

Soccer data base analysis

In this project I analyzed data of soccer data base from

(<https://www.google.com/url?q=https://www.kaggle.com/hugomathien/soccer&sa=D&ust=1532469042123000>)

This data contains 8 tables :

table	shape	dtypes
Player_attributes	(183978, 42)	float64(35), int64(3), object(4)
player	11060 rows x 7 columns	float64(1), int64(4), object(2)
Match	25979 rows x 115 columns	float64(96), int64(9), object(10)
league	11 rows x 3columns	int64(2), object(1)
country	(11, 2)	int64(1), object(1)
team	299 rows x 5 columns	float64(1), int64(2), object(2)
Team_Attributes	1458 rows x 25 columns	float64(1), int64(11), object(13)

After wrangling those tables we will ask the following questions :

Questions

What teams improved the most over the time period?

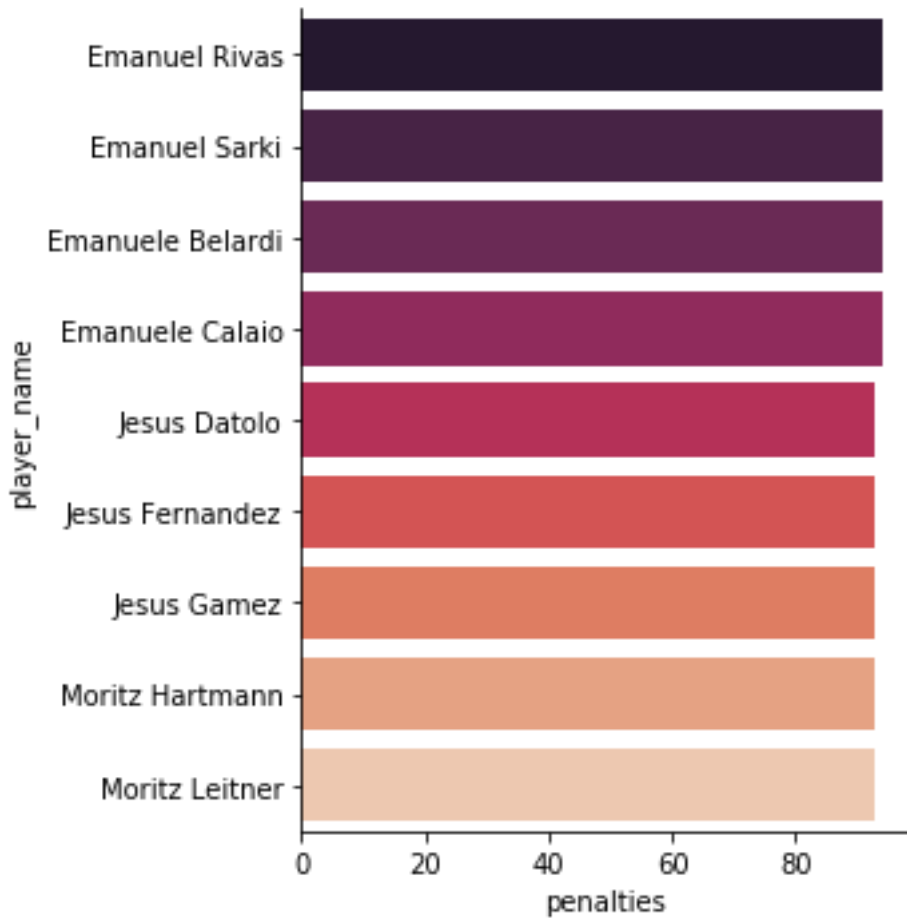
Which players had the most penalties?

What team attributes lead to the most victories?

what is the impact of playing in home land or remotely in different seasons ?

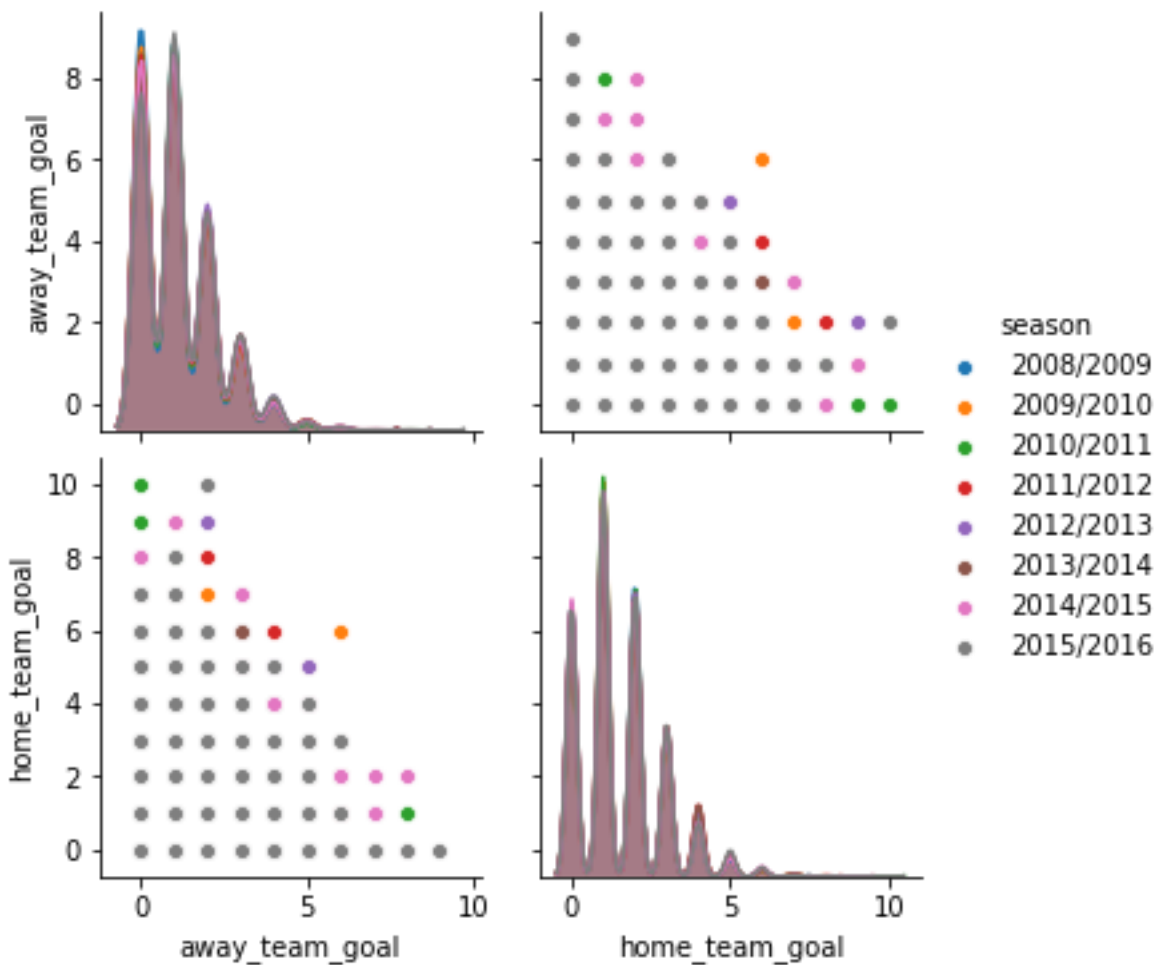
after asking these questions we get out with their answers which can be simply represented in visualizations

which player has max penalty ?



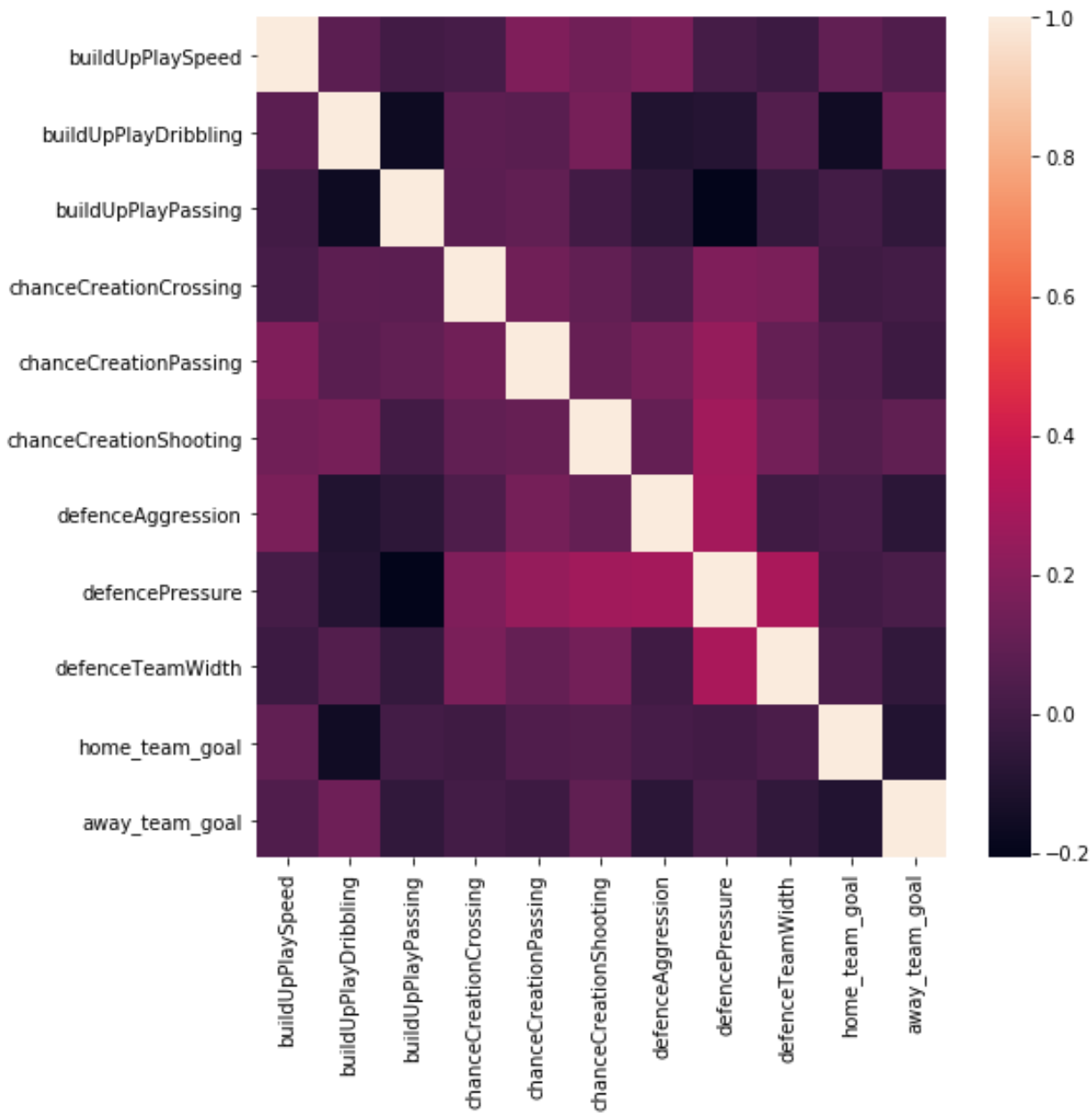
from this graph we notice that Emanuel Rivas , Emanuel Sarki , Emanuele Belardie and Emanuele Calaio had the highest penalties

what is the impact of playing in home land or remotly in different seasons ?



from this graph we notice that away team goals in 2015/2016 was extremely higher than home team goal and the home team goal in 2010/2011 was to high than the away team goal

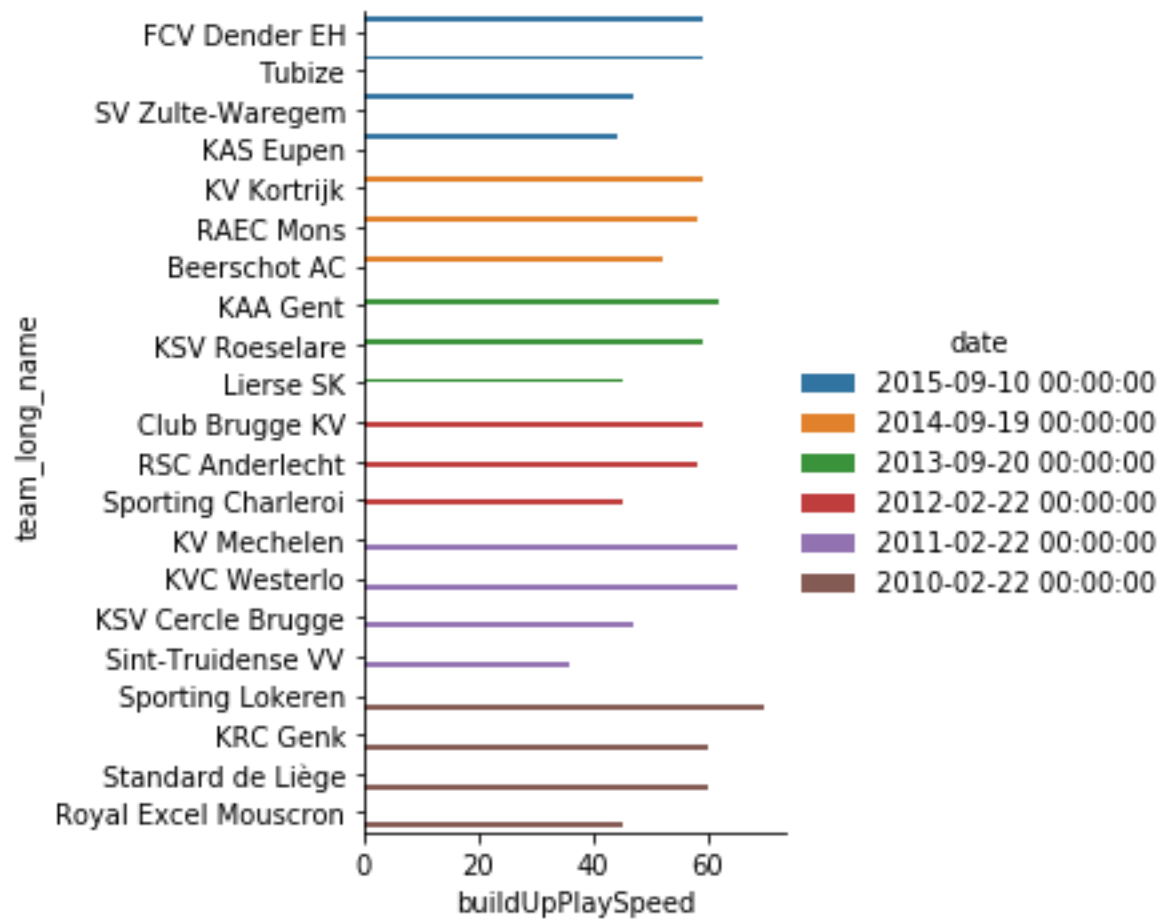
which team attribute lead to the most victory?



from heat map we find that buildUpPlayDribbling affects the most on home team goals

while defence aggression and buildUpPlayPassing affects the most on away team goals

which team improved with time?



we will notice from this graph that lately in 2015 FCV dender EH had the highest build up play speed
the highest build up play speed from 2015 to 2013 was from sporting lokeren team

In []: