Syracuse University

Memo

|  |  |
| --- | --- |
| To: | Dr. Landowski |
| From: | Martin Alonso |
| Date: | 2019-05-10 |
| Re: | Project Proposal |

**Topic: Analyze age effects on shots, goals, and Corsi on hockey players**

**Data Description:**

The main data set consists of 9 Comma-Separated value files that contain information for each National Hockey League (NHL) game between 2011 and 2018 for all active teams during the seven-year period.

|  |  |  |
| --- | --- | --- |
| Field | Description | Example |
| Play\_id | Play identification number | 2011030221\_1 |
| Game\_id | Game identification number | 2018021259 |
| Player\_id | Player identification number | 8477939 |
| Home\_team\_id | Home team identification number | 10 |
| Away\_team\_id | Away team identification number | 1 |
| timeOnIce | Total time player spent on the ice (in seconds) | 1058 |
| Goals | Total number of goals player scored | 1 |
| Shots | Total number of shots player had | 3 |
| Assists | Total number of assists player had | 0 |
| Saves | Total number of saves player had (Goalie only) | 0 |
| birthdate | Player date of birth | 5/1/1996 |
| firstName | Player first name | William |
| lastName | Player last name | Nylander |
| primaryPosition | Player’s primary position as listed on the team roster | C |
| teamName | Team official name | Toronto Maple Leafs |
| abbrev | Team official abbreviation | TOR |

Among the files, I also have data at the team level, play level, and identity of all other players on the ice during the shift the play was made, which would allow me to identify not only age effects of the player responsible for the shot, goal, or save, but also the ages and effect of the other players on the result of the given play.

I am also able to scrape data from NHL.com to try and get each player’s position on the ice at the time of the goal, shot, or save; though integrating this into the data might prove more time consuming as it would need to include where on the ice each player was at any given moment.

**Research Questions:**

* Does player age affect the number of goals, assists, or saves a player can make in any given season?
* How does Corsi (for Goalies) change with age?
* How are player shifts organized? Is there an effect of the combined player’s age for each shift on the number of goals and assists the shift has?
* Are statistics at the team level constant through time or is there a learning curve for younger teams as opposed to older teams?

**Data Preparation Plan**

1. Load the data into Python
2. Merge the datasets with the data regarding age, goals, assits, and saves; along with other players on the ice, teams that were playing, shift, and goalies.
3. Create a shift identification number based on the players in each team’s shift.
4. Explore the data for anomalies, missing data, adjusted duration, etc.
5. Depending on the amount of data missing, decide whether to replace data with average or median values for player at the season level or shift level.
6. Aggregate data by player-shfits-age, and model through season.
7. Group results by player/shift-age to notice any trends or effects.