A game of football has much more scope for analysis than the commonly used statistics like number of goals, number of cards etc. To analyze the events and their impact and also the betting stats, lets look at a secondary dataset.

	id_odsp	id_event	sort_order	time	text	event_type	event_type2	side	event_team
0	UFot0hit/	UFot0hit1	1	2	Attempt missed. Mladen Petric (Hamburg) left f...	1	12.0	2	Hamburg SV
1	UFot0hit/	UFot0hit2	2	4	Corner, Borussia Dortmund. Conceded by Dennis...	2	NaN	1	Borussia Dortmund

Out[4]: (941009, 22)

So we have almost a million events in the dataset

We have one more file in the dataset. Lets look at it as well

	id_odsp	link_odsp	adv_stats	date	league	season	country	ht
0	UFot0hit/	/soccer/germany/bundesliga-2011-2012/dortmund-...	True	2011-08-05	D1	2012	germany	Borussia Dortmund
1	Aw5DflLH/	/soccer/germany/bundesliga-2011-2012/augsburg-...	True	2011-08-06	D1	2012	germany	FC Augsburg
2	bkjpaC6n/	/soccer/germany/bundesliga-2011-2012/werder-br...	True	2011-08-06	D1	2012	germany	Werder Bremen
3	CzPV312a/	/soccer/france/ligue-1-2011-2012/paris-sg-lori...	True	2011-08-06	F1	2012	france	Paris Saint-Germain
4	GUOdmtIl/	/soccer/france/ligue-1-2011-2012/caen-valencie...	True	2011-08-06	F1	2012	france	Caen

	id_odsp	date	league	season	country	odd_h	odd_d	odd_a	fthg	ftag	prob_h	prob_
0	UFot0hit/	2011-08-05	D1	2012	germany	1.56	4.41	7.42	3	1	2.56	8.4
1	Aw5DflLH/	2011-08-06	D1	2012	germany	2.36	3.60	3.40	2	2	3.36	4.4
2	bkjpaC6n/	2011-08-06	D1	2012	germany	1.83	4.20	4.80	2	0	2.83	5.8
3	CzPV312a/	2011-08-06	F1	2012	france	1.55	4.50	9.40	0	1	2.55	10.4
4	GUOdmtII/	2011-08-06	F1	2012	france	2.50	3.40	3.45	1	0	3.50	4.4

This dataset uses cryptic names like D1, F1 etc for the leagues. Lets replace the values to have the same set of names that we used in the original match result dataset

Out[9]:

league	country	id_odsp	date	league	season	country	odd_h	odd_d	odd_a	fthg	ftag	pro
D1	germany	1690	526	1	6	1	358	240	500	10	8	
E0	england	2120	553	1	6	1	358	227	498	9	7	
F1	france	2107	594	1	6	1	347	224	507	7	9	
I1	italy	2106	541	1	6	1	361	268	516	8	8	
SP1	spain	2089	681	1	6	1	390	304	572	11	9	

So E0 means EPL, SP1 means La Liga, I1 means Serie A, F1 means Ligue1 and D1 means Bundesliga

Out[11]:

league	country	id_odsp	date	league	season	country	odd_h	odd_d	odd_a	fthg	ftag
Bundesliga	germany	1690	526	1	6	1	358	240	500	10	8
EPL	england	2120	553	1	6	1	358	227	498	9	7
LaLiga	spain	2089	681	1	6	1	390	304	572	11	9
Ligue1	france	2107	594	1	6	1	347	224	507	7	9
SerieA	italy	2106	541	1	6	1	361	268	516	8	8

Lets merge the two dataframes together

	id_odsp	id_event	sort_order	time	text	event_type	event_type2	side	event_team
0	UFot0hit/	UFot0hit1	1	2	Attempt missed. Mladen Petric (Hamburg) left f...	1	12.0	2	Hamburg SV
1	UFot0hit/	UFot0hit2	2	4	Corner, Borussia Dortmund. Conceded by Dennis...	2	NaN	1	Borussia Dortmund

The data has a lot of columns with cryptic numeric values. Lets use the dictionary file to replace the cryptic values with meaningful data instead

Lets look at the dataset once again now

	id_odsp	id_event	sort_order	time	text	event_type	event_type2	side	event_team
0	UFot0hit/	UFot0hit1	1	2	Attempt missed. Mladen Petric (Hamburg) left f...	Attempt	Key Pass	Away	Hamburg SV
1	UFot0hit/	UFot0hit2	2	4	Corner, Borussia Dortmund. Conceded by Dennis...	Corner	NaN	Home	Borussia Dortmund
2	UFot0hit/	UFot0hit3	3	4	Corner, Borussia Dortmund. Conceded by Heiko ...	Corner	NaN	Home	Borussia Dortmund
3	UFot0hit/	UFot0hit4	4	7	Foul by Sven Bender (Borussia Dortmund).	Foul	NaN	Home	Borussia Dortmund
4	UFot0hit/	UFot0hit5	5	7	Gokhan Tore (Hamburg) wins a free kick in the ...	Free kick won	NaN	Away	Hamburg SV

Now the event dataset is ready for analysis

Lets visualize the number of occurences for each type of event

```
<Figure size 432x288 with 0 Axes>
```
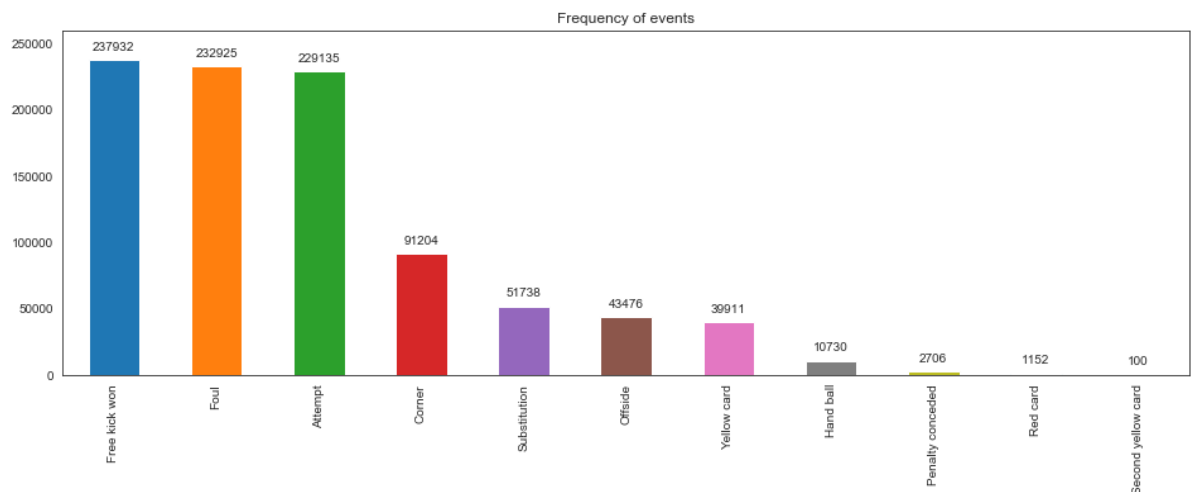
Overall, the most frequent events are :

1. Free kick won
2. Foul
3. Attempt

Lets look at a league wise view of the number of events

a) **English Premier League**

<Figure size 432x288 with 0 Axes>
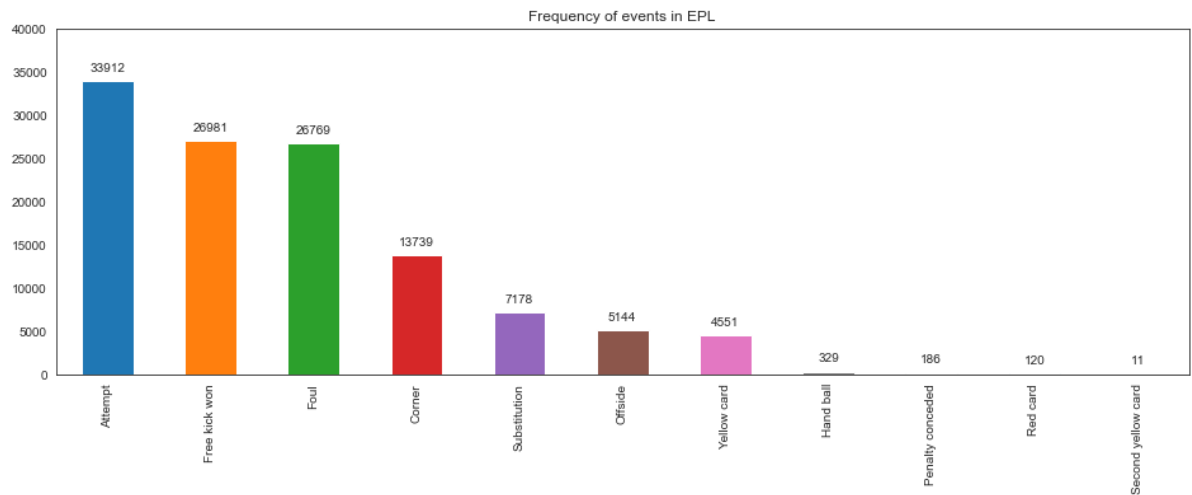
Foul moves from no 2 to no 3 position. This is consistent with our observation from the previous dataset that EPL has lesser number of fouls.

