Project Proposal

Proposal Details A proposal of your intended project is due by the above deadline. A proposal of your intended project should include the following.

1. The names of the students who will be contributing to the group project.

Jeffrey Yao, Martha Dixon, Nisha Sen

2. A tentative title for the project.

The Average Goals per Player per Season in the NHL (time period TBD)

3. Description of the data file (what they contain including number of variables and number of records). You do not necessarily have to list all the variables, but at least mention those of greatest importance.

- 28 variables
- 37, 825 observations total
- Some Key Variables: Age, Team (Tm), Position (Pos), Games Played (GP), Goals (G), Points (PTS)
- Response Variable of Interest: Goals (G)
- Categorical Explanatory Variable: Position (Pos)
- Continuous Numeric Explanatory Variables: Time on Ice (TOI), Penalties in Minutes (PIM), Age

4. Background information on the data sets, including specific citation of their source (so that I can also access it).

Data can be found here: http://inalitic.com/datasets/nhl%20player%20data.html

This dataset encompasses the NHL skater statistics of all players in the NHL from 1940 to 2018. While doing our analysis, we likely will not use the entirety of the dataset, and we will select a time period (for example, 2000-2018, or something similar).

5. A brief statement of the business, science, research, or personal interest you have in the data set which you hope to explore.

We are really big hockey fans, so we are interested in seeing which variables are most important in predicting the number of goals a player has in a season. It will be really interesting to see how the different variables affect the goals scored per player in recent years, and it will be interesting to see whether any trends are affected based on the time period that we choose to analyze during the course of this project.

6. Evidence that the data can be loaded into R. Load the data, and print the first few values of the response variable as evidence. Simply submit a single document (pdf, doc, html, . . .) outlining the proposal.

```
Console   Terminal ×   Jobs ×                                                              ▬▢

~/Desktop/ASRM 450 /Stats 420 Project/ ⇗

> head(skaters)
                       X Season              Player Age    Tm  Pos GP     G    GPG
1 2018Justin Abdelkader    2018 Justin Abdelkader   30  DET   LW  75    13  0.1733
2       2018Pontus Aberg    2018      Pontus Aberg   24  TOT   LW  53     4  0.0755
3       2018Noel Acciari    2018      Noel Acciari   26  BOS    C  60    10  0.1667
4    2018Kenny Agostino    2018    Kenny Agostino   25  BOS   LW   5   -   0.0000
5     2018Sebastian Aho    2018     Sebastian Aho   20  CAR   RW  78    29  0.3718
6     2018Sebastian Aho    2018     Sebastian Aho   21  NYI    D  22     1  0.0455
    A  PTS X...   PIM   EVG  PPG   SHG   GWG   EVA   PPA   SHA     S    S.
1  22   35  -11    78     9    4    -     -    17     5    -    110    12
2  12   16    9    10     4    -     -     3    11     1    -     70     6
3   1   11   -6     9     9    -     1    -     1    -    -     66    15
4   1    1   -1     4    -     -     -     -    -     1    -     11    -
5  36   65    4    24    21     8    -     4    28     8    -    200    15
6   3    4   -5     6    -     1    -     -     3    -    -     24     4
    TOI   ATOI BLK HIT FOwin FOloss   FO.
1  1,241  16:33  40 174    47     50 48.5
2    645  12:10   8  24     4      8 33.3
3    775  12:55  41 152    42     51 45.2
4     60  12:03   1   4     0      1  0.0
5  1,398  17:55  17  65    78     94 45.3
6    357  16:15  13  10     0      1  0.0
>
```