**Extract, Transform, and Load (ETL) Project**

## Project Report:

In this group project, we downloaded two Comma Separated Values (CSV) files from the NHL Game Data on Kaggle. Specifically, we looked at one file with the record of each NHL game in the dataset. The other file looked at the team information, such as names. The purpose of this project was to see how each NHL team performed during the playoffs from 2012 to 2017 using the ETL process.

Sources of data: game.csv and team\_info.csv

Link: <https://www.kaggle.com/martinellis/nhl-game-data>

**Extract and cleaning:** To start, we imported our dependencies (Pandas, Sqlalchemy, MySQLdb) into Jupyter Notebook and extracted raw data from the website. After we exported game data and team info data, we extracted these CSV files into two separate data frames. With proper data cleansing, we were able to combine columns that didn't need to be split and and filter through columns in each data frame to make sure we were only looking at playoff data.

**Transforming:** Next, we transformed the game data frame and team\_info data frame and merged them as one data frame. In addition to this, we used the information in the team\_info data to use team names instead of unknown team numbers. This allowed us to determine which team was the winner for each specific game. The final playoffs data frame contained the columns: game\_id, season, away\_team, home\_team, away\_goals, home\_goals, outcome, winning\_team.

**Load:** After these steps, the single data frame was uploaded to a MySQL relational database named “playoff\_db". Each row of data has all of the required information, so it made the most sense to use a relational database. The database features information including game id, season (year), away team name, home team name, away goals, home goals, outcome (win or loss), and winning team name.

The sql files attatched will allow you to recreate the table loaded into MySQL and run any queries you desire of the playoff data.

Some of our sample queries consisted of:

1. Playoff\_data table,
2. The winner of the NHL playoffs (2012-2013) season,
3. How many goals each team made in the 2012-2013 playoff season in descending order,
4. The winner of the NHL playoffs in 2016-2017 season,
5. How many goals each team made in the 2016-2017 playoff season in descending order.