**NHL Data: Linear Regression**

The dataset **nhl\_2223.csv** contains the final standings of the 2022/2023 NHL season, along with team level statistics. The dataset contains many variables, but this worksheet explores the relationship between various team level statistics such as goals for, goals against, and points.

If regulation time ends and the score is tied, the game goes into overtime where the next goal wins. If no goal is scored in overtime, the game goes to shootout, where players take breakaways against the other team’s goalie. A win in regulation is worth 2 points, a regulation loss is 0 points, and an overtime/shootout loss gets 1 point.

|  |  |
| --- | --- |
| **Variable** | **Description** |
| PPG | Points won per game in regulation time or overtime (# of non-shootout points / # of non-shootout games) |
| PTS | Total points earned (regulation, overtime, shootout) |
| GF | Total goals scored |
| GA | Total goals allowed |
| DIFF | Goals for (GF) – goals against (GA) |

1. Do you think that teams who score more goals (higher *GF*) win more games (higher PTS)?
2. Plot *PTS vs GF*. Does there appear to be a relationship between the two variables? If so, describe it.
3. Fit a linear model using *GF* to predict *PTS*.
   1. Write down the prediction equation.
   2. Interpret the slope in the context of what the equation is predicting.
   3. Is *GF* an effective predictor of *PTS*? What part of the output did you use to make this decision?
   4. Calculate the residual for the Toronto Maple Leafs (*Team* = TOR).
4. Fit a linear model to predict *PTS* using *GA.* 
   1. Write down the prediction equation*.*
   2. Based on the coefficients, describe the relationship between *PTS* and *GA*.
   3. How many points is a team the lets in 219 goals predicted to win?
5. Fit a linear model using *DIFF* to predict *PTS*.
   1. Write down the prediction equation.
   2. What is the R2 value? Interpret this in context.
6. Which of the three models is the most effective for predicting *PTS*? Explain your answer using evidence from your output.