

# fifa2

July 16, 2024

```
[1]: import pandas as pd
import numpy as np
import datetime
```

```
[8]: d1_matches=pd.read_csv(r"C:\Users\thala\Downloads\FIFA WC data\\WorldCupMatches.
↪csv")
d2_players=pd.read_csv(r"C:\Users\thala\Downloads\FIFA WC data\WorldCupPlayers.
↪csv")
d3_cups=pd.read_csv(r"C:\Users\thala\Downloads\FIFA WC data\WorldCups.csv")
```

```
[9]: d1_matches.head()
```

```
[9]:
```

	Year	Datetime	Stage	Stadium	City \
0	1930.0	13 Jul 1930 - 15:00	Group 1	Pocitos	Montevideo
1	1930.0	13 Jul 1930 - 15:00	Group 4	Parque Central	Montevideo
2	1930.0	14 Jul 1930 - 12:45	Group 2	Parque Central	Montevideo
3	1930.0	14 Jul 1930 - 14:50	Group 3	Pocitos	Montevideo
4	1930.0	15 Jul 1930 - 16:00	Group 1	Parque Central	Montevideo

  

	Home Team Name	Home Team Goals	Away Team Goals	Away Team Name \
0	France	4.0	1.0	Mexico
1	USA	3.0	0.0	Belgium
2	Yugoslavia	2.0	1.0	Brazil
3	Romania	3.0	1.0	Peru
4	Argentina	1.0	0.0	France

  

	Win conditions	Attendance	Half-time Home Goals	Half-time Away Goals \
0		4444.0	3.0	0.0
1		18346.0	2.0	0.0
2		24059.0	2.0	0.0
3		2549.0	1.0	0.0
4		23409.0	0.0	0.0

  

	Referee	Assistant 1 \
0	LOMBARDI Domingo (URU)	CRISTOPHE Henry (BEL)
1	MACIAS Jose (ARG)	MATEUCCI Francisco (URU)
2	TEJADA Anibal (URU)	VALLARINO Ricardo (URU)
3	WARNKEN Alberto (CHI)	LANGENUS Jean (BEL)

4 REGO Gilberto (BRA) SAUCEDO Ulises (BOL)

	Assistant 2	RoundID	MatchID	Home Team	Initials \
0	REGO Gilberto (BRA)	201.0	1096.0		FRA
1	WARNKEN Alberto (CHI)	201.0	1090.0		USA
2	BALWAY Thomas (FRA)	201.0	1093.0		YUG
3	MATEUCCI Francisco (URU)	201.0	1098.0		ROU
4	RADULESCU Constantin (ROU)	201.0	1085.0		ARG

	Away Team Initials
0	MEX
1	BEL
2	BRA
3	PER
4	FRA

```
[10]: d2_players.head()
```

	RoundID	MatchID	Team	Initials	Coach Name	Line-up	Shirt Number \
0	201	1096	FRA	CAUDRON Raoul (FRA)	S	0	
1	201	1096	MEX	LUQUE Juan (MEX)	S	0	
2	201	1096	FRA	CAUDRON Raoul (FRA)	S	0	
3	201	1096	MEX	LUQUE Juan (MEX)	S	0	
4	201	1096	FRA	CAUDRON Raoul (FRA)	S	0	

	Player Name	Position	Event
0	Alex THEPOT	GK	NaN
1	Oscar BONFIGLIO	GK	NaN
2	Marcel LANGILLER	NaN	G40'
3	Juan CARRENO	NaN	G70'
4	Ernest LIBERATI	NaN	NaN

```
[11]: d3_cups.head()
```

	Year	Country	Winner	Runners-Up	Third	Fourth \
0	1930	Uruguay	Uruguay	Argentina	USA	Yugoslavia
1	1934	Italy	Italy	Czechoslovakia	Germany	Austria
2	1938	France	Italy	Hungary	Brazil	Sweden
3	1950	Brazil	Uruguay	Brazil	Sweden	Spain
4	1954	Switzerland	Germany FR	Hungary	Austria	Uruguay

	GoalsScored	QualifiedTeams	MatchesPlayed	Attendance
0	70	13	18	590.549
1	70	16	17	363.000
2	84	15	18	375.700
3	88	13	22	1.045.246
4	140	16	26	768.607

```
[12]: d1_matches = d1_matches.replace('Germany FR', 'Germany')
      d2_players= d2_players.replace('Germany FR', 'Germany')
      d3_cups = d3_cups.replace('Germany FR', 'Germany')
```

```
[13]: d1_matches.describe()
```

```
[13]:
```

