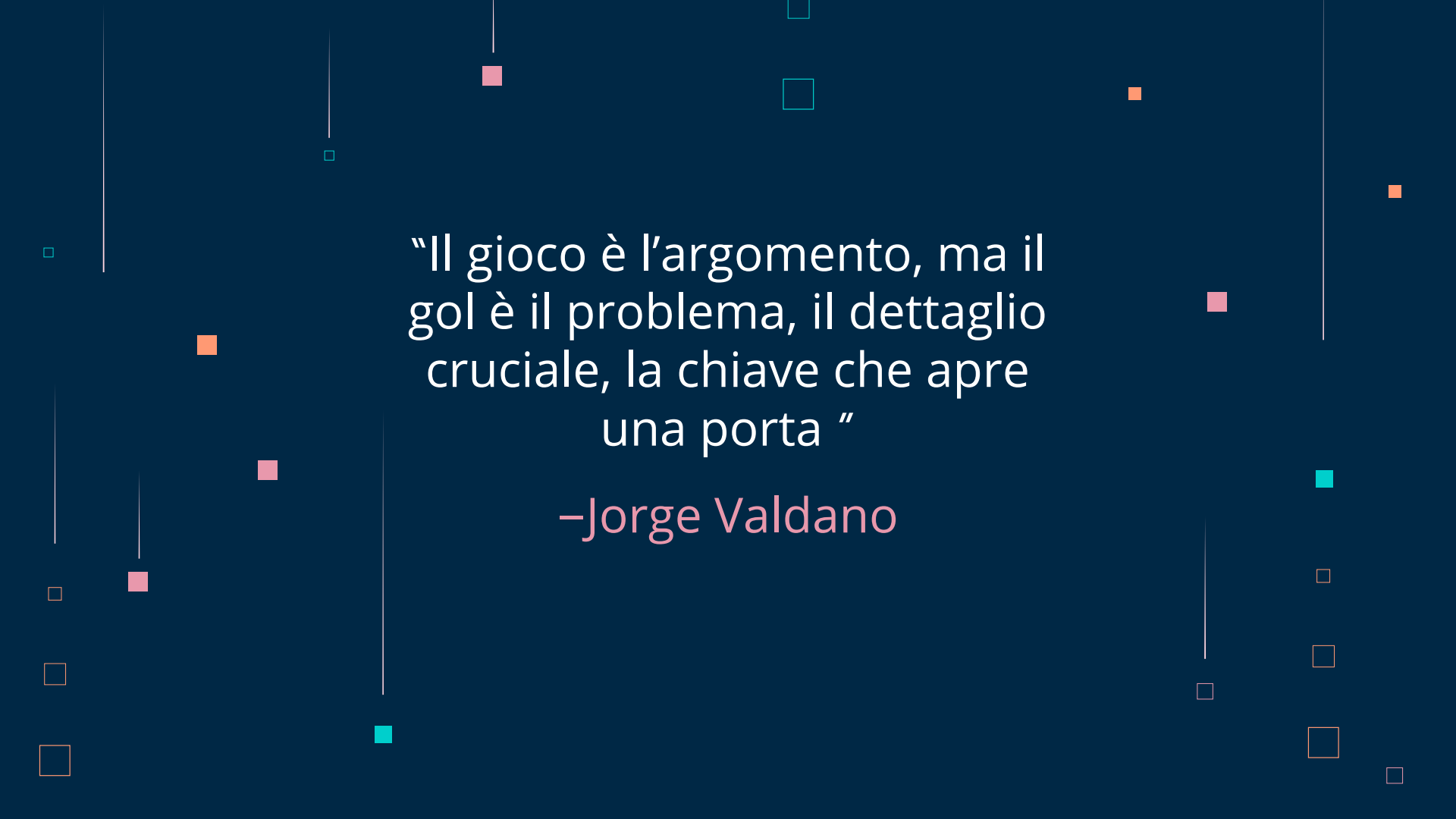


SoccerStats

La statistica del gol



The background is a dark blue field decorated with a sparse, abstract pattern of small squares and thin vertical lines. The squares are in various colors including light blue, orange, pink, and teal, and are scattered across the frame. Some thin white vertical lines extend from the top or bottom edges towards the center.

“Il gioco è l’argomento, ma il
gol è il problema, il dettaglio
cruciale, la chiave che apre
una porta ”

–Jorge Valdano

Contesto di riferimento

Setting: Premier League, Ligue 1, Bundesliga, LaLiga, Serie A

Timing: 2014-2020

Focus:
modello di previsione GG/NG,
ruolo fattore casa/fuori casa

Fonte: <https://www.kaggle.com/datasets/sanjeetsinghnaik/football-data-top-5-leagues>



IL PROCESSO DI ANALISI

Analisi esplorativa
e data preparation

STEP 1

STEP 2

Regressione
logistica

Cluster analysis

STEP 3

STEP 4

Principal
component
analysis



TAVOLA DEI CONTENUTI

Step 1: Analisi esplorativa e
data preparation

Step 2

Step 3

Step 4



PROCESSO

Analisi descrittiva delle
variabili
Preparazione data set
divisi per anno



OBIETTIVO

Comprendere i dati
a disposizione
Ottenere i data set
da utilizzare

PANORAMICA DESCRITTIVA VARABILI

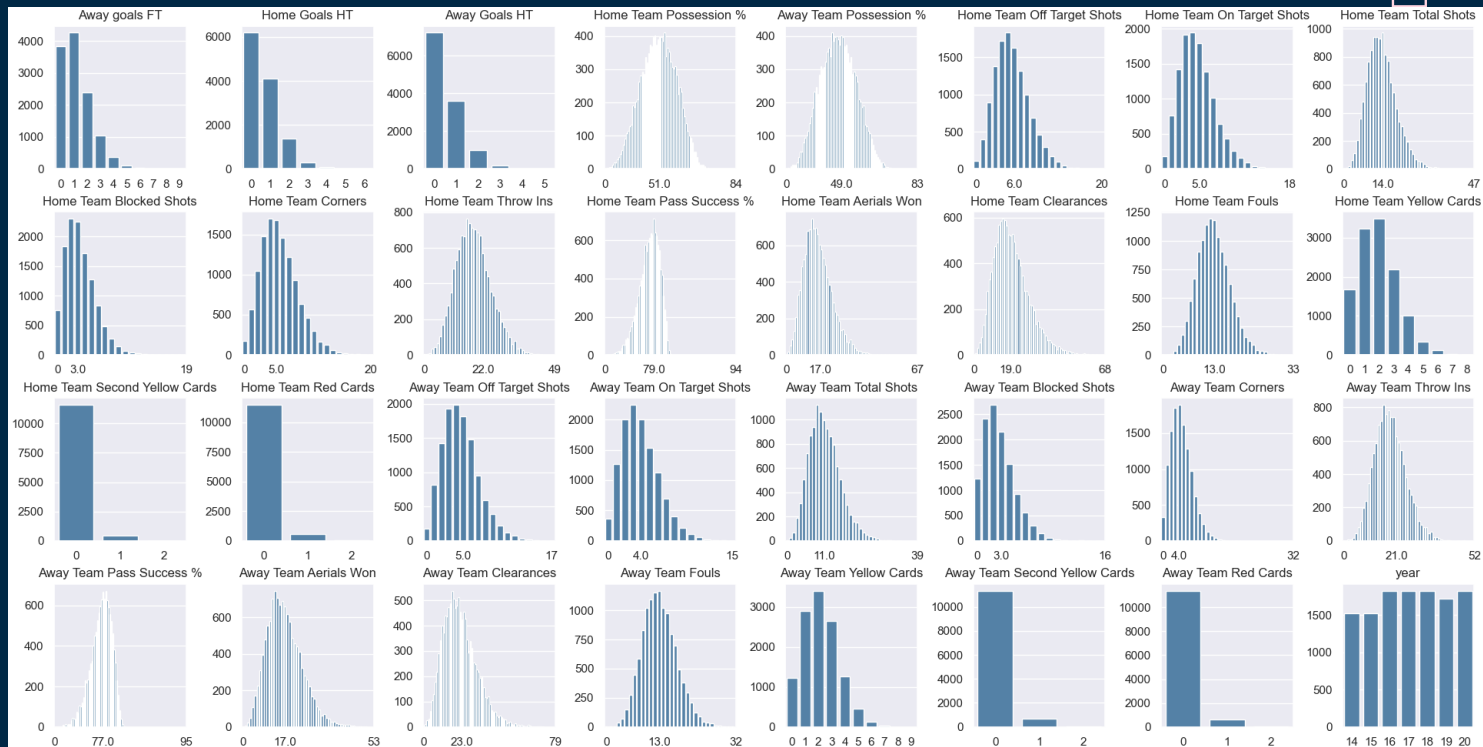


TAVOLA DEI CONTENUTI

Step 1

Step 2: Regressione logistica
training & testing

Step 3

Step 4



PROCESSO

Analisi mediante
regressione logistica
(link logit)
Testing e selezione del
threshold



OBIETTIVO

Trovare il giusto fit
tra significatività e
bontà di adattamento
Ottenere un buon
modello previsionale
per il GG/NG

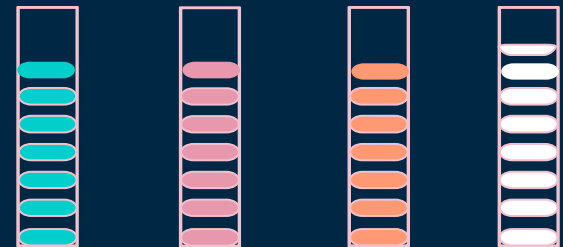
Regressione logistica training & testing

Variabile	Coefficienti	P-Value	Significatività
Intercept	-2.957862	1.24e-12	☆☆☆
Home Team Goal HT	0.716503	< 2e-16	☆☆☆
Away Team Goal HT	1.024221	< 2e-16	☆☆☆
Home Team on target shots	0.133108	< 2e-16	☆☆☆
Home Team Blocked Shots	0.021176	0.07352	●
Home Team Corners	-0.020621	0.03959	☆
Home Team Throw Ins	0.007675	0.06183	●
Home Team Pass Success	-0.006453	0.12389	●
Home Team Aerials Won	0.006996	0.09062	●
Home Team Yellow Cards	0.081078	1.17e-05	☆☆☆
Home Team Second Yellow Cards	0.140932	0.24210	●
Away Team On Target Shots	0.208522	< 2e-16	☆☆☆
Away Team Blocked Shots	0.008772	0.49189	●
Away Team Corners	-0.015008	0.16594	●
Away Team Aerials Won	0.005076	0.23560	●
Away Team Clearances	0.006830	0.02269	☆
Away Team Fouls	0.025253	3.45e-05	☆☆☆
Away Team Yellow Cards	0.095398	5.44e-07	☆☆☆
Away Team Red Cards	0.258435	0.00813	☆☆

Min 1 qrt Median 3 qrt Max

-3.37 -0.93 0.31 0.95 2.14

Residui



70.5% Accuracy 70.2% Specificity 70.8% Sensitivity 0.77 Area Roc

Predizione

VIF < 2

Test LR (p-value) 0

R-quadro di McFadden 0.181345

Threshold 0.4956

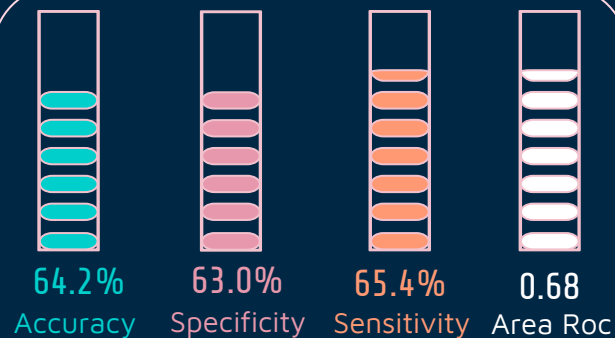
Test

Regressione logistica training & testing - Home

Variabile	Coefficienti	P-Value	Significatività
Intercept	0.608961	0.10956	
Home Team Goal HT	0.617959	< 2e-16	☆☆☆
Home Team Possession	-0.004038	0.18253	
Home Team on target shots	0.085245	< 2e-16	☆☆☆
Home Team Blocked Shots	0.033947	0.00172	☆☆
Home Team Corners	-0.023586	0.00630	☆☆
Home Team Throw Ins	0.006060	0.11363	
Home Team Pass Success	-0.014723	0.00227	☆☆
Home Team Clearance	-0.014736	4.00e-08	☆☆☆
Home Team Foul	-0.005116	0.35608	
Home Team Red Card	0.179570	0.05339	●
Home Team Yellow Cards	0.469020	1.88e-13	☆☆☆
Home Team Second Yellow Cards	0.140932	3.03e-05	☆☆☆

Min	1 qrt	Median	3 qrt	Max
-2.57	-1.08	0.66	1.10	1.65

Residui



Predizione

VIF	< 2.73
Test LR (p-value)	0
R-quadro di McFadden	0.060508
Threshold	0.4956

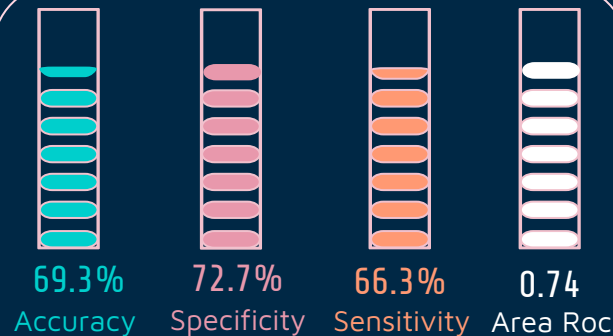
Test

Regressione logistica training & testing - Away

Variabile	Coefficienti	P-Value	Significatività
Intercept	-0.327021	0.440701	
Away Team Goal HT	0.975631	< 2e-16	☆☆☆
Away Team Possession	-0.007052	0.040387	☆
Away Team Off Target Shots	0.017011	0.084711	●
Away Team On Target Shots	0.163160	< 2e-16	☆☆☆
Away Team Blocked Shots	0.008217	0.502940	
Away Team Corners	-0.009785	0.314695	
Away Team Throw Ins	0.003404	0.399335	
Away Team Pass Success	-0.005811	0.280691	
Away Team Aerials Won	-0.003795	0.329080	
Away Team Clearances	-0.010259	6.07e-05	☆☆☆
Away Team Fouls	0.003988	0.485536	
Away Team Yellow Cards	0.106655	2.65e-09	☆☆☆
Away Team Second Yellow Cards	0.131170	0.151711	
Away Team Red Cards	0.297323	0.000975	☆☆☆

Min	1 qrt	Median	3 qrt	Max
-3.37	-0.99	0.46	1.02	1.73

Residui



Predizione

VIF	< 3.45
Test LR (p-value)	0
R-quadro di McFadden	0.110360
Threshold	0.4956

Test

TAVOLA DEI CONTENUTI

Step 1

Step 3

Step 3: Cluster analysis

Step 4



PROCESSO

Preparazione dataset
per anno & squadra &
home/away
Analisi silhouette
Cluster gerarchico e non

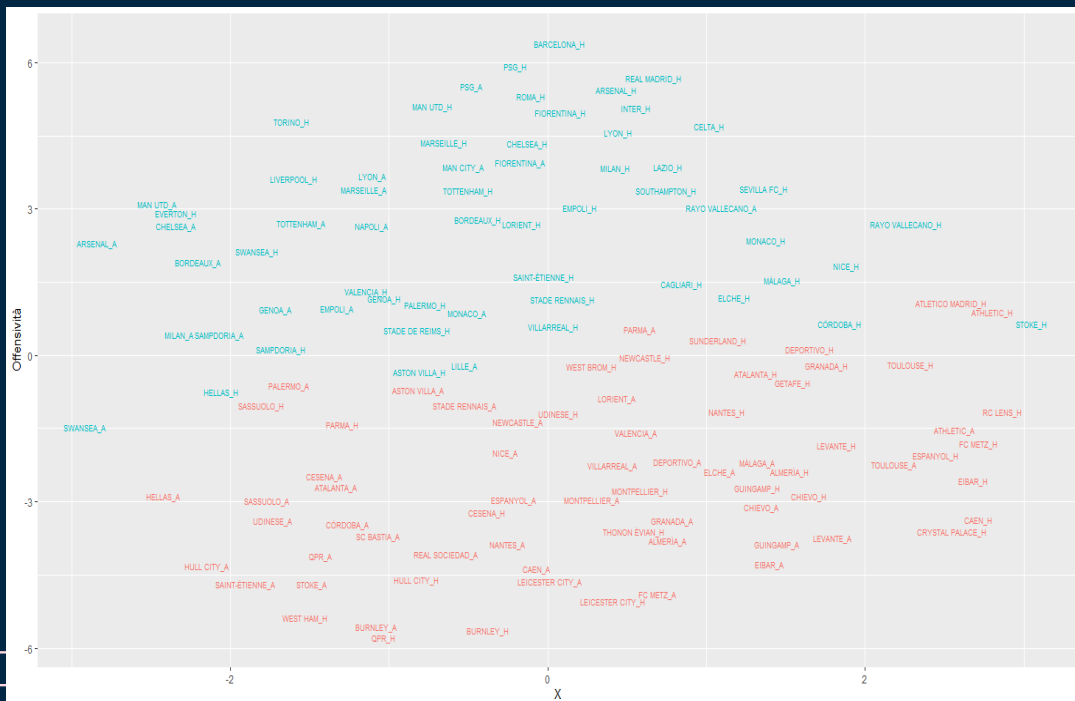


OBIETTIVO

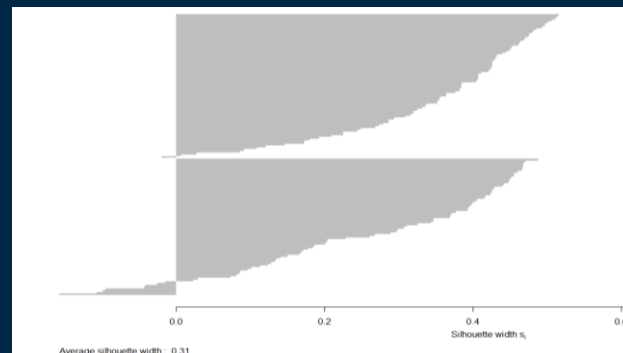
Ricerca del numero
ottimo di gruppi
Identificazione dei
cluster
Dendrogramma

Cluster non gerarchico 2014

Rappresentazione cluster mediante t-sne



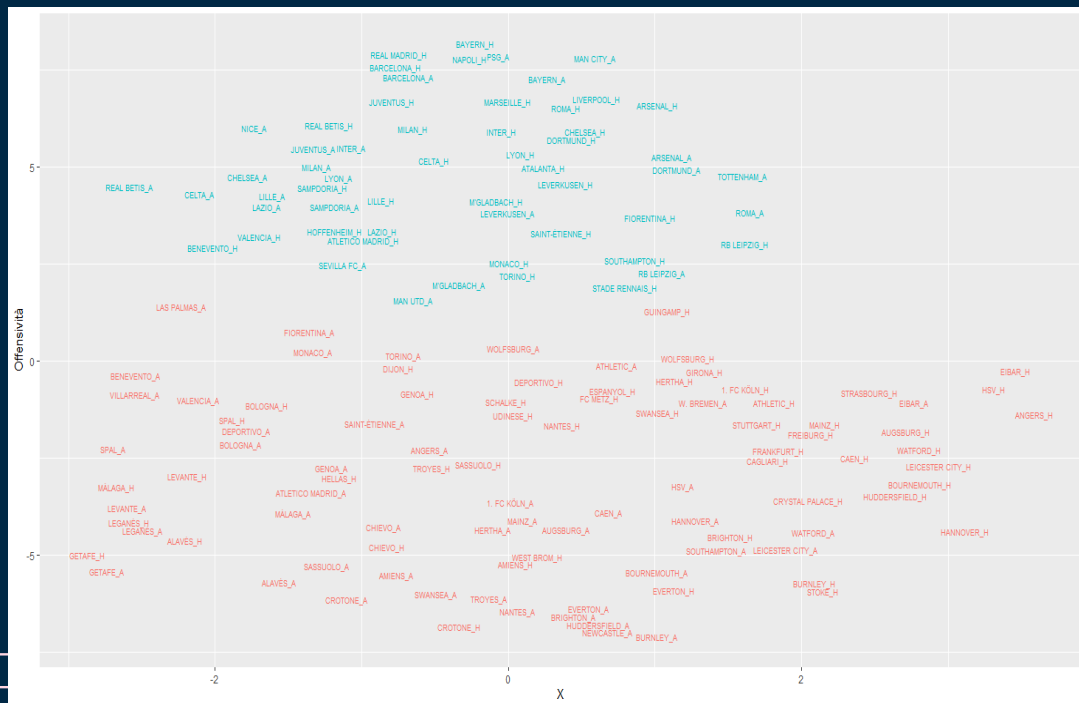
Silhouette e correlazioni



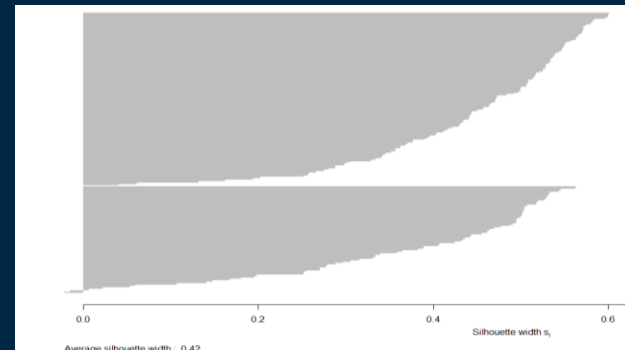
Variabile	X	Variabile	Offensivita
Throw Ins	0.61406	Possession	0.90959
Clearances	-0.54810	On target sh	0.75326
Aereal Won	0.35483	Goals FT	0.68980
Pass Success	-0.35273	Clearances	-0.66584

