

TRANSFER BUSINESS OF EPL CLUBS

SURAAJ SHRESTHA

ABOUT DATA

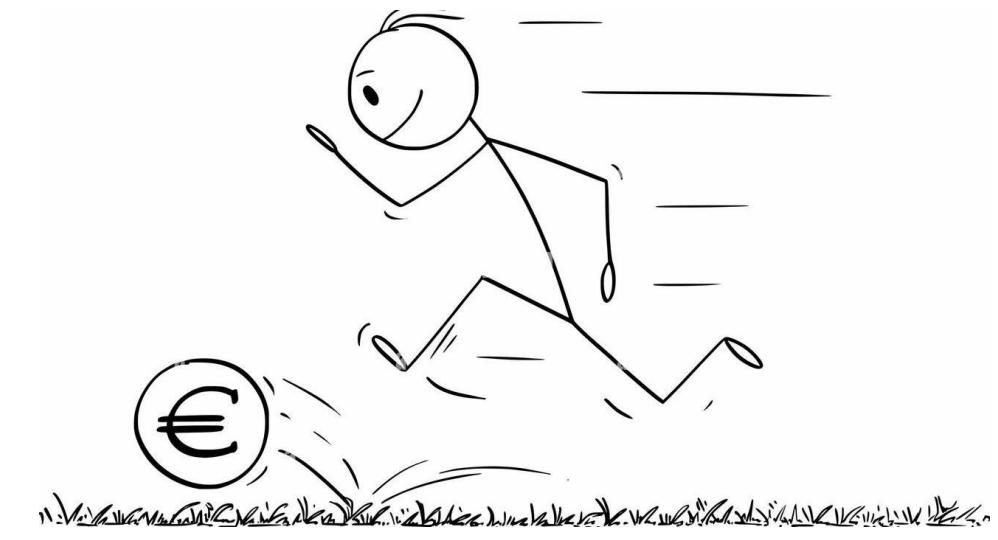
- DATASET FROM KAGGLE
- TRANSFER LISTINGS FROM SEASON 1992/1993 UNTIL NOW.

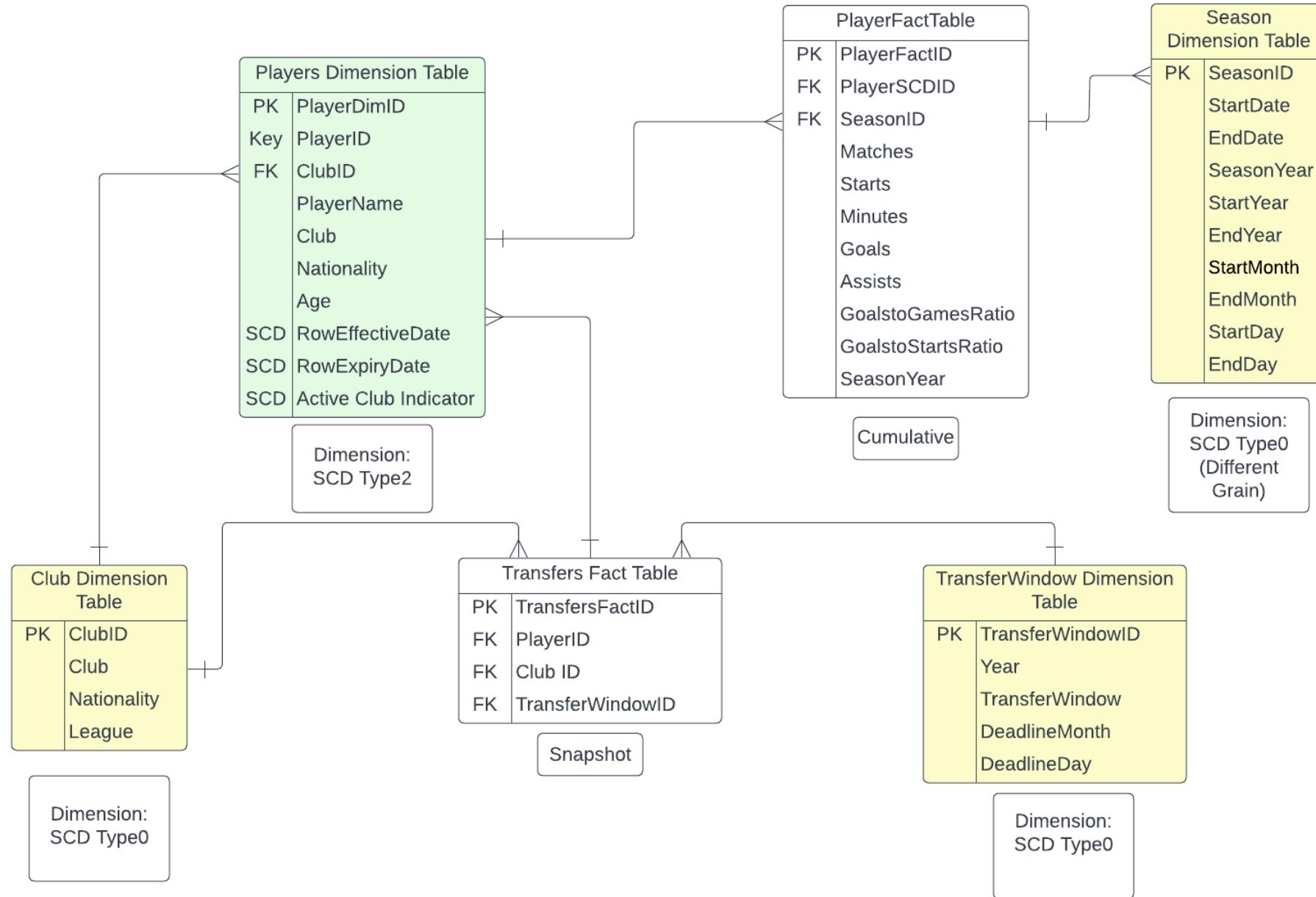
club_name	player_name	age	position	club_involve	fee	transfer_mo	transfer_per	fee_cleaned	league_name	year	season
Middlesbrou	Tommy Wrig	26	Left Winger	Leicester	,£'910Th.	in	Summer	0.91	Premier Leag	1992	1992/1993
Middlesbrou	Jonathan Git	28	defence	Southampton	,£'250Th.	in	Summer	0.25	Premier Leag	1992	1992/1993
Middlesbrou	Chris Morris	28	Right-Back	Celtic	?	in	Summer	NA	Premier Leag	1992	1992/1993
Middlesbrou	Ben Roberts	17	Goalkeeper	Boro U18	-	in	Summer	NA	Premier Leag	1992	1992/1993
Middlesbrou	Andy Todd	17	Centre-Back	Boro U18	-	in	Summer	NA	Premier Leag	1992	1992/1993
Middlesbrou	Stuart Ripley	24	Right Wing	Blackburn	,£'2.00m	out	Summer	2	Premier Leag	1992	1992/1993
Middlesbrou	Gary Parkins	24	Right-Back	Southend Un	loan transfer	out	Summer	NA	Premier Leag	1992	1992/1993
Middlesbrou	Bernie Slave	31	Centre-Forw	Port Vale	?	out	Summer	NA	Premier Leag	1992	1992/1993
Middlesbrou	Andy Payton	24	attack	Celtic	?	out	Summer	NA	Premier Leag	1992	1992/1993
Middlesbrou	Ian Arnold	20	Centre-Forw	Carlisle Unit	?	out	Summer	NA	Premier Leag	1992	1992/1993
Middlesbrou	Dwight Mars	27	attack	Plymouth	End of loan	out	Summer	NA	Premier Leag	1992	1992/1993
Norwich City	Mark Robins	22	Centre-Forw	Man Utd	,£'1.00m	in	Summer	1	Premier Leag	1992	1992/1993
Norwich City	Gary Meggsor	33	Defensive M	Man City	free transfer	in	Summer	0	Premier Leag	1992	1992/1993
Norwich City	Mark Peters	19	defence	Man City	free transfer	in	Summer	0	Premier Leag	1992	1992/1993
Norwich City	Darren Eadie	17	Left Midfield	Norwich U18	-	in	Summer	NA	Premier Leag	1992	1992/1993
Norwich City	Andy Johnson	18	Attacking Mi	Norwich U18	-	in	Summer	NA	Premier Leag	1992	1992/1993
Norwich City	Lee Power	20	Centre-Forw	Charlton	End of loan	in	Summer	NA	Premier Leag	1992	1992/1993
Norwich City	Tim Sherwo	23	Central Midf	Blackburn	,£'750Th.	out	Summer	0.75	Premier Leag	1992	1992/1993
Norwich City	Paul Blades	27	Right-Back	Wolves	,£'378Th.	out	Summer	0.378	Premier Leag	1992	1992/1993
Norwich City	Steve Ball	22	midfield	Cambridge U	free transfer	out	Summer	0	Premier Leag	1992	1992/1993
Norwich City	Adrian Penn	21	midfield	Bournemouth	,£'36Th.	out	Summer	0.036	Premier Leag	1992	1992/1993

- PLAYER STATS FROM SEASONS 2020/2021 & 2021/2022.