	Year	Home Team Goals	Away Team Goals	Attendance \
count	852.000000	852.000000	852.000000	850.000000
mean	1985.089202	1.811033	1.022300	45164.800000
std	22.448825	1.610255	1.087573	23485.249247
min	1930.000000	0.000000	0.000000	2000.000000
25%	1970.000000	1.000000	0.000000	30000.000000
50%	1990.000000	2.000000	1.000000	41579.500000
75%	2002.000000	3.000000	2.000000	61374.500000
max	2014.000000	10.000000	7.000000	173850.000000

  

	Half-time Home Goals	Half-time Away Goals	RoundID	MatchID
count	852.000000	852.000000	8.520000e+02	8.520000e+02
mean	0.708920	0.428404	1.066177e+07	6.134687e+07
std	0.937414	0.691252	2.729613e+07	1.110572e+08
min	0.000000	0.000000	2.010000e+02	2.500000e+01
25%	0.000000	0.000000	2.620000e+02	1.188750e+03
50%	0.000000	0.000000	3.370000e+02	2.191000e+03
75%	1.000000	1.000000	2.497220e+05	4.395006e+07
max	6.000000	5.000000	9.741060e+07	3.001865e+08

```
[14]: d2_players.describe()
```

```
[14]:
```

	RoundID	MatchID	Shirt Number
count	3.778400e+04	3.778400e+04	37784.000000
mean	1.105647e+07	6.362233e+07	10.726022
std	2.770144e+07	1.123916e+08	6.960138
min	2.010000e+02	2.500000e+01	0.000000
25%	2.630000e+02	1.199000e+03	5.000000
50%	3.370000e+02	2.216000e+03	11.000000
75%	2.559310e+05	9.741000e+07	17.000000
max	9.741060e+07	3.001865e+08	23.000000

```
[15]: d3_cups.describe()
```

```
[15]:
```

	Year	GoalsScored	QualifiedTeams	MatchesPlayed
count	20.000000	20.000000	20.000000	20.000000
mean	1974.800000	118.950000	21.250000	41.800000
std	25.582889	32.972836	7.268352	17.218717
min	1930.000000	70.000000	13.000000	17.000000
25%	1957.000000	89.000000	16.000000	30.500000
50%	1976.000000	120.500000	16.000000	38.000000

75%	1995.000000	145.250000	26.000000	55.000000
max	2014.000000	171.000000	32.000000	64.000000

```
[17]: d1_matches.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4572 entries, 0 to 4571
Data columns (total 20 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Year                  852 non-null   float64
1   Datetime              852 non-null   object
2   Stage                 852 non-null   object
3   Stadium               852 non-null   object
4   City                  852 non-null   object
5   Home Team Name        852 non-null   object
6   Home Team Goals       852 non-null   float64
7   Away Team Goals       852 non-null   float64
8   Away Team Name        852 non-null   object
9   Win conditions        852 non-null   object
10  Attendance            850 non-null   float64
11  Half-time Home Goals  852 non-null   float64
12  Half-time Away Goals  852 non-null   float64
13  Referee               852 non-null   object
14  Assistant 1           852 non-null   object
15  Assistant 2           852 non-null   object
16  RoundID               852 non-null   float64
17  MatchID               852 non-null   float64
18  Home Team Initials    852 non-null   object
19  Away Team Initials    852 non-null   object
dtypes: float64(8), object(12)
memory usage: 714.5+ KB
```

```
[22]: d1_matches['Datetime'] = pd.to_datetime(d1_matches['Datetime'],errors='coerce')
d1_matches.head()
```

```
[22]:
```

	Year	Datetime	Stage	Stadium	City \
0	1930.0	1930-07-13 15:00:00	Group 1	Pocitos	Montevideo
1	1930.0	1930-07-13 15:00:00	Group 4	Parque Central	Montevideo
2	1930.0	1930-07-14 12:45:00	Group 2	Parque Central	Montevideo
3	1930.0	1930-07-14 14:50:00	Group 3	Pocitos	Montevideo
4	1930.0	1930-07-15 16:00:00	Group 1	Parque Central	Montevideo

  

	Home Team Name	Home Team Goals	Away Team Goals	Away Team Name \
0	France	4.0	1.0	Mexico
1	USA	3.0	0.0	Belgium
2	Yugoslavia	2.0	1.0	Brazil
3	Romania	3.0	1.0	Peru