Attempt moves from no 3 to no 1 position. This is consistent with our observation from the other dataset that EPL has higher number of shots at goal as compared to other leagues

b) **La Liga**
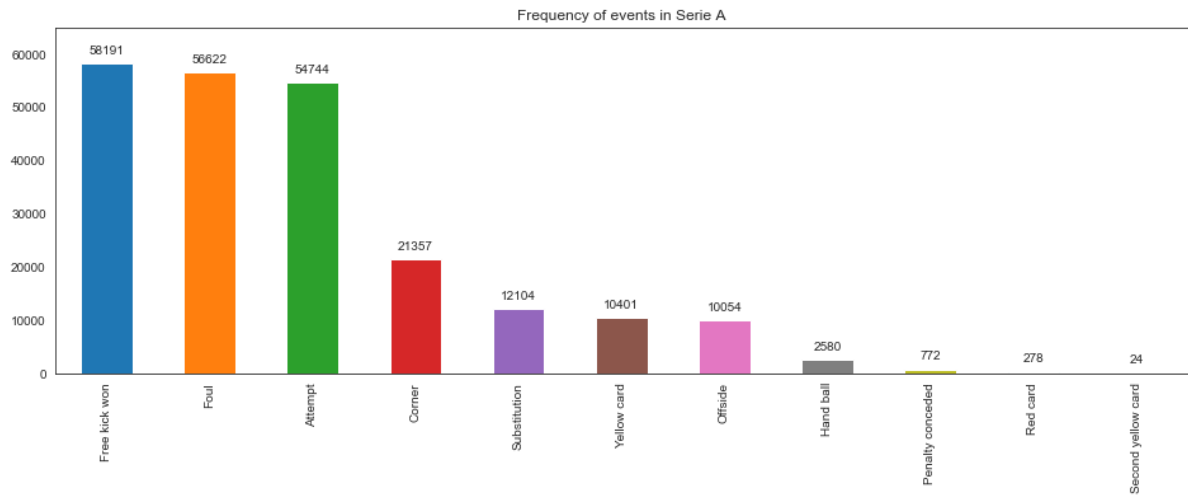
```
<Figure size 432x288 with 0 Axes>
```


Frequency of events in La Liga

c) Serie A

```
<Figure size 432x288 with 0 Axes>
```


Frequency of events in Serie A

d) Bundesliga

```
<Figure size 432x288 with 0 Axes>
```


Frequency of events in Bundesliga

e) **Ligue One**

```
<Figure size 432x288 with 0 Axes>
```


Frequency of events in Ligue 1
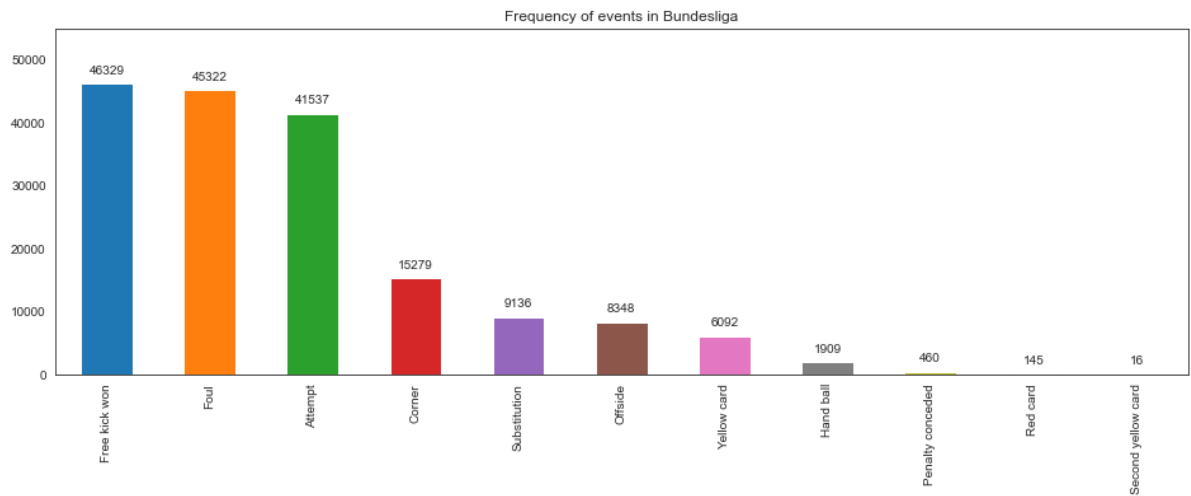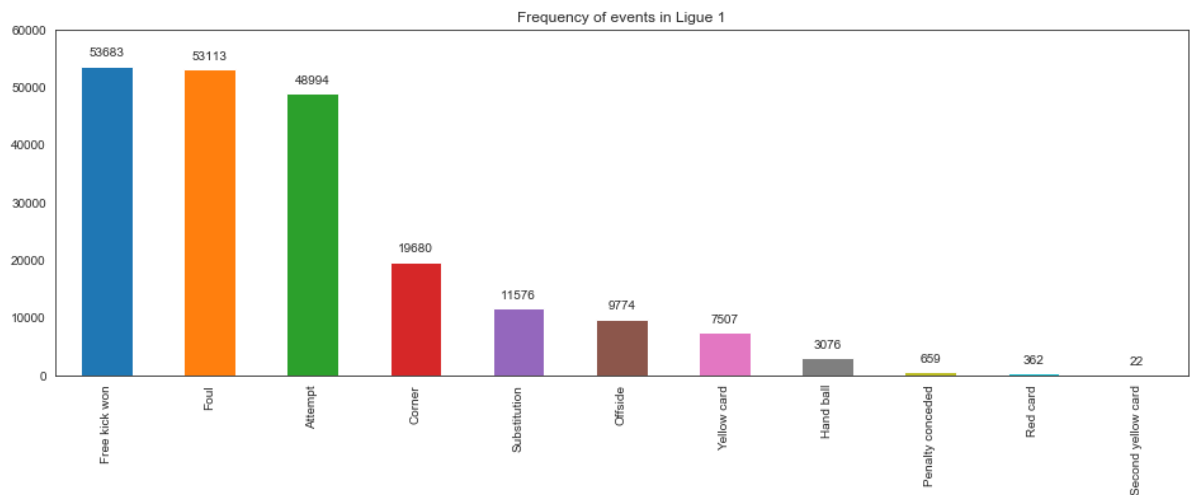
So the top 3 events in every league are Free kick won, Fouls and Attempts. EPL witnesses almost half the number of Fouls as LaLiga even though both the leagues involve equal number of games per season

Where to shoot if you want to score a goal ?

Lets look at the shots that resulted in goals

```
<Figure size 432x288 with 0 Axes>
```
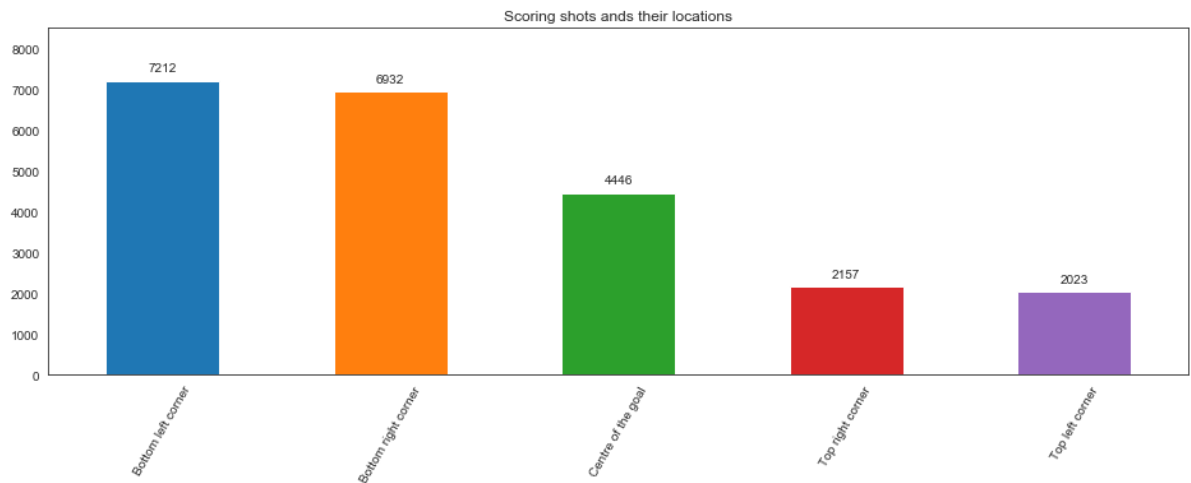

The data suggests that its more effective to shoot at the bottom than to shoot at the top or centre

Lets have a look at the missed attempts as well

```
<Figure size 432x288 with 0 Axes>
```
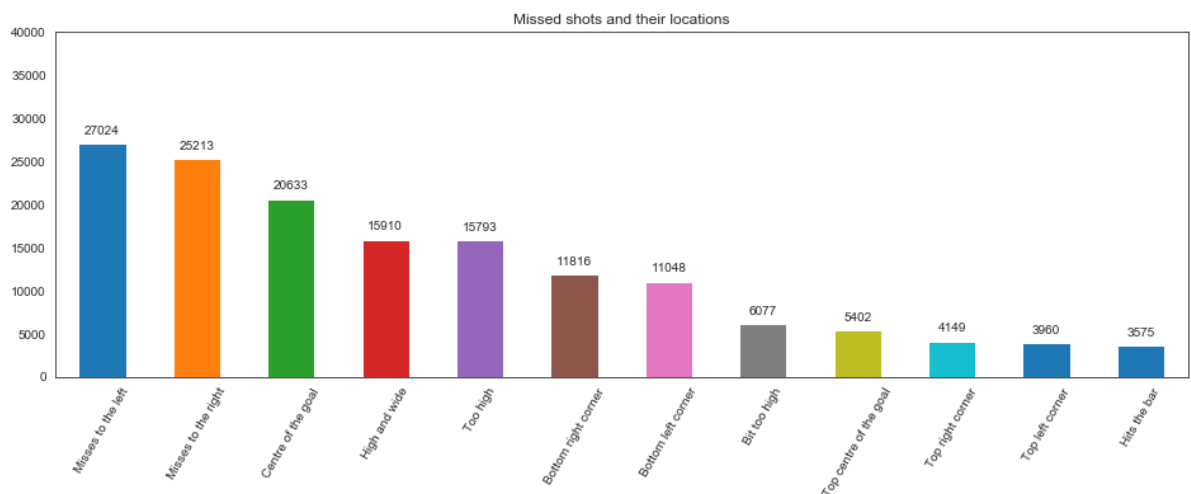

The number of saves in bottom left/right corner is quite low as opposed to centre of the box

There are fewer saves at top left/right corner but the chances of missing are high (might hit the bar or may be too high)

So, we can conclude that bottom right or left corners are the best locations to aim at while trying to score a goal

Which teams and players are the most prolific scorers ?

Lets have a look at the most heavily scoring teams first

`<Figure size 432x288 with 0 Axes>`
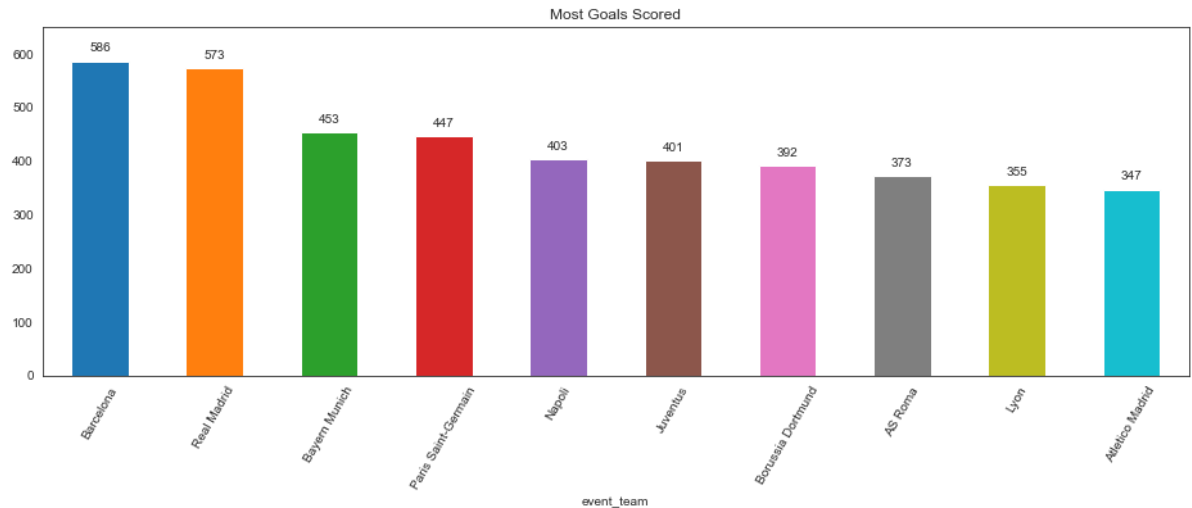
Barcelona and Real Madrid are at the top. This is also in line with our observation that La Liga witnesses high number of goals.

Interestingly none of the EPL teams make it to this list

Lets have a look at most prolific individual scorers now

`<Figure size 432x288 with 0 Axes>`
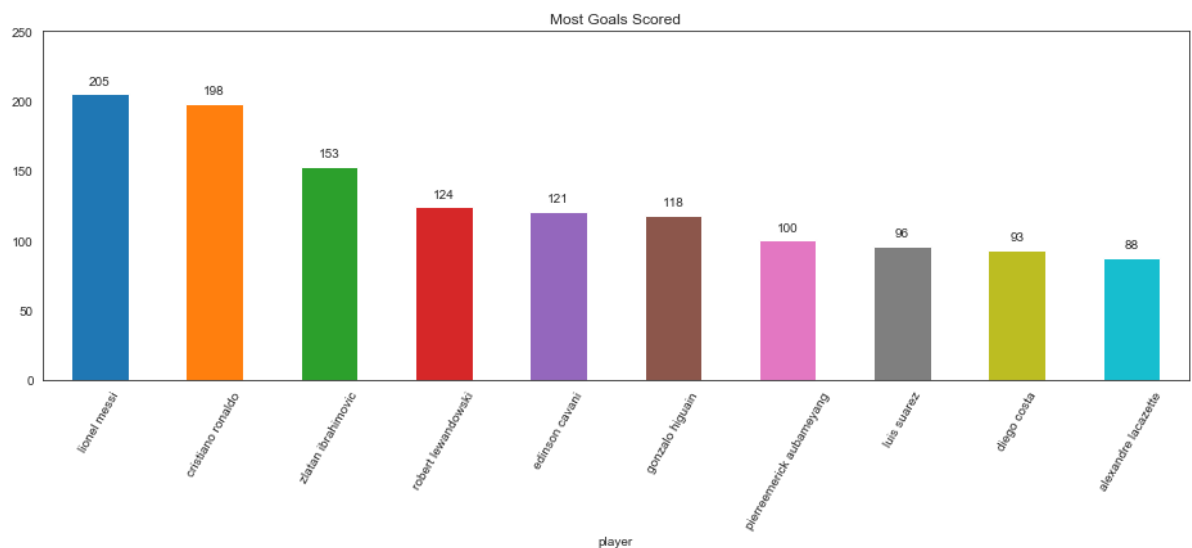
No surprises there in the top two names !!!

Zlatan Ibrahimovic is the only player to make it to this list after playing in three different leagues during this period

The players in the above chart might have high number of misses as well. Lets look at the players with maximum number of missed attempts

`<Figure size 432x288 with 0 Axes>`
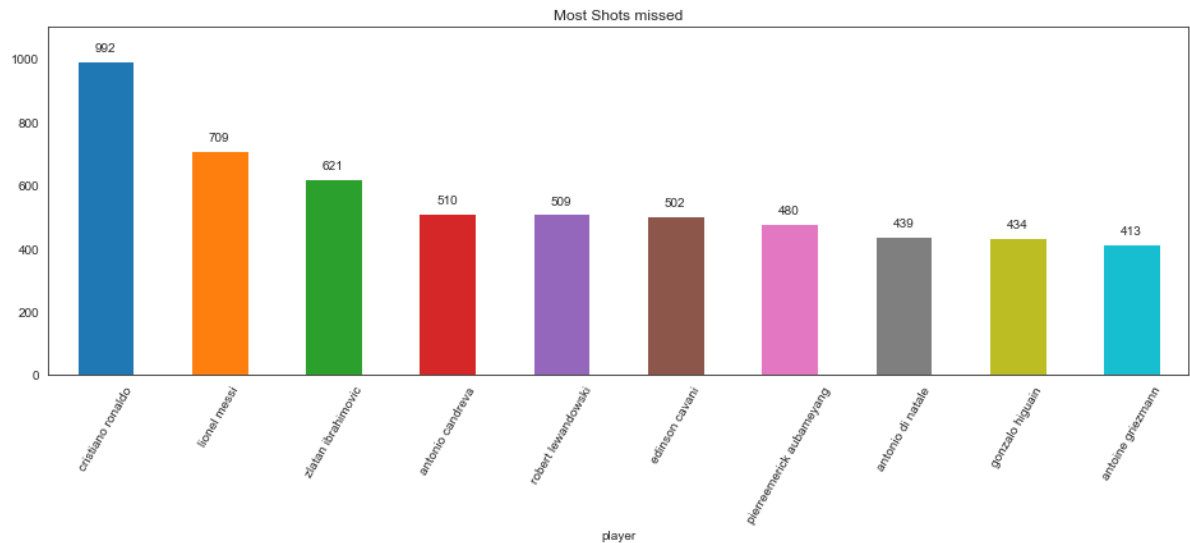
The top 3 are still Cristiano Ronaldo, Lionel Messi and Zlatan Ibrahimovic

Let's have a look at total number of Attempts as well

`<Figure size 432x288 with 0 Axes>`
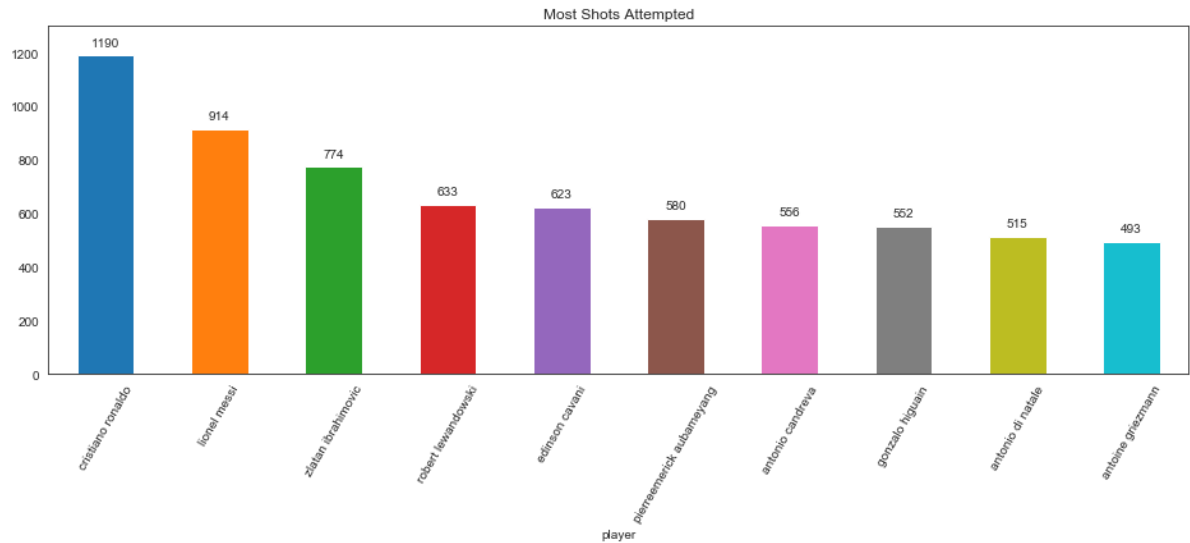
No wonder Christiano Ronaldo is considered as one of the greatest forwards of all time. He is almost 20 % ahead of Messi at second position

```
<Figure size 432x288 with 0 Axes>
```
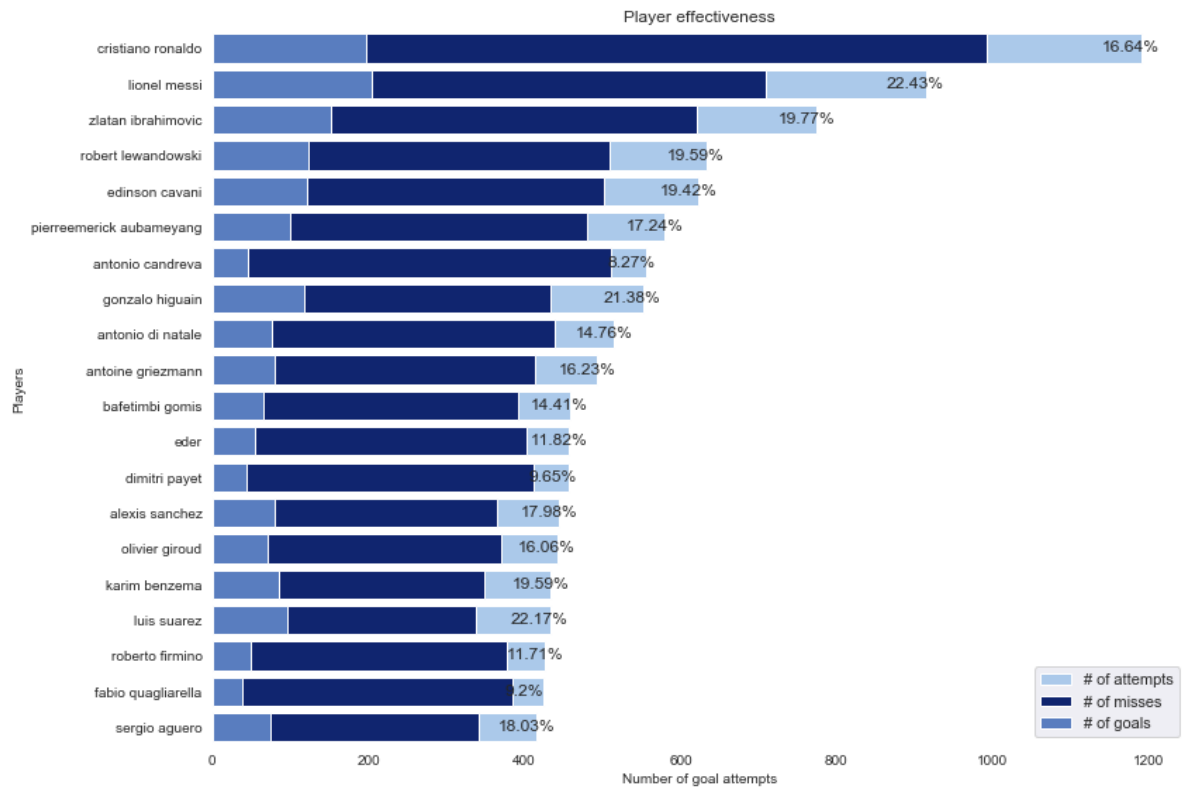

Clearly Christiano Ronaldo and Lionel Messi are considered among greatest off all times because they have been consistently creating chances.

Their teammates like Suarez and Higuain have better attempt-to-goal-conversion-rate but doesn't create half as many chances as Messi and Ronaldo
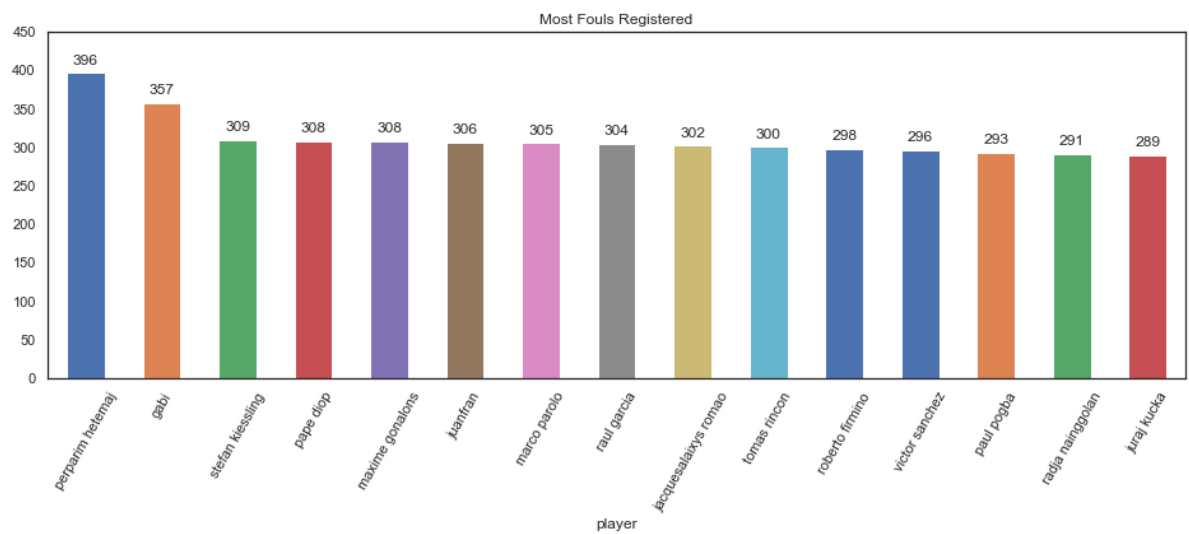
Which teams and players commit the maximum number of fouls ?

<Figure size 432x288 with 0 Axes>
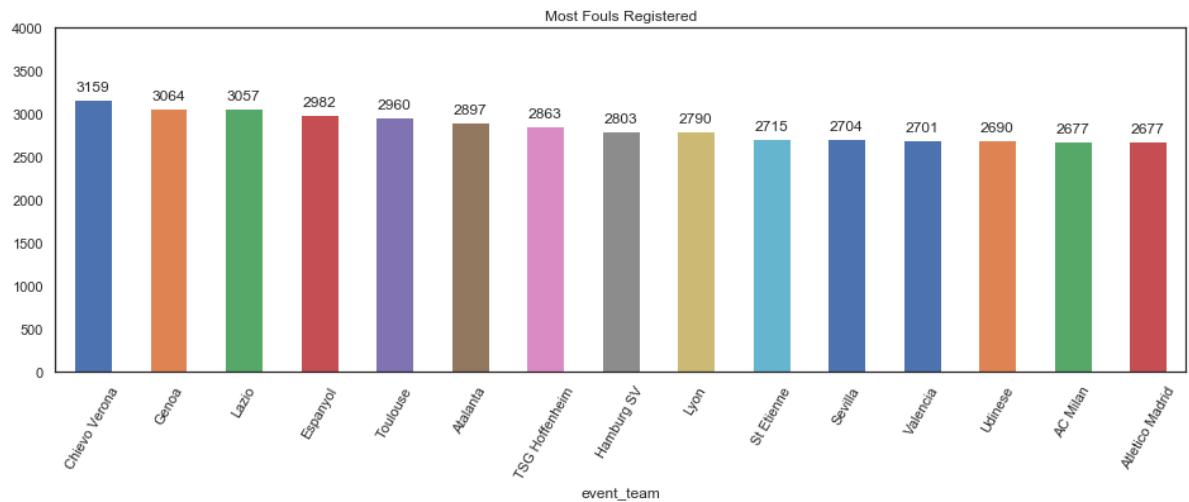
More fouls are being committed by lower ranked teams in the league. Lyon and Lazio are the only top tier teams with very high number of fouls

<Figure size 432x288 with 0 Axes>

Surprising to not see Sergio Ramos among top 15

Lets look the card data as well, there should be some correlation between the two lists
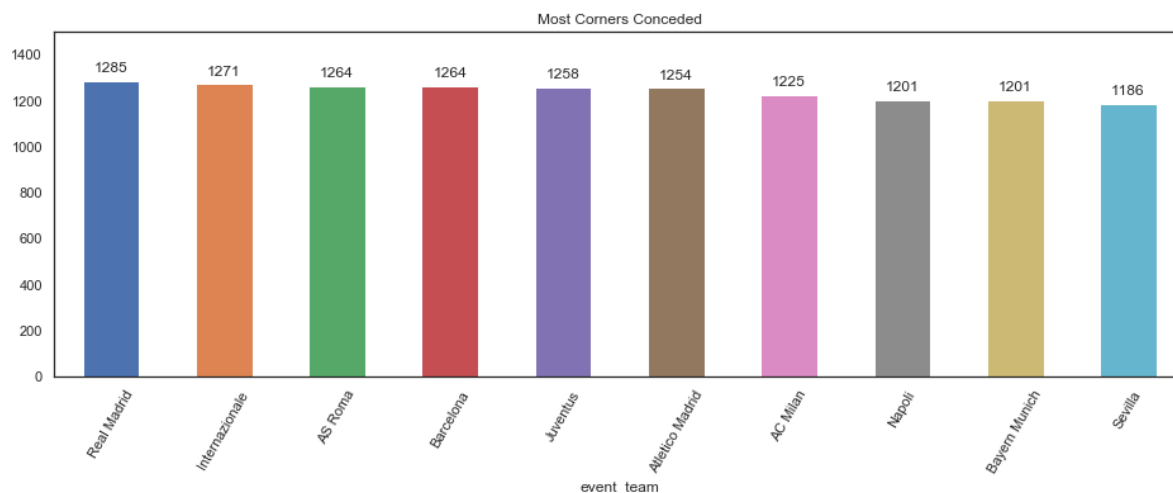
<Figure size 432x288 with 0 Axes>
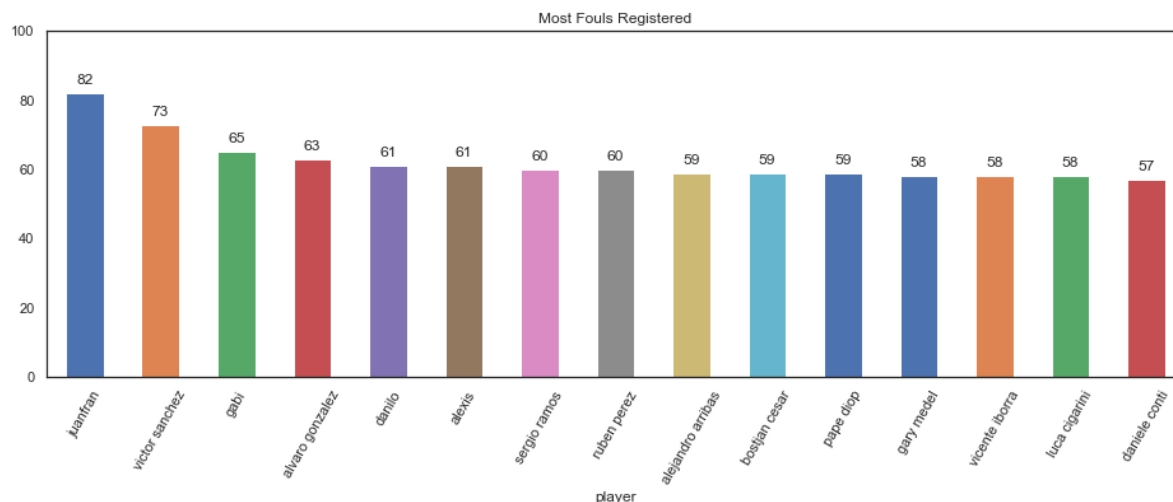
Gabi from Athletico Madrid is among the worst offenders - he features in top 3 in terms of both the number of fouls committed as well as the number of cards

Which teams and players concede the maximum number of corners ?

<Figure size 432x288 with 0 Axes>

The top 10 list is dominated by La Liga (Real Madrid, Barcelona, Athletico Madrid, Sevilla) and Serie A (Internazionale, AS Roma, Juventus, AC Milan, Napoli)
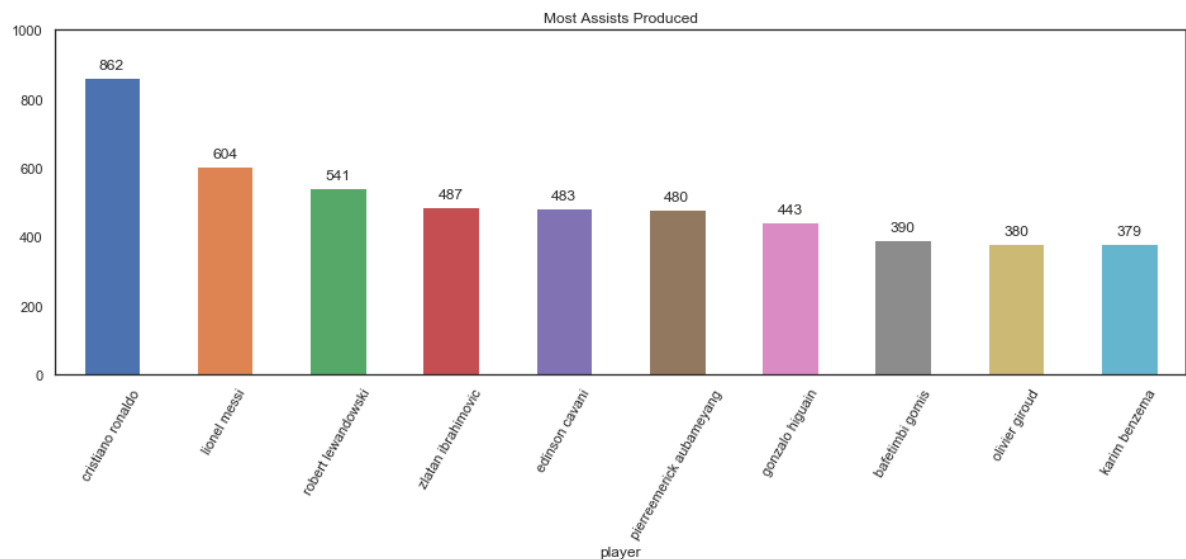
So we can clearly conclude that teams in La Liga and Serie A concede more corners compared to other leagues

`<Figure size 432x288 with 0 Axes>`
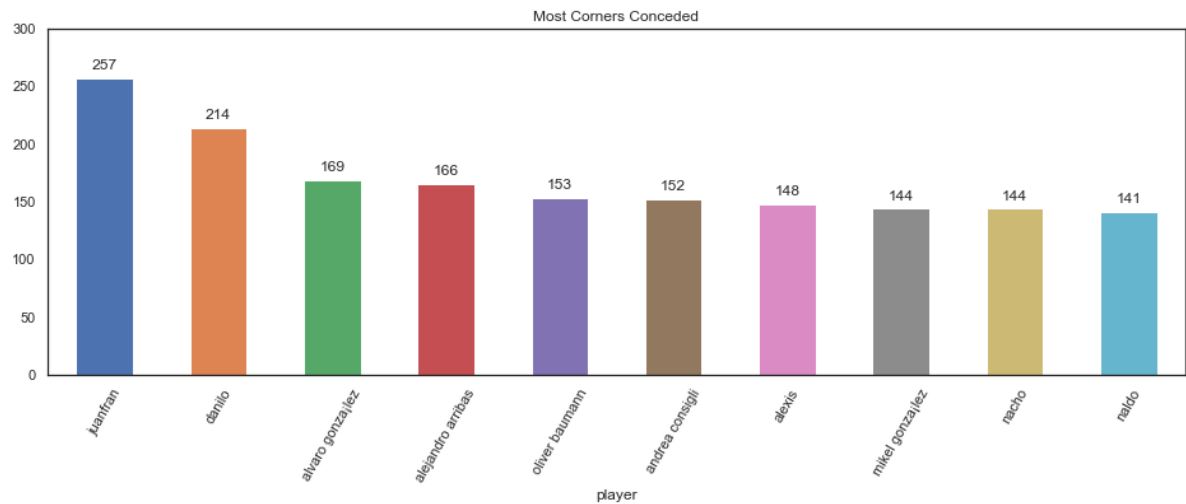
Which players contribute the most through assists ?

`<Figure size 432x288 with 0 Axes>`

Again we have the two usual suspects at the top. Cristiano Ronaldo tops the number of assists as well (and is ahead of Messi at no 2 by 30 %)

Higuain and Benzema also appear in the top 10 list indicating that they played a significant role behind Christiano Ronaldo's success as a goalscorer

Messi vs Ronaldo - What does data indicate ?

Lets build a player comparator

	Goals	Misses	Attempts	Assists	Cards	Fouls	Offsides
lionel messi	205	709	914	604	17	68	89
cristiano ronaldo	198	992	1190	862	29	100	206
zlatan ibrahimovic	153	621	774	487	29	248	225
gonzalo higuain	118	434	552	443	21	155	190
luis suarez	96	337	433	322	22	113	161
karim benzema	85	349	434	379	4	91	137
gareth bale	50	251	301	234	9	61	52
harry kane	65	288	353	278	9	85	78
neymar	58	260	318	248	23	106	69

When are the most goals scored during a football match ?

```
<Figure size 432x288 with 0 Axes>
```
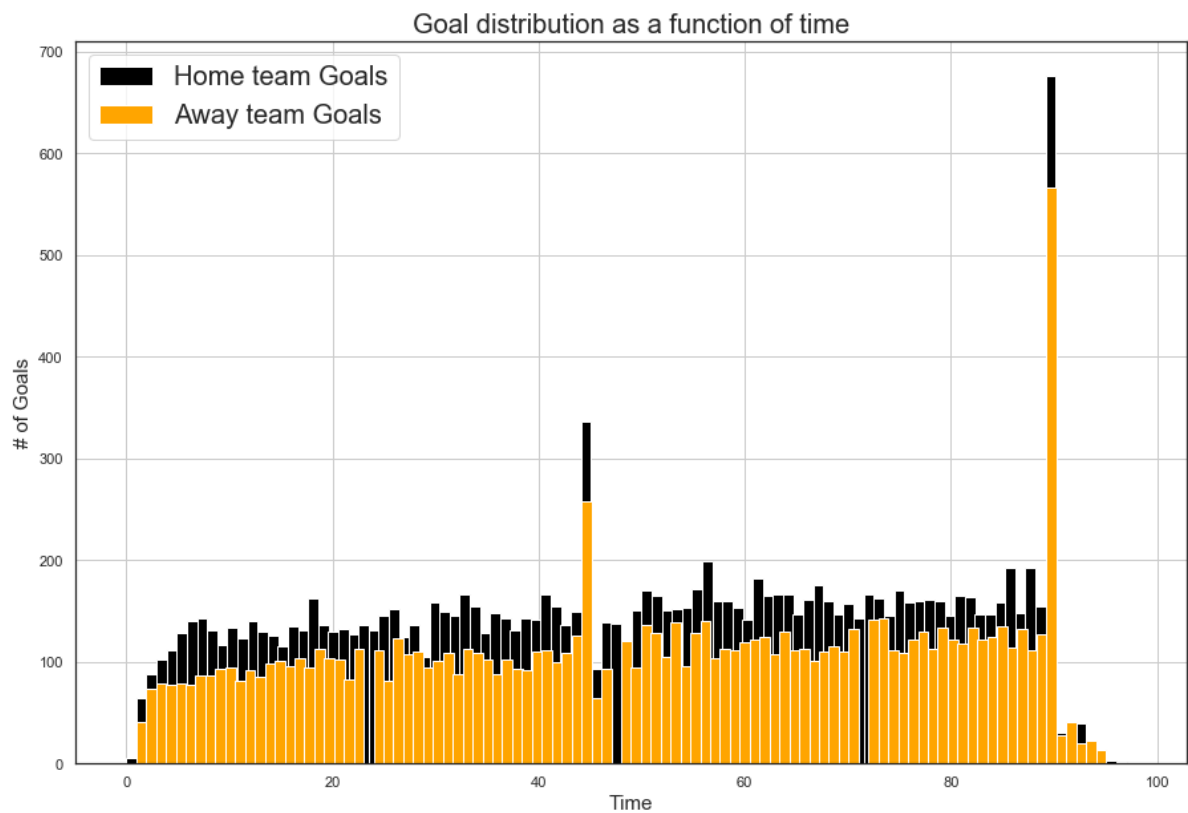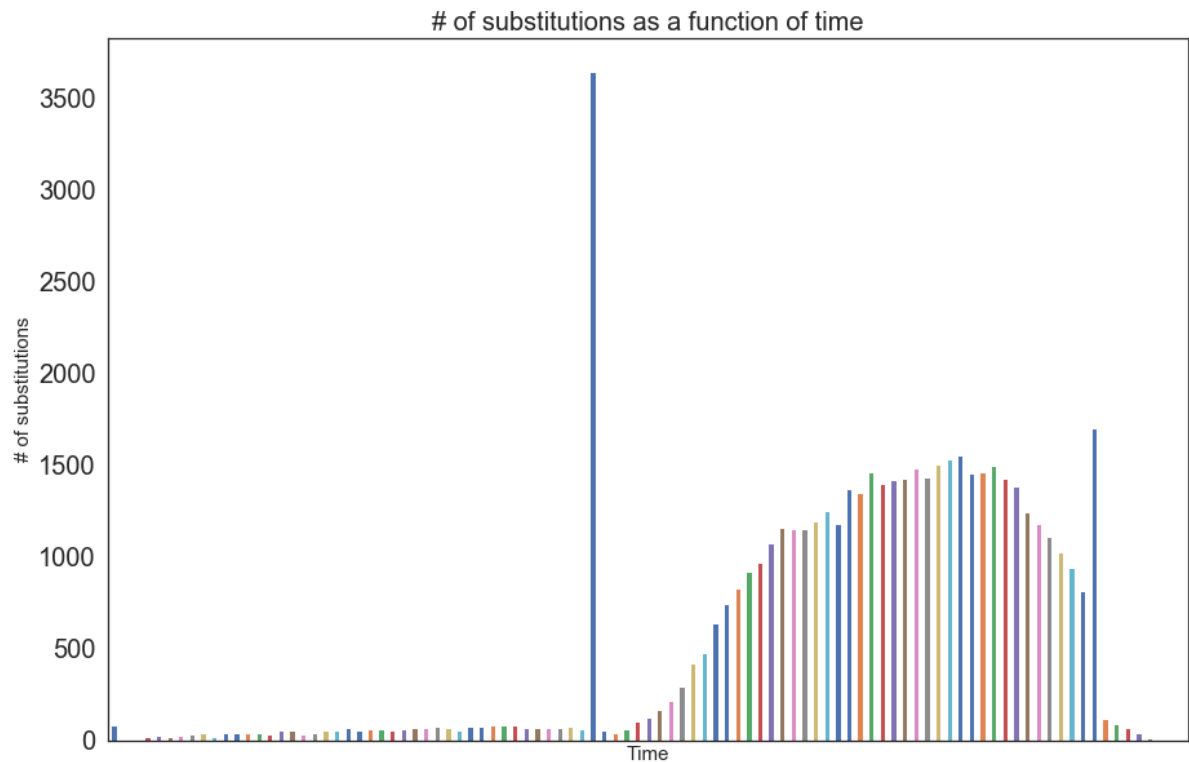

Interestingly there's a massive spike in number of goals just before half time and full time

When do the most substitutions occur ?
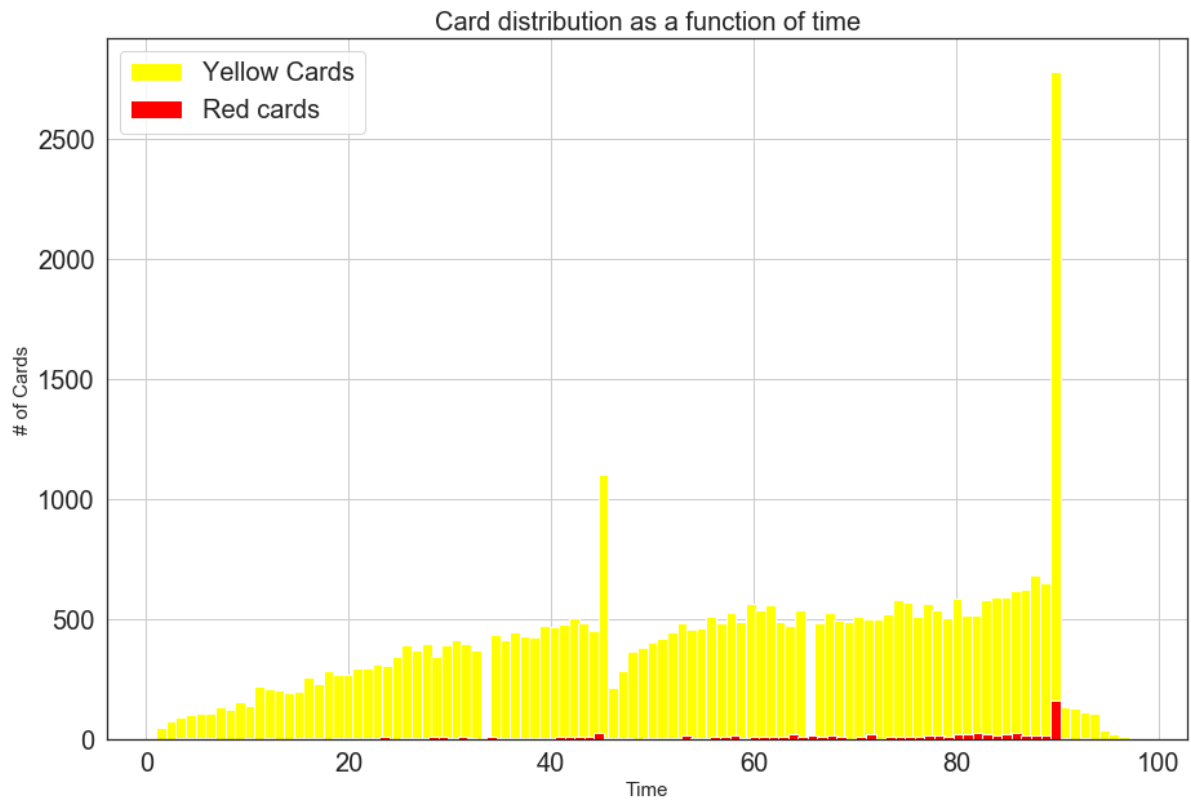
`<Figure size 432x288 with 0 Axes>`

There's a massive spike in number of substitutions at around half time More substitutions happen during the second half of a match which makes logical sense

When are the most number of red/yellow cards shown during a football match ?

```
<Figure size 432x288 with 0 Axes>
```

Card distribution as a function of time

There's a spike in number of yellow cards just before half time and full time Most red cards are shown in last few minutes of a match (possibly when opponent is trailing by a goal and desperate to score and thus, the stakes are high))

Which players provide the most number of early breakthroughs ?

For our analysis we'll consider a goal scored in first 10 minutes of a match as an early breakthrough

`<Figure size 432x288 with 0 Axes>`

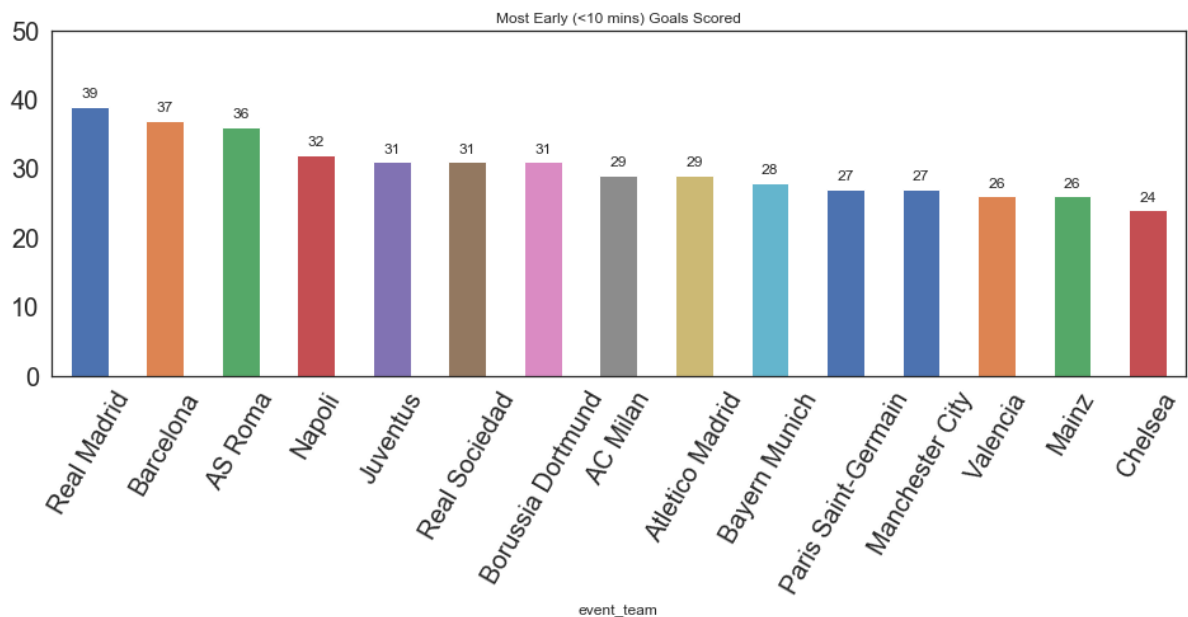Most Early (<10 mins) Goals Scored
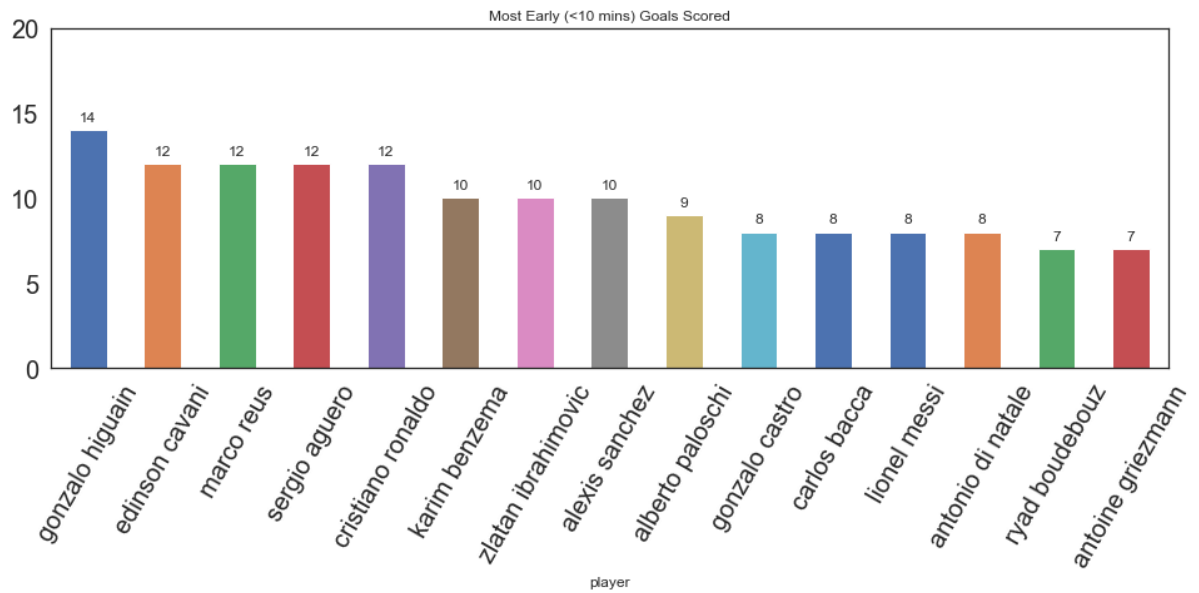
There are very few goals that are scored in first 10 minutes of a match. Real Madrid has 3 players in the top 5 list
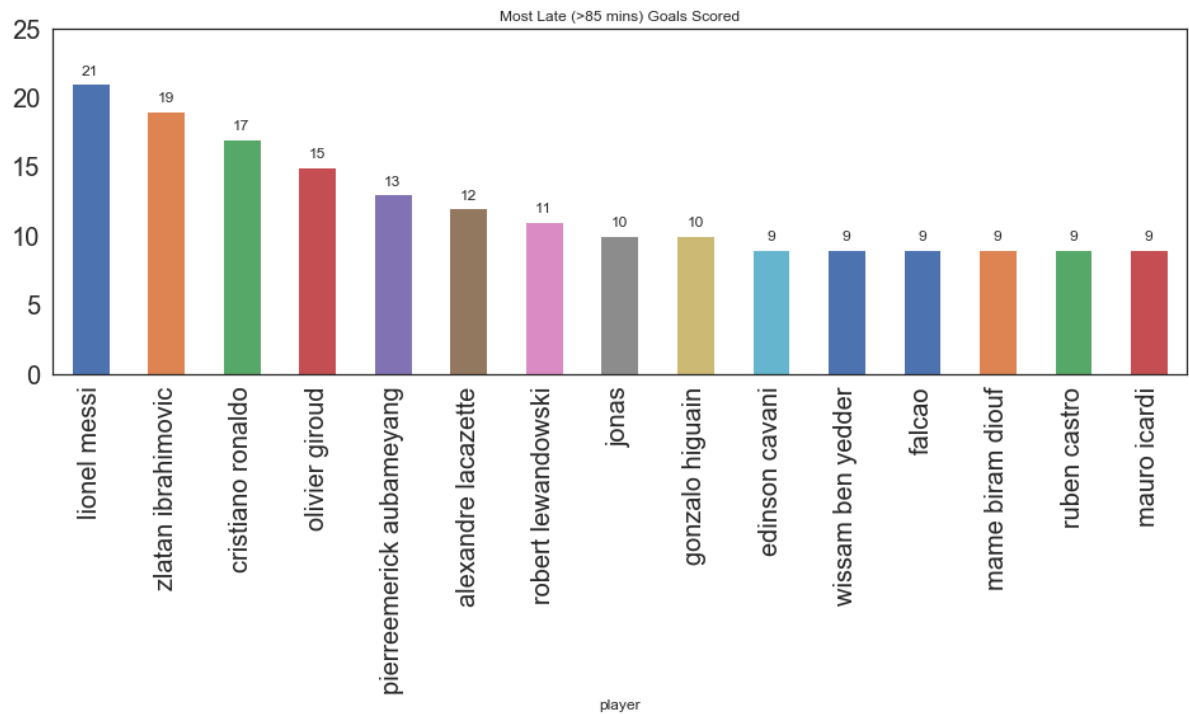
Let's look the corresponding team data as well

`<Figure size 432x288 with 0 Axes>`

Most Early (<10 mins) Goals Scored

No surprises there. Real Madrid scores the maximum number of goals in first 10 minutes of a match

Which players tend to score the most during last few minutes ?

