STARS LINE OPTIMIZATION

An statistical report on the optimization of the 22-23 Dallas Stars Lines

Zach McCombs, Jaelee McCleland, Chuck Gunderson

OBA 465 – Sports Analytics, Group Hockey

Executive Summary

This reports goal is to identify player pairings that maximize offensive and defensive efficiency to inform optimal line deployment strategies for future games and roster decisions.

Approach:
Using data from the 2022–2023 NHL season, combined summary and play-by-play datasets to evaluate player combinations at even strength (5v5). This report applied advanced statistical measures including Point Shares (PS), shooting efficiency, goal/shot likelihood models, and With-or-Without-You (WOWY) differentials. Analyses were conducted using both R and Python for calcualtions, table manipulations and visualizations to inform decision-making.

Key Findings:
<u>Top Forward Line:</u>
The line of Robertson–Pavelski–Hintz was the most offensively effective unit. This trio led in Offensive Point Shares (OPS), shooting efficiency, and posted exceptional z-scores. Each pair within the trio recorded positive deltas in GF and SF per 60 minutes, with Robertson–Hintz reaching a goal likelihood of 22.4% in a standard 250-second window.
<u>Underutilized Duos:</u>
Duos like Klingberg–Hintz and Benn–Dellandrea showed high goal conversion rates despite minimal usage, suggesting hidden offensive value that can be better leveraged through targeted deployment.
<u>Top Defensive Pairings:</u>
Heiskanen–Miller and Lindell–Hakanpää ranked highest in Defensive Point Shares (DPS) and PS, with statistically significant z-scores. Guiarnov-Robertson and Sekera-Benn standout with the lowest likihoods of letting up a goal per game.
<u>Optimal Starting Unit Recommendation:</u>
Based on combined offensive and defensive metrics, the optimal starting five consists of Robertson, Pavelski, Hintz, Heiskanen, and Miller.

Implications for Leadership:
- Deploy data-backed starting units for improved lineup efficiency.
- Reassess ice time allocation to underused but high-impact duos.
- Use boost metrics (OB/DB) to identify players who elevate teammates' performance.
- Continue integrating WOWY and likelihood analysis into pre-game matchups and scouting reports.

Next Steps:
To expand this analysis, data recommendations include:
- Incorporating opponent context and score-state data to evaluate clutch performance.
- Segmenting special teams data (PP/PK) to optimize non-even-strength scenarios.
- Including spatial and possession metrics for a more nuanced view of zone impact.

Description and Discussion of the Development of the Analytics Question

"*What pairings would be the most offensively/defensively effective for the Dallas Stars*?"

This report seeks to answer whether there's a way with a variety of calculations and comparisons if Dallas can implement optimized lines for offense and defense.

Discussion of Variables Available and Variables Needed to Examine the Analytics Question

This project evaluates offensive efficiency for Dallas Stars player duos using play-by-play and season summary data from the 2022–2023 NHL season. The fastRhockey dataset provides several core variables that support this analysis. After filtering for all Stars' games and these variables are as follows:

Variable	Description
game_id	Provides a unique code for each game in the NHL season
period	Identifies the period in which each event occurred
strength_state	Identifies the strength of the event team (e.g., 5v5, 4v5, 6v5, etc)
event_type	Categorizes all events into one of 17 types (e.g., "SHOT", "GOAL", etc)
DAL_loc	Identifies the location of the Dallas Stars for each unique game_id
event_team	Indicates whether the event is attributed to the home or away team
description	Provides a brief summary of the event and involved players
skater1–skater6	Names of Dallas Stars skaters on the ice during the event

This play-by-play data allows us to derive shooting efficiency (goals per shot, conduct a "With or Without", likelihood Calculations for goals for and against.

Evaluation of offensive efficiency between Dallas Stars player duos requires both observed and derived variables. These components from the play by play enable the construction of base statistics for each player pair. From these values, normalized metrics such as goals for per 60 minutes (GF per-60), goals against per 60 (GA per-60), and shots per 60 (SF per-60) are calculated to account for variation in deployment time. A standardized expected goals metric (λ) is also determined based on a fixed 250-second shift length, which corresponds to the average ice time observed across all pairs per game. Comparative metrics delta goals for per 60 (ΔGF per-60), delta goals against per 60 (ΔGA per-60), and delta shots per 60 (ΔSF per-60), measure performance changes when players are paired with specific teammates versus their performance with all other combinations, used in a "With or Without You" Analysis (WOWY).

Using the season summary data this report identified a potential new variable Point Shares. Point shares (PS) is an advanced statistic for hockey that evaluates an individual skater's impact on their team's performance. It can be calculated using the formula from Hockey Reference (Kubatko). It combines offensive point shares (OPS) and defensive point shares (DPS). The OPS value represents a skater's offensive impact, the DPS value represents a skater's defensive impact, and the PS value represents the overall impact the skater has on the team's performance. The higher each of those values are, the more impact the skater has. Most of the required variables of PS were source able from the season summary data, to fill the gaps this report employed the play-by-play data for calculations.

Construction of New Variables

PS uses a wide range of common variables such as goals, assists, team goals, and more. MoneyPuck tracks most of these variables ("Download Player and Team Data"). Leaving only team totals, goals created, assists, position data, and league totals to be sourced or calculated via Play by Play.

Goals created is a variable needed for PS and needs to be calculated for each line/pair and their team. Hockey Reference provided this formula ("NHL Stats Glossary").
Team data is needed for team goals created and for the PS calculation; data such as team goals for and against, team time on ice, team games, team assists, and team shots against per minute; Hockey Reference also has this data ("2022-23 Dallas Stars"). Team data was used to isolate defensive players and offensive players in order to calculate ice time and plus/minus for each position.
Hockey Reference provided league data that's needed for PS, such as league points per goal, league goals per game, league points, and league shots against per minute ("2022-23 NHL Summary").
Two final variables need to be calculated, plus/minus and assists for each line/pair. Hockey Reference details how to calculate plus/minus ("NHL Stats Glossary"), which can be done for the line/pair with the Money Puck data. A Python formula is created to calculate it for every line/pair, and another Python formula is created to calculate assists.
To get play-by-play data for PS, fastRhockey was used to create a CSV file that could be worked on in Python. Play-by-play data can be used to calculate assists; when there is a goal, an assist is credited to a line/pair if the whole line/pair is on the ice and one of the skaters in the line/pair gets an assist. The "zoomed" out calculations for Point Share and for attributive calculation see Figure 1:

$$OPS = \frac{Marg. GF\ Skater}{Marg. Goals\ per\ point}$$

$$DPS = \frac{Marg. GA\ Skater}{Marg. Goals\ per\ point}$$

$$Point\ Shares = OPS + DPS$$

From the delta per-60 calculations 2 new variables can be constructed, Offensive Boost (OB) and Defensive Boost (DB). OB is the change in offensive output (in terms of scoring rate) for other players when paired with the given player. The same logic is applied to DB instead the change in defensive output, the scoring rate against. The calculation is the average of the delta goals for/against per 60 minutes for an individual player.