4	Argentina	1.0	0.0	France
---	-----------	-----	-----	--------

  

	Win conditions	Attendance	Half-time Home Goals	Half-time Away Goals	\
0		4444.0	3.0	0.0	
1		18346.0	2.0	0.0	
2		24059.0	2.0	0.0	
3		2549.0	1.0	0.0	
4		23409.0	0.0	0.0	

	Referee	Assistant 1	\
0	LOMBARDI Domingo (URU)	CRISTOPHE Henry (BEL)	
1	MACIAS Jose (ARG)	MATEUCCI Francisco (URU)	
2	TEJADA Anibal (URU)	VALLARINO Ricardo (URU)	
3	WARNKEN Alberto (CHI)	LANGENUS Jean (BEL)	
4	REGO Gilberto (BRA)	SAUCEDO Ulises (BOL)	

	Assistant 2	RoundID	MatchID	Home Team Initials	\
0	REGO Gilberto (BRA)	201.0	1096.0	FRA	
1	WARNKEN Alberto (CHI)	201.0	1090.0	USA	
2	BALWAY Thomas (FRA)	201.0	1093.0	YUG	
3	MATEUCCI Francisco (URU)	201.0	1098.0	ROU	
4	RADULESCU Constantin (ROU)	201.0	1085.0	ARG	

	Away Team Initials
0	MEX
1	BEL
2	BRA
3	PER
4	FRA

```
[23]: d1_matches.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4572 entries, 0 to 4571
Data columns (total 20 columns):
#   Column              Non-Null Count  Dtype
---  -
0   Year                852 non-null   float64
1   Datetime            842 non-null   datetime64[ns]
2   Stage               852 non-null   object
3   Stadium             852 non-null   object
4   City                852 non-null   object
5   Home Team Name      852 non-null   object
6   Home Team Goals     852 non-null   float64
7   Away Team Goals     852 non-null   float64
8   Away Team Name      852 non-null   object
9   Win conditions      852 non-null   object
10  Attendance          850 non-null   float64
```

```

11 Half-time Home Goals 852 non-null float64
12 Half-time Away Goals 852 non-null float64
13 Referee              852 non-null object
14 Assistant 1          852 non-null object
15 Assistant 2          852 non-null object
16 RoundID              852 non-null float64
17 MatchID              852 non-null float64
18 Home Team Initials   852 non-null object
19 Away Team Initials   852 non-null object
dtypes: datetime64[ns](1), float64(8), object(11)
memory usage: 714.5+ KB

```

```
[24]: d1_matches.isnull().sum()
```

```

[24]: Year                3720
      Datetime            3730
      Stage              3720
      Stadium            3720
      City               3720
      Home Team Name      3720
      Home Team Goals     3720
      Away Team Goals     3720
      Away Team Name      3720
      Win conditions      3720
      Attendance          3722
      Half-time Home Goals 3720
      Half-time Away Goals 3720
      Referee             3720
      Assistant 1         3720
      Assistant 2         3720
      RoundID             3720
      MatchID             3720
      Home Team Initials   3720
      Away Team Initials   3720
      dtype: int64

```

```
[25]: d1_matches = d1_matches.dropna()
```

```
[27]: d1_matches.isnull().sum()
```

```

[27]: Year                0
      Datetime            0
      Stage              0
      Stadium            0
      City               0
      Home Team Name      0
      Home Team Goals     0
      Away Team Goals     0

```

```

Away Team Name      0
Win conditions      0
Attendance          0
Half-time Home Goals 0
Half-time Away Goals 0
Referee            0
Assistant 1         0
Assistant 2         0
RoundID            0
MatchID            0
Home Team Initials  0
Away Team Initials  0
dtype: int64

```

```
[29]: d1_matches.shape
```

```
[29]: (840, 20)
```

```
[30]: d1_matches = d1_matches.drop_duplicates()
```

```
[31]: d1_matches.shape
```

```
[31]: (825, 20)
```

```
[32]: d1_matches.columns
```

```
[32]: Index(['Year', 'Datetime', 'Stage', 'Stadium', 'City', 'Home Team Name',
          'Home Team Goals', 'Away Team Goals', 'Away Team Name',
          'Win conditions', 'Attendance', 'Half-time Home Goals',
          'Half-time Away Goals', 'Referee', 'Assistant 1', 'Assistant 2',
          'RoundID', 'MatchID', 'Home Team Initials', 'Away Team Initials'],
          dtype='object')
```

```
[34]: del d1_matches['RoundID']
```

```
[35]: d1_matches.columns
```

```
[35]: Index(['Year', 'Datetime', 'Stage', 'Stadium', 'City', 'Home Team Name',
          'Home Team Goals', 'Away Team Goals', 'Away Team Name',
          'Win conditions', 'Attendance', 'Half-time Home Goals',
          'Half-time Away Goals', 'Referee', 'Assistant 1', 'Assistant 2',
          'MatchID', 'Home Team Initials', 'Away Team Initials'],
          dtype='object')
```

```
[38]: d1_matches['Goals'] = d1_matches['Home Team Goals'] + d1_matches['Away Team_
      ↪Goals']
```

```
[39]: d1_matches
```

