

"Oeuvre-nalysis" Project - Phase 2

In []:

```
In [1]: import pandas as pd
#~~~~~
import numpy as np
#~~~~~
import os, path
```

```
In [2]: import matplotlib.pyplot as plt
#~~~~~
import seaborn as sns
#~~~~~
from adjustText import adjust_text
```

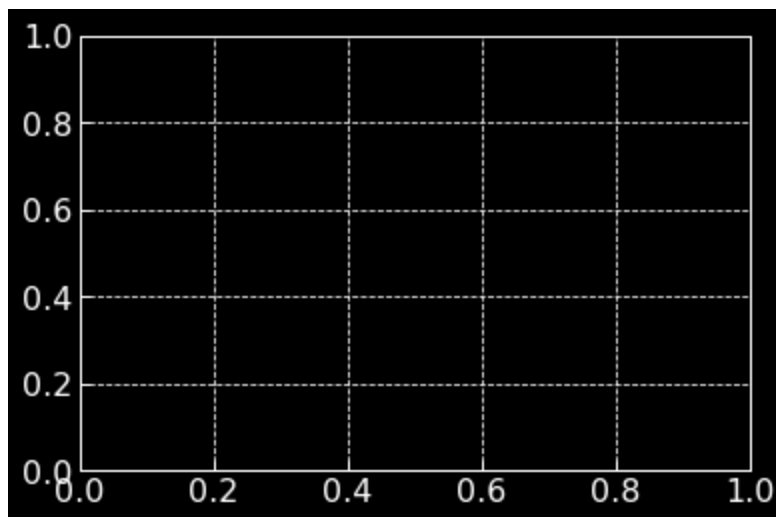
In [3]: `plt.rcParams['font.family'] = 'serif'`

In []:

```
In [4]: plt.style.use(
        [
            'bmh',
            'dark_background'
        ]
    )
```

```
In [5]: sns.set_context(
        'paper',
        font_scale = 1.8
    )

fig, ax = plt.subplots()
```



LOAD the "cleaned" data output from this

1. STANDARD_stats.csv
2. SHOOTING_stats.csv
3. GCA_and_SCA.csv
4. PASSING.csv
5. PASS_TYPES.csv
6. POSSESSION.csv
7. DEFENSIVE_ACTIONS.csv

CSV (1): "STANDARD_stats"

```
In [6]: data_folder_path = ('~/SBNation_Articles_Soccer_Data/' +  
                             'Player_Comparisons/' +  
                             'Big_5_players_by_position/' +  
                             # 'DATA_clean_grouped_by_position_20_21/'  
                             'CLEAN_but_ungrouped_DATA/'  
                             )  
data_folder_path
```

```
Out[6]: '~/SBNation_Articles_Soccer_Data/Player_Comparisons/Big_5_players_by_
position/CLEAN_but_ungrouped_DATA/'
```

```
In [7]: file_path = (data_folder_path +  
                      'STANDARD_stats.csv'  
                      )  
file_path
```

```
Out[7]: '~/SBNation_Articles_Soccer_Data/Player_Comparisons/Big_5_players_by_
position/CLEAN_but_ungrouped_DATA/STANDARD_stats.csv'
```

```
In [8]: STANDARD_stats = pd.read_csv(file_path)
```

```
In [9]: STANDARD_stats
```

```
Out[9]:
```

Player	Nation	Pos	Squad	Comp	Age	Born	MP	Starts	Min	...	x
--------	--------	-----	-------	------	-----	------	----	--------	-----	-----	---

	Player	Nation	Pos	Squad	Comp	Age	Born	MP	Starts	Min	...	x
0	Ismael Aaneba	FRA	FWDF	Strasbourg	Ligue 1	21.0	1999.0	2	0	11	...	0
1	Patrick van Aanholt	NED	DF	Crystal Palace	Premier League	29.0	1990.0	22	20	1777	...	1
2	Issah Abbas	GHA	DFFW	Mainz 05	Bundesliga	21.0	1998.0	2	0	18	...	0
3	Yunis Abdelhamid	MAR	DF	Reims	Ligue 1	32.0	1987.0	33	33	2889	...	1
4	Sabit Abdulai	GHA	MF	Getafe	La Liga	21.0	1999.0	3	0	60	...	0
...
2818	Steven Zuber	SUI	DFMF	Eint Frankfurt	Bundesliga	28.0	1991.0	20	6	585	...	0
2819	Martín Zubimendi	ESP	MF	Real Sociedad	La Liga	21.0	1999.0	31	17	1882	...	0
2820	Martin Ødegaard	NOR	MF	Arsenal	Premier League	21.0	1998.0	14	9	866	...	1
2821	Martin Ødegaard	NOR	MFFW	Real Madrid	La Liga	21.0	1998.0	7	3	234	...	0

```
In [10]: STANDARD_stats_df = STANDARD_stats[
        STANDARD_stats['Min']
        ] >= 500
```

