AS Roma & Winning Elevens

Danny Moncada

September 29th, 2019

Business Problem:

We have been hired by AS Roma as their resident data scientist to find patterns that our coach can exploit to increase success on the field and decrease failure. Specifically, he is looking for non-obvious patterns because he has learned all of the obvious ones through experience. He has asked us to use associations rules for this analysis and we have been provided data (euro_soccer.sqlite) to conduct our tests.

We have decided to take a multi-stepped approach to our proposed challenge. First, we want to determine what traits are specific to "successful" teams and what traits are specific to "losing" teams. We are going to do this by analyzing the top 4 teams in the top 4 leagues (16 teams total) and by analyzing the bottom 4 teams in the top 4 leagues (16 teams). We have chosen to focus on the top 4 leagues for determining these overall trends because AS Roma is part of Serie A, which is a competitive league. If we had focused on lower-tier leagues, there would be the potential for associations to appear that could not be applied to AS Roma.

After determining what "successful" and "losing" teams do in the top leagues (formations, # of shots, penalties, defense focus, offense focus, etc.), we will shift our attention to what makes great players. Though we would all love to say to simply hire Ronaldo or Messi, we understand that this simple solution is not feasible for AS Roma. However, we may learn insights into what types of players "successful" teams are most looking for. These characteristics may be speed, attacking rating, defense rating, crossing, heading, short pass, etc. After determining what characteristics make up the best players, which in turn make the best teams, we will move onto compiling our conclusions.

With these recommendations, we hope to tell our coach what hidden trends can help AS Roma become a more successful team. We hope to identify key characteristics in players, as well as key insights into team play that can lead AS Roma to victory.

Specifically, we hope to come up with actions that the Manager of Roma can take to get Roma in the top 4 teams in their Series so that they can go onto the Championship.

Definition of Success:

The top 4 teams in each league go on to the Championship which is tied to financial reward and prestige; finishing each season in the top 4 teams is what our team is defining as success. While it is hard to guarantee a top 4 finish, we are aiming to greatly increase the likelihood of Roma finishing there.

An Overview of team statistics

```
## Warning: package 'ggplot2' was built under R version 3.6.2
```

```
## Warning: package 'dplyr' was built under R version 3.6.2
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

## Warning: package 'arules' was built under R version 3.6.2

## Loading required package: Matrix

##
## Attaching package: 'arules'

## The following object is masked from 'package:dplyr':
##
##     recode

## The following objects are masked from 'package:base':
##
##     abbreviate, write

## Warning: package 'lubridate' was built under R version 3.6.2

##
## Attaching package: 'lubridate'

## The following object is masked from 'package:base':
##
##     date

## Warning: package 'sqldf' was built under R version 3.6.3

## Loading required package: gsubfn

## Warning: package 'gsubfn' was built under R version 3.6.2

## Loading required package: proto

## Warning: package 'proto' was built under R version 3.6.2

## Loading required package: RSQLite
```

```r
## Get country and league names and have a final master match table

master_df <- sqldf("SELECT m.id, c.name AS country, l.name AS league, season, date, home_team, away_tea
                   game_winner, game_loser, winner_goals, loser_goals, year, month
                   FROM match_away_teams m
                   INNER JOIN country_df c
                   ON m.country_id = c.id
                   INNER JOIN league_df l
                   ON m.league_id = l.id")

head(master_df)
```

```
##   id country                 league    season       date         home_team
## 1  1 Belgium Belgium Jupiler League 2008/2009 2008-08-17          KRC Genk
## 2  2 Belgium Belgium Jupiler League 2008/2009 2008-08-16  SV Zulte-Waregem
## 3  3 Belgium Belgium Jupiler League 2008/2009 2008-08-16 KSV Cercle Brugge
## 4  4 Belgium Belgium Jupiler League 2008/2009 2008-08-17          KAA Gent
## 5  5 Belgium Belgium Jupiler League 2008/2009 2008-08-16     FCV Dender EH
## 6  6 Belgium Belgium Jupiler League 2008/2009 2008-09-24       KV Mechelen
##           away_team home_team_goal away_team_goal game_result total_goals
## 1      Beerschot AC              1              1         tie           2
## 2  Sporting Lokeren              0              0         tie           0
## 3    RSC Anderlecht              0              3        away           3
## 4         RAEC Mons              5              0        home           5
## 5 Standard de Liège              1              3        away           4
## 6    Club Brugge KV              1              1         tie           2
##   game_winner game_loser winner_goals loser_goals year month
## 1         tie        tie            1           1 2008     8
## 2         tie        tie            0           0 2008     8
## 3        8635       9984            3           0 2008     8
## 4        9991       9998            5           0 2008     8
## 5        9985       7947            3           1 2008     8
## 6         tie        tie            1           1 2008     9
```

This is a base table for different types of analyses. We can use this to get counts of winners by Season/League - who are top performing teams?

```
## Show the top four leagues - by home winners
best_home_league[1:4,]
```

```
## # A tibble: 4 x 2
##   league                     n
##   <chr>                  <int>
## 1 England Premier League   160
## 2 France Ligue 1           160
## 3 Italy Serie A            160
## 4 Spain LIGA BBVA          160
```

Interpretation:

The 4 leagues with the most home team wins are England Premier League, France Ligue 1, Italy Serie A, and Spain LIGA BBVA.

```
## Next we'll filter our matches data to only have matches from these leagues
## We don't care about anything else

best_leagues <- c(best_home_league$league[1:4])

best_league_matches <- sqldf("SELECT * FROM master_df
                             WHERE league IN ('England Premier League', 'France Ligue 1', 'Italy Serie A',
                             'Spain LIGA BBVA')")

head(best_league_matches)
```

```
##     id country                 league    season       date
```

```
## 1 1729 England England Premier League 2008/2009 2008-08-17
## 2 1730 England England Premier League 2008/2009 2008-08-16
## 3 1731 England England Premier League 2008/2009 2008-08-16
## 4 1732 England England Premier League 2008/2009 2008-08-16
## 5 1733 England England Premier League 2008/2009 2008-08-17
## 6 1734 England England Premier League 2008/2009 2008-08-16
##           home_team            away_team home_team_goal away_team_goal
## 1 Manchester United     Newcastle United              1              1
## 2           Arsenal West Bromwich Albion              1              0
## 3        Sunderland            Liverpool              0              1
## 4   West Ham United       Wigan Athletic              2              1
## 5       Aston Villa      Manchester City              4              2
## 6           Everton     Blackburn Rovers              2              3
##   game_result total_goals game_winner game_loser winner_goals loser_goals
## 1         tie           2         tie        tie            1           1
## 2        home           1        9825       8659            1           0
## 3        away           1        8650       8472            1           0
## 4        home           3        8654       8528            2           1
## 5        home           6       10252       8456            4           2
## 6        away           5        8655       8668            3           2
##   year month
## 1 2008     8
## 2 2008     8
## 3 2008     8
## 4 2008     8
## 5 2008     8
## 6 2008     8
```

```
## OK!  Now we're getting somewhere.  We have a data set with just the league matches
## that we care about - our top four leagues.

## Now we need to see who our top four teams for each league are - we can do this
## by season

## Let's just get counts of wins - regardless of home or away - and sort by team
all_wins_df <- sqldf("SELECT season, league, CASE WHEN game_result = 'home' then home_team WHEN game_re
                     FROM best_league_matches")

head(all_wins_df)
```

```
##      season                 league           result
## 1 2008/2009 England Premier League              tie
## 2 2008/2009 England Premier League          Arsenal
## 3 2008/2009 England Premier League        Liverpool
## 4 2008/2009 England Premier League  West Ham United
## 5 2008/2009 England Premier League      Aston Villa
## 6 2008/2009 England Premier League Blackburn Rovers
```