[39]:	Year	Datetime	Stage	Stadium \
0	1930.0	1930-07-13 15:00:00	Group 1	Pocitos
1	1930.0	1930-07-13 15:00:00	Group 4	Parque Central
2	1930.0	1930-07-14 12:45:00	Group 2	Parque Central
3	1930.0	1930-07-14 14:50:00	Group 3	Pocitos
4	1930.0	1930-07-15 16:00:00	Group 1	Parque Central
..	...	...	...	...
831	2014.0	2014-07-05 13:00:00	Quarter-finals	Estadio Nacional
832	2014.0	2014-06-29 13:00:00	Round of 16	Estadio Castelao
833	2014.0	2014-06-29 17:00:00	Round of 16	Arena Pernambuco
834	2014.0	2014-07-01 13:00:00	Round of 16	Arena de Sao Paulo
835	2014.0	2014-07-01 17:00:00	Round of 16	Arena Fonte Nova

	City	Home Team Name	Home Team Goals	Away Team Goals \
0	Montevideo	France	4.0	1.0
1	Montevideo	USA	3.0	0.0
2	Montevideo	Yugoslavia	2.0	1.0
3	Montevideo	Romania	3.0	1.0
4	Montevideo	Argentina	1.0	0.0
..	...	...	...	...
831	Brasilia	Argentina	1.0	0.0
832	Fortaleza	Netherlands	2.0	1.0
833	Recife	Costa Rica	1.0	1.0
834	Sao Paulo	Argentina	1.0	0.0
835	Salvador	Belgium	2.0	1.0

	Away Team Name	Win conditions	Attendance \
0	Mexico		4444.0
1	Belgium		18346.0
2	Brazil		24059.0
3	Peru		2549.0
4	France		23409.0
..	...	...	...
831	Belgium		68551.0
832	Mexico		58817.0
833	Greece	Costa Rica win on penalties (5 - 3)	41242.0
834	Switzerland	Argentina win after extra time	63255.0
835	USA	Belgium win after extra time	51227.0

	Half-time Home Goals	Half-time Away Goals	Referee \
0	3.0	0.0	LOMBARDI Domingo (URU)
1	2.0	0.0	MACIAS Jose (ARG)
2	2.0	0.0	TEJADA Anibal (URU)
3	1.0	0.0	WARNKEN Alberto (CHI)
4	0.0	0.0	REGO Gilberto (BRA)
..	...	...	...
831	1.0	0.0	Nicola RIZZOLI (ITA)



832	0.0	0.0	PROENCA Pedro (POR)
833	0.0	0.0	Ben WILLIAMS (AUS)
834	0.0	0.0	ERIKSSON Jonas (SWE)
835	0.0	0.0	HAIMOUDI Djamel (ALG)

	Assistant 1	Assistant 2	MatchID \
0	CRISTOPHE Henry (BEL)	REGO Gilberto (BRA)	1096.0
1	MATEUCCI Francisco (URU)	WARNKEN Alberto (CHI)	1090.0
2	VALLARINO Ricardo (URU)	BALWAY Thomas (FRA)	1093.0
3	LANGENUS Jean (BEL)	MATEUCCI Francisco (URU)	1098.0
4	SAUCEDO Ulises (BOL)	RADULESCU Constantin (ROU)	1085.0
..	...	...	...
831	Renato FAVERANI (ITA)	Andrea STEFANI (ITA)	300186504.0
832	MIRANDA Bertino (POR)	TRIGO Jose (POR)	300186508.0
833	CREAM Matthew (AUS)	ANAZ Hakan (AUS)	300186459.0
834	KLASENIUS Mathias (SWE)	WARNMARK Daniel (SWE)	300186503.0
835	ACHIK Redouane (MAR)	ETCHIALI Abdelhak (ALG)	300186497.0

	Home Team Initials	Away Team Initials	Goals
0	FRA	MEX	5.0
1	USA	BEL	3.0
2	YUG	BRA	3.0
3	ROU	PER	4.0
4	ARG	FRA	1.0
..	...	...	...
831	ARG	BEL	1.0
832	NED	MEX	3.0
833	CRC	GRE	2.0
834	ARG	SUI	1.0
835	BEL	USA	3.0

[825 rows x 20 columns]

[40]: d2\_players.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 37784 entries, 0 to 37783
Data columns (total 9 columns):
#   Column          Non-Null Count  Dtype
---  -
0   RoundID         37784 non-null  int64
1   MatchID         37784 non-null  int64
2   Team Initials   37784 non-null  object
3   Coach Name      37784 non-null  object
4   Line-up         37784 non-null  object
5   Shirt Number    37784 non-null  int64
6   Player Name     37784 non-null  object
7   Position        4143 non-null   object
```

```
      8    Event          9069 non-null    object
dtypes: int64(3), object(6)
memory usage: 2.6+ MB
```

```
[41]: d2_players.describe()
```

```
[41]:
```

	RoundID	MatchID	Shirt Number
count	3.778400e+04	3.778400e+04	37784.000000
mean	1.105647e+07	6.362233e+07	10.726022
std	2.770144e+07	1.123916e+08	6.960138
min	2.010000e+02	2.500000e+01	0.000000
25%	2.630000e+02	1.199000e+03	5.000000
50%	3.370000e+02	2.216000e+03	11.000000
75%	2.559310e+05	9.741000e+07	17.000000
max	9.741060e+07	3.001865e+08	23.000000

```
[42]: d2_players.isnull().sum()
```

```
[42]: RoundID          0
      MatchID        0
      Team Initials   0
      Coach Name      0
      Line-up         0
      Shirt Number    0
      Player Name     0
      Position        33641
      Event           28715
      dtype: int64
```

```
[43]: d2_players.fillna('Not Available', inplace=True)
```

```
[44]: d2_players.isnull().sum()
```

```
[44]: RoundID          0
      MatchID        0
      Team Initials   0
      Coach Name      0
      Line-up         0
      Shirt Number    0
      Player Name     0
      Position        0
      Event           0
      dtype: int64
```

```
[48]: d1_matches = d1_matches.replace(['Germany FR'], 'Germany')
      d1_matches.head(30)
```

[48] :	Year	Datetime	Stage	Stadium \
0	1930.0	1930-07-13 15:00:00	Group 1	Pocitos
1	1930.0	1930-07-13 15:00:00	Group 4	Parque Central
2	1930.0	1930-07-14 12:45:00	Group 2	Parque Central
3	1930.0	1930-07-14 14:50:00	Group 3	Pocitos
4	1930.0	1930-07-15 16:00:00	Group 1	Parque Central
5	1930.0	1930-07-16 14:45:00	Group 1	Parque Central
6	1930.0	1930-07-17 12:45:00	Group 2	Parque Central
7	1930.0	1930-07-17 14:45:00	Group 4	Parque Central
8	1930.0	1930-07-18 14:30:00	Group 3	Estadio Centenario
9	1930.0	1930-07-19 12:50:00	Group 1	Estadio Centenario
10	1930.0	1930-07-19 15:00:00	Group 1	Estadio Centenario
11	1930.0	1930-07-20 13:00:00	Group 2	Estadio Centenario
12	1930.0	1930-07-20 15:00:00	Group 4	Estadio Centenario
13	1930.0	1930-07-21 14:50:00	Group 3	Estadio Centenario
14	1930.0	1930-07-22 14:45:00	Group 1	Estadio Centenario
15	1930.0	1930-07-26 14:45:00	Semi-finals	Estadio Centenario
16	1930.0	1930-07-27 14:45:00	Semi-finals	Estadio Centenario
17	1930.0	1930-07-30 14:15:00	Final	Estadio Centenario
18	1934.0	1934-05-27 16:30:00	Preliminary round	Stadio Benito Mussolini
19	1934.0	1934-05-27 16:30:00	Preliminary round	Giorgio Ascarelli
20	1934.0	1934-05-27 16:30:00	Preliminary round	San Siro
21	1934.0	1934-05-27 16:30:00	Preliminary round	Littorale
22	1934.0	1934-05-27 16:30:00	Preliminary round	Giovanni Berta
23	1934.0	1934-05-27 16:30:00	Preliminary round	Luigi Ferraris
24	1934.0	1934-05-27 16:30:00	Preliminary round	Nazionale PNF
25	1934.0	1934-05-27 16:30:00	Preliminary round	Littorio
26	1934.0	1934-05-31 16:30:00	Quarter-finals	Stadio Benito Mussolini
27	1934.0	1934-05-31 16:30:00	Quarter-finals	San Siro
28	1934.0	1934-05-31 16:30:00	Quarter-finals	Giovanni Berta
29	1934.0	1934-05-31 16:30:00	Quarter-finals	Littorale

	City	Home Team Name	Home Team Goals	Away Team Goals \
0	Montevideo	France	4.0	1.0
1	Montevideo	USA	3.0	0.0
2	Montevideo	Yugoslavia	2.0	1.0
3	Montevideo	Romania	3.0	1.0
4	Montevideo	Argentina	1.0	0.0
5	Montevideo	Chile	3.0	0.0
6	Montevideo	Yugoslavia	4.0	0.0
7	Montevideo	USA	3.0	0.0
8	Montevideo	Uruguay	1.0	0.0
9	Montevideo	Chile	1.0	0.0
10	Montevideo	Argentina	6.0	3.0
11	Montevideo	Brazil	4.0	0.0
12	Montevideo	Paraguay	1.0	0.0
13	Montevideo	Uruguay	4.0	0.0

14	Montevideo	Argentina	3.0	1.0
15	Montevideo	Argentina	6.0	1.0
16	Montevideo	Uruguay	6.0	1.0
17	Montevideo	Uruguay	4.0	2.0
18	Turin	Austria	3.0	2.0
19	Naples	Hungary	4.0	2.0
20	Milan	Switzerland	3.0	2.0
21	Bologna	Sweden	3.0	2.0
22	Florence	Germany	5.0	2.0
23	Genoa	Spain	3.0	1.0
24	Rome	Italy	7.0	1.0
25	Trieste	Czechoslovakia	2.0	1.0
26	Turin	Czechoslovakia	3.0	2.0
27	Milan	Germany	2.0	1.0
28	Florence	Italy	1.0	1.0
29	Bologna	Austria	2.0	1.0

	Away Team Name	Win conditions	Attendance \
0	Mexico		4444.0
1	Belgium		18346.0
2	Brazil		24059.0
3	Peru		2549.0
4	France		23409.0
5	Mexico		9249.0
6	Bolivia		18306.0
7	Paraguay		18306.0
8	Peru		57735.0
9	France		2000.0
10	Mexico		42100.0
11	Bolivia		25466.0
12	Belgium		12000.0
13	Romania		70022.0
14	Chile		41459.0
15	USA		72886.0
16	Yugoslavia		79867.0
17	Argentina		68346.0
18	France	Austria win after extra time	16000.0
19	Egypt		9000.0
20	Netherlands		33000.0
21	Argentina		14000.0
22	Belgium		8000.0
