Hockey is on Ice: Stochastic Simulation to the Rescue!

Andrew Rippy

May 2020

1 The Stanley Cup

There are currently 31 Hockey teams across two conferences, split into a total of 4 divisions. The Eastern Conference is split into the Atlantic Division and the Metropolitan Division, and the Western Conference is split into the Central Division and the Pacific Division. Each Division has its own sub-rankings, and each conference has two wildcards. A wildcard is a team within a conference with the best record that did not finish top 3 in their division. For the Stanley Cup playoffs, the brackets are organized in this manner. Each match-up is the best of 7 games. For the Division Semi-finals; 1st vs. Wildcard and 2nd vs. 3rd. The winners of these two match-ups play against each each other in the Division Finals, resulting in a division winner. The two division winners per conference play against each other in the Conference Finals, resulting in a conference winner. Finally, the two conference winners go head to head in the Finals for the glory of winning the Stanley Cup. For a visual representation, see Fig. 1. Due to the COVID-19 outbreak, the season was cut short, by about 11 games per team. If the teams went into the playoffs right when COVID-19 hit, these would have been the match-ups.

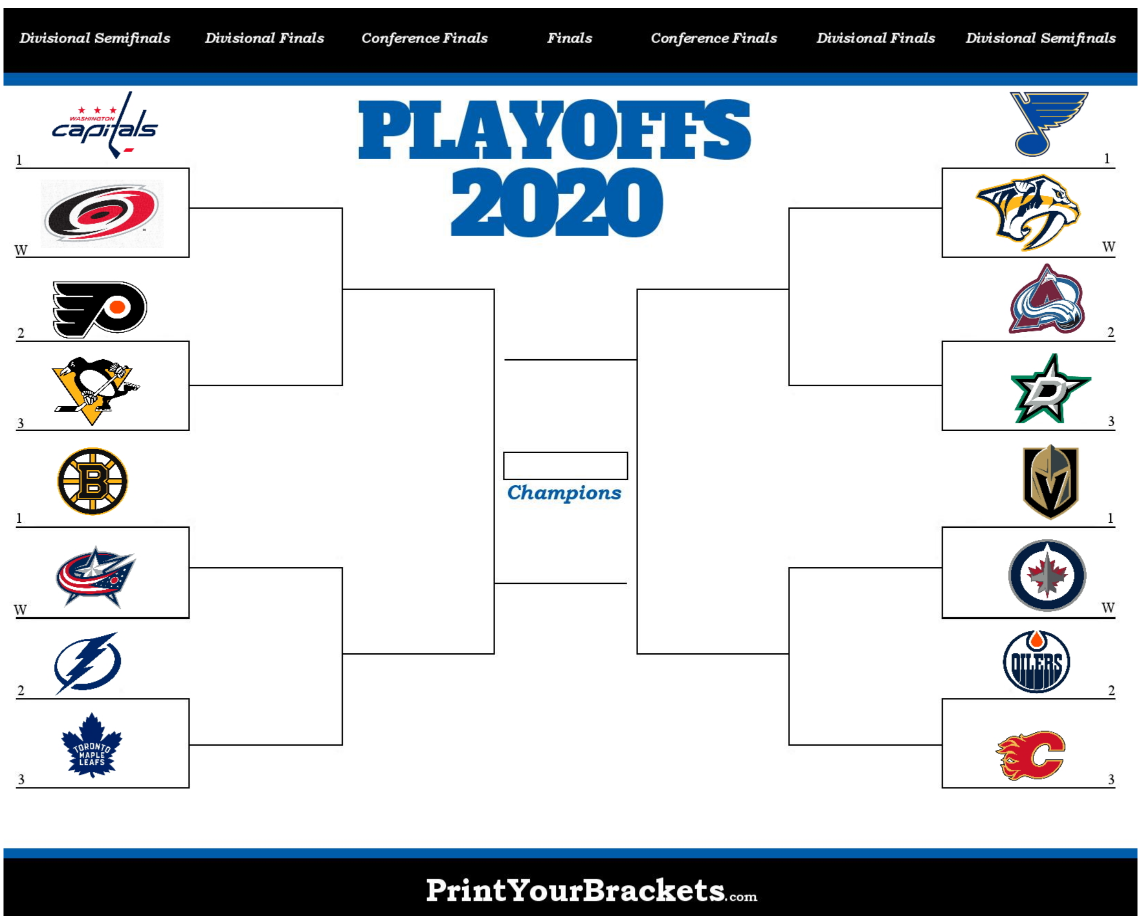

Figure 1: The Stanley Cup Playoffs Bracket

2 Models and Statistics

In order to determine who would have won these Stanley Cup playoffs, we can employ Stochastic Simulation to predict the outcome. All data and statistics for the models are provided by the official NHL stats website.

