

Homework 1
Association Rules and Sports Analytics

Vijay R Dhulipala
Qian Fu
Yidan Gao
Arivarasu Perumal
Yi-Li Yu
Chuchen Xiong

29th September 2019

Contents

Background and The Business Problem.	3
Our Task	3
Our Approach	3
Understanding Data	4
Reading the data	4
Filtering EPL Data	4
Exploring EPL Data	4
Missing Data	6
Data Munging and Transformation	7
Final Cleanup and Aesthetics	12
An Overview of the English Premier League	12
How does league position inform our performance?	12
How does Home and Away affect our performance?	14
How does player statistics influence?	15
Do implied probabilities measure the game outcome?	16
Summary of EPL	17
Analysis of opportunities to qualify into Champions League	17
Implied probability differentiator	17
High winning probability analysis	19
What team are we losing to with a high probability	20
Combination for win	21
Summary	23
Influence of Internal and External factors	23
Influence of external factors in the team's performance	23
Influence of internal factors in the team's performance	24
Summary	27
Final Takeaways	27
What have we found?	27
Final Recommendation	27

Background and The Business Problem.

The sports industry has seen a lucrative rise in stature and has now become an important contributor to the global economy. A huge amount of finance is being invested, resulting in an exponential rise in the amount of data generated by sports. This volume of data, together with the development in sports technology, has resulted in the need for informed decisions to be made, more than ever before. Since Michael Lewis' book "Moneyball" has been published, highlighting the method employed by the GM of Oakland Athletics (American baseball Club) in which he used statistical analysis to build a team on a low budget [1].

Sports Analytics has become a reality and a requirement. When the idea of working with sports data was formulated, the focus was decided to be placed upon association football (hereafter referred to as football) in particular. Football Analytics is a fascinating area. Since early 2000's football club have indulged themselves in work abreast with the recent development in the field of analytics to help improvise team performance and to make an informed decisions.

This is the case with Liverpool F.C., a professional football club in Liverpool, England, that competes in the Premier League, the top tier of English football. The club has won six European Cups, more than any other English club, three UEFA Cups and four UEFA Super Cups, also English records, eighteen League titles, seven FA Cups, a record eight League Cups, and fifteen FA Community Shields.

Our Task

Hired by the analytics team of Liverpool F.C., with the motivation to help the club to perform better and find patterns that tangibly contribute the clubs success using historical data. In order to increase the efficacy of our analysis and aid the manager/club in making informed decisions, we sought to provide a high-level overview of the game, our league and further proceed with team analysis. The challenge is that there are a lot of metrics to consider when evaluating a football team. Constraining the lens with which we explored, the data provides a set of valuable and concise insights.

Our Approach

Football is a beautiful game with fans from every corner of the world. It's a game where there's no doubting fans' commitment to their teams. It becomes cardinal to develop a loyal fanbase to become successful. Thus to drive our analysis, we first defined our success metric to be a victory in a match that will eventually score the club points. We think that this pinpointed view of success is valuable because it explains a greater amount by bringing the plethora of factors driving the game to a single dimension. This perspective of success directly influences the club's position on the leaderboard which influences the fans' sentiment.

Further, considering the English Premier League (EPL), in normal circumstances, four UEFA Champions League places are given to English clubs who fulfill the following criteria: The top four teams in the Premier League qualify for the group stage proper [1]. Aiming to be in the top 4 positions of EPL, help our club to qualify for UEFA Champions League. With the above definition of success metric, we make a subjective question objective with the purpose of qualifying into the UEFA Champions League.

The best place to understand our success is to find a relation/association of the list of metrics that maximize our likelihood to win a game.

In our analysis, we tried to explore the win probability from the betting odd and avoid losing in high probability win matches. Further, we understood what factor-like home v away, player stats and stage affect our lose or draw in high probability win matches. Finally, we suggest improvement in the lineup i.e. in terms of defense, midfield and forwards for Liverpool F.C. for these high probability games.

[1]. European qualification for UEFA competitions explained, <https://www.premierleague.com/european-qualification-explained>].

Understanding Data

Load dependencies

```
# Importing required libraries
library(ggplot2)
theme_set(theme_classic())

library(reshape2) ; library(ggrepel) ; library(RSQLite)
library(XML) ; library(knitr) ; library(dplyr)
library(naniar) ; library(gridExtra) ; library(magrittr)
library(lubridate) ; library(RSQLite) ; library(tidyr)
library(data.table) ; library(RColorBrewer) ; library(stringr)
library(arules) ; library(arulesViz)

opts_chunk$set(tidy.opts=list(width.cutoff=60),tidy=TRUE)
```

Reading the data

```
con = dbConnect(SQLite(), dbname='euro_soccer.sqlite') # Database connection

# Queries to all seven data tables
country = dbGetQuery(con, "SELECT * FROM country")
league = dbGetQuery(con, "SELECT * FROM league")
team = dbGetQuery(con, "SELECT * FROM team")

matches = dbGetQuery(con, "SELECT * FROM match WHERE league_id = 1729")
player = dbGetQuery(con, "SELECT * FROM player")
player_attributes = dbGetQuery(con, "SELECT * FROM player_attributes")
team_attributes = dbGetQuery(con, "SELECT * FROM team_attributes")
```

With the objective of qualifying into the UEFA Champions League by scoring points, we are only interested in the EPL teams and players.

Filtering EPL Data

```
unique_players <- unique(as.numeric(unlist(matches[,56:77])))
unique_teams <- unique(as.numeric(unlist(matches$home_team_api)))

player <- player %>% filter(player_api_id %in% unique_players)
player_attributes <- player_attributes %>% filter(player_api_id %in% unique_players)
team_attributes <- team_attributes %>% filter(team_api_id %in% unique_teams)
```

Exploring EPL Data

```
cat('Matches are from',min(matches$date),'to',max(matches$date),'\n')
```

```
## Matches are from 2008-08-16 00:00:00 to 2016-05-17 00:00:00
```

```
cat('Total # of seasons',length(unique(matches$season)),'\n')
```

```
## Total # of seasons 8
```

```
matches %>% group_by(season) %>% summarize(count=n())
```

```
## # A tibble: 8 x 2
##   season    count
##   <chr>     <int>
## 1 2008/2009    380
## 2 2009/2010    380
## 3 2010/2011    380
## 4 2011/2012    380
## 5 2012/2013    380
## 6 2013/2014    380
## 7 2014/2015    380
## 8 2015/2016    380
```

```
# EPL TEAMS
```

```
cat('Total # of teams in EPL',length(unique(matches$home_team_api_id)),'\n')
```

```
## Total # of teams in EPL 34
```

```
filter(team,team_api_id %in% unique(matches$home_team_api_id))$team_long_name
```

```
## [1] "Manchester United"      "Newcastle United"
## [3] "Arsenal"                "West Bromwich Albion"
## [5] "Sunderland"            "Liverpool"
## [7] "West Ham United"        "Wigan Athletic"
## [9] "Aston Villa"            "Manchester City"
## [11] "Everton"                "Blackburn Rovers"
## [13] "Middlesbrough"          "Tottenham Hotspur"
## [15] "Bolton Wanderers"       "Stoke City"
## [17] "Hull City"              "Fulham"
## [19] "Chelsea"                "Portsmouth"
## [21] "Birmingham City"       "Wolverhampton Wanderers"
## [23] "Burnley"                "Blackpool"
## [25] "Swansea City"           "Queens Park Rangers"
## [27] "Norwich City"           "Southampton"
## [29] "Reading"                "Crystal Palace"
## [31] "Cardiff City"           "Leicester City"
## [33] "Bournemouth"            "Watford"
```

```
cat('Total # of matches in EPL per Team per Season',+
    nrow(filter(matches,home_team_api_id == 8650,season == '2008/2009')),'\n')
```

```
## Total # of matches in EPL per Team per Season 19
```

