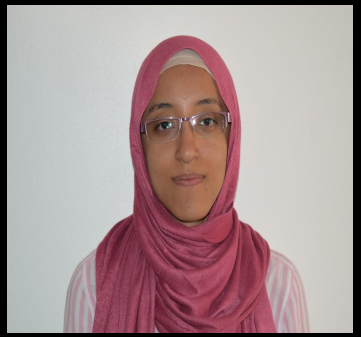
Football Dataset Analysis

B.K.L.M.S

Group Members

Belona S.

Keziah N.



Sarah E.

Lionel T.

Montaser M.

Agenda

- Introduction
- Statistical Data Analysis
- Predictive Models
- Challenges & Lessons Learned
- Conclusion

Introduction

- Problem Statement.
- Objectives.
- Dataset Description.

Problem Statement

- The importance of analysing football events has emerged.
- Teams management needs to attract investing entities.
- Predicting future results.

Objectives

- Provide decision makers with useful insights.
- Spot weaknesses and strengths in the teams/players in order to help them;
    - Look at the performance of each player.
    - Explore which kind of game piece the best team are using to win.

Dataset Description

Game_info.csv

ID: game id

General: league, season, date and host country.

Teams: home and away teams.

Results: home and away goals.

Odds: odds on (home win, away win and draw).

Events.csv

ID: game id and event id.

Teams: playing team and opponent.

Player: players involved in the event, bodypart, assist method and situation.

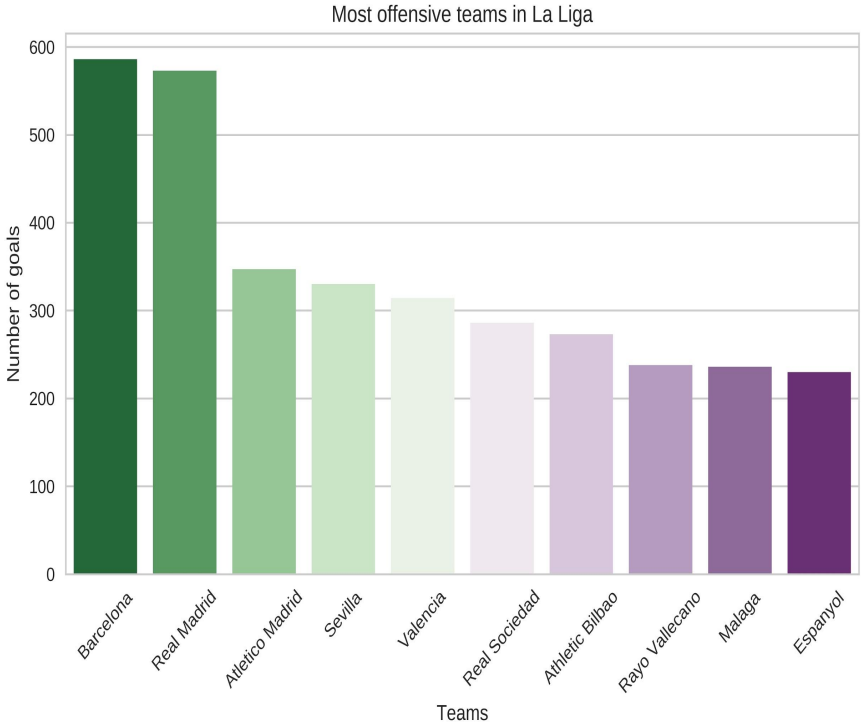
Shot: location (in the pitch), outcome, place, and is_goal or not.

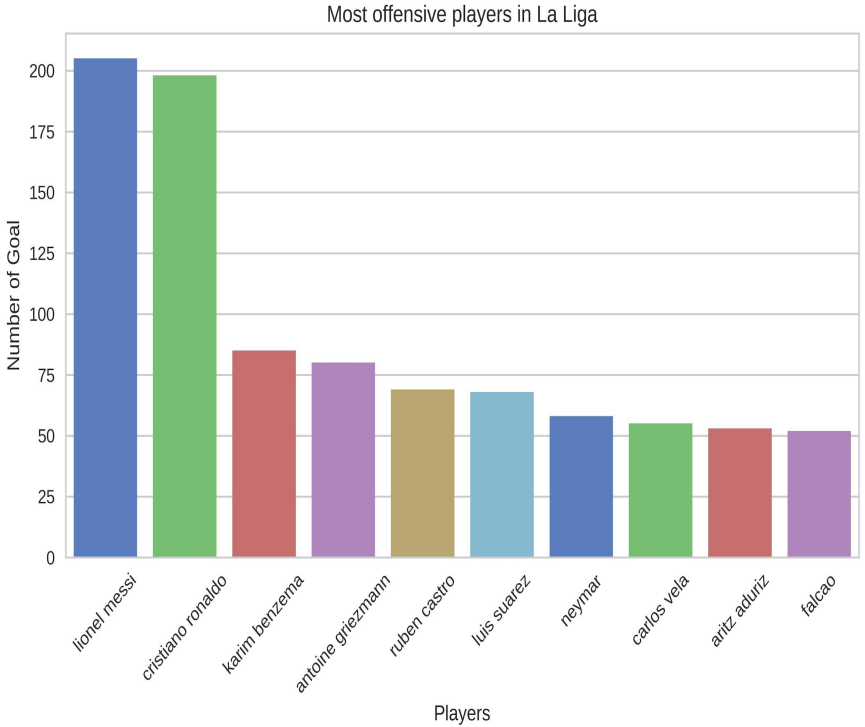
Statistical Data Analysis

- Detailed Analysis of Goals
- Detailed Analysis of Cards
- Evaluation of Characteristics of Teams