23	Brazil		21000.0
24	USA		25000.0
25	Romania		9000.0
26	Switzerland		12000.0
27	Sweden		3000.0
28	Spain		35000.0

	Half-time Home Goals	Half-time Away Goals	Referee \
0	3.0	0.0	LOMBARDI Domingo (URU)
1	2.0	0.0	MACIAS Jose (ARG)
2	2.0	0.0	TEJADA Anibal (URU)
3	1.0	0.0	WARNKEN Alberto (CHI)
4	0.0	0.0	REGO Gilberto (BRA)
5	1.0	0.0	CRISTOPHE Henry (BEL)
6	0.0	0.0	MATEUCCI Francisco (URU)
7	2.0	0.0	MACIAS Jose (ARG)
8	0.0	0.0	LANGENUS Jean (BEL)
9	0.0	0.0	TEJADA Anibal (URU)
10	3.0	1.0	SAUCEDO Ulises (BOL)
11	1.0	0.0	BALWAY Thomas (FRA)
12	1.0	0.0	VALLARINO Ricardo (URU)
13	4.0	0.0	REGO Gilberto (BRA)
14	2.0	1.0	LANGENUS Jean (BEL)
15	1.0	0.0	LANGENUS Jean (BEL)
16	3.0	1.0	REGO Gilberto (BRA)
17	1.0	2.0	LANGENUS Jean (BEL)
18	0.0	0.0	VAN MOORSEL Johannes (NED)
19	2.0	2.0	BARLASSINA Rinaldo (ITA)
20	2.0	1.0	EKLIND Ivan (SWE)
21	1.0	1.0	BRAUN Eugen (AUT)
22	1.0	2.0	MATTEA Francesco (ITA)
23	3.0	0.0	BIRLEM Alfred (GER)
24	3.0	0.0	MERCET Rene (SUI)
25	0.0	1.0	LANGENUS Jean (BEL)
26	1.0	1.0	BERANEK Alois (AUT)
27	0.0	0.0	BARLASSINA Rinaldo (ITA)
28	0.0	0.0	BAERT Louis (BEL)
29	1.0	0.0	MATTEA Francesco (ITA)

	Assistant 1	Assistant 2	MatchID \
0	CRISTOPHE Henry (BEL)	REGO Gilberto (BRA)	1096.0
1	MATEUCCI Francisco (URU)	WARNKEN Alberto (CHI)	1090.0
2	VALLARINO Ricardo (URU)	BALWAY Thomas (FRA)	1093.0
3	LANGENUS Jean (BEL)	MATEUCCI Francisco (URU)	1098.0
4	SAUCEDO Ulises (BOL)	RADULESCU Constantin (ROU)	1085.0
5	APHESTEGUY Martin (URU)	LANGENUS Jean (BEL)	1095.0
6	LOMBARDI Domingo (URU)	WARNKEN Alberto (CHI)	1092.0
7	APHESTEGUY Martin (URU)	TEJADA Anibal (URU)	1097.0
8	BALWAY Thomas (FRA)	CRISTOPHE Henry (BEL)	1099.0
9	LOMBARDI Domingo (URU)	REGO Gilberto (BRA)	1094.0
10	ALONSO Gualberto (URU)	RADULESCU Constantin (ROU)	1086.0
11	MATEUCCI Francisco (URU)	VALLEJO Gaspar (MEX)	1091.0

12	MACIAS Jose (ARG)	LOMBARDI Domingo (URU)	1089.0
13	WARNKEN Alberto (CHI)	SAUCEDO Ulises (BOL)	1100.0
14	CRISTOPHE Henry (BEL)	SAUCEDO Ulises (BOL)	1084.0
15	VALLEJO Gaspar (MEX)	WARNKEN Alberto (CHI)	1088.0
16	SAUCEDO Ulises (BOL)	BALWAY Thomas (FRA)	1101.0
17	SAUCEDO Ulises (BOL)	CRISTOPHE Henry (BEL)	1087.0
18	CAIRONI Camillo (ITA)	BAERT Louis (BEL)	1104.0
19	DATTILO Generoso (ITA)	SASSI Otello (ITA)	1119.0
20	BERANEK Alois (AUT)	BONIVENTO Ferruccio (ITA)	1133.0
21	CARRARO Albino (ITA)	TURBIANI Giuseppe (ITA)	1102.0
22	MELANDRI Ermenegildo (ITA)	BAERT Jacques (FRA)	1108.0
23	CARMINATI Ettore (ITA)	IVANCSICS Mihaly (HUN)	1111.0
24	ESCARTIN Pedro (ESP)	ZENISEK Bohumil (TCH)	1135.0
25	SCARPI Giuseppe (ITA)	SCORZONI Raffaele (ITA)	1141.0
26	MOHAMED Youssuf (EGY)	BAERT Jacques (FRA)	1143.0
27	MERCET Rene (SUI)	VAN MOORSEL Johannes (NED)	1129.0
28	ZENISEK Bohumil (TCH)	IVANCSICS Mihaly (HUN)	1122.0
29	ESCARTIN Pedro (ESP)	BIRLEM Alfred (GER)	1106.0

	Home Team Initials	Away Team Initials	Goals
0	FRA	MEX	5.0
1	USA	BEL	3.0
2	YUG	BRA	3.0
3	ROU	PER	4.0
4	ARG	FRA	1.0
5	CHI	MEX	3.0
6	YUG	BOL	4.0
7	USA	PAR	3.0
8	URU	PER	1.0
9	CHI	FRA	1.0
10	ARG	MEX	9.0
11	BRA	BOL	4.0
12	PAR	BEL	1.0
13	URU	ROU	4.0
14	ARG	CHI	4.0
15	ARG	USA	7.0
16	URU	YUG	7.0
17	URU	ARG	6.0
18	AUT	FRA	5.0
19	HUN	EGY	6.0
20	SUI	NED	5.0
21	SWE	ARG	5.0
22	GER	BEL	7.0
23	ESP	BRA	4.0
24	ITA	USA	8.0
25	TCH	ROU	3.0
26	TCH	SUI	5.0

27	GER	SWE	3.0
28	ITA	ESP	2.0
29	AUT	HUN	3.0