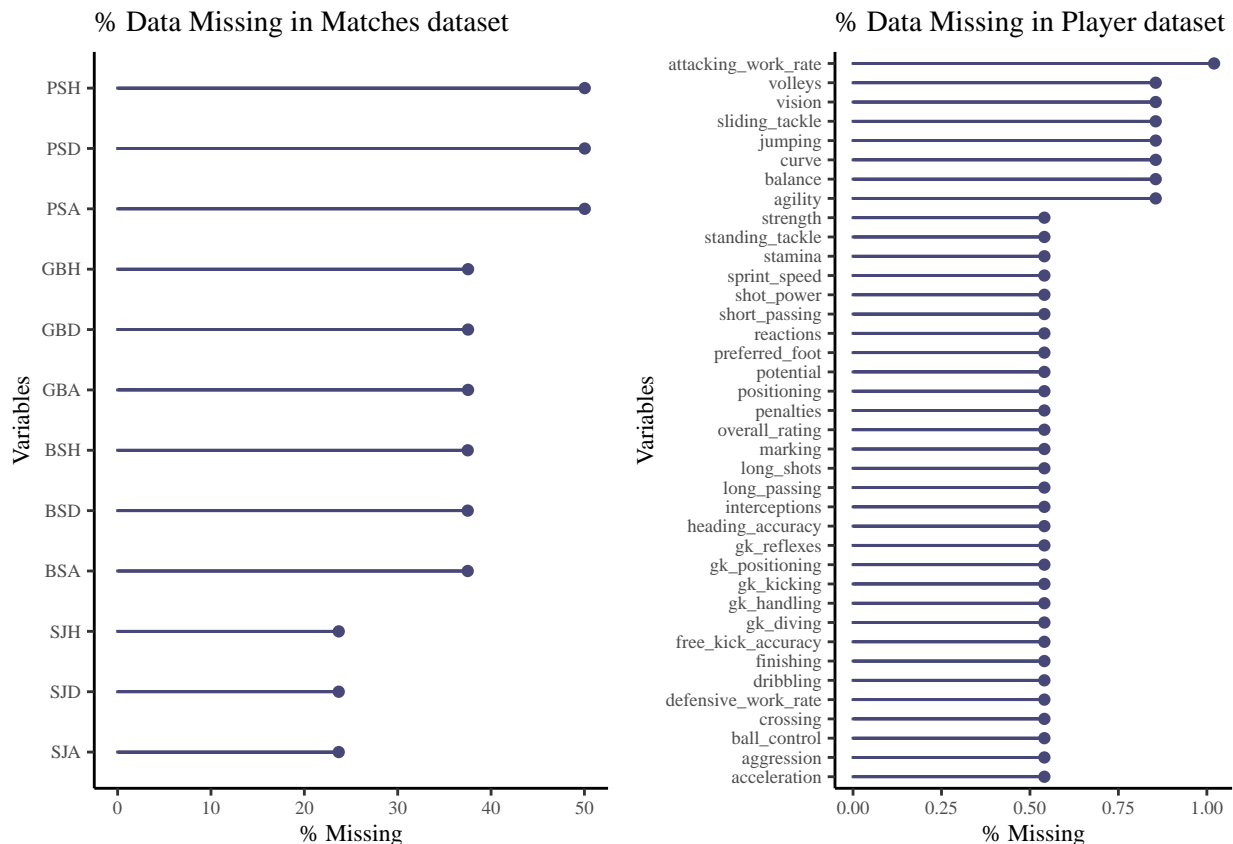
The Data contains information for 8 seasons from 2008-08-16 to 2016-05-17. So based on the above results we can conclude that each season has 380 matches in the premier league. Which serves properly because there are 20 teams. Each team will play in each round against another team and thus each team plays 38 matches in a season per round. Total Matches = 38 rounds * 10 matches per round. Thus 380 seems the right number.

```
# summary(player)
# summary(player_attributes)
# summary(team_attributes)
# summary(matches)
```

Let's have a look at the data contained in the EPL dataset and try to figure out if there is any missing data or additional cleaning is ecstasy.

Missing Data

```
p1 <- gg_miss_var((matches)[colSums(is.na(matches)) > 10], show_pct = TRUE) +
  labs(title = "% Data Missing in Matches dataset") + my_theme
p2 <- gg_miss_var((team_attributes)[colSums(is.na(team_attributes)) > 10], show_pct = TRUE) +
  labs(title = "% Data Missing in Team dataset") + my_theme
p3 <- gg_miss_var((player_attributes)[colSums(is.na(player_attributes)) > 10], show_pct = TRUE) +
  labs(title = "% Data Missing in Player dataset") + my_theme
grid.arrange(p1,p3, nrow = 1)
```



The first thing we discovered was the amount of missing data in all the datasets. We observed that a major chunk of our data was missing in the matches was related to the betting odds. “Pinnacle”, “Gamebookers”, “Blue Square” and “Stan James” betting odds are missing. When the betting odds are missing, it is observed that all the three i.e. home win, draw and away win betting odds are missing. Other variables contribute to less than 1% of missing data respective to each variable. Finally, player and team attributes have less than ~2% of missing data per variable.

The next step was to deal with these missing values. In order to deal with the missing values of the betting odds, we converted the betting odds to implied probability. Knowing how to convert betting odds can be very helpful. Understanding the implied probability behind the odds on offer is key to assessing the potential value in a particular betting market. Using the formula from betting sites [<https://www.bettingexpert.com/en-au/learn/understanding-betting-odds/how-to-convert-odds#gref>], we finally transformed our betting odds into implied probability by considering all the wins for home and away games. Hence, missing values can be dealt with in a superior fashion.

```
home_matches <- filter(matches,home_team_api_id == 8650)
away_matches <- filter(matches,away_team_api_id == 8650)

home_matches$liverpool_win_prob <- rowMeans(1 / home_matches[grepl( "H$"
                                , names( home_matches ))], na.rm = TRUE)
away_matches$liverpool_win_prob <- rowMeans(1 / away_matches[grepl( "A$"
                                , names( home_matches ))], na.rm = TRUE)
```

A quick check to see if there is any pattern to the missing betting values

```
missing <- is.na(home_matches$PSH)
nulls <- home_matches$liverpool_win_prob[missing]
not_nulls <- home_matches$liverpool_win_prob[!(missing)]
t.test(nulls, not_nulls)

##
##  Welch Two Sample t-test
##
## data:  nulls and not_nulls
## t = 0.86031, df = 149.4, p-value = 0.391
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.02676267  0.06803756
## sample estimates:
## mean of x mean of y
## 0.6455335 0.6248960
```

From our assumption, that the missing betting odds won't affect the implied probability. We performed a t-test with other betting odds and found that the odds are not so different indeed. The mean of the implied probability is roughly the same as our non-null values as illustrated above in the t-test. These values were classified as missing completely at random since their missingness was not correlated with the other features. The assumption that we must still make is that the rows with nulls were not tainted in any other way.

Data Munging and Transformation

There are many cases where the data can be further cleaned and different data set can be merged. First, we rename the columns for both home and away from data of Liverpool F.C. and calculate the betting odd as our assumption holds true.

```

home_player_index_begin <- match('home_player_1', names(matches))
home_player_index_end <- match('home_player_11', names(matches))
home_player_index <- c(home_player_index_begin:home_player_index_end)

away_player_index_begin <- match('away_player_1', names(matches))
away_player_index_end <- match('away_player_11', names(matches))
away_player_index <- c(away_player_index_begin:away_player_index_end)

home_team_index <- match('home_team_api_id', names(matches))
away_team_index <- match('away_team_api_id', names(matches))

home_goal_index <- match('home_team_goal', names(matches))
away_goal_index <- match('away_team_goal', names(matches))

```

Secondly, we derive the formation of Liverpool and thier opponent's for every match. This can be done by the X and Y positions of the players from the data. Finally, we merge the home and away matches.

```

# Add liverpool and opponent's formation

# Funcations to calculate formations
home_position_f = function(x, index = home_position_index) {
  df = data.frame(position=as.numeric(x[index]))
  dt = as.data.table(df)
  dt = unique(dt[,position_count:=.N, by=position])
  dt[,position_count]
}

away_position_f = function(x, index = away_position_index) {
  df = data.frame(position=as.numeric(x[index]))
  dt = as.data.table(df)
  dt = unique(dt[,position_count:=.N, by=position])
  dt[,position_count]
}

# Garthering player Y position index
home_position_index_begin <- match('home_player_Y1', names(matches))
home_position_index_end <- match('home_player_Y11', names(matches))
home_position_index <- c(home_position_index_begin:home_position_index_end)

away_position_index_begin <- match('away_player_Y1', names(matches))
away_position_index_end <- match('away_player_Y11', names(matches))
away_position_index <- c(away_position_index_begin:away_position_index_end)

# Getting formation
home_matches$liverpool_formation <- apply(home_matches, 1, home_position_f)
home_matches$opponent_formation <- apply(home_matches, 1, away_position_f)
away_matches$liverpool_formation <- apply(away_matches, 1, away_position_f)
away_matches$opponent_formation <- apply(away_matches, 1, home_position_f)

# Merging home and away matches
liverpool_matches <- merge(home_matches, away_matches, all = TRUE)

```

We converted the implied probability into factor i.e. high, medium and low. Next. we determined which team won the game based on the goal difference. In addition, we converted the stage into factors such as

