DATA 603 Project Checkpoint

# Question 1: Group members

1. Mark Dodd
2. Raymond Wong
3. Dustin Tang
4. Michael Ellsworth

# Question 2: Research Topic

**If you have a research topic in mind for this project, please briefly describe the topic and any background info I may need to understand the topic.**

Our group’s research topic is National Hockey League (NHL) statistics and what statistics influence the outcomes of games. The intention of this project is to evaluate a number of hockey statistics using Logistic Regression to build prediction models that can accurately predict the winning team prior to the game being played.

The majority of the statistics that will be used for this prediction model are related to shots. For the purposes of this research, a shot is defined as a player directing the puck at the net with the intention of scoring. Shot related statistics are grouped based on shot location and whether it was on net, blocked or missed.

# Question 3: Data

**Please briefly describe the data you have (or plan to acquire) to help answer the research topic above. Include: what type of variable or variables are included (quantitative, qualitative, etc.), how the variable or variables are measured (the measurement scale), and any other general info you may have on the variable(s).**

The data that we will be using is structured event data from naturalstattrick.com. Natural Stat Trick has logged shot related data for every NHL game in some form since 2007. For the sake of this project, we will be using data from 2016, 2017 and 2018.

The data that will be used is all quantitative. Some statistics will be counts while others will be in the form of a percentage of total. For example, after completion of the game, if team A had 31 shots while team B had 20 shots, the shots for percentage, or SF%, for team A would be 62% while 38% for team B. The quantitative data that we use is collected as a % as to normalize the datasets from game to game.

The majority, if not all of the variables that will be evaluated will be related to shots.

# Question 4: Data permission

**Is this your own data set (or the data of someone in the group) or is it “open”or “shared” data?**

The data acquired from Natural Stat Trick is open data. They only restrict use for commercial purposes. We have indicated to them that the use of this data will be for academic purposes only which removes any restrictions on its usage.

A copy of their message granting permission to use the data can be included in the final project report if needed.