What is the Home Ice Advantage in the National Hockey League?

W200 – Project 2
December 2018
Robert Bergan, Brent Biseda, Michael Hardy

Contents

Table of Figures 1
Overview 2
Focusing Questions 2
Dataset 3
Data Validation and Preparation 3
Exploratory Questions 4
Section 1: What is the home ice advantage and how does it vary between season? in the regular season vs. playoff season? 4
Overall and Regular Season vs. Playoffs 4
Trends by Time (Across Seasons and Within Seasons) 4
Advantage by Team 5
Section 2: How does the home-ice advantage manifest itself into different aspects of the game, such as face-offs, power-plays or venue? 9
Time Zone 9
Game Play Variables 9
Analysis of Team Level Statistics 10
Power Play Detail 11
Face Off Detail 12
Section 3: How do individual players contribute to home-away advantage? 12
Analysis of Top Performing Players 12
Analysis of Players by Team and by Season 14
Conclusion 15

Table of Figures

Figure 1 - Distribution of Goals Scored (Home vs. Away) 2
Figure 2: Home Win Rate by Season 4

Figure 3: Home vs Away Goals by Team 6
Figure 5: Home vs. Away Goal Differential Mean and 1 Std and Home vs. Away Win Differential 6
Figure 6: Win Differential by Good vs. Bad Teams 7
Figure 7: Goals Scored vs. Given Up, Home vs Away 8
Figure 8: Goal Differential by Team by Year 8
Figure 9: Impact of Time Zone Differential on Goal Differential 9
Figure 10: Impact of Venue on Goal Differential ..... **Error! Bookmark not defined.**
Figure 11: Home Win Correlation Heat Map 10
Figure 12: Game Statistics by Category (Home v. Away) 11
Figure 13: Powerplay Goals Explains Home and Away wins 11
Figure 14: Faceoff Wins vs. Goals 12
Figure 15: Top 50 Player Statistics by Game Category 13

Overview

In all almost all sports there is a phenomenon whereby the "home" team typically wins a higher percentage of games than the "away" team. This home-team advantage is predictable over a large enough sample size such that it is a major factor in sports betting. Here we seek to investigate this home-ice advantage for the National Hockey League ("NHL") in order to quantify it, identify the drivers, and to see where the advantage manifests in various facets of the game. We are interested specifically in the NHL because hockey is known as the "most-random" of the major US sports, making it the least-analyzed sport from the perspective of data science. As can be seen in the figure below, there exists a demonstrable difference between the distribution of goals scored at home vs. away. As seen by table 1, through time there is roughly a 0.3 goal differential in favor of the home team.

Table 1 - Goals at Home vs Away

year	goals_home	goals_away
2013	2.802888	2.476534
2014	2.830806	2.525140
2015	2.740941	2.516577
2016	2.803089	2.548263
2017	2.944694	2.646487
2018	3.068365	2.843164

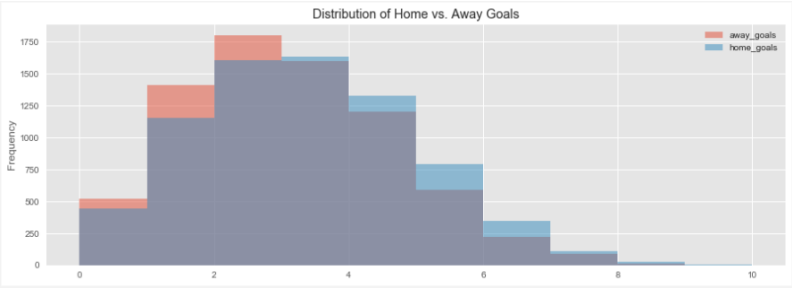

Figure 1 - Distribution of Goals Scored (Home vs. Away)

Focusing Questions

Our focusing questions divide our report into three sections: a high-level look at the home-ice advantage, a deeper look at the drivers and manifestations of home-ice advantage, and an examination of the impact of individual players:

- What is the magnitude of the home-ice advantage and can we see any differences or trends across seasons, during different parts of the season or between teams?
- How does the home-ice advantage manifest itself into different aspects of the game, such as face-offs, power-plays, hits and other factors?
- What factors help in explaining home-ice advantage and is there an impact from individual players?

Dataset

The dataset we used in our analysis was sourced from Kaggle, but its original source is directly from the NHL's website API. The data includes all data collected for the six seasons from 2012/2013 season – 2017/2018 season.

The dataset includes nine separate tables, all in csv format, that include different data or summaries of data. Our analysis focuses primarily on game data (as opposed to data on individual players). The main tables used in this analysis are summarized below:

- 'game' table:
  - Summary data on each regular season and playoff game including date/time, season, teams involved (home and away), outcome, type (playoff vs. regular season), venue, goals scored
  - 7,441 rows, 16 columns
- 'game_teams_stats' table:
  - Additional detail on each game. Includes 1 row for each team in each game (so 2x rows of 'game' table, including coach, penalties, giveaways, takeaways, face-offs
  - 14,882 rows, 14 columns

Data Validation and Preparation

As the dataset is bounded by the number of games per season and the associated statistics for those games, we can validate our data by cross-checking the aggregated data against these known bounds. In doing so we can confirm that all games and teams are present within the dataset for both the regular and playoff seasons. This data validation shown in the tables below, allows us to conclude that there are no missing games as each team played 82 games (except for 2012-2013 which was a shortened season) and that the detailed game statistics seem reasonable. We performed these types of analyses across our variables and could not find any issues with our source data.

Table 3 - Example Validation Table

season	games	teams	goals	games_per_team	goals_per_game	goals_per_team_per_game
20122013	720	30	3919	48.0	5.443056	2.721528
20132014	1230	30	6751	82.0	5.488618	2.744309
20142015	1230	30	6719	82.0	5.462602	2.731301
20152016	1230	30	6672	82.0	5.424390	2.712195
20162017	1230	30	6803	82.0	5.530894	2.765447
20172018	1271	31	7552	82.0	5.941778	2.970889

Table 2 - Missing Values

season	0
type	0
date_time	0
away_team_id	0
home_team_id	0
away_goals	0
home_goals	0
outcome	0
home_rink_side_start	116
venue	0
venue_link	0
venue_time_zone_id	0
venue_time_zone_offset	0
venue_time_zone_tz	0
away_team	0
home_team	0
side_won	0
month	0

dtype: int64

Additionally, we looked for missing values across both tables and found no missing values in the "game_teams_stats" table and 116 in the "games" table. However, the missing variables in the "games" table does not impact our analysis as the variable is not useful to exploring our questions.

