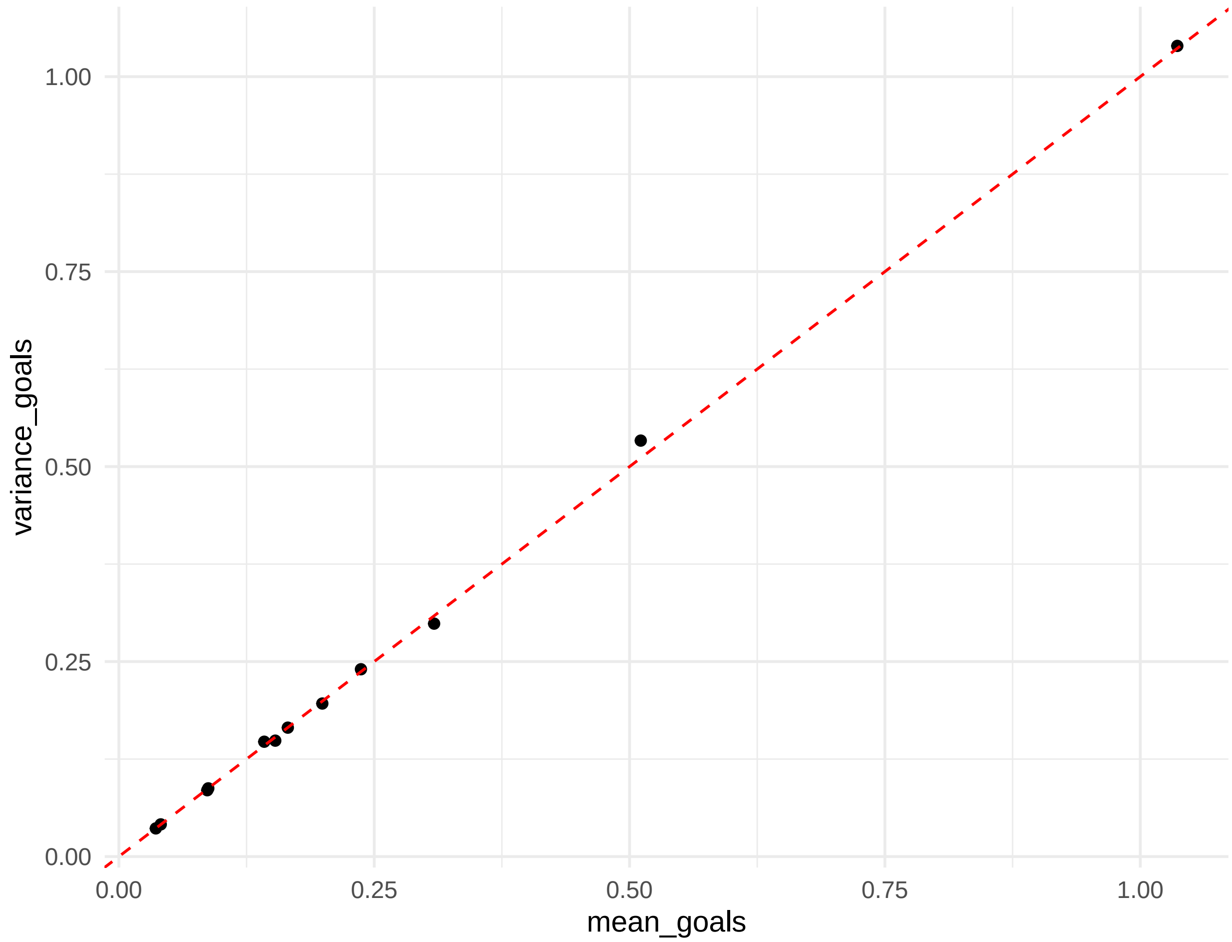
- d is the number of dimensions (here, $d = 2$ for `xCordAdjusted` and `yCordAdjusted`).
- $|\boldsymbol{\Sigma}_k|$: Determinant of the covariance matrix.
- $(\mathbf{x} - \boldsymbol{\mu}_k)^\top \boldsymbol{\Sigma}_k^{-1} (\mathbf{x} - \boldsymbol{\mu}_k)$: Mahalanobis distance from the data point $\mathbf{x}$ to the mean $\boldsymbol{\mu}_k$.

This Gaussian Mixture Model uses the shot coordinates xCordAdjusted and yCordAdjusted to determine spatial clusters, representing distinct zones on the ice. To account for the relative scoring probability of each shot, the standardized xGoal_standardized values are incorporated as weights, prioritizing shots with a higher likelihood of resulting in a goal during clustering. By specifying 12 clusters ($K = 12$), the model captures the diversity of offensive-zone shot patterns, providing a precise foundation for analyzing goalie strengths and weaknesses across different areas of the ice.

Here is a visual representation of what the clusterd zones look like:



To ensure that the assumptions for Poisson regression are met, we need to confirm that the variance of goals scored in each cluster is approximately equal to the mean. The plot below compares the mean and variance of goals for each cluster. The red dashed line represents where the mean equals the variance, which is the key assumption for Poisson regression.



The plot shows that the variance of goals in each cluster closely aligns with the mean, as most points lie near the red dashed line. This suggests that the Poisson regression assumptions are satisfied, making it an appropriate choice for modeling the data.

Modeling/Analysis

Zero-Inflated Poisson Model Analysis

Model Description

A **zero-inflated Poisson (ZIP)** regression model was used to analyze the data. The ZIP model is appropriate for count data with an excess of zeros, where the zeros can arise from two distinct processes: 1. Structural zeros (e.g., no goals occurred because no valid opportunities existed). 2. Zeros that arise naturally from a Poisson process.

The ZIP model combines: - A **count model** (Poisson regression) for positive counts. - A **zero-inflation model** (logistic regression) that predicts the probability of structural zeros.

Assumptions

- **Poisson Distribution**: The count component assumes that the mean equals the variance. If this is not satisfied (e.g., overdispersion), the ZIP model may need adjustment (e.g., to a negative binomial).
- **Independence of Observations**: Each observation is assumed to be independent.
- **Excess Zeros**: The model assumes that some zeros come from a different process than the Poisson distribution.

Observations, Predictors, and Outcome

- **Observations (Rows of X)**:
  Each observation represents a unique combination of game ID, shooting team, goalie team, cluster (shot zone), and game date.

- **Predictors (Columns of X)**:
  Predictors included variables related to team and goalie performance, pre-game averages, and shot statistics, such as:

  - `shotsOnGoal.offense`, `xGoals.offense`, `xFroze.offense`, `savePct.goalie`, `shots.goalie`, etc.

  - Categorical variable `home_or_away` to account for home vs. away performance.

- **Outcome (y)**:

  - **Count Component**: `Goals` (number of goals scored per cluster).

  - **Zero-Inflation Component**: Binary variable `zero_inflated` (1 if goals = 0, 0 otherwise).

Selecting Predictors

To identify the most important predictors for the model, LASSO (Least Absolute Shrinkage and Selection Operator) regression was applied. LASSO is a regularization technique that selects predictors by shrinking less important coefficients to zero, effectively reducing the model complexity and preventing overfitting. This method was used separately for both the count and zero-inflation components of the ZIP model to select predictors relevant to each process.

After applying LASSO, an initial model was fit using the selected predictors. To refine the model further, predictors with a significance level above 0.25 (based on p-values) were removed. This threshold was chosen to ensure that the final model focused on predictors that contributed meaningfully to explaining the variation in the data, while still allowing for exploratory insights. By iteratively refining the model, the final ZIP regression captured both the structural and count-based aspects of goal-scoring effectively, leading to more robust and interpretable results.

Zero-Inflated Poisson Model Formula

The ZIP model can be described as follows:

$$P(Y_i = y) = \begin{cases} \pi_i + (1-\pi_i)e^{-\lambda_i}, & \text{if } y = 0 \\ (1-\pi_i)\frac{\lambda_i^y e^{-\lambda_i}}{y!}, & \text{if } y > 0 \end{cases}$$

Where: - π_i: Probability of structural zeros, modeled using a logistic regression.
- λ_i: Mean of the Poisson distribution for positive counts.

Binary Component (Zero-Inflation)

The zero-inflation component predicts the probability of structural zeros:

$$\begin{aligned} \text{logit}(\pi_i) = -1 &+ \beta_1(\text{shotsOnGoal.offense}) + \beta_2(\text{xFroze.offense}) + \beta_3(\text{xPlayContinuedInZone.offense}) \\ &+ \beta_4(\text{shotsOnGoal.goalie}) + \beta_5(\text{xGoals.goalie}) + \beta_6(\text{xFroze.goalie}) \end{aligned}$$

Count Component

The count component models the expected number of goals:

$$\begin{aligned} \log(\lambda_i) = -1 &+ \gamma_1(\text{shotsOnGoal.offense}) + \gamma_2(\text{xFroze.offense}) + \gamma_3(\text{xPlayContinuedInZone.offense}) \\ &+ \gamma_4(\text{shotsOnGoal.goalie}) + \gamma_5(\text{xGoals.goalie}) + \gamma_6(\text{xFroze.goalie}) \\ &+ \gamma_7(\text{xGoals.offense}) + \gamma_8(\text{xShotWasOnGoal.offense}) + \gamma_9(\text{shots.goalie}) \\ &+ \gamma_{10}(\text{xPlayStopped.offense}) + \gamma_{11}(\text{goals\_scored}) + \gamma_{12}(\text{savePct.offense}) \\ &+ \gamma_{13}(\text{xPlayContinuedInZone.goalie}) + \gamma_{14}(\text{xPlayStopped.goalie}) + \gamma_{15}(\text{xShotWasOnGoal.goalie}) \\ &+ (1|\text{cluster:goalieNameForShot}) \end{aligned}$$

Coefficients Interpretation

- **Count Component**:
  Coefficients for predictors like `xGoals.offense` or `shots.goalie` represent the change in the expected number of goals for a one-unit change in the predictor, holding other variables constant.
  Example: A positive coefficient for `xGoals.offense` indicates that higher offensive expected goals increase the likelihood of actual goals scored.
- **Zero-Inflation Component**:
  Coefficients here influence the probability of a structural zero. Positive coefficients increase the likelihood of zero goals being structural (e.g., low shot quality increases zero outcomes).

Model Appropriateness The ZIP model is appropriate for this dataset because: 1. There is an observed excess of zeros in the goals data, which the ZIP model explicitly accounts for. 2. Poisson regression alone cannot distinguish between structural zeros and naturally occurring zeros, making the ZIP model a better fit. 3. The inclusion of both team and goalie-specific predictors, along with clusters, ensures that spatial and contextual information is properly captured.

Interpretability For a technical audience, the coefficients and model components (count and zero-inflation) are straightforward to explain. A detailed discussion of how predictors influence structural zeros versus actual goals can provide deep insights. For a non-technical audience, it may be helpful to focus on the overall implications: the model identifies key factors (e.g., shot quality, home/away status) that influence the likelihood of scoring and explains why certain games or clusters result in no goals. Visualizations of predicted vs. actual goals or spatial goal distributions can aid understanding.

With this regression model, we can examine the predicted scores to see if the predicted scores are similar to the training data.

```
[1] "Mean of predicted goals scored per game: 2.67"
```

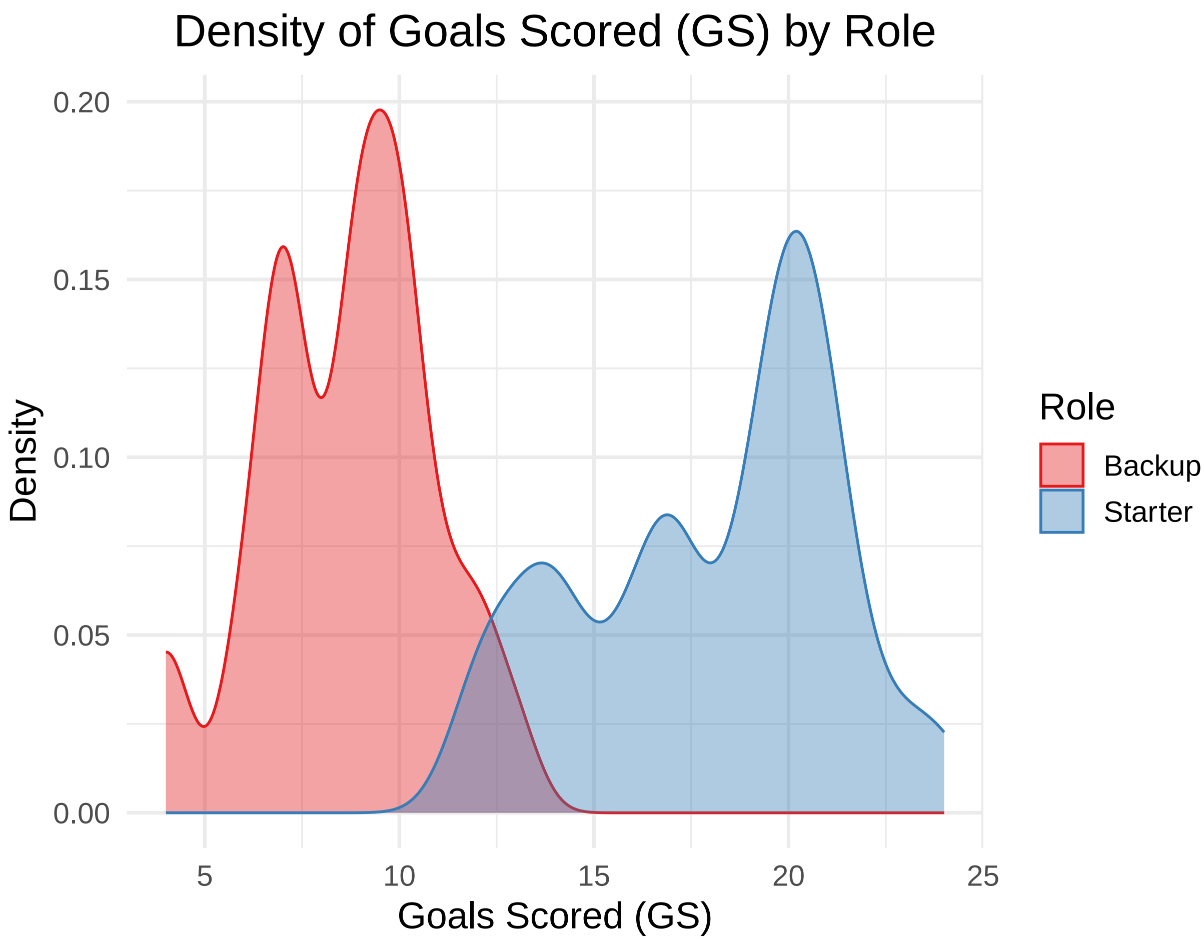
```
[1] "Variance of predicted goals scored per game: 2.24"
```

As we can see, while both the mean and variance of the predicted scores are slightly lower than those observed in the training data, the difference between the mean and variance remains small. Although there is a slight underdispersion, it would not be inappropriate to say that the predicted scores still follow a roughly Poisson distribution, as the mean and variance remain reasonably close. This suggests that the model is capturing the overall goal-scoring trends fairly well.

Visualization and interpretation of the results

Now that we have developed and evaluated our model, we can apply it to the 2024 season to gain further insights into goalie performance. A key point of interest is comparing starting goalies to backup goalies to see how their performance differs in terms of goals prevented or conceded. Understanding these differences can provide valuable insights into the impact of goalie depth on team success.

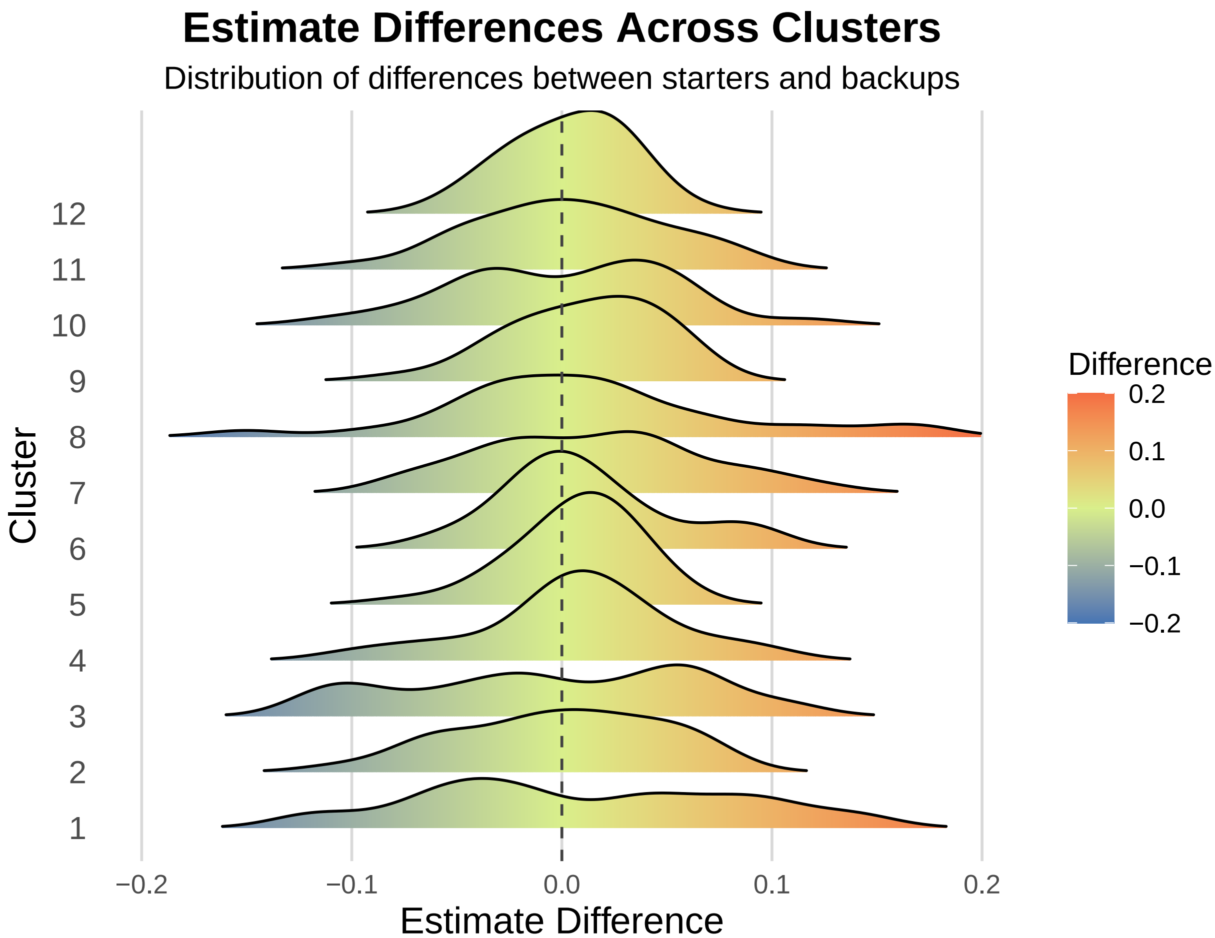
The following graph visualizes the distribution of goals scored (GS) for games involving starting and backup goalies. By examining these distributions, we can identify patterns in performance, such as differences in consistency or shot-stopping ability between the two groups.



The graph shows that starting goalies tend to have a tighter distribution with peaks at lower goals scored, indicating more consistent and stronger performance in preventing goals. In contrast, backup goalies exhibit a wider spread and higher density at larger goals scored, suggesting greater variability and potentially weaker performance. This highlights the importance of reliable goalie depth for teams to maintain defensive stability.

To further analyze the differences between starting and backup goalies, we calculate the Estimate Differences within each spatial cluster. These differences are derived from the mixed effects for each goalie by cluster, capturing zone-specific performance variations. Positive values indicate where starting goalies outperform backups, while values near zero suggest similar performance between the two groups.

The following ridge plot displays the distribution of estimate differences across clusters, with the dashed vertical line at zero representing no difference. This visualization allows us to identify specific clusters where starting goalies provide a measurable advantage over backups.



The plot shows the distribution of Estimate Differences between starting and backup goalies across spatial clusters. Positive differences (orange/red) indicate clusters where starting goalies outperform backups, while negative differences (blue) suggest the opposite. Most clusters show values centered near zero, with slight positive skews, indicating that starting goalies generally perform slightly better, especially in certain zones.

After analyzing the estimate differences across clusters, the next step is to simulate goalie performance in the 2024 season to evaluate the difference between starting and backup goalies. For teams with two goalies that have sufficient data (starter and backup), I created hypothetical matchups against every team in the league. Using the predicted goals scored from the previously fitted zero-inflated Poisson (ZIP) model, I calculate the total predicted score for both the starter and the backup goalie.

To determine how much of an impact goalies make, I compare their predicted performance by team. For each matchup, the starter wins if the predicted score (goals conceded) is lower than that of the backup goalie. I then calculate the following metrics for each team: - Starter Win Percentage: The proportion of games in which the starter outperformed the backup.
- Average Score Difference: The average difference in predicted goals conceded between the starter and backup across all matchups.

This approach allows us to quantify the advantage of starting goalies over backups, providing a team-by-team breakdown of goalie performance. The results will be summarized in a table to highlight the frequency of starter wins and the magnitude of the performance gap.

```
[1] "Mean of predicted goals scored in simulated games: 0.49"
```

The mean predicted score of 0.49 likely results from the zero-inflated Poisson (ZIP) model, which may overestimate the probability of structural zeros or produce conservative predictions. Despite this, the relative comparison between starter and backup goalies remains valid because both goalies are evaluated under the same conditions and model assumptions, ensuring that differences in predicted performance are meaningful and consistent.

goalie_team	total_games	starter_wins	starter_win_percentage	average_difference
NSH	31	31	100	-0.2202361
WPG	31	31	100	-0.2062398
TBL	31	31	100	-0.1917454
BOS	31	31	100	-0.1791310
NYR	31	31	100	-0.1760295
VAN	31	31	100	-0.1698142
COL	31	31	100	-0.1639276
DAL	31	31	100	-0.1590960
MTL	31	31	100	-0.1559849
MIN	31	31	100	-0.1548993
BUF	31	31	100	-0.1511851
NYI	31	31	100	-0.1504957
FLA	31	31	100	-0.1465153
CHI	31	31	100	-0.1404564
STL	31	31	100	-0.1302363
EDM	31	31	100	-0.1297051
CBJ	31	31	100	-0.1123792
SJS	31	31	100	-0.1092976
SEA	31	31	100	-0.1083547
CAR	31	31	100	-0.1077337
OTT	31	31	100	-0.1026880
DET	31	31	100	-0.0973297
ANA	31	31	100	-0.0912744
NJD	31	31	100	-0.0771872
TOR	31	31	100	-0.0634776
WSH	31	31	100	-0.0441448
LAK	31	0	0	0.0069121
PIT	31	0	0	0.0263289

Conclusions and Recommendations

The results of the simulated games clearly demonstrate that starting goalies consistently outperform backups, as evidenced by the 100% starter win percentage across nearly all teams. This suggests that teams are generally correct in selecting their starters, regardless of the offensive strategies or strengths of their opponents. The average score differences further emphasize this trend, with starters allowing fewer predicted goals than backups in every case except for specific anomalies, such as the Los Angeles Kings and Pittsburgh Penguins

For the Kings, the starting goalie is severely underperforming this season, with a noticeably lower save percentage and higher goals-against average compared to the backup. In contrast, the Penguins starting goalie situation is unconventional, as their “backup” has played more games than the designated starter. Tristan Jarry, a high-performing goalie who had a rough start to the 2024 season and was briefly sent to the AHL, has since returned to form and now significantly outperforms his teammate.

Future Directions

While this analysis strongly indicates that starters are almost always the superior option, it also raises important questions for further exploration:
1. Does opponent offensive strategy matter?
While starters perform better overall, further work could investigate whether specific types of offenses (e.g., high shot volume, low shot quality) influence goalie performance more than others.

2. Are current clusters optimal?
   The spatial clustering approach allows for evaluating goalie strengths and weaknesses by zone. However, refining clusters (e.g., considering shot type, game state, or shooter talent) could further improve the analysis.

3. Goalie evaluation over time
   Future iterations of this model could dynamically update goalie rankings throughout the season to account for performance trends, fatigue, or injuries, ensuring more responsive decision-making for teams.

Broader Implications

Because NHL goalies represent the absolute best talent in hockey, their skills are highly polished, and their weaknesses are minimal, making it challenging to find significant gaps in performance. However, this type of analysis could be even more impactful when applied to junior hockey leagues (e.g., BCHL, USHL), NCAA-level play, or minor professional leagues, where goalie skill gaps and spatial strengths/weaknesses may be more pronounced. At these levels, goalies are still developing their technical abilities and game awareness, which could reveal clearer insights into areas of improvement and provide valuable information for player development and scouting.

References

MoneyPuck

Hockey Reference

The GitHub repository will all code and data can be found here.