

# Selección de variables mediante PCA

## 1. Introducción

PCA (Principal Component Analysis) es una herramienta estadística utilizada para reducir la dimensionalidad de un conjunto de variables. En esencia, utilizando PCA sobre un dataset con un elevado número de variables explicativas podemos hacer que el conjunto de variables explicativas quede representado por un número más pequeño de variables, siendo estas nuevas variables combinaciones lineales de las originales.

La obtención de estas nuevas variables no deja de ser un procedimiento matemático conocido basado en la obtención de los valores y vectores propios de la matriz de correlación entre las variables originales. El problema de fondo viene en cómo definir o explicar estas nuevas variables de forma que se subrayen cuáles son las variables originales que más influyen en estas nuevas, es un proceso que no es tan fácil como en un principio pudiera parecer.

El siguiente análisis utiliza un dataset obtenido por medio de webscrapping de FBref.com para la temporada 2020-21 y los datos de las 5 grandes ligas europeas, esto es, 97 equipos y 2710 jugadores. Para el análisis se han eliminado los porteros y aquellos jugadores con menos de 900 minutos de tiempo jugado, quedando al final un dataset de tamaño 1346 jugadores y 151 parámetros o variables explicativas.

Asimismo, me he enfocado en analizar los parámetros más relacionados con la posición de mediocampista/mediapunta siguiendo lo indicado en el siguiente artículo:

<https://careyanalytics.wordpress.com/2020/04/30/a-closer-look-into-european-midfielder-playing-styles/>

**Table 1: 25 event metrics entered into the Principal Components Analysis**

Short Passes Attempted	Medium Passes Attempted	Long Passes Attempted
Passes Received	Touches	Passes into Final Third
Tackles in Defensive Third	Tackles in Mid Third	Tackles in Attacking Third
Key Passes	Carries	Interceptions
Progressive Distance	Progressive Passes	Shots
Pressures in Defensive Third	Pressures in Mid Third	Pressures in Attacking Third
Dribbles Attempted	Through Balls	Crosses
Clearances	Blocks	Switches
Passes into Penalty Area		

*NB. Defensive statistics are possession adjusted, and all data adjusted per 90 mins and z-scored. See [FBref website](#) for glossary of all event metric definitions*

['player', 'position', 'team\_name', 'passes\_short\_Per90', 'passes\_medium\_Per90', 'passes\_long\_Per90', 'passes\_received\_Per90', 'touches\_Per90', 'passes\_into\_final\_third\_Per90', 'tackles\_def\_3rd\_PAdj\_Per90', 'tackles\_mid\_3rd\_PAdj\_Per90', 'tackles\_att\_3rd\_PAdj\_Per90', 'assisted\_shots\_Per90', 'carries\_Per90', 'interceptions\_PAdj\_Per90', 'passes\_progressive\_distance\_Per90', 'carry\_progressive\_distance\_Per90', 'progressive\_passes\_Per90', 'shots\_total\_Per90', 'pressures\_def\_3rd\_PAdj\_Per90', 'pressures\_mid\_3rd\_PAdj\_Per90', 'pressures\_att\_3rd\_PAdj\_Per90', 'dribbles\_Per90', 'through\_balls\_Per90', 'crosses\_Per90', 'clearances\_PAdj\_Per90', 'blocks\_PAdj\_Per90', 'passes\_switches\_Per90', 'passes\_into\_penalty\_area\_Per90']

Ad\*\* Parameter adjusted by Average Team Possession

Per\_90. Parameter adjusted by Minutes played (games played)

## **2. Obteniendo e interpretando los valores de PCA**

El análisis de PCA identificó 5 factores o patrones de estilo de juego a lo largo de las 5 ligas europeas y las 25 métricas mencionadas anteriormente.

La metodología usada para interpretar los componentes principales ha sido la de dar más valor a aquellos parámetros con coeficientes más alto en valor (positivo o negativo) Rojo fuerte si el valor, positivo, está tanto más alejado de cero , azul

oscuro si el valor, negativo, está más alejado de cero, blanco para valores cercanos a cero. La tabla de visualización fue obtenida utilizando la función de formato condicional en EXCEL

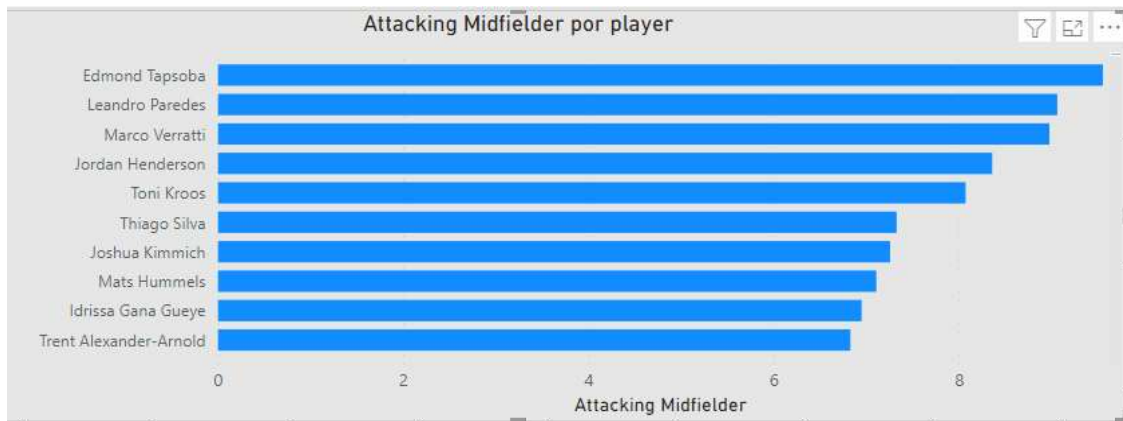
Parameter	PC1	PC2	PC3	PC4	PC5
passes_short_Per90	-0.541161263	0.577479867	0.168593028	0.127471756	-0.391871483
passes_medium_Per90	-0.944588625	0.002233243	-0.142553812	0.117998028	-0.08844224
passes_long_Per90	-0.837117002	-0.091630668	-0.119828944	-0.15036028	0.425179753
passes_received_Per90	-0.785789014	0.490628288	-0.139425274	0.228341525	-0.119529255
touches_Per90	-0.932736733	0.259961085	0.007796839	0.057555415	-0.124865788
passes_into_final_third_Per90	-0.77694643	0.352257824	0.116463412	0.303453484	0.168091979
tackles_def_3rd_PAdj_Per90	-0.436054651	-0.504576197	0.525496873	-0.204956449	-0.124138504
tackles_mid_3rd_PAdj_Per90	-0.175559795	-0.062882166	0.813420127	0.280451472	0.08497037
tackles_att_3rd_PAdj_Per90	0.292737054	0.323138309	0.570134369	0.168360634	-0.111194788
assisted_shots_Per90	0.194693667	0.797238568	0.168663006	-0.338341032	0.1761599
carries_Per90	-0.794533582	0.497285737	-0.046280588	0.158207479	-0.164434243
interceptions_PAdj_Per90	-0.434059092	-0.47592935	0.409863213	-0.13344653	0.03458928
passes_progressive_distance_Per90	-0.935553175	-0.062926403	-0.114435452	-0.110697375	-0.001871615
carry_progressive_distance_Per90	-0.66196776	0.443045212	-0.133841938	-0.067190863	-0.211502897
progressive_passes_Per90	-0.671612982	0.560234726	0.211004717	-0.010137519	0.087550943
shots_total_Per90	0.593226227	0.574746313	-0.135619949	0.051088372	0.172065217
pressures_def_3rd_PAdj_Per90	-0.345518348	-0.505499773	0.661836688	-0.129767804	-0.021578028
pressures_mid_3rd_PAdj_Per90	0.280123683	0.094082358	0.761533734	0.376260843	0.229315856
pressures_att_3rd_PAdj_Per90	0.746190798	0.424980511	0.202772539	0.166968525	0.041095818
dribbles_Per90	0.375266342	0.624815023	0.19125905	-0.188769469	-0.179073457
through_balls_Per90	-0.009045253	0.644863465	0.085045281	0.076192892	0.270369976
crosses_Per90	0.104451493	0.456755273	0.197551602	-0.740166746	-0.134545336
clearances_PAdj_Per90	-0.417870084	-0.720846333	-0.218940169	-0.17637561	0.087018063
blocks_PAdj_Per90	-0.264521816	-0.516221702	0.50626549	-0.292626042	-0.114427217
passes_switches_Per90	-0.559510324	0.213130684	0.003292136	-0.19221071	0.667732592
passes_into_penalty_area_Per90	0.004991778	0.796191878	0.200747099	-0.423710046	-0.012802569

### 3. Componente principal 1: Mediocampista ofensivo

***“Great shot and pressure in attacking”***

En el análisis global salen dos defensores con esas características además de jugadores de mediocampo. Tienen un valor alto de disparos a puerta.

player	position	team_name
Thiago Silva	DF	Chelsea
Edmond Tapsoba	DF	Leverkusen
Jordan Henderson	MF,DF	Liverpool
Toni Kroos	MF	Real Madrid
Leandro Paredes	MF	Paris S-G
Marco Verratti	MF	Paris S-G



Si escogemos solo mediocampistas:

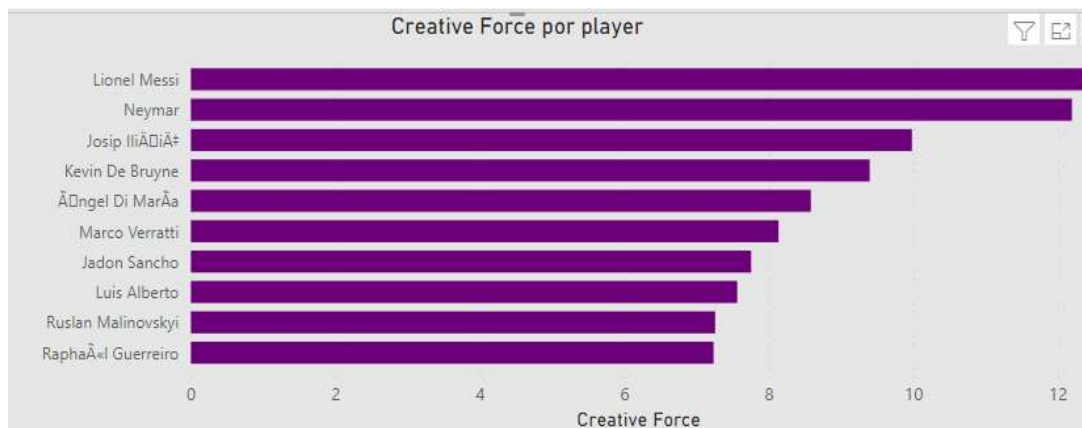
player	position	team_name
Joshua Kimmich	MF	Bayern Munich
Jordan Henderson	MF,DF	Liverpool
Toni Kroos	MF	Real Madrid
Idrissa Gana Gueye	MF	Paris S-G
Leandro Paredes	MF	Paris S-G
Marco Verratti	MF	Paris S-G

#### 4. Componente principal 2: Mediocampista en ataque- Creativo

Estos jugadores presión en la zona de ataque, son asistentes, meten balones en el area. Dependiendo del estilo del equipo podríamos encontrar mediocampista con orientación atacante o atacantes que asisten a gol, regatean, y disparan muchas veces a puerta. El top 6 de jugadores es:

player	position	team_name
Lionel Messi	FW,MF	Barcelona
Angel Di María	FW,MF	Paris S-G
Neymar	MF,FW	Paris S-G
Marco Verratti	MF	Paris S-G
Kevin De Bruyne	MF	Manchester City
Josip Ilicic	FW,MF	Atalanta

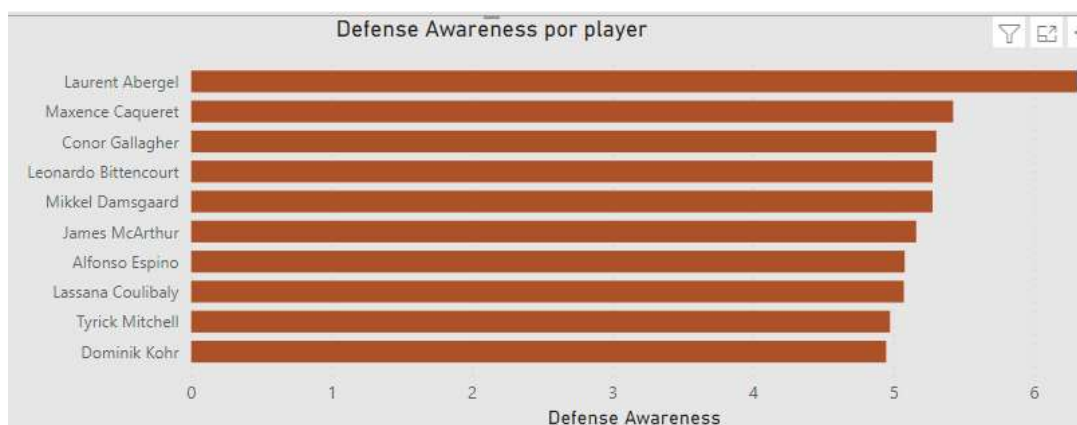
Sin mediocampistas, los mejores son Messi y Neymar Ninguna sorpresa.



### 5. Componente Principal 3. Concienciación defensiva

Jugador con mentalidad defensiva, cuya principal misión es parar los ataques rivales. Mucha presión en zona defensiva del campo y en zona de ataque, muchas entradas e intercepciones de juego.,

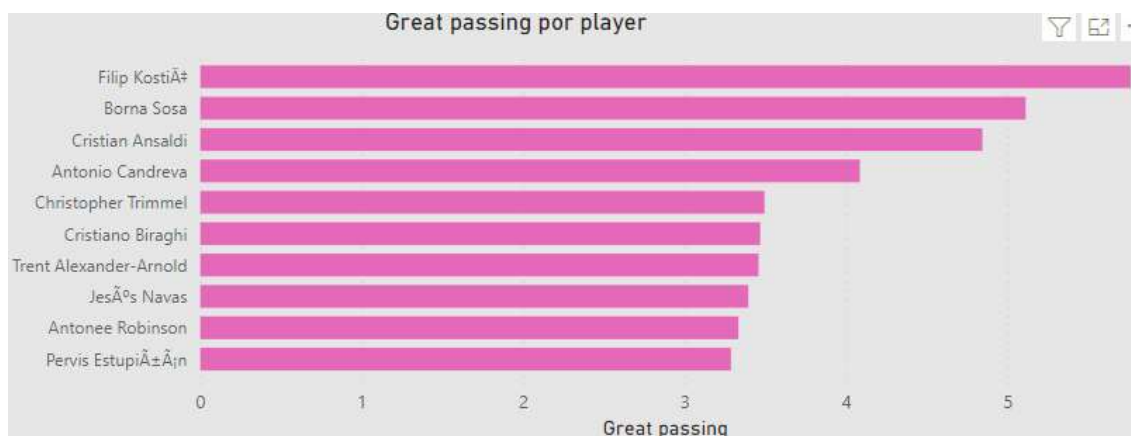
player	position	team_name
James McArthur	MF	Crystal Palace
Laurent Abergel	MF	Lorient
Conor Gallagher	MF	West Brom
Maxence Caqueret	MF	Lyon
Mikkel Damsgaard	MF,FW	Sampdoria
		Werder
Leonardo Bittencourt	MF,FW	Bremen



### 6. Componente principal 4: Gran volumen de pases y presión en campo atacante

Grandes pasadores y presionantes en zona de campo rival, así como mediocampo si el equipo sube líneas. Son laterales con gran progresión atacante. Este componente principal está altamente correlado con la posición de defensa..

player	Position	team_name
		Eint
Filip Kostia	DF	Frankfurt
Jesús Navas	DF	Sevilla
Antonee Robinson	DF	Fulham
Pervis Estupiñan	DF	Villarreal
Trent Alexander-Arnold	DF	Liverpool
Cristiano Biraghi	DF	Fiorentina
Christopher Trimmel	DF	Union Berlin
Cristian Ansaldi	DF	Torino
Borna Sosa	DF	Stuttgart
Antonio Candreva	MF	Sampdoria



Si solo escogemos mediocampistas:

Player	position	team_name
Lucas Digne	DF,MF	Everton
Rodrigo De Paul	MF	Udinese
Antonio Candreva	MF	Sampdoria
Franck Honorat	MF,FW	Brest
Josip Ilic	FW,MF	Atalanta
David Ferreira	MF,FW	Huesca

## 7. Componente principal 5. Juego directo

Jugador de precision en el pase largo, corto, cambios de juego y pases claves en el juego.

Mediocampista que dirige el juego atacante del equipo. No hay duda de que Kimmich y Kroos son los mejores ejemplos de este tipo de jugadores.

player	position	team_name
Dimitri Lionard	MF	Strasbourg
Joshua Kimmich	MF	Bayern Munich
Jonjo Shelvey	MF	Newcastle Utd
Maximilian Arnold	MF	Wolfsburg
Toni Kroos	MF	Real Madrid
Angel Montoro	MF	Granada



## 8. Comparativa de jugadores

El siguiente gráfico muestra todas estas características para 3 grandísimos mediocampistas. El parámetro de concienciación defensiva es el menos valorado para los 3, como cabía esperar de antemano, teniendo valores muy altos para los otros 3, excepto Verrati para el pase largo.