$$OB_i = \frac{1}{n}\sum_{j=1}^{n}\left(\Delta GF/60_{i,j}\right)$$

Where $\Delta GF/60_{i,j}$ is the "with or without you" differential in goals for per 60 minutes for player *i* when paired with player *j*, and *n* is the total number of unique teammates *j* player *i* played with. The calculation is the same for both OB and DB, substituting goals for in OB with goals against in DB. Using this to identify individual skaters that boost other players' performance that we can then match with our other data to identify pairs that possess pair synergy as well as an overall boost to performance to all teammates.

Flowchart of the Analytics Logic to Answer the Analytics Question
See Appendix A, Figure 2 for Flowchart visual. The general logic was as follows: Identify our question, from there identify data needs to solve that question. In our case it was online sourced season summary data and play by play data. After cleaning and revisions both were prepared for analysis. From summary data was the logical progression of calculations for Point Shares to evaluate lines and pairs. During the course of PS calculations, it required assist numbers which weren't on the season summary data and thus were calculated from the play-by-play data. From the play-play data logic for efficiency, delta, likelihood, and boost calculations were implemented via R. Using these calculations this report is able to evaluate and pinpoint Line Optimizations for the Dallas Stars under varying conditions, offensive and defense.

Rationale for Utilizing the Identified Analytics Tools

Python will be used to calculate PS in this report due to prior experience with Python. Python lets users filter, sort, display, explore, and manipulate data faster, which is beneficial when you need to sort through and work with large amounts of data, like the play-by-play data needed to find assists. Python also tends to take less time to execute lines of code for different functions (BasuMallick, 2022).

R was the source for play-by-play data from the package fastRhockey. This data enabled us to conduct our analysis and calculations. Due to the data being sourced from R all calculations, table manipulations, and visual aids that revolved around the play-by-play data were conducted in R for continuity.

Description of Conversion from Analytics Logic to Analytics Tool(s)

The development logic structures from conceptual logic of identifying ideal pairs and lines is as follows.

For the duo efficiency analysis in R, logic was implemented to:

- Isolate GOAL and SHOT events where both players in a pair were present and calculate shot totals, goal counts, and efficiency (goals per shot).
- Classification logic was also added to:
  - Compare each pair’s efficiency against the team average.
  - Assign color-coded recommendations (“Keep Deploying,” “Neutral,” “Drop”) based on performance bands.
  - Visualize results through annotated ggplot2 bar charts and summarize findings in publication-ready tables using the get package.

For the Pair Event Likelihood Model (GF/GA Likelihood per 250s) the logic that was implemented into R is as follow:

- Calculating the per-60 rates for both GF and GA
- Per-60 rates for GF and GA are scaled to a 250-second time window:
  - o lambda = rate_per_60 * (250 / 3600)
- Poisson probability of at least one goal occurring in that window is calculated:
  - o P = 1 - exp(-lambda), providing prob_goal_for_250 and prob_goal_against_250
- A heatmap matrix is built, after sorting and alignment, with as.matrix() and melt() from reshape2.
- Conditional filtering ensures that only pairs with ≥1250 seconds of TOI are included.

For the WOWY Delta Analysis Logic (With or Without You Differentials) the logic that was implemented into R is as follows:

- "Without" metrics are derived by subtracting each pair’s contribution from the player’s total:
  - o (total_goals_for - goals_for) / (total_ice_time - ice_time) yields the rate for Player A without Player B.
- Deltas are then calculated by subtracting these from the original pairwise rate:
  - o delta_goals_for_per_60 = goals_for_per_60 - goals_for_without, and likewise for others.
- The top 10 offensive-impact and defensive-impact pairs are visualized with a horizontal bar chart
- Player-specific averages across all pairings are calculated and sorted for:
  - o Offensive boost (avg delta_GF/60)
  - o Defensive boost (avg delta_GA/60)

Point Share (PS) calculations were executed in Google Colab using Python. After importing data from MoneyPuck and Hockey Reference:

- Functions were created to calculate required inputs for PS, such as goals, assists, and goals created.

- Additional functions computed OPS and DPS separately, before summing them to calculate each line's total PS value.
- A z-test was performed across all PS values to identify lines or pairs that outperformed the mean by more than one standard deviation.
- The resulting metrics, including z-scores, were visualized to identify the most impactful forward lines and defensive pairs.

While R logic emphasized event-level filtering and matchup-based efficiency scoring, Python logic focused on stat aggregation and standardized scoring of player impact. Together, these conversions enabled consistent, replicable analytics transforming raw NHL data into plug and play logic to answer a variety of line-based questions.

Primary Findings

Using shooting efficiency, shot volume, overall point shares (OPS, DPS, and PS), Pair Event Likelihoods (GF and GA in 250s), WOWY(Delta GF, GA, and SF per 60), and Boost (OB and DB) this report evaluates forward lines and defensive pairs that will maximize Stars' winning potential.

<u>Top Offensive Line Performance:</u>

Using OPS and OB to identify strong lines then comparing identified efficiencies, GF likelihood, and delta GF per-60 and SF per-60. Additionally noting any z-scores that identify a pair as statically outstanding.

The trio of Jason Robertson, Joe Pavelski, and Roope Hintz demonstrated the highest level of offensive productivity among all forward combinations analyzed. This line recorded an OPS of 16.0[1] and a DPS of 1.6,[2] resulting in a PS of 17.6[3]. A z-score of 4.3 [4]indicated statistically exceptional performance when compared to other skater groups. Within the pairwise analysis, both the Robertson–Pavelski and Robertson–Hintz duos exceeded the team average shooting efficiency of 0.100, registering values of 0.114 and 0.117[5], respectively. Goal-for likelihoods for these duos also surpassed the team average of 5.4%, reaching 18.9% and 22.4%, respectively[6].