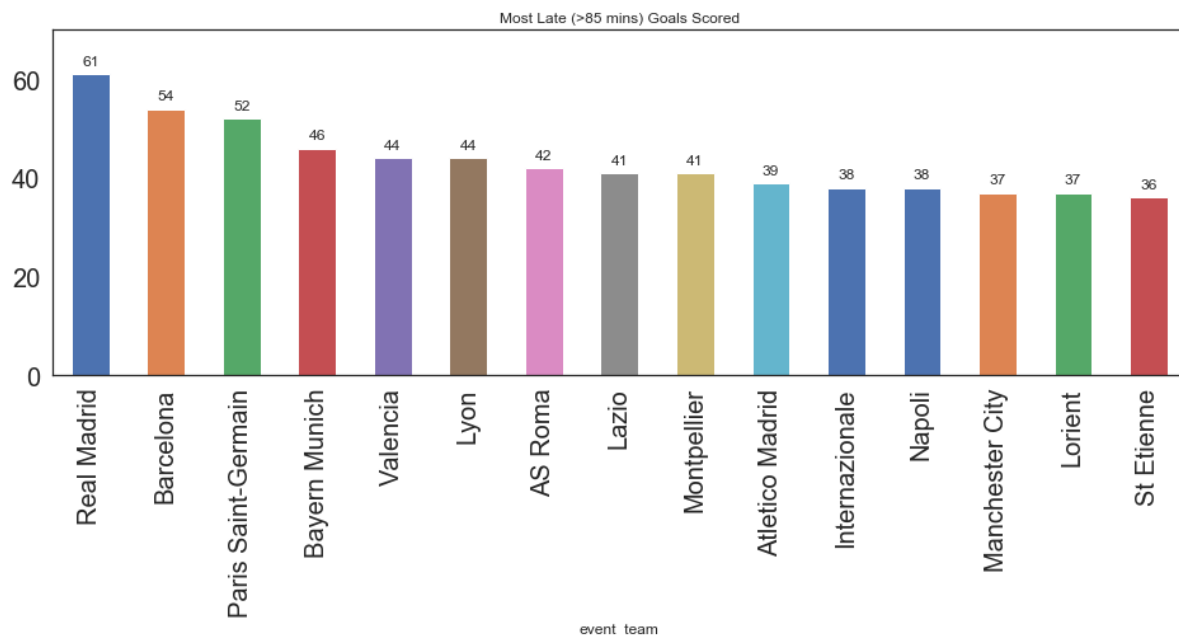For this, we will only consider goals scored after 85 minutes

<Figure size 432x288 with 0 Axes>

Again we have the same trio at top 3

Let's look at team level data as well
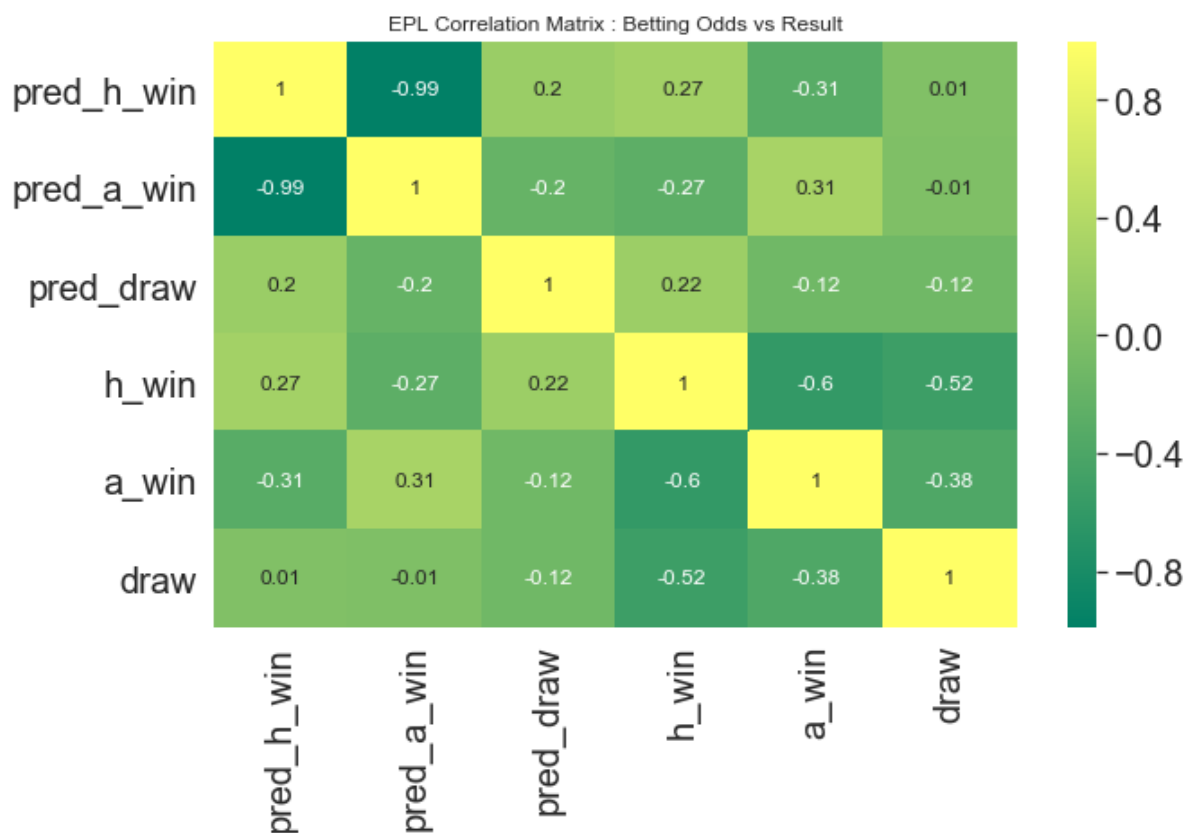
```
<Figure size 432x288 with 0 Axes>
```


Just like early goals, Real Madrid leads in terms of number of late goals scored too

Whats the best league to bet if you are a rookie ?

Let's do a league-wise analysis

A. **EPL**

<Figure size 432x288 with 0 Axes>

EPL Correlation Matrix : Betting Odds vs Result
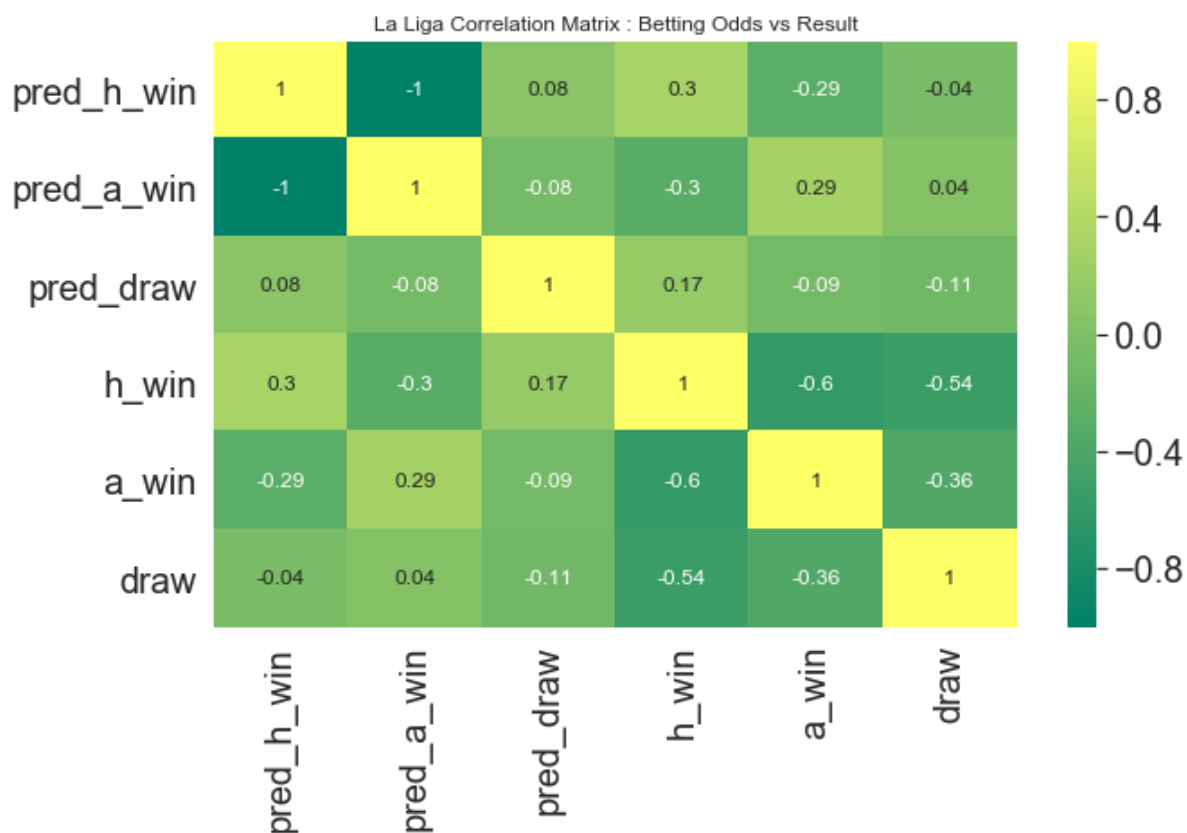
The betting odds have the following correlation with the actual results:

a) Home Team Win : 0.38

b) Away Team Win : 0.37

c) Draw : 0.14

b) **La Liga**

```
<Figure size 432x288 with 0 Axes>
```
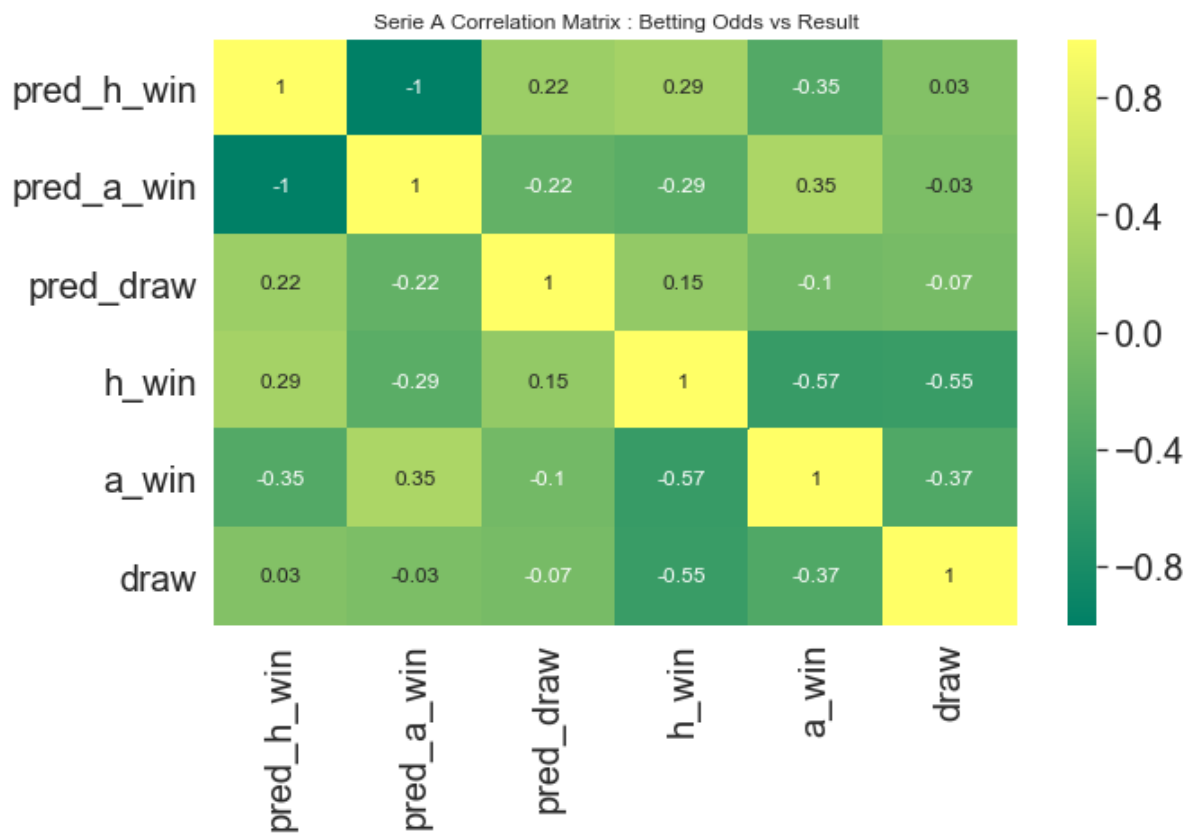

La Liga Correlation Matrix : Betting Odds vs Result

The betting odds have the following correlation with the actual results:

a) Home Team Win : 0.41

b) Away Team Win : 0.38

c) Draw : 0.16

c) **Serie A**

<Figure size 432x288 with 0 Axes>
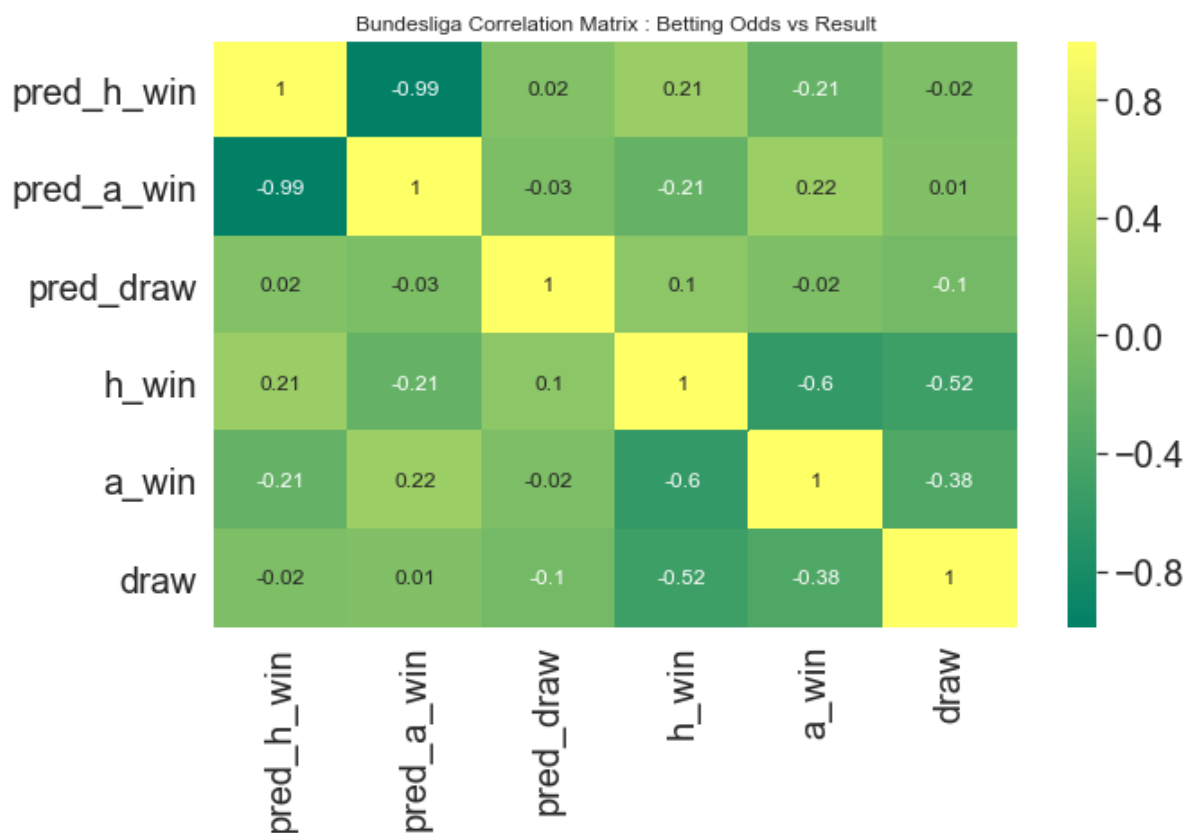
Serie A Correlation Matrix : Betting Odds vs Result

The betting odds have the following correlation with the actual results:

a) Home Team Win : 0.36

b) Away Team Win : 0.39

c) Draw : 0.1

d) **Bundesliga**

`<Figure size 432x288 with 0 Axes>`
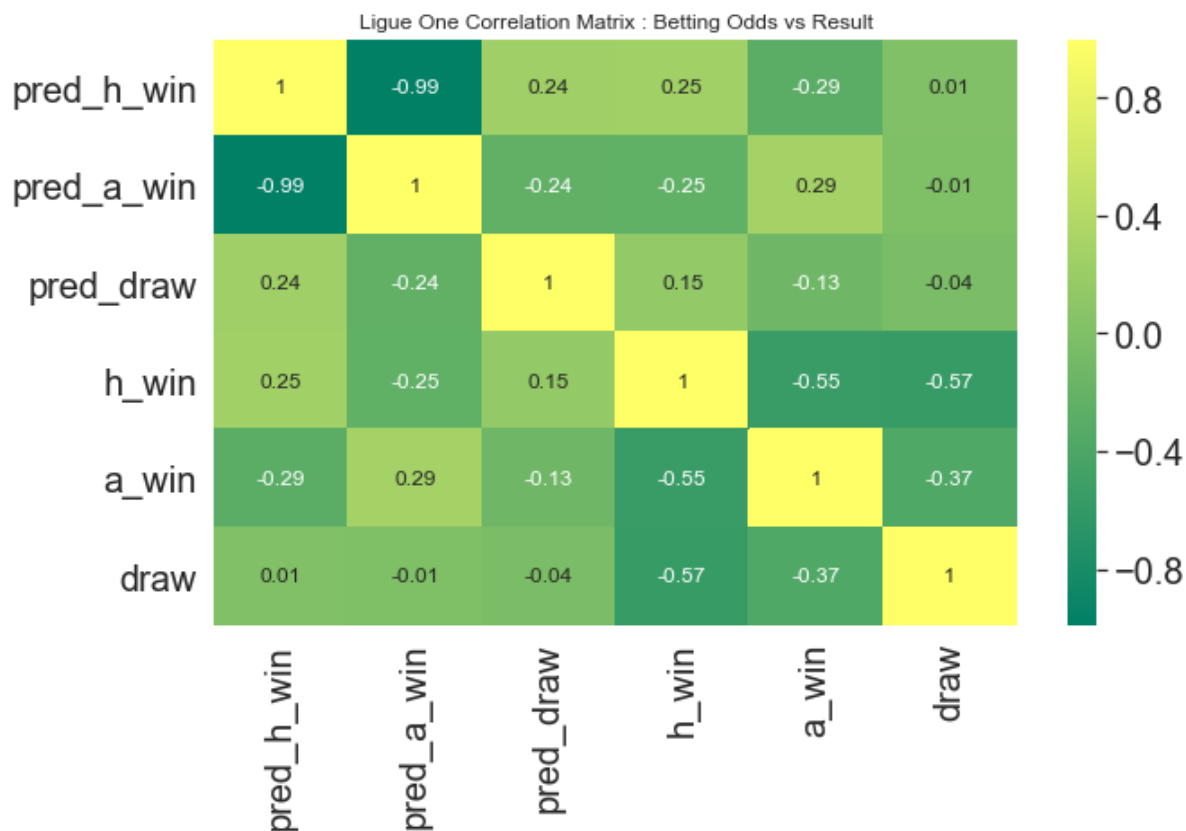

Bundesliga Correlation Matrix : Betting Odds vs Result

The betting odds have the following correlation with the actual results:

a) Home Team Win : 0.32

b) Away Team Win : 0.32

c) Draw : 0.13

e) **Ligue One**

```
<Figure size 432x288 with 0 Axes>
```


Ligue One Correlation Matrix : Betting Odds vs Result

The betting odds have the following correlation with the actual results:
a) Home Team Win : 0.3
b) Away Team Win : 0.33
c) Draw : 0.1

In general, the bookie odds are not a very good indicator of actual results

The odds for draw are the worst

Moral of the story : Dont rely on betting odds if you aim to get rich through betting !!