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

W200 – Project 2

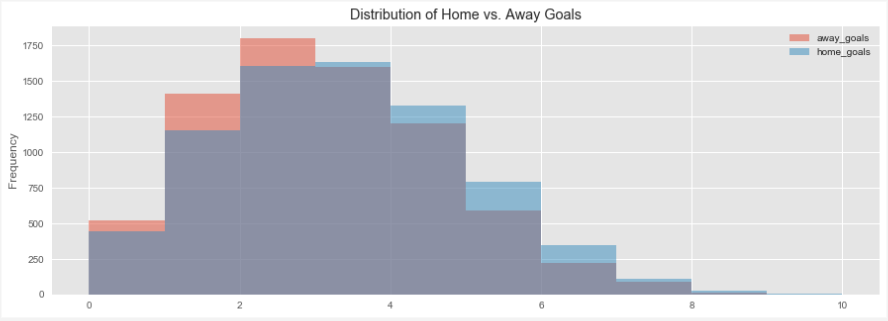
December 2018

Robert Bergan, Brent Biseda, Michael Hardy

**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. Our project seeks 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 a data science perspective.

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



**Focusing Questions**

Our focusing questions divide our report into two sections: a high-level look at the home-ice advantage and a deeper look at the drivers and manifestations of home-ice advantage.

* 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 or venue?

**Dataset**

The dataset we used in our analysis was sourced from Kaggle, but it was originally source is directly from the league’s website APIs. The data includes all of the data that the league collects 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 we used are:

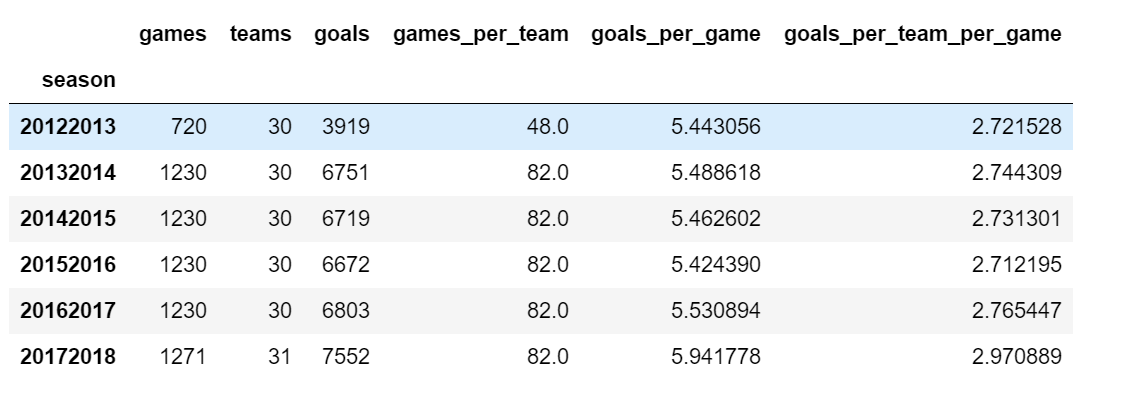
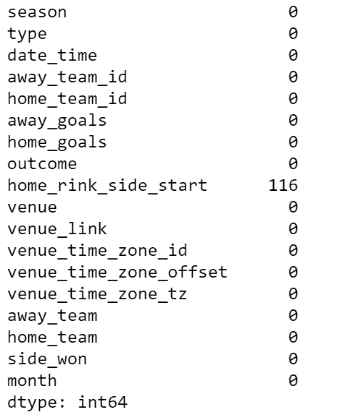
* ‘game’ table:
  + Summary data on each regular season and playoff game including date/time, season, teams involved (home and away), outcome, type (playoff vs. regular season), venue, goals scored
  + 7,441 rows, 16 columns
* ‘game\_teams\_stats’ table:
  + Additional detail on each game. Includes 1 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 our dataset is bounded the number of games per season and the associated statistics with those games we could understand and validate our data by cross-checking our aggregated data against the known bounds. In doing so we could check to see if all of the games and teams are included in the dataset for the regular season and playoffs and then also check some of the detailed statistics for reasonableness on a per game or per season basis. This analysis, as shown in Figure 1, leads us to conclude that there were 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 data points.