From a preparation perspective, our main task was to join / merge tables. Throughout our analysis we merged these two main tables, along with additional information tables such as the teams table to get the team names. The tables included keys for merging so we were able to merge or aggregate data across tables relatively seamlessly.

Exploratory Questions

Section 1: What is the home ice advantage and how does it vary between season? in the regular season vs. playoff season?

Overall and Regular Season vs. Playoffs

We knew there would be a home-ice advantage in the NHL, but we did not know the magnitude. From our data we can see that, on average, between the 2012-2013 season and the 2017-2018 season the home team has won 54.95% of all games. As a point of comparison, in the NFL, the average is closer to 60%. Intuitively, it makes sense that the NHL has less of a home-team advantage due to it being thought of as more random than the NFL and thus any advantage (home-ice, skill, etc.) should be smaller.

Of interest though is that the playoff advantage (54.84%) is almost 2 percentage points greater the regular season advantage (56.42%). However, as we will see below, the playoff percentage varies significantly by season, likely due to whether the championship team, which plays more games, wins more on the road or at home. In fact, there is most certainly bias in that the best teams have the opportunity to play additional games at home.

Trends by Time (Across Seasons and Within Seasons)

In all seasons, the regular season percentage has hovered in a range between 52.9% (2015-2016) and 56.81% (2012-2013, the lockout season) while the playoff percentage has fluctuated more dramatically between 68.6% in 2012-2013 and 47.6% in 2017-2018 (see Figure 2). Though it is hard to tell how substantial the regular season difference is without doing further statistical analysis. We do note that a lot of the variance seems driven by a few outlier teams. For instance, in 2012-2013 the LA Kings and San Jose Sharks both had a greater than 40% difference in their home win rate vs. their road win rate, an amazingly high differential. We will discuss time zone impact later in our analysis but note that any travel related impact could be magnified in the lockout shortened season of 2012-2013 when the games were compressed into a short period of time.

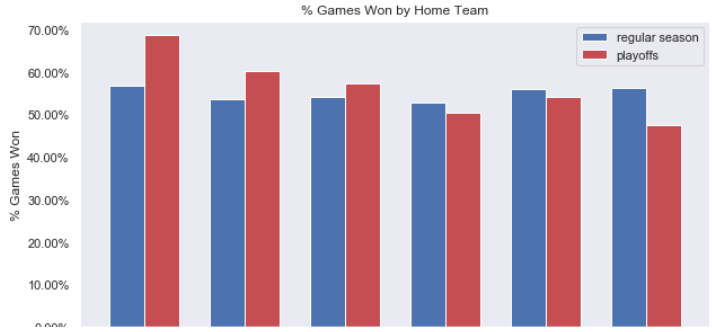

Figure 2: Home Win Rate by Season

For the playoff data we assert that with a small number of games the percentage is skewed significantly by the team that has success in that particular year. For instance, the LA

Kings played the most playoff games in 2012-2013 because they won the championship that year and had a significantly higher home win-rate, while the Washington Capitals won in 2017-2018 and played better on the road.

While there is a clear home ice advantage, that advantage has varied, even decreased, for playoff times over time, to the point where, compared to other sports, hockey's home playoff advantage doesn't even seem a sure thing[1]. This is unusual given the fact that home hockey teams have the right to make the final line change before the puck drops, giving them a consistent edge in matchups. Nevertheless, the variability makes hockey unique among the professional sports.

Table 4: Playoff Home Advantage

	HOME TEAMS IN REG. SEASON		HOME TEAMS IN PLAYOFFS		
LEAGUE	WIN PERCENT	BOOST*	WIN PERCENT	BOOST	PLAYOFF DIFF.
NFL	57.1%	+7.0	64.7%	+11.8	+4.8
NBA	59.9	+9.8	64.5	+13.9	+4.1
MLB	54.0	+4.0	54.2	+4.0	+0.0
NHL	55.1	+5.1	55.3	+4.8	-0.3

One of our main takeaways from this portion of the analysis is that while we need to investigate the time zone impact, there is no specific trend across seasons that we need to further investigate. In addition to checking trends across seasons, we examined trends within seasons and note that there appears to be little variance by the month of the season.

Advantage by Team

After exploring home ice advantage across the league, it is constructive to explore home ice advantage by team as this provides insight into other factors such as venue, style of play, time zone, and team skill level. As shown in the figures below, there is substantial variation among teams at home compared to on the road both in the number of goals scored and win-rate. As an overview, we visually represented all the teams' home goals vs away goals, highlighting the differential by color to identify teams most impacted by home ice advantage.

[1] Data adapted from https://fivethirtyeight.com/features/a-home-playoff-game-is-a-big-advantage-unless-you-play-hockey/

We then explored deeper, by comparing the differential (indicated by color in Figure 3) across team and season. There are two interesting insights from these charts: first, the advantage does not appear to be based on the performance of the team as we know anecdotally that there are a number of teams that have been playoff contenders, such as the Blues that have a low differential while other playoff contenders, such as the Penguins have a high differential; second, the order of the teams by home vs. away goal differential does not always align to the order by win differential. For instance the Lightning, Maple Leafs, Penguins and Kings have a high win rate differential but a middling-to-low goal differential while the Golden Knights rank at the top in both categories. However, the Golden Knights are a new NHL expansion team with limited play history, another example of bias. This suggests to us that while teams may perform better or worse at home, they may do so in different ways i.e. they may play better defense on the road and let in fewer goals.