Out[11]:

	Player	Nation	Pos	Squad	Comp	Age	Born	MP	Starts	Min	...	xG
2817	Igor Zubeldia	ESP	DF	Real Sociedad	La Liga	23.0	1997.0	24	21	1959	...	1.0
2818	Steven Zuber	SUI	DFMF	Eint Frankfurt	Bundesliga	28.0	1991.0	20	6	585	...	0.6
2819	Martín Zubimendi	ESP	MF	Real Sociedad	La Liga	21.0	1999.0	31	17	1882	...	0.6
2820	Martin Ødegaard	NOR	MF	Arsenal	Premier League	21.0	1998.0	14	9	866	...	1.1
2821	Florian Grillitsch	ESP	MF	Arsenal	Premier League	23.0	1999.0	20	27	2070	...	5.5

organize column variables into categories of related stats

```
In [12]: STANDARD_stats_COLUMNS = list(STANDARD_stats_df.columns)
```

```
STANDARD_stats_COLUMNS
```

```
Out[12]: ['Player',
          'Nation',
          'Pos',
          'Squad',
          'Comp',
          'Age',
          'Born',
          'MP',
          'Starts',
          'Min',
          '90s',
          'Gls',
          'Ast',
          'G-PK',
          'PK',
          'PKatt',
          'CrdY',
          'CrdR',
          'Gls_per90',
          'Ast_per90',
          'G+A_per90',
          'G-PK_per90',
          'G+A-PK_per90',
          'xG',
          'npxG',
          'xA',
          'npxG+xA',
          'xG_per90',
          'xA_per90',
          'xG+xA_per90',
          'npxG_per90',
          'npxG+xA_per90',
          'Name_no_spec_chars']
```

```
In [13]: demographic_vars = ['Nation',  
                             'Pos',  
                             'Squad',  
                             'Comp',  
                             'Age',  
                             'Born'  
                             ]
```

```
Out[13]: ['Nation', 'Pos', 'Squad', 'Comp', 'Age', 'Born']
```

```
In [14]: playing_time_vars = ['MP',  
                              'Starts',  
                              'Min',  
                              '90s'  
                              ]
```

```
Out[14]: ['MP', 'Starts', 'Min', '90s']
```

```
In [15]: goal_contribution_vars = ['Gls',  
                                   'Ast',  
                                   'G-PK',  
                                   'PK',  
                                   'PKatt'  
                                   ]
```

```
Out[15]: ['Gls', 'Ast', 'G-PK', 'PK', 'PKatt']
```

```
In [16]: goal_contrib_vars_PER_90 = ['Gls_per90',  
                                     'Ast_per90',  
                                     'G+A_per90',  
                                     'G-PK_per90',  
                                     'G+A-PK_per90'  
                                     ]
```

```
Out[16]: ['Gls_per90', 'Ast_per90', 'G+A_per90', 'G-PK_per90', 'G+A-PK_per90']
```

```
In [17]: expected_goal_contrib_vars = ['xG',  
                                       'npxG',  
                                       'xA',  
                                       'npxG+xA'  
                                       ]
```

```
Out[17]: ['xG', 'npxG', 'xA', 'npxG+xA']
```

```
In [18]: xG_and_xA_vars_PER_90 = ['xG_per90',  
                                   'xA_per90',  
                                   'xG+xA_per90',  
                                   'npxG_per90',  
                                   'npxG+xA_per90'  
                                   ]
```

```
Out[18]: ['xG_per90', 'xA_per90', 'xG+xA_per90', 'npxG_per90', 'npxG+xA_per90']
```

NOW organize ROWS into GROUPS:

...first by LEAGUE...

...and then by POSITION

Once all rows have already been sorted into their corresponding groups once, the results of this part can then be reused for each of the subsequent six CSV files remaining

In []:

```
In [19]: LEAGUE_group_value_counts = (STANDARD_stats_df["Comp"]
                                         .value_counts()
                                         .reset_index(name = 'Distinct_Value_Counts')
                                         .rename(columns = {'index': 'League_Name'})
                                         # .sort_values('Distinct_Value_Counts')
                                         )
LEAGUE_group_value_counts
```

Out[19]:

	League_Name	Distinct_Value_Counts
0	La Liga	416
1	Serie A	408
2	Ligue 1	399
3	Premier League	392
4	Bundesliga	347

In []:

In []:

In []:

```
In [20]: POSITION_group_value_counts = (STANDARD_stats_df["Pos"]
                                         .value_counts()
                                         .reset_index(name = 'Distinct_Value_Counts')
                                         .rename(columns = {'index': 'Position_Group'})
                                         )
```

POSITION_group_value_counts

Out[20]:

	Position_Group	Distinct_Value_Counts
0	DF	695
1	MF	395
2	FW	249
3	FWMF	200
4	MFFW	165
5	GK	132
6	DFMF	55
7	MFDF	45
8	DFFW	15
9	FWDF	11

In []:

```
In [21]: multi_position_groups_list = [  
        'FWMF',  
        'MFFW',  
  
        'DFMF',  
        'MFDF',  
  
        'DFFW',  
        'FWDF'  
    ]  
multi_position_groups_list
```

Out[21]: ['FWMF', 'MFFW', 'DFMF', 'MFDF', 'DFFW', 'FWDF']

In []:

In []:

```
In [22]: multi_POSITION_players_df = STANDARD_stats_df[  
        STANDARD_stats_df["Pos"  
        #  
        #  
        #  
        ].isin(  
        )  
    ]  
multi_POSITION_players_df
```

Out[22]:

	Player	Nation	Pos	Squad	Comp	Age	Born	MP	Starts	Min	...
7	Charles Abi	FRA	FWMF	Saint-Étienne	Ligue 1	20.0	2000.0	24	9	912	...
18	Tyler Adams	USA	DFMF	RB Leipzig	Bundesliga	21.0	1999.0	27	21	1850	...
25	Martin Agirregabiria	ESP	DFMF	Alavés	La Liga	24.0	1996.0	26	16	1558	...
30	Sergio Agüero	ARG	FWMF	Manchester City	Premier League	32.0	1988.0	12	7	559	...
31	Ruben Aguilar	FRA	DFFW	Monaco	Ligue 1	27.0	1993.0	33	27	2225	...
...
2804	Arber Zeneli	KVX	MFFW	Reims	Ligue 1	25.0	1995.0	28	14	1289	...
2807	Mehdi Zerkane	ALG	MFFW	Bordeaux	Ligue 1	21.0	1999.0	27	15	1187	...
2813	Hakim Ziyech	MAR	FWMF	Chelsea	Premier League	27.0	1993.0	23	15	1172	...
2818	Steven Zuber	SUI	DFMF	Eint Frankfurt	Bundesliga	28.0	1991.0	20	6	585	...
2822	Filip Đuričić	SRB	MFFW	Sassuolo	Serie A	28.0	1992.0	32	27	2078	...

491 rows x 33 columns

In []:

In []:

In []:

In []:

NOTE TO SELF

.

FBref's methodology for producing their "scouting report pages" is explained at this link:

<https://fbref.com/en/about/scouting-reports-explained#minutes> (<https://fbref.com/en/about/scouting-reports-explained#minutes>)

.

That explanation included some important takeaways:

FBref methodology - Takeaway (1)

"Position Sets"

.

Players are grouped into SIX different "position sets":

.

GOALKEEPERS

.

CENTER BACKS

.

FULLBACKS

.

MIDFIELDERS

.

ATTACKING MIDS and WINGERS

.

FORWARDS

(...basically just for STRIKERS...???)

.

In []:

In []:

In []:

FBref methodology - Takeaway (2)

"Comparison Pools"

Comparison pools are groups of players in each position set who are used when calculating percentiles.

To be included in a comparison pool, players need to reach a minimum number of minutes played at that position.

These minimums have been set to balance...

- ...the need to ensure that only players with an adequate amount of playing time are included...
- ...but also so that there are enough qualified players in the pool.

Specifically, for all of the Big Five leagues, the magic number currently for that "minimum" requirement...

...excludes any players who have NOT yet played at least 450 minutes

<https://fbref.com/en/about/scouting-reports-explained#minutes> (<https://fbref.com/en/about/scouting-reports-explained#minutes>)

Separating all Midfielders into four different categories...

- ...Defensive...
- ...Central...
- ...Attacking...
- ...and Wide...

...caused each resulting comparison pool to include too few players.

Instead, the people at FBref made the following important decisions:

the "position set" for "Midfielders" combines Defensive, Central and Wide Midfielders

a separate "position set" exists for Attacking Midfielders and Wingers

the "position set" for "Fullbacks" similarly includes Wingbacks

In []:

```
In [23]: messi_row_check = multi_POSITION_players_df[
                                multi_POSITION_players_df[
                                    ]
                                ]

messi_row_check
```

Out[23]:

	Player	Nation	Pos	Squad	Comp	Age	Born	MP	Starts	Min	...	xG	npG
1724	Lionel Messi	ARG	FWMF	Barcelona	La Liga	33.0	1987.0	35	33	3023	...	23.6	19.5

1 rows x 33 columns

In []:

...apparently perfect SPELLING matters:

```
In [24]: perisic_row = multi_POSITION_players_df[
                                multi_POSITION_players_df["Pla
                                ] ==
```

```
perisic_row
```

Out[24]:

Player	Nation	Pos	Squad	Comp	Age	Born	MP	Starts	Min	...	xG	npG	xA	npG+x
0 rows × 33 columns														

^ note how it failed without the correct "š" and "ć"

also note that it did NOT produce an error message...

...instead returned empty DF with same columns but 0 rows

```
In [25]: perisic_row = multi_POSITION_players_df[
        multi_POSITION_players_df["Player"] == "Ivan Perišić"]
perisic_row
```

Out[25]:

	Player	Nation	Pos	Squad	Comp	Age	Born	MP	Starts	Min	...	xG	npG	xA
2065	Ivan Perišić	CRO	DFFW	Inter	Serie A	31.0	1989.0	32	20	1795	...	4.4	4.4	4.4
1 rows × 33 columns														

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [26]: hybrid_position_counts_per_league = (
        multi_POSITION_players_df[
```



```

    ] == "Christian Erik
eriksen_row

```

Out[27]:

	Player	Nation	Pos	Squad	Comp	Age	Born	MP	Starts	Min	...	xG	npG	xA
821	Christian Eriksen	DEN	MF	Inter	Serie A	28.0	1992.0	26	17	1385	...	1.7	1.7	2.6

1 rows × 33 columns

In []:

```

In [28]: alberto_row = STANDARD_stats_df[STANDARD_stats_df["Player"
    ] == "Luis Alberto"
alberto_row

```

Out[28]:

	Player	Nation	Pos	Squad	Comp	Age	Born	MP	Starts	Min	...	xG	npG	xA
55	Luis Alberto	ESP	MF	Lazio	Serie A	27.0	1992.0	34	33	2618	...	5.5	5.3	8.2

1 rows × 33 columns

In []:

```

In [29]: calhanoglu_row = STANDARD_stats_df[
    STANDARD_stats_df[
        "Player"
    ] == "Hakan Çalhan
calhanoglu_row

```

Out[29]:

	Player	Nation	Pos	Squad	Comp	Age	Born	MP	Starts	Min	...	xG	npG
421	Hakan Çalhanoglu	TUR	MFFW	Milan	Serie A	26.0	1994.0	33	30	2614	...	5.5	4.8

1 rows × 33 columns

In []:

In []:

NOTE to self:

Eriksen and Alberto both simple, with "Pos" = "MF"

Initial Plan for Approaching Next Step:

start by just getting the averages/percentiles

...for each stat, only using rows listed as "MF"

the main point of this analysis is to get a sense of how well Çalhanoglu can adjust to the deeper LCM role, as opposed to his usual "trequartista" CAM role

comparing his CAM stats to everyone else's ACTUAL midfielder stats should of course be taken with a grain of salt as a result...

...that being said, it's still the most logical way to at least start out

In []:

```
In [30]: # | STANDARD_stats_df["Player"]  
# | == "Luis Alberto"
```

In []:

```
In [31]: MF_standard_stats_Big_5 = STANDARD_stats_df[  
    (STANDARD_stats_df["Pos"  
    ] == "M  
    ) | # "or"  
    (STANDARD_stats_df["Playe
```

```

)
] == "H
]
MF_standard_stats_Big_5

```

Out[31]:

	Player	Nation	Pos	Squad	Comp	Age	Born	MP	Starts	Min	...	xG	npk(
5	Mehdi Abeid	ALG	MF	Nantes	Ligue 1	27.0	1992.0	18	15	1320	...	1.0	1.0
6	Laurent Abergel	FRA	MF	Lorient	Ligue 1	27.0	1993.0	38	32	2963	...	1.3	1.0
22	Yacine Adli	FRA	MF	Bordeaux	Ligue 1	20.0	2000.0	35	25	2418	...	1.4	1.0
26	Lucien Agoume	FRA	MF	Spezia	Serie A	18.0	2002.0	12	9	800	...	0.0	0.0
33	Matteo Ahlinvi	BEN	MF	Nîmes	Ligue 1	21.0	1999.0	20	10	899	...	1.6	1.0
...
2794	Andre-Frank Zambo Anguissa	CMR	MF	Fulham	Premier League	24.0	1995.0	36	29	2587	...	1.8	1.0
2795	Niccolò Zanelato	ITA	MF	Crotone	Serie A	22.0	1998.0	27	18	1637	...	0.5	0.0
2809	Piotr Zieliński	POL	MF	Napoli	Serie A	26.0	1994.0	36	32	2523	...	5.4	5.0
2819	Martín Zubimendi	ESP	MF	Real Sociedad	La Liga	21.0	1999.0	31	17	1882	...	0.6	0.0
2820	Martin Ødegaard	NOR	MF	Arsenal	Premier League	21.0	1998.0	14	9	866	...	1.1	1.0

396 rows × 33 columns

In []:

In []:

now create separate DF with RANKS and...

...PERCENTILES for each corresponding column

In []:

In []:

```
In [32]: MF_big_5_std_NUMERICAL_stats = MF_standard_stats_Big_5.loc[:,  
                                     'Age' : 'np  
                                     ]  
MF_big_5_std_NUMERICAL_stats
```

Out[32]:

	Age	Born	MP	Starts	Min	90s	Gls	Ast	G-PK	PK	...	G+A- PK_per90	xG	np	xA
5	27.0	1992.0	18	15	1320	14.7	0	0	0	0	...	0.00	1.0	1.0	0.8
6	27.0	1993.0	38	32	2963	32.9	3	1	3	0	...	0.12	1.3	1.3	1.4
22	20.0	2000.0	35	25	2418	26.9	2	5	2	0	...	0.26	1.4	1.4	4.7
26	18.0	2002.0	12	9	800	8.9	0	0	0	0	...	0.00	0.0	0.0	0.1
33	21.0	1999.0	20	10	899	10.0	2	0	2	0	...	0.20	1.6	1.6	0.3
...
2794	24.0	1995.0	36	29	2587	28.7	0	3	0	0	...	0.10	1.8	1.8	2.2
2795	22.0	1998.0	27	18	1637	18.2	1	1	1	0	...	0.11	0.5	0.5	1.3
2809	26.0	1994.0	36	32	2523	28.0	8	10	8	0	...	0.64	5.4	5.4	5.4
2819	21.0	1999.0	31	17	1882	20.9	0	0	0	0	...	0.00	0.6	0.6	0.6
2820	21.0	1998.0	14	9	866	9.6	1	2	1	0	...	0.31	1.1	1.1	2.1

396 rows × 27 columns

In []:

```
In [33]: MF_big_5_std_NUMERICAL_stats.insert(  
        loc = 0,  
        column = 'Player',  
        value = MF_standard_stats_Big_5['P  
        ]  
        )  
MF_big_5_std_NUMERICAL_stats
```

Out[33]:

	Player	Age	Born	MP	Starts	Min	90s	Gls	Ast	G-PK	...	G+A- PK_per90	xG	np	xA
5	Mehdi Abeid	27.0	1992.0	18	15	1320	14.7	0	0	0	...	0.00	1.0	1	
6	Laurent Abergel	27.0	1993.0	38	32	2963	32.9	3	1	3	...	0.12	1.3	1	

	Player	Age	Born	MP	Starts	Min	90s	Gls	Ast	G-PK	...	G+A-PK_per90	xG	np
22	Yacine Adli	20.0	2000.0	35	25	2418	26.9	2	5	2	...	0.26	1.4	1
26	Lucien Agoume	18.0	2002.0	12	9	800	8.9	0	0	0	...	0.00	0.0	0
33	Matteo Ahlinvi	21.0	1999.0	20	10	899	10.0	2	0	2	...	0.20	1.6	1
...
2794	Andre-Frank Zambo Anguissa	24.0	1995.0	36	29	2587	28.7	0	3	0	...	0.10	1.8	1
2795	Niccolò Zanellato	22.0	1998.0	27	18	1637	18.2	1	1	1	...	0.11	0.5	0
2809	Piotr Zieliński	26.0	1994.0	36	32	2523	28.0	8	10	8	...	0.64	5.4	5
2819	Martín Zubimendi	21.0	1999.0	31	17	1882	20.9	0	0	0	...	0.00	0.6	0

In []:

In []:

```
In [34]: MF_big_5_std_NUMERICAL_stats.set_index('Player', inplace = True)
```

Out[34]:

[illegible]

In []:

```
In [ ]:
```

[illegible]

	Age	Born	MP	Starts	Min	90s	Gls	Ast	G-PK	PK	...	G+A-PK_per90	xG	npxG
Player														
Andre-Frank Zambo Anguissa	240	145	19	76	71	71	250	59	239	43	...	211	126	111
Niccolò Zanellato	310	48	204	225	224	224	150	164	141	43	...	202	301	295
Piotr Zieliński	169	180	19	39	87	87	12	4	8	43	...	6	22	12
Martín	346	28	127	238	184	184	250	259	239	43	...	319	287	281

In []:

In []:

In []:

In []:

```
In [36]: MF_big_5_std_stat_PERCENTILES = (MF_big_5_std_NUMERICAL_stats.  
#                                     rank(ascending = 1,  
                                           0 = False  
                                           method = 'min',  
                                           pct = True  
                                           )  
#                                     .astype(int)  
#                                     )  
MF_big_5_std_stat_PERCENTILES
```

Out[36]:

	Age	Born	MP	Starts	Min	90s	Gls	Ast	
Player									
Mehdi Abeid	0.578283	0.308081	0.141414	0.275253	0.305556	0.305556	0.002525	0.002525	0.0
Laurent Abergel	0.578283	0.371212	0.982323	0.863636	0.949495	0.949495	0.755051	0.351010	0.7
Yacine Adli	0.050505	0.934343	0.863636	0.623737	0.755051	0.755051	0.626263	0.896465	0.6
Lucien Agoume	0.022727	0.977273	0.022727	0.103535	0.093434	0.093434	0.002525	0.002525	0.0
Matteo Ahlinvi	0.080808	0.883838	0.186869	0.128788	0.126263	0.126263	0.626263	0.002525	0.6

	Age	Born	MP	Starts	Min	90s	Gls	Ast
Player								
...
Andre-Frank Zambo Anguissa	0.323232	0.550505	0.914141	0.770202	0.823232	0.815657	0.002525	0.752525
Niccolò Zanello	0.131313	0.810606	0.431818	0.404040	0.436869	0.436869	0.373737	0.351010
Piotr Zieliński	0.484848	0.436869	0.914141	0.863636	0.782828	0.777778	0.959596	0.987374
Martín Zubimendi	0.080808	0.883838	0.641414	0.343434	0.537879	0.537879	0.002525	0.002525

In []:

In []:

In []:

now extract two sub-DF's --> 3 rows each

.

...to see the RANKS first and then the...

.

...PERCENTILES in each column variable...

.

...for Eriksen, Alberto, and Çalhanoğlu

In []:

```
In [37]: playmaking_artists = [
            "Christian Eriksen",
            "Luis Alberto",
            "Hakan Çalhanoğlu"
        ]
playmaking_artists
Out[37]: ['Christian Eriksen', 'Luis Alberto', 'Hakan Çalhanoğlu']
```

In []:

first get their RANKS from...

"MF_in_Big_5_std_stat_RANKS"

```
In [38]: # playmaking_artist_RANKS = MF_in_Big_5_std_stat_RANKS[MF_in_Big_5_std_stat_RANKS
#
#
# playmaking_artist_RANKS
```

NOTE to self:

.

the ABOVE code chunk did NOT work, because the...

..."Player" column was set to be the INDEX now

.

the BELOW code chunk actually DID work

.

...however, I wanted to try reformatting in a separate cell in order to make it...

...more readable, but without the risk of messing up what had actually worked

.

```
In [39]: # playmaking_artist_RANKS = MF_in_Big_5_std_stat_RANKS[MF_in_Big_5_std_stat_RANKS
#
# playmaking_artist_RANKS
```

```
In [40]: playmaking_artist_RANKS = (
        MF_in_Big_5_std_stat_RANKS[
            MF_in_Big_5_std_stat_RANKS
                .index
                .isin(
                    playmaking_artist_RANKS
                )
            ]
        )
```

Out[40]: playmaking_artist_RANKS

	Age	Born	MP	Starts	Min	90s	Gls	Ast	G-PK	PK	...	G+A- PK_per90	xG	npG
Player														
Luis Alberto	139	251	56	28	65	64	8	100	4	43	...	30	20	13
Hakan Çalhanoğlu	169	180	75	66	68	68	46	7	48	24	...	27	20	17
Christian Eriksen	109	251	227	238	264	264	63	259	48	43	...	113	132	119

3 rows × 27 columns

In []:

In []:

In []:

then get the PERCENTILES from...

"MF_big_5_std_stat_PERCENTILES"

.

```
In [41]: playmaking_artist_PERCENTILES = (MF_big_5_std_stat_PERCENTILES[MF_big_5_std_stat_PERCENTILES.index
                                          .isin(p)]
                                          )
playmaking_artist_PERCENTILES
```

Out[41]:

	Age	Born	MP	Starts	Min	90s	Gls	Ast
Player								
Luis Alberto	0.578283	0.308081	0.815657	0.906566	0.838384	0.833333	0.974747	0.590909
Hakan Çalhanoğlu	0.484848	0.436869	0.747475	0.813131	0.830808	0.828283	0.845960	0.972222
Christian Eriksen	0.654040	0.308081	0.404040	0.343434	0.335859	0.335859	0.755051	0.002525

3 rows × 27 columns

In []:

In []:

In []:

now TRANSPOSE both mini-DataFrames

In []:

In []:

In []:

```
In [42]: RANKS_transposed = playmaking_artist_RANKS.transpose()
```

RANKS_transposed

Out[42]:

Player	Luis Alberto	Hakan Çalhanoğlu	Christian Eriksen
Age	139	169	109
Born	251	180	251
MP	56	75	227
Starts	28	66	238
Min	65	68	264
90s	64	68	264
Gls	8	46	63
Ast	100	7	259
G-PK	4	48	48
PK	43	24	43
PKatt	47	29	47
CrdY	142	199	346
CrdR	71	71	71
Gls_per90	14	84	47
Ast_per90	165	13	259
G+A_per90	39	29	123
G-PK_per90	7	102	35
G+A-PK_per90	30	27	113

Player	Luis Alberto	Hakan Çalhanoğlu	Christian Eriksen
xG	20	20	132
npG	13	17	119
xA	6	5	79
npG+xA	8	8	100
xG_per90	35	35	98
xA_per90	8	5	40
xG+xA_per90	10	9	64

In []:

In []:

In []:

In [43]: PERCENTILES_transposed = playmaking_artist_PERCENTILES.transpose()
PERCENTILES_transposed

Out[43]:

Player	Luis Alberto	Hakan Çalhanoğlu	Christian Eriksen
Age	0.578283	0.484848	0.654040
Born	0.308081	0.436869	0.308081
MP	0.815657	0.747475	0.404040
Starts	0.906566	0.813131	0.343434
Min	0.838384	0.830808	0.335859
90s	0.833333	0.828283	0.335859
Gls	0.974747	0.845960	0.755051
Ast	0.590909	0.972222	0.002525
G-PK	0.984848	0.782828	0.782828
PK	0.002525	0.896465	0.002525
PKatt	0.002525	0.886364	0.002525
CrdY	0.502525	0.363636	0.050505
CrdR	0.002525	0.002525	0.002525
Gls_per90	0.964646	0.777778	0.871212
Ast_per90	0.535354	0.967172	0.002525
G+A_per90	0.898990	0.921717	0.666667
G-PK_per90	0.984848	0.709596	0.901515
G+A-PK_per90	0.921717	0.934343	0.694444

Player	Luis Alberto	Hakan Çalhanoğlu	Christian Eriksen
xG	0.949495	0.949495	0.648990
npG	0.969697	0.959596	0.676768
xA	0.987374	0.989899	0.780303
npG+xA	0.979798	0.979798	0.742424
xG_per90	0.891414	0.891414	0.727273
xA_per90	0.977273	0.984848	0.893939
xG+xA_per90	0.969697	0.979798	0.833333
npG_per90	0.969697	0.959596	0.780303

In []:

In []:

In []:

the following CODE chunk, copied from...

my latest draft in the "Oeuvrenalysis" folder...

...successfully produced a FACET GRID with...

...every single variable for all three players!!!

```
In [44]: # ax = Stats_df.plot.barh(
#             subplots = True,
#             use_index = False
#             layout = (25, # rows
#                       1 # columns
#                       ),
#             figsize = (5,
#                        125
#                        ),
#             sharex = False
#             #
#             #
#             )
```

...HOWEVER, that was only possible/practical because...

...it used just THREE players... as opposed to 396 players

.

In []:

In []:

In []:

In []:

In []:

In [45]:

Out[45]:

	Player	Nation	Pos	Squad	Comp	Age	Born	MP	Starts	Min	...	xG	npv
5	Mehdi Abeid	ALG	MF	Nantes	Ligue 1	27.0	1992.0	18	15	1320	...	1.0	1.0
6	Laurent Abergel	FRA	MF	Lorient	Ligue 1	27.0	1993.0	38	32	2963	...	1.3	1.0
22	Yacine Adli	FRA	MF	Bordeaux	Ligue 1	20.0	2000.0	35	25	2418	...	1.4	1.0
26	Lucien Agoume	FRA	MF	Spezia	Serie A	18.0	2002.0	12	9	800	...	0.0	0.0
33	Matteo Ahlinvi	BEN	MF	Nimes	Ligue 1	21.0	1999.0	20	10	899	...	1.6	1.0
...
2794	Andre-Frank Zambo Anguissa	CMR	MF	Fulham	Premier League	24.0	1995.0	36	29	2587	...	1.8	1.0
2795	Niccolò Zanelato	ITA	MF	Crotone	Serie A	22.0	1998.0	27	18	1637	...	0.5	0.0
2809	Piotr Zieliński	POL	MF	Napoli	Serie A	26.0	1994.0	36	32	2523	...	5.4	5.0
2819	Martín Zubimendi	ESP	MF	Real Sociedad	La Liga	21.0	1999.0	31	17	1882	...	0.6	0.0
2820	Martin Ødegaard	NOR	MF	Arsenal	Premier League	21.0	1998.0	14	9	866	...	1.1	1.0

396 rows × 33 columns

In []:

In []:

In []:

```
In [46]: playmaking_artist_STATS = MF_standard_stats_Big_5[
                                                MF_standard_stats_Bi

                                                ]

playmaking_artist_STATS
```

Out[46]:

	Player	Nation	Pos	Squad	Comp	Age	Born	MP	Starts	Min	...	xG	npG
55	Luis Alberto	ESP	MF	Lazio	Serie A	27.0	1992.0	34	33	2618	...	5.5	5.3
421	Hakan Çalhanoğlu	TUR	MFFW	Milan	Serie A	26.0	1994.0	33	30	2614	...	5.5	4.8
821	Christian Eriksen	DEN	MF	Inter	Serie A	28.0	1992.0	26	17	1385	...	1.7	1.7

3 rows × 33 columns

In []:

```
In [47]: playmaking_artist_STATS.set_index(
                                                'Player',
                                                inplace = True
                                                )
```

In []:

```
In [48]: playmaking_artists_TRANSPOSED = playmaking_artist_STATS.transpose()

playmaking_artists_TRANSPOSED
```

Out[48]:

	Player	Luis Alberto	Hakan Çalhanoğlu	Christian Eriksen
	Nation	ESP	TUR	DEN
	Pos	MF	MFFW	MF
	Squad	Lazio	Milan	Inter
	Comp	Serie A	Serie A	Serie A
	Age	27	26	28

Player	Luis Alberto	Hakan Çalhanoğlu	Christian Eriksen
Born	1992	1994	1992
MP	34	33	26
Starts	33	30	17
Min	2618	2614	1385
90s	29.1	29	15.4
Gls	9	4	3
Ast	2	9	0
G-PK	9	3	3
PK	0	1	0
PKatt	0	1	0
CrdY	5	4	1
CrdR	0	0	0
Gls_per90	0.31	0.14	0.19
Ast_per90	0.07	0.31	0
G+A_per90	0.38	0.45	0.19
G-PK_per90	0.31	0.1	0.19
G+A-PK_per90	0.38	0.41	0.19
xG	5.5	5.5	1.7
npxG	5.3	4.8	1.7
xA	8.2	8.8	2.6
npxG+xA	13.5	13.5	4.3
xG_per90	0.19	0.19	0.11
xA_per90	0.28	0.3	0.17
xG+xA_per90	0.47	0.49	0.28
npxG_per90	0.18	0.16	0.11
npxG+xA_per90	0.46	0.47	0.28

In []:

In []:

In []:

SWARM + BOX plots removed for now

Instead, focus should be on placing the three playmaking artists within four different contexts of comparison:

(1) Serie A midfielders

(2) Big Five midfielders

(3) ALL players in Serie A

(4) ALL players in Big Five

Radars next to facet grids? For each data category...

3 “artists” times 8 data categories = 24 different radars!

In []:

In []:

In []:

In []:

In []:

```
In [49]: # demographic_vars = ['Nation',  
#                             'Pos',  
#                             'Squad',  
#                             'Comp',  
#                             'Age',
```

```
#             'Born'  
#             ]
```

Out[49]: ['Nation', 'Pos', 'Squad', 'Comp', 'Age', 'Born']

In []:

In []:

```
In [50]: # playing_time_vars = ['MP',  
#                             'Starts',  
#                             'Min',  
#                             '90s'  
#                             ]
```

playing_time_vars

Out[50]: ['MP', 'Starts', 'Min', '90s']

In []:

In []:

```
In [51]: # goal_contribution_vars = ['Gls',  
#                                   'Ast',  
#                                   'G-PK',  
#                                   'PK',  
#                                   'PKatt'  
#                                   ]
```

goal_contribution_vars

Out[51]: ['Gls', 'Ast', 'G-PK', 'PK', 'PKatt']

In []:

In []:

```
In [52]: # goal_contrib_vars_PER_90 = ['Gls_per90',  
#                                      'Ast_per90',  
#                                      'G+A_per90',  
#                                      'G-PK_per90',  
#                                      'G+A-PK_per90'  
#                                      ]
```

goal_contrib_vars_PER_90

Out[52]: ['Gls_per90', 'Ast_per90', 'G+A_per90', 'G-PK_per90', 'G+A-PK_per90']

In []:

In []:

```
In [53]: # expected_goal_contrib_vars = ['xG',
```

```
#           'npxG',
#           'xA',
#           'npxG+xA'
#       ]
```

Out[53]: ['xG', 'npxG', 'xA', 'npxG+xA']

In []:

In []:

```
In [54]: # xG_and_xA_vars_PER_90 = ['xG_per90',
#                                   'xA_per90',
#                                   'xG+xA_per90',
#                                   'npxG_per90',
#                                   'npxG+xA_per90'
#                                   ]
```

Out[54]: ['xG_per90', 'xA_per90', 'xG+xA_per90', 'npxG_per90', 'npxG+xA_per90']

In []:

In []:

In []:

In []:

In []:

REMEMBER:

the entire goal of THIS folder...

...and therefore THIS code notebook...

...is to produce a LOOKUP TABLE...

...filled with RANKS and PERCENTILES...

.

...to provide CONTEXT for LATER...

.

...and THEN using it to analyze...

.

...specific players (with visualizations)

In []:

In []:

In []:

In []:

In []:

(1):

.

Serie_A_midfield_RANKS

.

Serie_A_midfield_PERCENTILES

.

(2):

.

Serie_A_all_pos_RANKS

.

Serie_A_all_pos_PERCENTILES

.

(3):

.

Big_5_midfield_RANKS

.

Big_5_midfield_PERCENTILES

.

(4):

.

Big_5_all_pos_RANKS

.

Big_5_all_pos_PERCENTILES

In []:

In []:

In []:

In []:

In []:

In []: