

## 6.1 Data Profile

**Primary data summary:** The NHL (National Hockey League) is a long time sport with an abundance of stats that are in the beginning stages of modeling statistics. "It has been suggested that the best models can only predict the winner 62% of the time due to variances in talent and "puck luck" ([Kaggle inspiration](#)).

**Collection:** Data in the Kaggle site came from the NHL website.

**Datasets:** Overall there were 13 datasets with a photo of the relationship between tables

game.csv  
game\_goalie\_stats.csv  
game\_goals.csv  
game\_officials.csv  
game\_penalties.csv  
game\_plays.csv  
game\_plays\_players.csv  
game\_scratches.csv  
game\_shifts.csv  
game\_skater\_stats.csv  
game\_teams\_stats.csv  
player\_info.csv  
table\_relationships.JPG  
team\_info.csv

**Limitations:** Because there is an abundance of documents in the data there is a need to merge a good deal of them together to better understand what is possible with the data at hand. Hockey is also a very fast paced game so some of the data will come with a slight best guessing as hockey pucks travel at 80ish mph. This results in an average of 0.59 seconds for a puck to travel from the shooter to the goalie with a possible of the puck bouncing between players.

**Ethics:** There is no private data being used from this data so there is no ethics about private info becoming public. Ethics with this data will come from not allowing bias to occur from my perspective as a fan of the sport prior to working with the data.

## Data Cleaning

**Table:** game.csv

Column	Description	Issue	Action
Game id	ID of game	None	None
Season	Year the game was	None	None
Type	Regular or Playoff game	Not needed	Removed

<b>Date time</b>	Day and time of game	None	None
<b>Away team ID</b>	ID of away team	None	None
<b>Home team id</b>	ID of home team	None	None
<b>Away goals</b>	Goals of away team	None	None
<b>Home goals</b>	Goals of home team	None	None
<b>Outcome</b>	Which team won	None	None
<b>Home rink side start</b>	Which side the home team started on	Irrelevant	Removed
<b>Venue</b>	Stadium played at	None	None
<b>Venue link</b>	Link to stadium	Irrelevant	removed
<b>Venue time zone id</b>	Time zone of stadium		
<b>Venue time zone offset</b>	Stadium time zone in relation to UTC	Irrelevant	Removed
<b>Venue time zone tz</b>	Abrv of time zone	None	None

**Table:** Game team stats

Column	Description	Issue	Action
Game id	ID of game	None	None
Team id	ID of team	None	None
HoA	Home or Away	Renamed	Changed to Home or Away
Won	Who won	None	None
Settled in	Regulation or Overtime play	None	None
Head coach	Who is the head coach	Mixed data	Changed to str
Goals	Team goals	Needed renaming	Changed to team goals
Shots	Team shots	Needed renaming	Changed to team shots
Hits	Team hits	Needed renaming	Changed to team hits
Pim	Penalty in minutes	Needed renaming	Changed to team pim
Powerplayopportunities	Chances to score on powerplay	None	None
Powerplaygoals	Goals while on powerplay	Needed renaming	Changed to team powerplaygoals
Faceoffwinpercentage	Percent of faceoffs won	None	None
Giveaways	Times gave puck to other team	Needed renaming	Changed to team giveaways

Takeaways	Time team stole puck from other team	Needed renaming	Changed to team takeaways
Blocked	Times team blocked shots	Needed renaming	Changed to team blocked
startRinkSide	Side started on	Irrelevant	Removed

**Table:** Game skater stats

Column	Description	Issue	Action
Game id	ID of game	None	None
Player id	ID of player	None	None
Team id	ID of team	None	None
Timeonice	Time player skated during game	None	None
Assists	Times players assisted with goal	None	None
Goals	Times player scored	None	None
Shots	Times player shot on goal	None	None
Hits	Times player checked another player	Missing value	Added mean
Powerplaygoals	Times player scored while on powerplay	None	None
Powerplayassist	Times player assisted while on powerplay	None	None
Penaltyminutes	Number of minutes give in penalties	None	None
Faceoffwins	Number of faceoffs won	None	None
Faceofftaken	Number of faceoffs taken	None	None
Takeaways	Number of times taken puck from another player	Missing values	Added mean
Giveaways	Number of times player gave up puck to other team	Missing values	Added mean
Shorthandedassists	Numbers of assists while shorthanded	None	None
Shorthandedgoals	Number of goals while shorthanded	None	None
Blocked	Number of shots blocked	Missing values	Added mean

Plusminus	Goals for over goals against while on the ice	None	None
Eventimeonice	Time spend on ice while at even strength	None	None
Shorthandedtimeonice	Time spent on ice while killing a powerplay	None	None
Powerplaytimeonice	Time spent one ice while on the powerplay	None	None

**Table:** Game goalie stats

Column	Description	Issue	Action
Game id	ID of game	None	None
Player id	ID of player	None	None
Team id	ID of team	None	None
Timeonice	Time spent on ice	None	None
Assists	Times assisted a goal	None	None
Goals	Number of goals	None	None
Pim	Penalty in minutes	None	None
Shots	Number of shots on goal	None	None
Saves	Number of shots saved	None	None
powerPlaySaves	Number of saves while on powerplay	None	None
shortHandedSaves	Number of saves while short handed	None	None
evenSaves	Number of saves while even strength	None	None
shortHandedShotsAgainst	Number of shots attempted against while shorthanded	None	None
evenShotsAgainst	Number of shots attempted against while even strength	None	None
powerPlayShotsAgainst	Number of shots attempted against while on the powerplay	None	None

Decision	Result of the game	Missing values	Nothing as missing values were for goalies who didn't finish game
Savepercentage	Save percentage overall	Missing value	Removed
powerPlaySavePercentage	Save percentage while shorthanded	Missing values	No penalty against team so results are 0/0
evenStrengthSavePercentage	Save percentage while at even strength	Missing values	Removed

**Table:** Player info

Column	Description	Issue	Action
Player id	ID of player	None	None
Firstname	First name of player	None	None
Lastname	Last name of player	None	None
Nationality	Nation player is from	Mixed	Changed to str
Birthcity	City player is born in	Mixed	Changed to str
Primaryposition	Position player plays	None	None
Birthdate	Date of player birth	None	None
birthStateProvince	State player was born	Irrelevant	Only was applicable to North America players
Height	Height in inches	None	None
Height cm	Hight in centimeters	None	None
Weight	Weight of player	None	None
Shotcatches	Side player shoots from	Changed name	Changed to side shot from

**Table:** Team info

Column	Description	Issue	Action
Team id	ID of team	None	None
Franchiseid	ID of Franchise	None	None
Shortname	Name of city	None	None
Teamname	Name of team	None	None
Abbreviation	Abbreviation of team	None	None
Link	Link to teams	Irrelevant	Removed

Tables merged: Tables game team stats/team info/game doc/goalie stats/ player info merged to make game\_goalie\_merged doc.

Tables game team stats/team info/game doc/ player info/ game skater stats merged to make game\_team\_stats\_merged\_all.

**Question:**

Does the head coach matter to a team's success?

Does the team with the most blocks win more?

Does the team with more penalties lose more?

Does the side a player shoot from result in more goals?

Does the team who shoots more win more often?