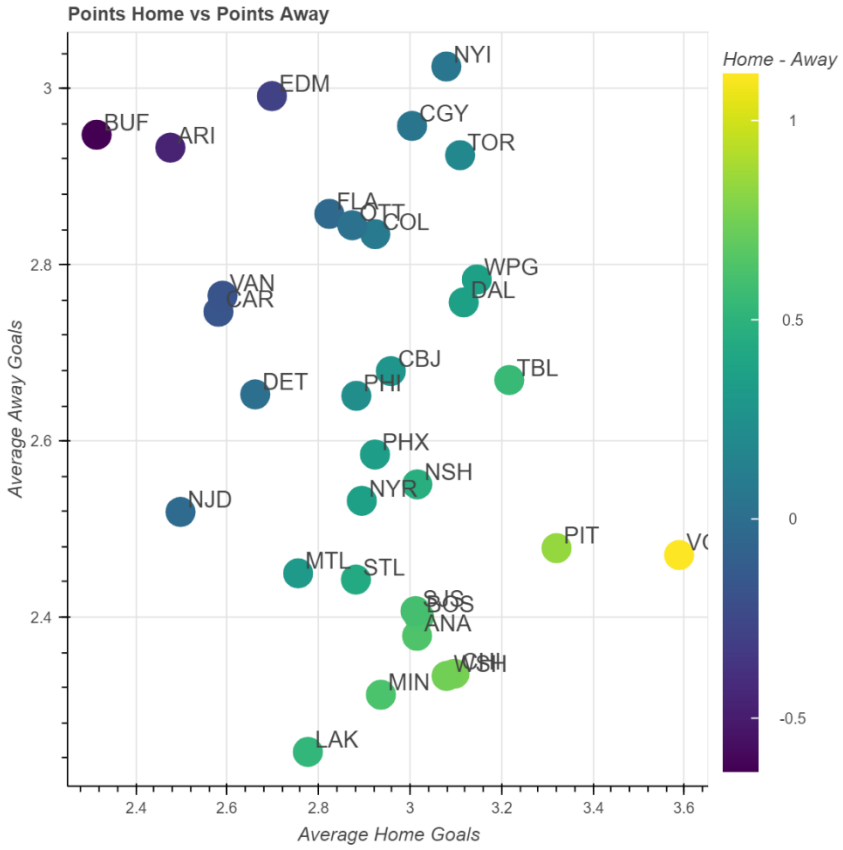

Figure 3: Home vs Away Goals by Team

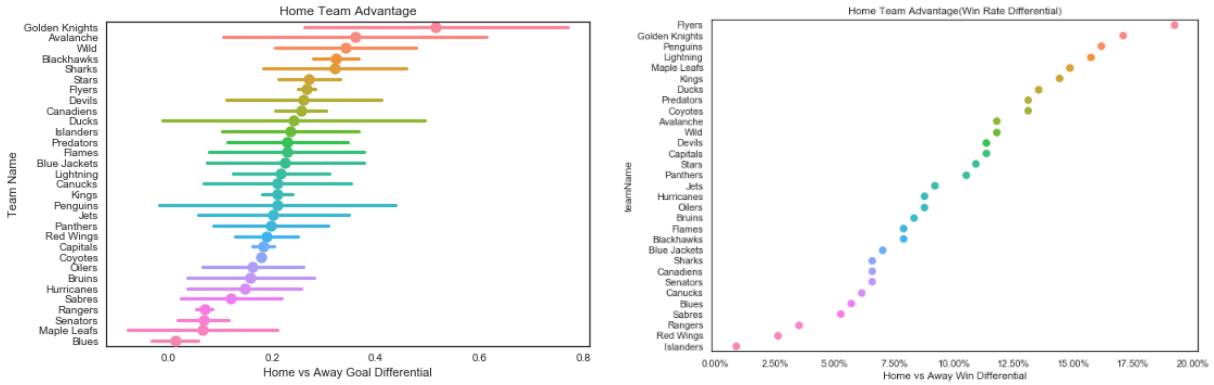

Figure 4: Home vs. Away Goal Differential Mean and 1 Std and Home vs. Away Win Differential

To further explore the advantage by team, first we will look at 'good' teams vs. 'bad' teams to verify whether overall performance is correlated with home vs. away differential. For this analysis we have defined good teams as the top 10 teams as ranked by win total. As shown in Figure 5 we do not see overall performance to be correlated with home win rate (note that 2012-2013 is the lockout season with fewer games and substantial outliers as noted above).

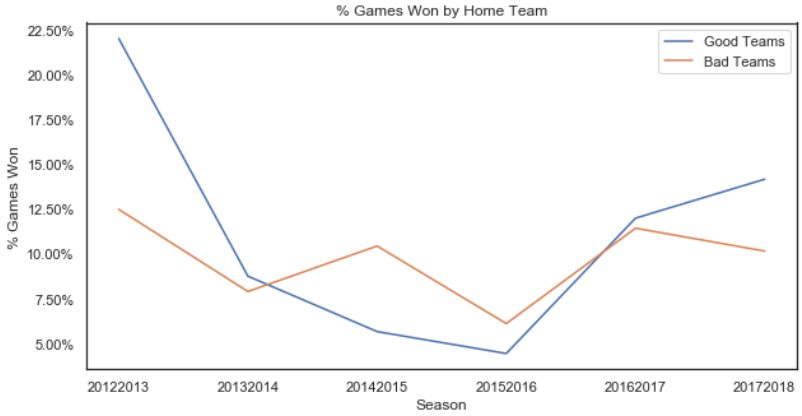

Figure 5: Win Differential by Good vs. Bad Teams

Third, we explore the advantage by team by looking at whether there is a larger difference in offense or defense when comparing home and away statistics. As a proxy for offense and defense we look at the difference in goals scored and the difference in goals given up at home vs. away. This may explain the difference discussed above in the home vs. away goal scored differential and the home vs. away win rate. Figure 6 below shows that there can be substantial difference in where the advantage comes from between teams. For instance, more than 100% of the Flames' advantage seems to stem from offense as the team scores more goals at home relative to the number of goals it gives up at home. The Canadians advantage seems to stem from defense as the team gives up fewer goals at home relative to the increased number of goals scored at home. There doesn't seem to be a very strong correlation between a team's relative home advantage and whether the source of the advantage is offense vs. defense. This is somewhat surprising to us as we would have expected that the potential mechanisms for home ice advantage in the rules (i.e. between face offs and last change) would have been meant a greater benefit for offense relative to defense. As an interesting side note, the Vegas Golden Knights were a new team in the 2017-2018 season and though it was a small sample size, they had the highest home win rate and home goal differential. In the media their home success was often portrayed as a function of 'the Vegas flu': teams would travel to Las Vegas for the first time and visiting players would be out partying all night and be very tired for the game. We can't see this in the data, but would not be surprised if that were the truth.