Name	Club	Nationality	Position	Age	Matches	Starts	Mins	Goals	Assists
Mason Mour	Chelsea	ENG	MF,FW	21	36	32	2890	6	5
Edouard Mer	Chelsea	SEN	GK	28	31	31	2745	0	0
Timo Werne	Chelsea	GER	FW	24	35	29	2602	6	8
Ben Chilwell	Chelsea	ENG	DF	23	27	27	2286	3	5
Reece James	Chelsea	ENG	DF	20	32	25	2373	1	2
CV@sar Azpi	Chelsea	ESP	DF	30	26	24	2188	1	2
N'Golo Kant	Chelsea	FRA	MF	29	30	24	2146	0	2
Jorginho	Chelsea	ITA	MF	28	28	23	2010	7	1
Thiago Silva	Chelsea	BRA	DF	35	23	23	1935	2	0
Kurt Zouma	Chelsea	FRA	DF	25	24	22	2029	5	0
Mateo Kovaf	Chelsea	CRO	MF	26	27	21	1815	0	1
Antonio Rvº	Chelsea	GER	DF	27	19	19	1710	1	0
Christian Pul	Chelsea	USA	FW,MF	21	27	18	1738	4	2
Kai Havertz	Chelsea	GER	MF,FW	21	27	18	1520	4	3
Andreas Chri	Chelsea	DEN	DF	24	17	15	1371	0	0
Hakim Ziyech	Chelsea	MAR	FW,MF	27	23	15	1172	2	3

Player	Team	Nation	Pos	Age	MP	Starts	Min	90s	Gls	Ast
Bukayo Saka	Arsenal	eng+ENG	FW,MF	19	38	36	2,978	33.1	11	7
Gabriel Dos	Arsenal	brt+BRA	DF	23	35	35	3,063	34	5	0
Aaron Rams	Arsenal	eng+ENG	GK	23	34	34	3,060	34	0	0
Ben White	Arsenal	eng+ENG	DF	23	32	32	2,880	32	0	0
Martin ýdega	Arsenal	not+NOR	MF	22	36	32	2,785	30.9	7	4
Granit Xhaka	Arsenal	cht+SUI	MF,DF	28	27	27	2,327	25.9	1	2
Thomas Part	Arsenal	ght+GHA	MF	28	24	23	2,028	22.5	2	1
Kieran Tierne	Arsenal	sct+SCO	DF	24	22	22	1,915	21.3	1	3
Emile Smith	Arsenal	eng+ENG	MF,FW	21	33	21	1,921	21.3	10	2
Martinelli	Arsenal	brt+BRA	FW,MF	20	29	21	1,860	20.7	6	6
Alexandre La	Arsenal	frt+FRA	FW,MF	30	30	20	1,782	19.8	4	7
Takehiro Ton	Arsenal	jpt+JPN	DF	22	21	20	1,679	18.7	0	1





DESIGN

ETL

- QUERY ONLY THE SEASON 2020/2021 & 2021/2022 AND ONLY THE TRANSFERS INTO THE CLUB
 - FIGURE OUT DIFFERENT TYPE OF TRANSFERS AND CLEAN THE FEE ACCORDINGLY

```
transferfile = "premier-league.csv"
transfer_df = pd.read_csv(transferfile)
transfer_df = transfer_df.query('season == "2020/2021" | season == "2021/2022"')
transfer_df = transfer_df.query('transfer_movement == "in"')
transfer_df = transfer_df[transfer_df['fee'].str.contains("End of loan") == False]

transfer_df['Transfer_Type'] = np.where(transfer_df['fee'] != 'loan transfer', 'Permanent Transfer', 'Loan Transfer')
transfer_df['Transfer_Type'].mask((transfer_df['fee'] == '-') & (transfer_df['club_involved_name'] != 'Without Club'), 

transfer_df.loc[(transfer_df.season == '2021/2022') & (transfer_df.transfer_period == 'Winter'), 'year'] = 2022

transfer_df.loc[(transfer_df.season == '2020/2021') & (transfer_df.transfer_period == 'Winter'), 'year'] = 2021
```

```
stats_df20['Club'].mask((stats_df20['Club'] == 'Chelsea') , 'Chelsea FC', inplace=True)
stats_df20['Club'].mask((stats_df20['Club'] == 'Arsenal') , 'Arsenal FC', inplace=True)
stats_df20['Club'].mask((stats_df20['Club'] == 'Everton') , 'Everton FC', inplace=True)
stats_df20['Club'].mask((stats_df20['Club'] == 'Southampton') , 'Southampton FC', inplace=True)
stats_df20['Club'].mask((stats_df20['Club'] == 'Brighton') , 'Brighton & Hove Albion', inplace=True)
stats_df20['Club'].mask((stats_df20['Club'] == 'Burnley') , 'Burnley FC', inplace=True)
stats_df20['Club'].mask((stats_df20['Club'] == 'Fulham') , 'Fulham FC', inplace=True)
```

```
stats_df20['Name'].mask((stats_df20['Name'] == 'Emile Smith-Rowe') , 'Emile Smith Rowe', inplace=True)
stats_df20['Name'].mask((stats_df20['Name'] == 'Gabriel Dos Santos') , 'Gabriel Magalhães', inplace=True)
stats_df20['Name'].mask((stats_df20['Name'] == 'Raphael Dias Belloli') , 'Raphinha', inplace=True)
stats_df20['Name'].mask((stats_df20['Name'] == 'Kostas Tsimikas') , 'Konstantinos Tsimikas', inplace=True)
stats_df20['Name'].mask((stats_df20['Name'] == 'Ferrán Torres') , 'Ferran Torres', inplace=True)
stats_df20['Name'].mask((stats_df20['Name'] == 'Valentino Livramento') , 'Tino Livramento', inplace=True)
stats_df20['Name'].mask((stats_df20['Name'] == 'Branislav Ivanović') , 'Branislav Ivanovic', inplace=True)
stats_df20['Name'].mask((stats_df20['Name'] == 'Pierre Höjbjerg') , 'Pierre-Emile Höjbjerg', inplace=True)
|
stats_df20.loc[(stats_df20.Name == 'Matty Cash'), 'Nationality'] = 'POL'
stats_df20.loc[(stats_df20.Name == 'Abdoulaye Doucouré'), 'Nationality'] = 'MLI'
stats_df20.loc[(stats_df20.Name == 'Hélder Costa'), 'Nationality'] = 'ANG'
stats_df20.loc[(stats_df20.Name == 'Hannibal Mejibri'), 'Nationality'] = 'TUN'
```

```
transfer_df['player_name'].mask((transfer_df['player_name'] == 'Thiago') , 'Thiago Alcántara', inplace=True)
transfer_df['player_name'].mask((transfer_df['player_name'] == 'Kiernan Dewsbury-Hall') , 'Kiernan Dewsbury Hall', inpl]
transfer_df['player_name'].mask((transfer_df['player_name'] == "Dan N'Lundulu") , 'Dan Nlundulu', inplace=True)
transfer_df['player_name'].mask((transfer_df['player_name'] == "Vladimir Coufal") , 'Vladimír Coufal', inplace=True)
transfer_df['player_name'].mask((transfer_df['player_name'] == "Nicolas N'Koulou") , 'Nicolas Nkoulou', inplace=True)
transfer_df['player_name'].mask((transfer_df['player_name'] == "Odsonne Edouard") , 'Odsonne Édouard', inplace=True)
transfer_df['player_name'].mask((transfer_df['player_name'] == "Tomas Soucek") , 'Tomáš Souček', inplace=True)
transfer_df['player_name'].mask((transfer_df['player_name'] == "Filip Krovinovic") , 'Filip Krovinović', inplace=True)
transfer_df['player_name'].mask((transfer_df['player_name'] == "Marçal") , 'Fernando Marçal', inplace=True)
transfer_df['player_name'].mask((transfer_df['player_name'] == "Toti") , 'Toti Gomes', inplace=True)
transfer_df['player_name'].mask((transfer_df['player_name'] == "Rayan Aït-Nouri") , 'Rayan Aít Nouri', inplace=True)
transfer_df['player_name'].mask((transfer_df['player_name'] == "Okay Yokuslu") , 'Okay Yokuşlu', inplace=True)
transfer_df['player_name'].mask((transfer_df['player_name'] == "Jaden Philogene-Bidace") , 'Jaden Philogene Bidace', in
transfer_df['player_name'].mask((transfer_df['player_name'] == "Peter Etebo") , 'Oghenekaro Etebo', inplace=True)
transfer_df['player_name'].mask((transfer_df['player_name'] == "Alex Kral") , 'Alex Král', inplace=True)
transfer_df['player_name'].mask((transfer_df['player_name'] == "Samir") , 'Samir Santos', inplace=True)
```

```
stats_df21['Team'].mask((stats_df21['Team'] == 'Arsenal') , 'Arsenal FC', inplace=True)
stats_df21['Team'].mask((stats_df21['Team'] == 'Fulham') , 'Fulham FC', inplace=True)
stats_df21['Team'].mask((stats_df21['Team'] == 'Chelsea') , 'Chelsea FC', inplace=True)
stats_df21['Team'].mask((stats_df21['Team'] == 'Burnley') , 'Burnley FC', inplace=True)
stats_df21['Team'].mask((stats_df21['Team'] == 'Brentford') , 'Brentford FC', inplace=True)
stats_df21['Team'].mask((stats_df21['Team'] == 'Everton') , 'Everton FC', inplace=True)
stats_df21['Team'].mask((stats_df21['Team'] == 'Liverpool') , 'Liverpool FC', inplace=True)
stats_df21['Team'].mask((stats_df21['Team'] == 'Southampton') , 'Southampton FC', inplace=True)
stats_df21['Team'].mask((stats_df21['Team'] == 'Watford') , 'Watford FC', inplace=True)

stats_df21['Nation'] = stats_df21['Nation'].str.replace('[a-z]', '')

stats_df21 = stats_df21.rename(columns={'Player': 'Name', 'Team': 'Club', 'Nation': 'Nationality'})
stats_df21['Name'].mask((stats_df21['Name'] == 'Łukasz Fabiański') , 'Łukasz Fabiański', inplace=True)
stats_df21['Name'].mask((stats_df21['Name'] == 'Tomáš Souček') , 'Tomáš Souček', inplace=True)
stats_df21['Name'].mask((stats_df21['Name'] == 'Nikola Vlačić') , 'Nikola Vlačić', inplace=True)
stats_df21['Name'].mask((stats_df21['Name'] == 'Przemysław Pacheta') , 'Przemysław Pacheta', inplace=True)
stats_df21['Name'].mask((stats_df21['Name'] == 'Nemanja Matić') , 'Nemanja Matić', inplace=True)
stats_df21['Name'].mask((stats_df21['Name'] == 'Halil Dervișoğlu') , 'Halil Dervisoglu', inplace=True)
stats_df21['Name'].mask((stats_df21['Name'] == 'İlkay Gündoğan') , 'İlkay Gündoğan', inplace=True)
stats_df21['Name'].mask((stats_df21['Name'] == 'Michał Karbowiak') , 'Michał Karbowiak', inplace=True)
stats_df21['Name'].mask((stats_df21['Name'] == 'Matěj Vydra') , 'Matěj Vydra', inplace=True)
stats_df21['Name'].mask((stats_df21['Name'] == 'Mateo Kovačić') , 'Mateo Kovačić', inplace=True)
```

```

TransferWindowDimTable = pd.DataFrame({
    'TransferDimID': ['1', '2', '3', '4'],
    'Year' : [2020, 2021, 2021, 2022],
    'TransferWindow' : ['Summer', 'Winter', 'Summer', 'Winter'],
    'DeadlineMonth': ['October', 'January', 'August', 'January'],
    'DeadlineDay': ['05', '31', '31', '31']})

```

TransferWindowDimTable

	TransferDimID	Year	TransferWindow	DeadlineMonth	DeadlineDay
0	1	2020	Summer	October	05
1	2	2021	Winter	January	31
2	3	2021	Summer	August	31
3	4	2022	Winter	January	31

import datetime

```

SeasonDimTable = pd.DataFrame({
    'SeasonID': ['1', '2'],
    'StartDate': [datetime.date(2020, 9, 12), datetime.date(2021, 8, 13)],
    'EndDate' : [datetime.date(2021, 5, 23), datetime.date(2022, 5, 22)],
    'SeasonYear': ['2020/2021', '2021/2022'],
    'StartYear': ['2020', '2021'],
    'EndYear': ['2021', '2022'],
    'StartMonth': ['September', 'August'],
    'EndMonth': ['May', 'May'],
    'StartDay': ['12', '13'],
    'EndDay': ['23', '22']})

```

SeasonDimTable

	SeasonID	StartDate	EndDate	SeasonYear	StartYear	EndYear	StartMonth	EndMonth	StartDay	EndDay
0	1	2020-09-12	2021-05-23	2020/2021	2020	2021	September	May	12	23
1	2	2021-08-13	2022-05-22	2021/2022	2021	2022	August	May	13	22

ETL

```

ClubStagingTable = pd.DataFrame({
    'Nationality': ['ENG']*23,
    'League': ['English Premier League']*23})
clubs = stats_df20.Club.unique()
clubs = np.append(clubs, stats_df21.Club.unique())
clubs

```

```

ClubStagingTable['Club'] = list(set(clubs))
ClubStagingTable=ClubStagingTable.sort_values(by="Club")

```

```

ClubDimTable = ClubStagingTable.reset_index(drop=True)
ClubDimTable.reset_index(inplace=True)

```

```

ClubDimTable = ClubDimTable.rename(columns={"index": "ClubID"})
ClubDimTable.ClubID = ClubDimTable.ClubID + 1

```

ClubDimTable

	ClubID	Nationality	League	Club
0	1	ENG	English Premier League	Arsenal FC
1	2	ENG	English Premier League	Aston Villa
2	3	ENG	English Premier League	Brentford FC
3	4	ENG	English Premier League	Brighton & Hove Albion
4	5	ENG	English Premier League	Burnley FC
5	6	ENG	English Premier League	Chelsea FC
6	7	ENG	English Premier League	Crystal Palace
7	8	ENG	English Premier League	Everton FC
8	9	ENG	English Premier League	Fulham FC
9	10	ENG	English Premier League	Leeds United
10	11	ENG	English Premier League	Leicester City
11	12	ENG	English Premier League	Liverpool FC
12	13	ENG	English Premier League	Manchester City
13	14	ENG	English Premier League	Manchester United

```
PlayerA = stats_df20[['Name', 'Nationality']].copy()
PlayerA.reset_index()

PlayerB = stats_df21[['Name', 'Nationality']].copy()
PlayerB.reset_index()

Player_df = pd.merge(PlayerA, PlayerB, how="outer", on=["Name", "Nationality"])

PlayerDistinct_df = Player_df.drop_duplicates()

PlayerDistinct = PlayerDistinct_df.reset_index(drop=True)
PlayerDistinct.reset_index(inplace=True)

PlayerDistinct = PlayerDistinct.rename(columns={"index": "PlayerID"})
PlayerDistinct.PlayerID = PlayerDistinct.PlayerID + 1

PlayerStagingA = stats_df20[['Name', 'Club', 'Nationality', 'Age']].copy()
PlayerStagingA.reset_index()

PlayerStagingB = stats_df21[['Name', 'Club', 'Nationality', 'Age']].copy()
PlayerStagingB.reset_index()

PlayerStaging = pd.merge(PlayerStagingA, PlayerStagingB, how="outer", on=["Name", "Club", "Nationality"])
PlayerStaging
PlayerStaging.reset_index(inplace=True)

PlayerStaging = PlayerStaging.rename(columns={"index": "PlayerDimID"})
PlayerStaging.PlayerDimID = PlayerStaging.PlayerDimID + 1

PlayerStaging['Age'] = PlayerStaging['Age_y'].fillna(PlayerStaging['Age_x'])
PlayerStaging = PlayerStaging.drop(['Age_x', 'Age_y'], axis=1)

PlayerDimTable = pd.merge(PlayerStaging, PlayerDistinct, how="left", on=["Name", "Nationality"])
```

ETL: PLAYER DIMENSION TABLE SCD TYPE2

ETL: PLAYER DIMENSION TABLE SCD TYPE2

```

idx = PlayerDimTable.groupby(['PlayerID'])['Age'].transform(max) == PlayerDimTable['Age']
PlayerDimTable['ActiveClubIndicator'] = idx

PlayerDimTable = pd.merge(PlayerDimTable, ClubDimTable.iloc[:,[0,-1]], how = "left", on = ["Club"])
|
PlayerDimTable.loc[np.isnan(PlayerDimTable.Age), 'ActiveClubIndicator'] = True
PlayerDimTable['RowEffectiveDate'] = pd.to_datetime("today").strftime("%m/%d/%Y")
PlayerDimTable['RowExpiryDate'] = np.where(PlayerDimTable['ActiveClubIndicator']!= True, pd.to_datetime("today").strfti

```

PlayerDimID		Name	Club	Nationality	Age	PlayerID	ActiveClubIndicator	ClubID	RowEffectiveDate	RowExpiryDate
21	22	Billy Gilmour	Chelsea FC	SCO	19.0	22	False	6	12/10/2022	12/10/2022
757	758	Billy Gilmour	Norwich City	SCO	20.0	22	True	16	12/10/2022	9999-12-31
788	789	Willy Caballero	Southampton FC	ARG	39.0	23	True	18	12/10/2022	9999-12-31
22	23	Willy Caballero	Chelsea FC	ARG	38.0	23	False	6	12/10/2022	12/10/2022
23	24	Ruben Loftus-Cheek	Chelsea FC	ENG	25.0	24	True	6	12/10/2022	9999-12-31
457	458	Ruben Loftus-Cheek	Fulham FC	ENG	24.0	24	False	9	12/10/2022	12/10/2022
274	275	Ross Barkley	Aston Villa	ENG	26.0	27	False	2	12/10/2022	12/10/2022
26	27	Ross Barkley	Chelsea FC	ENG	27.0	27	True	6	12/10/2022	9999-12-31
71	72	Axel Tuanzebe	Manchester United	ENG	22.0	72	False	14	12/10/2022	12/10/2022
559	560	Axel Tuanzebe	Aston Villa	ENG	23.0	72	True	2	12/10/2022	9999-12-31
762	763	Ozan Kabak	Norwich City	TUR	21.0	95	True	16	12/10/2022	9999-12-31
94	95	Ozan Kabak	Liverpool FC	TUR	20.0	95	False	12	12/10/2022	12/10/2022
103	104	Takumi Minamino	Liverpool FC	JPN	26.0	104	True	12	12/10/2022	9999-12-31
381	382	Takumi Minamino	Southampton FC	JPN	25.0	104	False	18	12/10/2022	12/10/2022
132	133	Demarai Gray	Leicester City	ENG	24.0	133	False	11	12/10/2022	12/10/2022
659	660	Demarai Gray	Everton FC	ENG	25.0	133	True	8	12/10/2022	9999-12-31
145	146	Jesse Lingard	West Ham United	ENG	27.0	146	False	22	12/10/2022	12/10/2022
732	733	Jesse Lingard	Manchester United	ENG	28.0	146	True	14	12/10/2022	9999-12-31
204	205	Ainsley Maitland-Niles	Arsenal FC	ENG	23.0	205	True	1	12/10/2022	9999-12-31
489	490	Ainsley Maitland-Niles	West Bromwich Albion	ENG	22.0	205	False	21	12/10/2022	12/10/2022

ETL: PLAYER STATS FACT TABLE

```

PlayerStatsStagingTable = stats_df20[['Name', 'Club', 'Nationality', 'Matches', 'Starts', 'Mins', 'Goals', 'Assists']].copy()
PlayerStatsStagingTable['SeasonYear'] = '2020/2021'
PlayerStatsStagingTable= PlayerStatsStagingTable.rename(columns={'Mins': 'Minutes'})

Stats21Staging = stats_df21[['Name', 'Club','Nationality', 'MP', 'Starts', 'Min', 'Gls', 'Ast']].copy()
Stats21Staging[ 'SeasonYear' ] = '2021/2022'
Stats21Staging=Stats21Staging.rename(columns={'Min': 'Minutes', 'MP':'Matches','Min':'Minutes','Gls':'Goals','Ast':'Assists'})

PlayerStatsStagingTable = PlayerStatsStagingTable.append(Stats21Staging)
PlayerStatsFactTable = PlayerStatsStagingTable.reset_index(drop=True)
PlayerStatsFactTable.reset_index(inplace=True)

PlayerStatsFactTable = PlayerStatsFactTable.rename(columns={"index": "PlayerFactID"})
PlayerStatsFactTable.PlayerFactID = PlayerStatsFactTable.PlayerFactID + 1
PlayerStatsFactTable = pd.merge(PlayerStatsFactTable, PlayerDimTable.loc[:,["PlayerDimID", "Name", "Club", "Nationality"]], how="left", left_index=True, right_index=True)
PlayerStatsFactTable = pd.merge(PlayerStatsFactTable, SeasonDimTable.loc[:,["SeasonYear", "SeasonID"]], how="left", left_index=True, right_index=True)

```

	PlayerFactID	Name	Club	Nationality	Matches	Starts	Minutes	Goals	Assists	SeasonYear	PlayerDimID	SeasonID
412	413	Aaron Connolly	Brighton & Hove Albion	IRL	17	9	791.0	2.0	1.0	2020/2021	413	1
670	671	Aaron Connolly	Brighton & Hove Albion	IRL	4	1	156.0	0.0	0.0	2021/2022	413	2
136	137	Aaron Cresswell	West Ham United	ENG	36	36	3170.0	0.0	8.0	2020/2021	137	1
1164	1165	Aaron Cresswell	West Ham United	ENG	31	31	2726.0	2.0	3.0	2021/2022	137	2
697	698	Aaron Lennon	Burnley FC	ENG	28	17	1551.0	2.0	0.0	2021/2022	630	2
534	535	Aaron Ramsdale	Arsenal FC	ENG	34	34	3060.0	0.0	0.0	2021/2022	533	2
505	506	Aaron Ramsdale	Sheffield United	ENG	38	38	3420.0	0.0	0.0	2020/2021	506	1
612	613	Aaron Ramsey	Aston Villa	ENG	0	0	NaN	NaN	NaN	2021/2022	572	2
52	53	Aaron Wan-Bissaka	Manchester United	ENG	34	34	3060.0	2.0	4.0	2020/2021	53	1
960	961	Aaron Wan-Bissaka	Manchester United	ENG	20	20	1793.0	0.0	0.0	2021/2022	53	2
241	242	Abdoulaye Doucouré	Everton FC	MLI	29	29	2443.0	2.0	3.0	2020/2021	242	1
781	782	Abdoulaye Doucouré	Everton FC	MLI	30	29	2537.0	2.0	4.0	2021/2022	242	2
469	470	Aboubakar Kamara	Fulham FC	MTN	11	2	315.0	0.0	0.0	2020/2021	470	1
1050	1051	Abu Kamara	Norwich City	ENG	0	0	NaN	NaN	NaN	2021/2022	779	2
1067	1068	Adam Armstrong	Southampton FC	ENG	23	17	1413.0	2.0	2.0	2021/2022	787	2
832	833	Adam Forshaw	Leeds United	ENG	22	17	1530.0	0.0	1.0	2021/2022	680	2
1038	1039	Adam Idah	Norwich City	IRL	17	6	625.0	1.0	1.0	2021/2022	767	2
407	408	Adam Lallana	Brighton & Hove Albion	ENG	30	16	1596.0	1.0	1.0	2020/2021	408	1
660	661	Adam Lallana	Brighton & Hove Albion	ENG	24	18	1561.0	0.0	1.0	2021/2022	408	2
1133	1134	Adam Masina	Watford FC	MAR	15	13	1077.0	0.0	1.0	2021/2022	825	2

ETL: PLAYER STATS FACT TABLE MEASURES

```
PlayerStatsFactTable['GoalstoGamesRatio'] = PlayerStatsFactTable['Goals'] / PlayerStatsFactTable['Matches']
PlayerStatsFactTable['GoalstoStartsRatio'] = PlayerStatsFactTable['Goals'] / PlayerStatsFactTable['Starts']
```

Name	Club	Nationality	Matches	Starts	Minutes	Goals	Assists	SeasonYear	PlayerDimID	SeasonID	GoalstoGamesRatio	GoalstoStartsRatio
Aaron Connolly	Brighton & Hove Albion	IRL	17	9	791.0	2.0	1.0	2020/2021	413	1	0.117647	0.222222
Aaron Connolly	Brighton & Hove Albion	IRL	4	1	156.0	0.0	0.0	2021/2022	413	2	0.000000	0.000000
Aaron Cresswell	West Ham United	ENG	36	36	3170.0	0.0	8.0	2020/2021	137	1	0.000000	0.000000
Aaron Cresswell	West Ham United	ENG	31	31	2726.0	2.0	3.0	2021/2022	137	2	0.064516	0.064516
Aaron Lennon	Burnley FC	ENG	28	17	1551.0	2.0	0.0	2021/2022	630	2	0.071429	0.117647
Aaron Ramsdale	Arsenal FC	ENG	34	34	3060.0	0.0	0.0	2021/2022	533	2	0.000000	0.000000
Aaron Ramsdale	Sheffield United	ENG	38	38	3420.0	0.0	0.0	2020/2021	506	1	0.000000	0.000000
Aaron Ramsey	Aston Villa	ENG	0	0	NaN	NaN	NaN	2021/2022	572	2	NaN	NaN
Aaron Wan-Bissaka	Manchester United	ENG	34	34	3060.0	2.0	4.0	2020/2021	53	1	0.058824	0.058824
Aaron Wan-Bissaka	Manchester United	ENG	20	20	1793.0	0.0	0.0	2021/2022	53	2	0.000000	0.000000
Abdoulaye Doucouré	Everton FC	MLI	29	29	2443.0	2.0	3.0	2020/2021	242	1	0.068966	0.068966
Abdoulaye Doucouré	Everton FC	MLI	30	29	2537.0	2.0	4.0	2021/2022	242	2	0.066667	0.068966
Aboubakar Kamara	Fulham FC	MTN	11	2	315.0	0.0	0.0	2020/2021	470	1	0.000000	0.000000
Abu Kamara	Norwich City	ENG	0	0	NaN	NaN	NaN	2021/2022	779	2	NaN	NaN
Adam Armstrong	Southampton FC	ENG	23	17	1413.0	2.0	2.0	2021/2022	787	2	0.086957	0.117647
Adam	Leeds United	ENG	22	17	1522.0	2.0	1.0	2021/2022	206	2	0.000000	0.000000

ETL: TRANSFERS FACT TABLE

```

TransferStagingTable = transfer_df[['club_name', 'player_name', 'fee_cleaned', 'year', 'transfer_period', 'Transfer_Type']

TransferStagingTable=TransferStagingTable.rename(columns={'club_name': 'Club', 'player_name':'Name','fee_cleaned':'TransferFee','year':'Year','transfer_period':'TransferWindow','Transfer_Type':'TransferType'})

TransferFactTable = TransferStagingTable.reset_index(drop=True)
TransferFactTable.reset_index(inplace=True)

TransferFactTable = TransferFactTable.rename(columns={"index": "TransferFactID"})
TransferFactTable.TransferFactID = TransferFactTable.TransferFactID + 1

TransferFactTable = pd.merge(TransferFactTable, TransferWindowDimTable.loc[:,["Year", "TransferWindow", "TransferDimID"]], how="left", on="Year")
TransferFactTable = pd.merge(TransferFactTable, ClubDimTable.loc[:,["Club", "ClubID"]], how="left", on="Club")
TransferFactTable = pd.merge(TransferFactTable, PlayerDimTable.loc[:,["Name", "Club", "PlayerDimID"]], how="left", on="Name")

TransferFactTable.drop(TransferFactTable.loc[np.isnan(TransferFactTable.PlayerDimID) == True].index, inplace=True)
TransferFactTable.sort_values(by=[ 'Name']).head(20)

```

TransferFactID	Club	Name	TransferFee	Year	TransferWindow	TransferType	TransferDimID	ClubID	PlayerDimID	
192	193	Burnley FC	Aaron Lennon	0.00	2021	Summer	Permanent Transfer	3	5	630.0
163	164	Arsenal FC	Aaron Ramsdale	28.00	2021	Summer	Permanent Transfer	3	1	533.0
92	93	Sheffield United	Aaron Ramsdale	20.50	2020	Summer	Permanent Transfer	1	17	506.0
40	41	Everton FC	Abdoulaye Doucouré	22.00	2020	Summer	Permanent Transfer	1	8	242.0
248	249	Southampton FC	Adam Armstrong	17.70	2021	Summer	Permanent Transfer	3	18	787.0
19	20	Brighton & Hove Albion	Adam Lallana	0.00	2020	Summer	Permanent Transfer	1	4	408.0
222	223	Leicester City	Ademola Lookman	NaN	2021	Summer	Loan Transfer	3	11	696.0
52	53	Fulham FC	Ademola Lookman	2.00	2020	Summer	Permanent Transfer	1	9	450.0
154	155	West Bromwich Albion	Ainsley Maitland-Niles	NaN	2021	Winter	Loan Transfer	2	21	490.0
165	166	Arsenal FC	Albert Sambi Lokonga	17.50	2021	Summer	Permanent Transfer	3	1	537.0
274	275	West Ham United	Alex Král	5.20	2021	Summer	Permanent Transfer	3	22	857.0
79	80	Manchester United	Alex Telles	15.00	2020	Summer	Permanent Transfer	1	14	69.0
39	40	Everton FC	Allan	23.60	2020	Summer	Permanent Transfer	1	8	245.0
276	277	West Ham United	Alphonse Areola	2.00	2021	Summer	Permanent Transfer	3	22	853.0
56	57	Fulham FC	Alphonse Areola	NaN	2020	Summer	Loan Transfer	1	9	448.0
147	148	Manchester United	Amad Diallo	21.30	2021	Winter	Permanent Transfer	2	14	74.0
15	16	Brighton & Hove Albion	Andi Zeqiri	4.00	2020	Summer	Permanent Transfer	1	4	419.0
209	210	Everton FC	Andros Townsend	0.00	2021	Summer	Permanent Transfer	3	8	661.0
208	209	Everton FC	Andy Lonergan	0.00	2021	Summer	Permanent Transfer	3	8	673.0
242	243	Norwich City	Angus Gunn	5.85	2021	Summer	Permanent Transfer	3	16	765.0

```
import psycopg2 as pg
import sqlite3
from sqlalchemy import create_engine


conn = pg.connect(database="PlayerStatTransfers",
                  user="postgres",
                  password="password",
                  host = "127.0.0.1",
                  port = "5432")
cursor = conn.cursor()
for index, row in PlayerDimTable.iterrows():
    cursor.execute("INSERT INTO PlayerDimTable (PlayerDimID, ClubID, PlayerID, Name, Club, Nationality, Age, RowEffect)
                    (str(row.PlayerDimID), str(row.ClubID), str(row.PlayerID), str(row.Name), str(row.Club), str(row.Nationality), str(row.Age), str(row.RowEffect))")
conn.commit()
cursor.close()

import psycopg2 as pg
import sqlite3
from sqlalchemy import create_engine


conn = pg.connect(database="PlayerStatTransfers",
                  user="postgres",
                  password="password",
                  host = "127.0.0.1",
                  port = "5432")
cursor = conn.cursor()
for index, row in PlayerStatsFactTable.iterrows():
    cursor.execute("INSERT INTO PlayerStatsFactTable (PlayerFactID, PlayerDimID, SeasonID, Name, Club, Nationality, Manager, RowEffect)
                    (str(row.PlayerFactID), str(row.PlayerDimID), str(row.SeasonID), str(row.Name), str(row.Club), str(row.Nationality), str(row.Manager), str(row.RowEffect))")
conn.commit()
cursor.close()

import psycopg2 as pg
import sqlite3
from sqlalchemy import create_engine


conn = pg.connect(database="PlayerStatTransfers",
                  user="postgres",
                  password="password",
                  host = "127.0.0.1",
                  port = "5432")
cursor = conn.cursor()
for index, row in TransferFactTable.iterrows():
    cursor.execute("INSERT INTO TransferFactTable (TransferFactID, TransferDimID, ClubID, PlayerDimID, Club, Name, TransferType, RowEffect)
                    (str(row.TransferFactID), str(row.TransferDimID), str(row.ClubID), int(row.PlayerDimID), str(row.Club), str(row.Name), str(row.TransferType), str(row.RowEffect))")
conn.commit()
cursor.close()
```

```

select * from SeasonDimTable
select * from TransferWindowDimTable
select * from ClubDimTable
select * from PlayerDimTable

```

output Messages Notifications

playerdimid [PK] integer	clubid integer	playerid integer	name character varying (255)	club character varying (255)	nationality character varying (255)	age character varying (255)	roweffective date date	rowexpirydate date	activeclubindicator boolean
1	6	1	Mason Mount	Chelsea FC	ENG	22.0	2022-12-09	9999-12-31	true
2	6	2	Edouard Mendy	Chelsea FC	SEN	29.0	2022-12-09	9999-12-31	true
3	6	3	Timo Werner	Chelsea FC	GER	25.0	2022-12-09	9999-12-31	true
4	6	4	Ben Chilwell	Chelsea FC	ENG	24.0	2022-12-09	9999-12-31	true
5	6	5	Reece James	Chelsea FC	ENG	21.0	2022-12-09	9999-12-31	true
6	6	6	César Azpilicueta	Chelsea FC	ESP	31.0	2022-12-09	9999-12-31	true
7	6	7	N'Golo Kanté	Chelsea FC	FRA	30.0	2022-12-09	9999-12-31	true
8	6	8	Jorginho	Chelsea FC	ITA	29.0	2022-12-09	9999-12-31	true
9	6	9	Thiago Silva	Chelsea FC	BRA	36.0	2022-12-09	9999-12-31	true
10	6	10	Kurt Zouma	Chelsea FC	FRA	26.0	2022-12-09	9999-12-31	true
11	6	11	Mateo Kovacic	Chelsea FC	CRO	27.0	2022-12-09	9999-12-31	true
12	6	12	Antonio Rüdiger	Chelsea FC	GER	28.0	2022-12-09	9999-12-31	true
13	6	13	Christian Pulisic	Chelsea FC	USA	22.0	2022-12-09	9999-12-31	true

ETL

```
select * from PlayerStatsFactTable
```

output Messages Notifications

erfactid integer	playerdimid integer	seasonid integer	name character varying (255)	club character varying (255)	nationality character varying (255)	matches double precision	starts double precision	minutes double precision	goals double precision
1	1	1	Mason Mount	Chelsea FC	ENG	36	32	2890	6
2	2	1	Edouard Mendy	Chelsea FC	SEN	31	31	2745	0
3	3	1	Timo Werner	Chelsea FC	GER	35	29	2602	6
4	4	1	Ben Chilwell	Chelsea FC	ENG	27	27	2286	3
5	5	1	Reece James	Chelsea FC	ENG	32	25	2373	1
6	6	1	César Azpilicueta	Chelsea FC	ESP	26	24	2188	1
7	7	1	N'Golo Kanté	Chelsea FC	FRA	30	24	2146	0
8	8	1	Jorginho	Chelsea FC	ITA	28	23	2010	7
9	9	1	Thiago Silva	Chelsea FC	BRA	23	23	1935	2
10	10	1	Kurt Zouma	Chelsea FC	FRA	24	22	2029	5
11	11	1	Mateo Kovačić	Chelsea FC	CRO	27	21	1815	0
12	12	1	Antonio Rüdiger	Chelsea FC	GER	19	19	1710	1
13	13	1	Christian Pulisic	Chelsea FC	USA	27	18	1738	4

ETL

```
select * from TransferFactTable
```

Output Messages Notifications

transferfactid [PK] integer	transferdimid integer	clubid integer	playerdimid integer	club character varying (255)	name character varying (255)	transferfee double precision	year integer	transferwindow character varying (255)	transfertype character varying (255)
1	1	1	193	Arsenal FC	Thomas Partey	50	2020	Summer	Permanent Transfer
2	1	1	191	Arsenal FC	Gabriel Magalhães	26	2020	Summer	Permanent Transfer
3	1	1	200	Arsenal FC	Pablo Marí	6	2020	Summer	Permanent Transfer
4	1	1	212	Arsenal FC	Rúnar Alex Rúnarsson	2	2020	Summer	Permanent Transfer
5	1	1	203	Arsenal FC	Cédric Soares	0	2020	Summer	Permanent Transfer
6	1	1	199	Arsenal FC	Willian	0	2020	Summer	Permanent Transfer
7	1	1	194	Arsenal FC	Emile Smith Rowe	NaN	2020	Summer	Academy Call Up
8	1	2	268	Aston Villa	Ollie Watkins	34	2020	Summer	Permanent Transfer
9	1	2	272	Aston Villa	Bertrand Traoré	18.4	2020	Summer	Permanent Transfer
10	1	2	265	Aston Villa	Emiliano Martínez	17.4	2020	Summer	Permanent Transfer
11	1	2	273	Aston Villa	Matty Cash	15.75	2020	Summer	Permanent Transfer
12	1	2	275	Aston Villa	Ross Barkley	NaN	2020	Summer	Loan Transfer
13	1	2	281	Aston Villa	Jacob Ramsey	NaN	2020	Summer	Academy Call Up

ETL

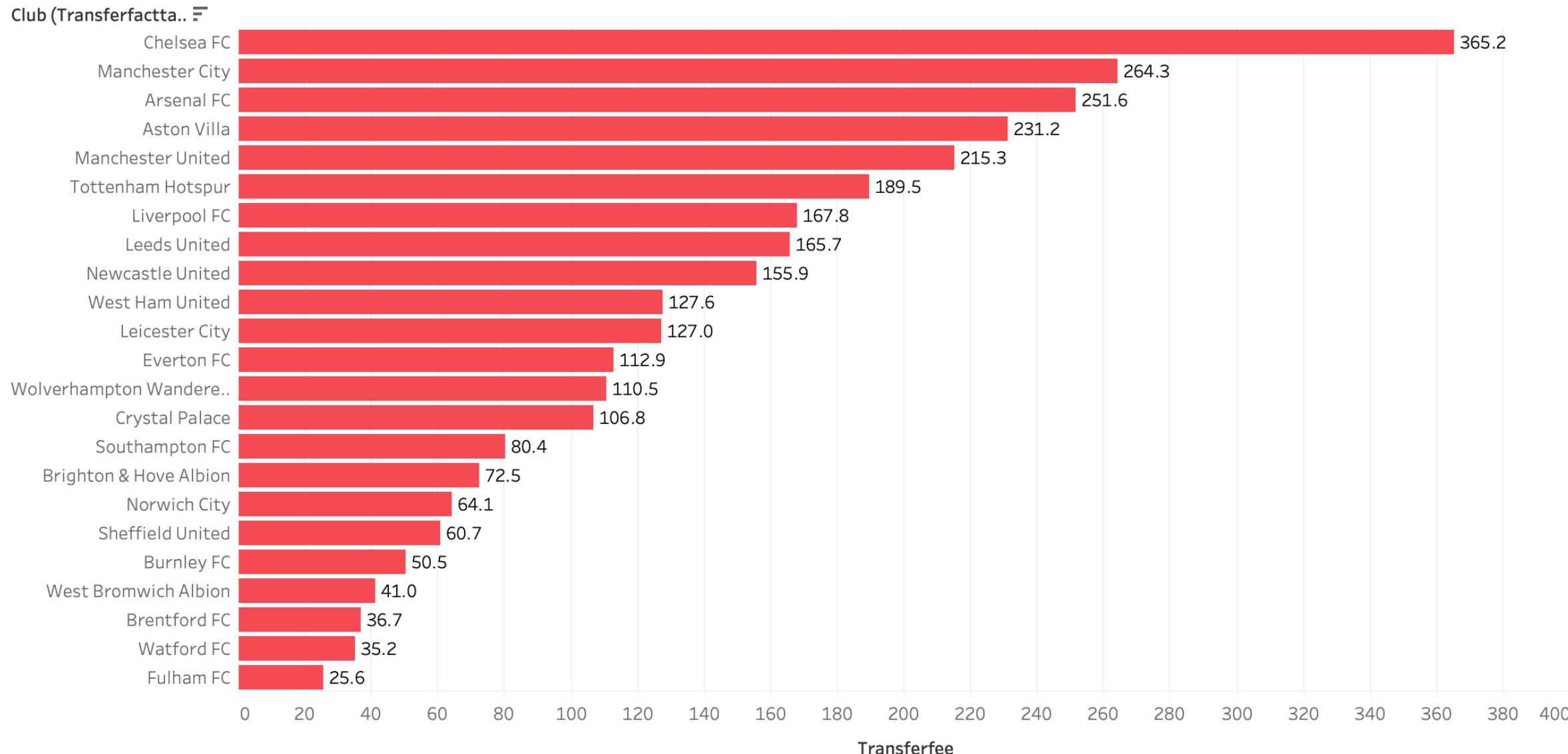
```

SELECT PlayerName, PlayerClub, Season, PlayerGoals, PlayerAssists, TotalGoalsOverall, TotalAssistsOverall, transferfacttable.transferfee,
DENSE_RANK() OVER (ORDER BY TotalGoalsOverall desc) AS TopScorersTransferredPlayersOverall
FROM (
    SELECT PlayerDimTable.playerdimid, PlayerDimTable.name AS PlayerName, ClubDimTable.club AS PlayerClub, seazondimtable.seasonyear AS Season,
    PlayerStatsFactTable.goals AS PlayerGoals, SUM(PlayerStatsFactTable.goals) OVER (PARTITION BY PlayerDimTable.playerdimid) TotalGoalsOverall,
    PlayerStatsFactTable.assists AS PlayerAssists, SUM(PlayerStatsFactTable.assists) OVER (PARTITION BY PlayerDimTable.playerdimid) TotalAssistsOverall
    FROM PlayerDimTable
    JOIN PlayerStatsFactTable ON PlayerStatsFactTable.playerdimid = playerdimtable.playerdimid
    JOIN transferfacttable ON transferfacttable.playerdimid = playerdimtable.playerdimid
    JOIN seazondimtable seazondimtable ON seazondimtable.seasonid = PlayerStatsFactTable.seasonid
    JOIN clubdimtable ON clubdimtable.clubid = playerdimtable.clubid
    WHERE PlayerStatsFactTable.goals != 'NaN'
    GROUP BY PlayerDimTable.name, PlayerDimTable.playerdimid, ClubDimTable.club, seazondimtable.seasonyear, PlayerStatsFactTable.goals, PlayerStatsFactTable.assists
    ORDER BY PlayerDimTable.name, SUM(PlayerStatsFactTable.goals) DESC
    JOIN transferfacttable ON transferfacttable.playerdimid = S.playerdimid
)

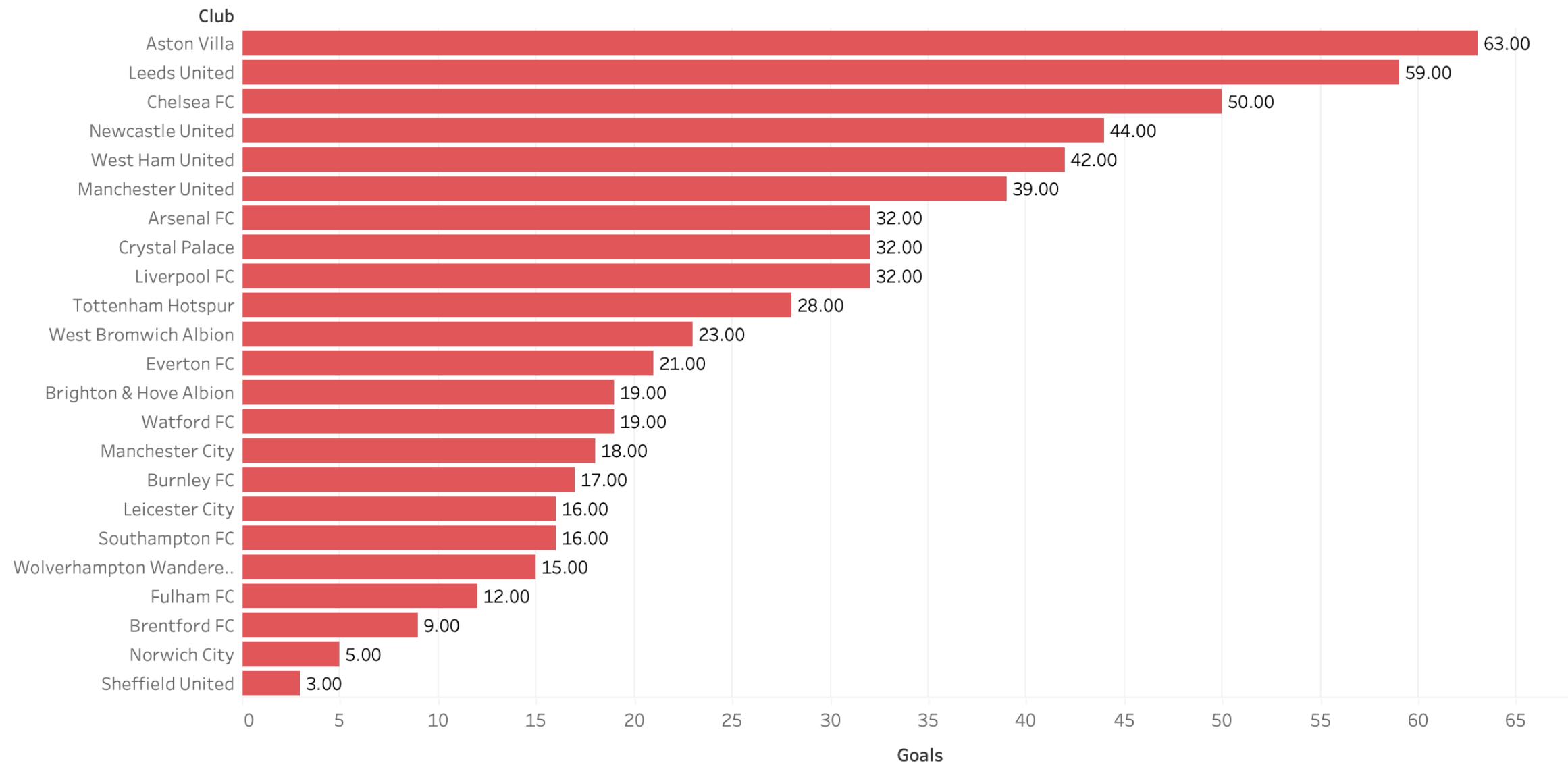
```

	playername character varying (255)	playerclub character varying (255)	season character varying (255)	playergoals double precision	playerassists double precision	totalgoalsoverall double precision	totalassistsoverall double precision	transferfee double precision	topscorertransferredplayersoverall bigint
1	Ollie Watkins	Aston Villa	2021/2022	11	2	25	7	34	1
2	Ollie Watkins	Aston Villa	2020/2021	14	5	25	7	34	1
3	Diogo Jota	Liverpool FC	2020/2021	9	0	24	4	44.7	2
4	Diogo Jota	Liverpool FC	2021/2022	15	4	24	4	44.7	2
5	Callum Wilson	Newcastle United	2021/2022	8	0	20	5	22.25	3
6	Callum Wilson	Newcastle United	2020/2021	12	5	20	5	22.25	3
7	Cristiano Ronaldo	Manchester United	2021/2022	18	3	18	3	15	4
8	Raphinha	Leeds United	2021/2022	11	3	17	12	18.6	5
9	Raphinha	Leeds United	2020/2021	6	9	17	12	18.6	5
10	Jack Harrison	Leeds United	2020/2021	8	8	16	9	12.8	6
11	Jack Harrison	Leeds United	2021/2022	8	1	16	9	12.8	6
12	Tomáš Souček	West Ham United	2020/2021	10	1	15	2	16.2	7
13	Tomáš Souček	West Ham United	2021/2022	5	1	15	2	16.2	7
14	Rodrigo	Leeds United	2020/2021	7	2	13	3	30	8
15	Rodrigo	Leeds United	2021/2022	6	1	13	3	30	8
16	Emile Smith Rowe	Arsenal FC	2021/2022	10	2	12	6	NaN	9
17	Edinson Cavani	Manchester United	2020/2021	10	3	12	4	0	9
18	Kai Havertz	Chelsea FC	2021/2022	8	3	12	6	80	9
19	Kai Havertz	Chelsea FC	2020/2021	4	3	12	6	80	9
20	Danny Welbeck	Brighton & Hove Albion	2021/2022	6	2	12	3	0	9
21	Danny Welbeck	Brighton & Hove Albion	2020/2021	6	1	12	3	0	9
22	Edinson Cavani	Manchester United	2021/2022	2	1	12	4	0	9
23	Emile Smith Rowe	Arsenal FC	2020/2021	2	4	12	6	NaN	9
24	Matheus Pereira	West Bromwich Albion	2020/2021	11	6	11	6	8.25	10
25	Gareth Bale	Tottenham Hotspur	2020/2021	11	2	11	2	NaN	10
26	Joe Willock	Newcastle United	2021/2022	2	0	10	0	29.4	11

Spending of Clubs



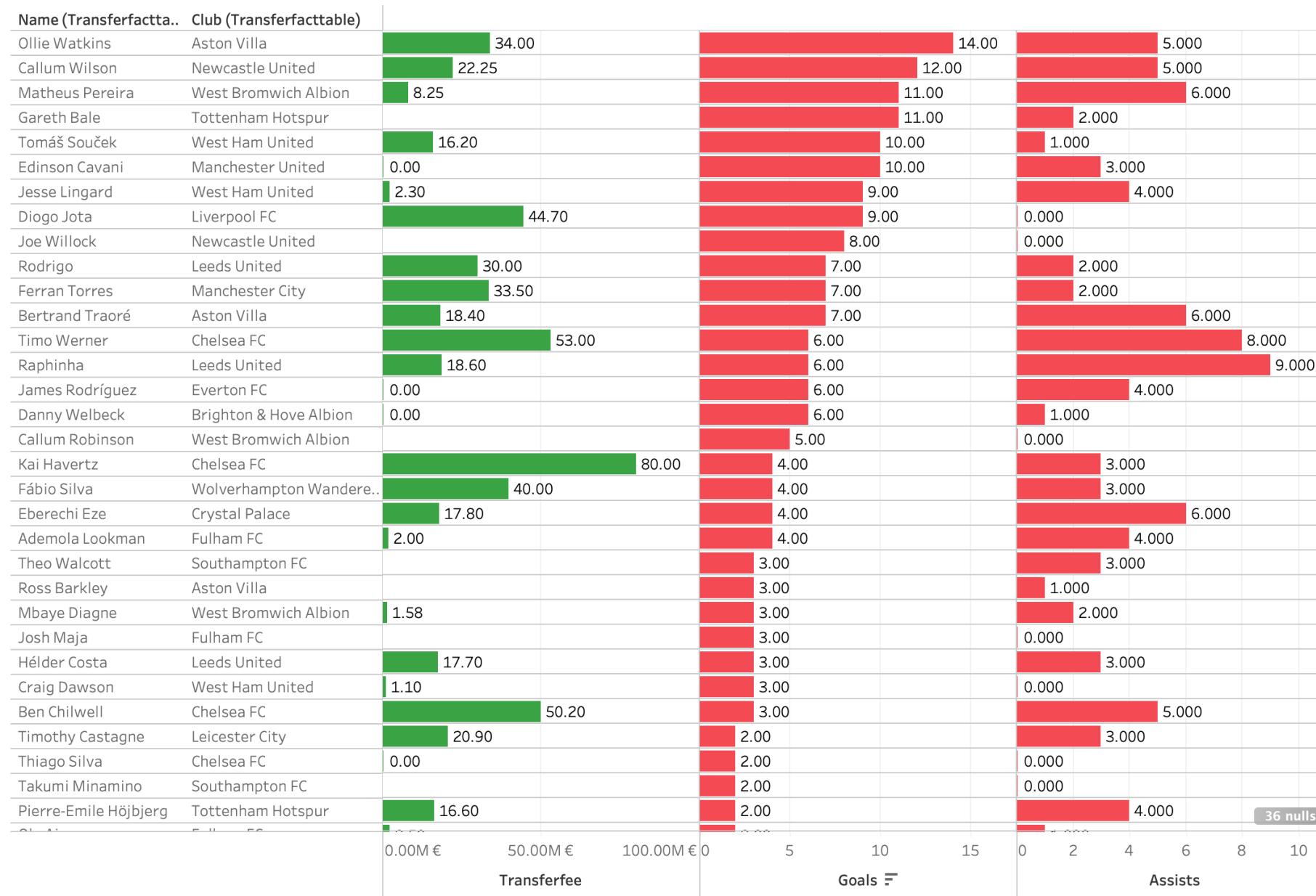
CLUBS SPENDING FOR 2020/2021 & 2021/2022



Sum of Goals for each Club. The data is filtered on Calculation1, which keeps Show.