```
[50]: d3_cups = d3_cups.replace(['Germany FR'], 'Germany')
d3_cups
```

```
[50]:
```

	Year	Country	Winner	Runners-Up	Third \
0	1930	Uruguay	Uruguay	Argentina	USA
1	1934	Italy	Italy	Czechoslovakia	Germany
2	1938	France	Italy	Hungary	Brazil
3	1950	Brazil	Uruguay	Brazil	Sweden
4	1954	Switzerland	Germany	Hungary	Austria
5	1958	Sweden	Brazil	Sweden	France
6	1962	Chile	Brazil	Czechoslovakia	Chile
7	1966	England	England	Germany	Portugal
8	1970	Mexico	Brazil	Italy	Germany
9	1974	Germany	Germany	Netherlands	Poland
10	1978	Argentina	Argentina	Netherlands	Brazil
11	1982	Spain	Italy	Germany	Poland
12	1986	Mexico	Argentina	Germany	France
13	1990	Italy	Germany	Argentina	Italy
14	1994	USA	Brazil	Italy	Sweden
15	1998	France	France	Brazil	Croatia
16	2002	Korea/Japan	Brazil	Germany	Turkey
17	2006	Germany	Italy	France	Germany
18	2010	South Africa	Spain	Netherlands	Germany
19	2014	Brazil	Germany	Argentina	Netherlands

	Fourth	GoalsScored	QualifiedTeams	MatchesPlayed	Attendance
0	Yugoslavia	70	13	18	590.549
1	Austria	70	16	17	363.000
2	Sweden	84	15	18	375.700
3	Spain	88	13	22	1.045.246
4	Uruguay	140	16	26	768.607
5	Germany	126	16	35	819.810
6	Yugoslavia	89	16	32	893.172
7	Soviet Union	89	16	32	1.563.135
8	Uruguay	95	16	32	1.603.975
9	Brazil	97	16	38	1.865.753
10	Italy	102	16	38	1.545.791
11	France	146	24	52	2.109.723
12	Belgium	132	24	52	2.394.031
13	England	115	24	52	2.516.215
14	Bulgaria	141	24	52	3.587.538
15	Netherlands	171	32	64	2.785.100
16	Korea Republic	161	32	64	2.705.197

17	Portugal	147	32	64	3.359.439
18	Uruguay	145	32	64	3.178.856
19	Brazil	171	32	64	3.386.810

```
[51]: d3_cups.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 20 entries, 0 to 19
Data columns (total 10 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Year            20 non-null    int64
1   Country         20 non-null    object
2   Winner          20 non-null    object
3   Runners-Up      20 non-null    object
4   Third           20 non-null    object
5   Fourth          20 non-null    object
6   GoalsScored     20 non-null    int64
7   QualifiedTeams  20 non-null    int64
8   MatchesPlayed   20 non-null    int64
9   Attendance       20 non-null    object
dtypes: int64(4), object(6)
memory usage: 1.7+ KB
```

```
[52]: d3_cups.describe()
```

```
[52]:
```

	Year	GoalsScored	QualifiedTeams	MatchesPlayed
count	20.000000	20.000000	20.000000	20.000000
mean	1974.800000	118.950000	21.250000	41.800000
std	25.582889	32.972836	7.268352	17.218717
min	1930.000000	70.000000	13.000000	17.000000
25%	1957.000000	89.000000	16.000000	30.500000
50%	1976.000000	120.500000	16.000000	38.000000
75%	1995.000000	145.250000	26.000000	55.000000
max	2014.000000	171.000000	32.000000	64.000000

```
[53]: d3_cups.isnull().sum()
```

```
[53]: Year            0
Country           0
Winner            0
Runners-Up       0
Third            0
Fourth           0
GoalsScored      0
QualifiedTeams    0
MatchesPlayed     0
Attendance       0
```



dtype: int64

```
[54]: d1_matches.to_excel("C:/Users/thala/Downloads/FIFA_MATCHS_1_Cleaned.
      ↪xlsx",index= False)
      d2_players.to_excel("C:/Users/thala/Downloads/FIFA_PLAYERS_1_Cleaned.
      ↪xlsx",index= False)
      d3_cups.to_excel("C:/Users/thala/Downloads/FIFA_CUPS_1_Cleaned.xlsx",index=
      ↪False)
```

```
[ ]:
```