Figure 2: Missing Values

Figure 3: Example Validation Table



Additionally, we looked for missing values across both of our 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 our 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 and in the regular season vs. playoffs?**

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.

Trends by Time (Across Seasons and Within Seasons)

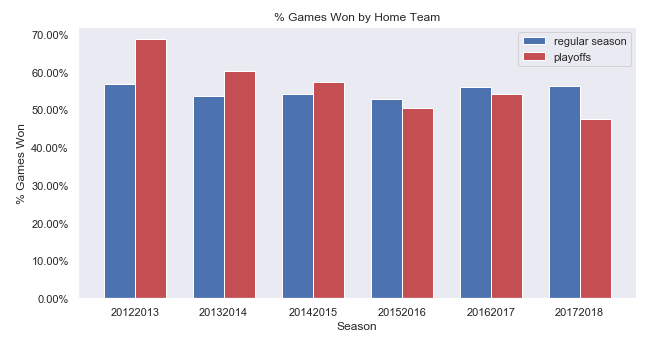
In all of the seasons in our data the regular season percentage has hovered in a range between 52.9% (2015-2016) and 56.81% (2012-2013, the lockout season) while the playoff percentage has fluctuated more dramatically between 68.6% in 2012-2013 and 47.6% in 2017-2018 (see Figure 4). Though it is hard to tell how substantial regular season difference is without doing more statistical analysis, we 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 4: 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.

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 figures 5 and 6, there is substantial variation among teams at home compared to on the road both in the number of goals scored and win-rate. There are two interesting insights from these charts: first, it appears that 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 good throughout the seasons, such as the Blues that have a low differential while other good teams, such as the Penguins have a high differential; second, the order of the teams by home vs. away goal differential (Figure 5) does not always align to the order by win differential (Figure 6). 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. 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.

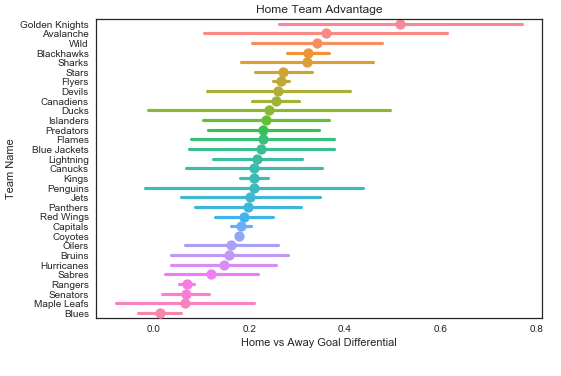
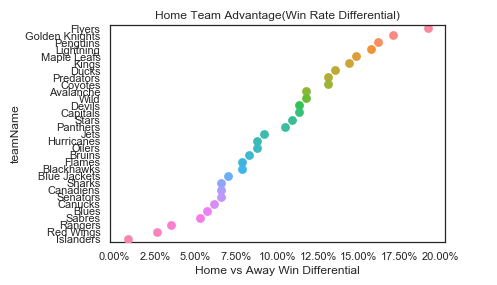


