Introduction

Given that we're in the middle of the FIFA world cup and that I enjoy watching football games (we don't use « soccer » in France for a game that involves a ball played with the feet, « football » seams more adapted), I choose the « soccer database » for this first project.

Football is a typical team game. Successful teams are not necessary the ones with the best individuals but often the teams that manage to play together based on their skill as a group of players. Therefore, I propose to study in the first part what are typical team attributes of successful and unsuccessful teams.

Then I would like to investigate how football leagues all over Europe compare. I often read that some leagues are more evenly disputed than others and I would like to see if the « soccer database » can bring elements to the discussion.

Finally, I often heard that playing home is an advantage for a soccer team. I would like to see that by myself with the « soccer database ». For that purpose I will plot some statistics of away vs home points, scored goals and conceded goals.

In order to carry out above described analysis, I build a table that summarizes team results by season. The table will contain following informations:

- Season
- League
- Team
- Team attributes (speed, dribbling, passing, positioning, crossing, shooting …)
- Rank
- Point (home, away, total)
- Scored goals (home, away, total)
- Conceded goals (home, away, total)

Then, I use the table and most appropriate data visualisation techniques in order to answer to following questions:

- What are team attributes of successful, average and unsuccessful teams?
- Are all soccer European championships as tight?
- Is it favorable to play at home?

Result by national championship and by season

Table results is first used to edit a document that provides result of football national championship by country and by season. The document is saved as « result_by_country_and_season.pdf »
An example of such table is provided below.

Result of football national championship in France
season 2015/2016

rank	name	win	lose	draw	goal +	goal -	difference	point
1	PSG	30	6	2	102	19	83	96
2	LYO	19	8	11	67	43	24	65
3	MON	17	14	7	57	50	7	65
4	NIC	18	9	11	58	41	17	63
5	LIL	15	15	8	39	27	12	60
6	ETI	17	7	14	42	37	5	58
7	CAE	16	6	16	39	52	-13	54
8	REN	13	13	12	52	54	-2	52
9	ANG	13	11	14	40	38	2	50
10	BAS	14	8	16	36	42	-6	50
11	BOR	12	14	12	50	57	-7	50
12	MON	14	7	17	49	47	2	49
13	MAR	10	18	10	48	42	6	48
14	NAN	12	12	14	33	44	-11	48
15	LOR	11	13	14	47	58	-11	46
16	GUI	11	11	16	47	56	-9	44
17	TOU	9	13	16	45	55	-10	40
18	REI	10	9	19	44	57	-13	39
19	GAJ	8	13	17	37	58	-21	37
20	TRO	3	9	26	28	83	-55	18

First question

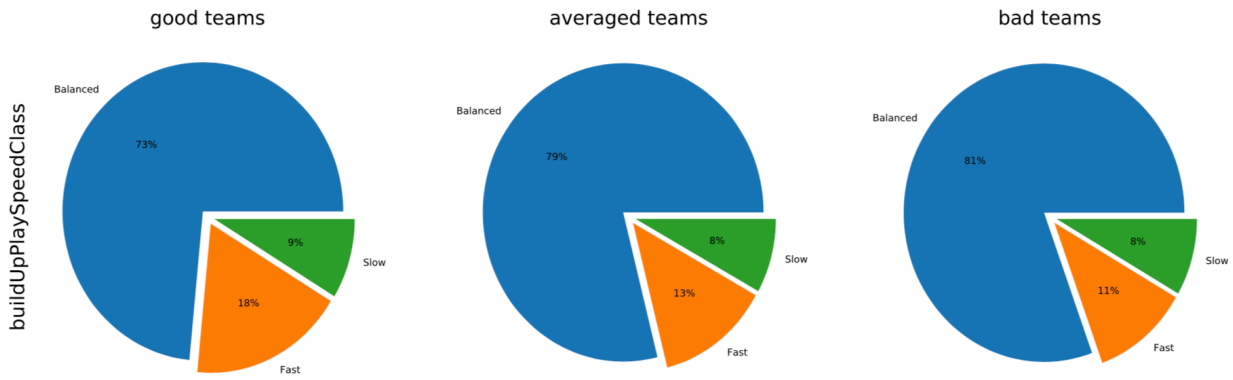
Table results is then used to answer the first question:
What are team attributes of successful, average and unsuccessful teams?

