Final Project\_MA 611

Predicting Player & Team Performance

Author

A group of baseball players on a field

Description automatically generated

# Introduction

The National Hockey League (NHL) is a professional ice hockey league in North America, currently comprising 31 teams: 24 in the United States and 7 in Canada. The NHL is considered to be the premier professional ice hockey league in the world, and one of the major professional sports leagues in United States and Canada. At its inception, the NHL had four teams—all in Canada, thus the adjective "National" in the league's name. The league expanded to the United States in 1924, when the Boston Bruins joined, and has since consisted of American and Canadian teams. From 1942 to 1967, the league had only six teams, collectively nicknamed the "Original Six". The NHL added six new teams to double its size at the 1967 NHL expansion. The league then increased to 18 teams by 1974 and 21 teams in 1979. Between 1991 and 2000, the NHL further expanded to 30 teams. It added its 31st team in 2017 and has approved the addition of a 32nd team in 2021.

## Abstract

The simplest sports bet one can make is guessing which team would win. Now each team playing in an NHL game can have one of the three possible outcomes: win, loss, or overtime loss. This makes prediction of how a team will perform and especially how the players will perform on the basis of certain trends a very lucrative objective for our project. In this project, we want to predict the performance of major NHL teams and players in the future. These are the two broad questions that we want to address through this project. In order to achieve this, we are considering two datasets.

The first dataset has information about the historical trend of all the 31 teams and details regarding the cap hit for the season , the different category of players selected (type of contract), players contract beginning and end date and also the percentage points scored by the team in a season (%PTS). We intend to use the data of past five years as validation data. We want to analyze the underlying trends in performance of teams and select the best model to predict the performance of the team for the next season. We will focus on describing a team’s in-season trend by using heat maps and polar maps. While utilizing the dataset we would identify the outliers and modify them. We would also compare the predictive accuracy of different models.

The second dataset will be used to address our second objective to predict a player’s performance. However, we intend to address a very interesting comparison of two teams “Boston Bruins” and “Toronto Maple Leafs”. We intend the analyze the relationships between a team’s performance and a player’s contract cap hit. We will also compare players between two teams in the similar playing positions (Centre, Left Wing, Right Wing, Defense, Goalie) and also the points scored by and against the team while the player is on the rink. The goal would be to compare different models and evaluate the model’s predictive accuracy. The business implication that we would like to address in this context, is to understand the difference in underlying strategy of the two teams as well as finding a critical annual salary for a player that balances risk taken by the team for a player underperforming and a player’s risk for overperforming.

In the end we would discuss the various limitations and areas of improvement in our research and the possible implications of our research.