Cluster non gerarchico 2017

Rappresentazione cluster mediante t-sne



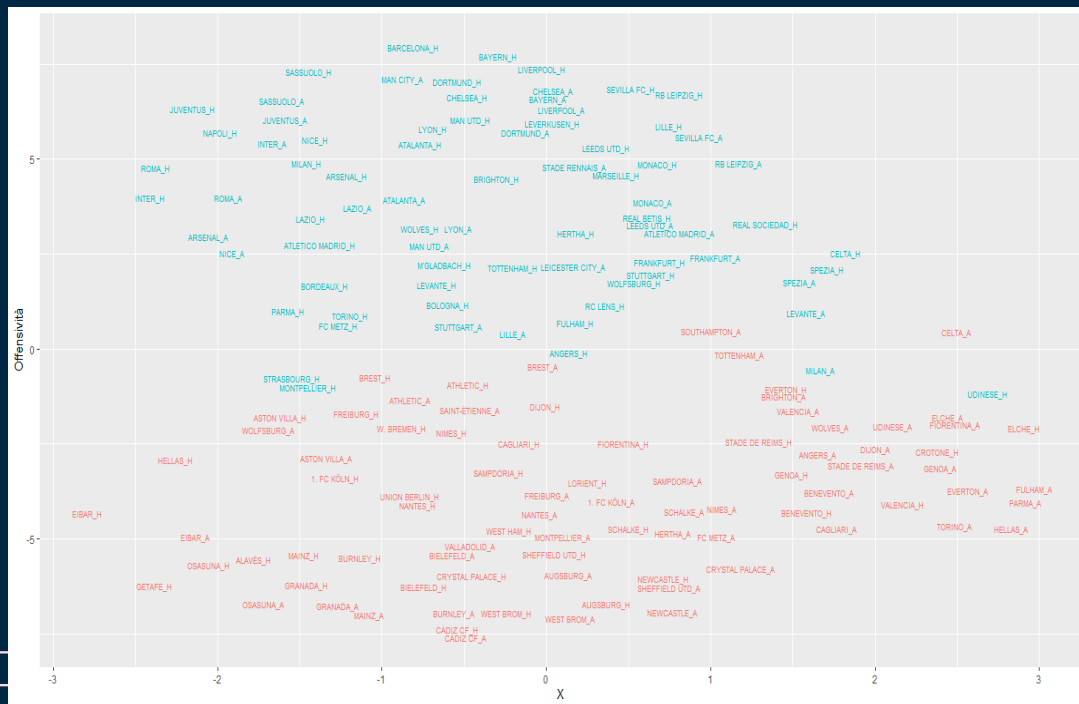
Silhouette e correlazioni



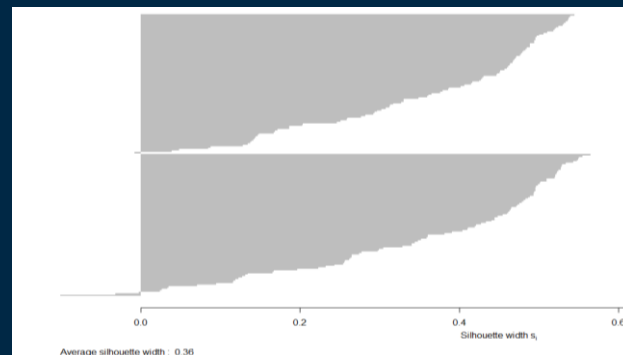
Variabile	X	Variabile	Offensività
Aerial Won	0.68648	Possession	0.93294
Throw Ins	0.57719	Clearances	-0.78680
Yellow Cards	-0.47310	On Target Sh	0.77631
Corners	0.25532	Goals FT	0.72269

Cluster non gerarchico 2020

Rappresentazione cluster mediante t-sne



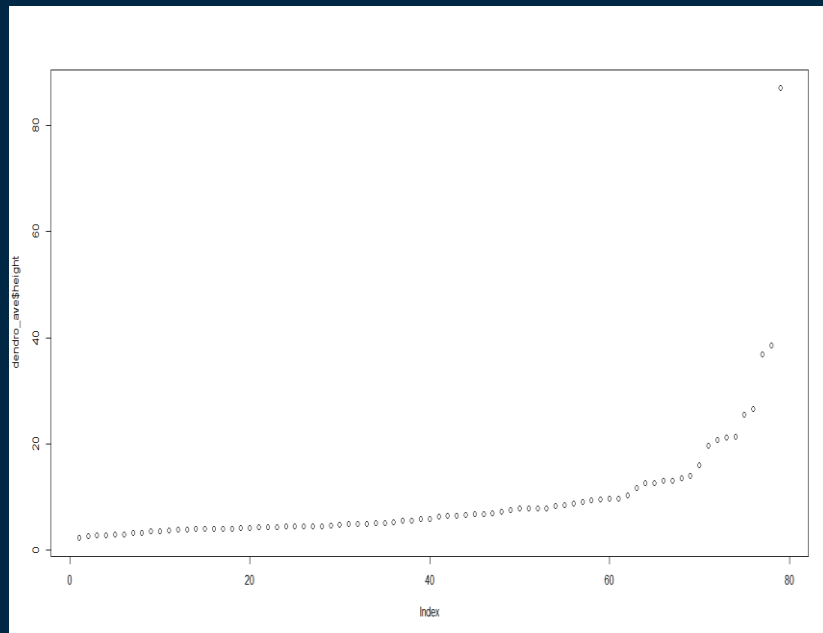
Silhouette e correlazioni



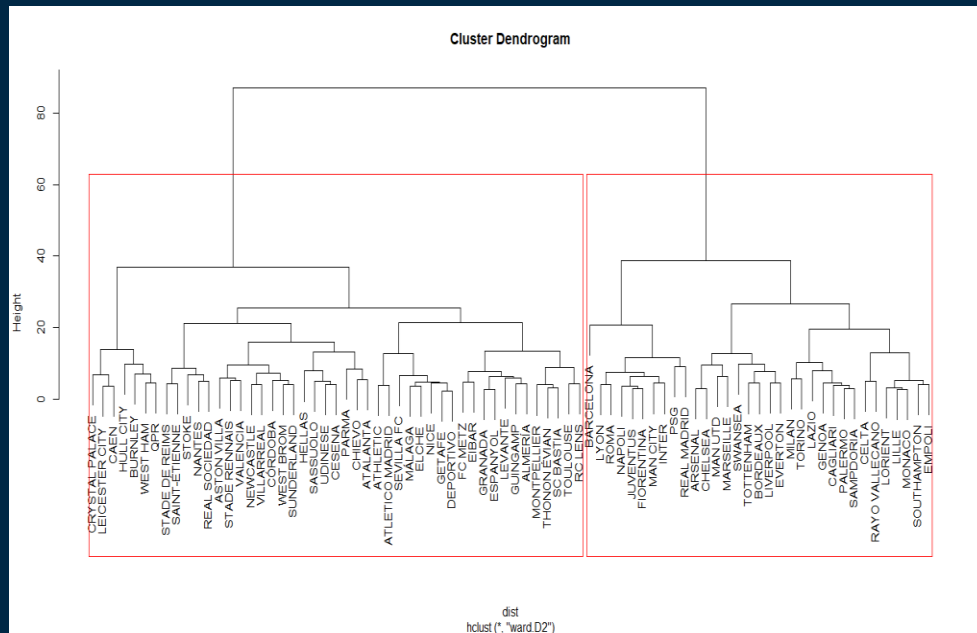
Variabile	X	Variabile	Offensività
Off Target Sh	-0.40932	Possession	0.95596
Blocked Sh	-0.35419	Clearances	-0.75680
On Target Sh	-0.34868	On Target Sh	0.73844
Clearances	0.33212	Corners	0.70210

Cluster gerarchico 2014

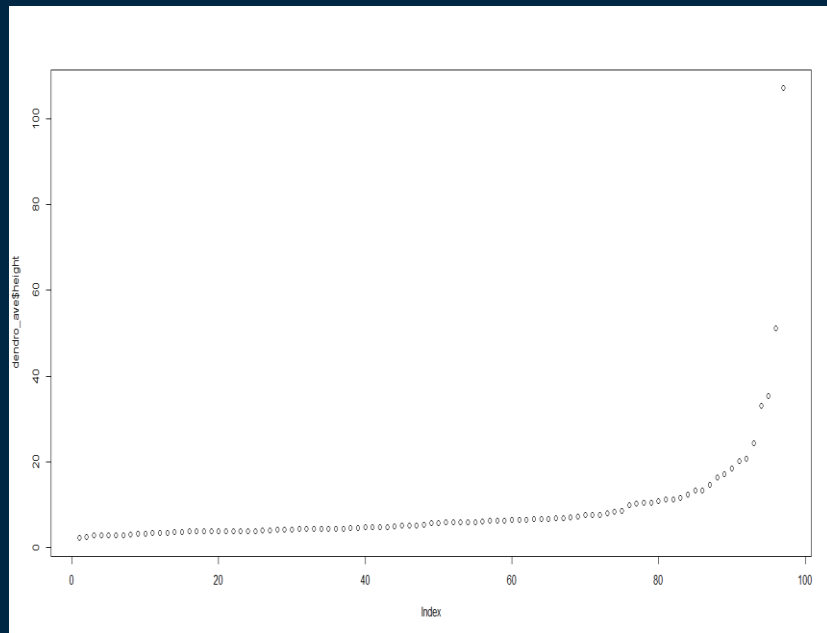
Elbow method



Dendrogramma



Elbow method

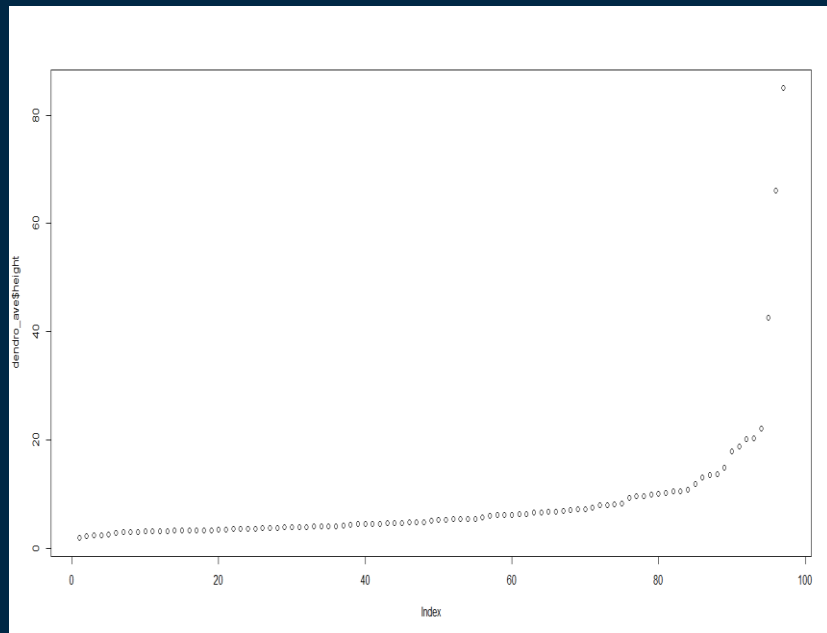


Dendrogramma



Cluster gerarchico 2020

Elbow method



Dendrogramma

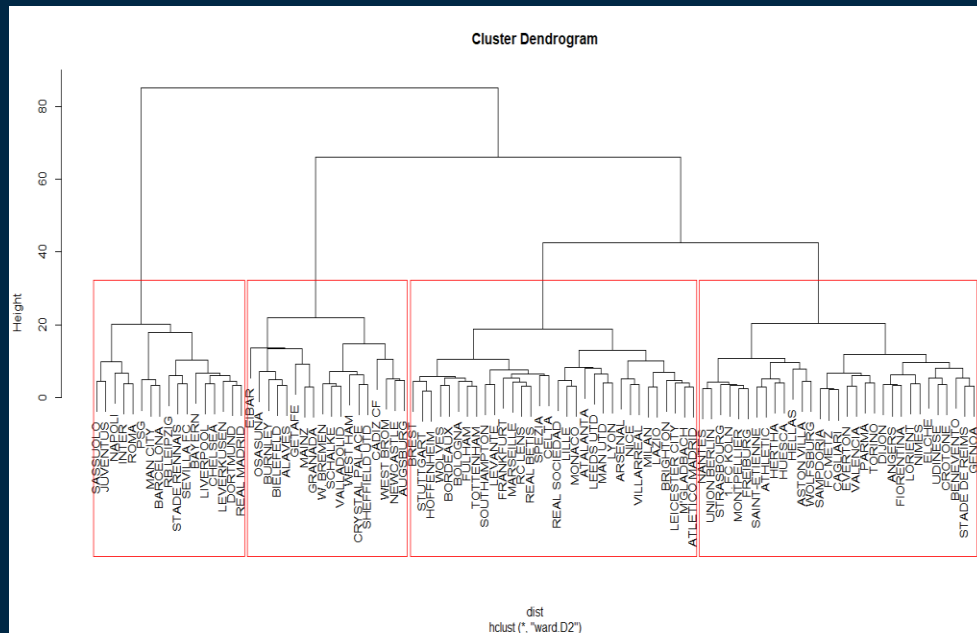


TAVOLA DEI CONTENUTI

Step 1

Step 3

Step 4

Step 4: Principal component analysis



PROCESSO

Individuazione ed interpretazione delle dimensioni della PCA
Raccolta statistiche per anno & squadra & home/away



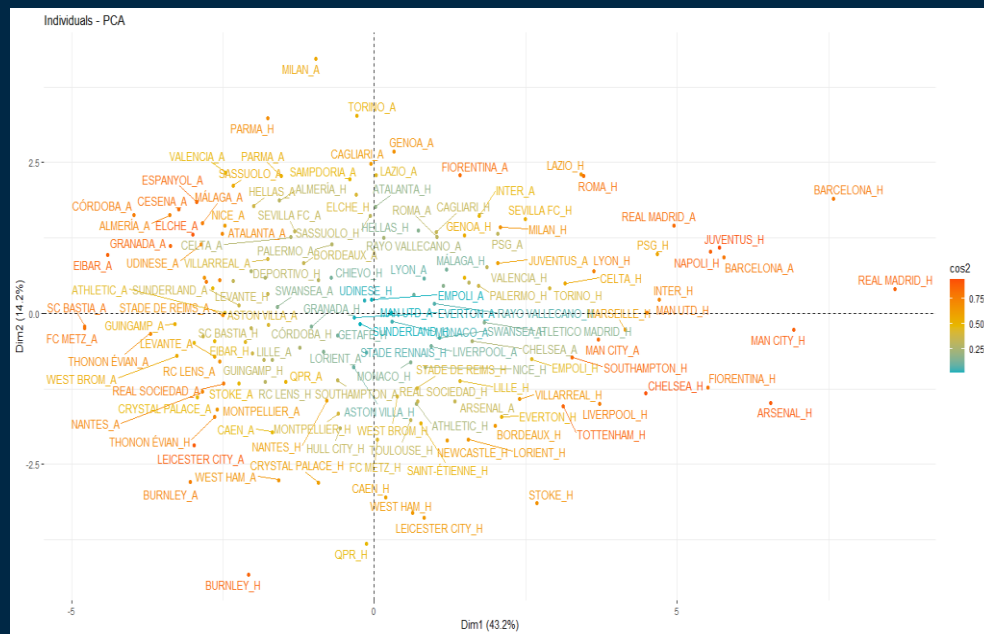
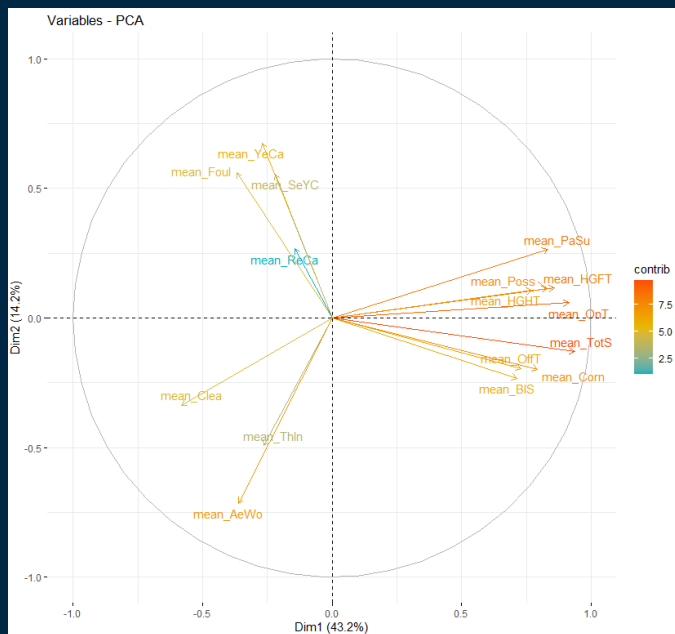
OBIETTIVO

Creazione grafici
Valutazione ruolo casa/fuori casa
Spiegazione dei trend di cambiamento nell'attitudine di gioco

Principal Component Analysis 2014

Direzione variabili

PCA



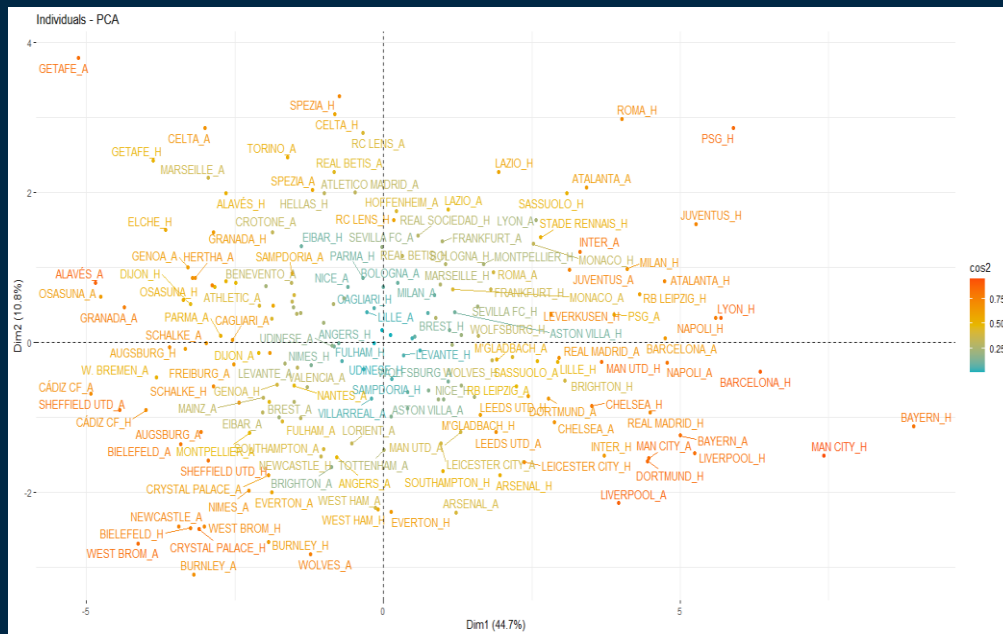
Direzione variabili

PCA

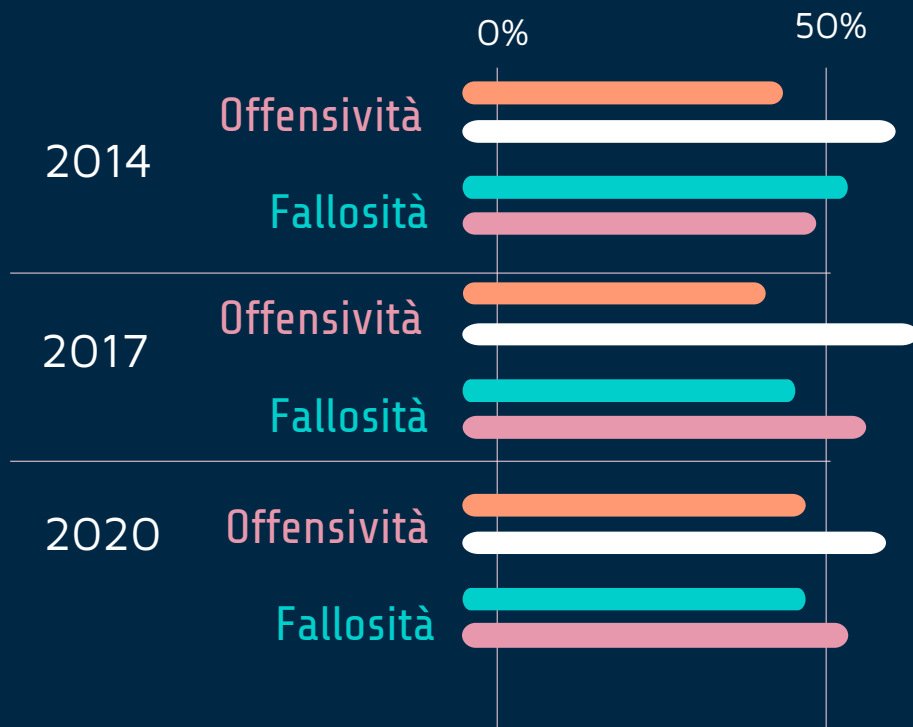


Direzione variabili

PCA



Trend



Anno	Offensive	Difensive	Fallose	Non fallose
2014	45.6%	54.4%	50.6%	49.4%
2015	44.4%	55.6%	50.6%	49.4%
2016	41.3%	58.7%	46.9%	53.1%
2017	43.9%	56.1%	48.0%	52.0%
2018	42.9%	57.1%	49.0%	51.0%
2019	39.8%	60.2%	44.9%	55.1%
2020	45.9%	54.1%	48.5%	51.5%

Do you have any questions?

jacobobonanno@outlook.it
S1110402@studenti.univpm.it

THANKS



Andrea Baldinini
Jacopo Bonanno