Results from the WOWY analysis supported the synergy among these three skaters. Each pair within the trio posted positive differentials in ΔGF/60 and ΔSF/60, indicating strong mutual influence on offensive outcomes. The Robertson–Pavelski pairing produced a ΔGF/60 of 2.8 and a ΔSF/60 of 38.8, while the Robertson–Hintz pairing achieved a ΔGF/60 of 2.5 and a ΔSF/60 of 41.2. The Pavelski–Hintz duo recorded a ΔGF/60 of 2.6 and a ΔSF/60 of 40.[7]1. These consistently high values across all metrics identified this trio as the most effective offensive unit observed in the dataset. All 3 players were in the top 10 offensive boost players.[8]

---

[1] See Figure 3 for OPS Table
[2] See Figure 4 for DPS Table
[3] See Figure 5 for PS Table
[4] See Figure 6 for Z-Score Table
[5] See Figure 7 for Efficiency table
[6] See Figure 8 GF likelihood table
[7] See Figures 10 and 11 for WOWY Goals for and Shots for Table
[8] See Figure 13 for OB table

Underused Offensive Duos:

Several duos demonstrated strong goal-to-shot conversion rates despite lower deployment, less than average ice time of all pairs 8502 seconds. Notably, John Klingberg-Roope Hintz and Jamie Benn-Ty Dellandrea exceeded the team's average shooting efficiency of 0.133, 0.165 and 1.000 respectively[9]. Highlighting an opportunity for coaching staff to leverage efficient but underused pairings.

Effective Defensive Pairings:

Defensive efficiency for player pairs was evaluated using DPS[10], PS[11], and standardized z-scores[12]. The pairing of Miro Heiskanen and Colin Miller recorded the highest values, with a DPS of 2.2, a PS of 17.3, and a z-score of 4.23. Esa Lindell and Jani Hakanpää followed closely, posting a DPS of 1.9, a PS of 15.6, and a z-score of 3.78. These metrics reflect the strong defensive impact of both duos and support their frequent deployment throughout the season.

Looking at pairwise GA probailitiy, player pairs Guiarnov-Robertson and Sekera-Benn standout with the lowest likihoods of letting up a goal per game, 4.2% and 4.3% respectively[13], well be below the average of 14.9%. The 4 skaters also maintain average (~2.77) DB to their teammates, the highest preforming DB teammate is Joel Hanley at 8.24[14].

Overall Starting Line Optimization:

The data supports a starting five of Robertson, Pavelski, Hintz, Heiskanen, and Miller as the most effective overall unit. These five skaters not only rank at the top in point share metrics but also exhibit strong efficiency in goal creation and suppression. This line provides the Stars with the highest offensive production while maintaining defensive stability.

Extensions of the Analytics Question and Recommended Data to Accomplish the Extensions

While the current analysis focuses on 5v5 even-strength goals, shots, and pairwise efficiency should we be able to obtain additonal metrics extensions to the question could be further proposed.

- **Opponent Context Adjustment**: Adjust efficiency metrics for opponent strength or team. Perhaps incorporating variables such as oppsoing team defensive ratings, win percentage against. This contextualizes duo performance across varying levels of competition. This could answer the extension of who should we play when playing against teams with varying play styles.
- **Special Teams Analysis**: Segment data by strength state to assess pair effectiveness on power plays and penalty kills. This reveals impact beyond even-strength play and wuld enable coaching staff to devolp strategic deployment strategies on special teams.
- **Score-State Context**: Introduce score differential at the time of each event to assess pair performance under pressure. This highlights which players elevate performance in critical or high-leverage situations.
- **Puck Possession and Location**: Integrate spatial and possession-based metrics to evaluate how effectively pairs generate offensive zone time or high-danger chances. This adds depth to raw efficiency analysis.

---

[9] See Figure 7 for efficiency table

[10] See Figure 4 for DPS Table

[11] See Figure 5 f0r PS table

[12] See Figure 6 for Z-score table

[13] See Figure 8 for Pairwise Probability for GA/60min

[14] See Figure 12 for top 10 DB skaters.

Appendix A

Appendix A, Figure 1, *Contributive Calculations to Offensive Point Shares (OPS) and Defensive Point Shares (DPS)*

$$Marg.GF = Team\ Goals - \left({}^{7}/_{12}\right)(Team\ Games)(Leauge\ Goals\ per\ game)$$

$$Marg.GA = \left(1 + \left({}^{7}/_{12}\right)\right)(Team\ GP)(Leauge\ Goals\ per\ game) - (Team\ GA)$$

$$Goals\ Created = \left(Goals\ Scored + 0.5(Assists)\right) * \left(\frac{Team\ Goals}{Team\ Goals\ + 0.5(Team\ Assists)}\right)$$

$$Marg.GF\ Skater = Goals\ Created - {}^{7}/_{12}\,(TOI)\left(\frac{Goals\ created\ by\ F\ or\ D}{TOI\ for\ F\ or\ D}\right)$$

$$Marg.Goals\ per\ Point = \frac{leauge\ goals}{leauge\ points}$$

$$Prop.of\ team\ TOI\ for\ skater\ = \frac{TOI}{team\ TOI\ for\ skaters}$$

$$Prop.of\ team\ marg.GA\ for\ skater = \left(7 - 2\left(\frac{team\ SA\ per\ min}{leauge\ SA\ per\ min}\right)\right)\Big/7$$

$$Positon\ Adjuistment\ (Defenseman) = {}^{10}/_{7}$$

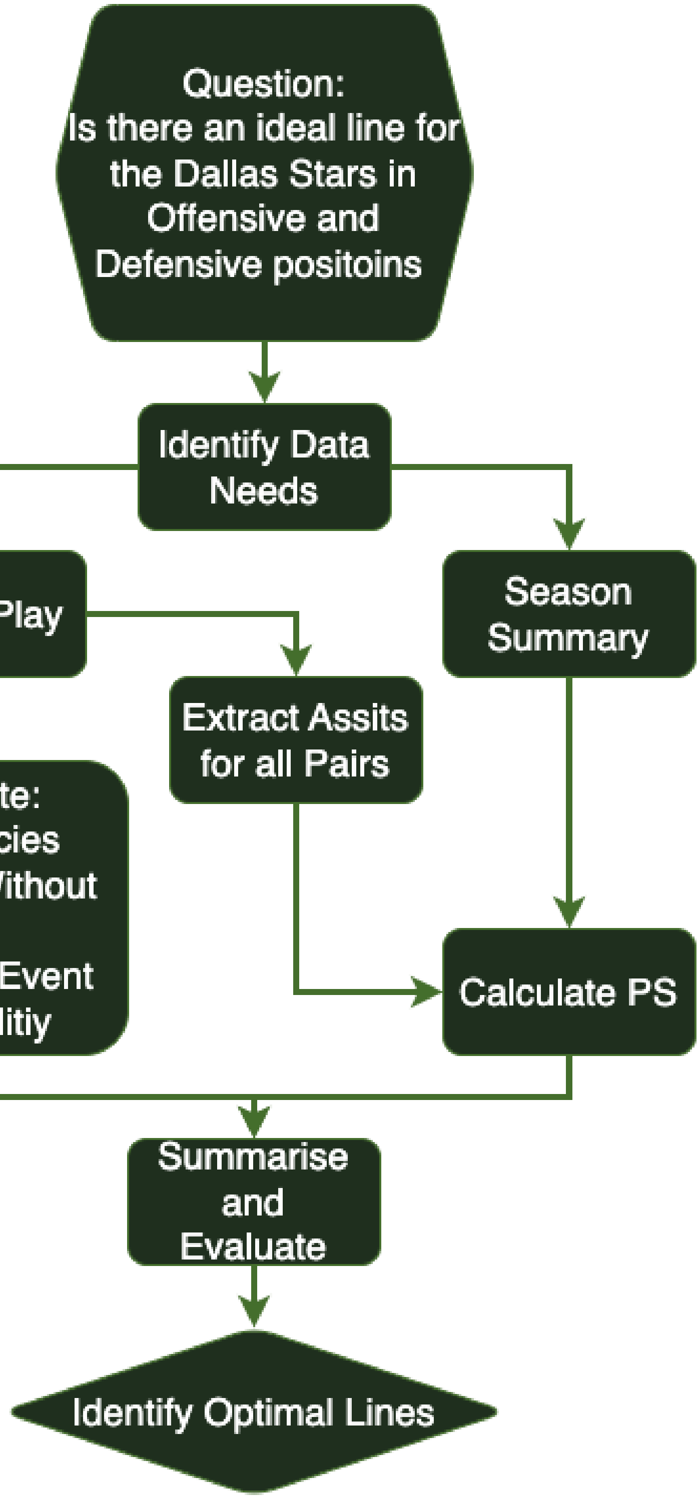
$$Positon\ Adjuistment\ (Forwards) = {}^{5}/_{7}$$

$$team\ Marg.GA = \left(1 + {}^{7}/_{12}\right)(Team\ GP)(Leauge\ Goals\ per\ G) - (Team\ GA)\backslash$$

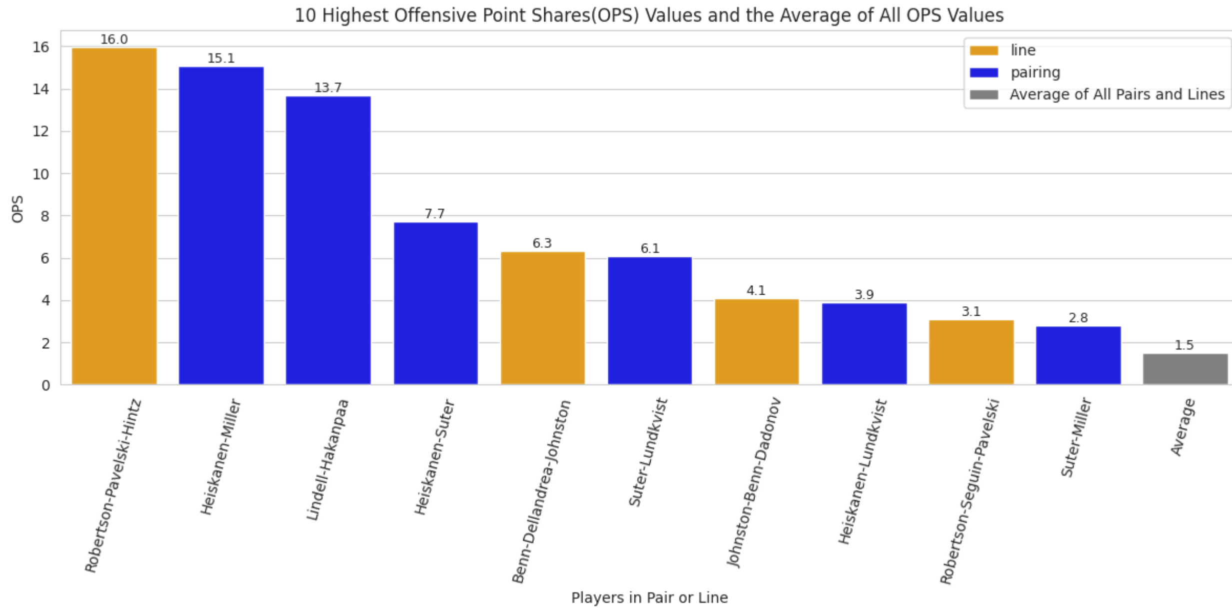
$$+/- Adjustment = \left({}^{1}/_{7}\right)(Pos.Adj.)\left((+/-) - TOI\left(\frac{Team\ +/-\ for\ pos.}{Team\ TOI\ for\ pos.}\right)\right)$$

$$Marg.GA = (Prop.team\ TOI)(Prop.team\ Marg.GA)(Pos.Adj.)(team\ Marg.GA) + (+/-\ ADJ.)$$

