# Proposal

*Jiazhang Cai*

*4/5/2020*

## Dataset

The data is about the NFL stadium attendence from "https://github.com/rfordatascience/tidytuesday/tree/master/data/2020/2020-02-04". There are three tables in the dataset. The first one is the overview of the attendence:

```
## Parsed with column specification:
## cols(
##   team = col_character(),
##   team_name = col_character(),
##   year = col_double(),
##   total = col_double(),
##   home = col_double(),
##   away = col_double(),
##   week = col_double(),
##   weekly_attendance = col_double()
## )
```

```
##      team team_name year  total   home   away week weekly_attendance
## 1 Arizona Cardinals 2000 893926 387475 506451    1             77434
## 2 Arizona Cardinals 2000 893926 387475 506451    2             66009
## 3 Arizona Cardinals 2000 893926 387475 506451    3                NA
## 4 Arizona Cardinals 2000 893926 387475 506451    4             71801
## 5 Arizona Cardinals 2000 893926 387475 506451    5             66985
## 6 Arizona Cardinals 2000 893926 387475 506451    6             44296
```

the dictionary of this dataset is:

| variable | class | description |
|---|---|---|
| team | character | team city |
| team_name | character | team name |
| year | integer | season year |
| total | double | total attendence across 17 weeks (1 week = no game) |
| home | double | total home attendence |
| away | double | total away attendence |
| week | character | week number (1-17) |
| weekly_attendence | double | weekly attendence |

The second one is the information about each team:

```
## Parsed with column specification:
## cols(
##   team = col_character(),
##   team_name = col_character(),
##   year = col_double(),
##   wins = col_double(),
##   loss = col_double(),
##   points_for = col_double(),
##   points_against = col_double(),
##   points_differential = col_double(),
##   margin_of_victory = col_double(),
##   strength_of_schedule = col_double(),
##   simple_rating = col_double(),
##   offensive_ranking = col_double(),
##   defensive_ranking = col_double(),
##   playoffs = col_character(),
##   sb_winner = col_character()
## )
```

```
##           team team_name year wins loss points_for points_against
## 1        Miami  Dolphins 2000   11    5        323            226
## 2 Indianapolis     Colts 2000   10    6        429            326
## 3     New York      Jets 2000    9    7        321            321
## 4      Buffalo     Bills 2000    8    8        315            350
## 5  New England  Patriots 2000    5   11        276            338
## 6    Tennessee    Titans 2000   13    3        346            191
##   points_differential margin_of_victory strength_of_schedule simple_rating
## 1                  97               6.1                  1.0           7.1
## 2                 103               6.4                  1.5           7.9
## 3                   0               0.0                  3.5           3.5
## 4                 -35              -2.2                  2.2           0.0
## 5                 -62              -3.9                  1.4          -2.5
## 6                 155               9.7                 -1.3           8.3
##   offensive_ranking defensive_ranking    playoffs    sb_winner
## 1               0.0               7.1    Playoffs No Superbowl
## 2               7.1               0.8    Playoffs No Superbowl
## 3               1.4               2.2 No Playoffs No Superbowl
## 4               0.5              -0.5 No Playoffs No Superbowl
## 5              -2.7               0.2 No Playoffs No Superbowl
## 6               1.5               6.8    Playoffs No Superbowl
```

the dictionary of this dataset is:

| variable | class | description |
|---|---|---|
| team | character | team city |
| team_name | character | team name |
| year | integer | season year |
| wins | double | wins (0-16) |
| loss | double | losses (0-16) |
| points_for | double | points for offensive performance |
| points_against | double | points for defensive performance |
| points_differential | double | points_for-points_against |
| margin_of schedule | double | (points scored-points allowed)/game played |
| strength_of_schedule | double | average quality of opponent as measured as measured by SRS |
| simple rating | double | team quality relative to average as measured by SRS |

| variable | class | description |
|---|---|---|
| offensive_ranking | double | team offense quality relative to average as measured by SRS |
| defensive_ranking | double | team defense quality relative to average as measured by SRS |
| playoffs | character | made playoffs or not |
| sb_winner | character | won superbowl or not |

The last one is the information of every games:

```
## Parsed with column specification:
## cols(
##   year = col_double(),
##   week = col_character(),
##   home_team = col_character(),
##   away_team = col_character(),
##   winner = col_character(),
##   tie = col_character(),
##   day = col_character(),
##   date = col_character(),
##   time = col_time(format = ""),
##   pts_win = col_double(),
##   pts_loss = col_double(),
##   yds_win = col_double(),
##   turnovers_win = col_double(),
##   yds_loss = col_double(),
##   turnovers_loss = col_double(),
##   home_team_name = col_character(),
##   home_team_city = col_character(),
##   away_team_name = col_character(),
##   away_team_city = col_character()
## )
```

```
##   year week           home_team            away_team               winner  tie
## 1 2000    1   Minnesota Vikings        Chicago Bears    Minnesota Vikings <NA>
## 2 2000    1  Kansas City Chiefs   Indianapolis Colts   Indianapolis Colts <NA>
## 3 2000    1 Washington Redskins    Carolina Panthers  Washington Redskins <NA>
## 4 2000    1     Atlanta Falcons  San Francisco 49ers      Atlanta Falcons <NA>
## 5 2000    1 Pittsburgh Steelers     Baltimore Ravens     Baltimore Ravens <NA>
## 6 2000    1    Cleveland Browns Jacksonville Jaguars Jacksonville Jaguars <NA>
##   day         date     time pts_win pts_loss yds_win turnovers_win yds_loss
## 1 Sun September 3 13:00:00      30       27     374             1      425
## 2 Sun September 3 13:00:00      27       14     386             2      280
## 3 Sun September 3 13:01:00      20       17     396             0      236
## 4 Sun September 3 13:02:00      36       28     359             1      339
## 5 Sun September 3 13:02:00      16        0     336             0      223
## 6 Sun September 3 13:02:00      27        7     398             0      249
##   turnovers_loss home_team_name home_team_city away_team_name away_team_city
## 1              1        Vikings      Minnesota          Bears        Chicago
## 2              1         Chiefs    Kansas City          Colts   Indianapolis
## 3              1       Redskins     Washington       Panthers       Carolina
## 4              1        Falcons        Atlanta          49ers  San Francisco
## 5                 1       Steelers     Pittsburgh          Ravens      Baltimore
## 6                 1         Browns      Cleveland         Jaguars   Jacksonville
```

the dictionary of this dataset is:

| variable | class | description |
|---|---|---|
| year | integer | season year |
| week | character | week number (1-17 and playoffs) |
| home_team | character | home team |
| away_team | character | away team |
| winner | character | winning team |
| tie | character | same for both team |
| day | character | day of week |
| date | character | date without year |
| time | character | time of game start |
| pts_win | double | points by winning team |
| pts_loss | double | points by lossing team |
| yds_win | double | yards by winning team |
| turnovers_win | double | turnovers by winning team |
| yds_loss | double | yards by losing team |
| turnovers_loss | double | turnovers by losing team |
| home_team_name | character | home team name |
| home_team_city | character | home team city |
| away_team_name | character | away team name |
| away_team_name | character | away team city |

The additional data is from "https://www2.census.gov/programs-surveys/popest/datasets/2010-2019/national/totals/", the *United States Census* website. The data is about the population and population change in every state of the United States.

# Plan

The data I found have three aspects: the attendence, the teams and the games. The describe of this dataset is aiming to study the attendence performance. However, I think we can expand the thought because of the abundance of the information we have.

First, we can study the attendence performance with the time, the location, the team played and something like that as the description of the dataset. To study deeper of the information about the location, I also find the data of population and population change in every state. In this part, I plan to get a map of the performance of the attendence in each state and maybe a model of predicting the attendence as well.

Second, we can study the team strength with the performance of each team in all the games, like the points they got, the yards they ran and maybe there history performance. I hope the output is a model of estimating the strength using score like the quality measured by SRS (Simple Rating System) or maybe a model of predicting if the team will win the super bowl or make the playoffs, which may use the logistic model.

Finally, we can also study the result of the games. There should be more effects to the result of the game excpet the strength of the teams. The time, the location, the attendence all could influnce the performance of the team. In this part, the ideal output is to get a model to predict the result of the game using all the information we have.

It's important that the three parts study are independent with each other. For example, in the study of the team strength, we assume the strength is the response and the results of the games are parts of the effects.

However, in the study of the games result, we assume the result is the response and the strength of team are parts of the effects. So we must be aware that we can't use the conclusion beyond the study we forcus on.

## Presentation

It would be a systematic study, so I prefer a longer report instead of the presentation.

Thank you!