

Analyzing and estimating annual National Hockey League players cap hit based on their yearly performances

Louis Popovic - Processing Big Data

Abstract—The goal of this research is to analyze the salary of NHL (National Hockey League) players and try to determine how much a player should be paid based on their on-ice performances.

I. INTRODUCTION

The NHL imposes a salary cap which limits the amount of money a team can spend on its players each year. The salary impact of a player on the team salary cap is defined by the average of his annual income throughout the duration of his contract. This is called a **Cap Hit**. For example, Shea Weber had a salary of \$8M last year but had a cap hit of \$7.75M. Due to these limitations and rules, contract negotiation is important and plays a big role in teams' success.

The purpose of this project is to determine if players cap hits are related to their yearly statistics, and if so, find the factors that influence their annual cap hit. This way, one can determine the expected salary of a player and find out if he is over, under or adequately paid based on his on-ice performances. For this research, only skaters (forwards and defence players) will be evaluated. In addition, as they both have different roles, they will have to be analyzed separately.

II. DATA

The data used in this project comes from two full seasons: 2016-17 and 2017-18. It is presented as an Excel file and contains all statistics offered by the NHL¹. 48 statistics per player were taken from these data sets. There are many different statistics such as the number of hits made, time spent on the ice, number of scoring chances, etc.

The data will have to be pre-processed to facilitate its analysis. The players with a cap hit equal to or lower than the rookie salary (\$0.925M), have been withdrawn since their salaries are not based on their performances but rather on a regulation of the league. Players who played fewer than 35 games were also removed (82 games in a season) because their total time on ice is too small to draw conclusions. There are also missing values that will need to be addressed later. The final Data Set contains 791 rows (players) and 51 columns (48 features, the Player's Name and a One Hot Encoding of their position).

III. GRAPHICAL ANALYZE AND DATA EXPLORATION

The Principal Component Analysis was applied on the full data set (forwards and defence players together) in order to find patterns. As it can be seen on figure 1, higher cap

hits tend to have a PC1 higher than smaller cap hits. If we separate the points according to the player's position, we obtain figure 2. Here a separation between the two positions can be seen: defence players tend to have a higher PC2 than forwards. This confirms the decision of evaluating the two positions separately. Another way to have a look at the data is to analyze the correlations between the players' statistics. This assessment will be based on the forward's statistics only. Defence players will be evaluated later. With the correlation map obtained from this analysis (See fig 3) we can observe two things :

- 1) There is a strong correlation between the TOI (*Time On Ice*) and other statistics. This is explained by the fact that the more a player is on the ice, the more likely he is to increase his other statistics. It could be interesting to convert the players' annual statistics into hourly statistics (Goals/60mins, Assists/60mins, etc.). This will be addressed later.
- 2) There are many features that are correlated with the Cap Hit. The strongest correlation is with the Team Corsi (textitCF), which represents all shot attempts made by the players' team while he was on the ice (shots on goal, shots that missed the net or hit the post and shots that were blocked by the opposition². By plotting this feature with the Cap Hit, we obtain figure 4. Another assumption would be that the more a player scores goals, the higher his cap hit should be. By plotting these two features we get fig 5. Again, we can see a certain trend between the number of goals scored and the cap hit of a player. It is not strong enough to draw any conclusions yet.

Based on these first observations, one can suppose that there is indeed some correlation between the statistics of a player and his cap hit.

IV. RESEARCH QUESTIONS

It is important to mention that the purpose of this research is not to predict the future salary of a player, but rather to estimate his yearly cap hit, strictly based on his statistics. Therefore, the objectives of this research are :

- 1) Determine if yearly statistics of players are good indicators of their cap hit,
- 2) if so, build a model that calculates the *Expected* cap hit of a player.

¹<http://www.hockeyabstract.com>

²<https://www.jewelsfromthecrown.com>

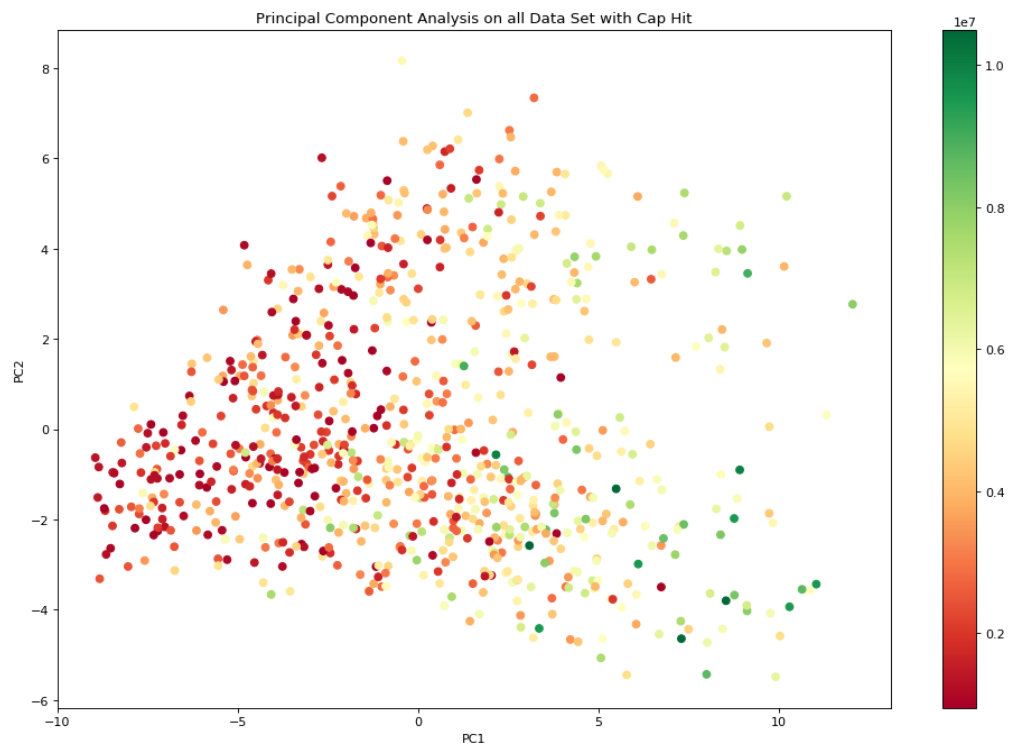


Fig. 1. Principal Components and players' Cap Hit

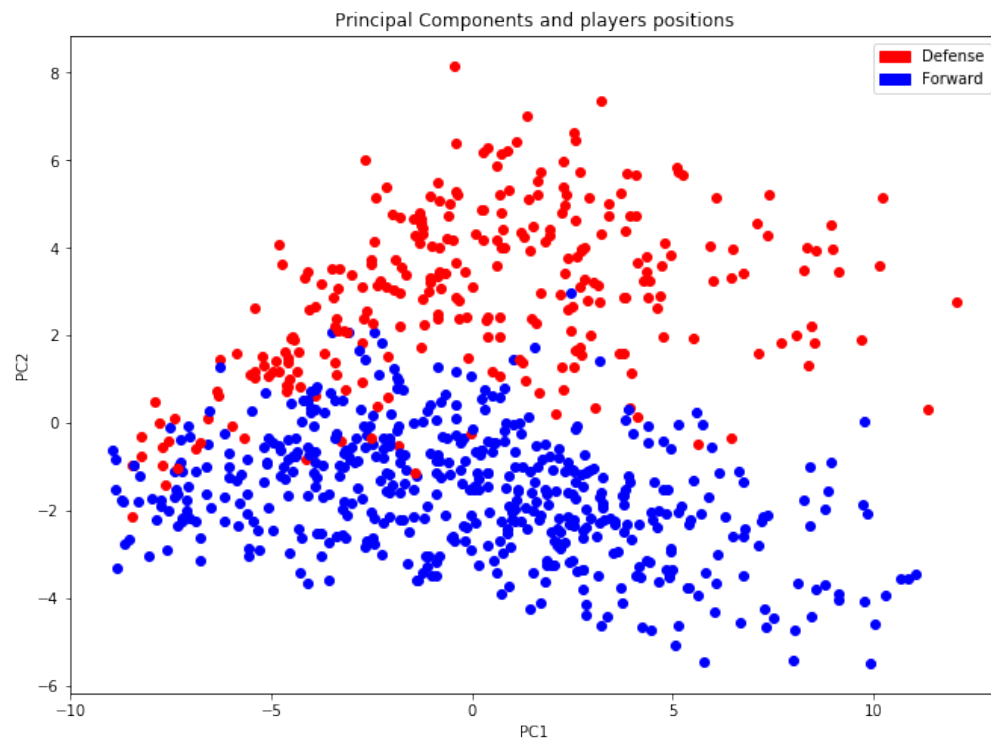


Fig. 2. Principal Components separated by players position

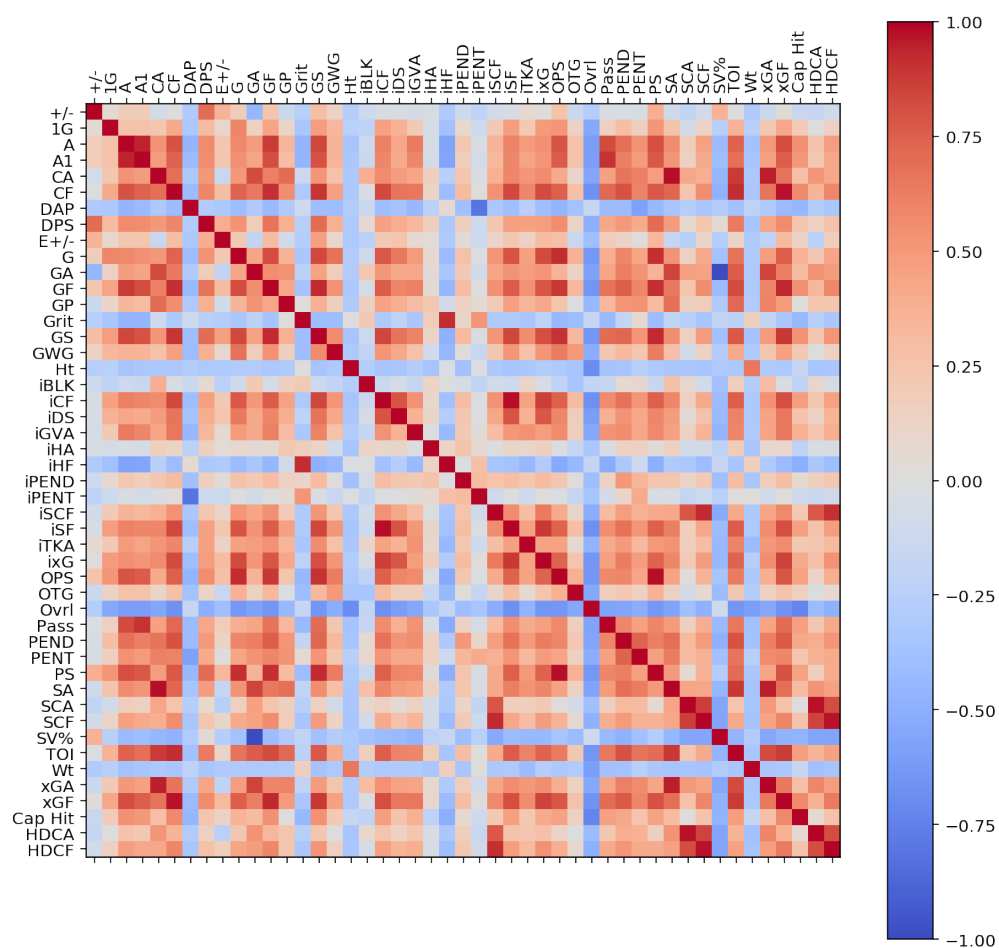


Fig. 3. Correlations between features (Forwards)