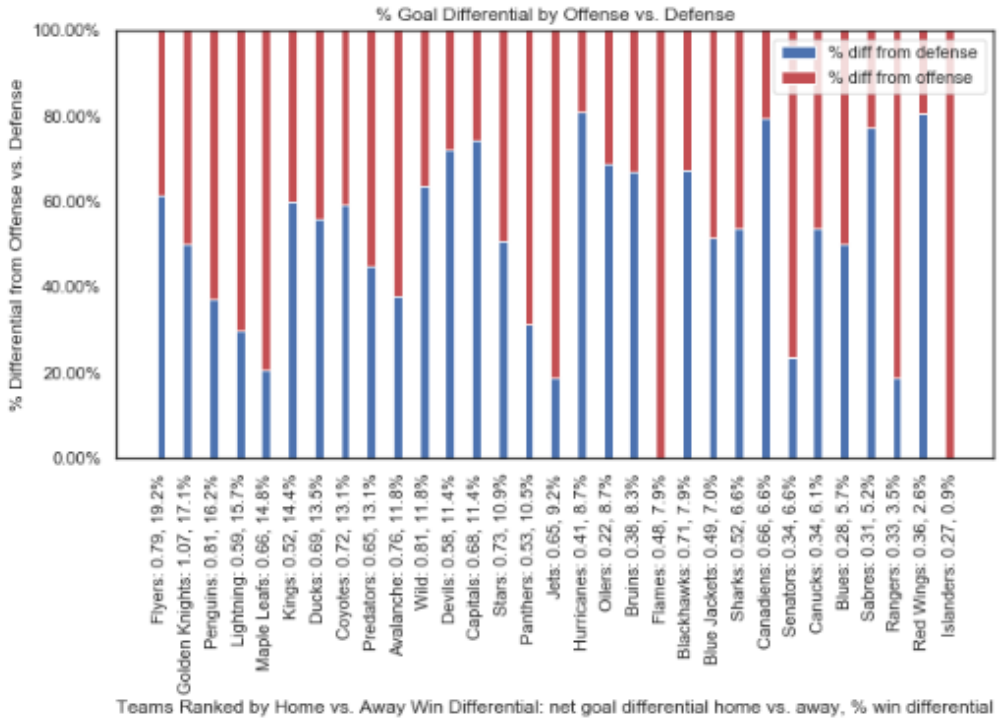

Figure 6: Goals Scored vs. Given Up, Home vs Away

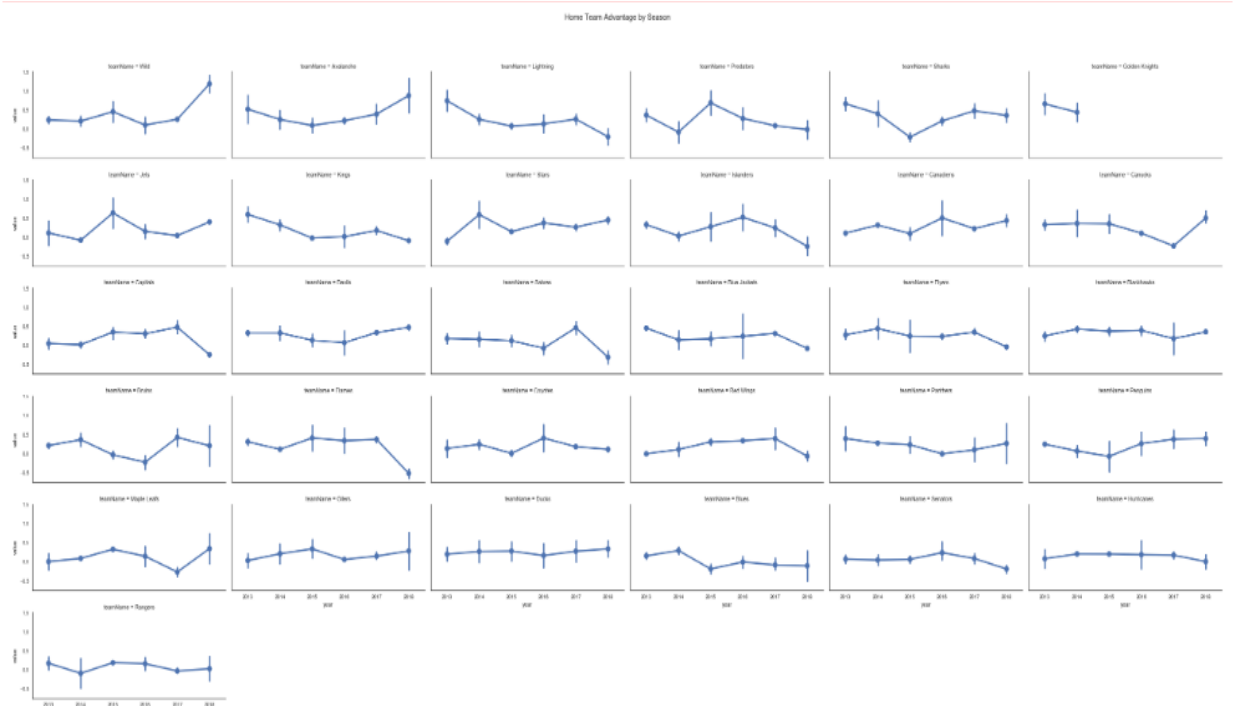

Figure 7: Goal Differential by Team by Year

Based on our analysis in section 1 we can conclude that there is a persistent home ice advantage of approximately 10% (i.e. 55% home team wins, 45% away team wins), but that there do not seem to be any specific trends related to the change in the advantage either between seasons or within seasons.

There does however seem to be a difference in the advantage between teams, as shown both in the outliers that indicate potential time-zone impact and in the differences between goal and win-rate differentials. In the next section we will take a deeper dive into detailed game statistics to examine whether we can find any factors that explain the advantage and the differences between teams.

Section 2: How does the home-ice advantage manifest itself into different aspects of the game, such as face-offs, power-plays or venue?

Time Zone

The figure below shows the impact that time zone differential has on the home team advantage. A time zone difference of -3, corresponds to a West Coast team traveling to the East Coast, while a time zone difference of +3 corresponds to an East Coast team traveling to the West Coast. The expectation would be that West Coast teams playing at home against an East Coast team would have the greatest advantage as the evening games could potentially last until 2 AM EST. In the figure below, we see that variance is higher as a result of travel in all cases outside of staying within a time zone (0), which, suggests that some of the variation in goal scoring can be attributable to travel.

