SSC 442 Final Project

Group 17 Bushong's Boys

Jacob Loukota, Michael St. Jean, Matthew McGavin, Brendan Collier

For our project we will be analyzing NHL data to find relationships between players performance, team performance, and player salaries. We will put ourselves in the shoes of the Vancouver Canuck's GM in the 2021 signing of rookie Quinn Hughes. Hughes is currently in a breakout rookie season and is looking at substantial payout after his entry level contract expires. We want to into look how much Hughes is worth to the organization and whether this paycheck is worth it.

We will start off by building a linear regression model predicting player salary based on a wide variety of statistics gathered in the 2016/2017 season and use it to predict a fair 2021 salary offer for Hughes. Afterwords we will investigate a variety of the questions listed below such as the success of teams with high paid players. We will also investigate and attempt to find interesting relationships that could factor into decisions within organizations.

The data will be sourced from https://www.kaggle.com/camnugent/predict-nhl-player-salaries#train.csv and https://www.kaggle.com/martinellis/nhl-game-data#team_info.csv. The first data set is nicely formatted and gives us a ton of information about around 1000 players. We will attempt to pull in some information about the overall performance of the teams from the more complex second dataset.

The data should allow us to answer a lot of interesting questions but we will start by answering the following:

1. What should a player be paid based on a variety of statistics.
2. Is it better to have a couple star players or a more even roster?
3. Which teams should sign Quinn Hughes?
4. Which players are under/overperforming based on their salaries?
5. Are certain positions over/underpaid on average?

Overall, this project will be an in-depth analysis into a fairly complex dataset, and we hope to mimic and better understand the work of data scientists already working in athletic organizations.