Successful teams (or good teams) are defined as the top 3 teams of each European championship for each year.
Unsuccessful teams (or bad teams) are defined as the bottom 3 teams of each European championship for each year.
Average teams takes all team into account.

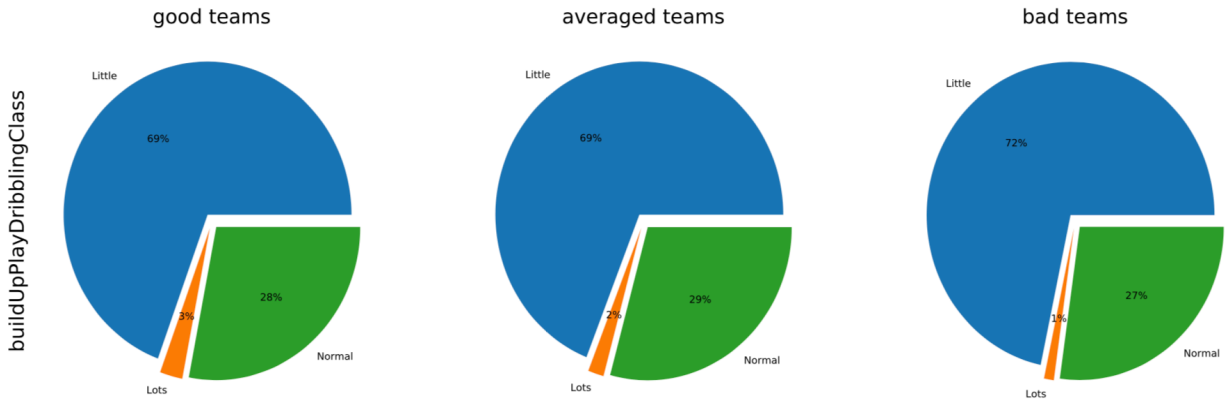
Pie plot visualisation is used to illustrate what proportion of each attribute is found in each population (successful, average and unsuccessful).
Pie plot results are saved in « team_attribute_analysis.pdf ».

Analysis of each attribute is the following:

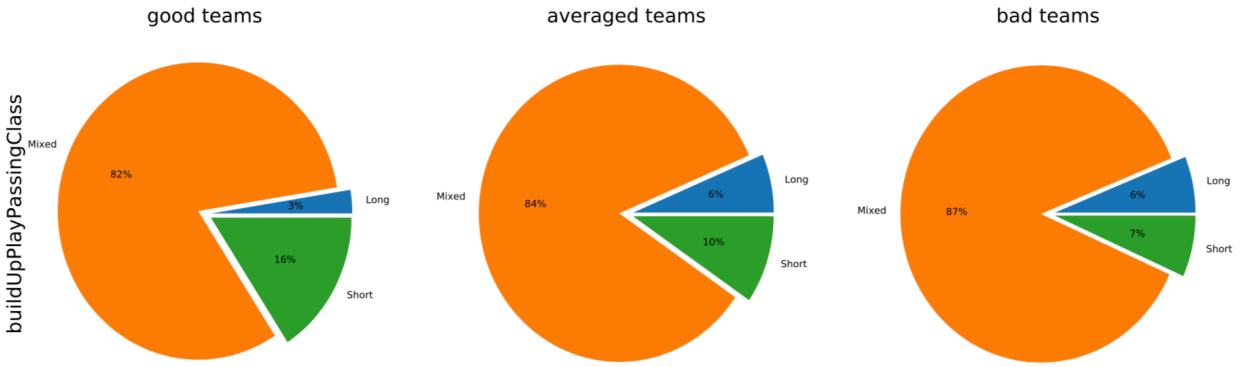
PlaySpeed attribute: good teams tend to play faster than bad teams