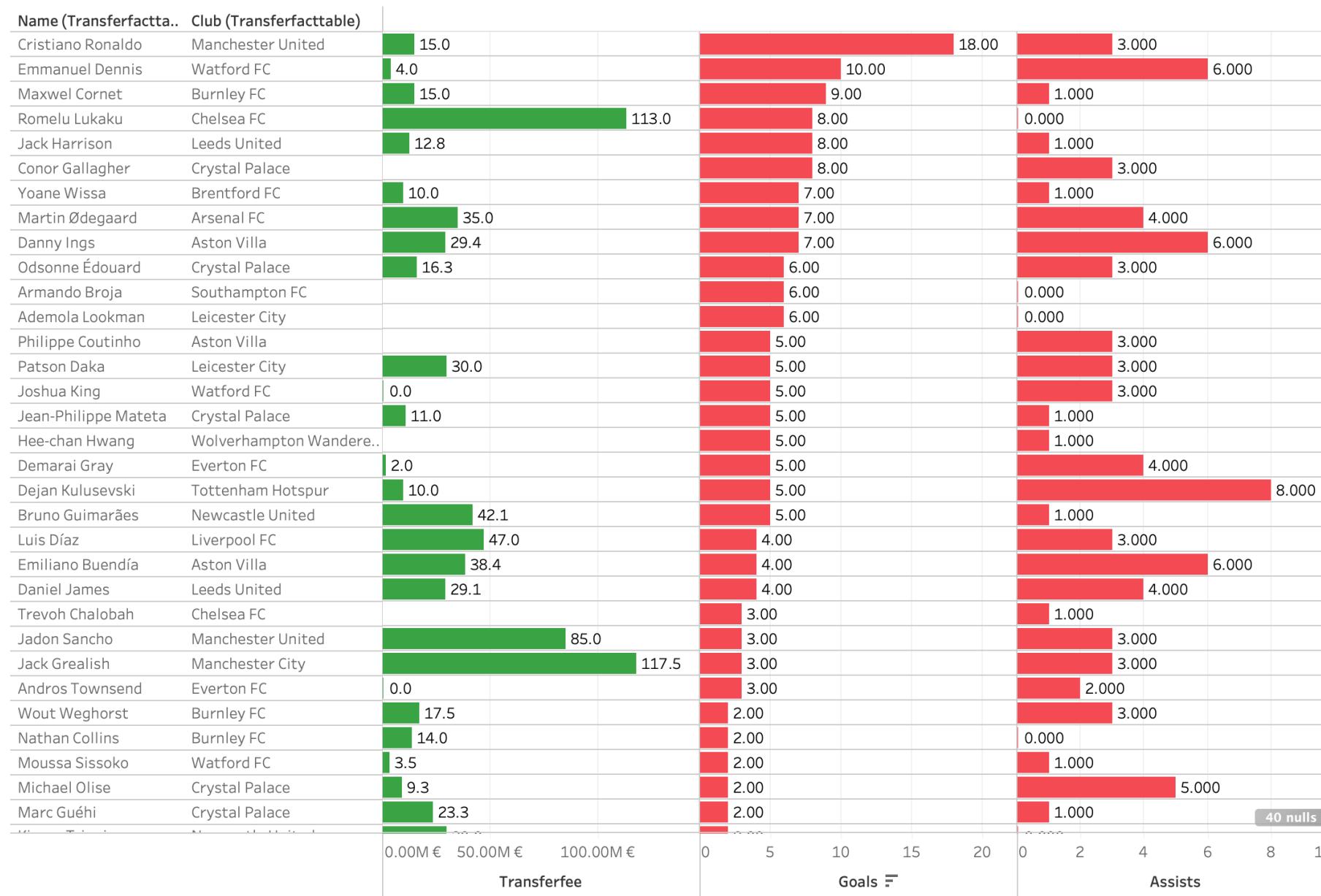
GOALS SCORED BY TRANSFER INS 2020/2021 & 2021/2022