2.1 Model 1

For each team, we have "Total Goals For" (G_F), which are the total goals they scored for their team, "Total Goals Against" (G_A), which are the total goals that were scored by other teams against them, and then "Average Goals For per game" and "Average Goals Against per game", g_F and g_A respectively. The PythagenPuck[1] probability statistic is a formula developed for hockey, derived from baseball roots. It is an empirical model that is fit to a large aggregate of statistics. The equation gives the individual probability a team will win like so:

$$Pr(\text{Win}) = \frac{G_f^E}{G_f^E + G_A^E} \tag{1}$$

Where $E = (g_F + g_A)^P$, and P is an empirically derived constant, 0.458. From this equation, and the stats provided on the official NHL website, we can calculate the individual probability to win for each team. The model then employs a "dice roll" simulation. A uniform random variable is used in conjunction with the calculated win probability to "roll" each team. If Team one rolls a win, and Team two rolls a lose, team one wins. If both teams roll a win, or both teams roll a loss, the simulated game goes into overtime. The statistics change in overtime, as both teams are tired, and it becomes sudden death. A different set of overtime win probabilities is used for each team's probability to win in overtime. These are based on their overtime Win-Loss performance during the normal season. The dice are then rolled continuously until a tie is broken, and one team wins. Up to 7 games are run per match-up, until one team has won 4 games in total. This team moves on and repeats the process until a cup winner is determined. The outcome for this model is seen in Fig. 2.

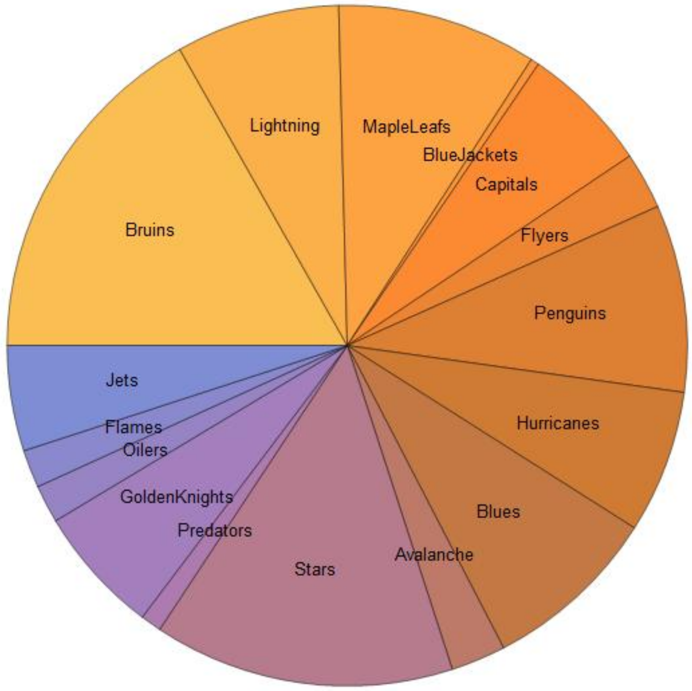

Figure 2: The Win Percentages for each eligible team

The win percentages for the top 3 teams are as follows:

$$\text{Bruins: } 16.78\% \pm 0.68\% \text{ (95\%CI)}$$
$$\text{Stars: } 14.32\% \pm 0.53\% \text{ (95\%CI)}$$
$$\text{Penguins: } 9.7\% \pm 0.43\% \text{ (95\%CI)}$$

The confidence intervals were determined by running the simulation 20 times, with each simulation being 1000 playoffs.

Figure 3: The Win Percentages for each eligible team

2.2 Model 2

For this model, the stats are gathered based on specific team match-ups. Each team has a different win-loss ratio against different division teams, which can allow us to weight the probabilities based on different match-ups. For example, a team could have a 70% win rate against Metropolitan Division teams, 60% against Central, 80% against Atlantic, and 90% against Pacific. When the team is matched against other teams, the team win-rate would change based on the opponent's division. The overtime stats remain general. The same "dice roll" simulation in Model 1 is used, with a tie resulting in using the overtime probabilities to find a winner. The outcome for this model is seen in Fig. 3. The win percentages for the top 3 teams are as follows:

$$\begin{aligned} \text{Stars: } & 17.1\% \pm 0.65\% \text{ (95\%CI)} \\ \text{Bruins: } & 14.7\% \pm 0.33\% \text{ (95\%CI)} \\ \text{Penguins: } & 11.2\% \pm 0.41\% \text{ (95\%CI)} \end{aligned}$$

The confidence intervals were determined by running the simulation 20 times, with each simulation being 1000 playoffs.

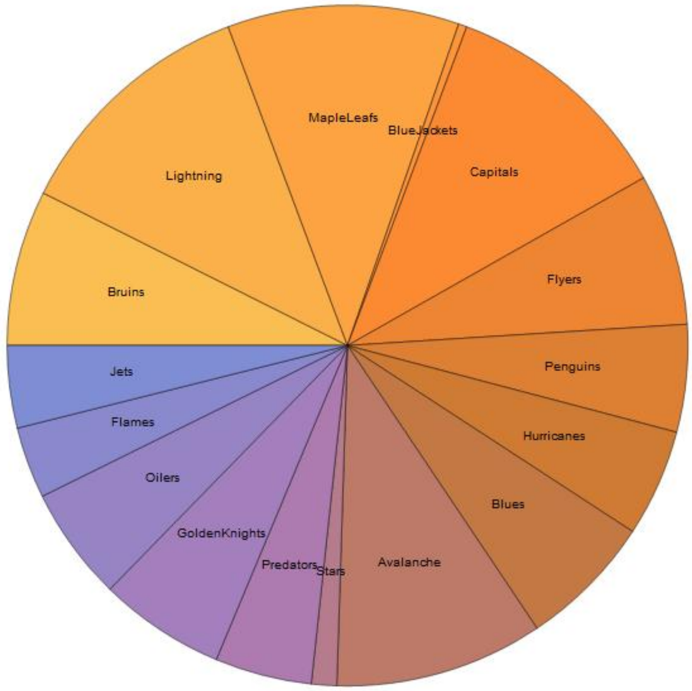

Figure 4: The Win Percentages for each eligible team

2.3 Model 3

For this model, goal scoring is modeled as a Poisson process. Over the course of 1 full game, each team has an expected number of goals they will make, and time between goals is modeled by a Poisson random variable determined by their expected goals. If by the end of the game, both teams are tied, their expected goal rate changes based on their overtime goals stats, and still remains a Poisson process. For this case, since only 1 goal is ever made in overtime, it becomes the number of overtime games won divided by the total overtime games played. Up to 7 games are played per final, and the winning team moves on, until a winner is determined in the cup finals. The outcome for this model is seen in Fig. 4.

The win percentages for the top 3 teams are as follows:

$$\begin{aligned} \text{Lightning: } & 12.5\% \pm 0.66\% \text{ (95\%CI)} \\ \text{Capitals: } & 11.3\% \pm 0.41\% \text{ (95\%CI)} \\ \text{Maple Leafs: } & 10.2\% \pm 0.34\% \text{ (95\%CI)} \end{aligned}$$

The confidence intervals were determined by running the simulation 20 times, with each simulation being 1000 playoffs.

3 Results

Models 1 and 2 have a clear top 3 teams they favor. However, since the season wasn't entirely completed, the stats may be skewed for or against specific teams, not taking into account possible 'bad games' that may have further balanced the stats. Model 3 sees a much more even split of possible winners. Many fall in the 8-10% range. Based on Models 1 and 2, I expect the Bruins, Stars, or Penguins to win. Based on Model 3, I expect the Lightning, Capitals, or Maple Leafs to win. However, the Bruins and Penguins still have a sizeable chance in Model 3, whereas the Stars have a much smaller chance to win. Since Model 3 was so spread out as far as possible winners, by comparing all 3 models together, the most likely would-be winner of the Stanley cup is the Bruins, since they were the most likely, consistent prediction between all 3 models.

References

[1] Alan Ryder. *Win Probabilites*, 2004, Click Here

[2] Official NHL Stats. Click Here