Figure 5: Home vs. Away Goal Differential

Figure 6: Home vs. Away Win Differential

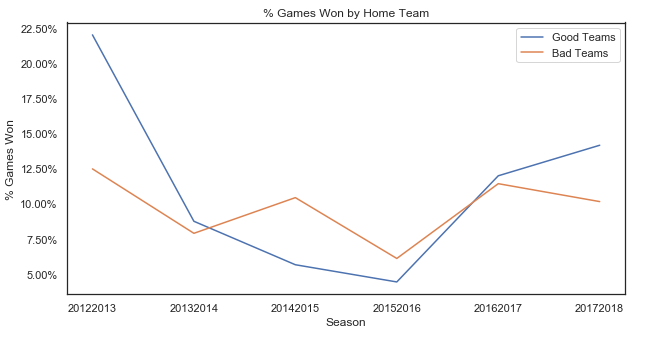
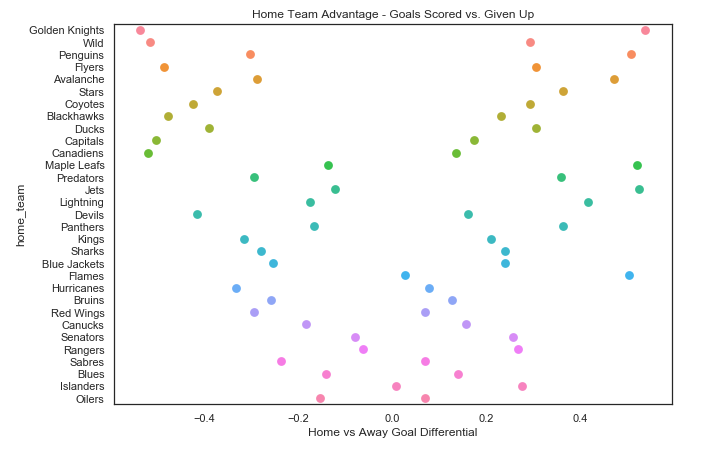


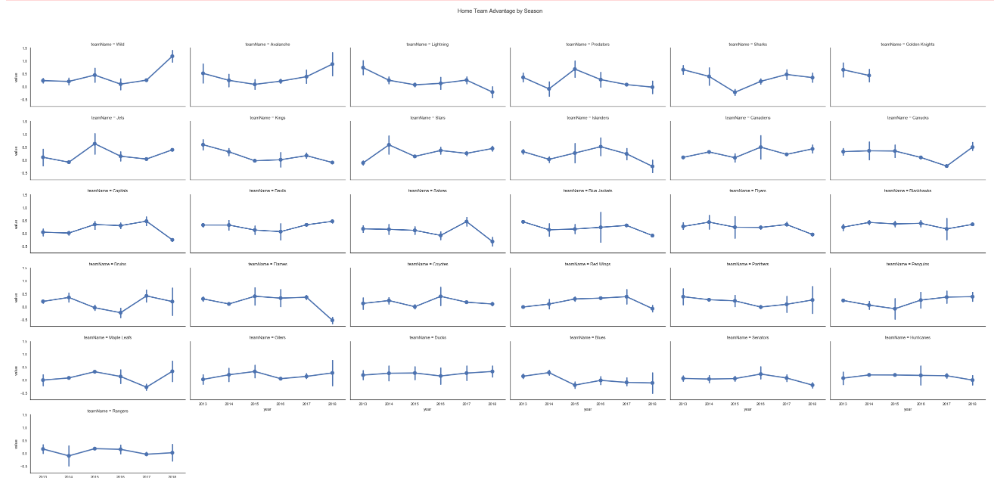
Figure 7: Win Differential by Good vs. Bad Teams

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 7 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).

Second, we explore the advantage by team by looking at whether there is more a difference in offense or defense in 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 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 8 below shows that there can be substantial difference in where the advantage comes from between teams. For instance, the Maple Leafs’ advantage seems to stem from offense as the team scores more goals at home relative to the number of goals it gives up. The Canadians advantage seems to stem from defense as it gives up fewer goals at home relative to the increased number of goals it scores at home. [Vegas flu]

Figure 8: Goals Scored vs. Given Up, Home vs Away





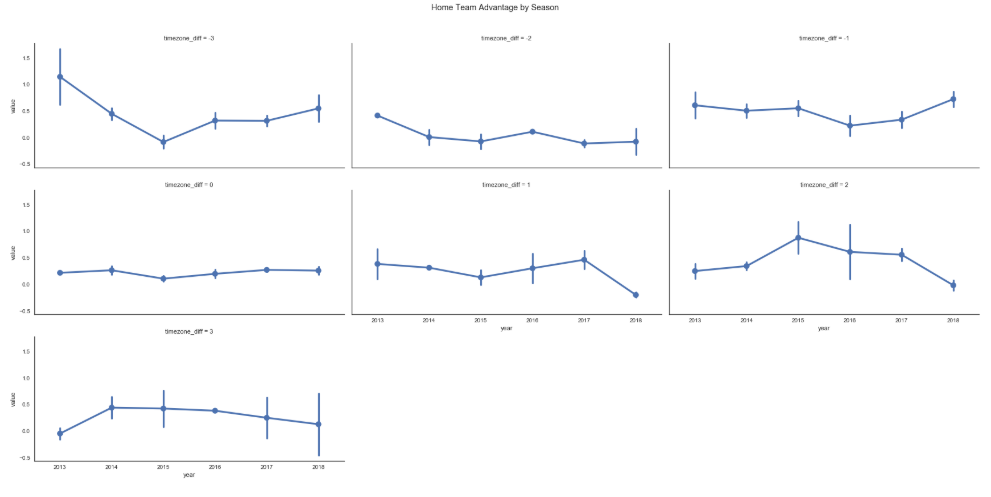
Based on our analysis in section 1 we can conclude that there is a persistent home ice advantage of approximately 10% (i.e. 55% home team wins, 45% away team wins), but that there don’t seem to be any specific trends related to the change in the advantage either between seasons or within seasons. There 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 look into more detailed game statistics to try to examine whether we can find any factors that affect 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?**

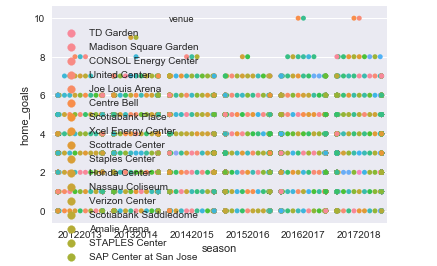
Time Zone

Describe time zone and charts

Interesting – doesn’t seem to be much of a difference other than in 2012-2013 (lockout) season?

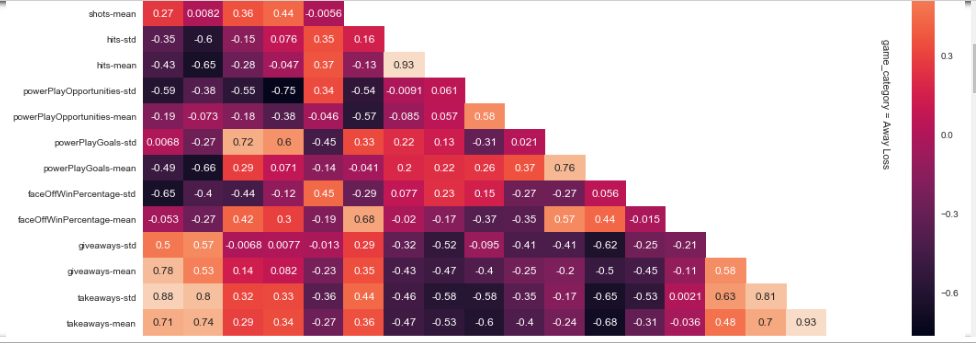


By Venue?



Other Variables

Is there a way we can summarize this in a smaller chart? Somehow do correlations with home vs. away win rate or goal differential?



* Deeper dives (i.e. 1-2 charts each) for:
  + Faceoffs?
  + Powerplays?
  + Hits?
  + Shots?

Player level summary (i.e. do top 50 players score more goals at home vs. away?)