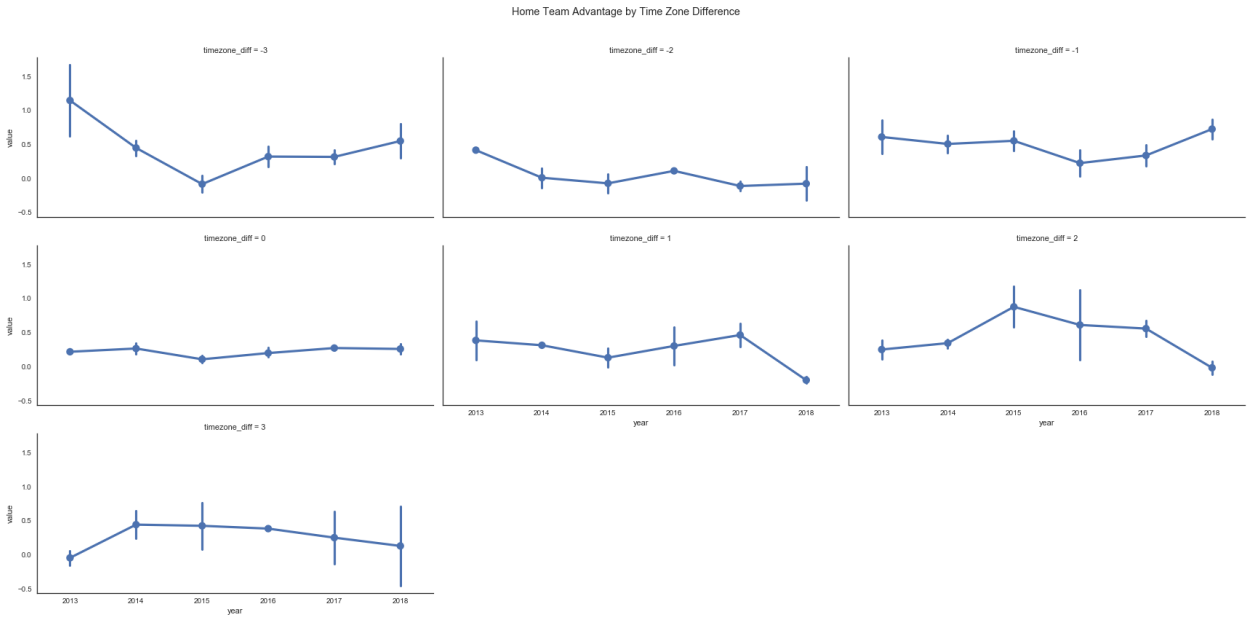

Figure 8: Impact of Time Zone Differential on Goal Differential

Game Play Variables

From each team's perspective we can bin each game into one of four categories: home win, home loss, away win, or away loss. Based off of aggregated stats for each of these categories, we can produce a correlation matrix in order to determine which statistics are the principal contributors to the ultimate success of each team. For the sake of brevity, only the correlation table for home wins is shown in the figure below.

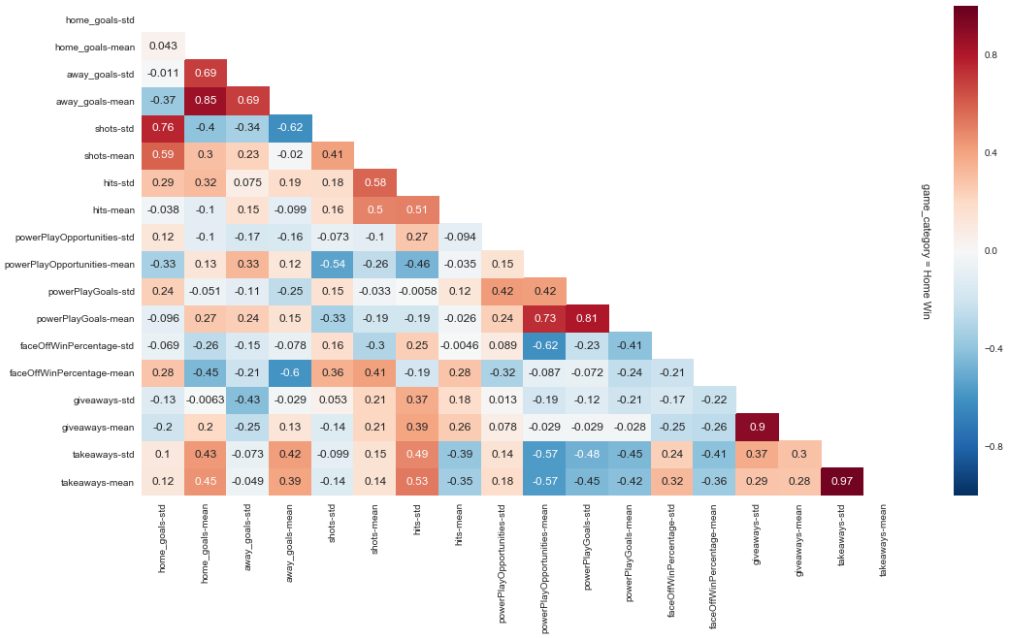

Figure 9: Home Win Correlation Heat Map

Based off the correlation matrix, we may conclude that the mean number of home goals scored is affected by the number of takeaways (0.45), and the number of goals the away team scores (0.85). Interestingly, this phenomenon is understandable as teams seem to try harder if they are losing, and when they have a large lead, allow their top players to rest.

Analysis of Team Level Statistics

Each of the stats present in the data source were analyzed when grouped into the four game categories. From the figures below, it becomes evident that there is no single statistic that is driving the home team advantage. Rather, the home team simply outperforms relative to the away across almost all of the stats, even when the home team loses!

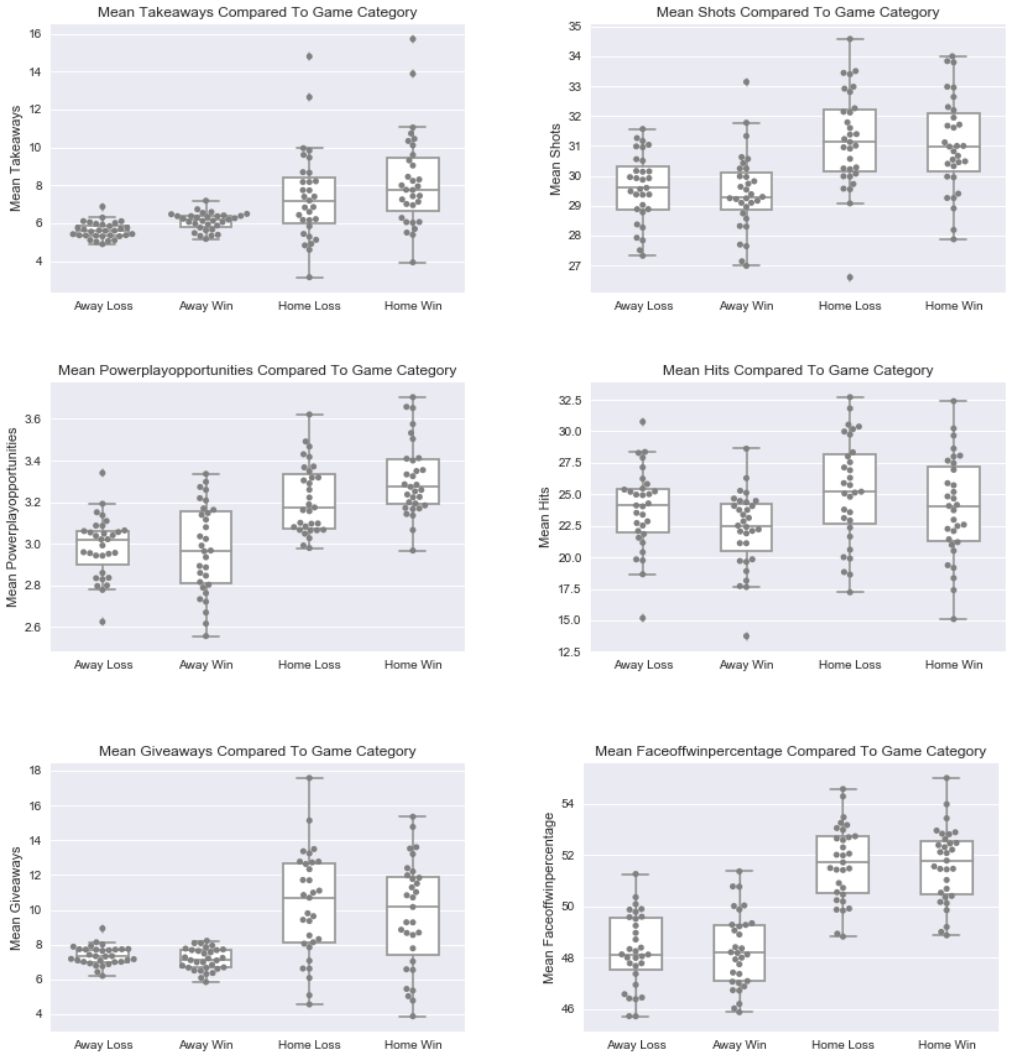

Figure 10: Game Statistics by Category (Home v. Away)

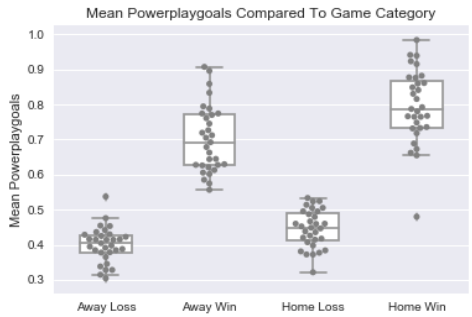

Figure 11: Powerplay Goals Explains Home and Away wins

Of interest is the figure shown to the left, which actually helps to identify one substantial driver in both home and away wins, penalties. This is the most impactful statistic to explain what can lead a team to win when playing away. This suggests that avoiding penalties while playing at home could be a winning strategy to further enhance the edge that playing at home imparts.

Power Play Detail

A power play in hockey is when the opposing team has a penalty and is forced to play without one player for a period of time (usually 2 minutes). According to our analysis, when a team has a power play it scores approximately 19% of the time and power play goals account for 20% of all goals scored. Interestingly

the home team has an average of .28 more power play opportunities than the away team (see related box plot in Figure 10 above). At a 19% conversion rate, this means that it accounts for approximately .05 goals per game for the home team vs. the away team. As the overall advantage is only .27 goals on average, the difference in power play opportunities accounts for approximately 20% of the overall home team advantage in goals scored.

Though we don't know the reason why power play opportunities are different, we can think of two possible explanations. First, referees could be biased by the home town crowd to call more penalties on the away team then the home team or second, the away team could be playing differently, either due to playing from behind, from playing a rougher style on the road, or out of frustration either at the crowd or from being away from home. We don't know the reason, but we think it is an interesting difference.

Face Off Detail

A face off is when the referee drops the puck between two opposing players to start play. In hockey the home team has a slight advantage per the rules as the home team player gets to put his stick down last and thus can get in a better position. Based on our calculations the home team does have an advantage in face offs, winning 51.7% at home vs. 48.3% on the road. This can be seen in the last box plot of Figure 10. Though a face off impacts which team gets the puck and controls the play, it is hard to determine the impact on goals scored. There may be some slight skew towards more goals from a higher face off win percentage (Figure 12), but it is certainly not definitive. There simply too many variables between when a face off wins and when a goal is scored, making any inference challenging.

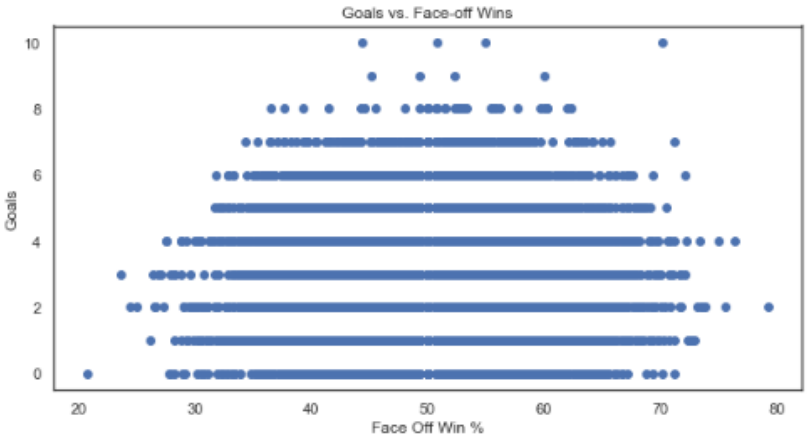

Figure 12: Faceoff Wins vs. Goals

Section 3: How do individual players contribute to home-away advantage?

Analysis of Top Performing Players

A similar analysis to that shown above was performed on the statistics available for the top fifty players in the league. This ranking was determined using lifetime Goals + Assists available within the data set. These statistics were then plotted using the methodology above where game outcome was split into the four categories: home win, home loss, away win, and away loss. Many more statistics are available for the players than the teams. Only a portion of the statistics that were shown to impact the home ice advantage are shown in this document for the sake of brevity.

As can be seen in the figures on the following page, Goals and Assists are the principal drivers of wins at home, an obvious conclusion. Again, another important contributor to victories, both home and away is the plus minus statistic. This statistic is a measure of the players on the ice scoring even handed or at a

man disadvantage as a positive value. Therefore, this is an example of selection bias where the winning categories by definition had more goals scored. Interestingly, when comparing players to teams, players have different factors that can predict the outcome of the game. The top 50 players hit more frequently while at home, and it manifests itself as a contributing factor to home wins. Teams could again try to improve their home ice advantage by focusing on reducing the frequency of giveaways.

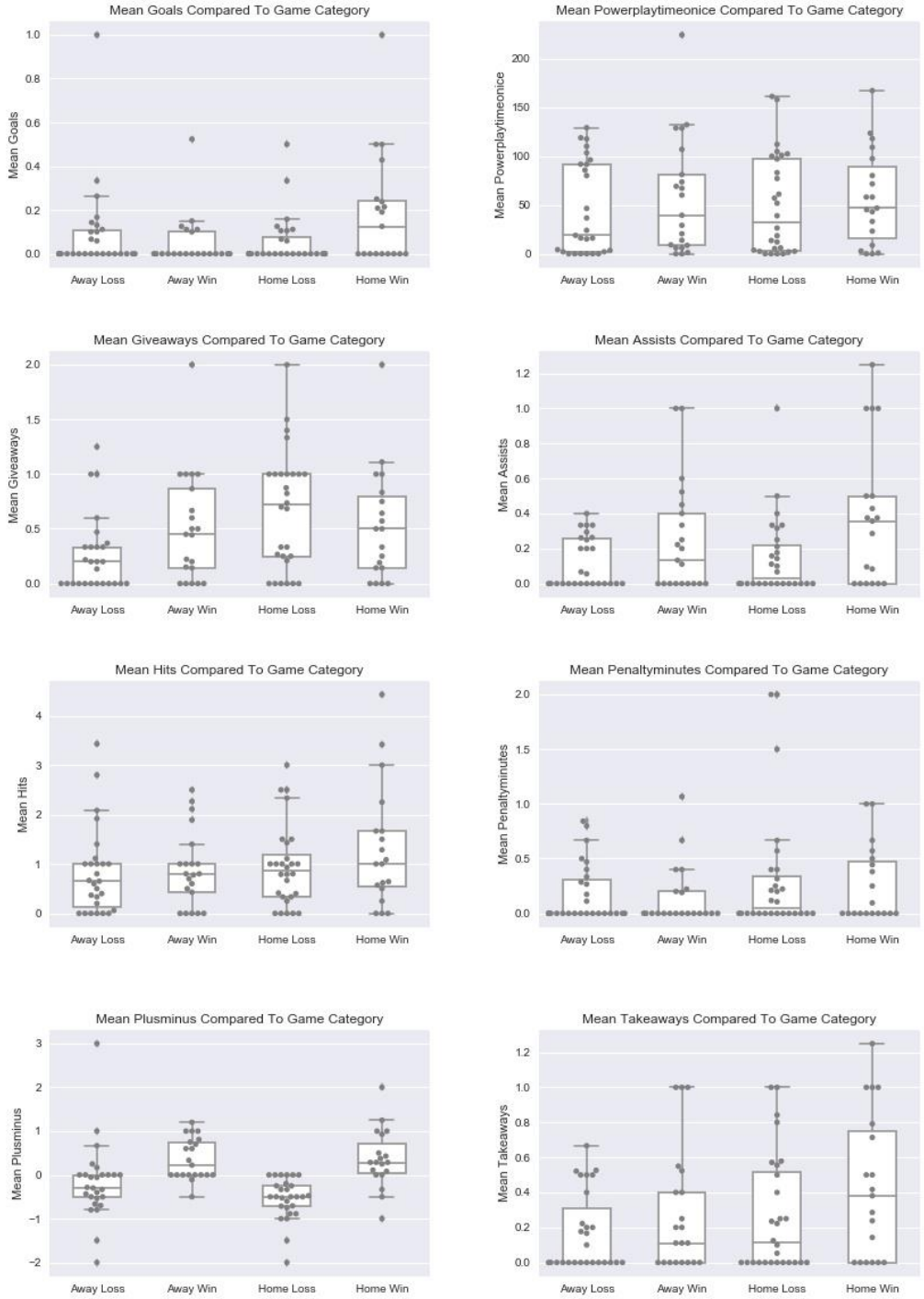

Figure 13: Top 50 Player Statistics by Game Category

Analysis of Players by Team and by Season

To assess whether or not individual players were the key to understanding home ice advantage, we looked at the extent to which an individual player had more/less impact on their team's Home-Away goal differential. We wondered whether, along with the clear advantage of not having to travel, the home-ice advantage may also have something to do with the players' psychological response to their fanbase. If so, it could be that some players have a stronger response than others, and we wanted to see if these players contributed disproportionately to the home ice advantage.

To do this, we created a new metric, "shots per minute on ice", to assess each team member's relative contribution to offense. We then took the differential of this statistic for each player's home game performance minus their away game performance. This difference will allow us to see which players were most impacted by home vs away. To compare this data, we took the teams by season who had the greatest goal differential, and then plotted them along with the each of their players who had a positive impact as measured by their shots-per-minute differential.

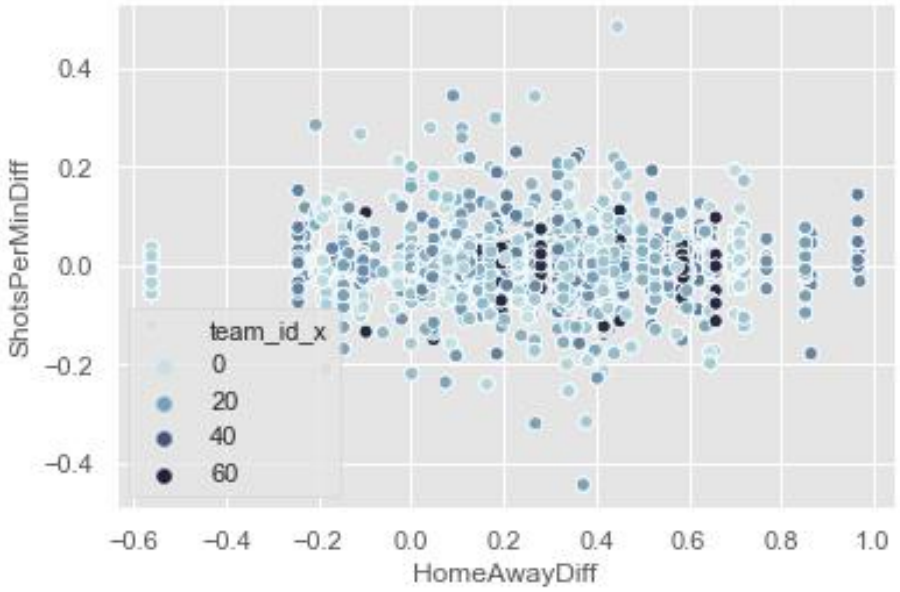

Figure 14: Player Offense Differential by Team Goal Differential

We first plotted all the players, across all teams and seasons, to explore any obvious trends. We would expect that if single players were mostly responsible for the uptick in the goals-differential, we would expect to see players as more noticeable outliers for shots-per-minute differential as team goals differential increases.

However, we see the opposite, as described in the figure below. This suggests that if there is psychological impact from the fan base, its effect is generally broader across the whole team, not acting upon just a few single players.

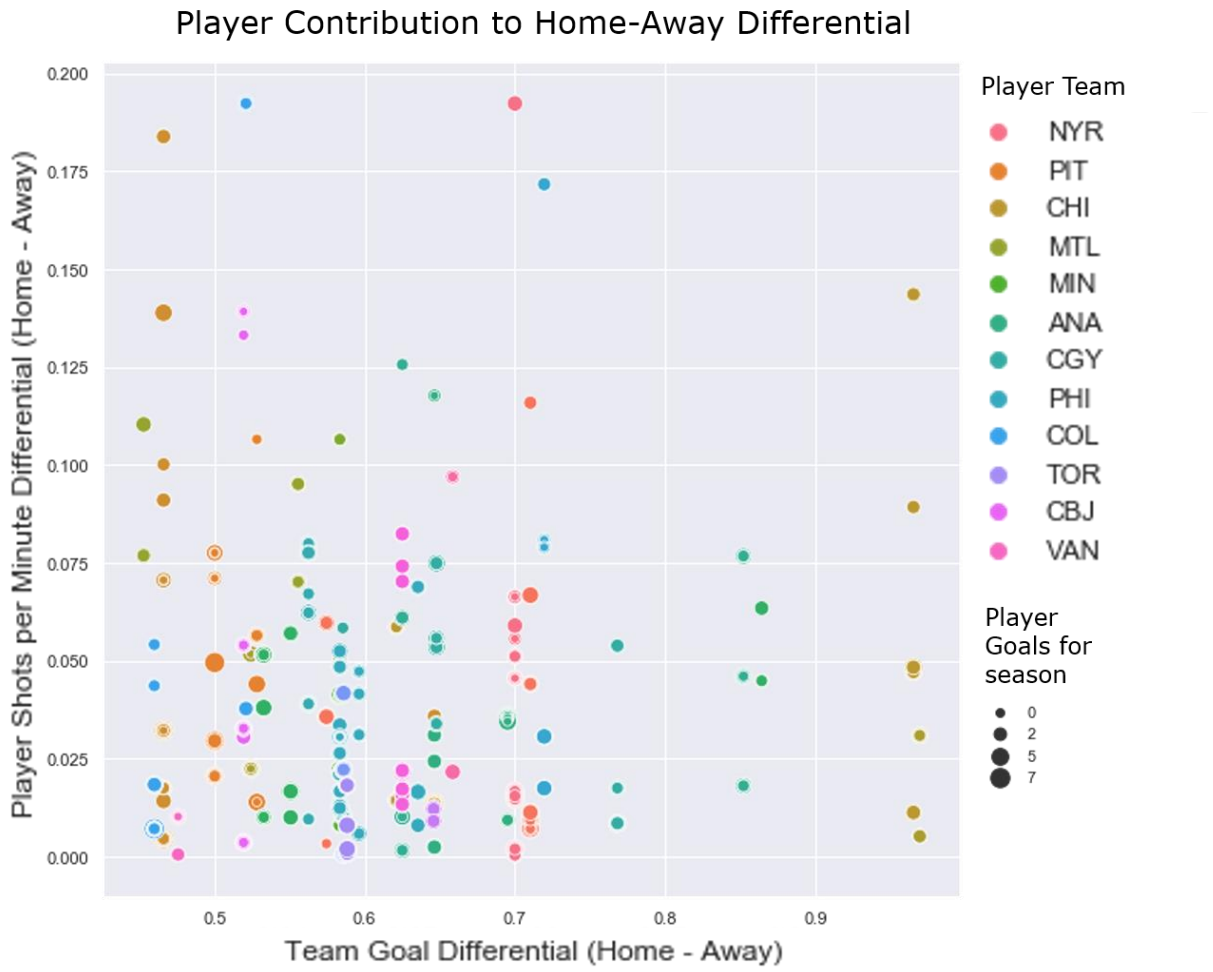

Figure 15: Top Player Shot Frequency Differential for Teams with Top Goal Differentials

Conclusion

Home-ice Advantage persists in the NHL, as it does is many sports, but the variables surrounding the advantage remain complicated. While the advantage varies much more wildly during the playoffs, in general, teams generally score 0.3 more goals when playing at home than away, which results in a win-rate 5% higher than expected. Some teams maintain a stronger advantage as a franchise, however, this varies from season to season. Some of this advantage might be ascribed to the distance the away team travels; for some teams, the home ice advantage is stronger in offense than on defense. But the highest home team differential goes to the uniquely situated Vegas Golden Knights, whose city of residence offers novel distractions to visiting teams.

Home ice advantage as measured by goals also has a relationship with certain variables: teams at home typically have above average rates of mean shots taken, mean Power Play opportunities, mean takeaways, mean faceoff wins, mean giveaways, and mean hits. Depending, we suppose, on whether

the home team is winning, there are correlations between home goals and visiting team goals (for when home teams are working to catch up), and home goals and takeaways (when the home team is ahead and allows their main players to rest).

Our research has shown that the following factors do not seem to be a strong indicator of home ice advantage: overall success of a team or exceptional performances of individual players at home. Among the top 50 players in the league, their performance home and away was not as significant as the difference in the whole team's performance differential. Likewise, when looking at individual player contributions to offense (both home and away), the individual responsiveness to the home ice advantage becomes less and less significant.

In conclusion, while we can expect superior performance from the home team, the underlying causes are manifold and not easily distilled into any single metric.