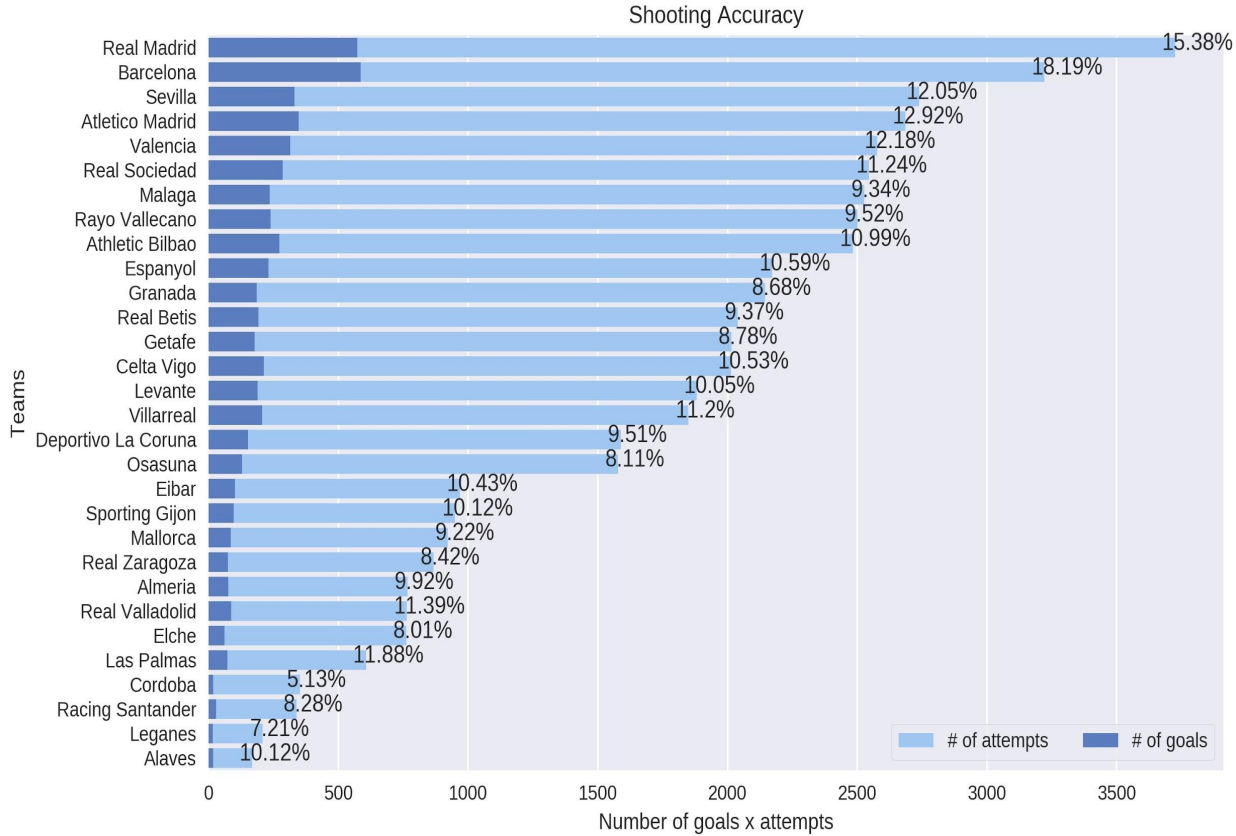
Detailed Analysis of Goals

Most offensive Team and Player in La Liga

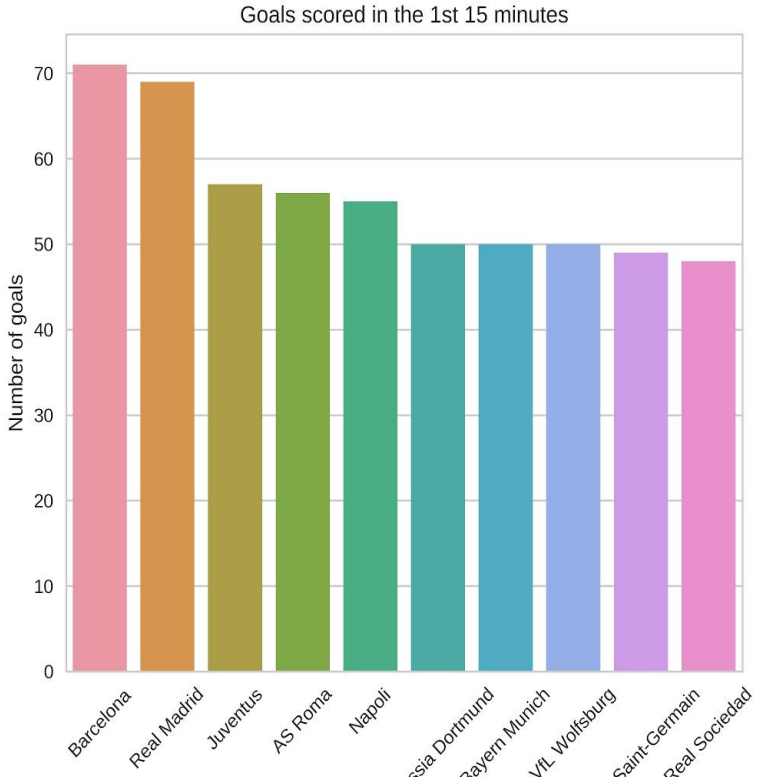


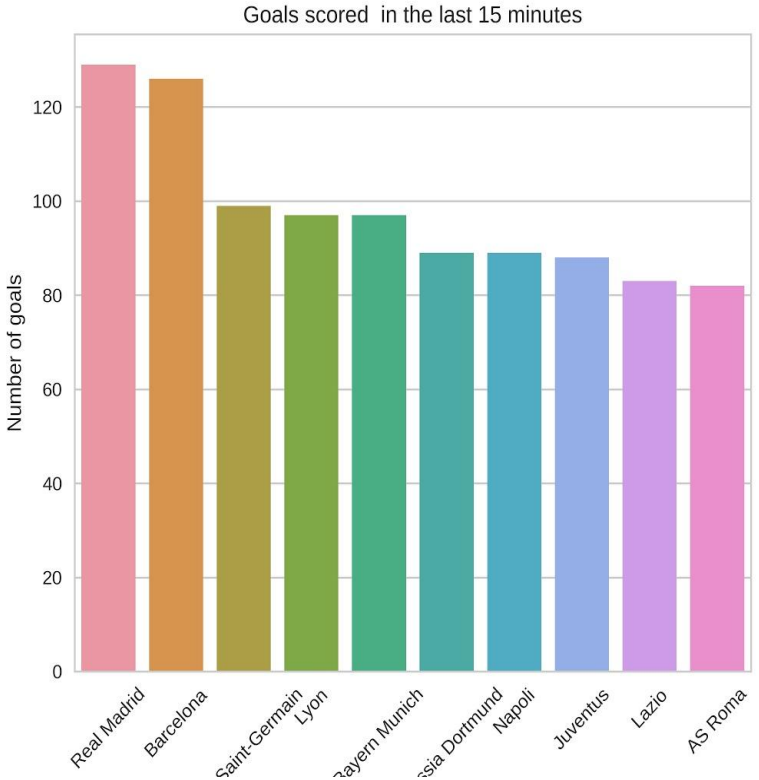


Shooting Accuracy



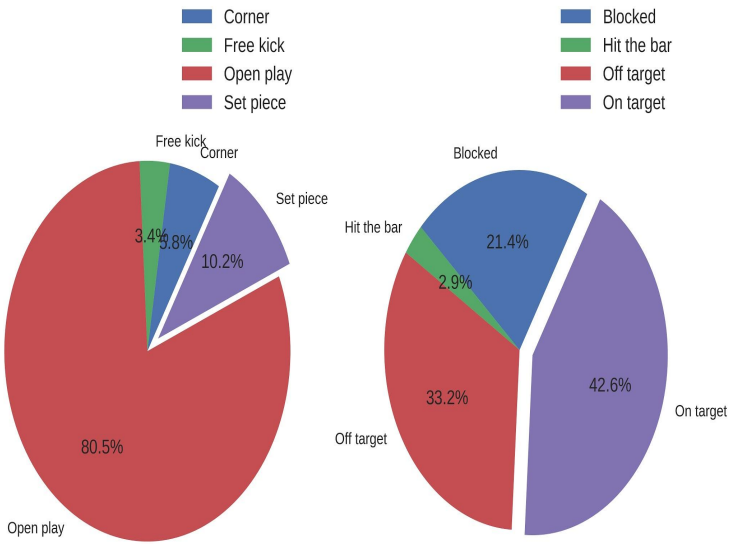
Goals scored (<15min, >75min)





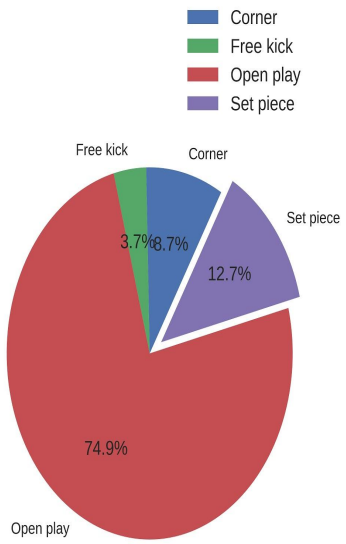
Goal Situation & Shot Outcome

Percentage of goals situations for Barcelona

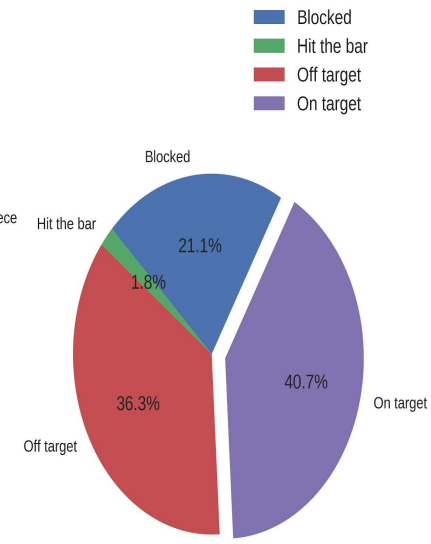


Percentage of shot outcome for Barcelona

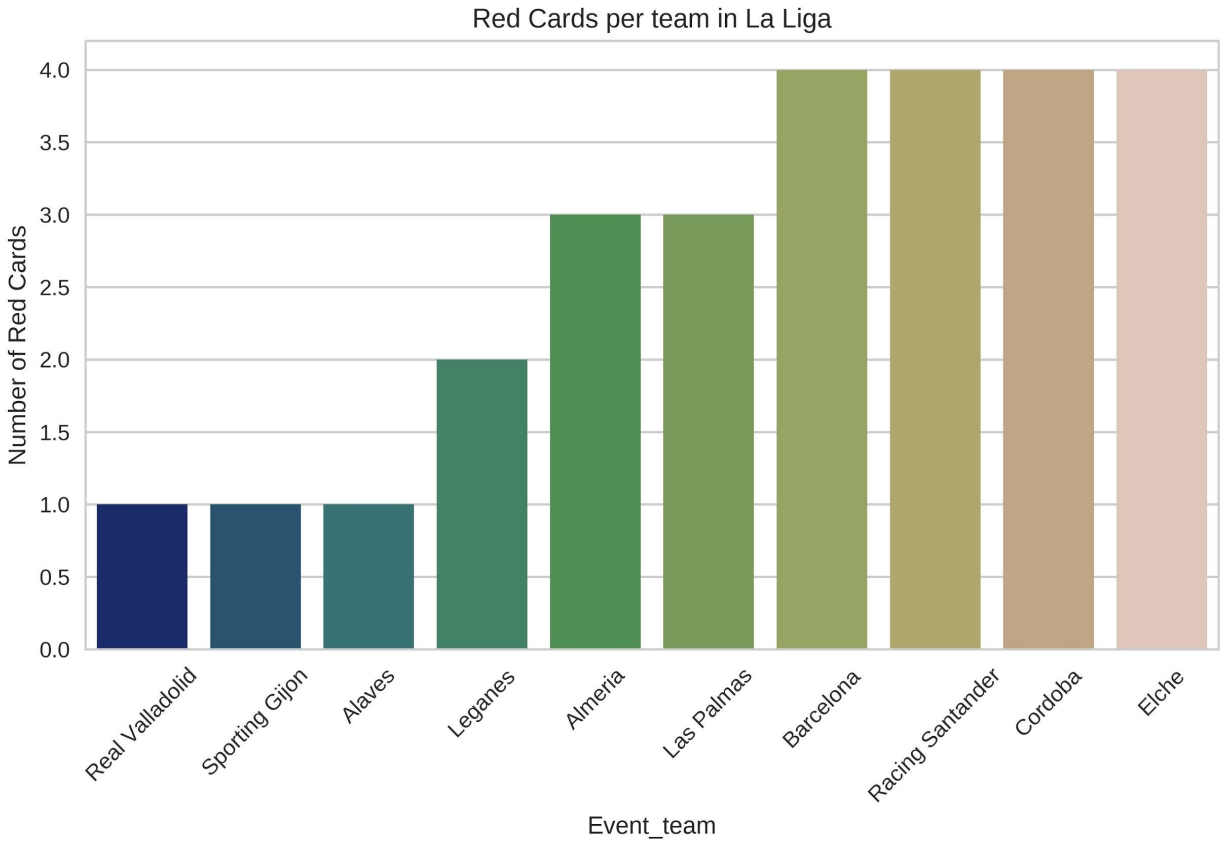
Percentage of goals situations for Real Madrid



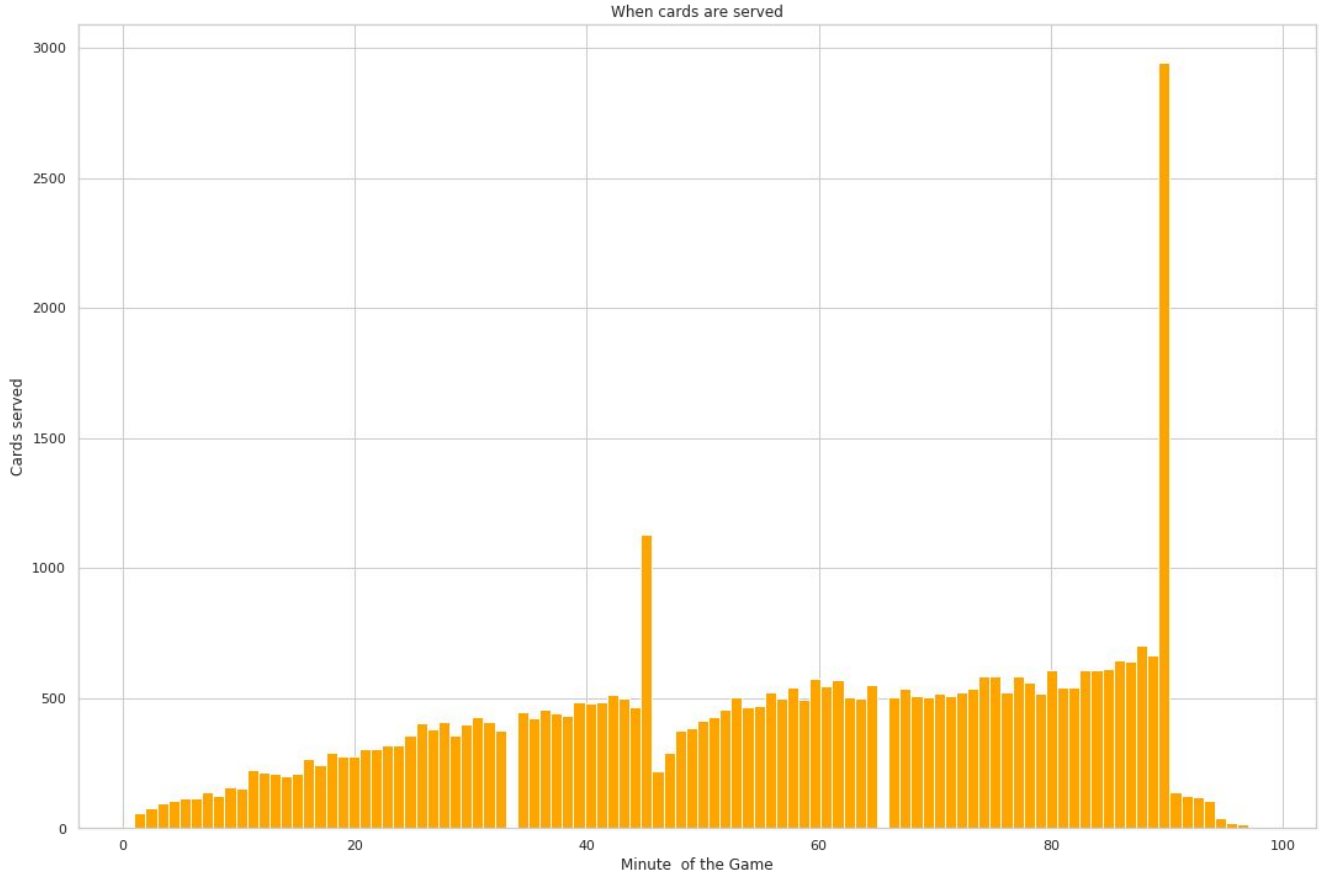
Percentage of shot outcome for Real Madrid



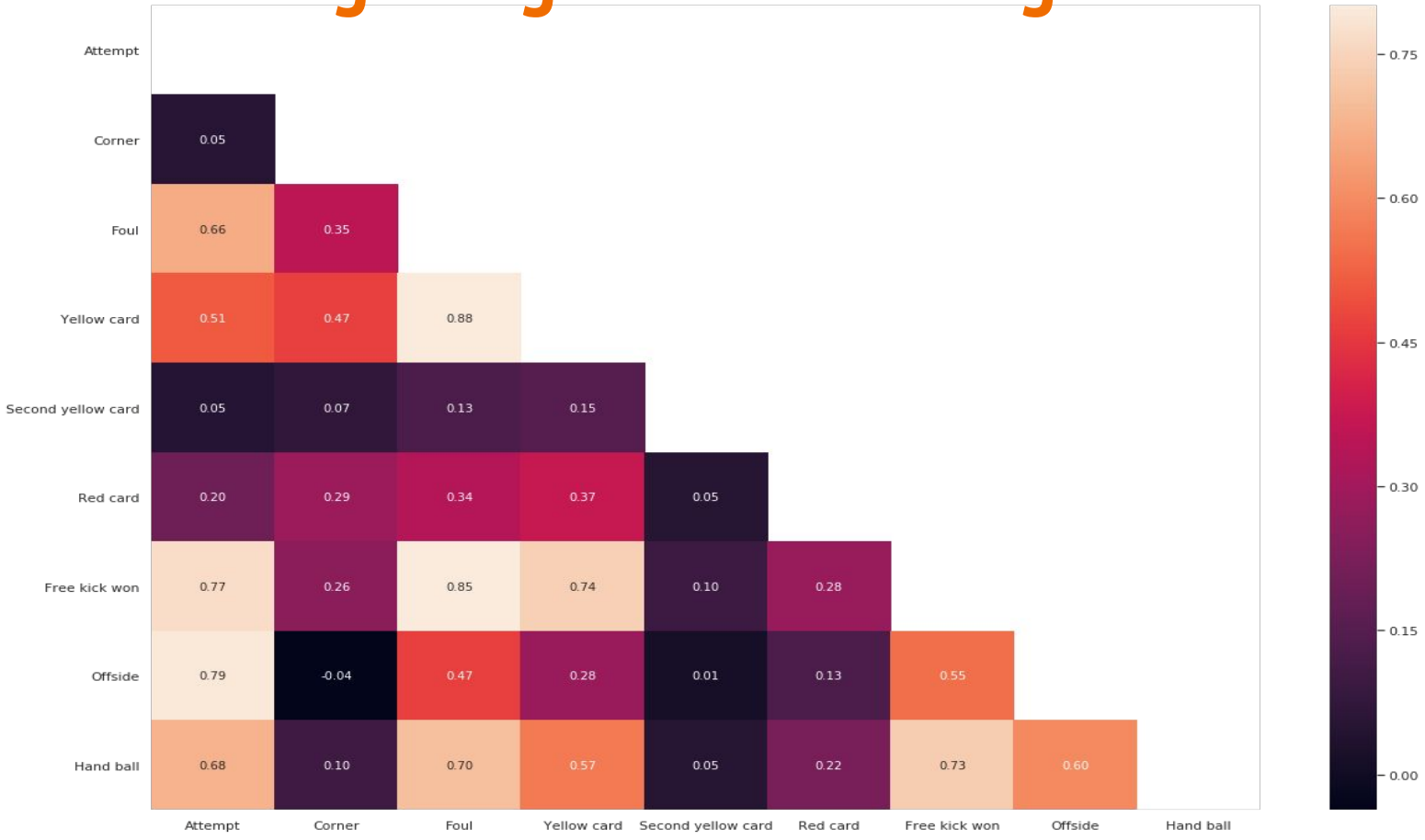
Why no Title?


Red Cards per team in La Liga
Number of Red Cards
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Real Valladolid
Sporting Gijon
Alaves
Leganes
Almeria
Las Palmas
Barcelona
Racing Santander
Cordoba
Elche
Event_team

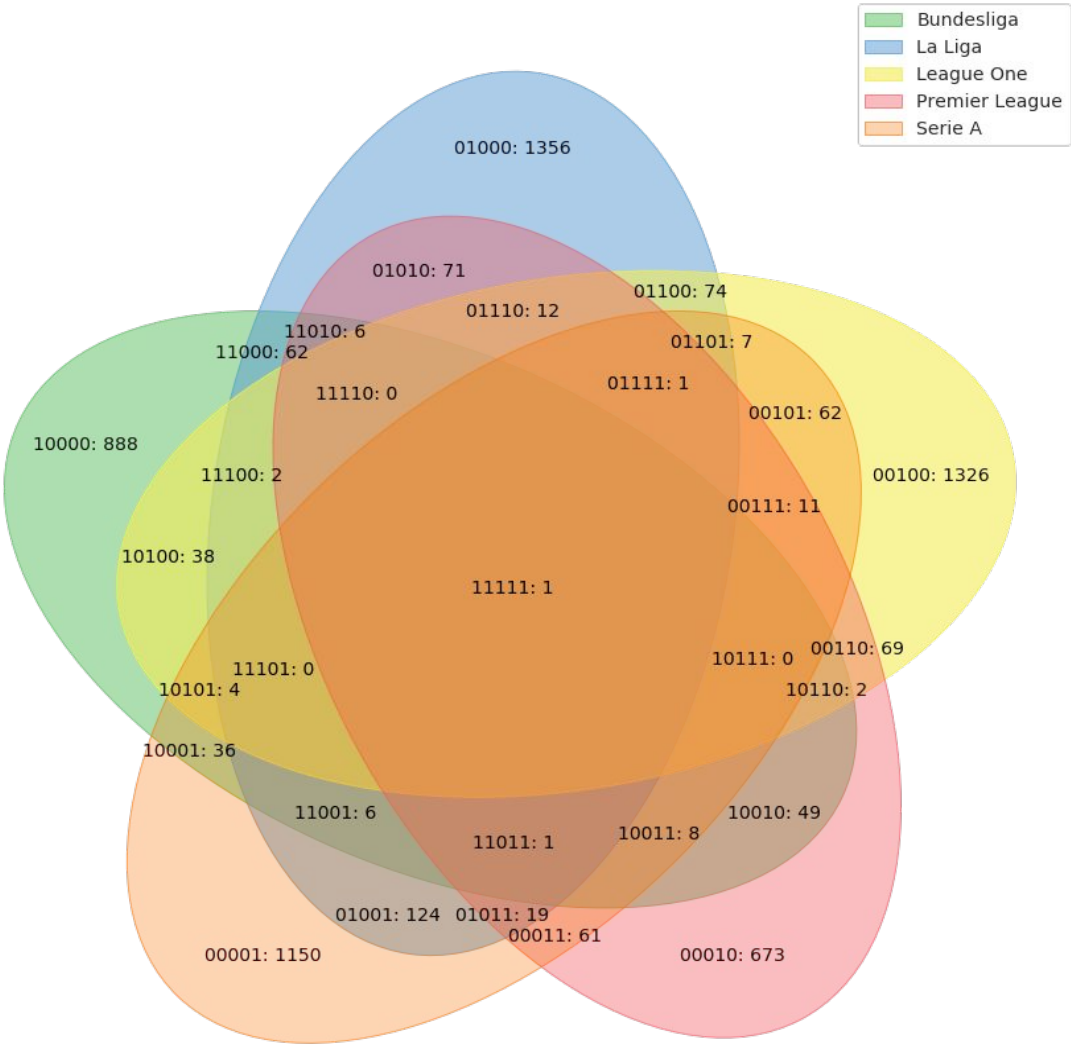
Time when are cards likely to be served



Effect of getting served on the game



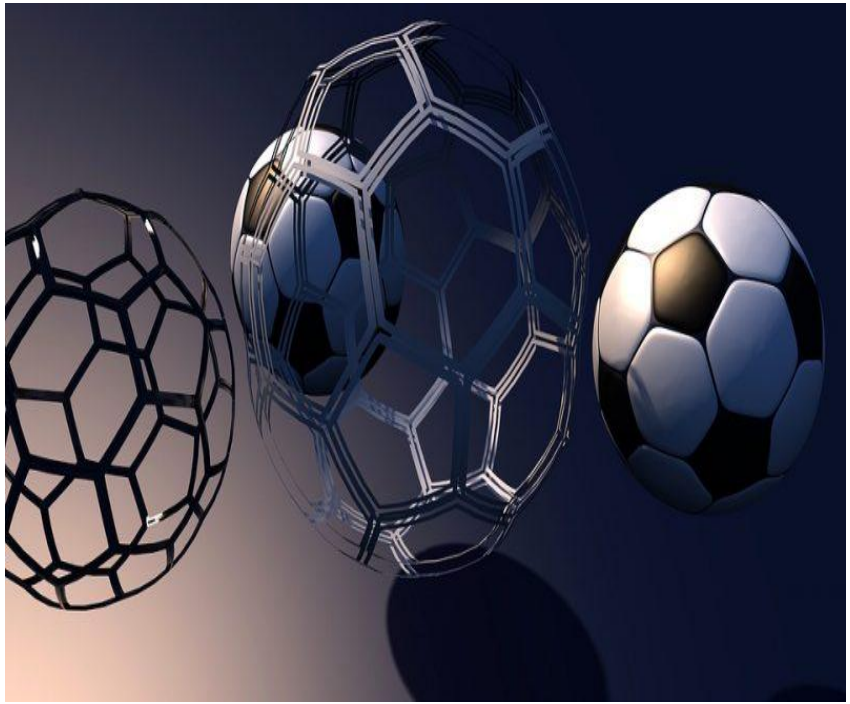
Players-League composition



Evaluation of Characteristics of Teams

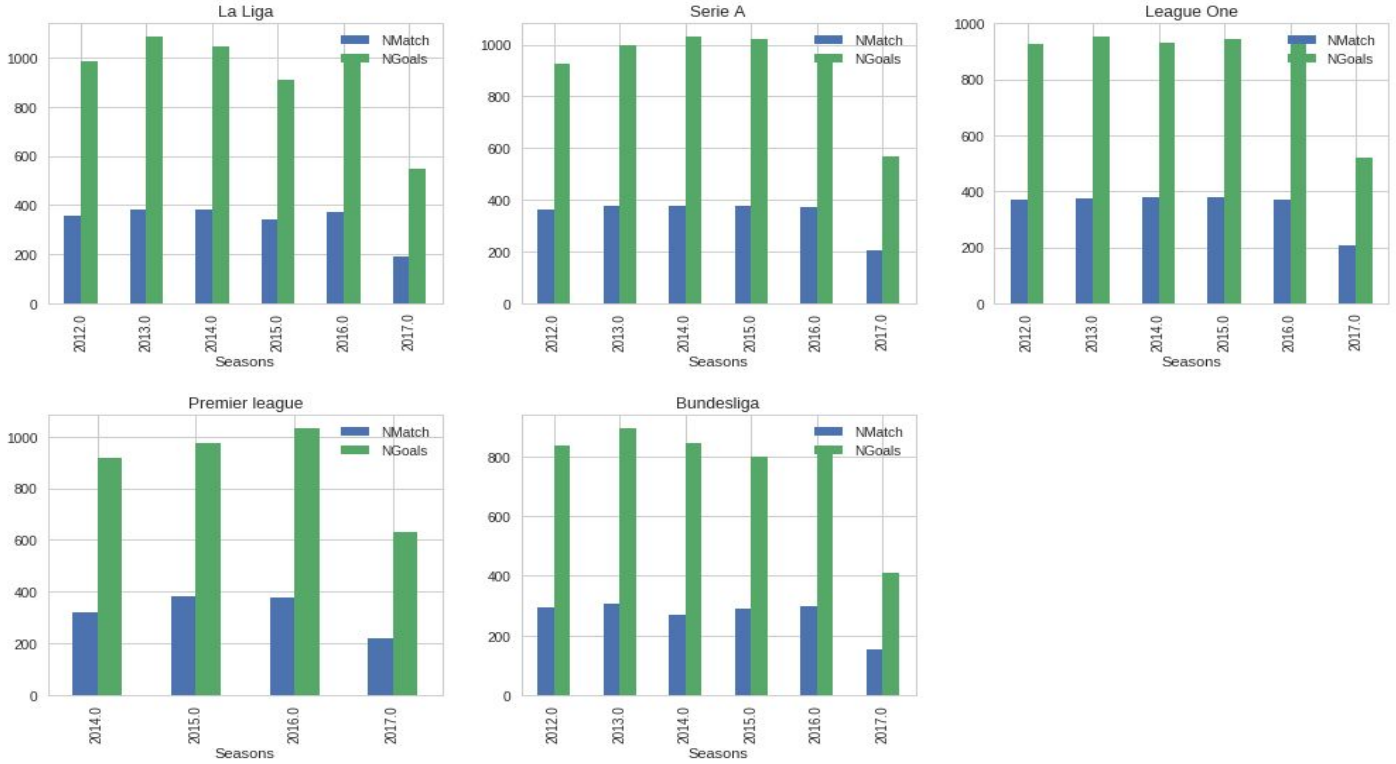
How to Evaluate Teams

- What kind of League do the team belong to?
- The League is balanced?
- What are the dominants teams per league?
- What are the characteristics of the best Teams?
- What about the worst Teams



What kind of League?

Figure : The Number of Goals per League



What kind of League?

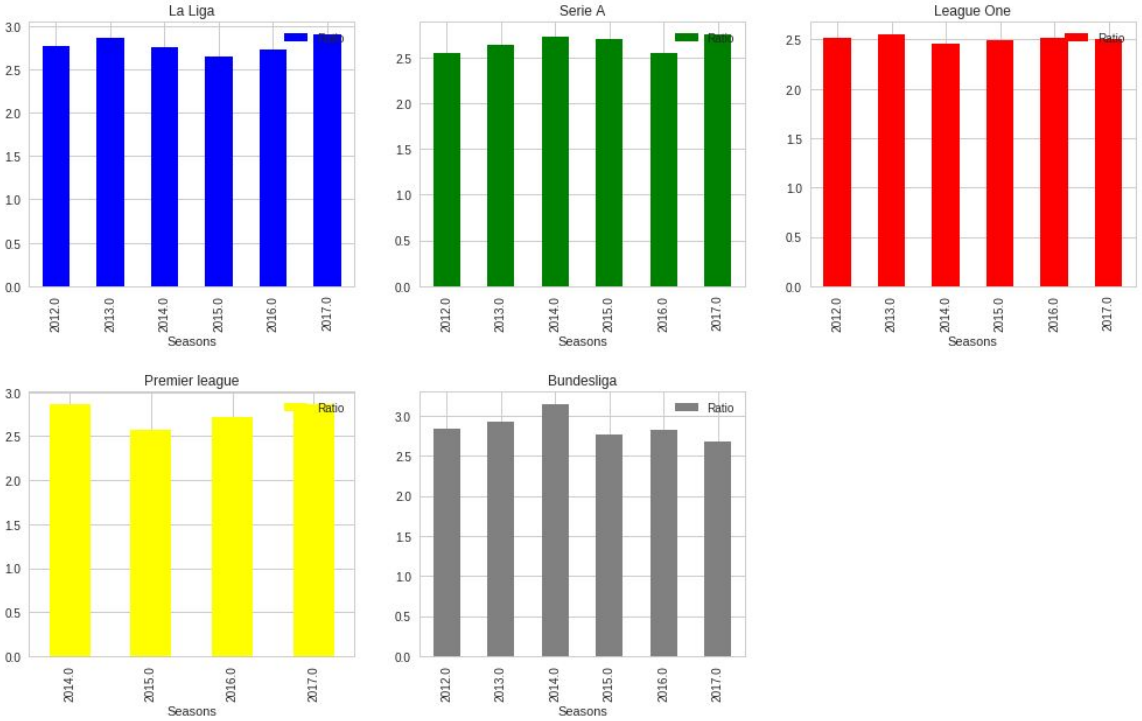


Figure :Average of Goals per League

The League is Balanced?

```
#########################################################
## the Informations about the League One 2012 ######
#########################################################
```

	ATeam	ButsEncaisses	ButsMarques	RatioE	RatioM	RatioME
6	Montpellier	33	65	0.87	1.71	1.97
0	Paris Saint-Germain	39	73	1.05	1.97	1.87
17	Lille	38	71	1.0	1.87	1.87
11	Lyon	46	58	1.31	1.66	1.26
8	Bordeaux	38	48	1.03	1.3	1.26
7	Marseille	33	40	0.94	1.14	1.21
14	Stade Rennes	42	48	1.14	1.3	1.14
16	Toulouse	32	34	0.86	0.92	1.06
12	St Etienne	43	44	1.16	1.19	1.02
19	Evian Thonon Gaillard	53	50	1.43	1.35	0.94
18	AJ Auxerre	53	45	1.39	1.18	0.85
4	Nice	41	33	1.17	0.94	0.8
15	Valenciennes	49	39	1.32	1.05	0.8
2	Brest	35	28	0.95	0.76	0.8
5	AS Nancy Lorraine	44	34	1.22	0.94	0.77
3	AC Ajaccio	52	39	1.44	1.08	0.75
13	Lorient	48	32	1.26	0.84	0.67
10	Sochaux	57	38	1.5	1.0	0.67
1	Caen	57	36	1.58	1.0	0.63
9	Dijon FCO	56	34	1.51	0.92	0.61

```
#########################################################
## the Informations about the League One 2017 ######
#########################################################
```

	ATeam	ButsEncaisses	ButsMarques	RatioE	RatioM	RatioME
1	AS Monaco	19	61	0.9	2.9	3.21
19	Paris Saint-Germain	15	39	0.71	1.86	2.6
8	Nice	14	34	0.67	1.62	2.43
10	Lyon	23	35	1.15	1.75	1.52
18	St Etienne	15	19	0.71	0.9	1.27
17	Guingamp	21	24	1.0	1.14	1.14
16	Toulouse	22	22	1.05	1.05	1.0
9	Marseille	25	24	1.19	1.14	0.96
6	Dijon FCO	29	26	1.38	1.24	0.9
5	Bordeaux	22	19	1.05	0.9	0.86
11	Stade Rennes	24	20	1.14	0.95	0.83
3	Montpellier	33	27	1.57	1.29	0.82
7	AS Nancy Lorraine	20	16	1.0	0.8	0.8
0	Bastia	24	18	1.14	0.86	0.75
15	Lille	23	17	1.1	0.81	0.74
13	Angers	26	17	1.24	0.81	0.65
2	Caen	32	20	1.6	1.0	0.62
14	Lorient	40	21	1.9	1.0	0.53
4	Metz	34	17	1.7	0.85	0.5
12	Nantes	28	13	1.33	0.62	0.46

Figure : Evolution of League One from 2012 to 2017

The League is Balanced?

```
#######################################################
## the Informations about the League One 2013 ######
#######################################################
                 ATeam ButsEncaisses ButsMarques RatioE RatioM RatioME
4     Paris Saint-Germain           21          66   0.57   1.78    3.14
2             St Etienne           30          59   0.79   1.55    1.97
10                Lyon             38          60    1.0   1.58    1.58
6                 Lille            36          56   0.97   1.51    1.56
17                Nice             44          55   1.19   1.49    1.25
14             Bordeaux            32          37   0.86    1.0    1.16
15            Marseille            36          40   0.95   1.05    1.11
7              Toulouse            45          47   1.18   1.24    1.04
9               Lorient            54          55   1.42   1.45    1.02
0            Montpellier           50          51   1.35   1.38    1.02
12         Valenciennes            53          48   1.43    1.3    0.91
3   Evian Thonon Gaillard          50          43   1.32   1.13    0.86
1           Stade Rennes           57          46   1.54   1.24    0.81
5         Stade de Reims           41          33   1.08   0.87     0.8
16              Troyes             55          42   1.49   1.14    0.76
19              Sochaux            52          39   1.41   1.05    0.75
13           AC Ajaccio            50          37   1.35    1.0    0.74
11              Bastia             63          45    1.7   1.22    0.71
18     AS Nancy Lorraine           55          36   1.53    1.0    0.65
8                Brest             60          30   1.62   0.81     0.5
```

```
#######################################################
## the Informations about the League One 2014 ######
#######################################################
                 ATeam ButsEncaisses ButsMarques RatioE RatioM RatioME
17    Paris Saint-Germain           22          83   0.58   2.18    3.77
16              AS Monaco           31          61   0.82   1.61    1.97
4                Lille              25          45   0.66   1.18     1.8
11            St Etienne            34          54   0.89   1.42    1.59
13             Marseille            38          49   1.03   1.32    1.29
6               Bordeaux            40          49   1.05   1.29    1.23
3                  Lyon             44          52   1.16   1.37    1.18
5           Stade Rennes            44          45   1.16   1.18    1.02
18              Lorient             52          48   1.37   1.26    0.92
14             Toulouse             50          46   1.32   1.21    0.92
7              Guingamp             38          33    1.0   0.87    0.87
0            Montpellier            51          44   1.34   1.16    0.86
12        Stade de Reims            50          43   1.32   1.13    0.86
19               Nantes             41          34   1.11   0.92    0.83
1   Evian Thonon Gaillard           49          38   1.29    1.0    0.78
15                Nice              40          30   1.05   0.79    0.75
10              Bastia              56          40   1.51   1.08    0.71
9               Sochaux             59          35   1.59   0.95    0.59
2            Valenciennes           65          34   1.71   0.89    0.52
8             AC Ajaccio            71          36   1.87   0.95    0.51
#######################################################
```

Figure : Evolution of League One from 2012 to 2017

The League is Balanced?

```
#########################################################
## the Informations about the League One 2015 ######
#########################################################
```

	ATeam	ButsEncaisses	ButsMarques	RatioE	RatioM	RatioME
13	Paris Saint-Germain	34	79	0.89	2.08	2.32
9	Lyon	33	70	0.87	1.84	2.12
8	AS Monaco	24	49	0.63	1.29	2.04
18	Marseille	38	72	1.0	1.89	1.89
17	St Etienne	30	48	0.79	1.26	1.6
6	Montpellier	38	44	1.0	1.16	1.16
19	Bordeaux	43	46	1.13	1.21	1.07
3	Lille	41	42	1.08	1.11	1.02
10	Caen	54	50	1.42	1.32	0.93
15	Stade Rennes	40	34	1.05	0.89	0.85
16	Lorient	48	41	1.26	1.08	0.85
7	Nice	51	41	1.34	1.08	0.8
1	Bastia	43	34	1.13	0.89	0.79
2	Guingamp	52	41	1.37	1.08	0.79
5	Nantes	38	28	1.0	0.74	0.74
4	Evian Thonon Gaillard	57	41	1.5	1.08	0.72
0	Stade de Reims	63	44	1.66	1.16	0.7
14	Toulouse	61	41	1.61	1.08	0.67
11	Lens	60	32	1.58	0.84	0.53
12	Metz	60	31	1.58	0.82	0.52

```
#########################################################
## the Informations about the League One 2016 ######
#########################################################
```

	ATeam	ButsEncaisses	ButsMarques	RatioE	RatioM	RatioME
17	Paris Saint-Germain	19	95	0.51	2.57	5.0
8	Lyon	38	65	1.03	1.76	1.71
4	Nice	35	54	0.95	1.46	1.54
0	Lille	26	38	0.7	1.03	1.46
6	Marseille	39	47	1.05	1.27	1.21
12	St Etienne	34	39	0.94	1.08	1.15
1	Montpellier	43	49	1.16	1.32	1.14
10	AS Monaco	48	53	1.3	1.43	1.1
14	Angers	35	36	0.95	0.97	1.03
11	Stade Rennes	50	49	1.35	1.32	0.98
7	Bordeaux	53	49	1.43	1.32	0.92
13	Guingamp	52	44	1.41	1.19	0.85
2	Bastia	40	32	1.08	0.86	0.8
16	Lorient	57	45	1.54	1.22	0.79
5	Nantes	40	31	1.08	0.84	0.78
9	Toulouse	52	40	1.41	1.08	0.77
15	Caen	52	37	1.41	1.0	0.71
18	Stade de Reims	55	38	1.49	1.03	0.69
19	GFC Ajaccio	55	36	1.53	1.0	0.65
3	Troyes	80	26	2.16	0.7	0.33

Figure : Evolution of League One from 2012 to 2017

The Dominant Teams

In BundesLiga we have

- Barcelona
- Real Madrid
- Atletico Madrid

```
#########################################################
## the Informations about the La Liga 2012 ######
#########################################################
              ATeam ATot AToth BWin BWinH CDefeat CDefeatH Null Points
9       Real Madrid   34    17   29    15       1        1    4     91
8         Barcelona   35    18   26    17       2        2    7     85
1          Valencia   36    18   16    10       9        6   11     59
6   Atletico Madrid   37    19   15    11      12        9   10     55
17          Levante   35    18   16    11      12        8    7     55
#########################################################
## the Informations about the La Liga 2013 ######
#########################################################
              ATeam ATot AToth BWin BWinH CDefeat CDefeatH Null Points
5         Barcelona   38    19   31    18       2        2    5     98
4       Real Madrid   37    19   25    16       5        5    7     82
19  Atletico Madrid   38    19   23    14       8        5    7     76
13    Real Sociedad   37    18   18    10       7        5   12     66
17          Valencia   38    19   19    13      11        8    8     65
#########################################################
## the Informations about the La Liga 2014 ######
#########################################################
              ATeam ATot AToth BWin BWinH CDefeat CDefeatH Null Points
18  Atletico Madrid   38    19   28    15       5        5    5     89
3         Barcelona   38    19   27    16       5        4    6     87
5       Real Madrid   38    19   27    16       5        3    6     87
11  Athletic Bilbao   38    19   20    13       8        6   10     70
4           Sevilla   38    19   18    11      12        8    8     62
#########################################################
## the Informations about the La Liga 2015 ######
#########################################################
              ATeam ATot AToth BWin BWinH CDefeat CDefeatH Null Points
7         Barcelona   34    17   27    15       3        2    4     85
9       Real Madrid   34    18   26    15       6        5    2     80
10         Valencia   34    17   21    15       3        3   10     73
0           Sevilla   34    17   21    12       6        5    7     70
15  Atletico Madrid   34    16   20    11       5        3    9     69
#########################################################
## the Informations about the La Liga 2016 ######
#########################################################
              ATeam ATot AToth BWin BWinH CDefeat CDefeatH Null Points
14        Barcelona   37    19   28    16       5        3    4     88
12      Real Madrid   37    19   27    16       4        2    6     87
2   Atletico Madrid   37    18   25    13       6        5    6     81
10       Villarreal   37    19   18    12       9        6   10     64
13       Celta Vigo   37    19   17     9      11        7    9     60
#########################################################
```

Figure :Dominant Teams of La Liga

The Dominant Teams

In La Liga we have

- Borussia Dortmund
- Bayern Munich
- Bayer Leverkusen

Figure :Dominant Teams of Bundesliga

```
#######################################################
## the Informations about the Bundesliga 2012 ######
#######################################################
                ATeam ATot AToth BWin BWinH CDefeat CDefeatH Null Points
0         Borussia Dortmund   33    16   24    13       3        2    6     78
8             Bayern Munich   32    16   21    13       7        5    4     67
9                Schalke 04   33    17   19    13      10        7    4     61
14  Borussia Monchengladbach   34    17   17     9       8        7    9     60
16          Bayer Leverkusen   33    17   15     8      10        5    8     53
#######################################################
## the Informations about the Bundesliga 2013 ######
#######################################################
               ATeam ATot AToth BWin BWinH CDefeat CDefeatH Null Points
16      Bayern Munich   34    17   29    14       2        1    3     90
0   Borussia Dortmund   34    17   19    10       5        1   10     67
12   Bayer Leverkusen   34    17   20    13       7        5    7     67
13         Schalke 04   33    17   16    10      11        7    6     54
5         SC Freiburg   34    17   14     8      11        6    9     51
#######################################################
## the Informations about the Bundesliga 2014 ######
#######################################################
               ATeam ATot AToth BWin BWinH CDefeat CDefeatH Null Points
5       Bayern Munich   30    15   26    13       2        1    2     80
3   Borussia Dortmund   30    14   19     9       7        3    4     61
15         Schalke 04   29    15   18    11       6        3    5     59
14      VfL Wolfsburg   30    15   16     9       8        5    6     54
7    Bayer Leverkusen   29    14   15     7       9        6    5     50
#######################################################
## the Informations about the Bundesliga 2015 ######
#######################################################
                ATeam ATot AToth BWin BWinH CDefeat CDefeatH Null Points
0             Bayern Munich   33    17   24    14       4        2    5     77
8   Borussia Monchengladbach   34    17   19    12       6        4    9     66
10            VfL Wolfsburg   33    17   18    12       5        5   10     64
14          Bayer Leverkusen   33    17   17    10       7        6    9     60
9                FC Augsburg   33    17   15     9      14       10    4     49
#######################################################
## the Informations about the Bundesliga 2016 ######
#######################################################
                ATeam ATot AToth BWin BWinH CDefeat CDefeatH Null Points
0             Bayern Munich   33    16   27    14       2        1    4     85
4          Borussia Dortmund   33    16   24    14       4        4    5     77
5           Bayer Leverkusen   33    16   16     8      10        6    7     55
16  Borussia Monchengladbach   33    17   16    13      13       10    4     52
9             Hertha Berlin   33    17   14     9      11        8    8     50
#######################################################
```

Characteristics of Best Teams

Figure :Best Teams of La Liga

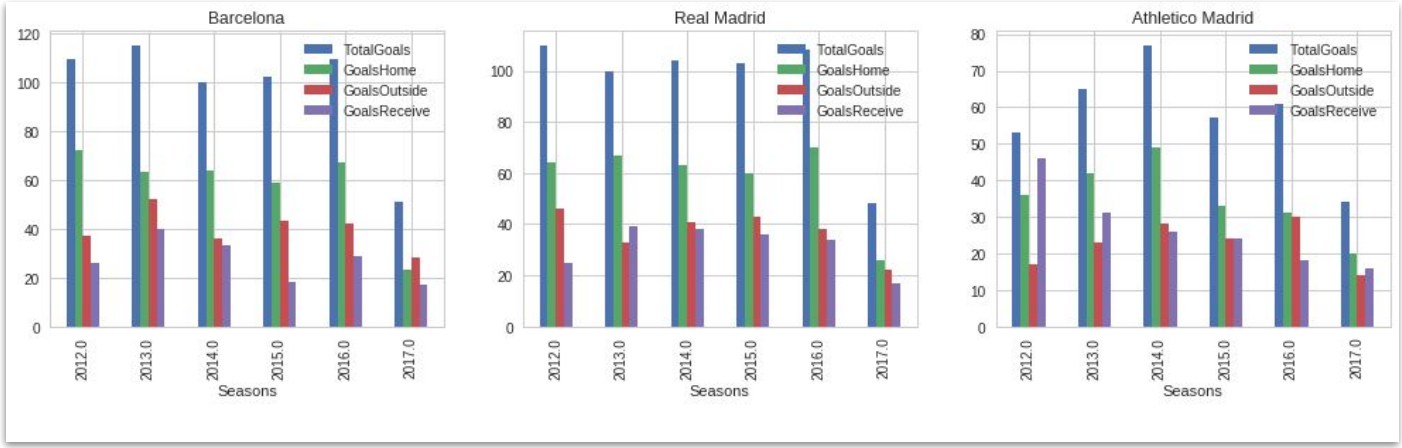
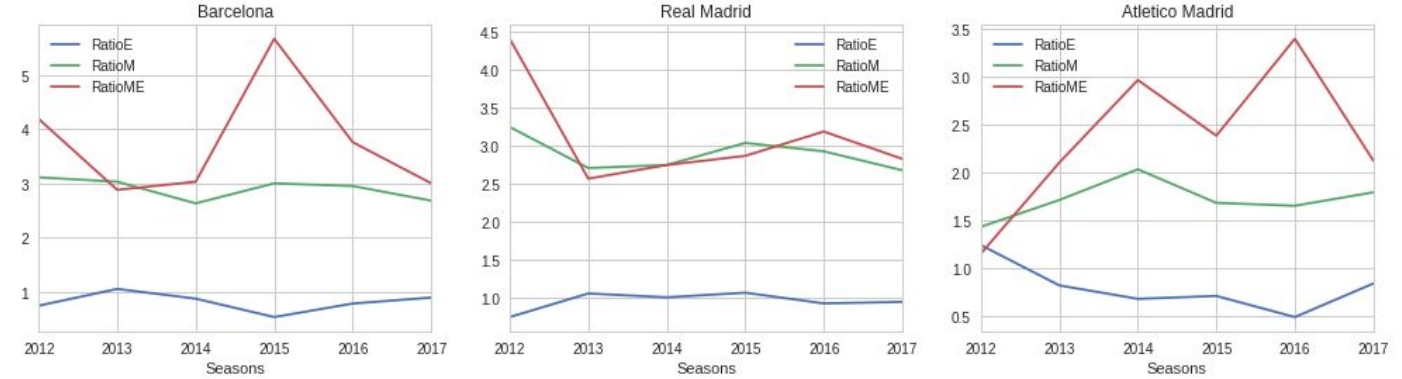


Figure :Ratio of Best Teams of La Liga



Characteristics of worst Teams

Figure :Worst Teams of Bundesliga

```
#######################################################
## the Informations about the Bundesliga 2012 ######
#######################################################
            ATeam ATot ATotH BWin BWinH CDefeat CDefeatH Null Points
1      FC Augsburg   31    15    7     5      12        8   12     33
13      Hamburg SV   32    16    7     3      13        7   12     33
3    Hertha Berlin   33    16    7     4      16        7   10     31
4       FC Cologne   32    16    8     5      19       12    5     29
17  Kaiserslautern   33    17    4     2      19        9   10     22
#######################################################
## the Informations about the Bundesliga 2013 ######
#######################################################
                ATeam ATot ATotH BWin BWinH CDefeat CDefeatH Null Points
9          Werder Bremen   34    17    9     5      16        9    9     36
4             FC Augsburg   34    17    8     5      17       10    9     33
11        TSG Hoffenheim   34    17    8     5      18       11    8     32
10    Fortuna Dusseldorf   34    17    7     5      18       12    9     30
3   SpVgg Greuther Furth   34    17    5     0      21        8    8     23
#######################################################
## the Informations about the Bundesliga 2014 ######
#######################################################
                      ATeam ATot ATotH BWin BWinH CDefeat CDefeatH Null Points
8               Werder Bremen   30    16    8     5      13        8    9     33
1               VfB Stuttgart   31    16    8     5      15        8    8     32
9   TSG Eintracht Braunschweig   30    15    6     5      17       10    7     25
16                   Nurnberg   30    15    5     3      16        8    9     24
13                Hamburg SV   30    15    6     4      19       11    5     23
#######################################################
## the Informations about the Bundesliga 2015 ######
#######################################################
            ATeam ATot ATotH BWin BWinH CDefeat CDefeatH Null Points
3   Hertha Berlin   33    17    9     6      17       10    7     34
11  VfB Stuttgart   33    17    8     5      16        8    9     33
15     Hamburg SV   33    17    8     6      18       12    7     31
7    SC Paderborn   33    17    6     4      17       10   10     28
16    SC Freiburg   18     1    2     0       8        8    8     14
#######################################################
## the Informations about the Bundesliga 2016 ######
#######################################################
                ATeam ATot ATotH BWin BWinH CDefeat CDefeatH Null Points
12       TSG Hoffenheim   33    16    9     6      14       10   10     37
15  Eintracht Frankfurt   33    17    9     6      15       10    9     36
6          Werder Bremen   33    16    9     4      16        9    8     35
8          VfB Stuttgart   33    17    9     6      18        8    6     33
10           Hannover 96   33    17    7     4      22        9    4     25
```

Characteristics of worst Teams

Figure :Goals of Worst Teams of La Liga

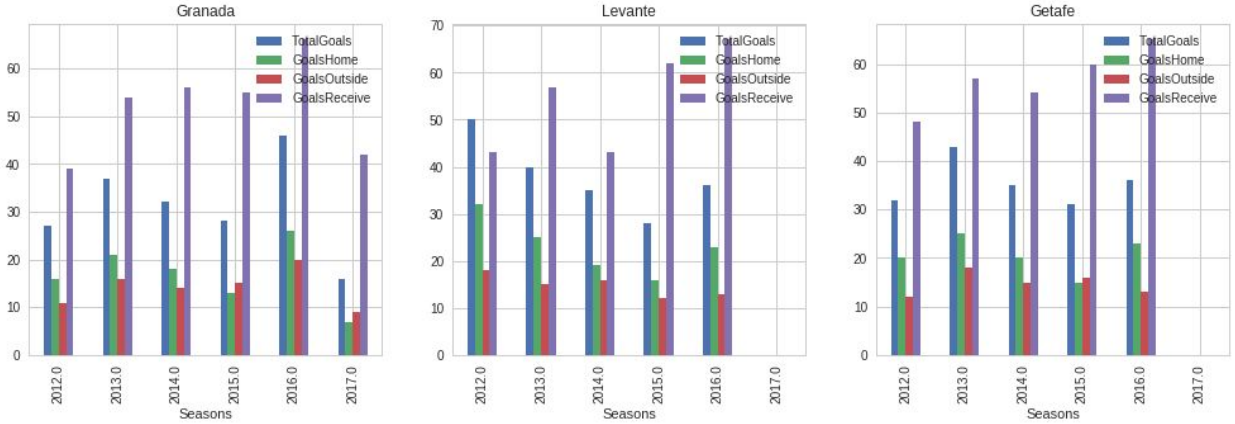
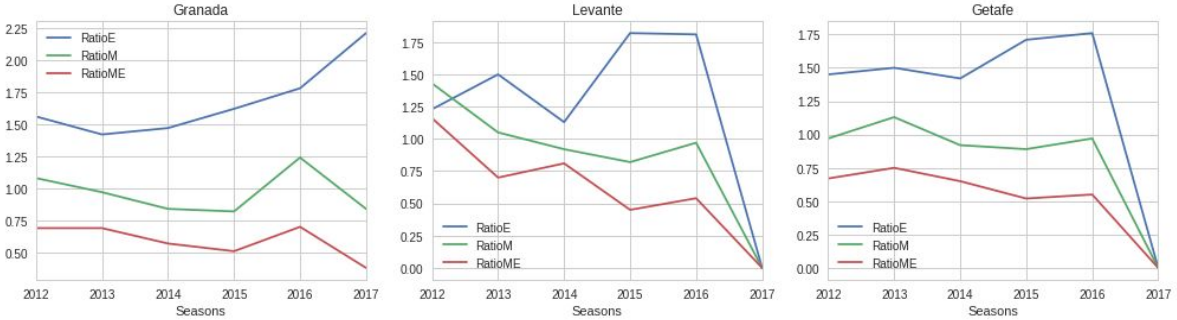


Figure :Ratio of Worst Teams of La Liga



Predictive Models

1- Predicting a Goal from Events

2- Predicting Number of Goals in a Match

1- Predicting a Goal from Events

Is it a Goal?

- Data Cleaning.
- Feature Selection.
- Used Models.
- Interpreting Results.

Data Cleaning

- Missing values:
  - some features have >75% missing.
  - Two methods: fill with 'UNK', or remove rows with nulls and work with smaller data.
- Data types:
  - Date.
  - Categorical instead of str object.

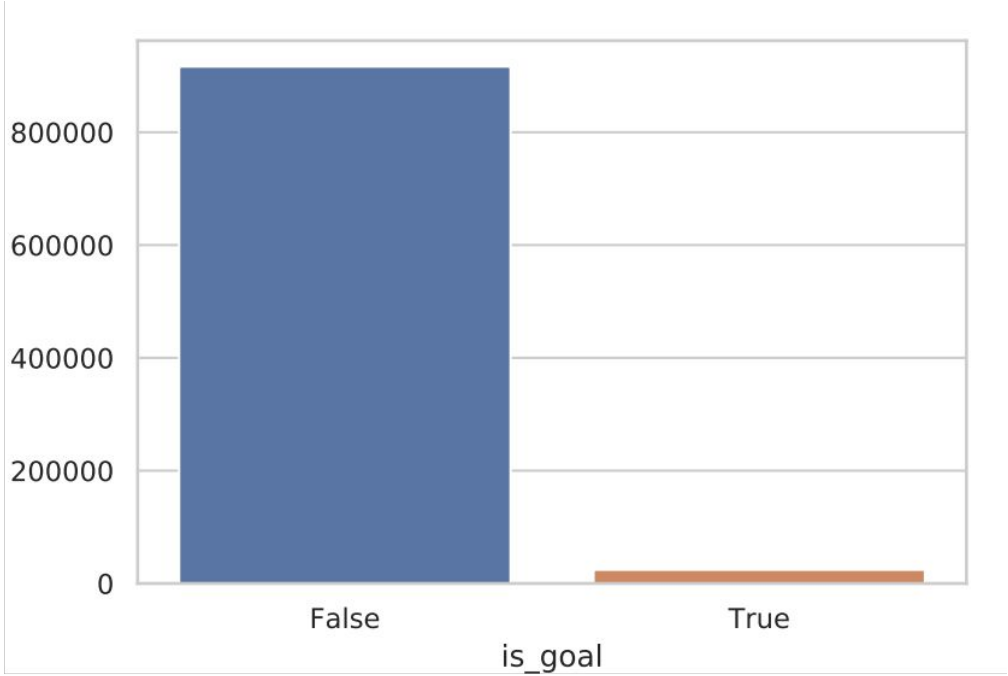
Feature name	Number of Missing Values	% of Total Values
player_in	889294	94.5
player_out	889271	94.5
odd_over	842329	89.5
odd_under	842329	89.5
odd_bts	842329	89.5
odd_bts_n	842329	89.5
assist_method	773104	82.2
event_type2	726716	77.2
shot_place	713550	75.8
shot_outcome	712511	75.7
situation	711872	75.6
bodypart	711824	75.6
player2	649699	69.0
location	473942	50.4
player	61000	6.5

Feature Selection

- Manually chosen features:
  - Odds data.
  - Properties of the shot.

Feature name	Information
odd_h	941009 non-null float16
odd_d	941009 non-null float16
odd_a	941009 non-null float16
assist_method	941009 non-null category
location	941009 non-null category
side	941009 non-null category
shot_place	941009 non-null category
situation	941009 non-null category
bodypart	941009 non-null category
first_half	941009 non-null bool

Imbalanced Classes



Imbalanced Classes

- Used: balanced accuracy instead of accuracy.
- ‘UNK’ filled + Random Under Sampling.

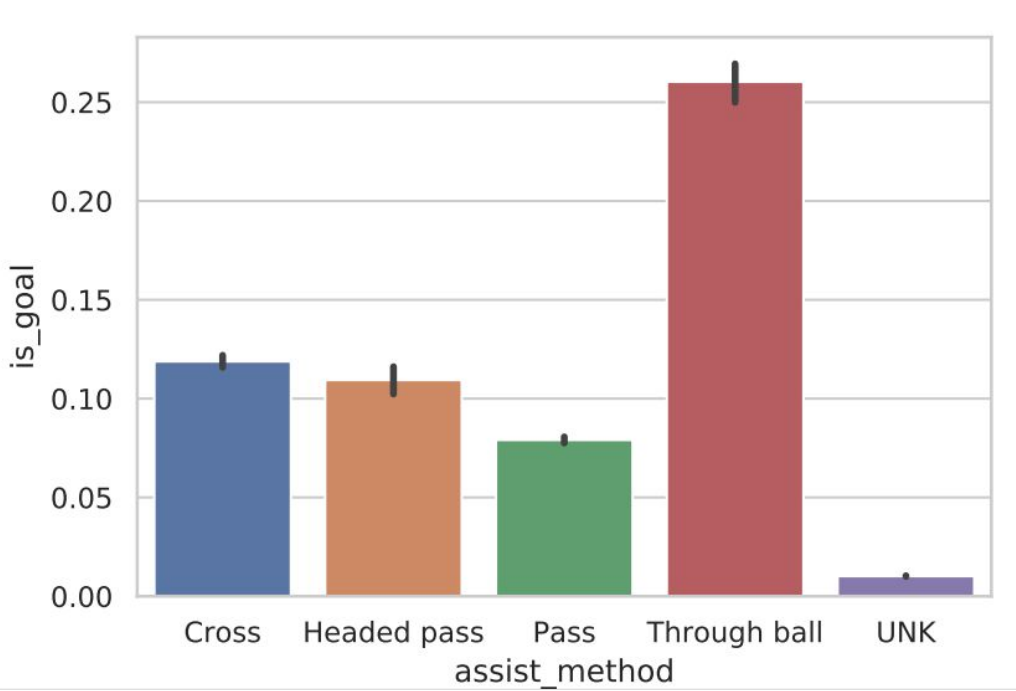
	Imbalanced classes		Balanced Classes with Random Under Sampling	
Model name	Accuracy	Balanced Accuracy	Accuracy	Balanced Accuracy
LR	98.42 %	80.37 %	94.51 %	97.18 %
RF	98.2 %	78.08 %	95.14 %	97.01 %
GB	98.39 %	75.77 %	94.48 %	97.14 %

Imbalanced Classes

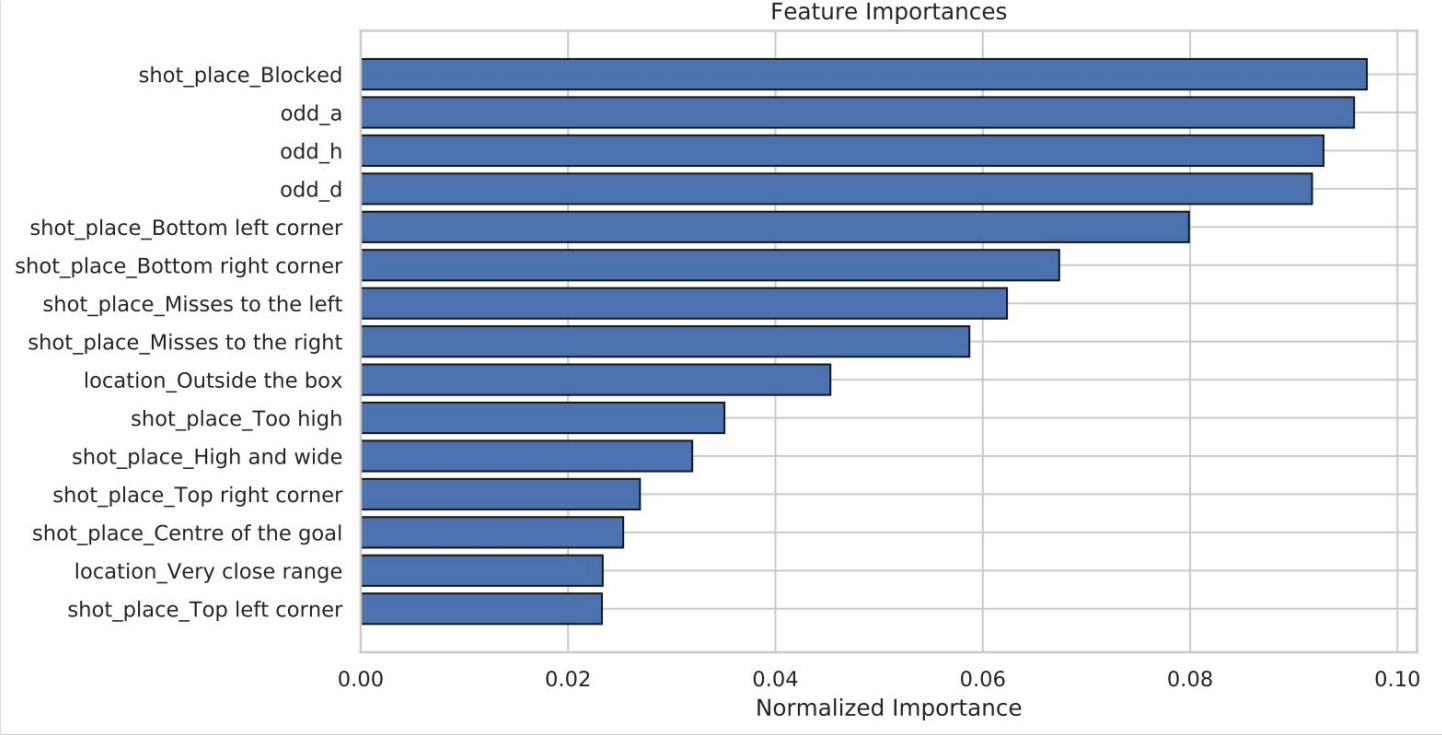
- Remove null rows + Over sampling techniques:
  - SMOTENC.
  - Random.

	Balanced Classes with SMOTENC		Balanced Classes with Random Over Sampling	
Model name	Accuracy	Balanced Accuracy	Accuracy	Balanced Accuracy
LR	83.68%	88.87%	83.11%	89.28%
RF	89.21%	80.07%	90.91%	75.25%
GB	80.96%	88.60%	81.36 %	88.82%

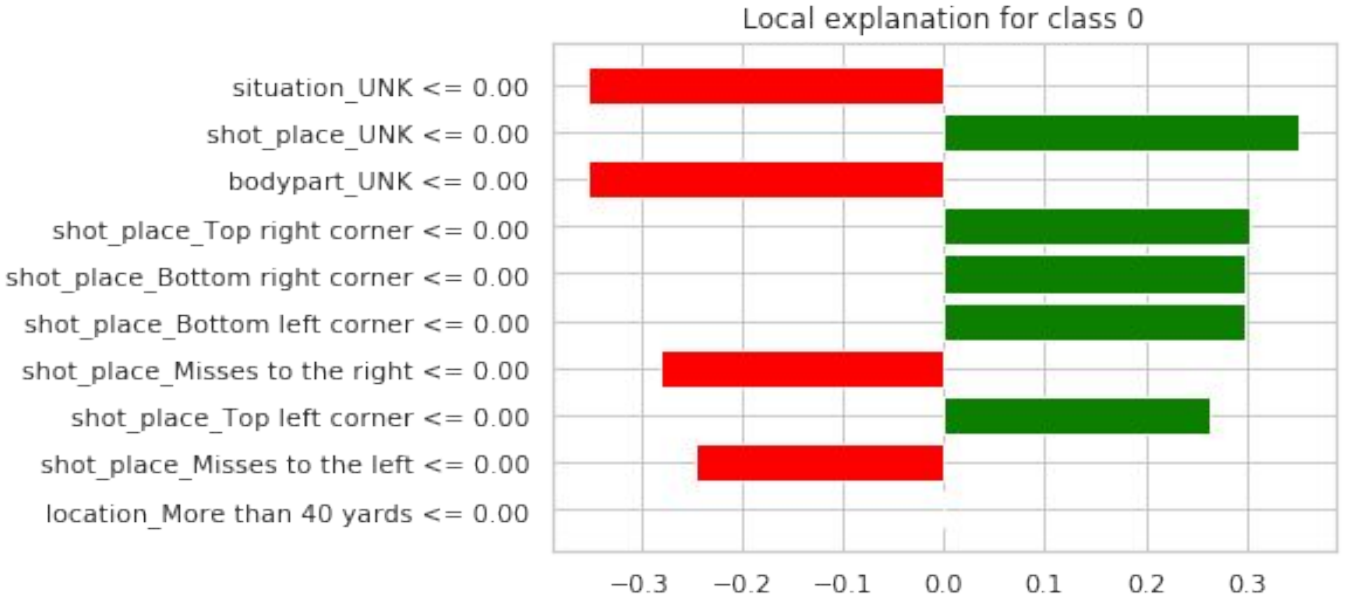
Why Remove Nulls



Interpreting Models

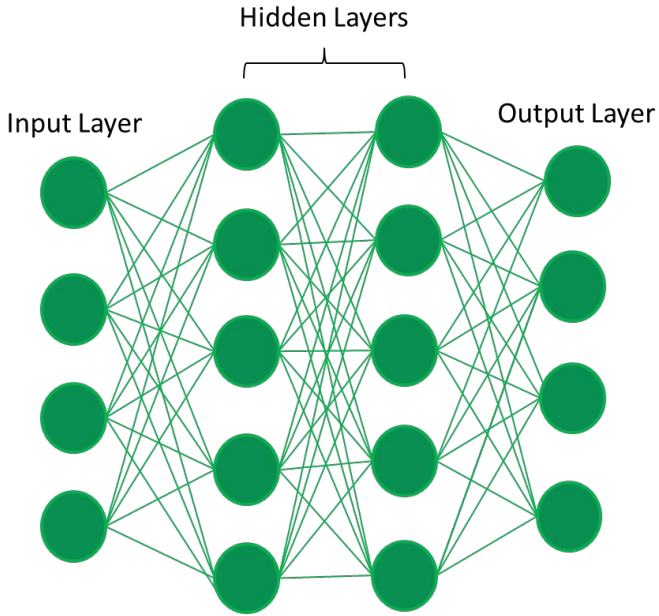


Interpreting Models



Deep Learning

- Using fully connected 2-layers with 200 and 100 units in each respectively.
- Using embeddings instead of one-hot-encoding for categorical features.
- Experiments (all done after balancing classes):
  - Adding more features like: player, team.
  - Both over-sampling and undersampling techniques.
- **Results**: always more than 98% accuracy.



Game Result from the Odds

- Odds are the amount of money that a better will receive for every 1 $ he bets on that result.
- We tried to predict game result from the odds but the accuracy was very low.

Model name	Accuracy
LR	54.8 %
RF	48 %
GB	54.2 %
Deep learning	52 %

2- Predicting Number of Goals in a Match

Predicting Number of Goals in a Match

Motivation: build a model predict the number of goals in a match according to match events.

To build this model we select 6 features:

1. `Event_type`
2. `Event_type2`
3. `Shot_place`
4. `Shot_outcome`
5. `Location`
6. `Side`

Our model use RNN as a classifier

- Inputs: 6 features
- Labels: numbers of goals in a match (home and away matches)

Predicting Number of Goals in a Match

Data pre-processing:

1. Home and away goals: separate each event.
2. One hot encoding.
3. Missing data: we put any missing data equal (-1).
4. Vectored: We select 5 features, In every match we have 180 events, so we have 2-d array size (5 X 180), we convert it into 1-d array (1 X 900).

Predicting Number of Goals in a Match Experiments:

➔ At every experiment we use one-hot encoding and without one-hot encoding

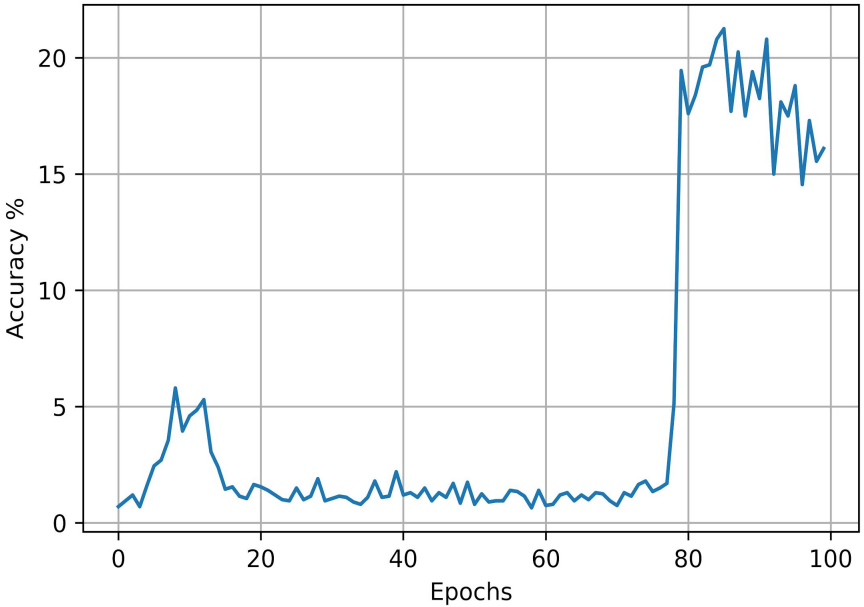
1- Combined Matches: Home and Away matches in one match:

- Add all events which shared the same game ID
- Summation number of goals in home and away matches.

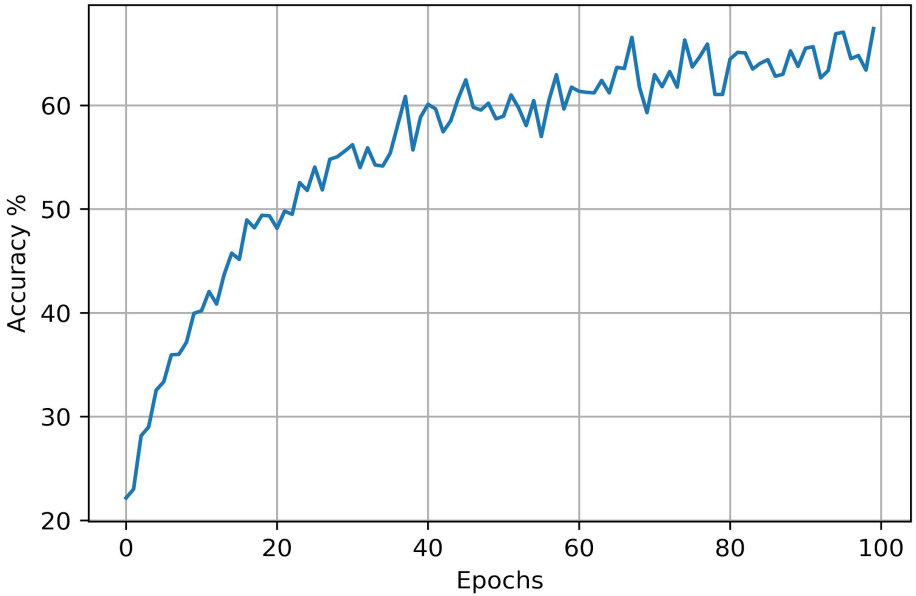
2-Separated Matches: Home Match and Away Match:

1. Home matches without one-hot encoding.
2. Away matches without one-hot encoding.
3. Home matches with one-hot encoding.
4. Away matches with one-hot encoding.

Predicting Number of Goals in a Match Results:

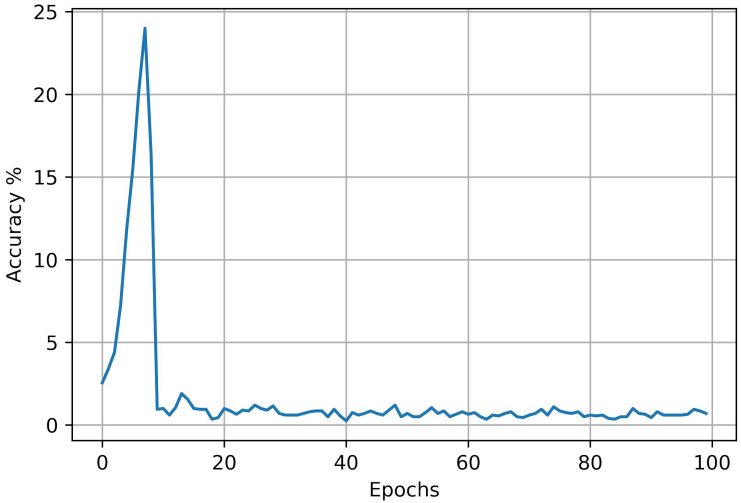


Combined Matches: without one-hot encoding

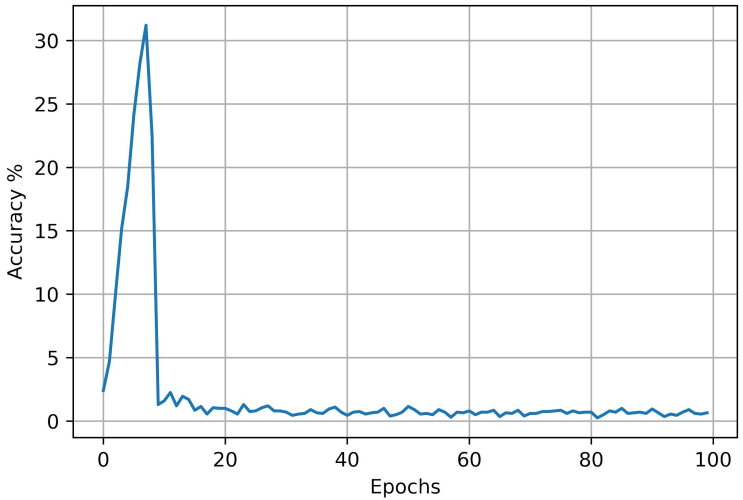


Combined Matches: with one-hot encoding

Predicting Number of Goals in a Match Results:

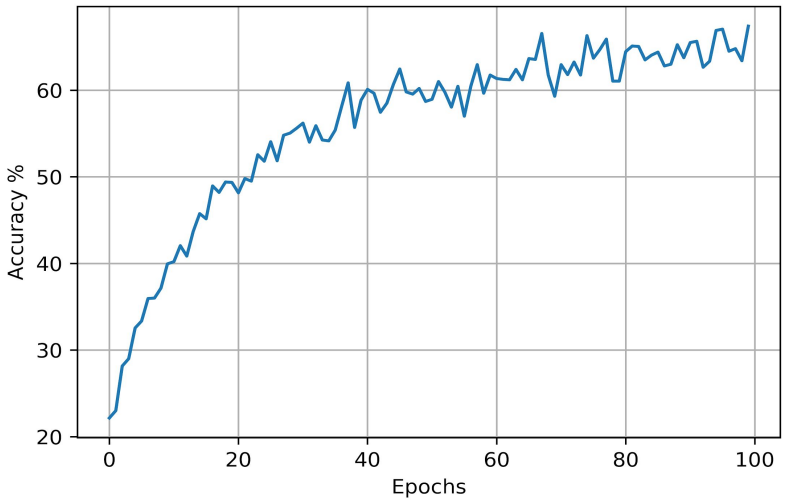


Home matches without one-hot encoding

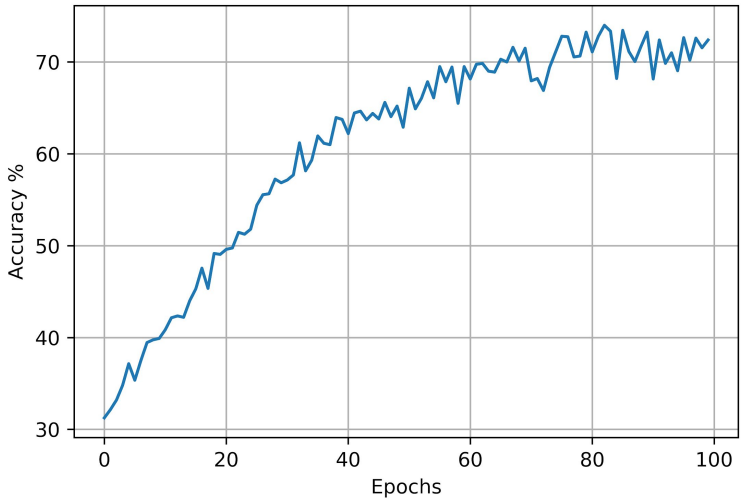


Away matches without one-hot encoding.

Predicting Number of Goals in a Match Results:



Home matches with one-hot encoding



Away matches with one-hot encoding

Predicting Number of Goals in a Match Results:

Combined Matches

	With One-hot encoding	Without One-hot encoding
Accuracy	**5.1 %**	**56.02 %**
Training time (in Minutes)	**108.25**	**84**

Predicting Number of Goals in a Match Results: Separated Matches

	Home with one-hot	Away with one-hot	Home without one-hot	Away without one-hot
Accuracy	1.726 %	2.27 %	56.017 %	60.937 %
Training time (in Minutes)	84.06	83.046	83.076	80.479

Challenges Faced & Lessons Learned

Challenges Faced

- Feature engineering
- Missing data values
- Imbalanced data
- Inefficient team collaboration

Solutions to the Challenges

- Studied the data set as a team, researched about the problem domain and used feature importance to select the best features.
- Depending on the task at hand, we either removed the rows with missing values or replaced them with placeholders.
- Generated Synthetic Samples.
- Created a GitHub repo, Google doc and overleaf documents to make collaboration faster and easier.

Lessons Learned

- Use of Git and GitHub
- Kaggle competitions structure
- Online collaborations
- Agile project development
- Application of theory to real world applications

Conclusion

- Football data analysis is a wide area of research that affects a lot of people: spectators, coaches, teams, players, and football investors.
- In this project, we aimed at finding the strengths and weaknesses of players and teams in a league and classified them according to strong and weak teams.
- This was mainly meant to provide some insight for the decision makers.
- We also predicted the total number of goals in a match and the probability of goal being scored given the events.
- We learned a lot from the project development and hope to pass the knowledge on in future machine learning projects.