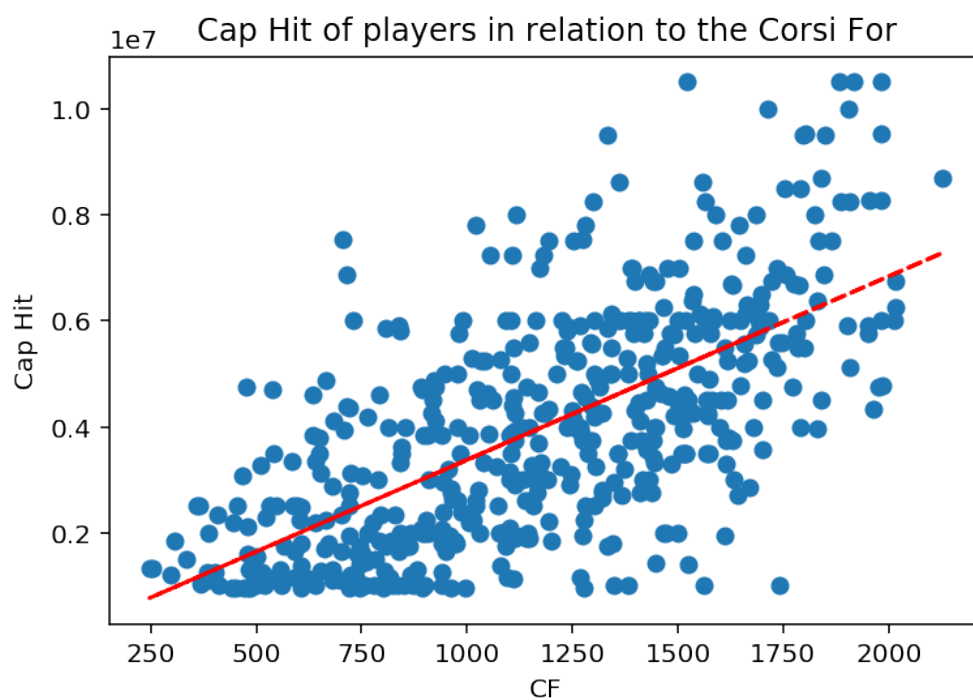


Fig. 4. Cap Hit and Corsi For (Forwards)

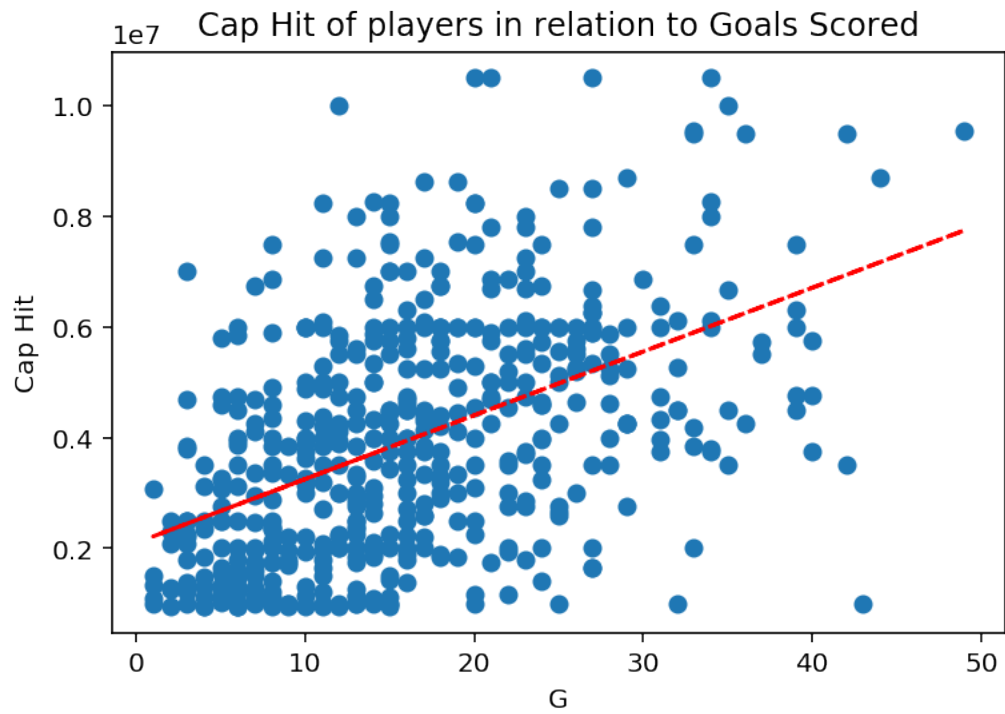


Fig. 5. Cap Hit and Goals Scored (Forwards)