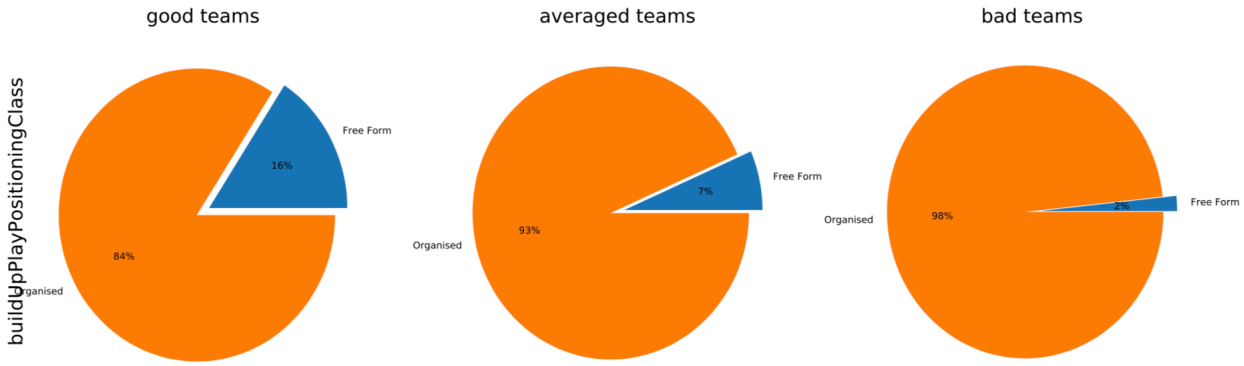
PlayDribbling attribute: no significant difference among the three populations



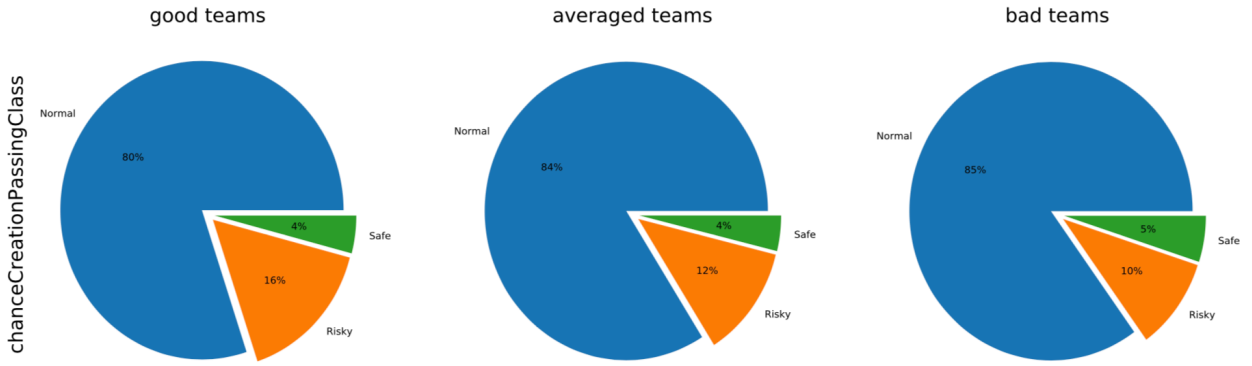
PlayPassing attribute: good teams tend to do more short pass and less long pass than bad teams



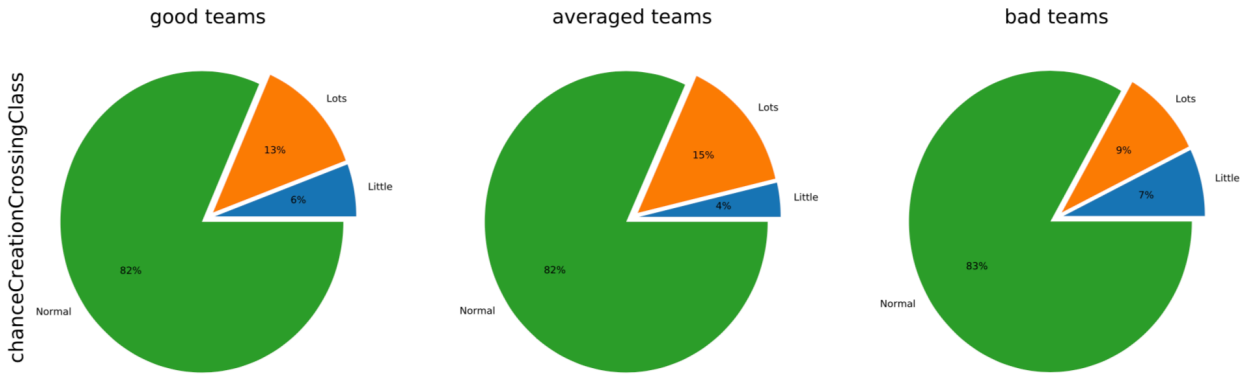
PlayPositionning attribute: good teams tend to be more creative than bad teams (more free form position on the field)



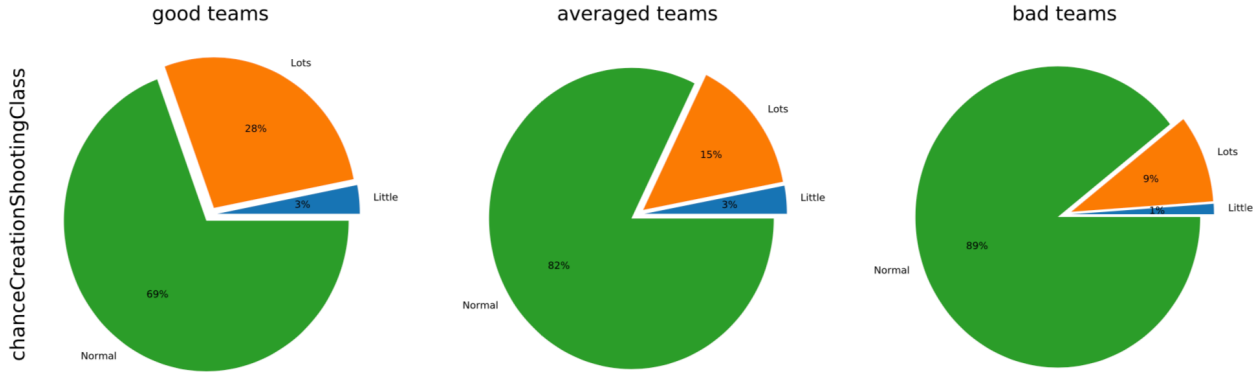
CreationPassing attribute: good teams tend to take more risk in the ball transmission (more risky pass and less safe pass)



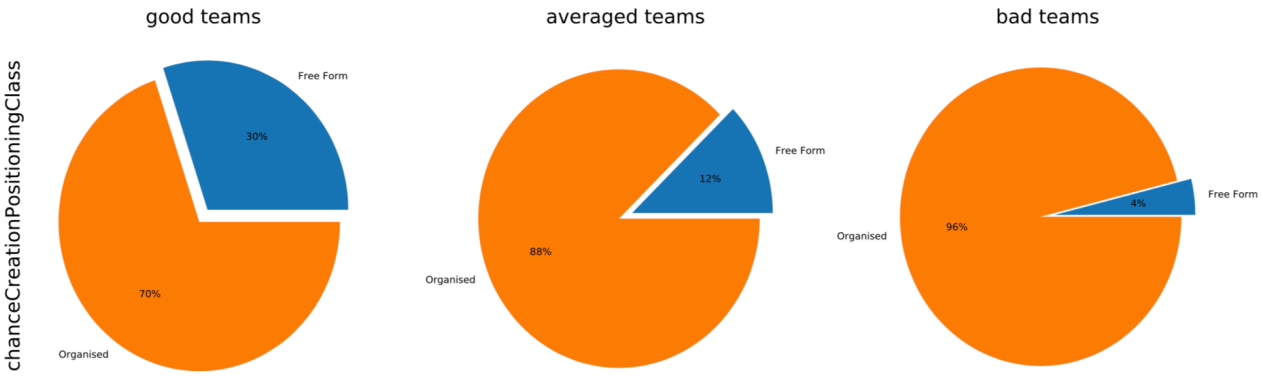
CreationCrossing attribute: no significant difference among the three populations