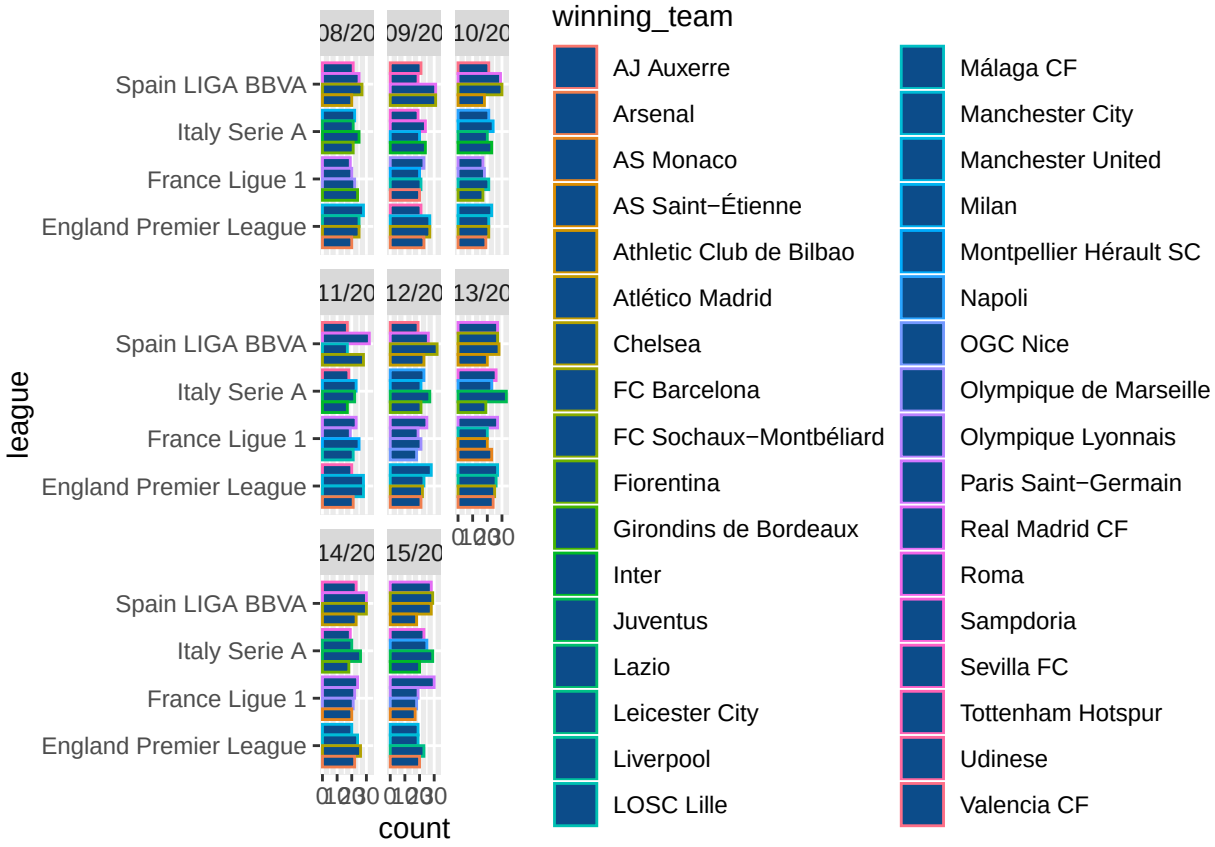
```
## Get the winning teams for each season and league
top_teams_by_wins_df <- sqldf("SELECT season, league, result as winning_team, count(result) as num_wins
                              FROM all_wins_df
                              WHERE result != 'tie'
                              GROUP by season, league, winning_team
```

```
                            ORDER BY season, league, num_wins desc")

## Create a league position column so that we can just grab the toup four of each season
top_teams_by_wins_df <- top_teams_by_wins_df %>%
  group_by(season, league) %>%
  mutate(league_position = rank(-num_wins, ties.method = "first"))

top_four_teams <- filter(top_teams_by_wins_df, league_position <= 4)

ggplot(top_four_teams) +
  aes(x = league, colour = winning_team, weight = num_wins) +
  geom_bar(position = "dodge", fill = "#0c4c8a") +
  scale_color_hue() +
  coord_flip() +
  theme_gray() +
  facet_wrap(vars(season))
```




```
## OK So I have a list of teams that are consistently peforming well, top four in each league
## I want to see what attributes their team consistents of
## Let's get the list of team names

## I subset my winning teams
winning_teams <- unique(top_four_teams$winning_team)

## only need the api_id and team_name
```

```r
team_cols <- c("team_api_id", "team_long_name")

## A lot of transformations at once!

## Filter out just the winning teams
winning_teams_df <- filter(team_df, team_long_name %in% winning_teams)

## Filter out on the columns that I want
winning_teams_df <- winning_teams_df[team_cols]

## Merge the data sets together - last thing I will do is a look-up to convert the date to a league sea
##  I want to filter on the league seasons where the team actually did well
## Then I will create summary statistics on this

## Look at the Team Attribute DF
head(team_atts_df)
```

```
##   id team_fifa_api_id team_api_id                date buildUpPlaySpeed
## 1  1              434        9930 2010-02-22 00:00:00                60
## 2  2              434        9930 2014-09-19 00:00:00                52
## 3  3              434        9930 2015-09-10 00:00:00                47
## 4  4               77        8485 2010-02-22 00:00:00                70
## 5  5               77        8485 2011-02-22 00:00:00                47
## 6  6               77        8485 2012-02-22 00:00:00                58
##   buildUpPlaySpeedClass buildUpPlayDribbling buildUpPlayDribblingClass
## 1              Balanced                   NA                     Little
## 2              Balanced                   48                     Normal
## 3              Balanced                   41                     Normal
## 4                  Fast                   NA                     Little
## 5              Balanced                   NA                     Little
## 6              Balanced                   NA                     Little
##   buildUpPlayPassing buildUpPlayPassingClass buildUpPlayPositioningClass
## 1                 50                   Mixed                   Organised
## 2                 56                   Mixed                   Organised
## 3                 54                   Mixed                   Organised
## 4                 70                    Long                   Organised
## 5                 52                   Mixed                   Organised
## 6                 62                   Mixed                   Organised
##   chanceCreationPassing chanceCreationPassingClass chanceCreationCrossing
## 1                    60                    Normal                     65
## 2                    54                    Normal                     63
## 3                    54                    Normal                     63
## 4                    70                     Risky                     70
## 5                    53                    Normal                     48
## 6                    45                    Normal                     70
##   chanceCreationCrossingClass chanceCreationShooting
## 1                      Normal                    55
## 2                      Normal                    64
## 3                      Normal                    64
## 4                        Lots                    70
## 5                      Normal                    52
## 6                        Lots                    55
##   chanceCreationShootingClass chanceCreationPositioningClass
```

```
## 1                     Normal                     Organised
## 2                     Normal                     Organised
## 3                     Normal                     Organised
## 4                       Lots                     Organised
## 5                     Normal                     Organised
## 6                     Normal                     Organised
##   defencePressure defencePressureClass defenceAggression
## 1              50               Medium                55
## 2              47               Medium                44
## 3              47               Medium                44
## 4              60               Medium                70
## 5              47               Medium                47
## 6              40               Medium                40
##   defenceAggressionClass defenceTeamWidth defenceTeamWidthClass
## 1                  Press               45                Normal
## 2                  Press               54                Normal
## 3                  Press               54                Normal
## 4                 Double               70                  Wide
## 5                  Press               52                Normal
## 6                  Press               60                Normal
##   defenceDefenderLineClass
## 1                    Cover
## 2                    Cover
## 3                    Cover
## 4                    Cover
## 5                    Cover
## 6                    Cover
```

```
## Reconfigure the date field so I can create a "season" flag
team_atts_df$date <- as.Date(team_atts_df$date, "%Y-%m-%d %H:%M:%S")
team_atts_df$year<-year(as.POSIXlt(team_atts_df$date, format="%Y-%m-%d"))
team_atts_df$month<-month(as.POSIXlt(team_atts_df$date, format="%Y-%m-%d"))

head(team_atts_df)
```

```
##   id team_fifa_api_id team_api_id       date buildUpPlaySpeed
## 1  1              434        9930 2010-02-22               60
## 2  2              434        9930 2014-09-19               52
## 3  3              434        9930 2015-09-10               47
## 4  4               77        8485 2010-02-22               70
## 5  5               77        8485 2011-02-22               47
## 6  6               77        8485 2012-02-22               58
##   buildUpPlaySpeedClass buildUpPlayDribbling buildUpPlayDribblingClass
## 1              Balanced                   NA                    Little
## 2              Balanced                   48                    Normal
## 3              Balanced                   41                    Normal
## 4                  Fast                   NA                    Little
## 5              Balanced                   NA                    Little
## 6              Balanced                   NA                    Little
##   buildUpPlayPassing buildUpPlayPassingClass buildUpPlayPositioningClass
## 1                 50                  Mixed                   Organised
## 2                 56                  Mixed                   Organised
## 3                 54                  Mixed                   Organised
## 4                 70                   Long                   Organised
```

```
## 5                 52                  Mixed                  Organised
## 6                 62                  Mixed                  Organised
##   chanceCreationPassing chanceCreationPassingClass chanceCreationCrossing
## 1                    60                    Normal                     65
## 2                    54                    Normal                     63
## 3                    54                    Normal                     63
## 4                    70                     Risky                     70
## 5                    53                    Normal                     48
## 6                    45                    Normal                     70
##   chanceCreationCrossingClass chanceCreationShooting
## 1                      Normal                    55
## 2                      Normal                    64
## 3                      Normal                    64
## 4                        Lots                    70
## 5                      Normal                    52
## 6                        Lots                    55
##   chanceCreationShootingClass chanceCreationPositioningClass
## 1                      Normal                      Organised
## 2                      Normal                      Organised
## 3                      Normal                      Organised
## 4                        Lots                      Organised
## 5                      Normal                      Organised
## 6                      Normal                      Organised
##   defencePressure defencePressureClass defenceAggression
## 1              50               Medium                55
## 2              47               Medium                44
## 3              47               Medium                44
## 4              60               Medium                70
## 5              47               Medium                47
## 6              40               Medium                40
##   defenceAggressionClass defenceTeamWidth defenceTeamWidthClass
## 1                  Press               45                Normal
## 2                  Press               54                Normal
## 3                  Press               54                Normal
## 4                 Double               70                  Wide
## 5                  Press               52                Normal
## 6                  Press               60                Normal
##   defenceDefenderLineClass year month
## 1                    Cover 2010     2
## 2                    Cover 2014     9
## 3                    Cover 2015     9
## 4                    Cover 2010     2
## 5                    Cover 2011     2
## 6                    Cover 2012     2
```

```
## Grab the season so that I can merge on successful teams and the season they were winners
team_att_season_df <- sqldf("SELECT DISTINCT t.id, team_fifa_api_id, t.team_api_id, team_long_name as te
                                buildUpPlayPassing, buildUpPlayPassingClass, buildUpPlayPositioningClass
                            chanceCreationCrossing, chanceCreationCrossingClass, chanceCreationShooting
                              defencePressure, defencePressureClass, defenceAggression, defenceAggressio
                            defenceTeamWidthClass, defenceDefenderLineClass, t.year, t.month, m.season
                              FROM team_atts_df t
                              INNER JOIN matches_df m
                              ON t.year = m.year
```

```
                              AND t.month = m.month
                              INNER JOIN winning_teams_df w
                              ON w.team_api_id = t.team_api_id
                              ORDER BY season, team")

## Take a look at the data and some sample columns

head(team_att_season_df)
```

```
##     id team_fifa_api_id team_api_id                   team
## 1  107               57        8583             AJ Auxerre
## 2  836               69        9829              AS Monaco
## 3 1119             1819        9853       AS Saint-Étienne
## 4   71                1        9825                Arsenal
## 5   89              448        8315 Athletic Club de Bilbao
## 6   95              240        9906        Atlético Madrid
##   buildUpPlaySpeed buildUpPlaySpeedClass buildUpPlayDribbling
## 1               30                  Slow                   NA
## 2               35              Balanced                   NA
## 3               65              Balanced                   NA
## 4               66              Balanced                   NA
## 5               60              Balanced                   NA
## 6               64              Balanced                   NA
##   buildUpPlayPassing buildUpPlayPassingClass buildUpPlayPositioningClass
## 1                 70                   Long                   Organised
## 2                 55                  Mixed                   Organised
## 3                 65                  Mixed                   Organised
## 4                 30                  Short                   Free Form
## 5                 45                  Mixed                   Organised
## 6                 30                  Short                   Free Form
##   chanceCreationPassing chanceCreationPassingClass chanceCreationCrossing
## 1                    35                    Normal                     65
## 2                    65                    Normal                     55
## 3                    55                    Normal                     60
## 4                    30                      Safe                     45
## 5                    35                    Normal                     70
## 6                    65                    Normal                     50
##   chanceCreationCrossingClass chanceCreationShooting
## 1                      Normal                    35
## 2                      Normal                    45
## 3                      Normal                    60
## 4                      Normal                    35
## 5                        Lots                    55
## 6                      Normal                    70
##   chanceCreationShootingClass chanceCreationPositioningClass
## 1                      Normal                      Organised
## 2                      Normal                      Organised
## 3                      Normal                      Organised
## 4                      Normal                      Free Form
## 5                      Normal                      Organised
## 6                        Lots                      Free Form
##   defencePressure defencePressureClass defenceAggression
## 1              30                 Deep                55
```

```
## 2              70                 High                65
## 3              30                 Deep                30
## 4              30                 Deep                40
## 5              30                 Deep                70
## 6              70                 High                34
##   defenceAggressionClass defenceTeamWidth defenceTeamWidthClass
## 1                  Press               30                Narrow
## 2                  Press               70                  Wide
## 3                Contain               30                Narrow
## 4                  Press               50                Normal
## 5                 Double               35                Normal
## 6                  Press               55                Normal
##   defenceDefenderLineClass year month    season
## 1                    Cover 2010     2 2009/2010
## 2                    Cover 2010     2 2009/2010
## 3             Offside Trap 2010     2 2009/2010
## 4                    Cover 2010     2 2009/2010
## 5                    Cover 2010     2 2009/2010
## 6             Offside Trap 2010     2 2009/2010
```

```
## OK I have a huge list of columns to the right hand side of season and team
## right hand side

## Let's make some rules

## Attacking Team Attributes - remove buildUpPlayDribbling due to NULL values
attack <- c('buildUpPlaySpeed', 'buildUpPlaySpeedClass', 'buildUpPlayPassing',
            'buildUpPlayPassingClass', 'buildUpPlayPositioningClass')

## Midfield Team Attributes
midfield <- c('chanceCreationPassing', 'chanceCreationPassingClass', 'chanceCreationCrossing', 'chanceC
              'chanceCreationShooting', 'chanceCreationShootingClass', 'chanceCreationPositioningClass'

## Defense Team Attributes
defense <- c('defencePressure', 'defencePressureClass', 'defenceAggression', 'defenceAggressionClass',
             'defenceTeamWidth', 'defenceTeamWidthClass', 'defenceDefenderLineClass')



attack_df <- team_att_season_df[, attack]

midfield_df <- team_att_season_df[, midfield]

defense_df <- team_att_season_df[, defense]

## Summary of top four teams' offensive numeric attributes
## Average build up play speed is 52/100 and passing (accuracy) is 45/100 - Roma can do better here!
summary(dplyr::select_if(attack_df, is.numeric))
```

```
##  buildUpPlaySpeed buildUpPlayPassing
##  Min.   :20.00    Min.   :20.0
##  1st Qu.:45.00    1st Qu.:36.0
##  Median :50.50    Median :47.0
##  Mean   :52.46    Mean   :45.6
```

```
##  3rd Qu.:63.25    3rd Qu.:52.0
##  Max.   :78.00    Max.   :70.0
```

```
# buildUpPlaySpeed buildUpPlayPassing
# Min.   :20.00    Min.   :20.0
# 1st Qu.:45.00    1st Qu.:36.0
# Median :50.50    Median :47.0
# Mean   :52.46    Mean   :45.6
# 3rd Qu.:63.25    3rd Qu.:52.0
# Max.   :78.00    Max.   :70.0

## Balanced speed for building up play, mix of long/short passing for build up play
## build up position is organized - teams that are successful play with balance (can we exploit this wi
dplyr::select_if(attack_df, is.character) %>%
  group_by(buildUpPlaySpeedClass, buildUpPlayPassingClass, buildUpPlayPositioningClass) %>%
  summarise(Freq = n()) %>%
  arrange(-Freq)
```

```
## # A tibble: 14 x 4
## # Groups:   buildUpPlaySpeedClass, buildUpPlayPassingClass [9]
##    buildUpPlaySpeedCla~ buildUpPlayPassingCl~ buildUpPlayPositioning~  Freq
##    <chr>                <chr>                 <chr>                   <int>
##  1 Balanced             Mixed                 Organised                 132
##  2 Fast                 Mixed                 Organised                  17
##  3 Balanced             Mixed                 Free Form                  13
##  4 Balanced             Short                 Organised                  10
##  5 Balanced             Short                 Free Form                   8
##  6 Fast                 Long                  Organised                   5
##  7 Fast                 Mixed                 Free Form                   4
##  8 Fast                 Short                 Free Form                   3
##  9 Slow                 Mixed                 Organised                   3
## 10 Slow                 Short                 Free Form                   3
## 11 Slow                 Short                 Organised                   3
## 12 Balanced             Long                  Organised                   1
## 13 Fast                 Short                 Organised                   1
## 14 Slow                 Long                  Organised                   1
```

```
#    buildUpPlaySpeedClass buildUpPlayPassingClass buildUpPlayPositioningClass  Freq
#  1 Balanced              Mixed                   Organised                    132
#  2 Fast                  Mixed                   Organised                     17

head(midfield_df)
```

```
##   chanceCreationPassing chanceCreationPassingClass chanceCreationCrossing
## 1                    35                     Normal                     65
## 2                    65                     Normal                     55
## 3                    55                     Normal                     60
## 4                    30                       Safe                     45
## 5                    35                     Normal                     70
## 6                    65                     Normal                     50
##   chanceCreationCrossingClass chanceCreationShooting
## 1                      Normal                     35
## 2                      Normal                     45
```

```
## 3                      Normal                     60
## 4                      Normal                     35
## 5                        Lots                     55
## 6                      Normal                     70
##   chanceCreationShootingClass chanceCreationPositioningClass
## 1                      Normal                      Organised
## 2                      Normal                      Organised
## 3                      Normal                      Organised
## 4                      Normal                      Free Form
## 5                      Normal                      Organised
## 6                        Lots                      Free Form
```

```
## Summary of top four teams' midfield numeric attributes
## Average chance creation from passing is 53/100, from crosses is 55/100, and shooting from midfield i
summary(dplyr::select_if(midfield_df, is.numeric))
```

```
##  chanceCreationPassing chanceCreationCrossing chanceCreationShooting
##  Min.   :28.00         Min.   :20.00          Min.   :23.00
##  1st Qu.:46.00         1st Qu.:50.00          1st Qu.:50.00
##  Median :50.50         Median :56.00          Median :54.50
##  Mean   :53.19         Mean   :55.58          Mean   :55.75
##  3rd Qu.:65.00         3rd Qu.:65.00          3rd Qu.:67.00
##  Max.   :77.00         Max.   :80.00          Max.   :80.00
```

```
# chanceCreationPassing chanceCreationCrossing chanceCreationShooting
# Min.   :28.00         Min.   :20.00          Min.   :23.00
# 1st Qu.:46.00         1st Qu.:50.00          1st Qu.:50.00
# Median :50.50         Median :56.00          Median :54.50
# Mean   :53.19         Mean   :55.58          Mean   :55.75
# 3rd Qu.:65.00         3rd Qu.:65.00          3rd Qu.:67.00
# Max.   :77.00         Max.   :80.00          Max.   :80.00

## Successful teams create chances from midfield normally, from cross normal, and shooting opportunitie
## And are organized in there positioning.  Nothing too suprising about that.
dplyr::select_if(midfield_df, is.character) %>%
  group_by(chanceCreationPassingClass, chanceCreationCrossingClass, chanceCreationShootingClass, chance
  summarise(Freq = n()) %>%
  arrange(-Freq)
```

```
## # A tibble: 27 x 5
## # Groups:   chanceCreationPassingClass, chanceCreationCrossingClass,
## #   chanceCreationShootingClass [15]
##    chanceCreationP~ chanceCreationC~ chanceCreationS~ chanceCreationP~
##    <chr>            <chr>            <chr>            <chr>
##  1 Normal           Normal           Normal           Organised
##  2 Normal           Normal           Lots             Organised
##  3 Normal           Normal           Normal           Free Form
##  4 Normal           Lots             Normal           Organised
##  5 Normal           Normal           Lots             Free Form
##  6 Risky            Lots             Normal           Organised
##  7 Risky            Normal           Lots             Free Form
##  8 Risky            Normal           Normal           Organised
##  9 Risky            Normal           Lots             Organised
```

```
## 10 Normal           Little           Normal           Free Form
## # ... with 17 more rows, and 1 more variable: Freq <int>
```

```
#   chanceCreationPassingClass chanceCreationCrossingClass chanceCreationShootingClass chanceCreationPo
# 1 Normal                     Normal                     Normal                       Organised
# 2 Normal                     Normal                     Lots                         Organised

head(defense_df)
```

```
##   defencePressure defencePressureClass defenceAggression
## 1              30                 Deep                55
## 2              70                 High                65
## 3              30                 Deep                30
## 4              30                 Deep                40
## 5              30                 Deep                70
## 6              70                 High                34
##   defenceAggressionClass defenceTeamWidth defenceTeamWidthClass
## 1                  Press               30                Narrow
## 2                  Press               70                  Wide
## 3                Contain               30                Narrow
## 4                  Press               50                Normal
## 5                 Double               35                Normal
## 6                  Press               55                Normal
##   defenceDefenderLineClass
## 1                    Cover
## 2                    Cover
## 3             Offside Trap
## 4                    Cover
## 5                    Cover
## 6             Offside Trap
```

```
## Summary of top four teams' defense numeric attributes
summary(dplyr::select_if(defense_df, is.numeric))
```

```
##  defencePressure defenceAggression defenceTeamWidth
##  Min.   :23.00   Min.   :30.00     Min.   :30.00
##  1st Qu.:40.00   1st Qu.:45.00     1st Qu.:49.00
##  Median :49.00   Median :49.00     Median :53.50
##  Mean   :48.36   Mean   :50.67     Mean   :53.39
##  3rd Qu.:56.00   3rd Qu.:57.00     3rd Qu.:60.00
##  Max.   :70.00   Max.   :70.00     Max.   :70.00
```

```
# defencePressure defenceAggression defenceTeamWidth
# Min.   :23.00   Min.   :30.00     Min.   :30.00
# 1st Qu.:40.00   1st Qu.:45.00     1st Qu.:49.00
# Median :49.00   Median :49.00     Median :53.50
# Mean   :48.36   Mean   :50.67     Mean   :53.39
# 3rd Qu.:56.00   3rd Qu.:57.00     3rd Qu.:60.00
# Max.   :70.00   Max.   :70.00     Max.   :70.00

dplyr::select_if(defense_df, is.character) %>%
  group_by(defencePressureClass, defenceAggressionClass, defenceTeamWidthClass, defenceDefenderLineClass
```

```
  summarise(Freq = n()) %>%
  arrange(-Freq)
```

```
## # A tibble: 22 x 5
## # Groups:   defencePressureClass, defenceAggressionClass,
## #   defenceTeamWidthClass [14]
##    defencePressure~ defenceAggressi~ defenceTeamWidt~ defenceDefender~
##    <chr>            <chr>            <chr>            <chr>
##  1 Medium           Press            Normal           Cover
##  2 Medium           Press            Normal           Offside Trap
##  3 Medium           Press            Wide             Cover
##  4 Deep             Press            Normal           Cover
##  5 Medium           Double           Normal           Cover
##  6 Medium           Contain          Normal           Offside Trap
##  7 Deep             Double           Normal           Cover
##  8 High             Double           Wide             Cover
##  9 High             Double           Wide             Offside Trap
## 10 High             Press            Wide             Cover
## # ... with 12 more rows, and 1 more variable: Freq <int>
```

```
#   defencePressureClass defenceAggressionClass defenceTeamWidthClass defenceDefenderLineClass  Freq
# 1 Medium               Press                  Normal                Cover                      151
# 2 Medium               Press                  Normal                Offside Trap                12

## Now that we've analyzed our successful teams, we should check what the team attributes say about Rom
## We want to see if there is area for improvement, either in attack, midfield, or defense

## Get team attributes for AS Roma
roma_att_season_df <- sqldf("SELECT DISTINCT t.id, t.team_fifa_api_id, t.team_api_id, team_long_name as
                                buildUpPlayPassing, buildUpPlayPassingClass, buildUpPlayPositioningClas
                            chanceCreationCrossing, chanceCreationCrossingClass, chanceCreationShooting
                            defencePressure, defencePressureClass, defenceAggression, defenceAggression
                            defenceTeamWidthClass, defenceDefenderLineClass, t.year, t.month, m.season
                            FROM team_atts_df t
                            INNER JOIN matches_df m
                            ON t.year = m.year
                            AND t.month = m.month
                            INNER JOIN team_df tm
                            ON tm.team_api_id = t.team_api_id
                            WHERE tm.team_long_name = 'Roma'
                            ORDER BY season, team")


head(roma_att_season_df)
```

```
##     id team_fifa_api_id team_api_id team buildUpPlaySpeed
## 1 1104               52        8686 Roma               70
## 2 1105               52        8686 Roma               64
## 3 1106               52        8686 Roma               60
## 4 1107               52        8686 Roma               53
## 5 1108               52        8686 Roma               67
## 6 1109               52        8686 Roma               68
##   buildUpPlaySpeedClass buildUpPlayPassing buildUpPlayPassingClass
```

```
## 1                 Fast                30                 Short
## 2             Balanced                30                 Short
## 3             Balanced                40                 Mixed
## 4             Balanced                55                 Mixed
## 5                 Fast                38                 Mixed
## 6                 Fast                47                 Mixed
##   buildUpPlayPositioningClass chanceCreationPassing
## 1                   Free Form                    60
## 2                   Organised                    67
## 3                   Free Form                    65
## 4                   Organised                    77
## 5                   Organised                    73
## 6                   Organised                    71
##   chanceCreationPassingClass chanceCreationCrossing
## 1                     Normal                     60
## 2                      Risky                     50
## 3                     Normal                     35
## 4                      Risky                     75
## 5                      Risky                     64
## 6                      Risky                     54
##   chanceCreationCrossingClass chanceCreationShooting
## 1                      Normal                    70
## 2                      Normal                    65
## 3                      Normal                    50
## 4                        Lots                    56
## 5                      Normal                    56
## 6                      Normal                    66
##   chanceCreationShootingClass chanceCreationPositioningClass
## 1                        Lots                       Free Form
## 2                      Normal                       Free Form
## 3                      Normal                       Free Form
## 4                      Normal                       Free Form
## 5                      Normal                       Free Form
## 6                      Normal                       Free Form
##   defencePressure defencePressureClass defenceAggression
## 1              55               Medium                30
## 2              46               Medium                47
## 3              35               Medium                45
## 4              49               Medium                47
## 5              36               Medium                57
## 6              36               Medium                57
##   defenceAggressionClass defenceTeamWidth defenceTeamWidthClass
## 1                Contain               70                  Wide
## 2                  Press               50                Normal
## 3                  Press               50                Normal
## 4                  Press               53                Normal
## 5                  Press               34                Normal
## 6                  Press               48                Normal
##   defenceDefenderLineClass year month    season
## 1             Offside Trap 2010     2 2009/2010
## 2             Offside Trap 2011     2 2010/2011
## 3             Offside Trap 2012     2 2011/2012
## 4             Offside Trap 2013     9 2013/2014
## 5                    Cover 2014     9 2014/2015
```

```
## 6                   Cover 2015     9 2015/2016
```

```
## Subset each set of attributes similarly to the top 4 teams above
roma_attack_df <- roma_att_season_df[, attack]

## Summary of offensive numeric attributes
## Average build up play speed is 63/100 and passing (accuracy) is 40/100 - area for improvement?
summary(dplyr::select_if(roma_attack_df, is.numeric))
```

```
##  buildUpPlaySpeed buildUpPlayPassing
##  Min.   :53.00    Min.   :30.00
##  1st Qu.:61.00    1st Qu.:32.00
##  Median :65.50    Median :39.00
##  Mean   :63.67    Mean   :40.00
##  3rd Qu.:67.75    3rd Qu.:45.25
##  Max.   :70.00    Max.   :55.00
```

```
# buildUpPlaySpeed buildUpPlayPassing
# Min.   :53.00    Min.   :30.00
# 1st Qu.:61.00    1st Qu.:32.00
# Median :65.50    Median :39.00
# Mean   :63.67    Mean   :40.00
# 3rd Qu.:67.75    3rd Qu.:45.25
# Max.   :70.00    Max.   :55.00

## Roma tends to play fast, with a mixed set of passing, but keep an organized shape
dplyr::select_if(roma_attack_df, is.character) %>%
  group_by(buildUpPlaySpeedClass, buildUpPlayPassingClass, buildUpPlayPositioningClass) %>%
  summarise(Freq = n()) %>%
  arrange(-Freq)
```

```
## # A tibble: 5 x 4
## # Groups:   buildUpPlaySpeedClass, buildUpPlayPassingClass [4]
##   buildUpPlaySpeedCla~ buildUpPlayPassingCl~ buildUpPlayPositioningC~  Freq
##   <chr>                <chr>                 <chr>                   <int>
## 1 Fast                 Mixed                 Organised                   2
## 2 Balanced             Mixed                 Free Form                   1
## 3 Balanced             Mixed                 Organised                   1
## 4 Balanced             Short                 Organised                   1
## 5 Fast                 Short                 Free Form                   1
```

```
#   buildUpPlaySpeedClass buildUpPlayPassingClass buildUpPlayPositioningClass  Freq
# 1 Fast                  Mixed                   Organised                     2

## Get Roma's midfield attributes
roma_midfield_df <- roma_att_season_df[, midfield]

## Creative passing from midfield, not much crossing, 60/100 creating chances from shots
summary(dplyr::select_if(roma_midfield_df, is.numeric))
```

```
##  chanceCreationPassing chanceCreationCrossing chanceCreationShooting
##  Min.   :60.00         Min.   :35.00          Min.   :50.00
```

```
##  1st Qu.:65.50         1st Qu.:51.00          1st Qu.:56.00
##  Median :69.00         Median :57.00          Median :60.50
##  Mean   :68.83         Mean   :56.33          Mean   :60.50
##  3rd Qu.:72.50         3rd Qu.:63.00          3rd Qu.:65.75
##  Max.   :77.00         Max.   :75.00          Max.   :70.00
```

```
# chanceCreationPassing chanceCreationCrossing chanceCreationShooting
# Min.   :60.00         Min.   :35.00          Min.   :50.00
# 1st Qu.:65.50         1st Qu.:51.00          1st Qu.:56.00
# Median :69.00         Median :57.00          Median :60.50
# Mean   :68.83         Mean   :56.33          Mean   :60.50
# 3rd Qu.:72.50         3rd Qu.:63.00          3rd Qu.:65.75
# Max.   :77.00         Max.   :75.00          Max.   :70.00

## Found a much better way to summarize this data! Awesome.
## Risky chance creation, free form creation / positioning - fluid midfield
dplyr::select_if(roma_midfield_df, is.character) %>%
  group_by(chanceCreationPassingClass, chanceCreationCrossingClass, chanceCreationShootingClass, chance
      summarise(Freq = n()) %>%
          arrange(-Freq)
```

```
## # A tibble: 4 x 5
## # Groups:   chanceCreationPassingClass, chanceCreationCrossingClass,
## #   chanceCreationShootingClass [4]
##   chanceCreationP~ chanceCreationC~ chanceCreationS~ chanceCreationP~  Freq
##   <chr>            <chr>            <chr>            <chr>            <int>
## 1 Risky            Normal           Normal           Free Form            3
## 2 Normal           Normal           Lots             Free Form            1
## 3 Normal           Normal           Normal           Free Form            1
## 4 Risky            Lots             Normal           Free Form            1
```

```
#   chanceCreationPassingClass chanceCreationCrossingClass chanceCreationShootingClass chanceCreationPo
#  1 Risky                     Normal                       Normal                      Free Form

## Get defensive attributes for Roma
roma_defense_df <- roma_att_season_df[, defense]

## Summary of defensive numeric attributes
## Low pressure, low agression, 50/100 team width
summary(dplyr::select_if(roma_defense_df, is.numeric))
```

```
##  defencePressure defenceAggression defenceTeamWidth
##  Min.   :35.00   Min.   :30.00     Min.   :34.00
##  1st Qu.:36.00   1st Qu.:45.50     1st Qu.:48.50
##  Median :41.00   Median :47.00     Median :50.00
##  Mean   :42.83   Mean   :47.17     Mean   :50.83
##  3rd Qu.:48.25   3rd Qu.:54.50     3rd Qu.:52.25
##  Max.   :55.00   Max.   :57.00     Max.   :70.00
```

```
# defencePressure defenceAggression defenceTeamWidth
# Min.   :35.00   Min.   :30.00     Min.   :34.00
# 1st Qu.:36.00   1st Qu.:45.50     1st Qu.:48.50
```

```
# Median :41.00   Median :47.00     Median :50.00
# Mean   :42.83   Mean   :47.17     Mean   :50.83
# 3rd Qu.:48.25   3rd Qu.:54.50     3rd Qu.:52.25
# Max.   :55.00   Max.   :57.00     Max.   :70.00

## Summary of defensive categorical attributes
## they play the press, normal team shape and a high line due to their offside trap scheme - maybe this
dplyr::select_if(roma_defense_df, is.character) %>%
  group_by(defencePressureClass, defenceAggressionClass, defenceTeamWidthClass, defenceDefenderLineClass
  summarise(Freq = n()) %>%
  arrange(-Freq)
```

```
## # A tibble: 3 x 5
## # Groups:   defencePressureClass, defenceAggressionClass,
## #   defenceTeamWidthClass [2]
##   defencePressure~ defenceAggressi~ defenceTeamWidt~ defenceDefender~  Freq
##   <chr>            <chr>            <chr>            <chr>            <int>
## 1 Medium           Press            Normal           Offside Trap         3
## 2 Medium           Press            Normal           Cover                2
## 3 Medium           Contain          Wide             Offside Trap         1
```

```
#  defencePressureClass defenceAggressionClass defenceTeamWidthClass defenceDefenderLineClass  Freq
#1 Medium               Press                  Normal                Offside Trap                3
```