Created By: Ravi Gupta, Rohit Kulkarni

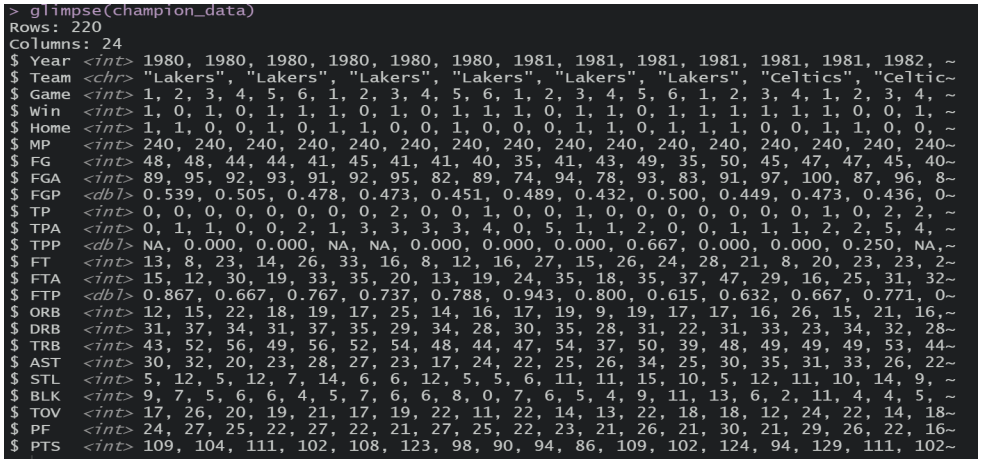
STAT108 Project Proposal

Background:

Every year, the National Basketball Association (NBA) ends the season with a series of championship games between the two best teams in the league. The series is a best-of-7, where the first team to win 4 games is the winner of the series. The championship is a very important accolade, both teams and players are compared by the number of championships they have won. In these comparisons, statistics like field goal percentage, offensive rebounds, steals, and blocks can be used to determine a team's performance and be an indicator of how much better one team is than another. Statistics in the NBA are highly analyzed, very often before a championship game to try to predict the outcome of the game. Our group wants to create a model that will be able to predict whether or not a team wins the NBA championship given the team's statistics during the NBA finals game. We want to find weights for each statistic that can show us how important each statistic is to predicting the overall winner, which helps our group conduct a greater analysis of how different focuses and strategies can affect the outcome of NBA games. Our group believes that certain statistics such as field goal percentage, turnovers, and offensive rebounds will be prevalent in winning NBA teams.

The dataset with which we want to create our model is the NBA Finals Team Stats dataset on Kaggle uploaded by Dave Rosenman. The dataset contains final data from 1980 to 2018, and is divided into two tables. The first table contains the data of each winning team and the second contains the losing team. Each observation includes data points like field goals made, field goals attempted, three point shots made, free throws made, total rebounds, assists, steals, turnovers, blocks, and many other statistics that will be covered in the data summary. It is important to note that each observation is one game in the championship series, teams can lose up to 3 games in the series and still win the championship, so even the championship game data contains losses.

The Data:



```
> glimpse(champion_data)
Rows: 220
Columns: 24
$ Year <int> 1980, 1980, 1980, 1980, 1980, 1980, 1981, 1981, 1981, 1981, 1981, 1981, 1982, ~
$ Team <chr> "Lakers", "Lakers", "Lakers", "Lakers", "Lakers", "Lakers", "Celtics", "Celtic~
$ Game <int> 1, 2, 3, 4, 5, 6, 1, 2, 3, 4, 5, 6, 1, 2, 3, 4, 5, 6, 1, 2, 3, 4, 1, 2, 3, 4, ~
$ Win  <int> 1, 0, 1, 0, 1, 1, 1, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1, 0, 1, 1, 1, 1, 1, 1, 0, 0, 1, ~
$ Home <int> 1, 1, 0, 0, 1, 0, 1, 1, 0, 0, 1, 0, 0, 0, 1, 1, 0, 1, 1, 1, 0, 0, 1, 1, 0, 0, ~
$ MP   <int> 240, 240, 240, 240, 240, 240, 240, 240, 240, 240, 240, 240, 240, 240, 240, 240~
$ FG   <int> 48, 48, 44, 44, 41, 45, 41, 41, 40, 35, 41, 43, 49, 35, 50, 45, 47, 47, 45, 40~
$ FGA  <int> 89, 95, 92, 93, 91, 92, 95, 82, 89, 74, 94, 78, 93, 83, 91, 97, 100, 87, 96, 8~
$ FGP  <dbl> 0.539, 0.505, 0.478, 0.473, 0.451, 0.489, 0.432, 0.500, 0.449, 0.473, 0.436, 0~
$ TP   <int> 0, 0, 0, 0, 0, 0, 0, 0, 2, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 2, 2, ~
$ TPA  <int> 0, 1, 1, 0, 0, 2, 1, 3, 3, 3, 3, 4, 0, 5, 1, 1, 2, 0, 0, 1, 1, 1, 2, 2, 5, 4, ~
$ TPP  <dbl> NA, 0.000, 0.000, NA, NA, 0.000, 0.000, 0.000, 0.667, 0.000, 0.000, 0.250, NA,~
$ FT   <int> 13, 8, 23, 14, 26, 33, 16, 8, 12, 16, 27, 15, 26, 24, 28, 21, 8, 20, 23, 23, 2~
$ FTA  <int> 15, 12, 30, 19, 33, 35, 20, 13, 19, 24, 35, 18, 35, 37, 47, 29, 16, 25, 31, 32~
$ FTP  <dbl> 0.867, 0.667, 0.767, 0.737, 0.788, 0.943, 0.800, 0.615, 0.632, 0.667, 0.771, 0~
$ ORB  <int> 12, 15, 22, 18, 19, 17, 25, 14, 16, 17, 19, 9, 19, 17, 17, 16, 26, 15, 21, 16,~
$ DRB  <int> 31, 37, 34, 31, 37, 35, 29, 34, 28, 30, 35, 28, 31, 22, 31, 33, 23, 34, 32, 28~
$ TRB  <int> 43, 52, 56, 49, 56, 52, 54, 48, 44, 47, 54, 37, 50, 39, 48, 49, 49, 49, 53, 44~
$ AST  <int> 30, 32, 20, 23, 28, 27, 23, 17, 24, 22, 25, 26, 34, 25, 30, 35, 31, 33, 26, 22~
$ STL  <int> 5, 12, 5, 12, 7, 14, 6, 6, 12, 5, 5, 6, 11, 11, 15, 10, 5, 12, 11, 10, 14, 9, ~
$ BLK  <int> 9, 7, 5, 6, 6, 4, 5, 7, 6, 6, 8, 0, 7, 6, 5, 4, 9, 11, 13, 6, 2, 11, 4, 4, 5, ~
$ TOV  <int> 17, 26, 20, 19, 21, 17, 19, 22, 11, 22, 14, 13, 22, 18, 18, 12, 24, 22, 14, 18~
$ PF   <int> 24, 27, 25, 22, 27, 22, 21, 27, 25, 22, 23, 21, 26, 21, 30, 21, 29, 26, 22, 16~
$ PTS  <int> 109, 104, 111, 102, 108, 123, 98, 90, 94, 86, 109, 102, 124, 94, 129, 111, 102~
```