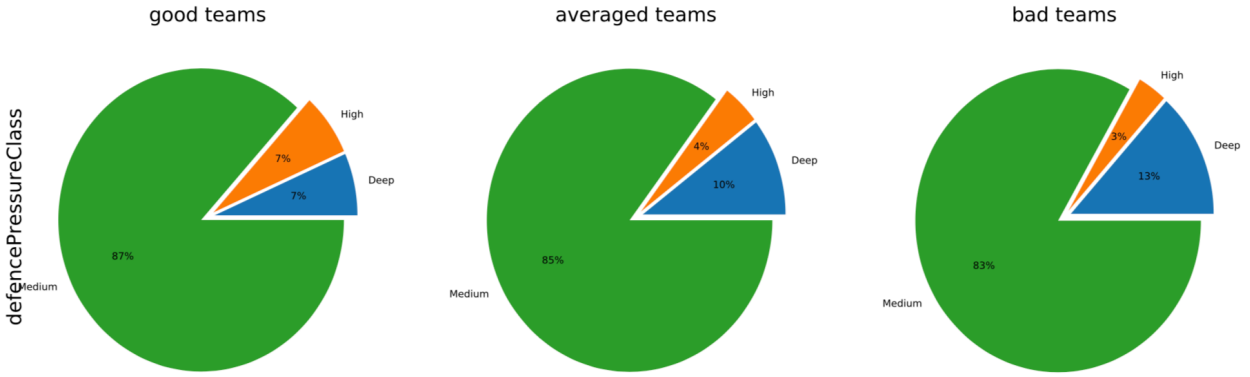
CreationShooting attribute: good teams tend to shoot a lot more on goals than bad teams



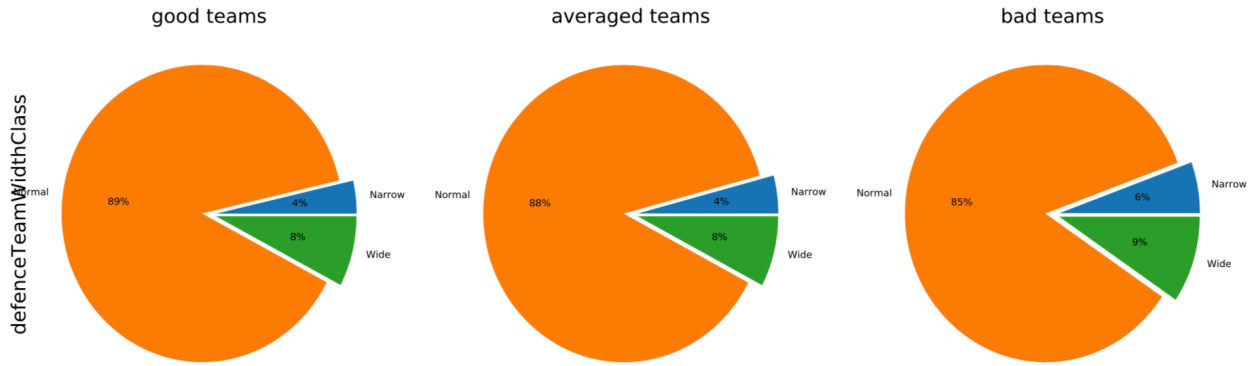
CreationPositioning attribute: again good teams tend to be more creative than bad teams (more free form position on the field)



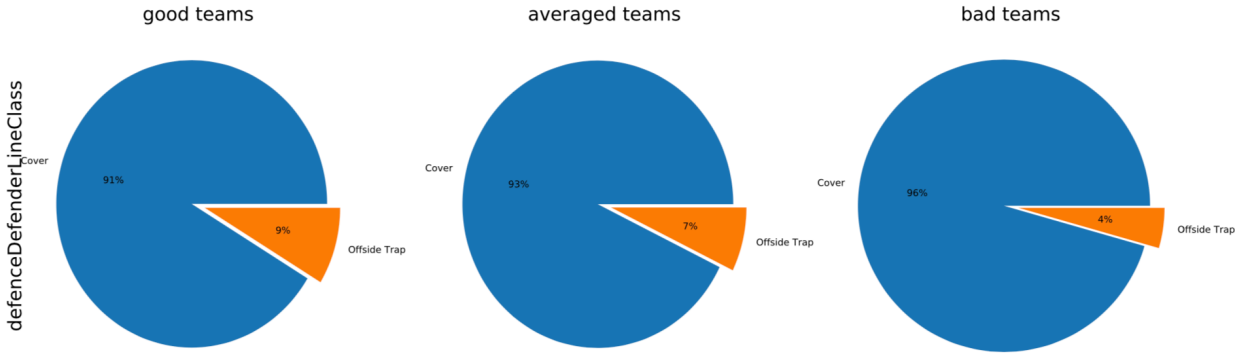
DefencePressure attribute: no significant difference among the three populations



DefenceTeamWidth attribute: no significant difference among the three populations



DefenderLine attribute: good team tend use more offside trap tactic



Second question

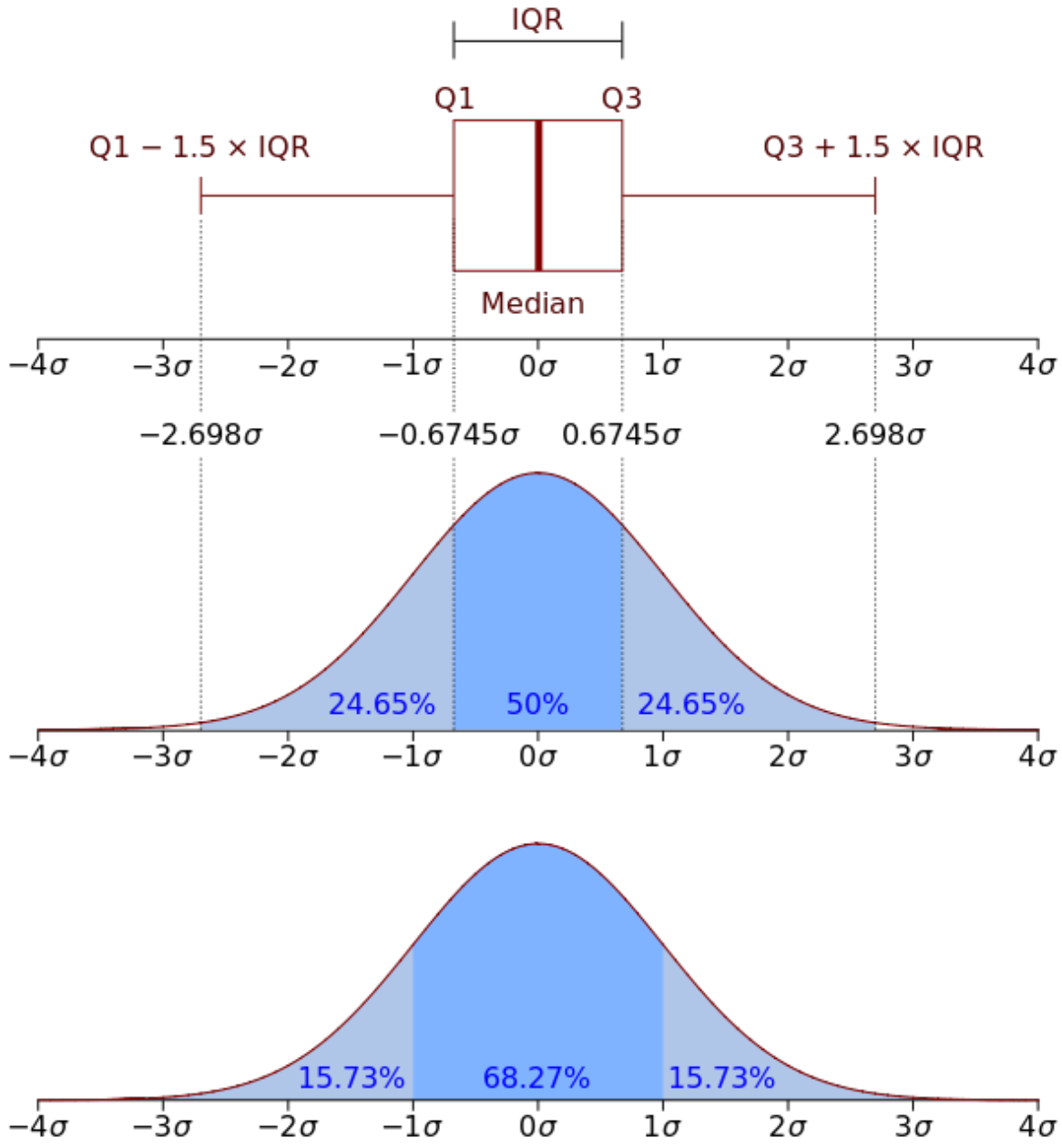
Table results is then used to answer the second question:
Are all soccer European championships as tight?

Box plot visualisation is used to illustrate how spread out are each championship results (by season).

For information box plot used in pandas has following definition:

- Box is the inter quartile range (IQR) and the band inside the box is always the second quartile (the median)
- The ends of the whiskers represent the lowest datum still within 1.5 IQR of the lower quartile, and the highest datum still within 1.5 IQR of the upper quartile (often called the Tukey boxplot)

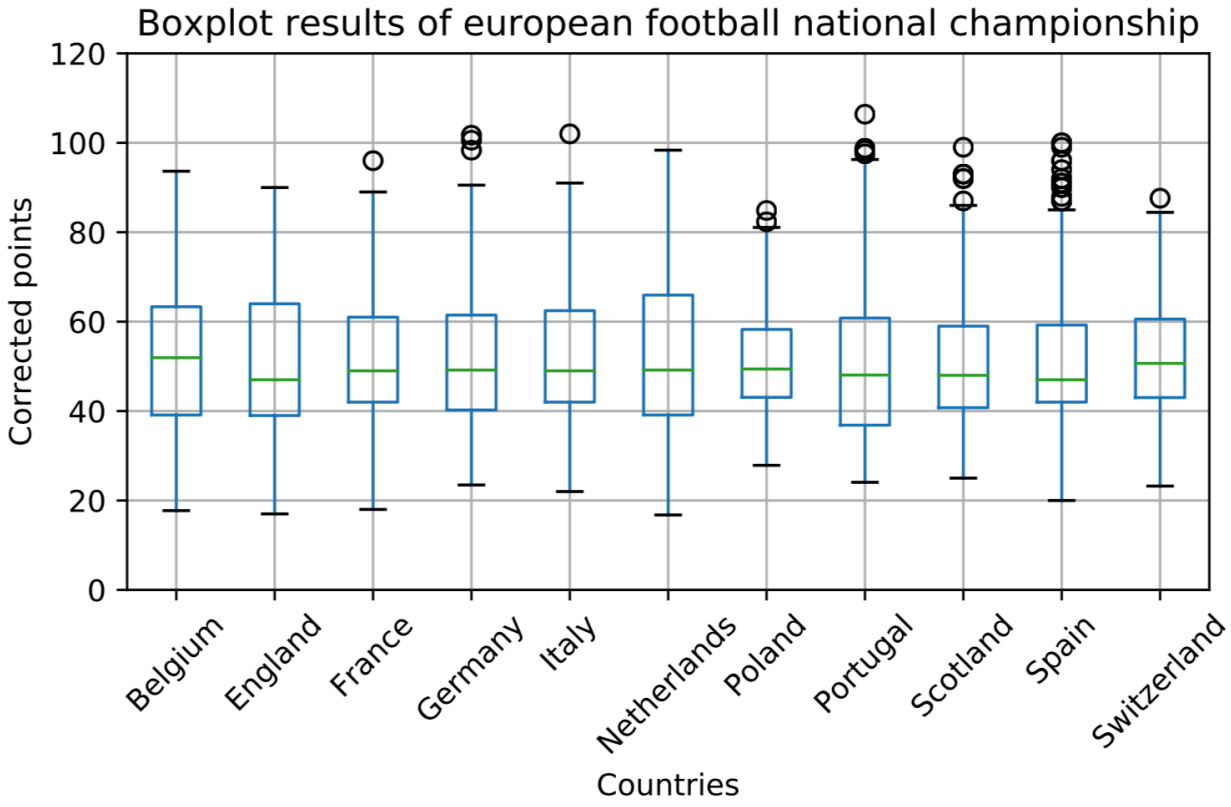
Figure below illustrates a boxplot and a probability density function (pdf) of a normal population.



Box plot is applied to the European championships results in the period 2008/2009 to 2015/2016, and saved in file « championship_analysis.pdf ». Same analysis is also carried out by years in file « championship_analysis_by_season.pdf ».

The box plot shows that the

- the most balanced championship in this period is Poland (lowest spread among the 11 countries)
- the most unbalanced championship in this period is Portugal (large spread and top teams tend to be largely above the other teams in the league)



Third question

Table results is finally used to answer the third question:
Is it favorable to play at home?

Histogram visualisation and some statistical calculation are used to illustrate European teams performance away compared to at home.

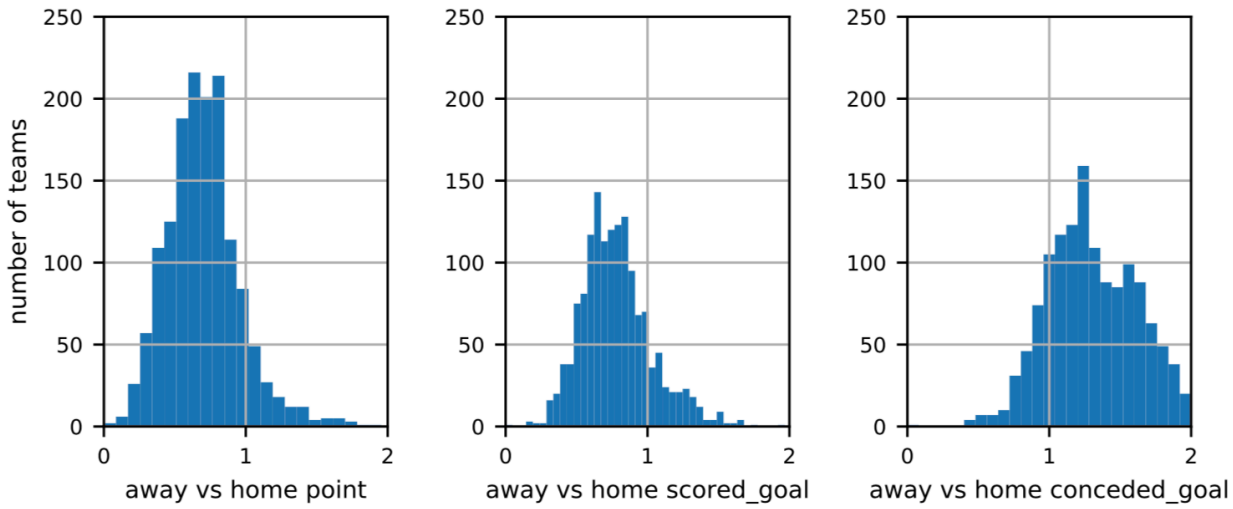
Three indicators are used:

- ratio of points scored away to points scored at home
- ratio of goals scored away to goals scored at home
- ratio of goals conceded away to goals conceded at home

Histogram plot and central statistics of these indicators are saved in file
'way_vs_home_analysis.pdf'

Based on European championships in the period 2008/2009 to 2015/2016, in average teams

- score 30% less points when playing away
- score 22% less goals when playing away
- concede 41% more goals when playing away



param	mean	median	std
away_vs_home_point	0.7	0.68	0.28
away_vs_home_scored_goal	0.78	0.76	0.25
away_vs_home_conceded_goal	1.41	1.32	0.47