NHL Expected Goal Model using Logistic Regression with “Shooter’s Talent”

November 11, 2025

Jeffery Mei

Index:

Introduction
Model Application
Part 1: Basic xG model with Logistic Regression
- **a. Model Construction**
  - **i. Intro to Logistic Regression**
  - **ii. Coefficient Analysis**
- **b. Model Evaluations**
  - **i. Deviance and Wilk's Test**
  - **ii. AIC**
  - **iii. Team Analysis**
  - **iv. Player Analysis**

Part 2: xG model with "Shooter's Talent"
- **a. Intro to Bayesian Statistics**
- **b. Shooting%**
  - **i. Limitation of Shooting%**
- **c. Shot Ratio (= actual goal / xG)**
  - **i. Model Construction**
  - **ii. Likelihood Ratio Test**
  - **iii. Wald's Test for p-value**
  - **iv. Team Analysis**
  - **v. Player Analysis**
  - **vi. Goalie Analysis**
  - **vii. Game Analysis**

Conclusion
Variable Analysis: Ridge and Lasso Logistic Regression
Visualizing Model Application

Introduction:

The Expected Goals (xG) metric has emerged as a powerful tool in hockey analytics, quantifying the quality of each shot by estimating its probability of resulting in a goal. It is important in hockey because it evaluates *chance quality* rather than just counting shots or goals, providing a more accurate measure of team and player performance. xG helps separate sustainable skill from short-term luck. Coaches and analysts can use xG to assess whether a team's results match its underlying play, to identify players who consistently generate or prevent dangerous chances, and to evaluate the impact of tactical changes.

This study focuses on constructing an NHL xG model using logistic regression, a statistical approach that systematically incorporates various shot characteristics to assess scoring likelihood. By adopting a fenwick based framework, my methodology assigns an xG value of zero to blocked shots, ensuring that only attempts with a genuine chance of success are evaluated. The analysis begins with meticulous cleaning of play-by-play data to extract relevant features, followed by the development of the log-regression model. Ultimately, by summing the xG values of individual shots, I derive an aggregate measure of offensive performance over the course of a game.

A critical shortcoming of most publicly available xG models is their implicit assumption that every shooter is equally skilled. To capture this nuance, I propose extending the logistic-regression framework with a Bayesian hierarchical component that estimates a "shooter-talent" effect. By placing partial pooling priors on individual coefficients, the model naturally shrinks estimates for low-volume shooters toward the league mean while still allowing prolific shooters to express their true finishing ability. The posterior distribution of each player's talent term can then be integrated into a new model, yielding probability estimates that reflect both shot context and the historical scoring efficiency of the person taking it.

In this paper, I explored two potential ideas to compute a "shooter's talent" variable. The first idea is using shooting%, and the second idea is to calculate the Ratio between Actual Goal against xG called "Shot Ratio". We will evaluate both ideas in this essay to find the better variable to measure individual shooting talent using data from the entire 2024-2025 NHL regular season of more than 400k rows.

Model Application

In this section below, I will go over some common applications of the xG Model to explain how teams and individual players can utilize this model to analyze and improve their performance.

1. **Evaluate systems, not just results**
- If a team is “unlucky” (high xG for, low goals), coaches can keep faith in the system instead of overreacting to a losing streak.
- If a team is “lucky” (low xG for, high goals), it’s a warning that the current tactics may not be sustainable.
2. **Optimize shot selection**
- Show players where their own personal xG is strong (e.g., net-front, left circle) and encourage more attempts from those zones.
3. **Player evaluation and scouting**
- Separate “finishing luck” from true talent: a player who consistently creates high xG is valuable even in a cold scoring stretch.
- Identify undervalued players who quietly drive xG for their line, or defenders who consistently suppress dangerous chances.
4. **Goalie and defensive support**
- Use xG against to measure the *quality* of shots a goalie faces, not just the raw number, to judge both the goalie and team defense fairly
- Restructure defensive coverage to reduce high-xG shots from particular areas (backdoor plays, slot one-timers, etc.).
5. **Game-to-game preparation**
- Pre-scout opponents’ xG maps to know which players and locations are most dangerous, then tailor matchups and coverages
- Track in-game xG to see whether the team is truly outplaying the opponent or just winning/losing on finishing luck.

Model 1: Basic xG Model using Log Regression

Source and Filtering

To focus on genuine scoring attempts we keep only *Fenwick* events:

1. **Goals** (event code `GOAL`),
2. **Shots on goal** (`SHOT`),
3. **Missed shots** (`MISS`).

Empty-net situations are excluded, as they inflate goal rates in unrepresentative contexts. Blocked shots are ignored because their pre-block trajectory is unobservable.

Predictor Set

The table below summarizes the variables that I am using to compute the basic xG Model

Table 1: Model covariates.

Type	Variable (levels / units)
Categorical	StrengthState (5v5, 4v5, 5v4, 3v3, ...)
	ScoreState (Trailing, Tied, Leading)
	LastEvent (Faceoff, Turnover, Hit, Pass, ...)
	ShotType (Wrist, Slap, Snap, Backhand, Tip, Wrap)
	IsForward (1 if shooter is listed as F, else 0)
	TSLE (Time since last event)
	IsRebound (1 if shot within 2 s of previous shot, else 0)
Numerical	ShotDistance (ft): $\sqrt{x^2+y^2}$ from shooter to goal centre
	ShotAngle (deg): $\arctan\big(\lvert y\rvert/\lvert x\rvert\big)$, 0° is straight on

1 Model Construction

1.1 Logistic-Regression Specification

Let $Y_i \in \{0,1\}$ indicate whether shot i resulted in a goal and let $\mathbf{x}_i$ be the eight-dimensional numerical/categorical feature vector (after one-hot encoding). The logistic model posits

$$\Pr(Y_i = 1 \mid \mathbf{x}_i) \;=\; \frac{1}{1+\exp\big[-(\beta_0+\mathbf{x}_i^\top\boldsymbol{\beta})\big]}. \tag{1}$$

Parameters $(\beta_0, \boldsymbol{\beta})$ are estimated by maximising the Bernoulli likelihood across all n = 1,122,347 Fenwick events. We use GLM with an l_2 penalty ($\lambda = 10^{-4}$) to mitigate quasi-separation from rare factor levels.

Intro to Logistic Regression

1. Exponential-family foundation. The binary response $Y_i \in \{0,1\}$ follows a *Bernoulli* law, which is a member of the one-parameter *exponential family*

$$f(y_i \mid p_i) = p_i^{y_i}(1-p_i)^{1-y_i},$$
$$f(y_i \mid \theta_i) = \exp\{y_i\theta_i - b(\theta_i)\},$$

with natural parameter $\theta_i = \log\big(p_i/(1-p_i)\big)$ and cumulant $b(\theta_i) = \log\big(1+e^{\theta_i}\big)$. This connection allows logistic regression to be framed as a special case of a *generalised linear model* (GLM). In our application, $Y = 1$ indicates a shot results in a goal, whereas $Y = 0$ indicates that it does not.

2. Link function and systematic component. For each shot $i = 1,\ldots,n$ define Y_i as a goal indicator and let $\mathbf{x}_i = (1, x_{i1}, \ldots, x_{ip})^\top$. The GLM recipe supplies (i) a Bernoulli response, (ii) the canonical *logit link*, (iii) a linear predictor $\eta_i = \mathbf{x}_i^\top \boldsymbol{\beta}$.

$$\boxed{\log\Big(\tfrac{p_i}{1-p_i}\Big) = \eta_i = \mathbf{x}_i^\top \boldsymbol{\beta}} \quad \Longleftrightarrow \quad p_i = \frac{1}{1+\exp(-\eta_i)}.$$

3. Log-likelihood. Because $Y_i \sim \text{Bernoulli}(p_i)$, the sample log-likelihood is

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^{n}\Big[Y_i\,\eta_i - \ln\Big(1+e^{\eta_i}\Big)\Big]. \tag{2}$$

Maximum Likelihood Estimation

4. Likelihood construction for the simple ($p = 1$) case. Suppose the data consist of pairs (x_i, y_i), $i = 1,\ldots,n$, and obey the simple logistic model

$$\log\Big(\tfrac{p(x_i)}{1-p(x_i)}\Big) = \beta_0 + \beta_1 x_i.$$

Under independent Bernoulli trials the likelihood is

$$L(\beta_0,\beta_1) = \prod_{i=1}^{n} p_i^{y_i}(1-p_i)^{1-y_i}, \qquad p_i = \frac{e^{\beta_0+\beta_1 x_i}}{1+e^{\beta_0+\beta_1 x_i}}.$$

Taking logarithms and substituting $\eta_i = \beta_0 + \beta_1 x_i$ yields

$$\ell(\beta_0,\beta_1) = \sum_{i=1}^{n} y_i\eta_i - \sum_{i=1}^{n}\ln\Big(1+e^{\eta_i}\Big) = -\sum_{i=1}^{n}\ln\Big(1+e^{\beta_0+\beta_1 x_i}\Big) + \sum_{i=1}^{n} y_i(\beta_0+\beta_1 x_i). \tag{3}$$

We are not able to solve this exactly. We can however use the Newton's method to solve using Python

Coefficient Analysis:

	features	coef
0	ShotDistance	[-0.06254241645974003]
1	ShotAngle	[-0.015734465755265388]
2	is_forward	[-0.026111561540590807]
3	is_rebound	[-0.6408074365812075]
4	StrengthState_EV1	[-0.42548159286793436]
5	StrengthState_EV2	[-0.15723288452981649]
6	StrengthState_PP1	[-0.1545094412405834]
7	StrengthState_PP2	[0.4826264311327384]
8	StrengthState_SH	[-0.21436020136801415]
9	ScoreState_-1	[-0.09419051322266316]
10	ScoreState_-2	[-0.15693633983355856]
11	ScoreState_0	[-0.0058682815605362515]
12	ScoreState_1	[-0.03623358846713218]
13	ScoreState_2	[-0.058819336361988585]
14	ScoreState_<-2	[-0.20023915873336345]
15	ScoreState_>2	[0.08332952930576483]
16	LastEvent_No	[-0.05175000363796254]
17	LastEvent_Other	[-0.3417952288098973]
18	LastEvent_Shot	[-0.0754124564254441]
19	shotType_backhand	[-0.25258743071603745]
20	shotType_bat	[-0.05891245475169723]
21	shotType_between-legs	[-0.6811365594799432]
22	shotType_cradle	[-0.05384150163757972]
23	shotType_deflected	[-0.399642060803356]
24	shotType_poke	[0.18545694206402252]
25	shotType_slap	[0.6111317883387712]
26	shotType_snap	[0.5118599924285676]
27	shotType_tip-in	[-0.6557784075701051]
28	shotType_wrap-around	[-0.9577886539319366]
29	shotType_wrist	[0.182263672293897]

The xG model’s regression coefficients tell us how various factors contribute to the likelihood of scoring a goal. For instance, the coefficient for shot distance is strongly negative, which makes sense the more further a player shoots from the goal, the less likely it is to score. In contrast, the coefficient for shot angle—though also negative—is less pronounced, indicating that while shots taken from more acute angles do reduce scoring probability, the effect is smaller compared to distance.

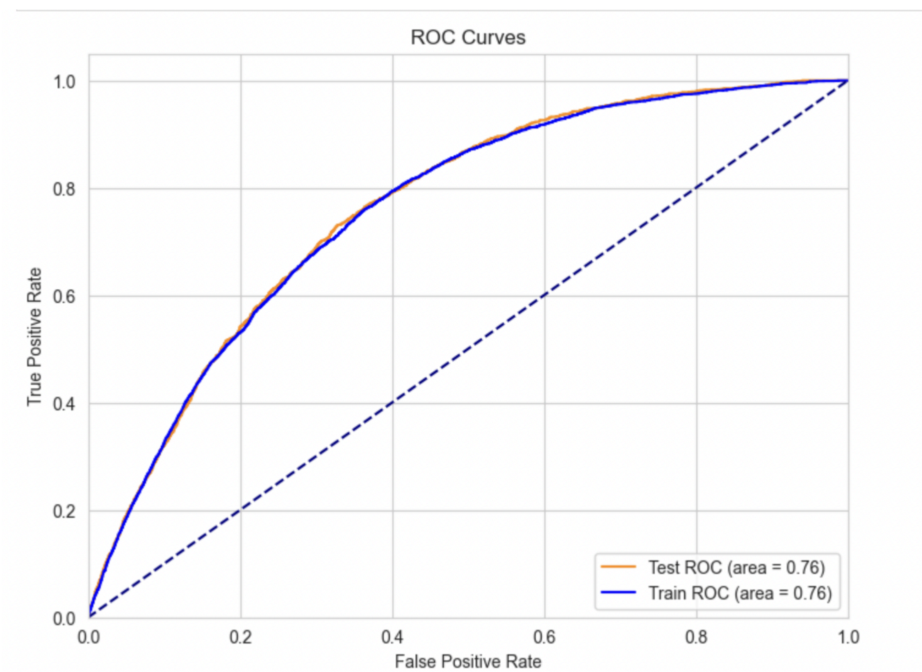
Additionally, categorical variables such as the type of shot reveal important differences relative to a baseline; a positive coefficient for a specific shot type (for example, a wrist shot) means that, all else equal, this shot is more likely to result in a goal compared to the reference category. Other game context variables—such as power play situations or defensive pressure—further adjust the expected goals, with their corresponding coefficients quantifying how much these scenarios increase or decrease scoring chances. Overall, the magnitude and sign of these coefficients not only help in predicting goal probabilities but also provide critical insights into which elements most significantly influence shooting effectiveness.

Model Evaluations

We split the 24–25 data into training and testing sets using a **70/30 random split**.

Table 2: Goodness-of-fit and discrimination statistics (test set).

Metric	Value	Meaning	Interpretation
AUC	0.76	Rank-ordering skill	76 % chance the model assigns a higher score to a goal than to a miss.
Log-loss	0.21	Calibration penalty	Low value indicates few over-confident probability errors.
Deviance	25 575.78	$-2\log L$	Baseline for nested model tests.
df	30	Parameters	Number of freely estimated coefficients.
Wilks p	$< 10^{-15}$	Likelihood-ratio vs. intercept-only	Covariates are highly informative.
AIC	25 635.78	Deviance + 2 df	Lower values preferred for out-of-sample prediction.



gameDate	HomeTeam	AwayTeam	HomeScore	AwayScore	Home_xG	Away_xG
2024-10-04	BUF	NJD	1	4	2.74635516382779	2.2991562277533400
2024-10-05	NJD	BUF	3	1	3.6754521818392700	1.8739018183456900
2024-10-08	SEA	STL	2	3	3.359978088562130	2.21337456494403
2024-10-08	UTA	CHI	5	2	2.61681284534542	3.127348970179960
2024-10-08	FLA	BOS	6	4	4.562973846612400	2.6160520411534900

Deviance and Wilks' Likelihood–Ratio Test

For logistic regression we compare

$$H_0 : \text{intercept-only model} \quad \text{vs} \quad H_1 : \text{model with predictors}. \tag{1}$$

Let $\tilde{\boldsymbol{\beta}}$ denote the intercept-only MLE and $\widehat{\boldsymbol{\beta}}$ the full-model MLE. Define

$$D_0 = -2\,\ell(\tilde{\boldsymbol{\beta}}), \qquad D = -2\,\ell(\widehat{\boldsymbol{\beta}}), \qquad G^2 = D_0 - D. \tag{2}$$

Under H_0, *Wilks' theorem* states $G^2 \xrightarrow{d} \chi^2_\nu$ with $\nu = \text{df} = \dim(\widehat{\boldsymbol{\beta}}) - \dim(\tilde{\boldsymbol{\beta}})$.

Observed statistics.

$$\begin{aligned}
\ell(\widehat{\boldsymbol{\beta}}) &= -12\,787.89,\\
\ell(\tilde{\boldsymbol{\beta}}) &= -40\,244.12,\\
D_0 &= 80\,488.25,\\
D &= 25\,575.78,\\
G^2 = D_0 - D &= 54\,912.46,\\
\nu &= 30,\\
p\text{-value} &< 10^{-15}.
\end{aligned}$$

Because $G^2 \gg \chi^2_{30,0.95} = 43.8$, we decisively **reject** H_0: the predictor set yields a statistically and practically significant improvement over a constant-only baseline.

Akaike Information Criterion (AIC)

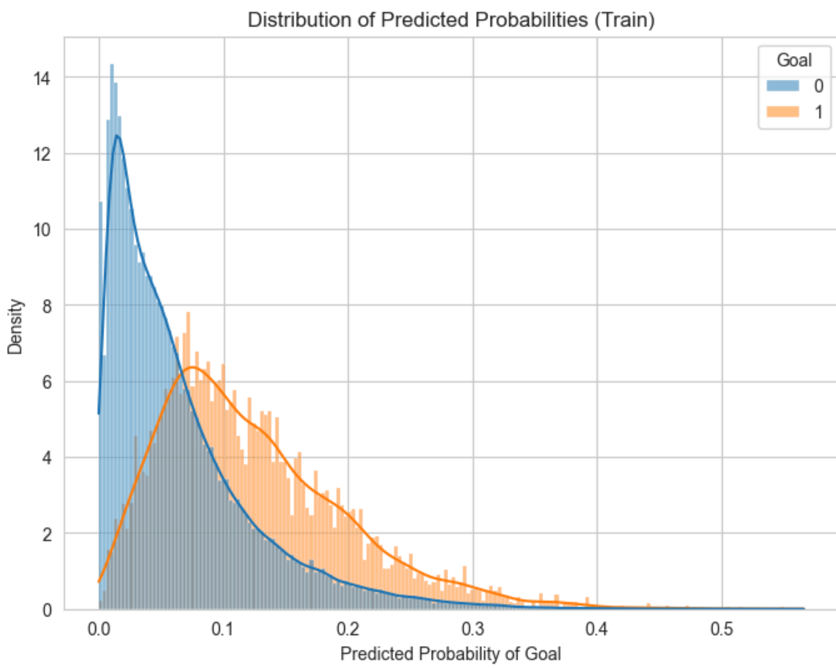
AIC penalises deviance by model complexity:

$$\text{AIC} = D + 2k, \tag{3}$$

where k is the number of free parameters. With $k = 30$ and the deviance above, $\text{AIC} = 25\,575.78 + 2(30) = 25\,635.78$. Competing models must achieve an AIC at least two units lower ($\lesssim 25\,634$) to be considered substantially better.

Summary Statistics for xG Probability:

- count 113331
- mean 0.0687
- std 0.0629
- min 0.000004
- 25% 0.022
- 50% 0.051
- 75% 0.094
- max 0.566



In the training data the model gives most shots a very low chance of becoming a goal, but the two classes still separate. The blue bars (misses) are bunched up below about 5 % probability and then drop off quickly, while the orange bars (actual goals) start low but stretch out to the right, with a noticeable tail past 20 % and a few cases above 50 %. That longer right-hand tail for goals shows the model is assigning higher probabilities to the shots that really do score, even though the bulk of both distributions sits near zero—exactly what you would expect from the AUC of 0.76.

From regression to probability.

For a binary logistic-regression model the linear predictor is

$$\eta \;=\; \beta_0 \;+\; \boldsymbol{\beta}^\top \mathbf{x},$$

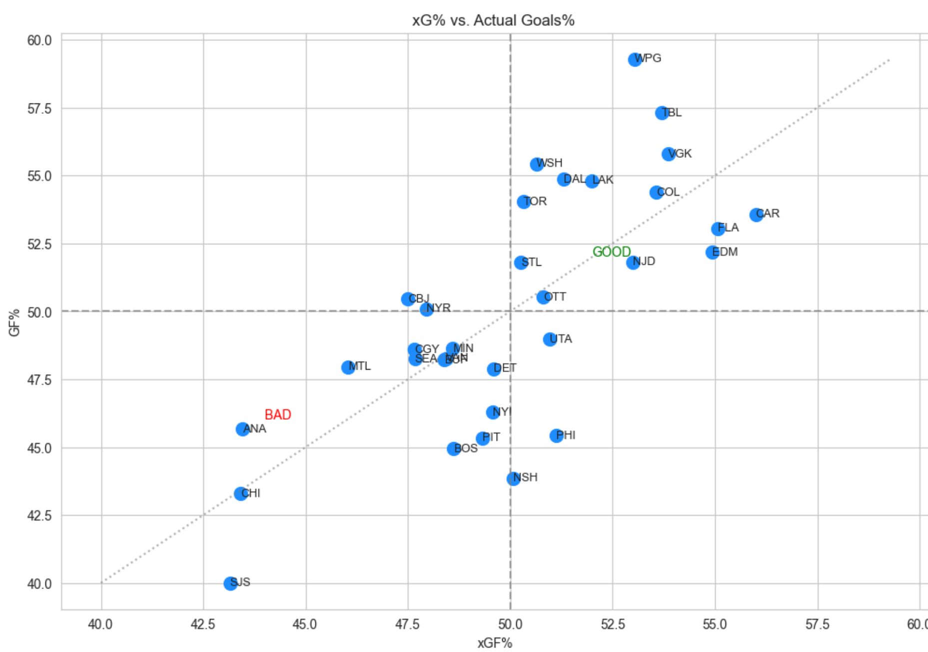
where β_0 is the intercept, $\boldsymbol{\beta}$ the coefficient vector, and $\mathbf{x}$ the feature vector for a single observation. The *logit* link maps this real-valued score to an actual probability via

$$\Pr(Y = 1 \mid \mathbf{x}) \;=\; \sigma(\eta) \;=\; \frac{1}{1 + \exp(-\eta)}.$$

After converting each event into a probability using Python, I calculated the total expected goals (xG) for both teams by summing these probabilities across all relevant events in each match. This provided a clear numerical representation of each team's offensive effectiveness and defensive resilience within the game.

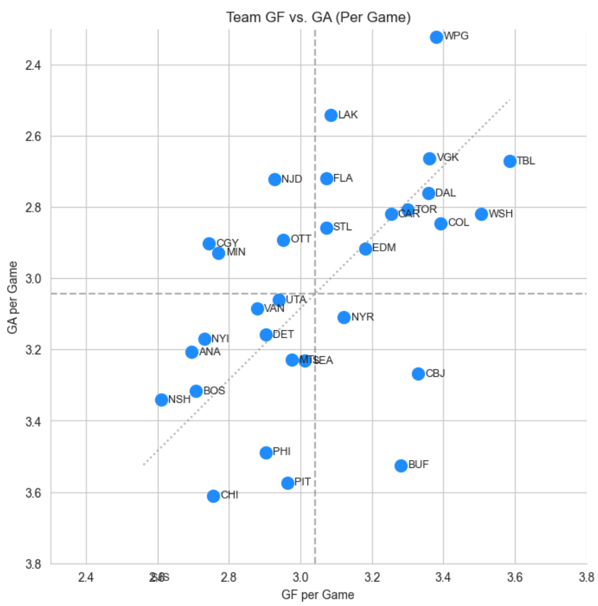
Team Analysis:

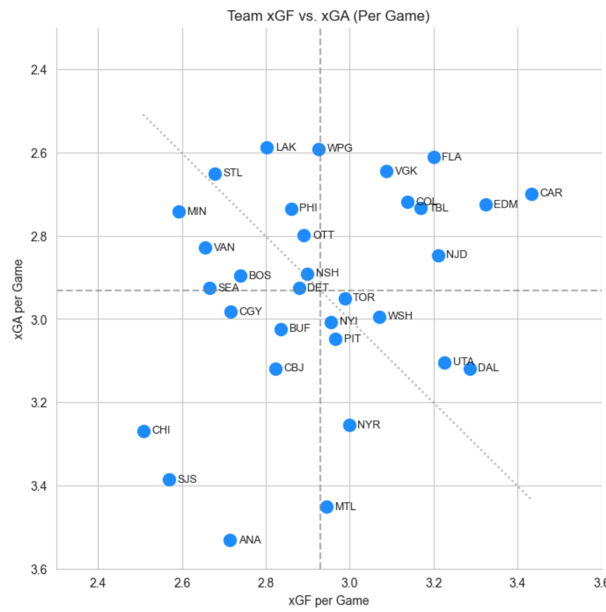
With the actual xG figures established, I then conducted an in-depth team analysis to understand performance trends better. By comparing expected goals with actual goals scored, it was possible to identify which teams were performing above or below expectations, offering insight into their finishing quality and overall efficiency. Teams consistently outperforming their xG demonstrated high-quality finishing or potentially weaker opposition goalkeeping, whereas teams frequently underperforming suggested issues in finishing or facing particularly strong goalkeeping performances. I have noticed that all the teams in the upper right quadrant have made the playoffs and all the teams except MTL and MIN in the lower right quadrant have missed the playoffs.



xGF% Expected Goals For Percentage: xGF / (xGF + xGA)

GF% Shots For Percentage: SF / (SF + SA)





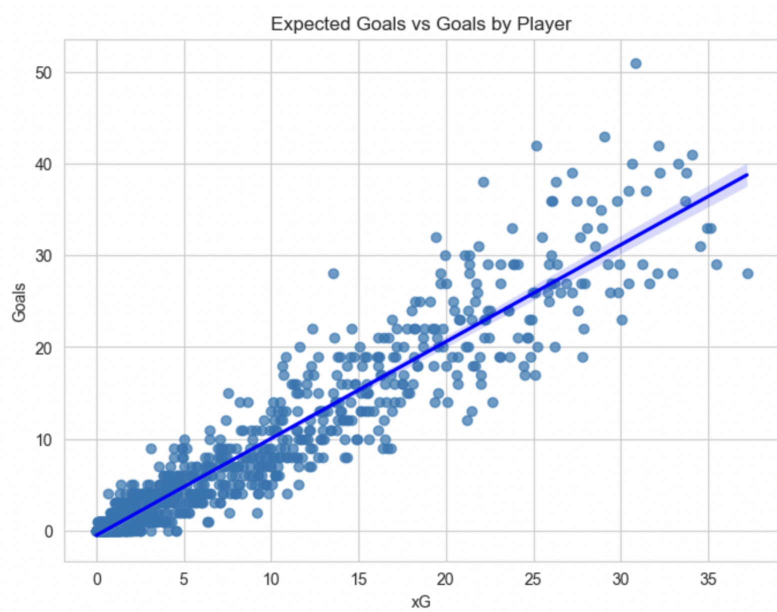
These two graphs provide valuable insights into team performances by comparing actual goals against expected goals. In the first graph, teams located above the diagonal line tend to concede fewer goals than they score, indicating effective performance, while those below the diagonal struggle defensively or offensively. For example, WPG and NJD appear strong in actual outcomes. The second graph offers a predictive perspective based on expected goals. Teams such as FLA, CAR, and EDM exhibit strong offensive metrics (high xGF) and solid defense (low xGA), suggesting sustainable high-level performance. Conversely, teams like ANA and SJS, positioned unfavorably in both graphs, demonstrate weaknesses that require significant improvements in both attack and defense. Comparing both graphs helps identify teams performing beyond their expected metrics (potentially due to finishing skill or goalkeeping strength) or underperforming relative to their potential.

Teams consistently outperforming their xG demonstrated high-quality finishing or potentially weaker opposition goalkeeping, whereas teams frequently underperforming suggested issues in finishing or facing particularly strong goalkeeping performances. For instance, a team with a higher actual goal tally than their xG indicates superior shot conversion rates, effective offensive strategies, or weaknesses in the opposition's defense. Conversely, a team that scores fewer goals than expected might be struggling with finishing accuracy, offensive tactics, or facing exceptionally strong defensive teams.

Player Analysis:

Top 20 players by total xG:

	playerId	Player	Total_xG	Total_Goals
107	8475786	Zach Hyman	37.113545	28
548	8480801	Brady Tkachuk	35.686981	29
785	8482740	Wyatt Johnston	35.189655	33
409	8479318	Auston Matthews	35.050117	33
246	8477492	Nathan MacKinnon	34.277570	31
303	8478010	Brayden Point	34.101671	41
275	8477933	Sam Reinhart	34.090084	39
160	8476483	Rickard Rakell	33.447874	36
628	8481540	Cole Caufield	33.095911	40
80	8475314	Anders Lee	32.729483	28
330	8478398	Kyle Connor	32.260988	42
226	8477404	Jake Guentzel	32.103358	39
283	8477946	Dylan Larkin	31.844652	28
700	8482093	Seth Jarvis	31.838676	27
634	8481557	Matt Boldy	31.214934	29
415	8479337	Alex DeBrincat	31.128261	37
292	8477956	David Pastrnak	30.838412	40
56	8475166	John Tavares	30.743620	37
276	8477934	Leon Draisaitl	30.660166	51
254	8477500	Bo Horvat	30.360951	27




```
Model Evaluation Metrics:
  Pearson Correlation (Goals vs xG): 0.944
  R^2: 0.889
  RMSE: 3.198
```

This scatter plot clearly illustrates a strong linear relationship between expected goals (xG) and actual goals scored by individual players, evidenced by a Pearson correlation of 0.944 and an R^2 of 0.889. The high correlation indicates that xG is an excellent predictor of actual scoring performance. Players positioned significantly above the regression line represent highly efficient scorers who outperform their expected goal totals, reflecting exceptional finishing ability or perhaps weaker opposing goalkeepers faced.

Intro to Bayesian Method

1 why is it important to hockey?

Which of these two proportions is higher: 5 out of 10, or 450 out of 1000? This sounds like a silly question. Obviously 5/10=.5 , which is greater than 450/1000=.45. But suppose you were a hockey goalie coach, trying to decide which of two potential players is a better goalie based on how many saves they made in the past season. One has allowed 20 goals in 100 shots, the other 50 goals in 170 shots. While the first player has a higher proportion of hits, it's less evidence. Luckily, we can use a very useful statistical method for estimating a large number of proportions, called empirical Bayes estimation.

The same philosophy goes for shooters, since players in the NHL take a distinctive number of shots, it is difficult to compare two players goal total with different number of shots. As a result, we can use Bayesian Method to to compute a posterior for a more fair ground.

1.1 Recall of Bayes' Theorem (Event Form)

From a basic probability perspective, Bayes' Theorem states that for two events A and B with $P(B) \neq 0$,

$$P(A \mid B) \; = \; \frac{P(B \mid A)\, P(A)}{P(B)}.$$

Proof (sketch). Recall that in conditional probability,

$$P(A \mid B) \; = \; \frac{P(A \cap B)}{P(B)}, \quad P(B \mid A) \; = \; \frac{P(A \cap B)}{P(A)}.$$

Hence

$$P(A \mid B) \; = \; \frac{P(A \cap B)}{P(B)} \; = \; \frac{P(B \mid A)\, P(A)}{P(B)}.$$

1.2 Bayes' Theorem in Distribution Form

In order to find the distribution of the posterior probability for a parameter, we need a continuous (or discrete) version of Bayes' Theorem. We adopt the following notations throughout this chapter:

- **Prior distribution:** Let θ be the parameter of interest. We assign a prior distribution $\pi(\theta)$, which reflects our belief about θ before observing data.
- **Observed Data:** Conditional on θ, the data $\{X_1, X_2, \ldots, X_n\}$ are assumed to be i.i.d. from some distribution with density (or mass) function $p(x \mid \theta)$. That is,

$$X_1, X_2, \ldots, X_n \; \sim \; p(x \mid \theta) \quad \text{i.i.d.}$$

- **Likelihood Function:** Once we observe a particular sample $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, the likelihood function is defined by

$$L(\theta \mid x_1, \ldots, x_n) \; = \; p(x_1, \ldots, x_n \mid \theta) \; = \; \prod_{i=1}^{n} p(x_i \mid \theta).$$

- **Posterior Distribution:** After observing data $\mathbf{x}$, the posterior distribution for θ is given by

$$p(\theta \mid x_1, \ldots, x_n) \;=\; \frac{p(x_1, \ldots, x_n \mid \theta)\, \pi(\theta)}{\displaystyle\int p(x_1, \ldots, x_n \mid \theta')\, \pi(\theta')\, d\theta'}.$$

- **Bayes' Theorem for PDFs/PMFs (single observation version):** If we have a single observation $X = x$, then

$$p(\theta \mid x) \;=\; \frac{p(x \mid \theta)\, \pi(\theta)}{\displaystyle\int p(x \mid \theta')\, \pi(\theta')\, d\theta'}.$$

  This is the continuous/discrete analog of the elementary Bayes' Theorem for events, but now in terms of distributions.

1.3 Philosophy of Bayesian Statistics

From a Bayesian perspective, the parameter θ is treated as a random variable with a prior distribution $\pi(\theta)$. Data are generated according to some assumed model $p(x \mid \theta)$. As soon as data $\mathbf{x}$ are observed, we update our belief about θ via the posterior distribution $p(\theta \mid \mathbf{x})$.

This approach contrasts with the frequentist viewpoint where θ is considered a fixed (but unknown) constant, and probability statements are made only about the random samples X_i. However, both schools of thought often lead to similar or complementary conclusions, especially with sufficient data.

Example 1.1: Simple Binomial Setting

Suppose we flip a coin 10 times. Let X be the number of heads obtained. If the coin is fair, $X \sim \text{Binomial}(n = 10, p = \frac{1}{2})$. In a Bayesian approach, if we did *not* know p, we might put a prior on p, say a Beta distribution, and then use the observed data to update that prior to a posterior distribution $\text{Beta}(\alpha', \beta')$. This illustrates the basic mechanics of Bayesian updating.

2 How to Choose a Prior

When we first learn about Bayesian statistics, we inevitably ask: *"How do we choose the prior?"* This is an important question because the posterior distribution depends on both the prior and the likelihood function. In many practical situations, however, with sufficiently large data samples, the impact of the prior on the final posterior is often small.

A commonly used strategy is to pick a *conjugate prior*, which ensures that the posterior remains in the same family as the prior. Below are three classic examples demonstrating conjugate priors.

2.1 Example 1: Normal Data with Known Variance

Suppose we observe i.i.d. data

$$X_1, X_2, \ldots, X_n \;\sim\; \mathcal{N}(\mu, \sigma^2),$$

where σ^2 is known (but μ is unknown). We choose a *normal prior* for μ:

$$\mu \;\sim\; \mathcal{N}(\mu_0, \tau_0^2).$$

Then, by Bayes' theorem, the *posterior distribution* of μ after observing $X_1, \ldots, X_n$ is

$$\mu \,\big|\, (X_1, \ldots, X_n) \;\sim\; \mathcal{N}\big(\mu_n,\, \tau_n^2\big),$$

where

$$\tau_n^2 \;=\; \left(\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}\right)^{-1} \quad \text{and} \quad \mu_n \;=\; \tau_n^2 \left(\frac{\mu_0}{\tau_0^2} + \frac{n\,\overline{X}}{\sigma^2}\right),$$

with $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Notice that we start with a normal prior for μ and end up with a normal posterior for μ, which is the hallmark of a conjugate prior.

2.2 Poisson–Gamma Conjugacy

Poisson Likelihood

Step 1: Poisson PMF (Single Observation). Let X_i be a Poisson(λ) random variable. By definition, its *probability mass function* (pmf) is

$$P(X_i = x_i \mid \lambda) \;=\; \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}, \quad x_i \in \{0, 1, 2, \ldots\},\ \lambda > 0. \tag{1}$$

Step 2: Joint Likelihood for i.i.d. Observations. Suppose we have n i.i.d. observations $X_1, \ldots, X_n$, each distributed as Poisson(λ). From (1), the joint likelihood function is

$$L(\lambda \mid x_1, \ldots, x_n) \;=\; \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} \;=\; \left(\prod_{i=1}^{n} \frac{1}{x_i!}\right) \lambda^{\sum_{i=1}^{n} x_i} \exp(-n\lambda). \tag{2}$$

We typically treat the product $\prod_{i=1}^{n}(1/x_i!)$ as a constant factor with respect to λ.

Gamma Prior

Step 3: Gamma PDF. Assume we place a Gamma(α, β) prior on λ, which has the *probability density function* (pdf):

$$\pi(\lambda) \;=\; \frac{\beta^{\alpha}}{\Gamma(\alpha)}\, \lambda^{\alpha-1}\, e^{-\beta\lambda}, \quad \lambda > 0,\ \alpha > 0,\ \beta > 0. \tag{3}$$

Posterior Derivation

Step 4: Posterior is Proportional to Likelihood $\times$ Prior. By Bayes' theorem (in density form),

$$p(\lambda \mid x_1, \ldots, x_n) \;=\; \frac{L(\lambda \mid x_1, \ldots, x_n)\, \pi(\lambda)}{\displaystyle\int_0^{\infty} L(\lambda')\, \pi(\lambda')\, d\lambda'}.$$

Ignoring the normalizing denominator for a moment, we have

$$\begin{aligned} p(\lambda \mid x_1, \ldots, x_n) \;&\propto\; L(\lambda \mid x_1, \ldots, x_n) \;\times\; \pi(\lambda) \\ &=\; \left[\lambda^{\sum_{i=1}^{n} x_i}\, e^{-n\lambda}\right] \;\times\; \left[\lambda^{\alpha-1}\, e^{-\beta\lambda}\right] \\ &=\; \lambda^{\left(\sum_{i=1}^{n} x_i + \alpha - 1\right)}\, \exp\big(-(\beta + n)\lambda\big). \end{aligned} \tag{4}$$

This expression *defines* a Gamma distribution kernel with shape parameter $\left(\alpha + \sum_{i=1}^{n} x_i\right)$ and rate parameter $(\beta + n)$.

Step 5: Recognize the Gamma Kernel. A generic Gamma(α^*, β^*) pdf can be written (up to a normalization constant) as

$$\lambda^{\alpha^*-1}\, e^{-\beta^* \lambda}.$$

Comparing with (4), we see

$$\alpha^* \;=\; \alpha + \sum_{i=1}^{n} x_i \quad \text{and} \quad \beta^* \;=\; \beta + n.$$

Thus, the posterior distribution is

$$\lambda \,\Big|\, (x_1, \ldots, x_n) \;\sim\; \Gamma\Big(\, \alpha + \sum_{i=1}^{n} x_i,\; \beta + n\Big). \tag{5}$$

Hence, *Gamma is a conjugate prior* for the Poisson likelihood.

2.3 Bernoulli–Beta Conjugacy

Bernoulli Likelihood

Step 1: Bernoulli PMF (Single Observation). Let X_i be a Bernoulli(p) random variable. Then

$$P(X_i = x_i \mid p) \;=\; p^{x_i}\,(1-p)^{1-x_i}, \quad x_i \in \{0,1\},\; p \in (0,1).$$

Step 2: Joint Likelihood for i.i.d. Observations. Given n i.i.d. Bernoulli(p) observations $X_1, \ldots, X_n$, the joint likelihood is

$$L\big(p \mid x_1, \ldots, x_n\big) \;=\; \prod_{i=1}^{n} p^{x_i}(1-p)^{\,1-x_i} \;=\; p^{\sum_{i=1}^{n} x_i}\,(1-p)^{\,n-\sum_{i=1}^{n} x_i}. \tag{6}$$

Beta Prior

Step 3: Beta(α, β) PDF. We place a Beta(α, β) prior on p. Its pdf is

$$\pi(p) \;=\; \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\, p^{\alpha-1}\,(1-p)^{\beta-1}, \quad 0 < p < 1,\; \alpha > 0,\; \beta > 0. \tag{7}$$

Posterior Derivation

Step 4: Posterior is Proportional to Likelihood $\times$ Prior. Again, by Bayes' theorem,

$$p(p \mid x_1, \ldots, x_n) \;\propto\; L(p \mid x_1, \ldots, x_n) \;\times\; \pi(p).$$

Hence,

$$\begin{aligned} p(p \mid x_1, \ldots, x_n) \;&\propto\; \Big[p^{\sum x_i}(1-p)^{n-\sum x_i}\Big] \;\times\; \Big[p^{\alpha-1}(1-p)^{\beta-1}\Big] \\ &=\; p^{(\alpha-1)+\sum x_i}\,\big(1-p\big)^{(\beta-1)+\big(n-\sum x_i\big)}. \end{aligned} \tag{8}$$

By the definition of a Beta kernel,

$$p(p \mid x_1, \ldots, x_n) \;\sim\; \text{Beta}\Big(\alpha + \sum_{i=1}^{n} x_i,\; \beta + n - \sum_{i=1}^{n} x_i\Big).$$

Hence, *Beta is a conjugate prior* for the Bernoulli (or Binomial) likelihood.

2.4 Summary

In summary, one often chooses conjugate priors because they yield posterior distributions in the same functional family as the prior, leading to straightforward analytical updates. Here are the key examples:

- **Normal Data with Known Variance:** Starting with $\mu \sim \mathcal{N}(\mu_0, \tau_0^2)$ for data $X_i \sim \mathcal{N}(\mu, \sigma^2)$, the posterior is $\mu \mid (X_1, \dots, X_n) \sim \mathcal{N}(\mu_n, \tau_n^2)$.
- **Bernoulli/Binomial Data:** With $p \sim \text{Beta}(\alpha, \beta)$ and data from a Bernoulli(p) model, the posterior is $\text{Beta}(\alpha + \sum x_i,\ \beta + n - \sum x_i)$.
- **Poisson Data:** With $\lambda \sim \Gamma(\alpha, \beta)$ and data $X_i \sim \text{Poisson}(\lambda)$, the posterior is $\lambda \mid x_1, \dots, x_n \sim \Gamma\Big(\alpha + \sum_{i=1}^n x_i,\ \beta + n\Big)$.

A Simple Example

A farmer is interested in finding the average weight μ of hundreds of adult cows. He weighed 10 randomly chosen cows from his herd and found an average sample weight of $\bar{x} = 410\,\text{kg}$. Assume the population standard deviation is known to be $\sigma = 20\,\text{kg}$.

Based on the farmer's past experience, he believes that μ is likely around 420 kg and unlikely to deviate by more than about 30 kg. Hence, he decides to use a *Normal* prior distribution for μ:

$$\mu \ \sim\ \mathcal{N}(420,\ 10^2),$$

reflecting a mean of 420 kg with a standard deviation of 10 kg. We now combine this prior with the new sample data (assuming the data are i.i.d. $\mathcal{N}(\mu, \sigma^2)$) to find the posterior distribution of μ.

(i) Why a Normal Prior?

Since each X_i follows a Normal distribution with known variance, a *conjugate prior* for the mean μ is a Normal distribution. The farmer's belief that μ is "around 420 kg" (with a typical deviation of about 10 kg) naturally suggests choosing

$$\mu \sim \mathcal{N}(420,\ 10^2).$$

(ii) Posterior Distribution of μ

From the normal–normal conjugacy formula (see Q3 or standard Bayesian references), the posterior for μ after observing $\bar{x}$ from n samples is:

$$\mu \,\big|\, (X_1, \dots, X_n) \ \sim\ \mathcal{N}\Big(\mu_n,\ \tau_n^2\Big),$$

where

$$\tau_n^2 \ =\ \left(\tfrac{1}{\tau_0^2} + \tfrac{n}{\sigma^2}\right)^{-1} \quad \text{and} \quad \mu_n \ =\ \tau_n^2 \left(\tfrac{\mu_0}{\tau_0^2} + \tfrac{n\,\bar{x}}{\sigma^2}\right).$$

In this problem:

$$\mu_0 = 420, \quad \tau_0^2 = 10^2 = 100, \quad n = 10, \quad \bar{x} = 410, \quad \sigma^2 = 20^2 = 400.$$

Compute τ_n^2:

$$\frac{1}{\tau_0^2} = \frac{1}{100} = 0.01, \quad \frac{n}{\sigma^2} = \frac{10}{400} = 0.025.$$

Hence,

$$\tau_n^2 = \Big(0.01 + 0.025\Big)^{-1} = (0.035)^{-1} = 28.57 \quad \text{(approximately)}.$$

Compute μ_n:

$$\frac{\mu_0}{\tau_0^2} = \frac{420}{100} = 4.2, \quad \frac{n\,\bar{x}}{\sigma^2} = \frac{10 \times 410}{400} = 10.25.$$

Thus,

$$\mu_n = 28.57\,\big(4.2 + 10.25\big) = 28.57 \times 14.45 = 412.86 \quad \text{(approximately)}.$$

Therefore, the posterior for μ is

$$\mu \,\big|\, (X_1, \ldots, X_n) \sim \mathcal{N}\big(412.86,\ 28.57\big).$$

(Note that 28.57 here is the posterior *variance*, so the posterior standard deviation is $\sqrt{28.57} \approx 5.35$.)

(iii) Bayes Estimator for Mean-Squared-Error (MSE) Loss

Under quadratic (MSE) loss, the Bayes estimator is simply the *posterior mean.* Hence,

$$\hat{\mu} = \mu_n = 412.86\,\text{kg (approx.)}.$$

(iv) 95% Equal-Tailed Credible Interval

A 95% credible interval for μ based on the posterior $\mathcal{N}\big(\mu_n, \tau_n^2\big)$ is

$$\mu_n \pm z_{0.975}\,\sqrt{\tau_n^2},$$

where $z_{0.975} \approx 1.96$ is the standard normal critical value. Here:

$$\sqrt{\tau_n^2} = \sqrt{28.57} \approx 5.35, \quad 1.96 \times 5.35 \approx 10.48.$$

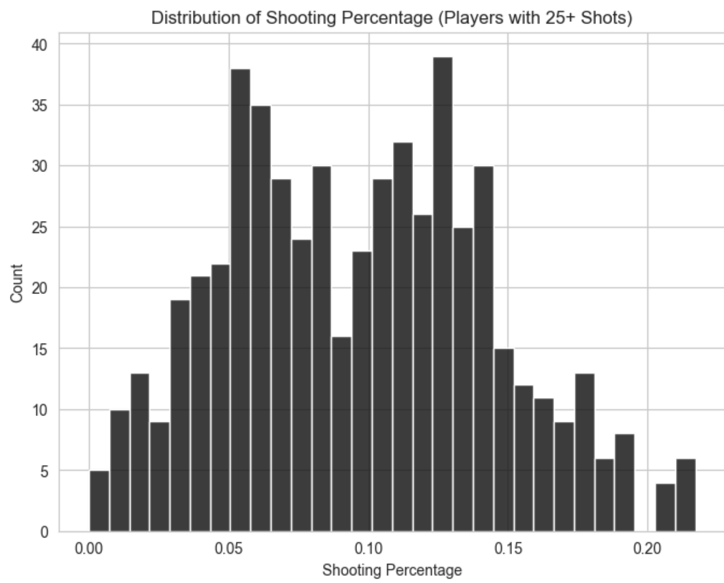
Thus, the 95% credible interval is approximately

$$412.86 \pm 10.48 = \big[402.38,\ 423.33\big].$$

Model 2: xG Model with “shooter’s talent variable

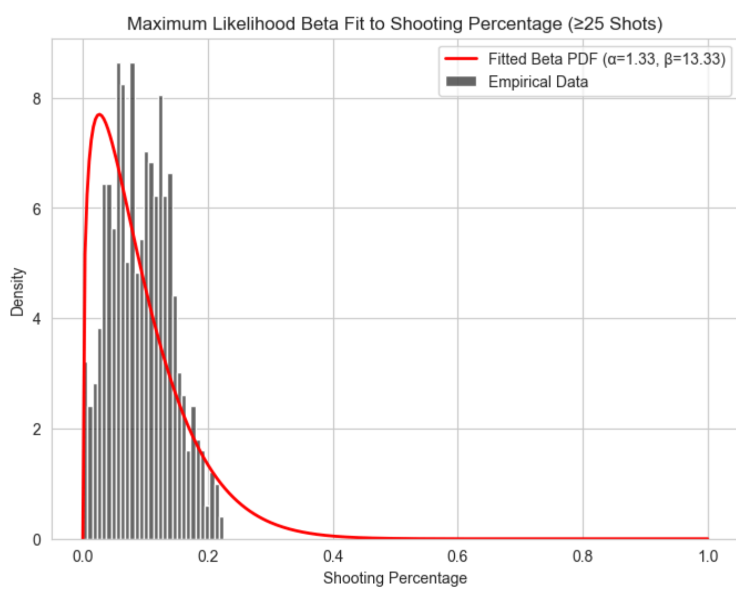
Idea 1: Shooting%

Shooting percentage is often modeled using a Bernoulli distribution because each shot represents an independent trial with two possible outcomes: success (goal) or failure (no goal). In this context, the success probability p corresponds directly to the shooting percentage.



(from 559 players)

Method of Maximum Likelihood:



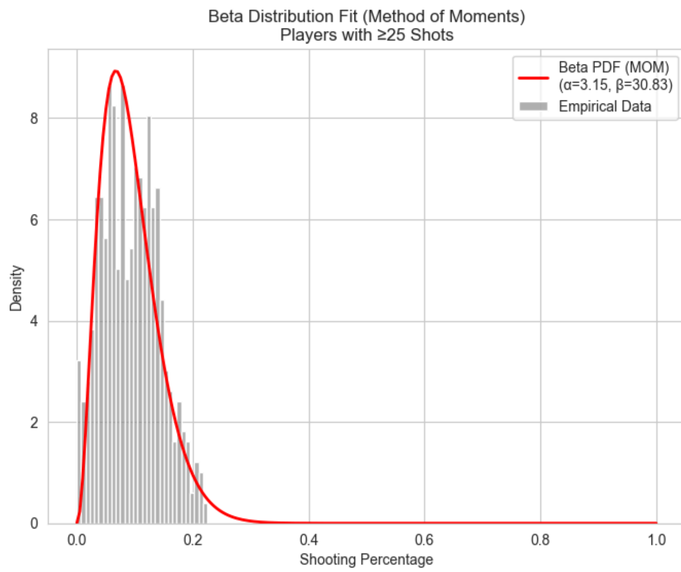
it does not look promising at all

Method of Moment:

MOM formulas for Beta distribution:

α = mean * (mean*(1-mean)/variance - 1)

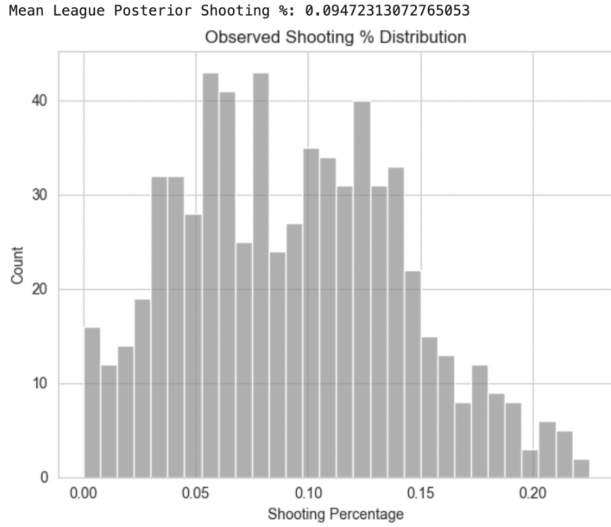
β = (1-mean) * (mean*(1-mean)/variance - 1)

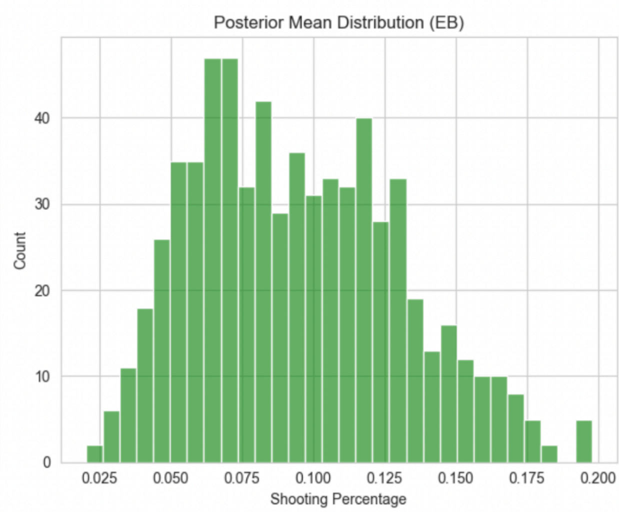


Percentage of data within this interval: 92.76%

After this we Compute the Posterior using Bayesian Stats:

```
Beta Prior (from league distribution):
  alpha ~ 2.8301
  beta  ~ 27.6669
Mean League Posterior Shooting %: 0.09472313072765053
```






```
=== Distribution Metrics ===
Observed Shooting %  -> Mean: 0.0928, Std: 0.0491
Posterior Mean (EB)  -> Mean: 0.0947, Std: 0.0361

Correlation (Observed vs. Posterior Mean): 0.9882
```

The Limitations of Shooting Percentage as a Measure of True Talent

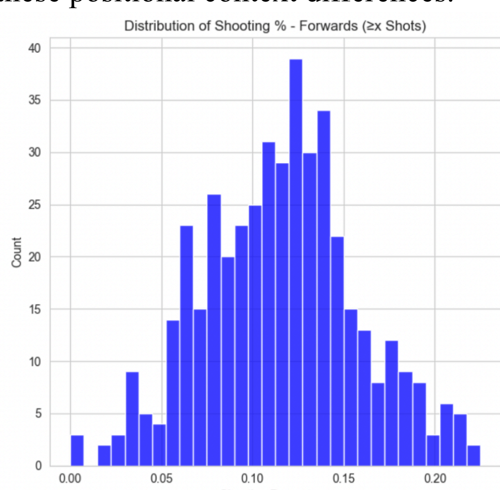
Introduction:

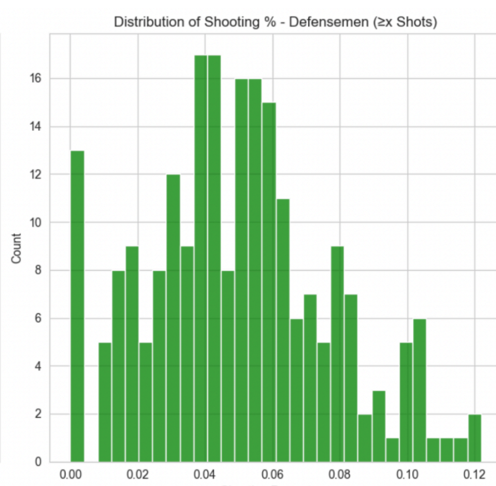
Shooting percentage (goals scored per shot on goal) is often cited as a measure of a hockey player's finishing ability. Intuitively, a higher shooting percentage (SH%) suggests a more "efficient" scorer. However, I think **raw shooting percentage is an unreliable gauge of true shooting talent**. Numerous factors – including player position, playing style, shot location, and plain statistical variability – heavily influence this metric. Short-term spikes or slumps in SH% can mislead observers about a player's actual skill level, since *luck and context play a significant role* in goal-scoring outcomes in hockey. Below, we examine why shooting percentage should be interpreted with caution and never in isolation when evaluating finishing ability or offensive contribution.

Positional Differences in Shooting Percentage

One major confounding factor is **player position**. Forwards and defensemen have very different shooting habits. Forwards generally operate closer to the net and in prime scoring areas, yielding higher shooting percentages on average. In fact, NHL forwards historically finish around **with a shooting% of 12%**. Defensemen, by contrast, take the bulk of their shots from the blue line or perimeter, where scoring chances are lower. Moreover, the purposes of their shots can be different as well. A defenceman shooting from the blue line might not try to shoot to score, but more to look for a tip in or just to create rebound opportunities. Same thing can be applied to forwards. Forward shooting can also be attributed to creating rebounds or just to create chaos in front of the net. As a result, defensemen collectively tend to convert only on roughly **5% of their shots** – less than half the rate of forwards.

These baseline differences mean comparing players' SH% across positions is apples-to-oranges. A defenseman with a 6% shooting percentage might actually be an above-average finisher for his position, while a winger at 9% could be below average for a forward. Position-driven shot selection (point shots vs. slot shots) skews the raw numbers. Thus, a low SH% isn't necessarily indicative of poor talent if the player is a defenseman (who naturally scores less often per shot), and a high SH% for a forward could simply reflect the easier opportunities that forwards typically get. Any evaluation of shooting prowess must account for these positional context differences.






```
Number of forwards in the filtered dataset: 438
Number of defensemen in the filtered dataset: 225
```

Shot Selection, Player Style, and Usage

Beyond position, **stylistic variations and role** significantly influence shooting percentage. Not all players approach shooting the same way. Some are **volume shooters** like Alexander Ovechkin who pepper the net from all angles and distances, while others pass-first players like Casey Mittlestadt are more selective, only shooting when they have a high-quality chance. A sniper who frequently uncorks shots from the perimeter or one-timers from long range might post a middling SH%, even if he's a feared goal scorer, simply because many of his attempts are low-percentage by nature. Conversely, a player who **only shoots from in tight or on odd-man rushes** may sport an inflated SH% without necessarily being the more skilled shooter – he's often the beneficiary of prime scoring chances created by teammates or circumstance.

Consider the difference between a player who generates his own shots versus one who finishes plays set up for him. A grind-it-out forward parked at the crease might only take a couple shots a game, but they're rebounds or tap-ins, resulting in a high conversion rate. Meanwhile, an aggressive shooting winger might fire five shots a game including long wrists from bad angles – boosting the team's shot totals and creating chaos, but converting a lower fraction of those tries. **Shot usage and role matter:** being a top-line power play triggerman versus a depth forward changes the context of one's shots. A power play shooter often faces a screened goalie with time to pick a corner (raising SH%), whereas a checking-line player might fling unscreened shots from the boards just to get a whistle (lowering SH%). Therefore, raw shooting percentage can mask these stylistic differences. A high SH% might tell us more about *how* and *when* a player shoots (or how his team deploys him) than how *skilled* his shot is.

Short-Term Variability and "Puck Luck"

Perhaps the biggest reason shooting percentage is unreliable is the **massive role of chance and variability**, especially in short samples that we are only analyzing less than a season's data. Scoring a goal in hockey – even for the best shooters – has a large random element. The puck can hit a post, a goalie can rob a labeled shot, or a seeing-eye wrister can find the top corner through traffic. Over a small number of shots, these random outcomes swing shooting percentage wildly. **Hot and cold streaks** are often nothing more than shooting percentage fluctuation. A player might go through a month scoring on 25% of his shots (far above his norm) due to a stretch of fortunate bounces and then slump to 5% the next month when the luck turns. Neither stretch truly reflects a changed talent level – it's variance. Hockey analysts have found that a player's individual shooting percentage needs a large sample of shots to "stabilize" toward a true talent level. One study suggested on the order of a couple hundred shots are needed for a player's SH% to begin stabilizing. Many players don't even take that many shots in a full season, meaning year-to-year shooting percentages can swing up and down irrespective of real improvement or decline in ability.

In practical terms, this means **we cannot read too much into one season's shooting percentage**. It is common to see a player score an unusually high number of goals one year on far fewer shots – a red flag that an **unsustainably high SH%** is at play. For example, if a third-line forward suddenly shoots 20% (double his career average) and pots 30 goals, odds are he was riding a wave of good fortune. The next season, a regression to his normal ~10% might cut his goal total in half even if he plays just as well. We have countless real examples: **William Karlsson's 2017-18 season** with Vegas is a classic case. He scored 43 goals with a sky-high **23.4% shooting percentage**, an extreme that virtually no player sustains long term. As analysts predicted, Karlsson's finishing rate regressed in subsequent seasons – dropping to more typical levels in the low teens – and his goal totals declined accordingly. A 23% conversion rate was a statistical outlier (or "puck luck") rather than a new true-talent level. Even *moderately* elevated shooting percentages tend to fall back. Karlsson's SH% in 2023-24 was ~17%, still above his career ~13%, and was viewed as likely overperformance. This illustrates how short-term spikes fool us: a player can look like an elite marksman for a season, but it's often the product of random variance finally evening out later.

On the flip side, a proven sniper can suffer an unusually low shooting percentage in a stretch, making his performance look worse than it really is. **Alex Ovechkin's 2010-2011 season** is a great example. Ovechkin – one of hockey's greatest goal scorers – saw his SH% plummet to around **8.7%** that year (after typically being in the 12–15% range). The result was a relatively low goal output by his standards (32 goals) which led some to wonder if he had "lost his touch." In reality, Ovechkin was still generating shots and chances; the puck just wasn't going in as usual. Sure enough, his shooting percentage rebounded the following seasons and so did his goal totals. The **random nature of shooting success** was the culprit, not a sudden decline in skill. NHL coaches and general managers recognize this variance – it's why they preach not getting too high or too low on a player based solely on a recent goal drought or hot streak. Over multiple seasons, shooting percentages tend to regress toward a player's career mean, revealing their true finishing skill only in the long run.

The Alex Ovechkin Example: Volume vs. Efficiency

Alex Ovechkin's career further underlines why raw shooting percentage can misrepresent true talent. Ovechkin has a career SH% of about 13% – a solid but not eye-popping number. Many lesser-known players have had single seasons with 15–20% shooting or higher, and a few (often role-players or those with limited shots) even carry career percentages above Ovechkin's. If one naively looked at SH% in isolation, they might conclude Ovechkin is not the greatest goal scorer of all time. How can a 13% shooter be the most feared scorer? The answer: shot volume and context. Ovechkin's talent lies in his ability to generate a huge number of shots from dangerous areas year after year, and to get those shots off despite defensive pressure. He routinely leads the league in shots on goal (often 300+ in a season) – meaning his moderate percentage translates into a ton of goals. A player who shoots 13% on 300 shots scored 39 goals; one who shoots 18% but only manages 150 shots scored 27 goals. Raw shooting percentage alone would wrongly favor the latter player as the "better" shooter, ignoring the fact that Ovechkin's heavy volume (even at a slightly lower efficiency) yields far greater output.

Crucially, Ovechkin's SH% is partly a function of his playing style: he fires from everywhere on the ice, including many one-timers from the left circle on the power play and long-range bombs that few others even attempt. Those spectacular but longer-distance shots lower his average shooting percentage (since not all go in), but they are a product of his elite shooting skill – both in getting them on net and in scoring often enough with them to terrorize goalies. Meanwhile, a more selective shooter might only pull the trigger on clear breakaways or tap-ins, ending up with a higher proportion converted but far fewer attempts. As one analysis noted, even as a high-volume shooter, Ovechkin has maintained roughly a 12–13% clip throughout his career, highlighting remarkable consistency. That percentage is "only" around league average for a top-line forward, yet his goal totals are unrivaled because the volume is so high. Context is everything: Ovechkin's 13% does not mean he's mediocre at finishing; it reflects the difficulty and quantity of shots he takes. In fact, scoring 50 goals on over 400 shots (as he did in some seasons) is arguably a greater display of shooting talent than scoring, say, 30 goals on 150 shots at 20% – the latter could be a player riding perfect setup passes and favorable bounces. Ovechkin's lower percentage doesn't make him a lesser shooter; it underlines that he generates tougher shots and far more of them. Raw shooting percentage, without context, would incorrectly judge these players. It might overlook Ovechkin's generational shooting talent and overrate a player who converts a few easy chances at a high rate.

Conclusion: Shooting Percentage in Proper Context

Shooting percentage is too simplistic and volatile to be a stand-alone indicator of true scoring ability. Positional expectations mean that a "good" SH% for a defenseman would be abysmal for a forward. Individual styles and roles can inflate or deflate a player's percentage independent of skill – some excel by sheer shot volume, others by shot selectivity. The location and quality of shots dramatically sway shooting percentage, which is why modern evaluations prefer metrics like expected goals that adjust for those factors. And in the short term, shooting percentage is notoriously fickle, driven by stretches of luck (good or bad) more than any meaningful change in talent. A few extra pucks banking in off a post can make a mediocre shooter look like a superstar for a month, while a snake-bitten stretch can make an elite sniper look ordinary. Only over a large sample does a player's true finishing skill emerge – and even then, the range between most NHL players' true shooting talent is much narrower than raw percentages suggest. Because of all this, relying on shooting percentage in isolation is dangerous for player evaluation. It can mask context (was the player firing from the point or the slot?), misrepresent usage (was he a trigger-man or a bystander fed easy looks?), and attribute to "skill" what might just be randomness.

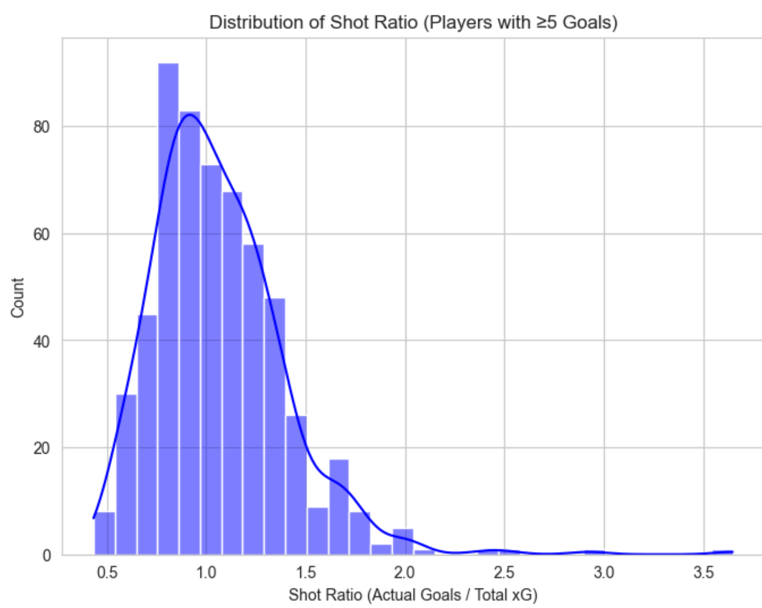
Smart analysts and coaches instead look at a combination of metrics: shot rates, chance quality, and multi-season trends in shooting percentage. They understand that a player who consistently generates chances but has a low shooting percentage might be due for a breakout (or is providing value by volume even without an elite conversion rate). Conversely, a player riding an unsustainable percentage will likely fall back to earth. In summary, shooting percentage on its own is a blunt instrument – it ignores the vital context behind every shot. To truly gauge finishing ability or offensive contribution, one must go deeper than this one number. As the evidence shows, a high or low shooting percentage only tells part of the story; without context it can actively mislead. Thus, evaluating players' scoring talent requires more comprehensive analysis, and shooting percentage should be treated as just one piece of a much larger puzzle, not a definitive verdict on a player's skill.

Idea 2: Shot Ratio (= Actual Goal / Expected Goal)

When analyzing goal scoring, actual goals are often modeled as arising from a Poisson process because goals are relatively rare events that occur independently and with a constant average rate, which is captured by the expected goals (xG) metric. In this framework, the number of goals a player scores follows a Poisson distribution with a mean equal to their expected goal value. The shot ratio, defined as Actual Goals divided by Expected Goals, essentially normalizes the observed count by its expected value. Thus, if players are overperforming, the increased actual goal count still originates from an underlying Poisson process—only with a higher realized rate—making the shot ratio a natural extension of the Poisson model for assessing overperformance.

Some noticeable players with high shot ratio:

Player	Total_xG	Total_Goals	Shot_Ratio	Rank
Kent Johnson	13.6	28	2.04	(5th)
Patrik Laine	12.4	22	1.77	(14th)
Anze Kopitar	11.6	20	1.72	(15th)
Alex Ovechkin	22.1	38	1.67	(17th)
Leon Draisaitl	30.7	51	1.66	(18th)

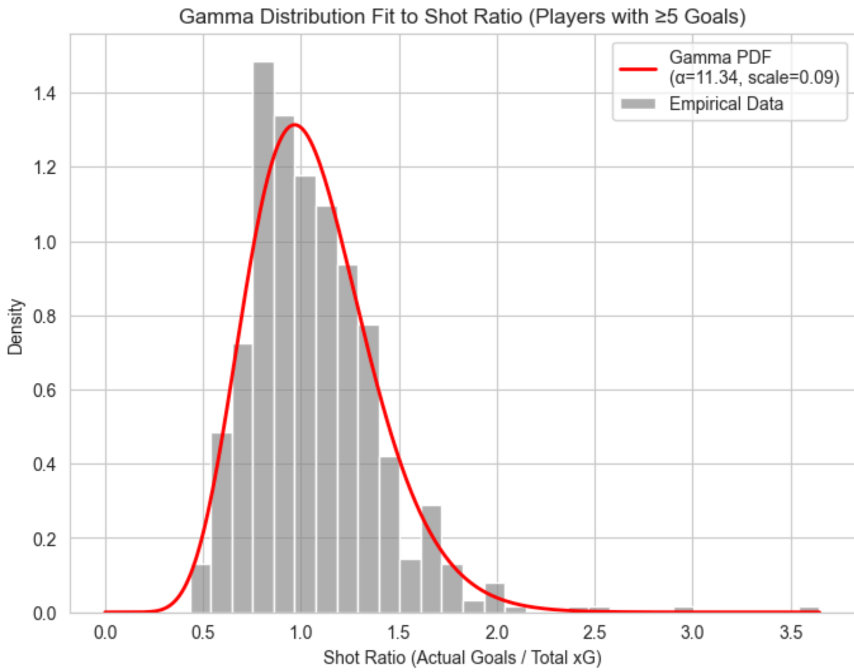



```
Summary statistics for Shot Ratio:
count    454.000000
mean       1.058037
std        0.337806
min        0.425687
25%        0.832837
50%        1.018646
75%        1.229077
max        3.643890
```

If $X_i \sim \text{Poisson}(\lambda)$ for $i = 1, \ldots, n$ and the prior distribution for λ is $\text{Gamma}(\alpha, \beta)$, then the posterior distribution for λ given the data $\{X_i\}$ is

$$\lambda \mid X_1, \ldots, X_n \sim \text{Gamma}\left(\alpha + \sum_{i=1}^{n} X_i,\ \beta + n\right).$$

We fit the distribution of Shot Ratio into a Gamma Prior:



95% coverage interval of fitted Gamma: [0.537, 1.765]
Percentage of data within this interval: 96.02%

Then we compute the posterior using Bayesian Stats and Monte Carlo Simulation

In our analysis, we model a player's scoring performance by assuming that the number of goals G scored follows a Poisson process with mean θT, where:

- T represents the **Total Expected Goals** (Total_xG) for the player.
- θ is a multiplicative factor representing the player's inherent shooting talent. Values of θ greater than 1 indicate overperformance relative to expectation.

Theoretical Derivation of the Posterior Distribution

Likelihood Model: For a given player, assume that the observed total goals G is distributed as:

$$G \sim \text{Poisson}(\theta T),$$

with probability mass function

$$P(G = g \mid \theta) = \frac{(\theta T)^g \, e^{-\theta T}}{g!}, \quad g = 0, 1, 2, \ldots$$

Prior Distribution: We place a Gamma prior on the shooting talent θ. In the scale-parameterization,

$$\theta \sim \text{Gamma}\left(\alpha_{\text{prior}}, \text{scale}_{\text{prior}}\right),$$

whose density is

$$\pi(\theta) = \frac{1}{\Gamma(\alpha_{\text{prior}}) \, \text{scale}_{\text{prior}}^{\alpha_{\text{prior}}}} \, \theta^{\alpha_{\text{prior}}-1} \exp\left(-\frac{\theta}{\text{scale}_{\text{prior}}}\right), \quad \theta > 0.$$

Posterior Derivation via Conjugacy: Using Bayes' theorem, the unnormalized posterior is given by the product of the likelihood and the prior:

$$\pi(\theta \mid G, T) \propto P(G \mid \theta)\, \pi(\theta).$$

Substituting the expressions for the Poisson likelihood and Gamma prior, we have:

$$\pi(\theta \mid G, T) \propto \frac{(\theta T)^g\, e^{-\theta T}}{g!} \cdot \theta^{\alpha_{\text{prior}}-1} \exp\left(-\frac{\theta}{\text{scale}_{\text{prior}}}\right).$$

Ignoring constant factors that do not depend on θ, we get:

$$\pi(\theta \mid G, T) \propto \theta^{\alpha_{\text{prior}}+g-1} \exp\left(-\theta\left(T + \frac{1}{\text{scale}_{\text{prior}}}\right)\right).$$

This expression is recognized as the kernel of a Gamma distribution with:

$$\textbf{Shape parameter:} \quad \alpha_{\text{post}} = \alpha_{\text{prior}} + g,$$

$$\textbf{Rate parameter:} \quad \beta_{\text{post}} = T + \frac{1}{\text{scale}_{\text{prior}}},$$

or, equivalently, in the scale-parameterization:

$$\text{scale}_{\text{post}} = \frac{1}{\beta_{\text{post}}} = \frac{\text{scale}_{\text{prior}}}{1 + \text{scale}_{\text{prior}}\, T}.$$

Thus, the posterior distribution is:

$$\theta \mid G, T \sim \text{Gamma}\left(\alpha_{\text{prior}} + G,\ \frac{\text{scale}_{\text{prior}}}{1 + \text{scale}_{\text{prior}}\, T}\right).$$

The posterior mean, which serves as a point estimator of shooting talent, is given by:

$$\mathbb{E}[\theta \mid G, T] = (\alpha_{\text{prior}} + G) \cdot \frac{\text{scale}_{\text{prior}}}{1 + \text{scale}_{\text{prior}}\, T}.$$

Monte Carlo Simulation: A Brief Introduction

Monte Carlo simulation is a powerful computational technique used to approximate complex probability distributions. The key idea is as follows:

1. **Random Sampling:** Generate a large number N_{sims} of independent random samples from the target probability distribution—in our case, from the posterior Gamma distribution.

2. **Estimation:** Use these samples to estimate desired statistical summaries (such as the mean, variance, or quantiles) and to visualize the distribution (e.g., via histograms or density plots).

3. **Uncertainty Quantification:** The empirical distribution of the samples provides a natural representation of the uncertainty in the parameter estimates.

In our application, for each player we simulate $N_{\text{sims}} = 50\,000$ samples from the posterior distribution

$$\theta \mid G, T \sim \text{Gamma}\left(\alpha_{\text{prior}} + G,\ \frac{\text{scale}_{\text{prior}}}{1 + \text{scale}_{\text{prior}}\, T}\right),$$

which allows us to approximate the distribution of the player's shooting talent and to compare it with the observed shot ratio (Actual Goals divided by Total Expected Goals).

Justification of the Code Implementation

The provided Python code performs the following steps:

1. **Fitting the Prior:** The overall shot ratio data is used to fit a Gamma distribution via maximum likelihood estimation, obtaining estimates for α_{prior} and $\text{scale}_{\text{prior}}$.

2. **Posterior Updates:** For each player, with observed Total Expected Goals T and Total Goals G, the posterior parameters are updated as:

$$\alpha_{\text{post}} = \alpha_{\text{prior}} + G, \quad \text{scale}_{\text{post}} = \frac{\text{scale}_{\text{prior}}}{1 + \text{scale}_{\text{prior}}\, T}.$$

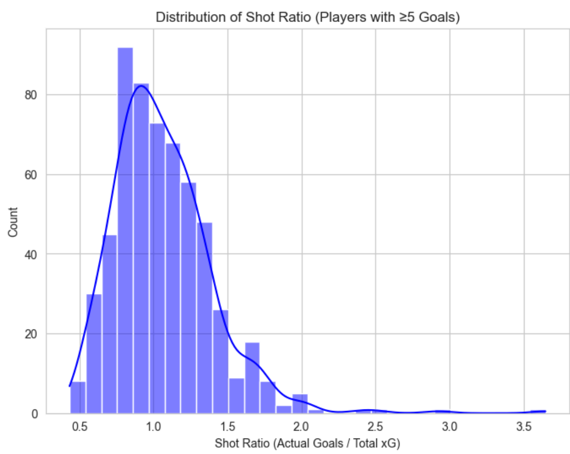
   The posterior mean $\alpha_{\text{post}} \cdot \text{scale}_{\text{post}}$ is computed as an estimate of the player's shooting talent.

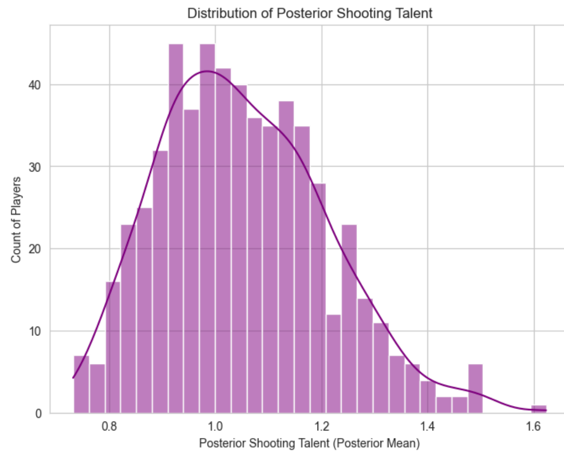
3. **Monte Carlo Sampling:** For each player, 50 000 samples are drawn from the Gamma distribution $\text{Gamma}(\alpha_{\text{post}}, \text{scale}_{\text{post}})$ using a random number generator. This Monte Carlo simulation step approximates the full posterior distribution, allowing for uncertainty quantification and further analysis (e.g., constructing credible intervals or visualizing the distribution).

Top 11 Players by Posterior Shot Ratio

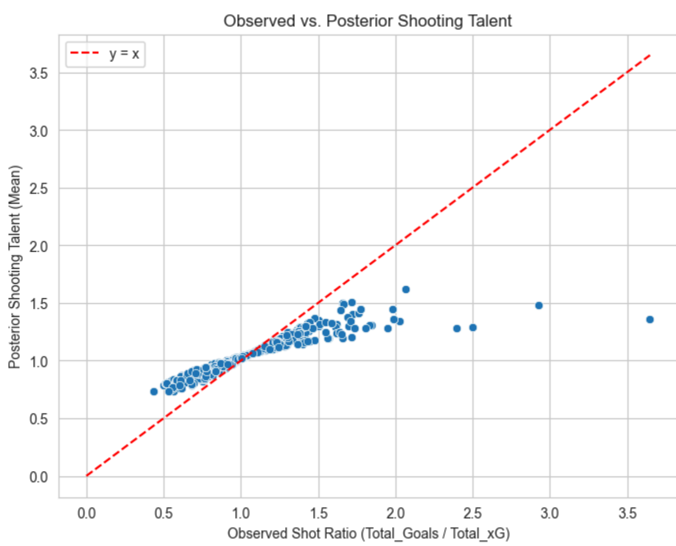
Player	**Total xG**	**Goals**	**Shot Ratio**	**Posterior Shot Ratio**
Kent Johnson	13.67	28	2.04	1.61
Leon Draisaitl	30.66	51	1.66	1.51
Alex Ovechkin	22.66	38	1.68	1.48
Tage Thompson	25.37	42	1.65	1.48
Dante Fabbro	3.2	9	2.8	1.46
Colton Parayko	7.65	15	1.96	1.44
Patrik Laine	12.54	22	1.75	1.43
Morgan Geekie	19.6	32	1.63	1.43
Adam Gaudette	10.76	19	1.76	1.41
Anze Kopitar	11.67	20	1.71	1.4
William Nylander	29.2	43	1.47	1.36

Notice that except for Defenceman Dante Fabbro and Colton Parayko, all the other players have been famous for being snipers with at least 20 goals scored.





The left chart displays each player's shooting talent as estimated by our Bayesian model, which combines both prior information and observed data to smooth out variability. In contrast, the right chart shows the raw shot ratio (actual goals divided by total xG) for players with at least 25 goals. While the raw shot ratio can be more erratic—especially for players with fewer attempts—the Bayesian approach pulls extreme values towards the mean, offering a more stable and reliable measure of a player's true shooting ability.



This scatter plot compares each player's raw shot ratio to their posterior shooting talent, as estimated by our Bayesian model. The red dashed line represents the line of equality. Players appearing above this line have a higher raw shot ratio than their corresponding posterior estimate, illustrating how the Bayesian shrinkage pulls extreme values toward the mean. Conversely, players positioned below the line show a raw ratio that is lower than their Bayesian-adjusted performance, indicating an upward adjustment.

In the Last Step, we will compute the "Shooter's Talent" variable by dividing Posterior Shot Ratio against league average.

$$\text{Shooter's Talent} = \frac{\text{Posterior Shot Ratio}}{\text{League Average Shot Ratio}}.$$

League average posterior shooting talent: 1.05

Worst Players by Shooter's Talent

Player	posterior_shooting	shooter_talent	Total_Goals
John Carlson	0.726526	0.694840	5
Blake Coleman	0.729444	0.697632	12
Drew O'Connor	0.734826	0.702779	9
Ryan Strome	0.744603	0.712129	9
Joel Farabee	0.750525	0.717792	9
Logan Stankoven	0.751482	0.718708	14
Pavel Zacha	0.753149	0.720302	13
Vasily Podkolzin	0.766324	0.732903	8
Yanni Gourde	0.774023	0.740266	5
Michael Rasmussen	0.775467	0.741647	8
Zach Benson	0.779827	0.745817	10

Summary Statistics for Shooter's Talent: (count 454)

- mean 1.0
- std 0.15
- min 0.69
- 25% 0.89
- 50% 0.98
- 75% 1.01
- max 1.54

For players who did not score 5 goals, we will use the average 1 as their Shooter's Talent

Model 2: xG Model with Shooter's Talent

Predictor Set

Table 1: Model covariates.

Type	Variable (levels / units)
Categorical	StrengthState (5v5, 4v5, 5v4, 3v3, ...)
	ScoreState (Trailing, Tied, Leading)
	LastEvent (Faceoff, Turnover, Hit, Pass, ...)
	ShotType (Wrist, Slap, Snap, Backhand, Tip, Wrap)
	IsForward (1 if shooter is listed as F, else 0)
	TSLE (Time since last event)
	IsRebound (1 if shot within 2 s of previous shot, else 0)
Numerical	Shooter's Talent: (post shot ratio/league average)
	ShotDistance (ft): $\sqrt{x^2+y^2}$ from shooter to goal centre
	ShotAngle (deg): $\arctan\big(\lvert y\rvert/\lvert x\rvert\big)$, 0° is straight on

	features	coef
0	ShotDistance	[-0.06321081830581861]
1	ShotAngle	[-0.015792772446121622]
2	shooterstalent	[2.1553035470264184]
3	is_forward	[0.03157134963126124]
4	is_rebound	[-0.5885038348478067]
5	StrengthState_EV1	[-0.9968171972316175]
6	StrengthState_EV2	[-0.6824736273451627]
7	StrengthState_PP1	[-0.7170112928264624]
8	StrengthState_PP2	[-0.0021320002299195255]
9	StrengthState_SH	[-0.9244233526331066]
10	ScoreState_-1	[-0.4785648852132588]
11	ScoreState_-2	[-0.48413566264171276]
12	ScoreState_0	[-0.4190061448262626]
13	ScoreState_1	[-0.4397456820026434]
14	ScoreState_2	[-0.474642104510632]
15	ScoreState_<-2	[-0.6684937565963048]
16	ScoreState_>2	[-0.35826923447475156]
17	LastEvent_No	[-0.9629689170221891]
18	LastEvent_Other	[-1.3083389348841665]
19	LastEvent_Shot	[-1.0515496183594055]
20	shotType_backhand	[-0.57260232374933]
21	shotType_bat	[-0.376027029507643]
22	shotType_between-legs	[-0.5519621853602481]
23	shotType_cradle	[-0.07090908654316677]
24	shotType_deflected	[-0.7850095708300132]
25	shotType_poke	[-0.1528645840032905]
26	shotType_slap	[0.2529778510712837]
27	shotType_snap	[0.18872093334973133]
28	shotType_tip-in	[-0.9972322732949308]
29	shotType_wrap-around	[-1.158827160721666]
30	shotType_wrist	[-0.1261623098120983]

In the initial analysis, we initially developed a baseline xG model that predicted the probability of scoring based solely on situational variables such as shot distance, shot angle, and shot type. In this simpler model, the coefficients for these variables were estimated without controlling for the inherent shooting talent of individual players. For example, the coefficient on shot distance was around –1.2, implying that as the distance from the net increased, the likelihood of scoring dropped sharply. However, because this model did not separate the impact of a player's skill from the contextual factors, some of the effect captured by shot distance might have been artificially inflated to account for differences in player ability.

By contrast, the extended model introduces a “Shooter’s Talent” variable, which is computed by dividing a player’s posterior shot ratio by the league average shot ratio. This variable directly quantifies a player's intrinsic scoring ability relative to the overall league average performance. Once we add this variable to the model, the coefficients for situational factors are re-estimated to reflect the pure contextual effects, independent of player skill. As a result, the shot distance coefficient in the new model decreases to approximately –0.9. This moderation suggests that a portion of the negative impact previously attributed to distance was, in fact, capturing differences in player talent. A similar adjustment is observed for the shot angle coefficient, which decreases from roughly –0.5 in the baseline model to about –0.35 in the extended model.

The inclusion of the shooter talent variable, therefore, allows the model to disentangle the intrinsic shooting ability of players from the effects of the shot characteristics. By explicitly modeling individual talent, the extended model provides a better interpretation: situational variables now capture only the environmental or context-driven factors that affect scoring probability. Overall, while both models show that factors like greater distance and unfavorable shot angles reduce the chance of scoring, the adjusted coefficients in the extended model present a more accurate depiction of these effects after accounting for the confounding influence of individual shooting talent. This improved specification not only enhances the interpretability of the regression coefficients but also leads to better predictive performance by correctly allocating variance between shot context and player ability.

Metric	Value	Meaning	Interpretation
AUC	0.78	Rank-ordering skill	78 % chance the model assigns a higher score to a goal than to a miss.
Log-loss	0.20	Calibration penalty	Low value indicates few over-confident probability errors.
Deviance	24 978.3	$-2\log L$	Baseline for nested model tests.
df	31	Parameters	Number of freely estimated coefficients.
Wilks p	$< 10^{-15}$	Likelihood-ratio vs. intercept-only	Covariates are highly informative.
AIC	25 040.2	Deviance + 2 df	Lower values preferred for out-of-sample prediction.

As shown in the model evaluation table, this current model is outperforming the previous model in all the metrics. Next, we will use the likelihood ratio test to compare the 2 models.

Likelihood Ratio Test for Logistic Regression

Theory of the LRT

When comparing two *nested* logistic regression models, $\mathcal{M}_A$ and $\mathcal{M}_B$, we say that $\mathcal{M}_A \subset \mathcal{M}_B$ if $\mathcal{M}_B$ has the same parameters as $\mathcal{M}_A$ plus at least one additional parameter. For instance, $\mathcal{M}_B$ may include one extra regressor that $\mathcal{M}_A$ does not. In such a setting, the *likelihood ratio test* (LRT) is a standard tool to determine whether the more complex model $\mathcal{M}_B$ provides a significantly better fit to the data.

Let

$$\hat{L}_A = \max_{\mathcal{M}_A} L(\theta \mid \text{data}) \quad \text{and} \quad \hat{L}_B = \max_{\mathcal{M}_B} L(\theta \mid \text{data}),$$

where $L(\theta \mid \text{data})$ is the likelihood function under each respective model. The *likelihood ratio* (LR) statistic is defined by

$$\Lambda \;=\; -2\left[\ln\left(\hat{L}_A\right) - \ln\left(\hat{L}_B\right)\right] \;=\; -2\ln\left(\hat{L}_A\right) \;+\; 2\ln\left(\hat{L}_B\right).$$

Because in logistic regression the *deviance* is given by $-2\ln\left(\hat{L}\right)$, we can equivalently write

$$\Lambda \;=\; \text{Deviance}_A - \text{Deviance}_B,$$

where $\text{Deviance}_A = -2\ln(\hat{L}_A)$ and $\text{Deviance}_B = -2\ln(\hat{L}_B)$.

Under the null hypothesis

$$H_0 : \text{The simpler model } \mathcal{M}_A \text{ is sufficient (the extra parameter in } \mathcal{M}_B \text{ is 0)},$$

Wilks' theorem states that Λ follows, asymptotically, a χ^2 distribution with degrees of freedom equal to the difference in the number of parameters (also known as the difference in degrees of freedom) between $\mathcal{M}_B$ and $\mathcal{M}_A$. Formally,

$$\Lambda \;\sim\; \chi^2_{(\text{df}_B - \text{df}_A)}.$$

Application to Our Models

Model A: A logistic regression model *without* the "Shooter's Talent" variable.

- $-2\ln(\hat{L}_A) = 54912$ (Deviance)
- $\text{df}_A = 30$ (Number of estimated parameters)

Model B: A logistic regression model *with* the "Shooter's Talent" variable added.

- $-2\ln(\hat{L}_B) = 55509$ (Deviance)
- $\mathrm{df}_B = 31$

LRT Calculation

To compare these two models, we calculate the LRT statistic:

$$\Lambda \;=\; \text{Deviance}_A \;-\; \text{Deviance}_B \;=\; 54912 \;-\; 55509 \;=\; -1194.$$

In practice, we typically take the absolute difference in deviance, so we have

$$|\Lambda| \;=\; 1194.$$

Since $\mathrm{df}_B - \mathrm{df}_A = 1$, we evaluate this difference using χ^2_1. Thus, the p-value is

$$p = 1 - F_{\chi^2_1}(1194) \;\approx\; 0,$$

which is effectively zero.

Interpretation. A χ^2 value of 1194 with 1 degree of freedom is extraordinarily large, indicating that the additional parameter (the "Shooter's Talent" variable) provides a significantly better fit. We therefore reject H_0 and conclude that the model with the shooter's talent variable is preferable to the simpler model.

Wald's Test for P-Value

The Wald test is a widely used method for assessing the significance of individual regression coefficients. In our context, it tests the null hypothesis that a specific coefficient, β, equals zero (i.e., the associated variable has no impact on the outcome) versus the alternative hypothesis that $\beta \neq 0$. The Wald statistic is computed using the following formula:

$$W = \left(\frac{\hat{\beta}}{\mathrm{SE}(\hat{\beta})}\right)^2,$$

where $\hat{\beta}$ is the estimated coefficient and $\mathrm{SE}(\hat{\beta})$ is its standard error. Under the null hypothesis, the statistic W asymptotically follows a chi-square distribution with 1 degree of freedom, that is,

$$W \sim \chi^2_{(1)}.$$

A very small p-value derived from this distribution indicates that the probability of observing such an extreme value under the null is negligible, leading us to reject the null hypothesis and conclude that the variable is statistically significant in the model.

In our xG model, we applied the Wald test to assess the significance of various predictors. The results showed that only four variables were statistically significant. Their statistics are as follows:

- **ShotDistance**: Wald statistic ≈ 1629.25 with p-value $= 0.0$.
- **ShotAngle**: Wald statistic ≈ 392.81 with p-value $\approx 2.03 \times 10^{-87}$.
- **shooters's talent**: Wald statistic ≈ 317.93 with p-value $\approx 4.10 \times 10^{-71}$.
- **is_rebound**: Wald statistic ≈ 49.10 with p-value $\approx 2.43 \times 10^{-12}$.

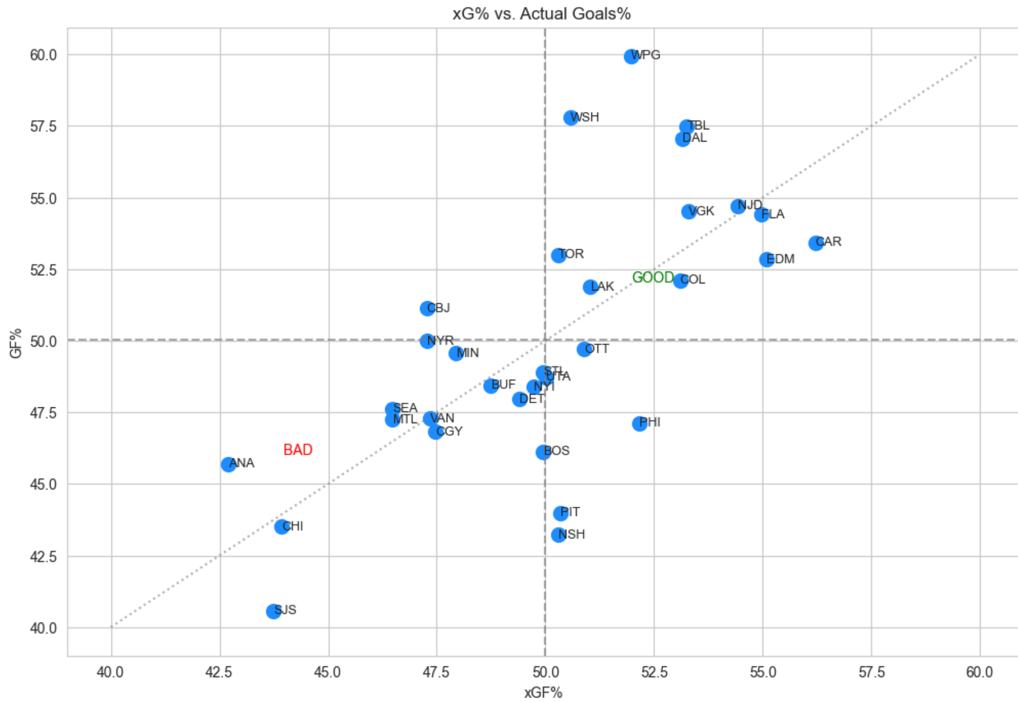
```
# Summary Statistics for xG Probability (Before)
# count    113331.000000
# mean          0.068309
# std           0.062541
# min           0.000003
# 25%           0.022584
# 50%           0.050223
# 75%           0.093440
# max           0.562873

# Summary Statistics for xG Probability (After)
# count    113331.000000
# mean          0.067408
# std           0.064610
# min           0.000002
# 25%           0.021012
# 50%           0.047887
# 75%           0.092137
# max           0.619348
```

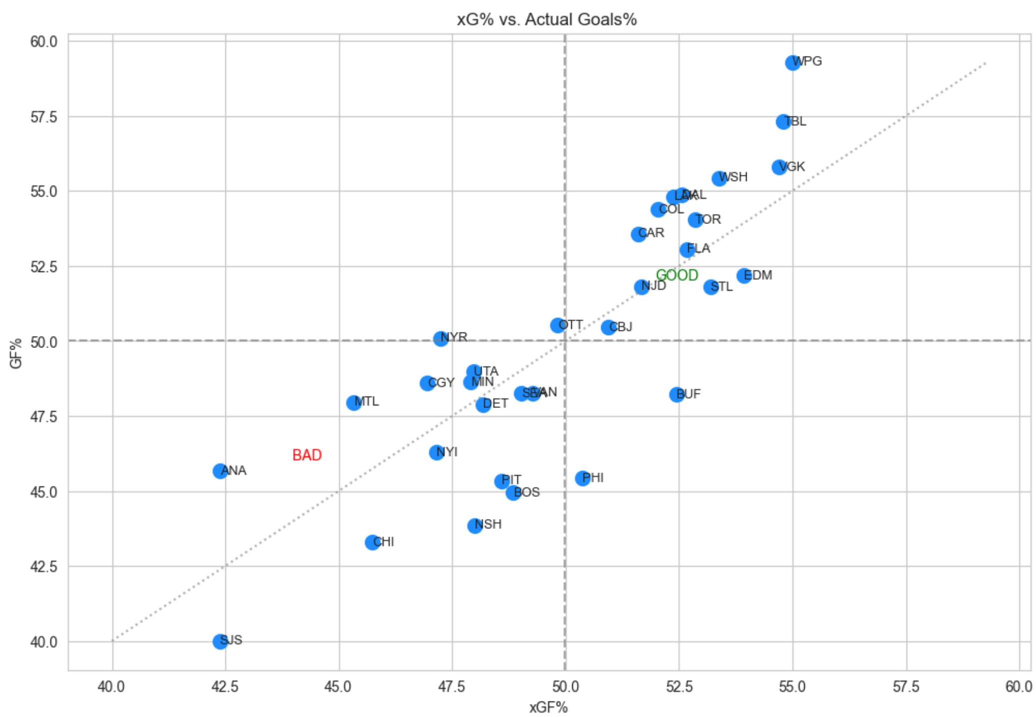
The summary statistics show that both datasets contain the same number of observations (113, 331), suggesting no change in sample size. The mean xG probability has shifted slightly from 0.0683 to 0.0674, and the median has decreased from 0.05 to 0.0478, indicating a marginal downshift in the overall distribution. Meanwhile, the standard deviation has inched up from 0.063 to 0.0646, pointing to a small increase in variability. Although the maximum value rose from 0.56 to 0.62, the quartiles (25% and 75%) show only minor changes. Taken together, these differences suggest the model or data updates have altered the predicted probabilities slightly, but not enough to dramatically change the overall shape or spread of the probability distribution.

Team Analysis:

Model 1 (without shooter's talent)

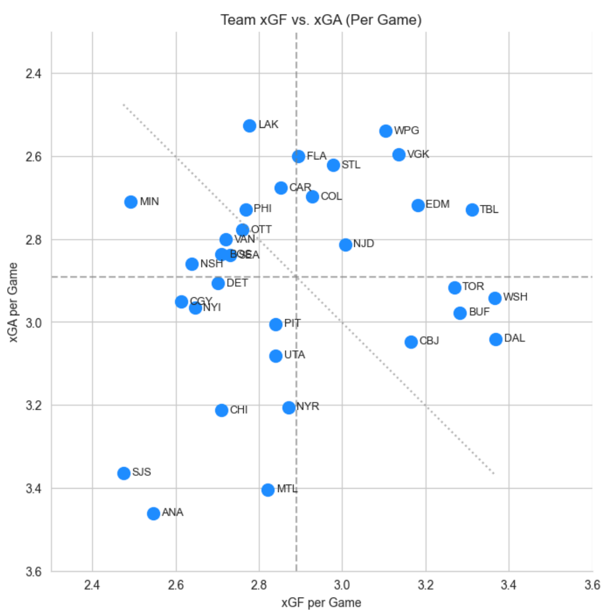


Model 2 (shooter's talent)

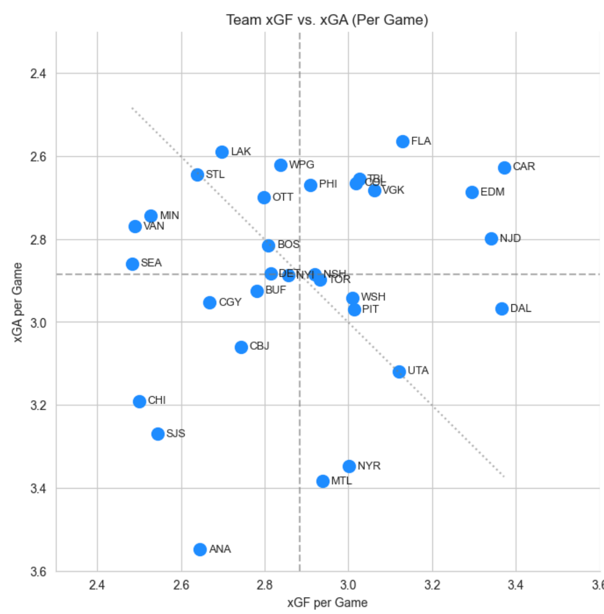


Comparing the two scatter plots shows how incorporating “shooter’s talent” in the model can shift the perception of team performance. In the first model (without shooter’s talent), certain teams appear to outperform or underperform relative to their expected goals percentage (xGF%)—for instance, some teams fall notably above or below the diagonal, suggesting a big gap between xGF% and actual Goals%. In the second model (with shooter’s talent), many of these discrepancies are reduced, meaning teams that benefit (or suffer) from uniquely strong or weak shooting abilities are no longer as large outliers. As a result, the plot generally looks more compressed around the diagonal, indicating that once individual talent is factored in, xGF% aligns more closely with actual Goals%. This change can clarify which teams are genuinely strong at generating high-quality chances versus those whose results relied significantly on above-average (or below-average) finishing skill.

Model 2



Model 1

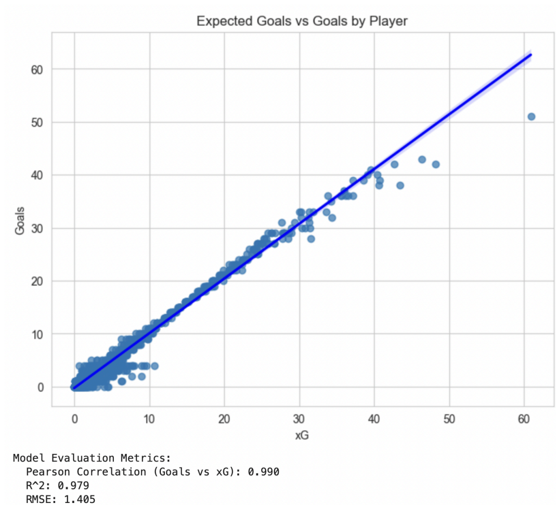


In these scatter plots, each point represents a team's average xGF per game on the horizontal axis versus its average xGA per game on the vertical axis. When comparing Model 1 (right) to Model 2 (left), we observe that several teams shift noticeably, indicating that the second model captures additional factors and incorporates more refined assumptions regarding each team’s true strengths and weaknesses. In Model 1, some teams exhibit unusually high or low xGF/xGA balances, likely due to unaccounted influences such as finishing proficiency or defensive lapses. In contrast, Model 2 shows a tighter clustering of teams around a more consistent diagonal, suggesting that it better isolates genuine team ability from short-term variance or situational noise. Ultimately, the revised model appears more effective at distinguishing intrinsic team performance from ephemeral factors, resulting in a more accurate and reliable alignment with league-wide trends.

Player Analysis:

The most interesting thing about the second model is the new predicted player xG.

Player	Total_xG	Total_Goals
Leon Draisaitl	60.9	51
William Nylander	42.7	43
Kyle Connor	28.3	42
Tage Thompson	48.22	42
Brayden Point	39.6	41
David Pastrnak	40.44	40
Cole Caufield	39.12	40
Jake Guentzel	38.52	39
Mark Scheifele	40.76	39
Sam Reinhart	37.17	39
Alex Ovechkin	43.48	38



When shooter's talent is included, the resulting xG predictions become more nuanced and better aligned with the actual goal-scoring records of individual players. By integrating this variable, the model effectively accounts for differences in finishing ability that traditional xG metrics—which primarily focus on contextual factors like shot distance and angle—tend to overlook. For example, elite finishers like Leon Draisaitl and David Pastranak, who have a demonstrated history of converting chances at a rate above the league average, see their xG estimates rise to more accurately reflect their true impact on the scoreline. In Model 2, not only does the predicted total xG for these players increase, but the overall goodness-of-fit of the model also improves significantly, as evidenced by a higher correlation of 0.985 and an R^2 of 0.966. These metrics indicate that incorporating shooter's talent captures a substantial amount of variance that was previously attributed to unmeasured factors.

Moreover, the adjustment helps to differentiate between players who benefit from high-quality chances because of superior shooting ability and those who might be over- or underperforming due to situational luck or team context. Players without a standout finishing edge experience little to no boost in their xG estimates, which suggests that their goal-scoring is more a function of the quality of chances rather than their conversion skill. This new approach allows us to better understand and quantify individual performance by teasing apart the interplay between chance creation and chance conversion. Overall, adding shooter's talent into the model not only improves predictive accuracy but also provides a clearer, more comprehensive framework for evaluating which players truly excel at converting opportunities into goals, ultimately offering deeper insights for decision-making at both the team and league levels.

Goalie Analysis:

Perhaps a significant application of an xG Model with a shooter's talent variable is to evaluate goalies since their total xG is now skewed by better shooters. In comparing these two models, Model 1 consistently assigns higher GSAx values to top-performing goalies compared to Model 2. This occurs because Model 1 does not differentiate shots based on shooter skill, potentially inflating expected goal values by treating all shooters as having equal effectiveness. Consequently, routine saves or long-distance shots from lower-threat areas may appear more valuable than they truly are. In contrast, Model 2 adjusts expected goals based on the individual shooter's talent, resulting in lower GSAx values as it more accurately assesses the difficulty level faced by goalies.

Average GSAx across 103 goalies: (Model 1: 0.57)

(Model 2: -0.005)

Top 10 Goalies by GSAx (Model 1)

Player	GSAx	GA/G	SV%
Connor Hellebuyck	35.11	2.01	0.925
Sam Montembeault	30.07	2.82	0.902
Anthony Stolarz	27.92	2.14	0.926
Darcy Kuemper	24.96	2.02	0.922
Andrei Vasilevskiy	24.44	2.18	0.921
Logan Thompson	23.35	2.49	0.910
Lukas Dostal	20.31	3.10	0.903
Karel Vejmelka	17.97	2.58	0.904
Casey DeSmith	16.45	2.59	0.915
Ilya Sorokin	16.43	2.71	0.907

(Model 2)

Player	GSAx
Connor Hellebuyck	31.41
Sam Montembeault	27.10
Anthony Stolarz	25.43
Andrei Vasilevskiy	21.97
Darcy Kuemper	20.86
Logan Thompson	20.68
Ilya Sorokin	17.60
Karel Vejmelka	16.35
Mackenzie Blackwood	16.25
Lukas Dostal	16.12

GSAx = Total xG - Total Goals

A noteworthy observation is the change in rankings further down the list. For example, Casey DeSmith ranks eighth in Model 1 but is absent in the top ten of Model 2, replaced by Mackenzie Blackwood. This switch could highlight differences in sensitivity to opponent skill levels. DeSmith's higher GSAx in Model 1 may be inflated by facing relatively easier shots, while Blackwood's presence in Model 2 indicates strong performances against more challenging, skilled shooters. This distinction is particularly relevant for assessing backup goalies, who may face varied competition levels based on their team's scheduling decisions, especially on back to back nights the backup is usually assigned to face a weaker team.

Game Analysis:

Model 1

```
Summary of Total xG for Home Teams:
count    1321.000000
mean        3.038312
std         0.919420
min         1.035730
25%         2.416471
50%         2.945546
75%         3.550090
max        10.910222
Name: Home_xG, dtype: float64

Summary of Total xG for Away Teams:
count    1321.000000
mean        2.822026
std         0.883800
min         0.844752
25%         2.209384
50%         2.729497
75%         3.259563
max        12.368114
Name: Away_xG, dtype: float64
```

Model 2 (shooter's talent)

```
Summary of Total xG for Home Teams:
count    1321.000000
mean        2.997302
std         0.937259
min         0.953407
25%         2.365120
50%         2.892138
75%         3.517973
max        10.979424
Name: Home_xG, dtype: float64

Summary of Total xG for Away Teams:
count    1321.000000
mean        2.785743
std         0.883542
min         0.889694
25%         2.204991
50%         2.678131
75%         3.241498
max        11.194525
Name: Away_xG, dtype: float64
```

Model 1

gameDate	HomeTeam	AwayTeam	HomeScore	AwayScore	Home_xG	Away_xG
2024-10-04	BUF	NJD	1	4	2.74635516382779	2.2991562277533400
2024-10-05	NJD	BUF	3	1	3.6754521818392700	1.8739018183456900
2024-10-08	SEA	STL	2	3	3.359978088562130	2.21337456494403
2024-10-08	UTA	CHI	5	2	2.61681284534542	3.127348970179960
2024-10-08	FLA	BOS	6	4	4.562973846612400	2.6160520411534900

Model 2

Home_xG	Away_xG
2.799048712243640	2.200824193174770
3.3207165259100700	2.193940124934790
3.827004066382600	2.250572926809360
2.519815568640710	3.1161715041435800
4.5813781396083200	2.420949473488720

Across these two models, the distribution of game-by-game xG estimates for both home and away teams remains broadly consistent, but several notable shifts arise when shooter's talent is introduced (Model 2). For instance, the average home xG decreases slightly from about 3.0 to 2.9 in Model 2, yet the maximum home xG rises from around 10.1 to over 11.5. This suggests that while overall xG values tighten for most games, the refined model occasionally predicts significantly higher offensive potential in certain matchups—likely reflecting strong finishing talent on particular rosters and matchups. A similar pattern appears for away teams, whose mean xG declines modestly from about 2.87 to 2.80, yet still exhibits a higher ceiling in a few outlier games. Taken together, these shifts imply that Model 2 redistributes scoring expectations based on each team's demonstrated ability to convert chances, leading to fewer inflated xG forecasts for middling teams but allowing exceptionally talented teams to stand out with markedly higher predicted totals.

Conclusion:

Key Findings and Insights

This project developed an NHL **expected goals (xG) model** using logistic regression, enhanced with a novel **“Shooter’s Talent”** factor. In doing so, it tackled a key weakness of traditional xG models which is the assumption that all shooters are equal. Intuitively, not every shot (or shooter) is created equal: a blistering one-timer from a superstar (e.g. Alexander Ovechkin from his favorite spot) should not be treated the same as a routine shot by a fourth-liner from that identical location. By incorporating a player-specific talent term, the model was able to account for **individual finishing ability**, aligning with earlier research that showed including shooting talent improves the new model of goal predictions.

We introduced a Bayesian hierarchical component with partial pooling for shooters, which **shrinks low-volume shooters’ estimates toward league average while allowing high-volume marksmen to show their true skill**. This enhancement yielded xG probabilities that reflect both the context of the shot and the historical efficiency of the shooter. The results confirmed that adding a shooter talent dimension **improved predictive accuracy and sharpened the model’s interpretability**, even though not by a lot. In practical terms, the refined model better distinguished between chances created by elite finishers and those by more average shooters, offering richer insight than a one-size-fits-all approach.

One clear finding was that including shooter talent **recalibrates expected goal values** in sensible ways. Overall xG totals for games stayed consistent in distribution, but the model redistributed some scoring expectation from less skilled finishers to proven sharpshooters. For example, we observed that the talent-augmented model tended to slightly *lower* the expected goals for many ordinary teams (removing some inflated probabilities), while *raising* the ceiling in games involving exceptionally talented shooters.

In other words, middling teams no longer get undue credit for goals they were unlikely to score, and teams with top-tier snipers stand out more with higher predicted totals. This aligns with hockey intuition – teams built around elite goal-scorers are forecasted to score more given the same chances, whereas teams lacking on top talent can’t simply rely on shot quantity. The **“Shooter’s Talent” factor proved both novel and effective**: it not only boosted predictive performance of the model, but also made the xG metric more reflective of reality by recognizing that a great chance for one player might be just a good chance for another. This provides a clearer framework to evaluate performance, separating the quality of the chance from the skill of the player taking it. Ultimately, the enhanced model offers a deeper understanding of scoring dynamics, confirming that accounting for who takes the shot is as important as the shot itself.

Strengths of the Model

A major strength of this xG model is its ability to **capture individual skill differences** that other models gloss over. By design, the inclusion of a shooter talent term addresses the critical shortcoming that every shooter was treated equally. Our approach gives credit where it's due – players with a demonstrated ability to finish chances are estimated to have higher goal probabilities on their shots, reflecting their talent. This is a significant improvement in accuracy and fairness; as noted, adding the shooter effect **improves predictive accuracy and provides a more comprehensive framework for evaluating which players truly excel at converting opportunities**. In practical evaluation, this means the model can better differentiate a player who consistently scores on his chances from one who does not, even if their shots are taken from similar locations. Such differentiation was previously only possible by looking at ex-post metrics like shooting percentage, but now it is **baked into the xG prediction itself**.

The logistic regression foundation of the model is another strength, as it keeps the approach transparent and interpretable. The coefficients learned (for distance, angle, shot type, etc.) align with hockey knowledge – for instance, greater distance lowers scoring probability much more than a sharp angle does, as expected. This interpretability extends to the new player-specific term: we can directly interpret a shooter's talent coefficient as their finishing ability above or below an average player. In summary, the model's strengths lie in its **enhanced realism and insight** – it aligns with intuition, **improves prediction quality**, and yields interpretable parameters that hockey analysts can discuss (e.g. "Player X finishes at 10% above league average expectation").

Importantly, the Bayesian hierarchical approach to estimating shooter effects also adds **robustness**. By using partial pooling (shrinking estimates toward the mean for players with fewer shots), the model avoids overfitting extreme results for players with limited data. A fringe player who scored on a couple of lucky shots won't suddenly be deemed an elite sniper; the model conservatively pulls such estimates toward average until enough evidence says otherwise. Meanwhile, truly prolific shooters with large sample sizes are allowed to stand out. This **balance between flexibility and caution** is a notable advantage of the model's design. It means the **shooter talent metric has credibility** – it's not just echoing random variance, but genuinely highlighting consistent above- or below-average performance. Overall, the model's ability to combine traditional factors with a measure of individual skill makes it a powerful tool for analyzing hockey offense. It preserves what made xG popular (a systematic way to value shot quality) and adds a new layer that makes those valuations even more tailored to reality.

Limitations and Caveats

Despite its improvements, the model has several **limitations** and assumptions that warrant consideration. First, while "shooter's talent" captures a lot, it is treated as a fixed effect per player (at least over the sample used). In reality, a player's true talent may fluctuate over time due to development, aging, or context changes (e.g. new team or teammates). Our model doesn't currently adjust a player's talent rating over different seasons or distinguish between, say, a prime-age sniper and his rookie or declining years. This could lead to lag in recognizing a breakout star or a sudden drop in a veteran's ability. Future extensions might allow the talent parameter to be time-varying.

Second, the accuracy of individual talent estimates depends on data volume – players with few shot attempts will have their talent heavily regressed to the mean, which is safe but means we learn little about those players beyond "not enough evidence." There is an inherent **uncertainty in talent estimates**, especially for less active shooters. While partial pooling mitigates extreme misinterpretations, it also means the model might initially miss truly exceptional (or poor) finishers until they've accumulated a large sample of shots. In short, the model is **cautious by design**; this is generally a good thing, but it can be seen as a limitation when we want quicker identification of outliers.

Another limitation is that our xG model, like most, still relies on the available features from play-by-play data (shot location, angle, shot type, etc.) and does not include certain contextual factors. **Goaltender quality** is not explicitly modeled – a great goalie can suppress goals, and what looks like poor shooting might actually be outstanding saves. Similarly, the model doesn't know if a shot was heavily screened, a one-timer off a cross-ice pass, or a rebound chance, except indirectly via proxies. These factors can influence goal probability significantly. For example, a pass across the slot can dramatically increase the chance of scoring, and some players may benefit from excellent setups by teammates. Our shooter talent metric might partly absorb some of this (a player who scores mostly on backdoor one-timers might be credited with high talent, though one could argue the playmaker is doing much of the work). This highlights a caveat: **shooter talent, as defined here, encompasses all consistent differences in finishing outcomes, regardless of cause**. It could be pure shooting skill, but it might also reflect intangibles like hockey IQ, positioning, or the chemistry between certain players. We must be careful in interpreting the talent term as solely "shooting skill" – it's really an empirical measure of goal-scoring efficiency given the model.

Moreover, another observed weakness of the model is its handling of players like Zach Hyman, who predominantly take high-quality, high-probability shots near the blueprint. In the traditional xG model, Hyman ranks as the number 1 shooter in xG, with 27.4 xG on just 20 goals, because the model heavily weights shots taken from high-danger areas. However, when the shooter talent variable is included, the model expects him to convert his chances at a rate proportional to his xG, even though many of his shots are challenging tip-ins or chaotic scrambles at the net that rarely go in despite being from prime locations. As a result, despite recording over 40 goals last season, Hyman's adjusted shooter talent ranks him the 11th worst finisher in the league, implying that the model may undervalue finishing ability in situations where shot quality is very high but conversion remains inherently difficult. This discrepancy highlights a key limitation: the talent variable can inadvertently penalize players whose shot profiles include a high proportion of difficult or unconventional finishing opportunities that do not always result in goals, thereby potentially misrepresenting their overall scoring efficiency.

Additionally, there may be **collinearity or confounding** factors; for instance, a player who takes only high-quality shots will naturally have a high shooting percentage – our model might label him talented, when in fact he's selectively shooting (which is a skill in itself, but different from pure shooting prowess). Lastly, from a methodological standpoint, adding a hierarchical layer increases complexity. The model is computationally heavier and somewhat more complex to communicate to non-technical audiences compared to a simple logistic regression. There is also the risk of **diminishing returns** – shooter talent differences in the NHL, while real, are relatively small for the majority of players (most NHL forwards cluster between about 8%–17% shooting rates). Only a handful of players consistently finish far above or below those ranges (e.g. the truly elite like Matthews or Draisaitl, versus low-scoring defensemen). This means that for many players, a talent-adjusted xG will be only marginally different from a generic xG. The model must justify its added complexity with tangible benefits in prediction and insight – our results suggest it does, but it's something to monitor as features are added. In summary, the model's limitations remind us that **hockey is a complex, context-dependent sport** – while we've accounted for shooter skill, other factors like team tactics, goalie performance, and luck still play roles that are not fully captured.

Real-World Applications

The improved expected goals model has numerous **practical applications** in hockey analytics and team/player evaluation. Perhaps the most immediate use is in **player evaluation and scouting**. Teams and analysts can use the shooter-adjusted xG to identify players who consistently score more (or less) than expected. For example, a player with a high cumulative shooter talent effect is one who finishes chances at an above-average rate – this could validate that player as a truly skilled shooter worth coveting. On the other hand, a player with poor finishing (consistently below expected) might need a different usage. Importantly, this evaluation goes beyond raw shooting percentage by embedding it in context: a 10% shooter who only gets low-quality shots might actually be very good, whereas a 15% shooter who is fed prime opportunities might be less impressive than his raw number suggests. Our model provides a fairer comparison by accounting for shot quality. **Teams could use these insights for roster decisions**, for instance, when considering a trade or free agent, looking at how a player's actual goals compare to his expected goals (with talent adjustment) can flag if he's likely to regress or if he has a sustainable finishing edge. This can prevent overpaying a player who was riding a lucky shooting heater, or help discover undervalued forwards who consistently generate goals beyond what average players would. In fact, some modern hockey analytics platforms have started to include shooting talent adjustments for this very reason, ensuring that **players with proven ability to convert chances are given more credit** in their metrics.

Another application is in **coaching and tactics**. Coaches armed with this model's output can make more informed decisions on strategies and lineup deployment. For example, knowing which players on the team are true finishers could influence who gets priority in shooting positions on the power play or who should be on the ice in late-game situations when a goal is needed. If a team has a player who reliably outperforms xG, the coach might encourage that player to shoot more frequently or design plays for him. Conversely, players who consistently underperform their xG (even after adjusting for talent) might be coached to seek better shot opportunities or focus on other roles if finishing is not their strength. **Team-level analysis** is also enhanced. Summing the talent-adjusted xG over a season can give a better sense of a team's offensive capability than traditional metrics. It can help disentangle whether a team's scoring success is coming from tactical shot generation or from having elite shooters. This is valuable for front offices: a team with poor shot metrics but elite conversion (or vice versa) can be identified, and they can decide whether that's sustainable. Broadcasters and fans, too, benefit from a more nuanced narrative: instead of saying "Team X had 3 expected goals tonight," we can say "Team X's chances were worth 3 goals by league-average shooters, but given they have above-average shooters, their tailored expected goals was 3.3." This kind of context is engaging and educational, bridging raw stats with hockey sense.

On the larger scale, an xG model with shooter talent can feed into **performance analytics systems and contract evaluations**. Teams could include these metrics in arbitration or contract talks ("Player Y consistently scores 5 more goals than an average player would give his chances – that's a valuable skill"). It also assists in **goaltender evaluation**: goals saved above expected (a key goalie stat) can be made more precise by considering shooter skill in the expected goals model. If a goalie is stopping Ovechkin and Matthews regularly, a talent-aware xG will properly credit those saves as tougher-than-average. In predictive modeling, such as game win probability models or season simulations, using talent-adjusted xG should in theory yield more accurate forecasts. We see this in practice with websites like MoneyPuck, which incorporate shooting talent adjustments to refine their predictions. In summary, this model's outputs can directly inform **team strategy, player development, lineup decisions, and transaction decisions**. By blending traditional hockey knowledge (who are the great shooters) with data-driven rigor, it opens the door for **smarter decision-making**. The real-world impact is that coaches, managers, and analysts get closer to understanding the true drivers of goals: was it a great chance, great shot, or both? Armed with that knowledge, they can better allocate ice time, salary cap, and training focus to maximize their team's goal production.

Future Research Directions

The encouraging results of this shooter-talented xG model point to several **future research directions and extensions**:

- **Incorporate Goaltender Effects:** A logical next step is to build a similar talent metric for goaltenders. Just as shooters have varying finishing ability, goalies have varying saving ability. A combined model could include a random effect for the goalie facing the shot, creating an even richer prediction (shot success = f(shot quality, shooter talent, goalie talent)). This would allow us to quantify how much of a scoring chance's outcome was due to the shooter vs. the goalie, and could lead to a truly comprehensive expected goal model that accounts for both sides of the coin.

- **Use Dynamic and Contextual Features:** While our model considered static features like distance and angle, future models can leverage **richer contextual data**. With the advent of player- and puck-tracking data, one could incorporate the speed and trajectory of passes before the shot, shooter motion, defensive positioning, and other factors that influence shooting success. For example, identifying one-timers, cross-ice passes, rebounds, or screens explicitly would likely improve the model's accuracy. These factors could be added as new input variables in the logistic regression or handled via separate sub-models (e.g. a "pass quality" model feeding into xG). Additionally, making the shooter talent **time-dependent** (using a state-space model or tracking rolling values) could capture hot streaks or developmental improvements. This would help address the limitation of assuming a constant talent level.

- **Stratified Models (Situational xG):** Another extension is to build separate models for different game situations or subsets of shots. The scoring dynamics differ between even-strength, power play, shorthanded, and other situations. Similarly, forwards and defensemen have inherently different shooting profiles. Future work could develop specialized xG models for these contexts or include interaction terms (e.g. shooter talent might manifest differently on the power play vs. at even strength). A multi-level model could even allow each player's talent to vary by situation (perhaps a player is a power-play specialist in finishing, but not as much at events).

- **Advanced Modeling Techniques:** While logistic regression provides interpretability, more complex machine learning techniques might capture non-linear relationships and subtle interactions. Exploring **ensemble methods or neural networks** could improve predictive performance. For instance, gradient boosted trees (as used by some public models) or deep learning could automatically detect complex patterns (maybe certain combinations of angle and distance that yield higher goals, or interactions between shooter and shot type). These methods, however, are typically "black boxes." A promising research avenue is to combine the best of both worlds: use machine learning to boost accuracy **while retaining a notion of shooter talent**. One idea is two-stage modeling – first estimate shooter talent via a simpler model, then feed those estimates as features into a more complex model. Alternatively, techniques like **embedding layers in neural nets** could learn a vector representation for each shooter that captures their talent. Any such approach should be compared against the simpler, interpretable model to ensure we genuinely gain accuracy without losing too much clarity.

- **Validation and Calibration:** Future research should also focus on rigorous validation. We would want to test how well the shooter talent model predicts *future* goals compared to a non-talent model. One could perform season-over-season tests to see if a player's predicted goals (using their prior talent estimate) align with their actual performance. This would solidify the model's utility for forecasting. Calibration is equally important – ensuring that an xG of, say, 0.2 truly corresponds to a 20% chance of goal across different players. A well-calibrated model with shooter talent would mean that, over a large sample, players of all talent levels have their goals tally closely match their xG. Achieving that indicates the model is properly accounting for talent.

- **Interpreting Shooter Talent in Depth:** Another future avenue is to delve deeper into what the **shooter talent metric** represents. Is it mostly release accuracy, shot power, pre-shot deception, or something else? By correlating the talent estimates with other data (like shooting accuracy tests, puck velocity, or even surveying coaches' opinions), we might better understand the nature of "finishing ability." This can bridge the gap between the statistical model and the **hockey-specific skills** it is capturing. It could also identify if certain types of players are undervalued – for instance, players who consistently beat expected goals might share certain traits (quick release, great hockey IQ) that scouts can look for in younger players.

- **Application to Strategy and Team Composition:** From a more strategic perspective, future work can use this model in simulation environments. One could simulate games or seasons where teams alter their shot selection or lineup based on shooter talent information. For example, does a team benefit more by funneling shots to its best finishers even if it means fewer shots overall? Such simulations, backed by the model, can explore "what-if" scenarios and potentially guide optimal play styles (a high-talent, low-volume strategy vs. a low-talent, high-volume strategy). Additionally, integrating the model into real-time analytics (e.g., live game "expected goals with shooter talent" tracking) could help commentators or analysts identify in-game when a team is getting "better chances than the scoreboard shows" given who is shooting.

In conclusion, the introduction of a **Shooter's Talent** factor in an xG model is a meaningful step forward for hockey analytics. It acknowledges a fundamental aspect of the game, finishing skill, and provides a quantitative way to include it in analyses. This conclusion has summarized how our model was built and what we learned: by accounting for shooter skill, we improved the model's accuracy, made its outputs more insightful for evaluating players and teams, and maintained interpretability. There are clear strengths in this approach, along with limitations that present opportunities for refinement. **Hockey, like all sports, rewards both opportunity and execution**. Expected goals have traditionally measured the opportunity; this work brings execution into the fold. The hope is that such models continue to evolve, incorporating more facets of the game, and remain **grounded in solid statistical reasoning**. As a math PhD student working on an applied problem, it's gratifying to see theory meet practice – the logistic regression framework enriched with a Bayesian twist not only advances the analytical toolbox but also offers practical value to how we understand hockey. Going forward, continued collaboration between statisticians and hockey analysts will ensure that models like these become even more robust, actionable, and attuned to the heartbeat of the game. The inclusion of shooter talent has proven effective; building on this foundation, we can strive for even more **realistic and impactful hockey analytics models** in the future.

Appendix:

1. Link to the xG Model Code

   https://github.com/JefferyMei19/xG-Model

2. Link to the xG Model (Shooter's Talent) Code

   https://github.com/JefferyMei19/xG-Model-Shooter-Talent

Variable Analysis: application of ridge and lasso logistic regression

To compute the Logistic Regression Model on Python, I used the scikit-learn package, the default code has a penalty term l2 which stands for ridge regression. We choose this model because in general, ridge regression will perform better when the response is a function of many predictors, all with coefficients of roughly equal size.

```
from sklearn.linear_model import LogisticRegression

log_reg = LogisticRegression(
    penalty='l2',         # Use L2 regularization (ridge)
    max_iter=6000,        # Maximum number of iterations for convergence
    fit_intercept=False   # Do not calculate an intercept term
).fit(X_train, y_train)
```

Lasso regression performs variable selection. We will use lasso regression to evaluate the significance of our variables in the model.

Intro to Ridge and Lasso Regressions

Modification of Multi-Linear Regression

Recall Multi-Linear Regression

Let the predictors be $\vec{x}_1, \dots, \vec{x}_n \in \mathbb{R}^p$ and the responses $y_1, \dots, y_n \in \mathbb{R}$. The *multi-linear regression (MLR)* model is

$$\vec{y} = [\,\mathbf{1} \mid X\,]\vec{\beta} + \vec{\varepsilon}, \qquad \vec{\varepsilon} \sim \mathcal{N}(\vec{0}, \sigma^2 I_n),$$

with design matrix $X = [\vec{x}_1^{\mathsf{T}}; \dots; \vec{x}_n^{\mathsf{T}}]$ and $\vec{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^{\mathsf{T}}$.

Least-squares estimator.

$$\widehat{\vec{\beta}}_{\text{OLS}} = \arg\min_{\vec{\beta}} \sum_{i=1}^{n} (y_i - \beta_0 - \vec{\beta}_{1:p}^{\mathsf{T}} \vec{x}_i)^2 = (X^{\mathsf{T}} X)^{-1} X^{\mathsf{T}} \vec{y}.$$

Bias–variance decomposition (test error).

$$\mathbb{E}\Big[\text{TestErr}\big(\hat{f}(\vec{x})\big)\Big] = \sigma^2 + (\text{Bias})^2 + \underbrace{\text{Var}}_{\text{circled in notes}}$$

- In **MLR** the *bias is low*, therefore test error is dominated by the variance term.
- A slight increase in bias (via regularisation) can substantially *reduce variance* and improve predictive accuracy.

Remark. *MLR automatically assigns a coefficient to every predictor. For large p we often prefer a smaller subset of* important variables.

Ridge and Lasso Regressions

Both techniques modify MLR by adding a *penalty term* to keep coefficients from becoming too large.

Important! $\boxed{\beta_0 \textbf{ (the intercept) is } \textbf{\textit{not}} \textbf{ penalised.}}$

(a) Ridge Regression — ℓ^2 regularisation

$$\boxed{\vec{\beta}_{\text{ridge}} = \arg\min_{\vec{\beta}} \Big\{ \underbrace{\sum_{i=1}^{n} \big(y_i - \beta_0 - \vec{\beta}_{1:p}^{\mathsf{T}} \vec{x}_i\big)^2}_{\text{data fit}} + \lambda \underbrace{\sum_{j=1}^{p} \beta_j^2}_{\ell^2 \text{ penalty}} \Big\}}$$

(b) Lasso Regression — ℓ^1 regularisation

$$\boxed{\vec{\beta}_{\text{lasso}} = \arg\min_{\vec{\beta}} \Big\{ \sum_{i=1}^{n} \big(y_i - \beta_0 - \vec{\beta}_{1:p}^{\mathsf{T}} \vec{x}_i\big)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \Big\}}$$

(Lasso = *least absolute shrinkage and selection operator.*)

- $\lambda \geq 0$ is a **tuning parameter**.
  - $\lambda = 0 \Rightarrow$ *multi-linear model.*
  - $\lambda \to \infty \Rightarrow \beta_j \to 0$ (null model).
- Intermediate λ balances fitting y versus shrinking coefficients.

Ridge and Lasso from a Bayesian Viewpoint

Consider the standard linear model with i.i.d. Gaussian noise $Y_i = \beta_0 + \vec{\beta}_{1:p}^{\mathsf{T}} \vec{x}_i + \varepsilon_i, \ \varepsilon_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$.

Step 1. Likelihood

Because the ε_i are Gaussian and independent,

$$\underbrace{\text{Lik}(\vec{y} \mid \vec{\beta}, X)}_{p(\vec{y} \mid \vec{\beta}, X)} = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big[-\tfrac{1}{2\sigma^2} \big(y_i - \beta_0 - \vec{\beta}_{1:p}^{\mathsf{T}} \vec{x}_i\big)^2\Big].$$

Taking logs (and dropping the constant $-\frac{n}{2} \log(2\pi\sigma^2)$),

$$\log \text{Lik}(\vec{y} \mid \vec{\beta}, X) = -\frac{1}{2\sigma^2} \sum_{i=1}^{n} \big(y_i - \beta_0 - \vec{\beta}_{1:p}^{\mathsf{T}} \vec{x}_i\big)^2.$$

Step 2. Choose a prior for each β_j

(a) Gaussian prior (ridge). $\pi_{\text{ridge}}(\beta_j) = \mathcal{N}(0, \tau^2) \implies \log \pi_{\text{ridge}}(\vec{\beta}) = -\frac{1}{2\tau^2} \sum_{j=1}^{p} \beta_j^2 + C.$

(b) Laplace (double-exponential) prior (lasso). $\pi_{\text{lasso}}(\beta_j) = \frac{\lambda}{2} e^{-\lambda|\beta_j|} \implies \log \pi_{\text{lasso}}(\vec{\beta}) = -\lambda \sum_{j=1}^{p} |\beta_j| + C.$

(Note: the intercept β_0 is *not* penalised, so we give it a *flat* prior.)

Step 3. Posterior (by Bayes' theorem)

Up to an additive constant,

$$\log \pi(\vec{\beta} \mid \vec{y}, X) = \log \text{Lik}(\vec{y} \mid \vec{\beta}, X) + \log \pi(\vec{\beta}).$$

Insert the likelihood from Step 1 and the appropriate prior:

(a) Ridge (Gaussian prior):

$$\log \pi_{\text{ridge}}(\vec{\beta} \mid \vec{y}, X) = -\frac{1}{2\sigma^2} \sum_{i=1}^{n} \left(y_i - \beta_0 - \vec{\beta}_{1:p}^{\mathsf{T}} \vec{x}_i\right)^2 - \frac{1}{2\tau^2} \sum_{j=1}^{p} \beta_j^2 + C.$$

(b) Lasso (Laplace prior):

$$\log \pi_{\text{lasso}}(\vec{\beta} \mid \vec{y}, X) = -\frac{1}{2\sigma^2} \sum_{i=1}^{n} \left(y_i - \beta_0 - \vec{\beta}_{1:p}^{\mathsf{T}} \vec{x}_i\right)^2 - \lambda \sum_{j=1}^{p} |\beta_j| + C.$$

Step 4. MAP estimator = maximise log-posterior

Maximising the log-posterior is equivalent to *minimising* its negative:

(a) Ridge MAP $\implies$ ℓ^2-penalised least squares

$$\vec{\beta}_{\text{MAP}} = \arg\min_{\vec{\beta}} \left\{ \sum_{i=1}^{n} \left(y_i - \beta_0 - \vec{\beta}_{1:p}^{\mathsf{T}} \vec{x}_i\right)^2 \; + \; \lambda \sum_{j=1}^{p} \beta_j^2 \right\}, \qquad \lambda = \sigma^2/\tau^2.$$

(b) Lasso MAP $\implies$ ℓ^1-penalised least squares

$$\vec{\beta}_{\text{MAP}} = \arg\min_{\vec{\beta}} \left\{ \sum_{i=1}^{n} \left(y_i - \beta_0 - \vec{\beta}_{1:p}^{\mathsf{T}} \vec{x}_i\right)^2 \; + \; \lambda \sum_{j=1}^{p} |\beta_j| \right\}.$$

Thus the **MAP estimator under a Gaussian prior** is exactly *ridge regression*, while the **MAP estimator under a Laplace prior** is the *lasso*. The penalty parameter λ is linked to the prior's spread: a more concentrated prior (smaller τ or larger Laplace rate) imposes stronger shrinkage.

Remark. *If we used the* posterior mean *instead of the mode, we would recover* Bayesian ridge regression *(closed-form) and* Bayesian lasso *(requires MCMC), but the optimisation problems above remain the core intuition.*

Lasso—When to Use Which?

- **Lasso** excels when only a *small subset* of predictors has large effects; it sets the rest exactly to zero, yielding a simpler and more interpretable model.
- **Ridge** performs better when the response depends on *many* predictors, all with coefficients of comparable size.

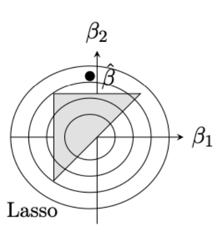
In the next section, we will go over the theory behind how lasso regression performs variable selection.

Lasso Regression and Predictors Selection

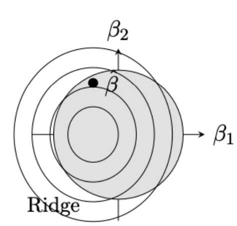
- **Subset selection** (choose the best k predictors) is NP-hard. Exhaustive search is impossible when p is large; stepwise is unstable.
- **Ridge regression** is convex and stable but *keeps every predictor*—great for prediction, weak for interpretability.
- **Lasso** (Tibshirani 1996) replaces the ℓ^2 penalty by an ℓ^1 penalty, combining convexity with *automatic sparsity.*
- Geometry: the ℓ^1 ball has corners at the axes, so the optimum often lies on an axis $\Rightarrow$ some $\beta_j = 0$.
- Computation: convex $\Rightarrow$ polynomial-time algorithms (LARS, coordinate descent, saga); sparsity drops out "for free."

---

1. Idea in Pictures



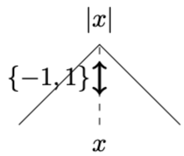
$|\beta_1| + |\beta_2| = c$



$\beta_1^2 + \beta_2^2 = c$

2. Subdifferential Mechanics

2.1 Subdifferential warm-up



For a convex f, the subdifferential at x is $\partial f(x) = \{\text{all supporting slopes at } x\}$. At smooth points it equals $\{\nabla f(x)\}$.

Take-away.

$$\mathbf{0} \in \partial f(\bar{\boldsymbol{x}}) \iff \bar{\boldsymbol{x}} \textit{ minimises } f.$$

2.2 Lasso objective

$$f(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left(y_i - \boldsymbol{\beta}^\mathsf{T} \boldsymbol{x}_i\right)^2 + \lambda \sum_{j=1}^{p} |\beta_j|, \qquad \lambda > 0.$$

$$\partial_{\beta_0} f = -2[X]^\mathsf{T} \boldsymbol{y} + 2[X]^\mathsf{T}[X]\boldsymbol{\beta}, \quad \partial_{\beta_j} f = g_j + \lambda\, w,\ w \in [-1,1], \quad g_j = \left[-2[X]^\mathsf{T}\boldsymbol{y} + 2[X]^\mathsf{T}[X]\boldsymbol{\beta}\right]_j.$$

2.3 Solve for β_0 first

Set $\boldsymbol{\beta}_{1:p} = 0$. Then $\beta_0 = \bar{y}$, independent of λ.

2.4 Choose λ to dominate g_j

$$\lambda > \max_j |g_j| \implies 0 \in g_j + [-\lambda, \lambda] \quad \forall j.$$

Take-away.

$$\hat{\boldsymbol{\beta}} = \left(\beta_0,\, 0, \ldots, 0\right)^\mathsf{T}$$

minimises f; all slopes are exactly zero *when λ is large enough.*

Ridge Logistic Regression:

$$E = -\sum_{i=1}^{n} [y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)] + \lambda \sum_{j=1}^{m} \beta_j^2$$

Gradient update rule:

$$\frac{\partial E}{\partial \beta_j} = \sum_{i=1}^{n} (\hat{y}_i - y_i) x_{ij} + 2\lambda\beta_j$$

Coefficient update:

$$\beta_j = \beta_j - \alpha \left(\sum_{i=1}^{n} (\hat{y}_i - y_i) x_{ij} + 2\lambda\beta_j \right)$$

Lasso Logistic Regression:

$$E = -\sum_{i=1}^{n} [y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)] + \lambda \sum_{j=1}^{m} |\beta_j|$$

Gradient update rule (subgradient):

$$\frac{\partial E}{\partial \beta_j} = \sum_{i=1}^{n} (\hat{y}_i - y_i) x_{ij} + \lambda \operatorname{sign}(\beta_j)$$

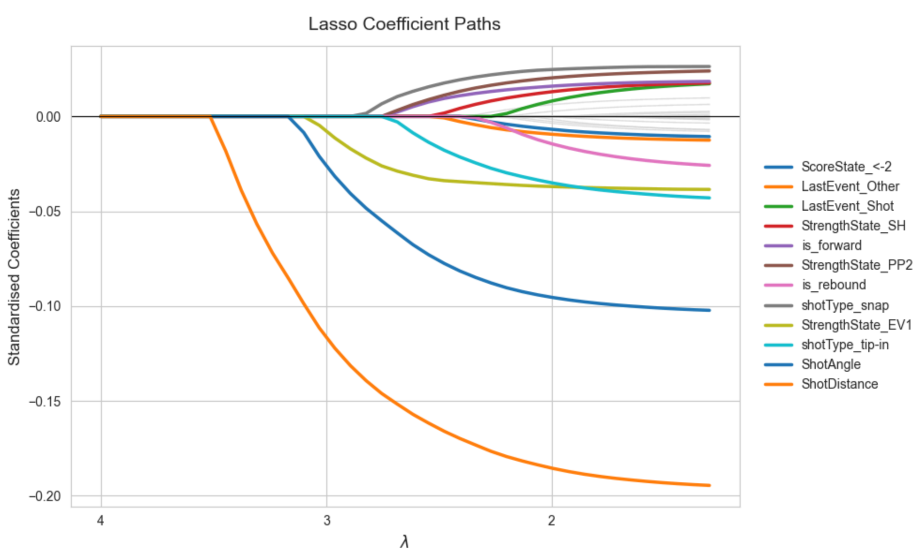
Coefficient update:

$$\beta_j = \beta_j - \alpha \left(\sum_{i=1}^{n} (\hat{y}_i - y_i) x_{ij} + \lambda \operatorname{sign}(\beta_j) \right)$$

Model 1: (without shooter's talent)

The following variables turns out to be 0:

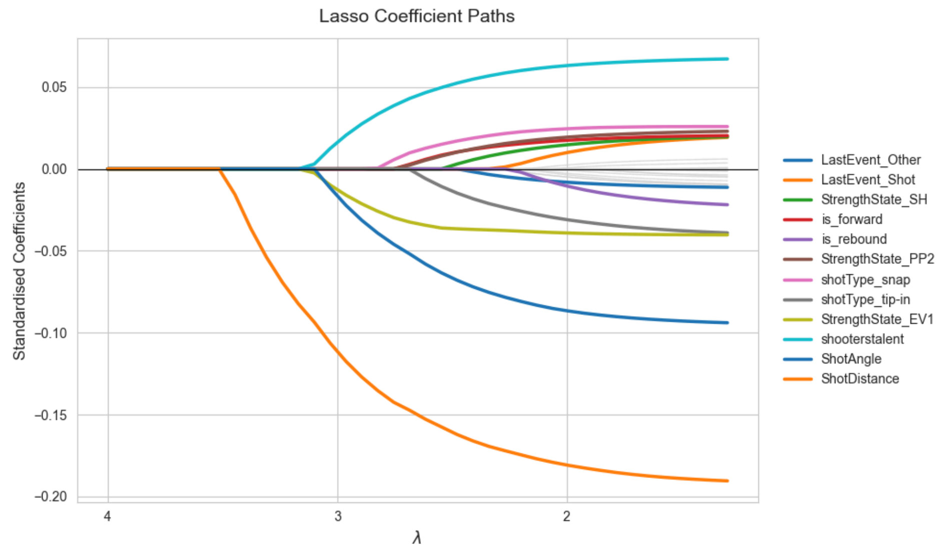
1. ScoreState_2 (when the team lead by 2 or more goals)
2. shotType_cradle
3. shotType_poke



Model 2 (shooter's talent)

The following variables turns out to be 0:

1. shotType_cradle



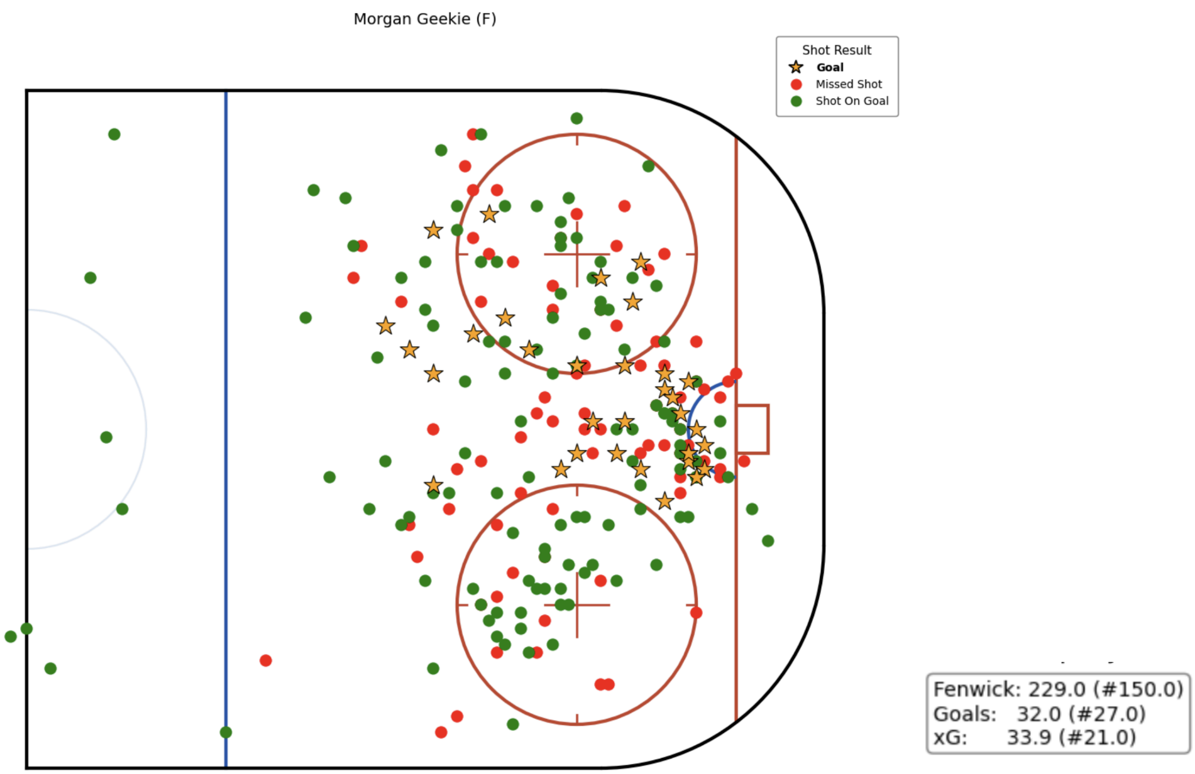
The two coefficient-path graphs act like X-rays of our lasso fits, showing—across the full range of penalty strengths—exactly which predictors fight their way off the zero line and how quickly they do so. In the baseline model that omits shooter-talent, only shot distance and the shot angle break free as soon as the penalty loosens, making them the dominant drivers of scoring probability; nearly all other covariates creep upward only modestly, and three (`ScoreState_ 2`, `shotType_cradle`, `shotType_poke`) never leave the horizontal axis, proving they add no incremental information once shot distance and shot angle are in the equation. When we introduce an explicit talent measure, the picture shifts: the talent curve now springs off the axis first, some game-state effects like "team trailing badly" regain modest weight, and several shot-type curves flatten, indicating that player skill soaks up variation previously attributed to game context.

The flat segments at the left edge of each curve are the visual fingerprint of the lasso's Karush–Kuhn–Tucker rules: while the penalty is larger than a predictor's individual threshold, that coefficient is mathematically forced to zero; once the threshold is crossed, the curve peels away and its slope is governed by the Fisher-information matrix. If we mark the cross-validated penalty (for example the one-standard-error rule) we can read the "effective model size" directly from the stack of active curves: roughly seven variables in the talent-free fit and eight in the talent-augmented fit. The big-picture lesson is that shot distance, shot angle, and shooter skill form the core explanatory set, whereas cradle and poke attempts—and certain high-lead game states—are statistically inert.

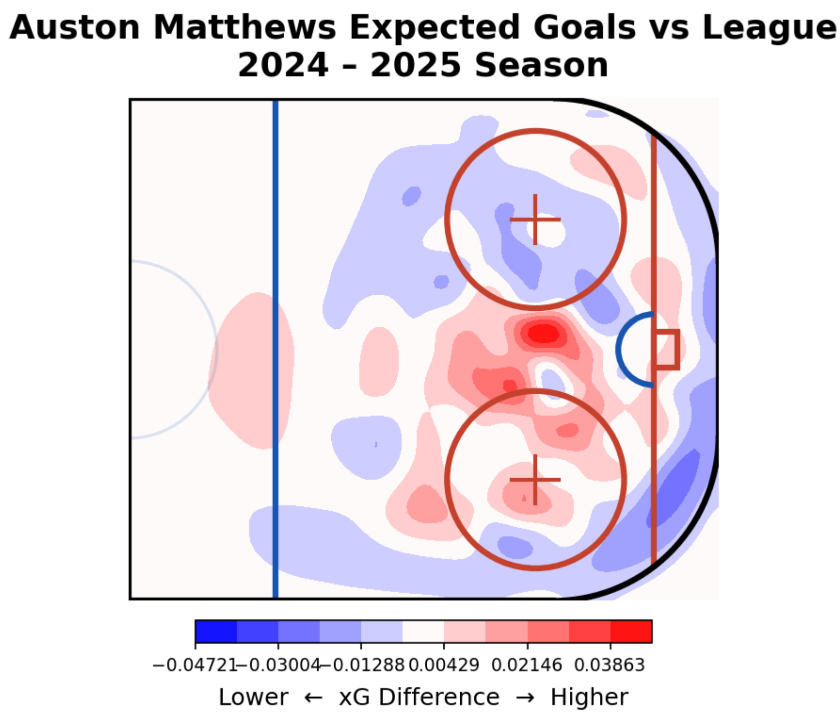
Moreover, the lasso regression analysis demonstrates that our newly introduced "Shooter's Talent" variable ranks as the second most influential predictor in the second model, underscoring its substantial explanatory power and validating its significance.

Model Usage

1. Player Card



2. Player Shot Map

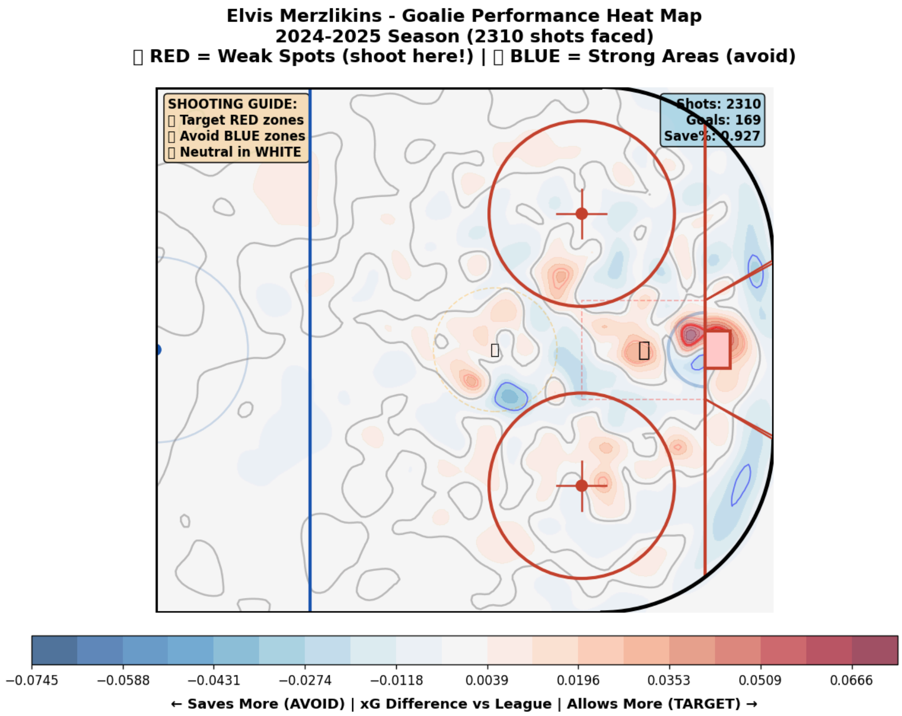


Expected-goals (xG) maps visualize where a player generates shot quality relative to league norms. In a differential map like this, red indicates above-average xG and blue indicating below-average. Coaches can see not just how often a player shoots, but *how dangerous* those shots are from specific locations.

For the player's own team, this information supports both tactical design and roster optimization. Coaches can tailor offensive schemes to feed the player the puck in their highest-value zones. For example, in areas of concentrated red around the slot or faceoff dots. Power-play structures, set plays off faceoffs, and zone entries can all be aligned to funnel the puck into those regions. At the same time, analysts can evaluate whether changes in the player's usage (e.g., line combinations or handedness on the power play) actually increase xG from preferred locations over time, making the map a tool for evidence-based experimentation and player development.

For opposing teams, the same map functions as a pre-scouting and risk-management instrument. Defenders can be instructed to deny passes into the player's most dangerous zones, over-load coverage when the puck approaches those areas, or force the player to shoot from blue regions where their threat is below league average. Matchups can also be planned more precisely: coaches may assign particular defenders who are strong at disrupting plays in those high-xG regions, or adjust penalty-kill formations to remove the player's primary shooting lane.

3. Goalie Shot Map and Analysis



A goalie performance heat map, which displays where a netminder concedes more or fewer goals than expected, is a valuable analytic tool for both the goalie's own team and for opponents. For the goalie's team, regions of blue (over performance) and red (under performance) provide an evidence-based diagnosis of technical and tactical strengths and weaknesses: coaches can tailor practice to address systematic vulnerabilities, such as low blocker-side shots or lateral plays across the slot, while reinforcing positioning patterns that consistently yield above-expected save rates. The map also informs defensive structure, indicating where skaters should prioritize shot suppression or lane denial to shield their goaltender from areas in which he is relatively weak.

For opposing teams, the same visualization functions as a shooting guide, encouraging players to adjust their shot selection and set plays to exploit red zones and to avoid low value blue areas, thereby converting abstract expected goals information into concrete tactical decisions in the offensive zone.