Top Goal Contributions of Transferred Players of 2020/2021



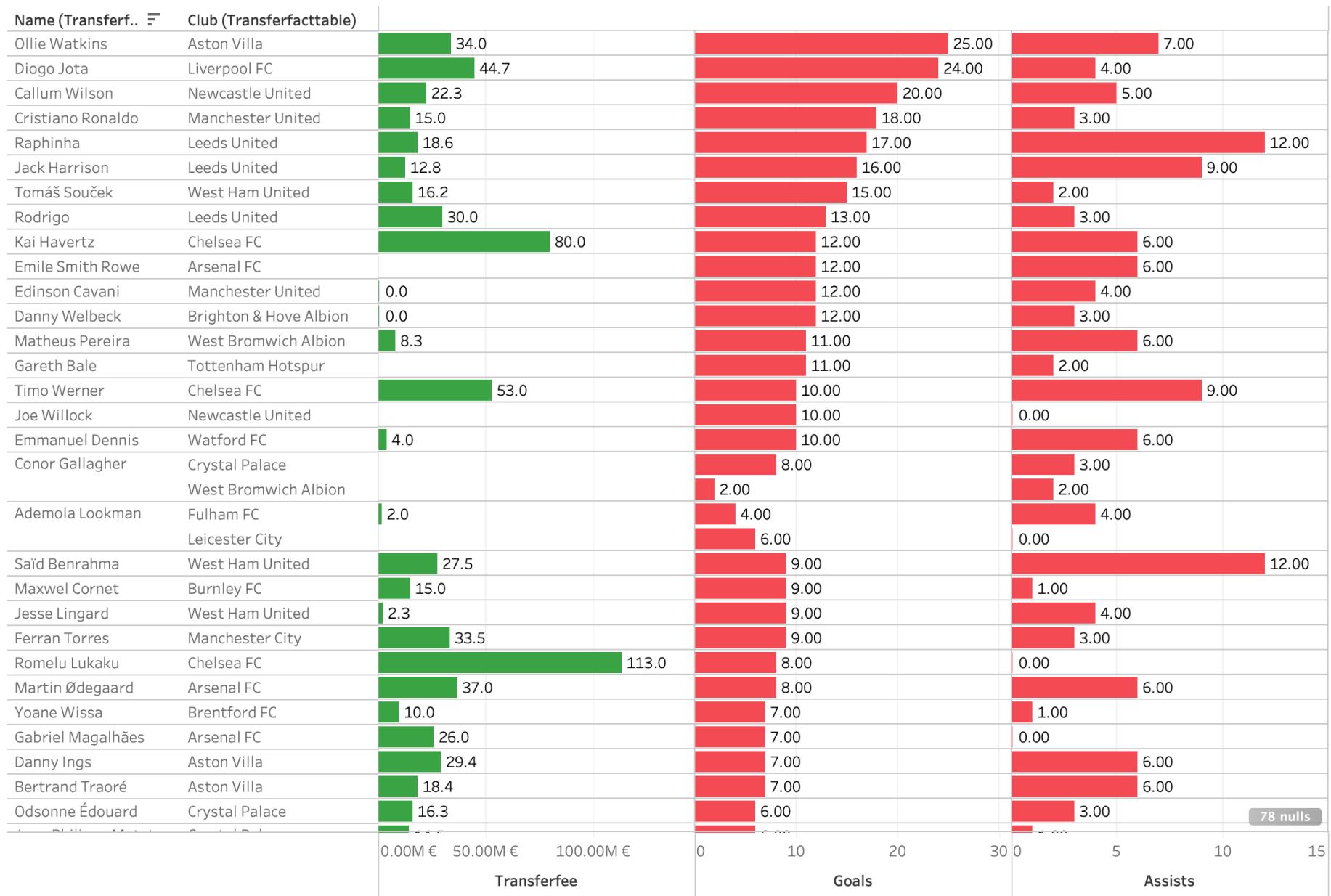
VISUALIZATION

Top Goal Contributions of Transferred Players of 2021/2022



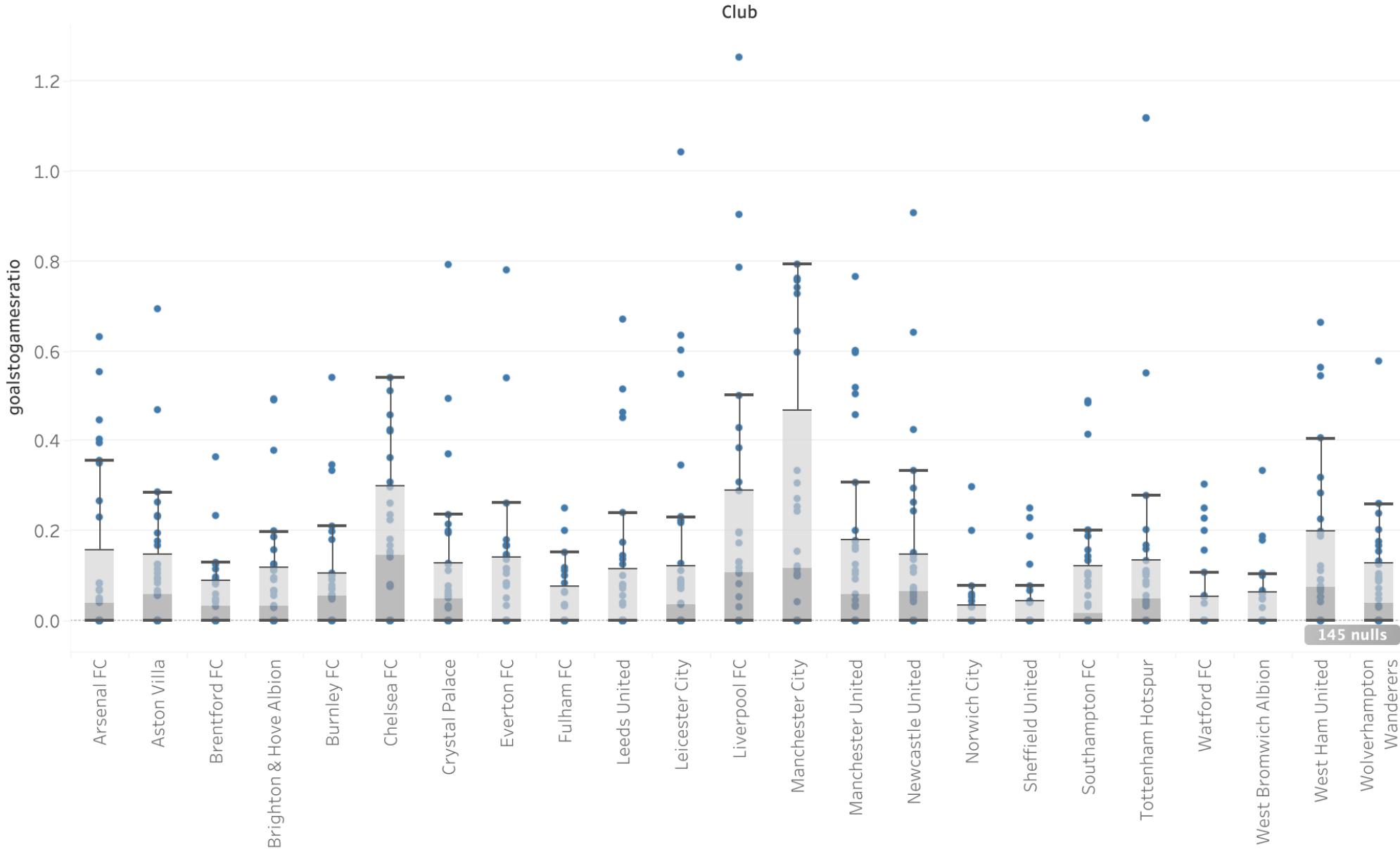
VISUALIZATION

Top Goal Contributions of Transferred Players of 2020/2021 and 2021/2022



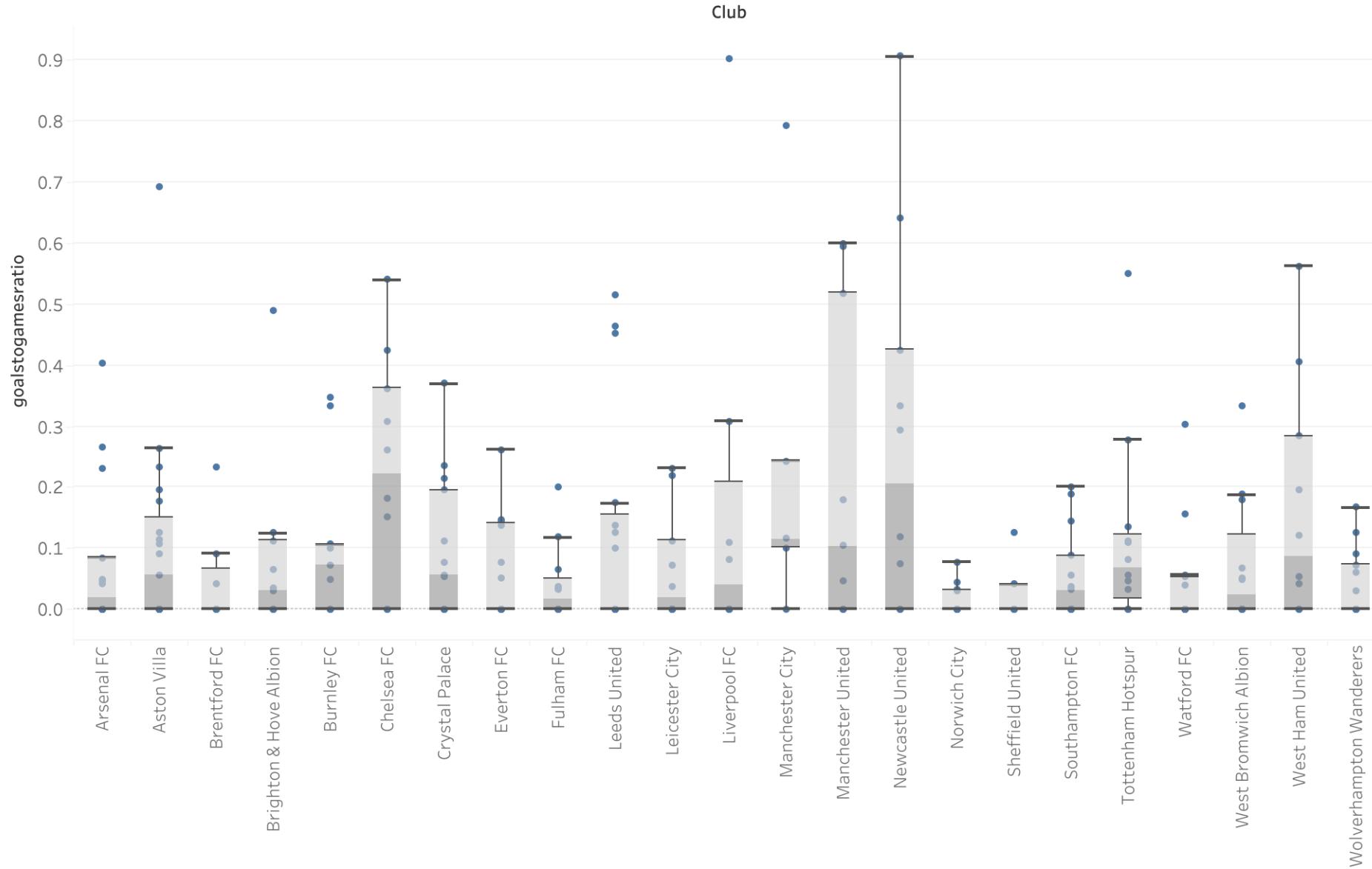
VISUALIZATION

GoalstoGameRatio



VISUALIZATION

GoalstoGameRatio Transferred Players

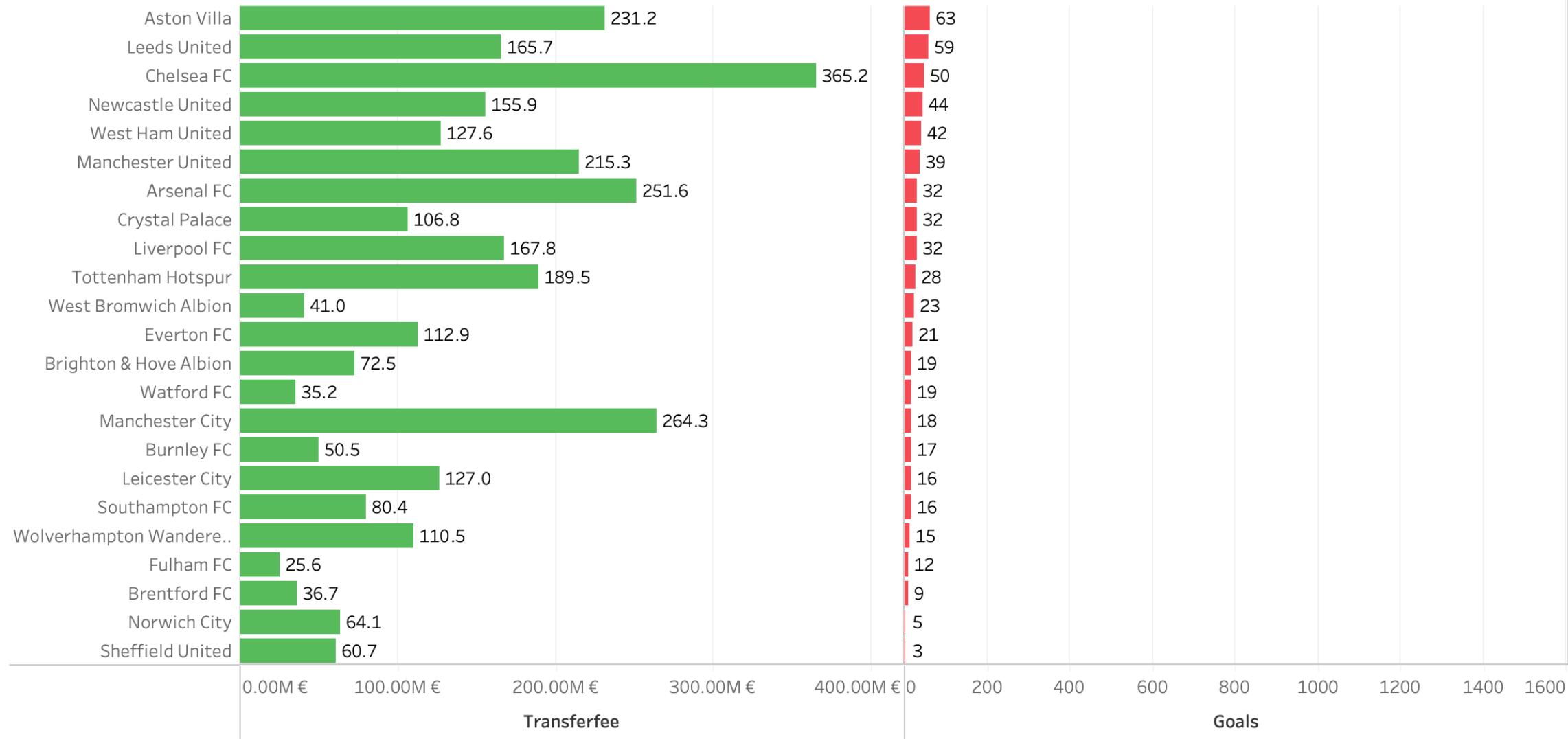


Sum of goalstogamesratio for each Club. Details are shown for Name. The data is filtered on Calculation1, which keeps Show.

VISUALIZATION

Club Spending

Club (Transferfactta..)



VISUALIZATION