



An Analysis of European Soccer Database Using R

Team 1
Assem, John, Joshua, Ern Chern



Do you like soccer?



Topic

Analysis of Soccer Database



 Dataset

European Soccer Database

25k+ matches, players & teams attributes for European Professional Football

 Hugo Mathien • updated 4 years ago (Version 9)

  3056

Research Questions:



PLAYERS

What are the most important features for successful **player**?



TEAM

What are the most important features for successful **team**?



MATCH

How to predict a probability of a **victory** in a match?

A blue-tinted photograph of two soccer players standing on a grassy field. A soccer ball is positioned on the ground between their feet. The player on the left is wearing dark shorts and dark socks with three white stripes. The player on the right is wearing light-colored shorts and light-colored socks with three dark stripes. The background is a blurred field.

Talent Wins Games, Teamwork Wins Championships

Q: Does **individual talent** directly translate to **team success**?

Dataset

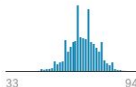
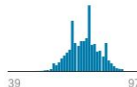

< Player_Attributes (184k rows)

Detail

Compact

Column

10 of 42 columns

# overall_rating	# potential	Δ preferred_foot	Δ attacking_work_r...	Δ defensive_work_r...	# crossing
		right 75% left 24% Other (836) 0%	medium 68% high 23% Other (16085) 9%	medium 71% high 15% Other (26091) 14%	

< Team_Attributes (1458 rows)

Detail

Compact

Column

10 of 25 columns

# buildUpPlaySpeed	Δ buildUpPlaySpee...	# buildUpPlayDribbl...	Δ buildUpPlayDribbl...	# buildUpPlayPassing	Δ buildUpPlayPassi...
	Balanced 81% Fast 12% Other (102) 7%		Little 69% Normal 30% Other (21) 1%		Mixed 85% Short 9% Other (94) 6%

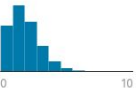
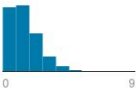
< Match (26.0k rows)

Detail

Compact

Column

6 of 115 columns

# home_team_goal	# away_team_goal	Δ goal	Δ card	Δ corner	Δ possession
		[null] 45% <goal /> 4% Other (13224) 51%	[null] 45% <card /> 2% Other (13776) 53%	[null] 45% <corner /> 22% Other (8464) 33%	[null] 45% <possession /> 22% Other (8419) 32%

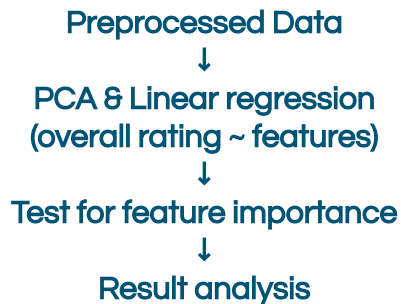
- Size: 298,59 MB
- +25,000 matches
- +10,000 players
- 11 European Countries with their lead championship
- Seasons 2008 to 2016
- Player and Team attributes

Methods



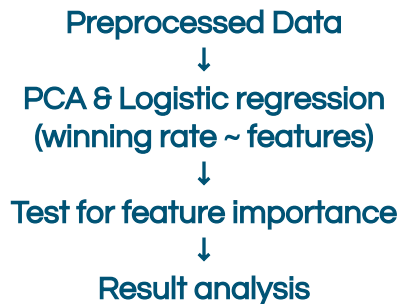
PLAYERS

What are the most important features for successful **player**?



TEAM

What are the most important features for successful **team**?



MATCH

How to predict a probability of a **victory** in a match?



Methods:

Talent Wins Games, Teamwork Wins Championships



Rank players



Spearman
correlation test



Rank teams

Schedule

	Week 10	Week 11	Week 12	Week 13	Week 14	Week 15
	Preprocessing	Featurization		Predictive Modeling		Presentation
John	Preprocess teams	LR/PCA on teams		LR to predict victory		Slides
Ern Chern	Preprocess players	LR/PCA on players		LR to predict victory		Graphs
Joshua	Preprocess match	Check Stats		Evaluate the model		Script
Assem	Create dataframe	Feature importance		Rank & Spearman		Speaker

Thank you