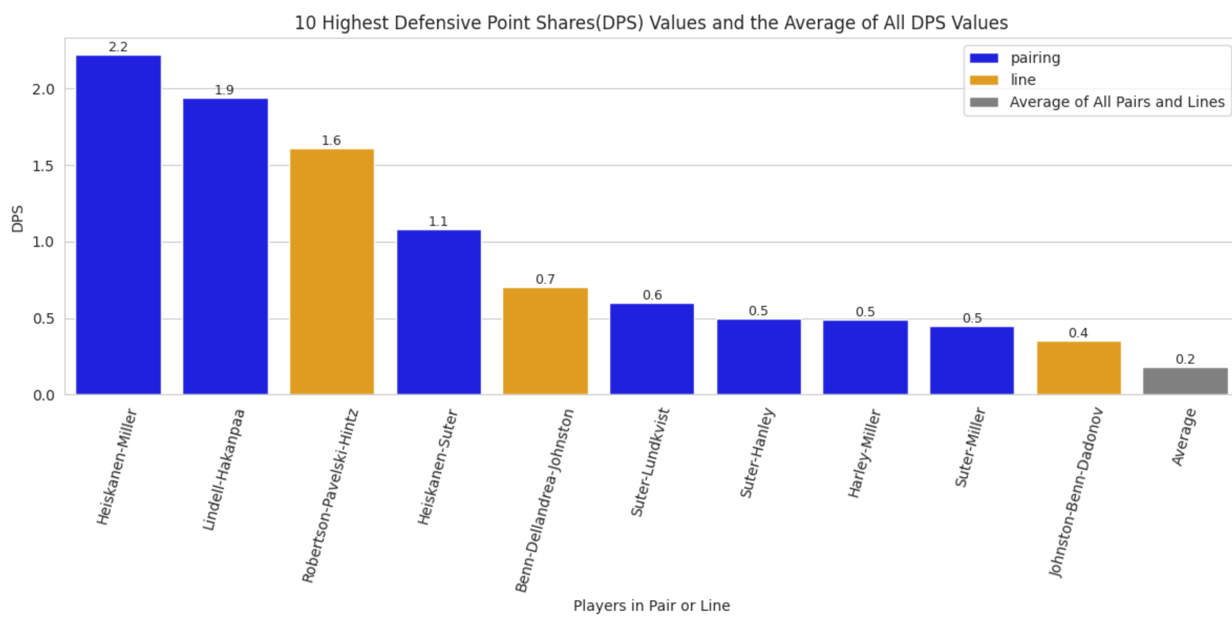
Appendix A, Figure 2, Flowchart of Analytic Logic



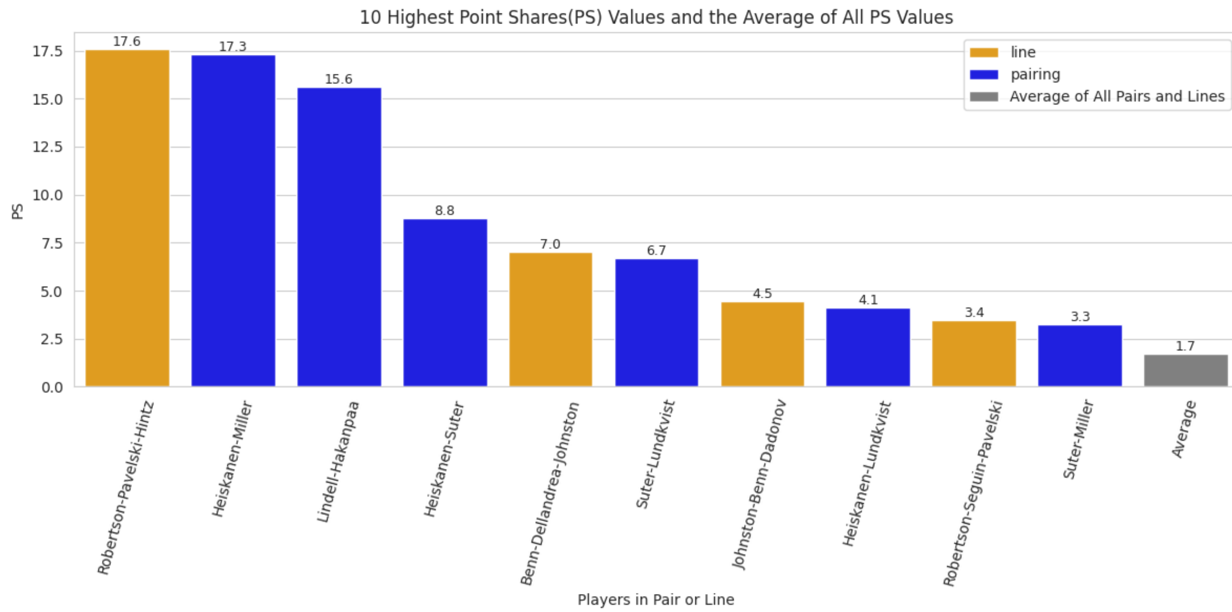
Appendix A, Figure 3, *Top 10 OPS Pairs*



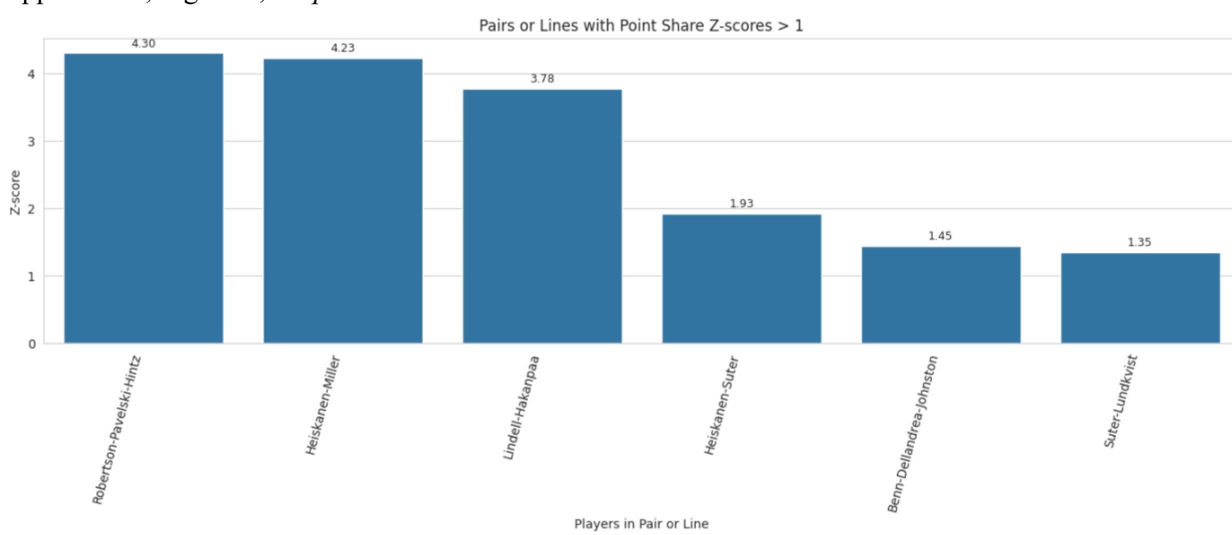
Appendix A, Figure 4, *Top 10 DPS Pairs*



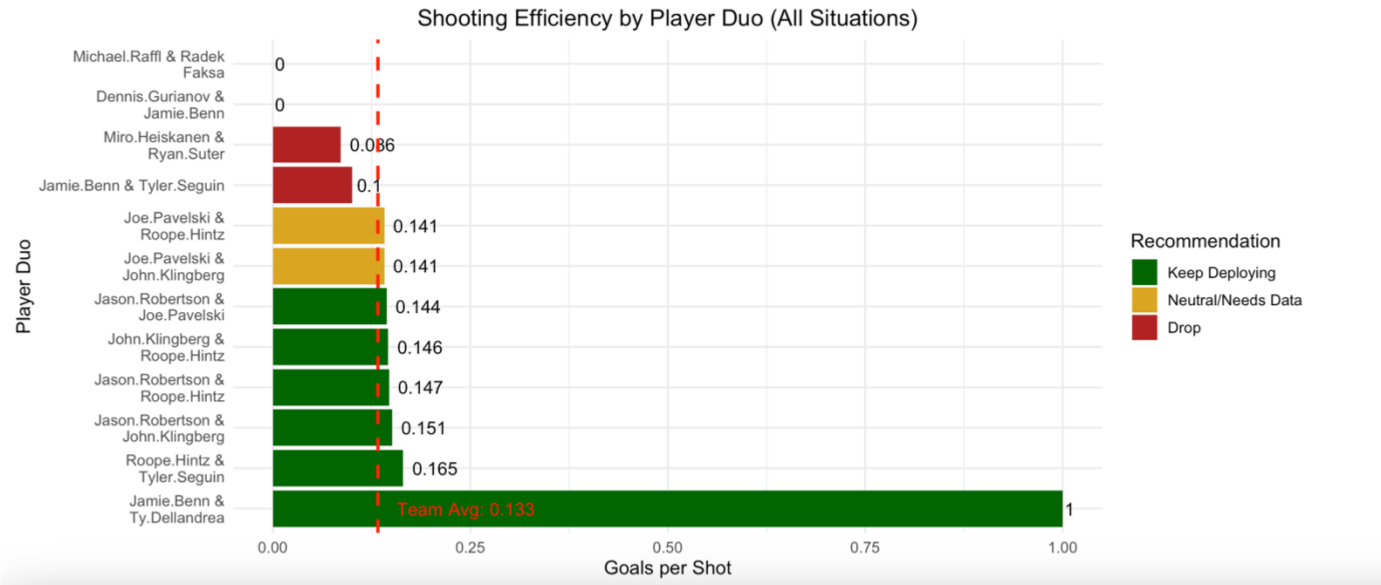
Appendix A, Figure 5, *Top 10 PS Pairs*



Appendix A, Figure 6, *All pairs or lines with PS z-scores > 1*



Appendix A, Figure 7.1, *Shooting effecieny by player duo*



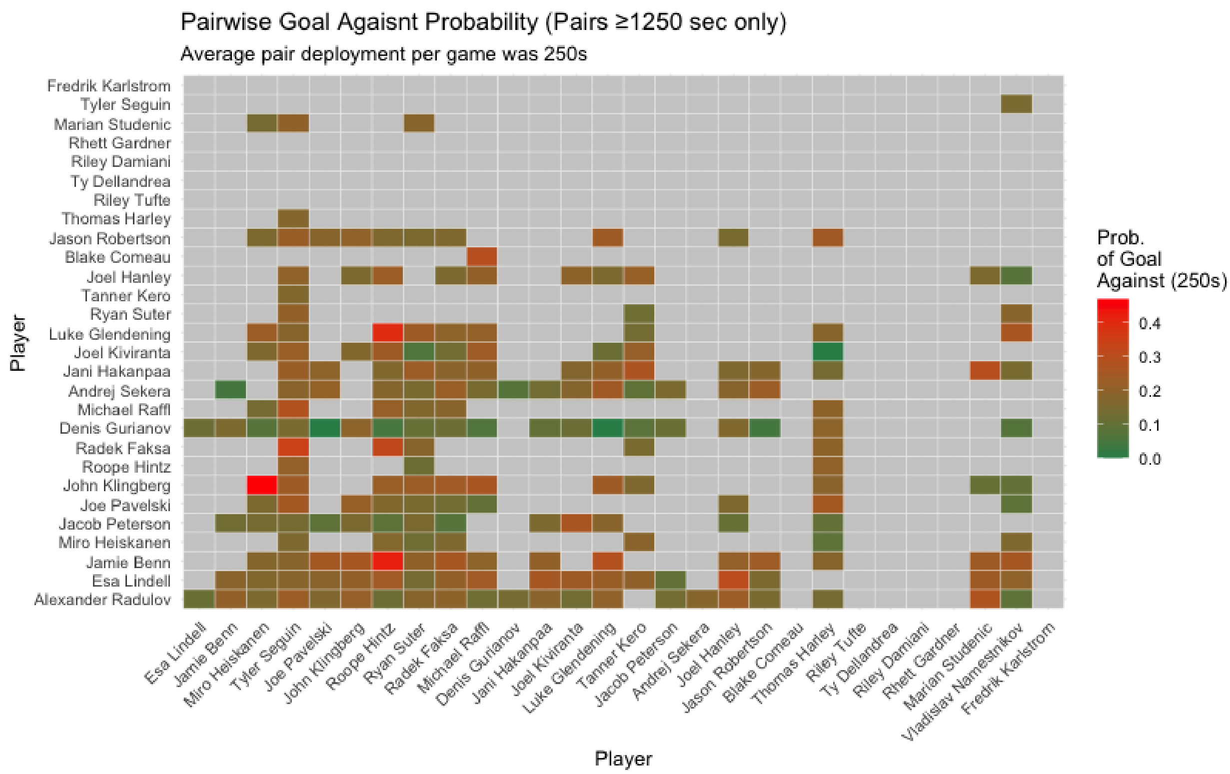
Appendix A, Figure 7.2, *Preformance Summary Table for Pairs observed in Figure 6*

Dallas Stars Duo Performance Summary (5v5)

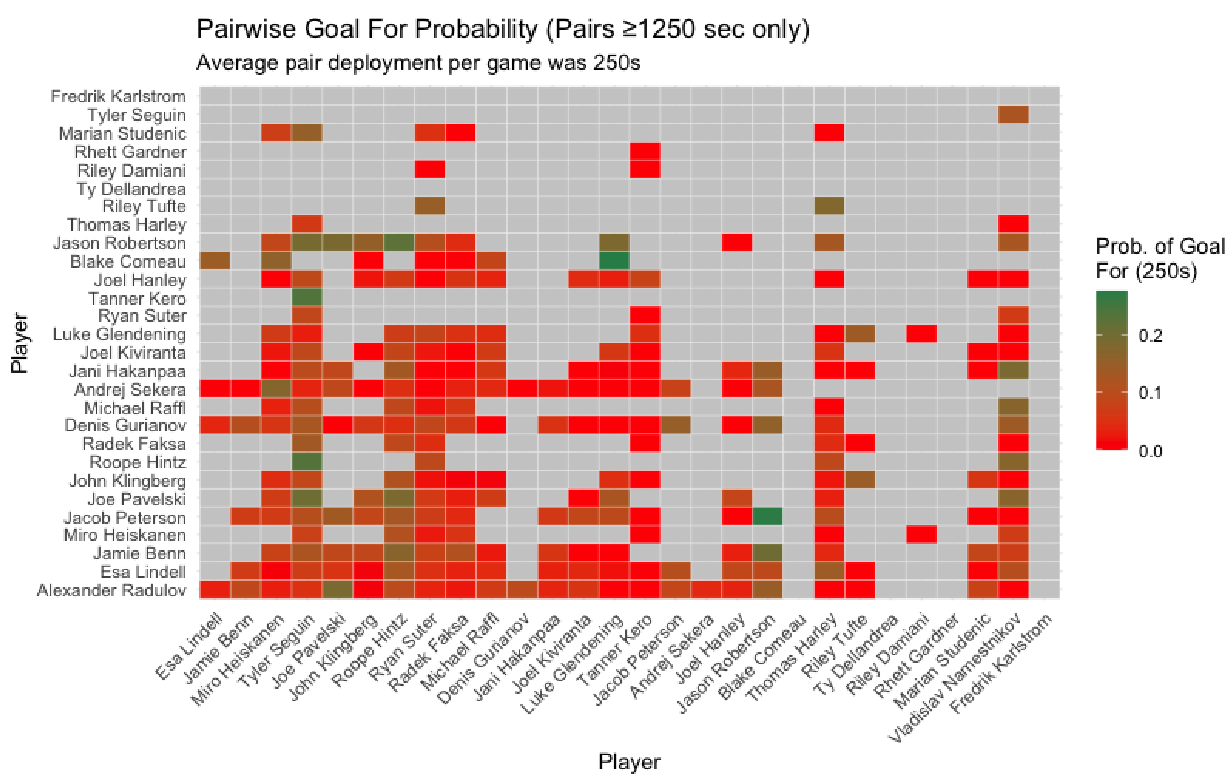
Team Average Efficiency: 0.1

Player Duo	Goals	Shots	Goals per Shot	Shot Volume	Recommendation
Jamie.Benn & Ty.Dellandrea	1	1	1.000	Low	Efficient but underused
Jason.Robertson & Roope.Hintz	58	496	0.117	High	Top-line, keep deploying
Jason.Robertson & Joe.Pavelski	61	533	0.114	High	Top-line, keep deploying
Joe.Pavelski & Roope.Hintz	56	508	0.110	High	Top-line, keep deploying
John.Klingberg & Roope.Hintz	23	227	0.101	Low	Efficient but underused
Jason.Robertson & John.Klingberg	25	260	0.096	Medium	Neutral or needs more data
Joe.Pavelski & John.Klingberg	26	274	0.095	Medium	Neutral or needs more data
Jamie.Benn & Tyler.Seguin	24	305	0.079	Medium	Neutral or needs more data
Miro.Heiskanen & Ryan.Suter	32	443	0.072	Medium	Neutral or needs more data
Michael.Raffl & Radek Faksa	0	0	0.000	Low	Neutral or needs more data
Dennis.Gurianov & Jamie.Benn	0	0	0.000	Low	Neutral or needs more data

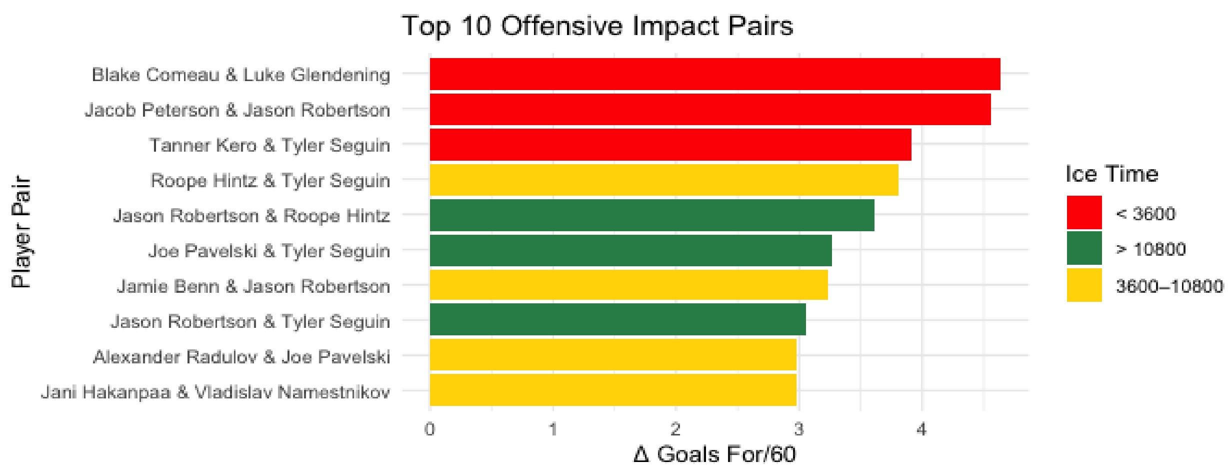
Appendix A, Figure 8, *Pairwise GA Probability within 250s for pairs TOI > 1250*



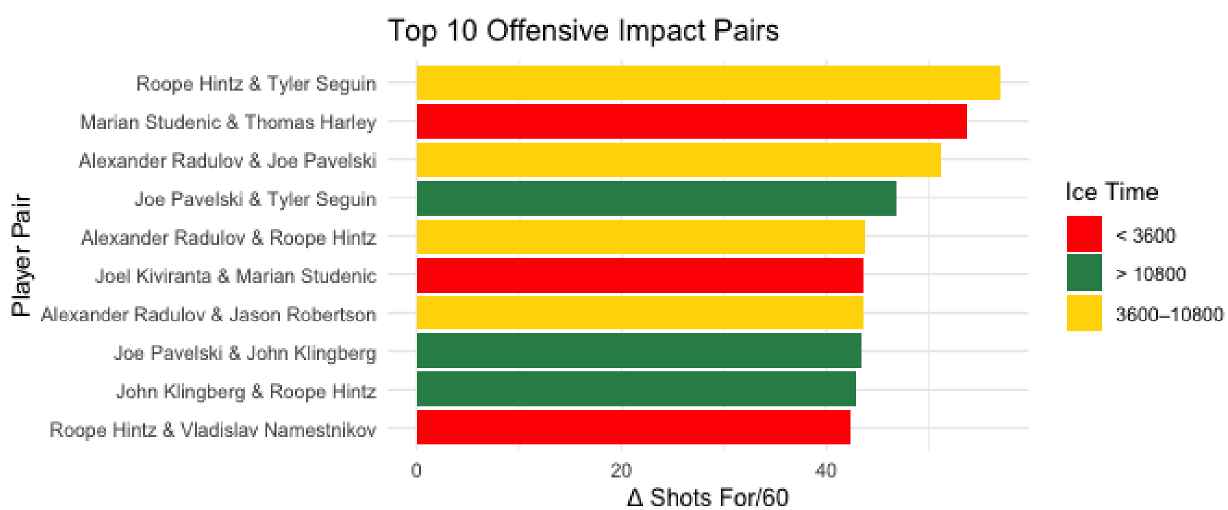
Appendix A, Figure 9, *Pairwise GF Probability within 250s for pairs TOI > 1250*



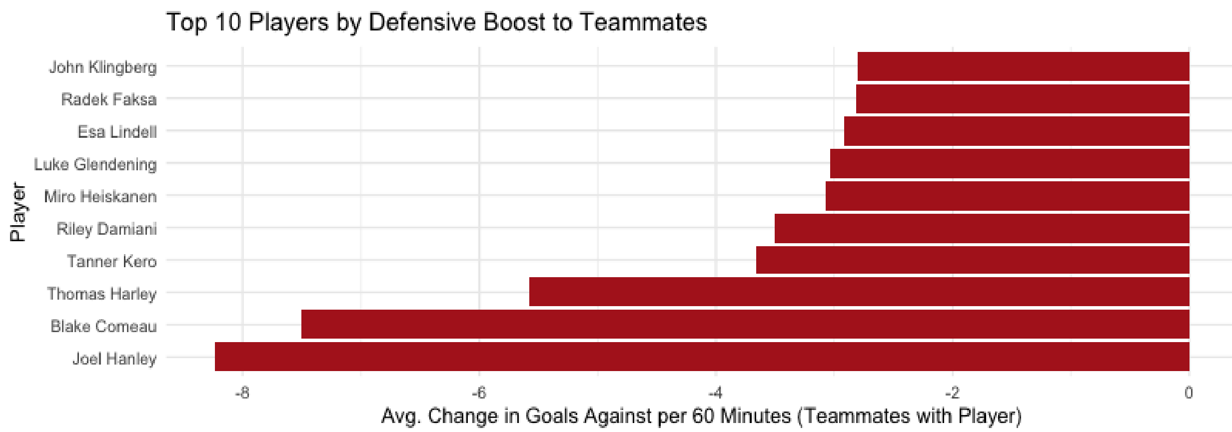
Appendix A, Figure 10, *WOWY Analysis for Delta GF/60*



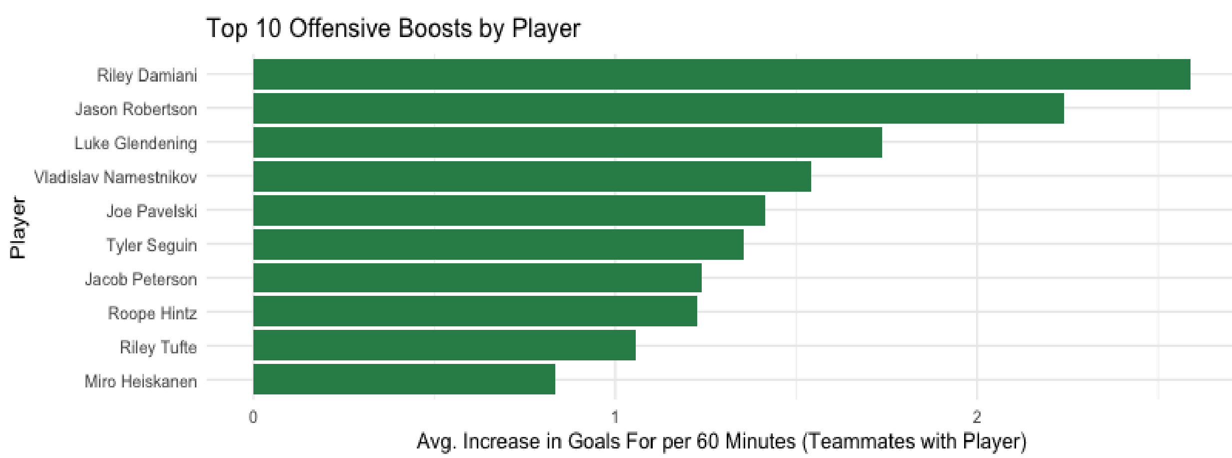
Appendix A, Figure 12, *WOWY Analysis for Delta SF/60*



Appendix A, Figure 12, *Players with highest calcuatated Defensive Boost*



Appendix A, Figure 13, *Players with highest calcuatated Offensive Boost*



Works Cited

"2022-23 Dallas Stars Roster and Statistics." *Hockey Reference*, www.hockey-reference.com/teams/DAL/2023.html.

"2022-23 NHL Summary." *Hockey Reference*, www.hockey-reference.com/leagues/NHL_2023.html.

BasuMallick, Chiradeep. "R vs. Python: 12 Key Comparisons." *Spiceworks*, 6 Dec. 2022, www.spiceworks.com/tech/devops/articles/r-vs-python/.

"Dallas Stars History." *Hockey Reference*, www.hockey-reference.com/teams/DAL/history.html.

"Download Player and Team Data." *MoneyPuck*, moneypuck.com/data.htm.

Kubatko, Justin. "Calculating Point Shares." *Hockey Reference*, www.hockey-reference.com/about/point_shares.html.

"NHL Stats Glossary." *Hockey Reference*, www.hockey-reference.com/about/glossary.html.