FIFA WORLD CUP ANALYSIS

```
# Importing the dependencies
import pandas as pd
import numpy as np
```

```
# Loading the Dataset
worldcup_matches = pd.read_csv('/content/WorldCupMatches.csv')
worldcup_players = pd.read_csv('/content/WorldCupPlayers.csv')
worldcup_info = pd.read_csv('/content/WorldCups.csv')
```

```
worldcup_matches.head()
```

	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	
0	1930.0	13 Jul 1930 - 15:00	Group 1	Pocitos	Montevideo	France	4.0	1.0	Mexico		4444.0	3.0	0.0	LOMBARDI Domingo (URU)	CF H
1	1930.0	13 Jul 1930 - 15:00	Group 4	Parque Central	Montevideo	USA	3.0	0.0	Belgium		18346.0	2.0	0.0	MACIAS Jose (ARG)	M
2	1930.0	14 Jul 1930 - 12:45	Group 2	Parque Central	Montevideo	Yugoslavia	2.0	1.0	Brazil		24059.0	2.0	0.0	TEJADA Anibal (URU)	V
3	1930.0	14 Jul 1930 - 14:50	Group 3	Pocitos	Montevideo	Romania	3.0	1.0	Peru		2549.0	1.0	0.0	WARNKEN Alberto (CHI)	L
4	1930.0	15 Jul 1930 - 16:00	Group 1	Parque Central	Montevideo	Argentina	1.0	0.0	France		23409.0	0.0	0.0	REGO Gilberto (BRA)	U

Next steps: Generate code with `worldcup_matches` View recommended plots

```
worldcup_players.head()
```

	RoundID	MatchID	Team Initials	Coach Name	Line-up	Shirt Number	Player Name	Position	Event
0	201	1096	FRA	CAUDRON Raoul (FRA)	S	0	Alex THEPOT	GK	NaN
1	201	1096	MEX	LUQUE Juan (MEX)	S	0	Oscar BONFIGLIO	GK	NaN

Next steps: Generate code with `worldcup_players` View recommended plots

```
worldcup_info.head()
```

	Year	Country	Winner	Runners-Up	Third	Fourth	GoalsScored	Qualifi
0	1930	Uruguay	Uruguay	Argentina	USA	Yugoslavia	70	
1	1934	Italy	Italy	Czechoslovakia	Germany	Austria	70	
2	1938	France	Italy	Hungary	Brazil	Sweden	84	
3	1950	Brazil	Uruguay	Brazil	Sweden	Spain	88	
4	1954	Switzerland	Germany	Hungary	Austria	Uruguay	140	

Next steps: Generate code with `worldcup_info` View recommended plots

```
# Data Integration: Merging Dataset
merged_data = pd.merge(worldcup_matches, worldcup_players, on = ['RoundID', 'MatchID'])
merged_data = pd.merge(merged_data, worldcup_info, on = 'Year')
```

```
# Checking the info the merged dataset
merged_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 39256 entries, 0 to 39255
Data columns (total 36 columns):
 #   Column                Non-Null Count  Dtype  
---  ------                --------------  -----  
 0   Year                  39256 non-null  float64
 1   Datetime              39256 non-null  object 
 2   Stage                 39256 non-null  object 
 3   Stadium               39256 non-null  object 
 4   City                  39256 non-null  object 
 5   Home Team Name        39256 non-null  object 
 6   Home Team Goals       39256 non-null  float64
 7   Away Team Goals       39256 non-null  float64
 8   Away Team Name        39256 non-null  object 
 9   Win conditions        39256 non-null  object 
 10  Attendance_x          39072 non-null  float64
 11  Half-time Home Goals  39256 non-null  float64
 12  Half-time Away Goals  39256 non-null  float64
 13  Referee               39256 non-null  object 
 14  Assistant 1           39256 non-null  object 
 15  Assistant 2           39256 non-null  object 
 16  RoundID               39256 non-null  float64
 17  MatchID               39256 non-null  float64
 18  Home Team Initials    39256 non-null  object 
 19  Away Team Initials    39256 non-null  object 
 20  Team Initials         39256 non-null  object 
 21  Coach Name            39256 non-null  object 
 22  Line-up               39256 non-null  object 
 23  Shirt Number          39256 non-null  int64  
 24  Player Name           39256 non-null  object 
 25  Position              4393 non-null   object 
 26  Event                 9561 non-null   object 
 27  Country               39256 non-null  object 
 28  Winner                39256 non-null  object 
 29  Runners-Up            39256 non-null  object 
 30  Third                 39256 non-null  object 
 31  Fourth                39256 non-null  object 
 32  GoalsScored           39256 non-null  int64  
 33  QualifiedTeams        39256 non-null  int64  
 34  MatchesPlayed         39256 non-null  int64  
 35  Attendance_y          39256 non-null  object 
dtypes: float64(8), int64(4), object(24)
memory usage: 11.1+ MB
```

```
# Checking missing value
merged_data.isnull().sum()
```

```
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_x              184
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
Team Initials               0
Coach Name                  0
Line-up                     0
Shirt Number                0
Player Name                 0
Position                34863
Event                   29695
Country                     0
Winner                      0
Runners-Up                  0
Third                       0
Fourth                      0
GoalsScored                 0
QualifiedTeams              0
MatchesPlayed               0
Attendance_y                0
dtype: int64
```

```
# Dropping irrelevant columns and null values
merged_data.dropna(inplace = True)
```

```
merged_data.isnull().sum()
```

```
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_x            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
Team Initials           0
Coach Name              0
Line-up                 0
Shirt Number            0
Player Name             0
Position                0
Event                   0
Country                 0
Winner                  0
Runners-Up              0
Third                   0
Fourth                  0
GoalsScored             0
QualifiedTeams          0
MatchesPlayed           0
Attendance_y            0
dtype: int64
```

```
merged_data.describe()
```

	Year	Home Team Goals	Away Team Goals	Attendance_x	Half-time Home Goals	Half-time Away Goals	
count	601.000000	601.000000	601.000000	601.000000	601.000000	601.000000	6.01(
mean	1993.693844	1.738769	1.242928	48733.011647	0.682196	0.497504	1.48(
std	19.165580	1.597908	1.125857	20758.991486	0.962895	0.728293	3.19-
min	1930.000000	0.000000	0.000000	2549.000000	0.000000	0.000000	2.01(
25%	1986.000000	1.000000	0.000000	35000.000000	0.000000	0.000000	3.09(
50%	1998.000000	2.000000	1.000000	45500.000000	0.000000	0.000000	1.02(
75%	2010.000000	2.000000	2.000000	62561.000000	1.000000	1.000000	2.55(
max	2014.000000	10.000000	7.000000	114600.000000	6.000000	5.000000	9.74(

```
# Exporting merged data to CSV
merged_data.to_csv('merged_data.csv', index = False)
```