1-12 are early round, 13-26 are medium rounds and 26-38 are the last rounds. The rational behind is that, as season progress based on the points score teams might suffer from emotional issues due to their confidence,

```
win_quant <- quantile(liverpool_matches$liverpool_win_prob, probs = c(0.33,0.66), na.rm = FALSE)

liverpool_matches$win_prob_class[liverpool_matches$liverpool_win_prob >= win_quant[2]] <- 'high'
liverpool_matches$win_prob_class[liverpool_matches$liverpool_win_prob < win_quant[2]] <- 'medium'
liverpool_matches$win_prob_class[liverpool_matches$liverpool_win_prob < win_quant[1]] <- 'low'

# Getting the result of liverpool's match
liverpool_matches$goal_diff <- liverpool_matches$liverpool_goal - liverpool_matches$opponent_goal
liverpool_matches$result[liverpool_matches$goal_diff > 0] <- "W"
liverpool_matches$result[liverpool_matches$goal_diff < 0] <- "L"
liverpool_matches$result[liverpool_matches$goal_diff == 0] <- "D"

# Changing stage to peirod
liverpool_matches$stage[(liverpool_matches$stage > 0) & (liverpool_matches$stage <= 13)] <- 'early round'
liverpool_matches$stage[(liverpool_matches$stage > 13) & (liverpool_matches$stage <= 26)] <- 'medium round'
liverpool_matches$stage[(liverpool_matches$stage > 26) & (liverpool_matches$stage <= 38)] <- 'last round'

# Changing date to year
liverpool_matches$date <- as.Date(liverpool_matches$date)
liverpool_matches$year <- as.numeric(format(liverpool_matches$date, format = "%Y"))

# Adding opponent name
liverpool_matches <- select(liverpool_matches, match_api_id, stage, year, date, liverpool_team_api_id,
                           opponent_team_api_id, goal_diff,result, location, win_prob_class
                           , liverpool_formation, opponent_formation)
liverpool_matches <- dplyr::rename(liverpool_matches, team_api_id = opponent_team_api_id)

opponent_name <- select(team, team_api_id, team_long_name)
liverpool_matches <- merge(x = liverpool_matches, y = opponent_name, by = "team_api_id") %>% dplyr::rename(team_api_id = opponent_team_api_id)

# Selecting the columns we want
liverpool_matches <- select(liverpool_matches, match_api_id, stage, year, date, liverpool_team_api_id,
                           opponent_name, goal_diff,result, location, win_prob_class
                           , liverpool_formation, opponent_formation)
```

We merged the Liverpool F.C. player data and their attributes into a single data set. The birthday and date column are having a timestamp which won't be utilized. Hence, we split the year from these columns.

```
lvp_matches <- matches %>% filter(home_team_api_id == 8650 | away_team_api_id == 8650)
lvp_unique_players <- unique(as.numeric(unlist(lvp_matches[,56:77])))

lvp_players <- player %>% filter(player_api_id %in% lvp_unique_players)
lvp_player_attributes <- player_attributes %>% filter(player_api_id %in% lvp_unique_players)

lvp_players <- merge(lvp_player_attributes, lvp_players, by = 'player_api_id')

lvp_players <- lvp_players %>% dplyr::mutate(year = lubridate::year(date))
lvp_players <- lvp_players %>% dplyr::mutate(birthyear = lubridate::year(birthday))
```

The player attributes are spanned over years. Few players only have data points for a specific year, hence

we decided to take the average across their skills over time. We assume that there attributes will remain constant or will over vary little over time. [Code hidden due to length]

After getting the player attributes, we consolidate the forward, center(midfield) and back (defense) attributes based on the player position and calculate the mean with respect to their position.

```
# Selecting the columns of player_Y
liverpool_matches_position <- lvp_matches %>%
  select(match_api_id, matches("home_player_Y[[:digit:]]"), matches("away_player_Y[[:digit:]]")) %>%
  gather(player, player_Y_number, -match_api_id)

# Changing the columns of player_Y to position
liverpool_matches_position$player_position[liverpool_matches_position$player_Y_number == 1] <- 'gk'
liverpool_matches_position$player_position[(1 < liverpool_matches_position$player_Y_number) & (liverpool_matches_position$player_Y_number < 5)] <- 'f'
liverpool_matches_position$player_position[(5 <= liverpool_matches_position$player_Y_number) & (liverpool_matches_position$player_Y_number < 9)] <- 'm'
liverpool_matches_position$player_position[(9 <= liverpool_matches_position$player_Y_number)] <- 'b'

# Cleaning position format
liverpool_matches_position$player = gsub("Y", "", liverpool_matches_position$player)
```

We calculate the attributes difference of Liverpool and opponent and change the result to factor. And finally, we get the table for opponent team attributes and player attributes difference.

```
# add liverpool and opponent prefix to attributes
logic_index <- ((lvp_matches_player_attr$home_away=='home') & (lvp_matches_player_attr$home_team_api_id == lvp_matches_player_attr$away_team_api_id))
lvp_matches_player_attr$position[logic_index] <- paste('liverpool', lvp_matches_player_attr$position)
lvp_matches_player_attr$position[is.na(lvp_matches_player_attr$position)] <- paste('opponent', lvp_matches_player_attr$position)
lvp_matches_player_attr <- unique(lvp_matches_player_attr)

# Getting the index of attributes
index <- c(match('overall_rating', names(lvp_matches_player_attr)):match('gk_reflexes', names(lvp_matches_player_attr)))

#Spreading the attributes table by match_api_id
result_tbl <- select(lvp_matches_player_attr, match_api_id) %>% unique()
for(i in index){

  temp_table <- select(lvp_matches_player_attr, match_api_id, position, colnames(lvp_matches_player_attr)[index[i]])
  temp_table <- spread(temp_table, position, colnames(lvp_matches_player_attr)[index[i]])
  result_tbl <- merge(result_tbl, temp_table, by = 'match_api_id')
}

# Renaming each columns
for(d in index){
  temp <- names(result_tbl)[(8*d-30):(8*d-23)]
  num <- c(1:8)
  for(p in num){
    names(result_tbl)[8*d-31+p] <- paste(temp[p], colnames(lvp_matches_player_attr)[d], sep='_')
  }
}

# Calculating the attributes difference
len <- c(1:length(index))

for(d in len){
```

```

start_index <- 8 * d - 6
for(n in c(1:4)){
  now_index <- start_index + n - 1
  result_tbl[now_index] <- result_tbl[now_index] - result_tbl[now_index + 4]
}
}

drop_index <- c(6:9)
for(d in len){
  result_tbl <- result_tbl[,-drop_index]
  drop_index <- drop_index + 4
}

# Changing attributes difference to factor
quant_index <- c(match('liverpool_back.x_overall_rating', names(result_tbl)):match('liverpool_gk_gk_ref', names(result_tbl)))
quants <- c(0.2, 0.4, 0.6, 0.8)
quants_result <- apply( result_tbl[quant_index] , 2 , quantile , probs = quants , na.rm = TRUE )

result_tbl[quant_index] <- lapply(result_tbl[quant_index], as.numeric)

for(d in quant_index){
  result_tbl[d][result_tbl[d] >= quants_result[4 + (d - 2) * 4]] <- 'very high'
  result_tbl[d][(result_tbl[d] < quants_result[4 + (d - 2) * 4]) & (result_tbl[d] >= quants_result[3 + (d - 2) * 4])] <- 'high'
  result_tbl[d][(result_tbl[d] < quants_result[3 + (d - 2) * 4]) & (result_tbl[d] >= quants_result[2 + (d - 2) * 4])] <- 'medium'
  result_tbl[d][(result_tbl[d] < quants_result[2 + (d - 2) * 4]) & (result_tbl[d] >= quants_result[1 + (d - 2) * 4])] <- 'low'
  result_tbl[d][result_tbl[d] < quants_result[1 + (d - 2) * 4]] <- 'very low'
}

for(c in quant_index){
  for (r in c(1:(nrow(result_tbl)))){
    result_tbl[r,c] <- ifelse(result_tbl[r,c] < 'a', 'high', result_tbl[r,c])
  }
}

# Dropping the redundant gk attributes
gk_re_index <- grepl( 'gk', names(result_tbl)) & !(grepl( 'gk.y_gk', names(result_tbl)) | grepl( 'gk.x_gk', names(result_tbl)))
gk_re_index <- !gk_re_index
result_tbl <- result_tbl[,gk_re_index]

# Changing column names
names(result_tbl) <- sub(".x", "", colnames(result_tbl))
names(result_tbl) <- sub(".y", "", colnames(result_tbl))
names(result_tbl) <- sub("_gk", "", colnames(result_tbl))
names(result_tbl) <- paste(names(result_tbl), 'diff', sep = '_')
names(result_tbl)[1] <- 'match_api_id'

# Getting the final table
result_tbl <- merge(liverpool_matches, result_tbl, by = 'match_api_id')

```

Final Cleanup and Aesthetics

An overview of our final data frame. We are selecting only two attributes. In total the table has 304 observation and 119 variables.

```
df <- select(result_tbl, stage, opponent_name, result, liverpool_formation, opponent_formation, liverpool_for  
head(df)
```

```
##           stage           opponent_name result liverpool_formation
## 1 medium round      Tottenham Hotspur      W      ( 4, 4, 1, 1)
## 2 medium round           Stoke City        D      ( 4, 4, 2)
## 3 medium round Wolverhampton Wanderers      D      ( 4, 4, 1, 1)
## 4 medium round      Bolton Wanderers      W      ( 4, 4, 1, 1)
## 5 medium round           Everton          W      ( 4, 2, 3, 1)
## 6 medium round           Arsenal          L      ( 4, 2, 3, 1)
##  opponent_formation liverpool_forward_reactions_diff
## 1      ( 4, 4, 2)                                low
## 2      ( 4, 4, 2)                                medium
## 3    ( 4, 3, 2, 1)                                low
## 4      ( 4, 5, 1)                                low
## 5      ( 4, 5, 1)                                very low
## 6      ( 4, 3, 3)                                very low
##  liverpool_center_reactions_diff
## 1                                low
## 2                                medium
## 3                                very high
## 4                                very high
## 5                                medium
## 6                                very low
```

An Overview of the English Premier League

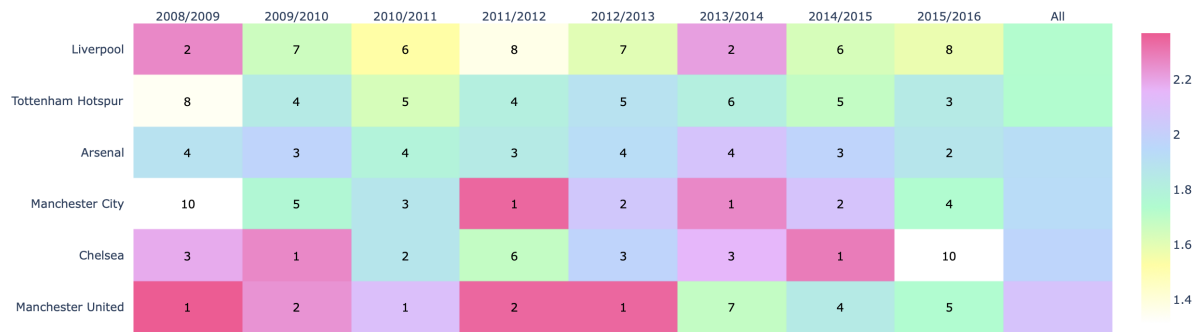
How does league position inform our performance?

Description and Rationale for the Chosen Analysis

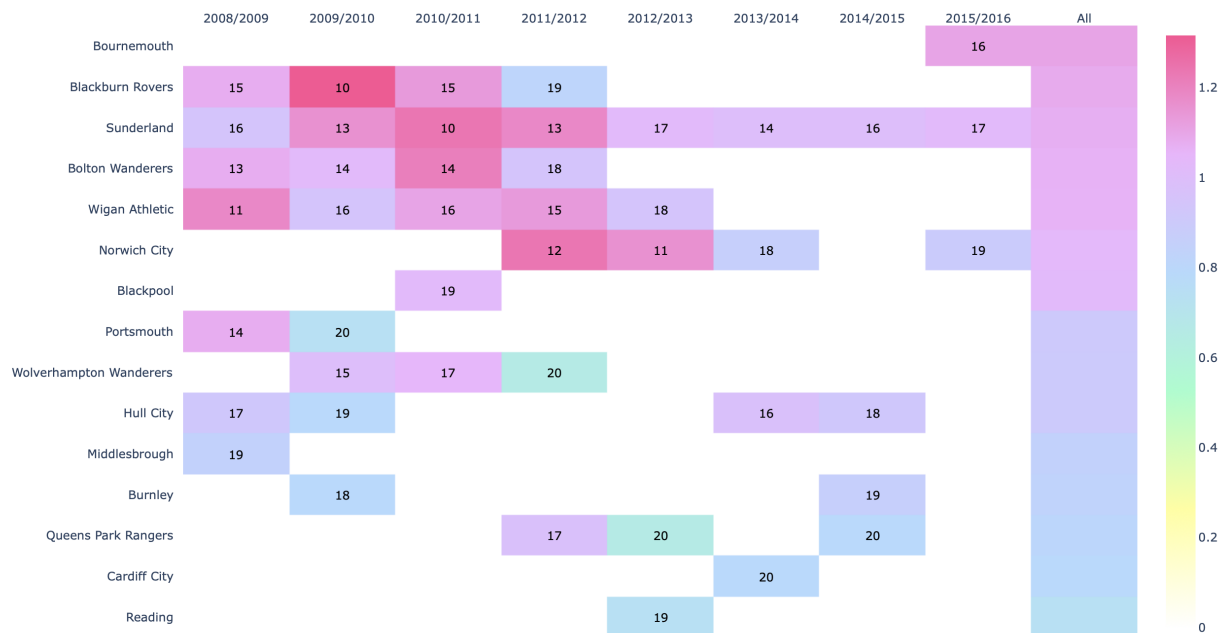
It is our primary task to understand how teams are performing in the EPL. In order to narrow down our analysis, we are interested in qualifying for the Champions League. Hence, we are interested to understand the team which has qualified into the Champions League.

Execution and Results (without code)

We calculated points for each team over the season to build a point table. Teams receive three points for a win and one point for a draw. No points are awarded for a loss. Teams are ranked by total points, then goal difference, and then goals scored.



Top 6 teams in the points tables



Bottom 15 teams in the points tables

Interpretation

Teams having a point average of less than 1 have a very high chance of relegation, having less than 0.9 (i.e. < 34 points) they are always relegated. Interestingly, in the 2010-11 season, teams having 39 points were also relegated. There are 9 teams that have maintained their status in the Premier League over the course of 8+ years. Arsenal has been the most consistent performers during these periods, though never winning the league in the same period. Manchester United have been the most successful during the 8 seasons, having won 3 league titles. After the 2012-13 season, there has been a drop off in their league consistency, which we can say might be due to the departure of their coach. Following United, Chelsea and Manchester City have been the most successful, each winning the league title twice. At the other end of the spectrum, Sunderland has been able to maintain its place in the premier league even after flirting with relegation during the 8 year period.

Finally, coming to Liverpool F.C., we qualified into the Champions League twice. But, have never won the EPL. In the year 2008 and 2013, we performed well. In addition, we maintain a position around #6 with average overall season points per match being 1.7303.

Conclusions

Understanding the team's position helped us with the knowledge about which teams are performing well and their trends. On average, we need 12-14 points more to qualify into the Champion league which is

equivalent to winning 3-4 more matches per season. Hence, we need to bridge the gap between our desired points and our current point. This indicates that we need to understand where we have the opportunity of scoring points and exploit these chances.

How does Home and Away affect our performance?

Description and Rationale for the Chosen Analysis

In football, the term home advantage describes the benefit that the home team is said to gain over the visiting team. Over time this benefit has been attributed to psychological effects supporting fans have on the competitors or referees; to psychological or physiological advantages of playing near home in familiar situations; to the disadvantages away teams suffer from changing time zones or climates, or from the rigors of travel. We are interested in looking into how the team is performing in-home ground vs away ground and compare it with Liverpool F.C.

Execution and Results (including code)

We calculated the # of the win per team across the season and observed the win count across home and away.

```
matches$goal_diff = matches$home_team_goal - matches$away_team_goal
home_win_matches <- matches %>% filter(goal_diff > 0)
home_win_matches <- merge(x = home_win_matches, y = team, by.x = 'home_team_api_id', by.y = 'team_api_id')

away_win_matches <- matches %>% filter(goal_diff < 0)
away_win_matches <- merge(x = away_win_matches, y = team, by.x = 'away_team_api_id', by.y = 'team_api_id')

home_win_matches$win_prob <- rowMeans(1 / home_win_matches[grepl("H$", names(home_win_matches))], na.rm = TRUE)
away_win_matches$win_prob <- rowMeans(1 / away_win_matches[grepl("A$", names(away_win_matches))], na.rm = TRUE)

home_win_matches <- select(home_win_matches, team_long_name, win_prob) %>%
  group_by(team_long_name) %>%
  summarize(home_win_prob = mean(win_prob))

away_win_matches <- select(away_win_matches, team_long_name, win_prob) %>%
  group_by(team_long_name) %>%
  summarize(away_win_prob = mean(win_prob))

win_matches <- merge(home_win_matches, away_win_matches, by='team_long_name')
win_matches <- win_matches %>% arrange(desc(home_win_prob)) %>% head(8)

dfm <- melt(win_matches[,c('team_long_name', 'home_win_prob', 'away_win_prob')], id.vars = 1)

ggplot(dfm, aes(x = team_long_name, y = value)) +
  geom_bar(aes(fill = variable), stat = "identity", position = "dodge") +
  labs(title = "Average winning probability of top teams across home and away", x = 'Teams', y = 'Average winning probability')
```



Interpretation

From the above plot, we can observe that the top team (selected the team who are consistently top on the points table) shows a significant difference between home and away games.

Conclusions

From this we can conclude that home team does have an advantage over away team.

How does player statistics influence?

Description and Rationale for the Chosen Analysis

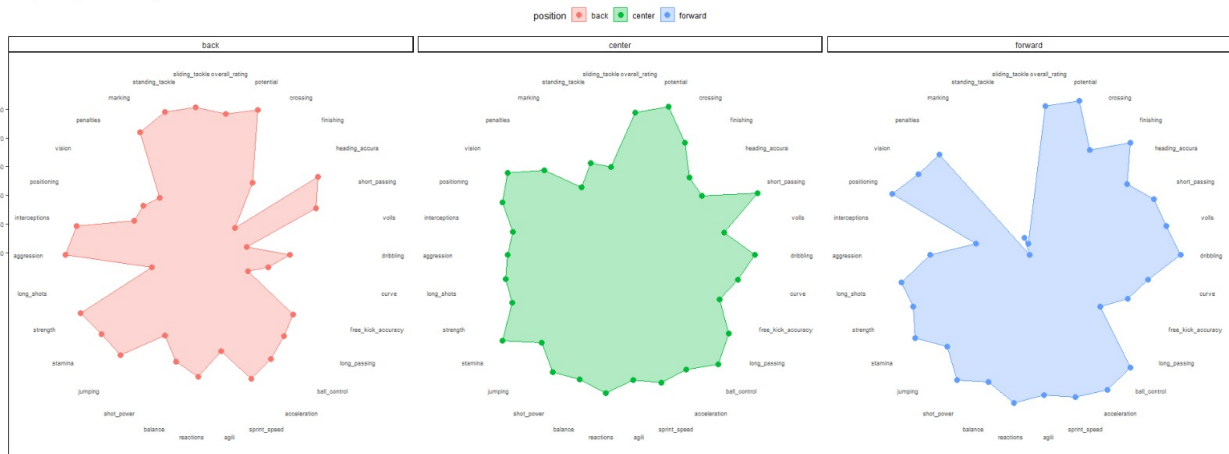
Ultimately, every team is a combinator of their players. Hence, it motivates us to observe our player stats and compare them with other teams. In order to observe player stats, we consolidate players based on forward, mid and defense to plot their strength and weakness

Execution and Results (including code)

[Commenting the code to save knitting time]

```
# spider_plot_table <- bind_rows(matches_forward_attr_plot, matches_center_attr_plot, matches_back_attr_plot)
#
# ggRadar(data=spider_plot_table, aes(color=position), interactive=FALSE, rescale = FALSE) +
#   facet_wrap(~position)
```

Liverpool Player Attributes by Position



Interpretation

In order to suggest changes, we need to know our team's strengths and weaknesses. Hence, we can use these stats to compare with another team in further analysis.

Conclusions

After having a look at the overall performance and formation, we need to understand what makes up the formation. Players! Understanding player statistics would give us insights into how teams are backing the formation with the player and might give us a lose the lead on the tactic which teams might be using.

Do implied probabilities measure the game outcome?

Description and Rationale for the Chosen Analysis

Since we converted our betting odd to winning probability, we need to understand how this factor is a good estimated of the math outcome. Hence, we perform this analysis to understand how teams are varying with implied winning probability.

Execution and Results (including code)

```
matches$goal_diff = matches$home_team_goal - matches$away_team_goal
home_win_matches <- matches %>% filter(goal_diff > 0) %>%
  group_by(home_team_api_id) %>% summarize(n = n())
home_win_matches <- merge(x = home_win_matches, y = team, by.x = 'home_team_api_id', by.y = 'team_api_id')

away_win_matches <- matches %>% filter(goal_diff < 0) %>% group_by(away_team_api_id) %>% summarize(n = n())
away_win_matches <- merge(x = away_win_matches, y = team, by.x = 'away_team_api_id', by.y = 'team_api_id')

home_win_matches <- select(home_win_matches, team_long_name, n)
away_win_matches <- select(away_win_matches, team_long_name, n)

win_matches <- merge(home_win_matches, away_win_matches, by='team_long_name')
win_matches <- win_matches %>% arrange(desc(n.x)) %>% head(8)

dfm <- melt(win_matches[,c('team_long_name', 'n.x', 'n.y')], id.vars = 1)

ggplot(dfm, aes(x = team_long_name, y = value)) +
  geom_bar(aes(fill = variable), stat = "identity", position = "dodge") +
  labs(title = "Win count of top teams across location", x = 'Teams', y = '# of games') + my_theme
```




Interpretation

We can observe that the home team has higher betting odd when compared to the away team. Better performing teams have signified higher betting odd when compared to other teams.

Conclusions From this analysis, we can corroborate our assumption that winning probability which was calculated from betting odds is a good indicator of the outcome. As the home team is expected to have a higher winning probability. In addition, the team which are at the top of the table have higher probability which also corroborates our assumption

Summary of EPL

From our analysis, we observe that there are particular set of teams that consistently perform well in the EPL. Further, there is a significant difference between home and away matches.

Focusing on Liverpool F.C., we observed that we qualified into Champions League only twice. Furthermore, we usually end up around the 6th position. Similarly, the win probability calculated from the betting odd is high for the home and good teams.

This gives us a motivation to focus on factors like home vs away, stage and betting odd to narrow own our analysis.

Analysis of opportunities to qualify into Champions League

Implied probability differentiator

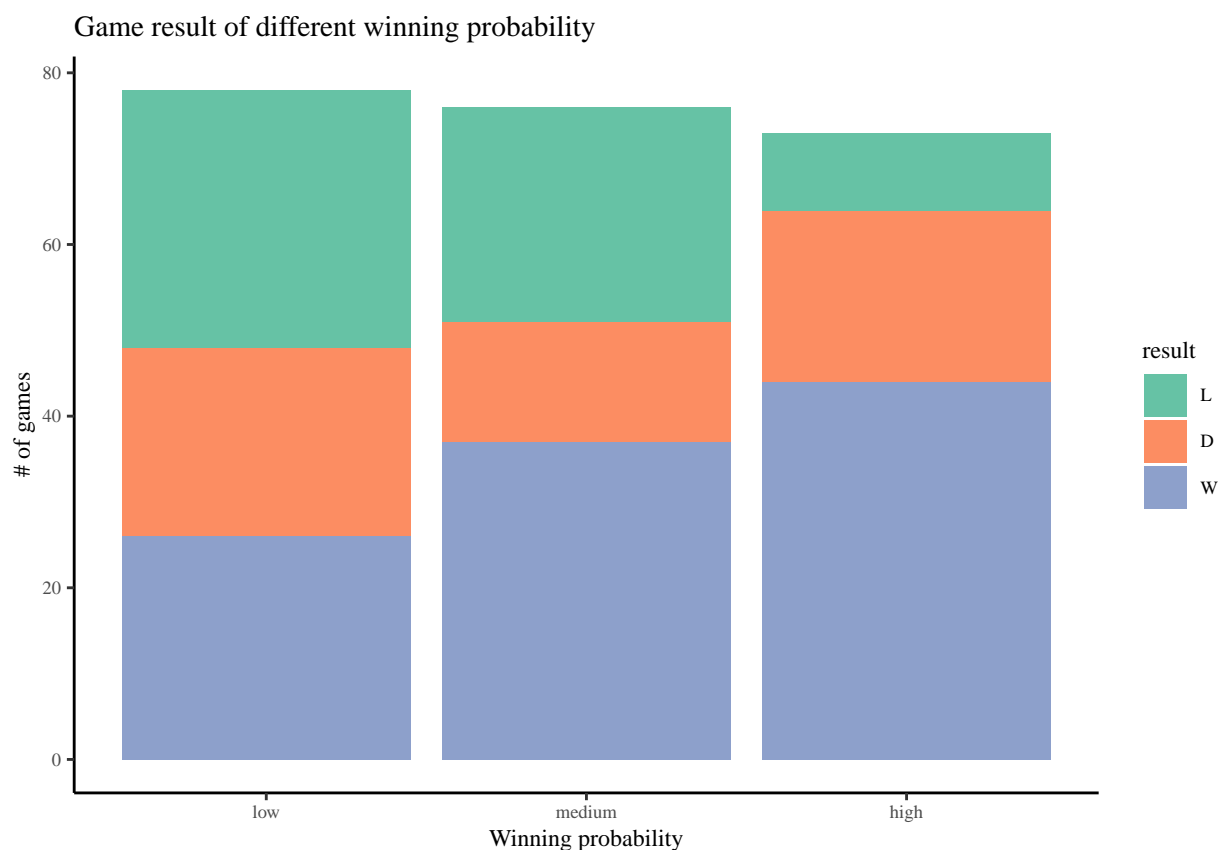
Description and Rationale for the Chosen Analysis

As we converted our betting odds into implied probability, where Implied Probability is simply the process

of converting betting odds into probabilities. When a Bookie sets the odds on a bet, he/she must first determine what chance he gives that bet of winning. We are interested in understanding if we are winning the matches where we have a high probability and seize an opportunity if it exists. Hence, as we divided the probabilities into high, medium and low; we need to visualize them.

Execution and Results (including code) We analyzed the win percentage across three winning probabilities for all Liverpool F.C. matches.

```
# result across win_prob_class
liverpool_matches$win_prob_class <- factor(liverpool_matches$win_prob_class, levels = c("low", "medium", "high"))
liverpool_matches$result <- factor(liverpool_matches$result, levels = c("L", "D", "W"))
ggplot(data = liverpool_matches, aes(x = win_prob_class, fill = result)) +
  scale_fill_brewer(palette = "Set2") +
  geom_bar() +
  labs(title = "Game result of different winning probability", x = 'Winning probability', y = '# of games')
```



Interpretation

On analyzing the bar plot, we understand that there is an equal number of matches distributed across all the probabilities. In addition, it can be understood that the loss and draw gap reduces when the probability increases from low to high, which is expected. However, the same trend is not seen when the probability increases from medium to high. In fact, we are converting more lose games into draw keep the increase in the winning matches the same.

Conclusions

In order to interpret the analysis, it is important to understand that probabilities are an external factor presents us with a prescription of how a team might or is expected to perform under normal conditions. Nevertheless, this helps us understand that we have an opportunity to win matches that have high probability yet ending up in a draw or lose. Hence, we are interested in converting these games into wins.

Finally, we make an assumption that all the betting odds are random samples i.e. bets are placed irrespective of location and have little external factor influence.

High winning probability analysis

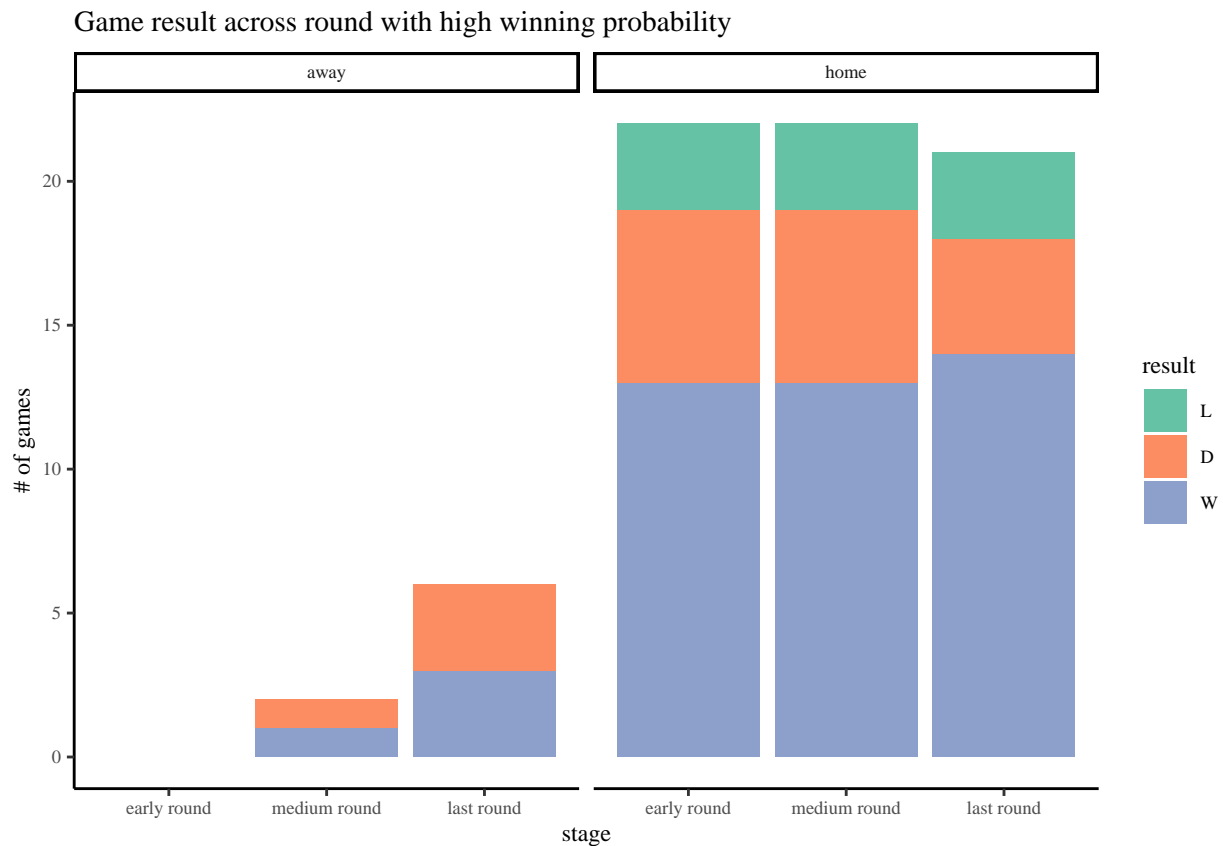
Description and Rationale for the Chosen Analysis

In order to understand how we can maximize the win count, it is important to understand our focus matches. The above analysis brings into light the matches we need to focus on. Naturally, we are interested in analysis these games by narrowing them. Hence, we chose to zoom into the high probability games across home & away and stages.

Execution and Results (including code)

We plotted the high probability games across different stages of the league i.e. 1-12 as early, 13-24 are mid and the others as final rounds. Further, from our understand about the home game sentiment we divided the analysis further into home vs away matches.

```
# result across win_prob_class
# result across round
liverpool_matches_win_prob_high <- filter(liverpool_matches, win_prob_class == 'high')
liverpool_matches_win_prob_high$stage <- factor(liverpool_matches_win_prob_high$stage, levels = c("early", "medium", "last"))
ggplot(data = liverpool_matches_win_prob_high, aes(x = stage, fill = result)) +
  scale_fill_brewer(palette = "Set2") +
  geom_bar() +
  facet_grid(.~location) +
  labs(title="Game result across round with high winning probability", y = '# of games') + my_theme
```



Interpretation

From our analysis, we made a significant number of observations. First, there is less number of away matches with high probability when compared to home ground. This corroborates that home game dose has various advantages. Further, Liverpool F.C. concludes many high probability games as draw in their home ground.

Conclusions

From focusing on all games, we narrowed it to a high probability game. Further, we found that we lose or draw many home games even with a high probability of winning.

What team are we losing to with a high probability

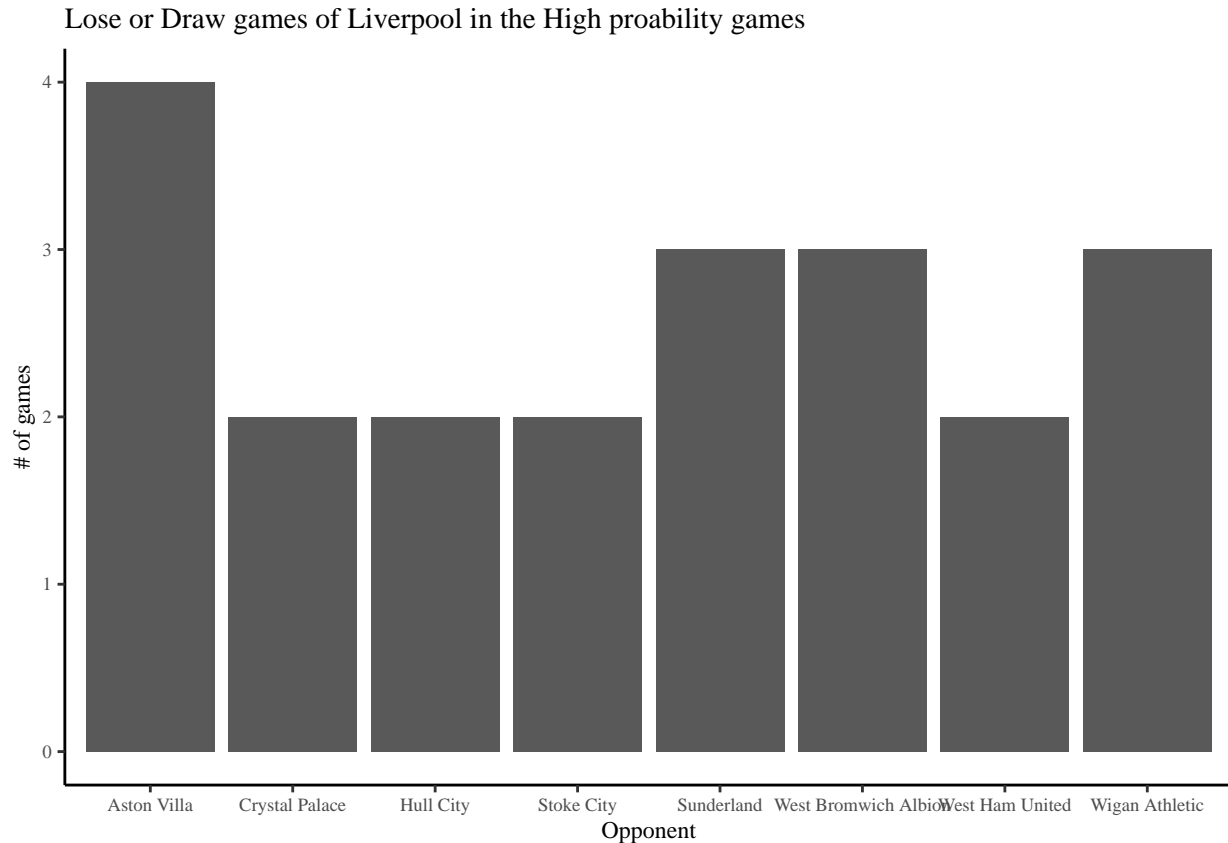
Description and Rationale for the Chosen Analysis

After understanding the game we are losing or draw, it is common to understand how are we losing? We are interested in quantifying these metrics. In order to observe patterns, we had to look into our opponents. Hence, we analyzed the opponent who we played and lost given a high probability.

Execution and Results (including code)

We calculated the # of times we are defeating an opponent and plotted the top opportunities.

```
liverpool_matches_opponent_name <- liverpool_matches_win_prob_high %>%  
  filter(win_prob_class == 'high' & (result == 'D' | result == 'L')) %>%  
  select(opponent_name, stage)  
  
liverpool_matches_opponent_name <- liverpool_matches_opponent_name %>%  
  group_by(opponent_name)%>%  
  summarize(n = n()) %>%  
  filter(n > 1)  
  
ggplot(liverpool_matches_opponent_name, aes(y=n,x=opponent_name)) +  
  scale_fill_brewer(palette = "Set2") +  
  geom_bar(stat = "identity") +  
  labs(title="Lose or Draw games of Liverpool in the High proability games",y = '# of games',x= 'Opp
```



Interpretation

From the plot, we can observe that there are around 9 teams with a total count of ~30 matches where we had the opportunity to win and we didn't.

Conclusions

On observation is that comparing these results from the points table we generated we can observe that most of the team are at the bottom 5-7 of the top 20 points table. This certainly paths the way to the perspective that these teams are having comparatively less overall team stats.

Combination for win

Description and Rationale for the Chosen Analysis

After we have focused our analysis to improve on high probability, home, and mid-round matches. We are interested in how to improve on these to account for a win. In order to win the game, we need to understand what factor can help us win the game. One set of factor which we are interested in looking are the player attributes. From our data transformation, we calculated the mean attributes scores of our defense, attacking, midfield and goalkeeper. Further, these mean differences of Liverpool F.C. and our opponents have been player attributes have been transformed into a scale of high, medium and low. Hence, we are interested in what difference in the attributes w.r.t the opponent would convert us to victory.

Execution and Results (including code)

In order to understand these factors, we applied association rules over these mean differences of attributes as LHS and win as RHS in the rule and observed the support, confidence and lift.

```
back_rule_table <- back_rule_table %>% filter(stage == 'medium round' & win_prob_class == 'high')
# drop stage, win_prob_class, location, and name columns
back_rule_table <- back_rule_table[, -c(1,2,4,5)]
```

```

back_rule_table = as(back_rule_table, "transactions")

back_rules_low <- apriori(back_rule_table, parameter=list(support =0.1, confidence =0.3, minlen=2, maxlen=10))
back_rules_low <- subset(subset(back_rules_low, subset = rhs %pin% "result=D"), lhs %pin% 'low')

back_rules_high <- apriori(back_rule_table, parameter=list(support =0.15, confidence =0.7, minlen=2, maxlen=10))
back_rules_high <- subset(subset(back_rules_high, subset = rhs %pin% "result=W"), lhs %pin% 'high')

# Center attributes
center_rule_table <- center_rule_table %>% filter(stage == 'medium round' & win_prob_class == 'high')
# drop stage, win_prob_class, location, and name columns
center_rule_table <- center_rule_table[,-c(1,2,4,5)]
center_rule_table = as(center_rule_table, "transactions")

center_rules_low <- apriori(center_rule_table, parameter=list(support =0.15, confidence =0.5, minlen=2, maxlen=10))
center_rules_low <- subset(subset(center_rules_low, subset = rhs %pin% "result=D"), lhs %pin% 'low')

center_rules_high <- apriori(center_rule_table, parameter=list(support =0.15, confidence =0.75, minlen=2, maxlen=10))
center_rules_high <- subset(subset(center_rules_high, subset = rhs %pin% "result=W"), lhs %pin% 'high')

# Forward attributes
forward_rule_table <- forward_rule_table %>% filter(stage == 'medium round' & win_prob_class == 'high')
# drop stage, win_prob_class, location, and name columns
forward_rule_table <- forward_rule_table[,-c(1,2,4,5)]
forward_rule_table = as(forward_rule_table, "transactions")

forward_rules_low <- apriori(forward_rule_table, parameter=list(support =0.2, confidence =0.4, minlen=2, maxlen=10))
forward_rules_low <- subset(subset(forward_rules_low, subset = rhs %pin% "result=D"), lhs %pin% 'low')

forward_rules_high <- apriori(forward_rule_table, parameter=list(support =0.22, confidence =0.4, minlen=2, maxlen=10))
forward_rules_high <- subset(subset(forward_rules_high, subset = rhs %pin% "result=W"), lhs %pin% 'high')

# Defending attributes
inspect(sort(back_rules_high, by = 'lift', decreasing = TRUE))

##      lhs                                rhs      support
## [1] {liverpool_back.x_strength=very high} => {result=W} 0.1818182
## [2] {liverpool_back.y_dribbling=high}    => {result=W} 0.1515152
## [3] {liverpool_back.x_jumping=high}      => {result=W} 0.2121212
##      confidence lift      count
## [1] 0.7500000  1.455882  6
## [2] 0.7142857  1.386555  5
## [3] 0.7000000  1.358824  7

# Mid-field attributes
inspect(sort(center_rules_high, by = 'lift', decreasing = TRUE))

##      lhs                                rhs      support confidence      lift count
## [1] {liverpool_center.y_interceptions=very high} => {result=W} 0.1818182      0.75 1.455882      6
## [2] {liverpool_center.y_potential=high}          => {result=W} 0.1818182      0.75 1.455882      6
## [3] {liverpool_center.x_agility=high}            => {result=W} 0.1818182      0.75 1.455882      6

```

```
# Forward attributes
inspect(sort(forward_rules_high, by = 'lift', decreasing = TRUE))
```

##	lhs	rhs	support	confidence	lift	count
## [1]	{liverpool_forward.x_strength=high}	=> {result=W}	0.2424242	0.7272727	1.411765	8
## [2]	{liverpool_forward.x_crossing=high}	=> {result=W}	0.2424242	0.5714286	1.109244	8
## [3]	{liverpool_forward.x_heading_accuracy=high}	=> {result=W}	0.3636364	0.5217391	1.012788	12

Interpretation & Conclusions

However, it is a known fact that if our team is strong it will increase the probability of a win. From the association rules across different positions i.e. forward, mid and defence we observe that if we have a strong defence, moderately strong midfielders and strong forwards when compared to opponents, it can help us to perform better in such conditions.

Summary

With an analysis of opportunities to qualify into Champions League, we observed a trend that we don't win high probability mid-stage matches. It is uncommon for us to lose matches at home with high winning probability. Finally, we came up with the recommendation to bridge a gap with a strong defence or mid, we have a higher probability to win 2-3 more matches per season which can place us in the top 4. we can exploit this opportunity when we play with low stats team.

Influence of Internal and External factors

Influence of external factors in the team's performance

Description and Rationale for the Chosen Analysis

External factors play an important role in identifying the opportunities and threats to the performance of the team. The performance majorly depends upon the sentiment of the player which can be determined by the team's previous performance in the league like team's winning/losing streak, the result of the last played match, opponent's previous performances, Home/Away match, winning probability and so on. These factors cannot be altered but have a huge indirect influence in deciding the team's performance. So we have taken these external factors and tried to find the actions which the team should take in order to perform well in different scenarios.

The below analysis explores the relationship between the external factors and the performance of the team.

Execution, Results (including code) and Interpretation

[Commenting the code to save knitting time]

We have calculated the winning streak, the previous performance of Liverpool and our opponents and considered various scenarios based on the combination of external factors to build an association rule to find out the optimal formations that will lead the team to win the match.

```
# match <- match[,c(12:16,18:20)]

# match <- as.data.frame(lapply(match,function(x){as.factor(x)}))

# rules <- apriori(match, parameter=list(support =0.01, confidence =0.05, minlen=2))
```

```
# rules <- rules%>%subset(subset = (rhs %pin% "Win_or_Loss=W"))%>%
# subset(subset = lhs %pin% "Home_or_Away" & lhs %pin% "Win_prob" & lhs %pin% "lvp_form")%>%
# subset(subset = !(lhs %pin% "Win_prob=high"))%>%
# subset(subset = !(lhs %pin% "lvp_con_win=0"))%>%
# sort(by = 'lift', decreasing = TRUE)
```

Home/Away	Win probability	Opp_Pre_Perf	Lvp_Pre perf	lvp formations	Win/loss	Support	Confidence	Lift	Count
H	low	1	0	(4,4,2)	W	0.0164	1.0000	2.0267	5
A	medium	0	1	(4,2,3,1)	W	0.0164	1.0000	2.0267	5
A	medium	0	0	(3,4,3)	W	0.0230	0.6364	1.2897	7
A	medium	1	0	(4,4,2)	W	0.0164	0.8333	1.6889	5

Home/Away	Win probability	Opp_Pre_Perf	Lvp_Pre perf
H	low	1	0

Case-1: The situation where our winning probability is low in a home match, the opponent has won their previous match and we lost our previous match. By analysing the value of lift obtained from the association rules, we can confidently say that we can win the match if we play with **(4,4,2)** formation for the above scenario

Home/Away	Win probability	Opp_Pre_Perf	Lvp_Pre perf
A	medium	0	1

Case-2: The situation where our winning probability is medium in an away match, the opponent has lost their previous match and we also won our previous match. By analysing the value of lift obtained from the association rules, we can confidently say that we can win the match if we play with **(4,2,3,1)** formation for the above scenario

Home/Away	Win probability	Opp_Pre_Perf	Lvp_Pre perf
A	medium	0	0

Case-3: The situation where our winning probability is medium in an away match, the opponent has won their previous match and we lost our previous match. By analysing the value of lift obtained from the association rules, with 29 percent confidence, we can say that we can win the match if we play with **(3,4,3)** formation for the above scenario

Home/Away	Win probability	Opp_Pre_Perf	Lvp_Pre perf
A	medium	1	0

Case-4: The situation where our winning probability is medium in an away match, the opponent has won their previous match and we lost our last match. By analysing the value of lift obtained from the association rules, with 69 percent confidence, we can say that we can win the match if we play with **(4,4,2)** formation for the above scenario

Conclusion

Based on the results obtained through the association rules, it is clearly visible that : 1.(4,4,2) is the best suited formation in Case-1

2.(4,2,3,1) is the best suited formation in Case-2

3.(3,4,3) is the best suited formation in Case-3

4.(4,4,2) is the best suited formation I Case-4

Influence of internal factors in the team's performance

Description and Rationale for the Chosen Analysis

To study for the internal factors of a match, we considered the team formation and the average difference of players' attributes between our team and the opponents' in the same position. We only focus on the away matches since the team lost way more than in-home courts. Hence, we started to focus on those we won in away, and if we can find out some association rules determining to win, we can apply the tactics and enhance our winning probability.

Execution and Results (including code)

We first divided the dataset into four different positions and filtered out those with away matches.

Association rules with the goal keepers

[Commenting the code to save knitting time]

```
# rules_gka <- apriori(matches_gk_attr, parameter=list(supp = 0.01, conf=0.01, minlen=4))
#
# rules_gka %>% subset(subset = (rhs %pin% 'goal_diff_class=close win ' | (rhs %pin% 'goal_diff_class=w
# subset(subset = !(lhs %pin% 'win_prob_class')) %>%
# sort(by = c("lift", 'support'),decreasing=TRUE) %>% head(n=20) %>% inspect()
```

lhs	rhs	supp	conf	lift	count
liverpool_formation=(4, 2, 3, 1) liverpool_gk_diving=high liverpool_gk_positioning=high liverpool_gk_handling=medium	close win	0.01315789	1	5.62963	2

Association rules with the forwards

[Commenting the code to save knitting time]

```
#
# rules_fa <- apriori(matches_forward_attr, parameter=list(supp = 0.01, conf=0.01, minlen=4))
#
# rules_fa %>% subset(subset = (rhs %pin% 'goal_diff_class=close win ' | (rhs %pin% 'goal_diff_class=wi
# subset(subset = !(lhs %pin% 'win_prob_class')) %>%
# sort(by = c("lift", 'support'),decreasing=TRUE) %>% head(n=20) %>% inspect()
```

lhs	rhs	supp	conf	lift	count
liverpool_forward_ball_control=high liverpool_forward_sliding_tackle=very high liverpool_forward_heading_accuracy=high liverpool_forward_marking=very high liverpool_forward_curve=very high	close win	0.03289474	1	5.62963	5

Association rules with the center/mid

[Commenting the code to save knitting time]

```
# rules_ca <- apriori(matches_center_attr, parameter=list(supp = 0.01, conf=0.01, minlen=4))  
#  
# rules_ca %>% subset(subset = (rhs %pin% 'goal_diff_class=close win ' | (rhs %pin% 'goal_diff_class=wi  
# subset(subset = !(lhs %pin% 'win_prob_class')) %>%  
# sort(by = c("lift", 'support'),decreasing=TRUE) %>% head(n=20) %>% inspect()
```

lhs	rhs	supp	conf	lift	count
liverpool_center_penalties=medium liverpool_center_long_shots=medium liverpool_center_ball_control=medium liverpool_center.x_aggression=very low	close win	0.04605263	0.875	4.925926	7

Association rules with the back/defence

[Commenting the code to save knitting time]

```
# rules_ba <- apriori(matches_back_attr, parameter=list(supp = 0.01, conf=0.01, minlen=4))  
#  
# rules_ba %>% subset(subset = (rhs %pin% 'goal_diff_class=close win ' | (rhs %pin% 'goal_diff_class=wi  
# subset(subset = !(lhs %pin% 'win_prob_class')) %>%  
# sort(by = c("lift", 'support'),decreasing=TRUE) %>% head(n=20) %>% inspect()
```

lhs	rhs	supp	conf	lift	count
liverpool_back_free_kick_accuracy=medium liverpool_back_curve=medium liverpool_back_heading_accuracy=medium	close win	0.04605263	0.5833333	3.283951	7

Interpretation

For the goal keepers, we can see that the combination of (4,2,3,1) and high defensive skills have a lift of 5.63. We can state we are confident to win in away matches with this kind of formation and goal keepers with better defense.

When choosing our forward players for the match, the results show that if players with high accuracy heading and nice defensive skills, it is more likely to win over.

For our center players and back players, we do not need them to perform extraordinarily better than the opponents. In the results we can see that even their attributes were just matching the opponents, we still have a higher chance to win an away game.

Conclusions

To conclude, we can say that the goalkeepers and the forward players are the more important positions on the court when we are playing away. By placing our good defensive goalkeepers and all-rounded forward players, we can win more games in a way courts then being in a better place in the league.

Summary

From the above analysis, we tried to find how external and internal factors impact the overall performance of the team. With the help of external factors, we were able to find the optimal formation required for playing matches on certain scenarios and with the help of internal factors, we determined the required player attributes compared to the opponent's attributes.

Final Takeaways

What have we found?

- 1.Liverpool performing better in home when compared to away matches.
- 2.We are not performing as expected in high and medium probability matches, especially in home.
- 3.When compared to early and later stages of the league, our performance is relatively bad in the mid stage matches.
- 4.By analysing the player and team attributes, we found that forward players and goalkeepers have huge influences in the team's performance
- 5.Previous performance of our team and the opponent has an impact in the sentiment of the players which inturns decides the performance of the team in that particular match.

Final Recommendation

To qualify for the champions league, We should focus more on performing well in the mid-stage matches. Within these important matches, under the scenarios when we play our opponents who have good previous performances, we should play more with formation of (4,4,2) for home matches and we should play with the formation of (4,2,3,1) for away matches, with having good defensive goalkeepers, moderately strong back and midfielders and all-rounded forward players on the field.