Diary Entries NM2207

Ariel Quek

2023-10-24

```
knitr::opts_chunk$set(echo = TRUE)
```

Diary Entry #1

(1) Finalized Topic for Final Project:

Premier League matches from 2021-2022 season!

In the Premier League 2021-2022, there were many surprises undoubtedly, and the predictions made during the season varied. There have been many debate about outstanding players as well, with many critics/commentators seeming to agree only on Kevin De Bruyne from Manchester City.

It seems interesting to look into the matches leading up to Man City's win, while investigating the consistency of the various club's play on the field.

(2) Data sources curated so far:

Data sources are derived from [Tidy tuesday](%5Bhttps://github.com/rfordatascience/tidytuesday#about-tidytuesday), as recommended via Canvas!

The CSV files are downloaded in my local desktop:

- soccer 21-22.csv

- weeklyrank.csv

The links to relevant data sets and websites can be found here:

- https://github.com/rfordatascience/tidytuesday#about-tidytuesday

- https://www.kaggle.com/datasets/evangower/premier-league-match-data

- end -

Diary Entry #2

Question 1:

Finalised Idea/ Theme:

- Topic: A quantitative analysis of the 20/21 Premier League season

- Question: "**How influential was the home advantage for the football clubs in the 20-21 Premier League Season?**"

- *Considering the performance analysis and the relationship with the match outcomes*

In order to answer the above question:

- First, look at the performance analysis by calculating and comparing the average number of points and goal differences for each team over the season.

- Second, look at the percentage of matches won by the home team and the away team, and compare.

- Lastly, look at the correlation between the two, especially looking at the correlation between perhaps the number of shots on target for home teams as opposed to away teams, etc.

Question 2

- This is an important question because it is commonly assumed that football teams are more likely to win if they're playing at home. However, the best way to answer this question accurately is through data analysis, as it allows us to use the precise data of goals scores by each time to measure the home advantage.

- This can be helpful in predicting the scores for future football matches, and identifying how relevant the fan atmosphere is for matches. This information can be very influential considering the post-Covid boom of spectators for various sports.

- On more personal fronts, as a football fan, this question will help with match analysis, especially with the ongoing Premier League season.

Data:

Question 3

- The data packages are as follows:

Downloading data package:

```
#Opening the relevant CSV files

##Reading soccer data from 21-22 Premier League

soccer <- read.csv("soccer21-22.csv")
head(soccer)
```

```
##         Date   HomeTeam        AwayTeam FTHG FTAG FTR HTHG HTAG HTR    Referee HS
## 1 13/08/2021  Brentford         Arsenal    2    0   H    1    0   H  M Oliver  8
## 2 14/08/2021 Man United           Leeds    5    1   H    1    0   H P Tierney 16
## 3 14/08/2021    Burnley        Brighton    1    2   A    1    0   H   D Coote 14
## 4 14/08/2021    Chelsea Crystal Palace    3    0   H    2    0   H    J Moss 13
## 5 14/08/2021    Everton     Southampton    3    1   H    0    1   A  A Madley 14
## 6 14/08/2021  Leicester          Wolves    1    0   H    1    0   H  C Pawson  9
##    AS HST AST HF AF HC AC HY AY HR AR
## 1 22   3   4 12  8  2  5  0  0  0  0
## 2 10   8   3 11  9  5  4  1  2  0  0
## 3 14   3   8 10  7  7  6  2  1  0  0
## 4  4   6   1 15 11  5  2  0  0  0  0
## 5  6   6   3 13 15  6  8  2  0  0  0
## 6 17   5   3  6 10  5  4  1  2  0  0
```

```
## Reading the weekly rankings of the FCs

weeklyrank <- read.csv("weeklyrank.csv")
head(weeklyrank)
```

```
##              Team GD Points Rank Week
## 1 Man United  4       3    1    1
## 2     Chelsea  3       3    2    1
## 3  Liverpool  3       3    3    1
## 4    West Ham  2       3    4    1
## 5     Everton  2       3    5    1
## 6  Brentford  2       3    6    1
```

From these data packages, the following columns and rows will be used:

- In "weeklyrank", I will be using the columns "Team", "Rank" and "Week"

- In "soccer", I will be using "HomeTeam", "AwayTeam", "FTHG", "FTAG", "FTR", "HTHG", "HTAG", "HTR", "HS", "AS", "HST", "AST".

- Clarifying acronyms

  - FTHG - full time home goals
  - FTAG - full time away goals
  - HTHG - half time home goals
  - HTAG - half time away goals
  - FTR - full time result
  - HTR - half time result
  - HS - home shots
  - AS - away shots
  - HST - home shots on target
  - AST - away shots on target

- end -