Data Collection and Preparation

```python
import pandas as pd

# Load datasets
matches_df = pd.read_csv('WorldCupMatches.csv')
players_df = pd.read_csv('WorldCupPlayers.csv')
cups_df = pd.read_csv('WorldCups.csv')

# Display initial structure of datasets
print(matches_df.head())
print(players_df.head())
print(cups_df.head())
```

```
     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
   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
   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
```

```
# Display the first few rows of the datasets
world_cups.head()
print(world_cup_matches.head())
print(world_cup_players.head())
```

```
     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
     RoundID  MatchID Team Initials                 Coach Name Line-up
\
35       201      1090            USA           MILLAR Bob (USA)       S

74       201      1093            BRA  DE CARVALHO Pindaro (BRA)       S

113      201      1098            PER        BRU Francisco (ESP)       S

415      201      1091            BRA  DE CARVALHO Pindaro (BRA)       S

468      201      1089            PAR   DURAND LAGUNA Jose (ARG)       S


     Shirt Number       Player Name Position      Event
35              0        Tom FLORIE        C       G45'
74              0         PREGUINHO        C       G62'
113             0   Placido GALINDO        C       R70'
415             0         PREGUINHO        C  G67' G83'
468             0  Luis VARGAS PENA        C       G40'


# Data cleaning: handle missing values, ensure consistency, etc.
world_cups.dropna(inplace=True)
world_cup_matches.dropna(inplace=True)
world_cup_players.dropna(inplace=True)
```

Key Metrics Identification:

Team Performance Metrics:

Goals scored and conceded.

Possession percentage.

Passing accuracy.

Defensive actions (tackles, interceptions).

```
# Calculate goals scored and conceded
team_performance = matches_df.groupby('Home Team Name').agg(
    total_goals_scored=('Home Team Goals', 'sum'),
    avg_goals_scored=('Home Team Goals', 'mean'),
    total_goals_conceded=('Away Team Goals', 'sum'),
    avg_goals_conceded=('Away Team Goals', 'mean')
).reset_index()
team_performance.head(50)
```

```
      Home Team Name  total_goals_scored  avg_goals_scored  \
0            Algeria                 5.0          0.833333
1             Angola                 0.0          0.000000
2          Argentina               111.0          2.055556
3          Australia                 7.0          1.166667
4            Austria                31.0          2.384615
5            Belgium                27.0          1.500000
6            Bolivia                 1.0          1.000000
7             Brazil               180.0          2.195122
8           Bulgaria                11.0          1.571429
9           Cameroon                11.0          1.000000
10            Canada                 0.0          0.000000
11             Chile                25.0          1.785714
12          China PR                 0.0          0.000000
13          Colombia                11.0          1.571429
14        Costa Rica                 7.0          1.000000
15           Croatia                 3.0          1.000000
16              Cuba                 5.0          2.500000
17    Czech Republic                 0.0          0.000000
18    Czechoslovakia                27.0          2.700000
19     C�te d'Ivoire                 5.0          1.666667
20           Denmark                13.0          1.857143
21           Ecuador                 4.0          1.000000
22           England                54.0          1.542857
23            France                68.0          2.193548
24         German DR                 3.0          1.000000
25           Germany                69.0          2.029412
26        Germany FR                99.0          2.302326
```

```
27             Ghana                 4.0          1.000000
28            Greece                 4.0          1.000000
29             Haiti                 0.0          0.000000
30          Honduras                 2.0          0.400000
31           Hungary                73.0          4.055556
32           IR Iran                 0.0          0.000000
33              Iran                 1.0          1.000000
34              Iraq                 1.0          0.500000
35             Italy                99.0          1.736842
36           Jamaica                 1.0          1.000000
37             Japan                 7.0          0.700000
38         Korea DPR                 2.0          0.666667
39    Korea Republic                18.0          1.285714
40            Mexico                22.0          1.375000
41           Morocco                 3.0          0.750000
42       Netherlands                51.0          1.593750
43       New Zealand                 1.0          1.000000
44           Nigeria                12.0          1.333333
45  Northern Ireland                 5.0          1.000000
46            Norway                 1.0          1.000000
47          Paraguay                14.0          1.272727
48              Peru                13.0          2.600000
49            Poland                27.0          1.687500

    total_goals_conceded  avg_goals_conceded
0                   10.0            1.666667
1                    1.0            1.000000
2                   44.0            0.814815
3                   11.0            1.833333
4                   17.0            1.307692
5                   16.0            0.888889
6                    3.0            3.000000
7                   78.0            0.951220
8                   10.0            1.428571
9                   23.0            2.090909
10                   1.0            1.000000
11                  11.0            0.785714
12                   2.0            2.000000
13                   6.0            0.857143
14                  10.0            1.428571
15                   6.0            2.000000
16                   4.0            2.000000
17                   4.0            2.000000
18                   8.0            0.800000
19                   3.0            1.000000
20                  13.0            1.857143
21                   3.0            0.750000
22                  20.0            0.571429
23                  31.0            1.000000
```

```
24                     2.0                 0.666667
25                    32.0                 0.941176
26                    36.0                 0.837209
27                     5.0                 1.250000
28                     6.0                 1.500000
29                     7.0                 7.000000
30                     8.0                 1.600000
31                    19.0                 1.055556
32                     0.0                 0.000000
33                     1.0                 1.000000
34                     3.0                 1.500000
35                    41.0                 0.719298
36                     3.0                 3.000000
37                    14.0                 1.400000
38                     4.0                 1.333333
39                    22.0                 1.571429
40                    11.0                 0.687500
41                     5.0                 1.250000
42                    21.0                 0.656250
43                     1.0                 1.000000
44                    14.0                 1.555556
45                     6.0                 1.200000
46                     0.0                 0.000000
47                    10.0                 0.909091
48                     4.0                 0.800000
49                    14.0                 0.875000
```

Player Performance Metrics:

Goals and assists by key players.

Player ratings and form.

Injuries and suspensions.

```
# Calculate goals and assists by players
player_performance = players_df.groupby('Player Name').agg(
    total_goals=('Event', lambda x: x.str.contains('G').sum()),  #
Assuming 'G' stands for goals
    total_assists=('Event', lambda x: x.str.contains('A').sum())  #
Assuming 'A' stands for assists
).reset_index()
player_performance.head(50)
```

```
         Player Name  total_goals  total_assists
0            ?URI?I?            0              0
1          A BAUTISTA           0              0
2              A COLE           0              0
3          A GUARDADO           0              0
```

```
4           A MEDINA            0              0
5      A. AL-DOSSARY            0              0
6      A. AL-GANOUBI            0              0
7         A. ALMEIDA            0              0
8            A. AYEW            2              0
9             A. BAK            0              0
10        A. BALANTA            0              0
11      A. CHOL HYOK            0              0
12           A. CRUZ            0              0
13           A. DAEI            0              0
14        A. DELGADO            2              0
15         A. DIARRA            0              0
16      A. FERNANDEZ            0              0
17    A. GARCIA ASPE            0              0
18       A. GONZALEZ            0              0
19       A. GUARDADO            1              0
20           A. GYAN            3              0
21       A. HAGHIGHI            0              0
22      A. HERNANDEZ            0              0
23        A. INIESTA            0              0
24           A. JOHN            0              0
25          A. KELLY            0              0
26         A. LATIFI            0              0
27          A. LOPEZ            0              0
28      A. M. NDIAYE            0              0
29         A. MADANI            0              0
30          A. MEJIA            0              0
31        A. MOKOENA            0              0
32        A. NAELSON            1              0
33        A. PEREIRA            0              0
34         A. PULIDO            0              0
35   A. R. ABEDZADEH            0              0
36  A. R. MANSOURIAN            0              0
37      A. RODRIGUEZ            0              0
38          A. ROJAS            0              0
39           A. SONG            0              0
40       A. SVENSSON            0              0
41       A. TALAVERA            0              0
42          A. TOURE            0              0
43       A. VALENCIA            0              0
44       A. YONG HAK            0              0
45      A. ZUBROMAWI            0              0
46  A.A. OSTAD ASADI            0              0
47         A.BORHANI            0              0
48            A.DAEI            0              0
49         A.INIESTA            2              0
```

Research and Findings

Compare successful and unsuccessful teams and provide insights.

Comparative Analysis

```python
# Compare metrics of winning teams with those of runners-up and semi-finalists
comparison = world_cup_matches[world_cup_matches['Stage'].isin(['Final', 'Semi-finals'])]
comparison_metrics = comparison.groupby(['Stage', 'Year'])[['Home Team Goals', 'Away Team Goals']].mean()
comparison_metrics.head(50)
```

```
                   Home Team Goals  Away Team Goals
Stage       Year
Final       1930.0             4.0              2.0
            1934.0             2.0              1.0
            1938.0             4.0              2.0
            1954.0             3.0              2.0
            1958.0             5.0              2.0
            1962.0             3.0              1.0
            1966.0             4.0              2.0
            1970.0             4.0              1.0
            1974.0             1.0              2.0
            1978.0             3.0              1.0
            1982.0             3.0              1.0
            1986.0             3.0              2.0
            1990.0             1.0              0.0
            1994.0             0.0              0.0
            1998.0             0.0              3.0
            2002.0             0.0              2.0
            2006.0             1.0              1.0
            2010.0             0.0              1.0
            2014.0             1.0              0.0
Semi-finals 1930.0             6.0              1.0
            1934.0             2.0              0.5
            1938.0             3.5              1.0
            1954.0             5.0              1.5
            1958.0             4.0              1.5
            1962.0             3.5              1.5
            1966.0             2.0              1.0
            1970.0             3.5              2.0
            1982.0             1.5              2.5
            1986.0             1.0              1.0
            1990.0             1.0              1.0
            1994.0             0.5              1.5
            1998.0             1.5              1.0
            2002.0              1.0               0.0
            2006.0              0.0               1.5
            2010.0              1.0               2.0
            2014.0              0.5               3.5
```

Stage-specific Insights

```
# Analyze differences in performance metrics between group stages and
knockout stages
group_stage = world_cup_matches[world_cup_matches['Stage'] == 'Group']
knockout_stage = world_cup_matches[world_cup_matches['Stage'] !=
'Group']

group_metrics = group_stage.groupby('Year')[['Home Team Goals', 'Away
Team Goals']].mean()
knockout_metrics = knockout_stage.groupby('Year')[['Home Team Goals',
'Away Team Goals']].mean()

print('Group Stage Metrics:')
group_metrics.head(50)
print('Knockout Stage Metrics:')
knockout_metrics.head(50)
```

```
Group Stage Metrics:
Knockout Stage Metrics:
```

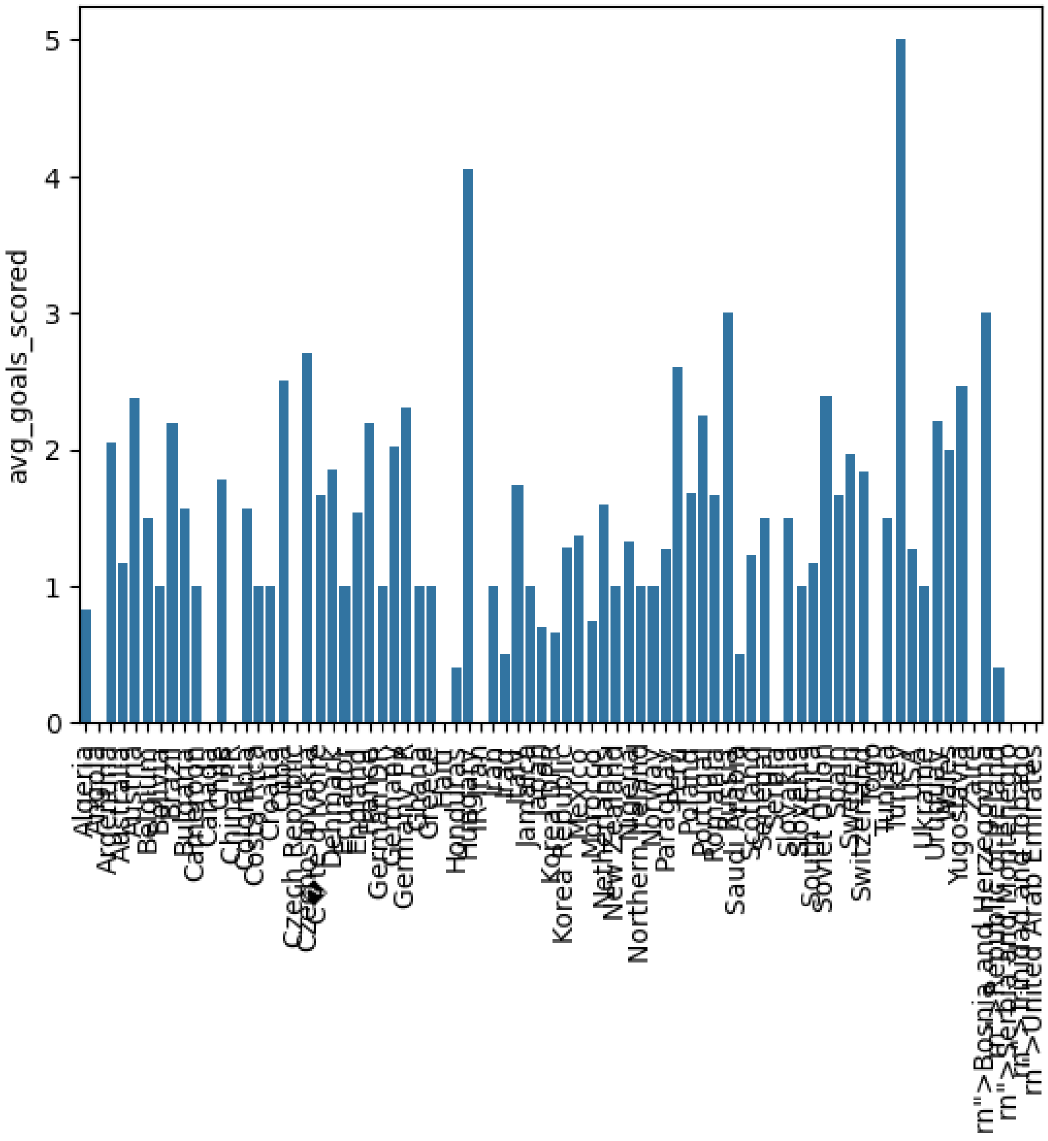
```
        Home Team Goals  Away Team Goals
Year
1930.0         3.277778         0.611111
1934.0         2.823529         1.294118
1938.0         3.388889         1.277778
1950.0         3.136364         0.863636
1954.0         4.192308         1.192308
1958.0         2.514286         1.085714
1962.0         2.156250         0.625000
1966.0         2.156250         0.625000
1970.0         2.250000         0.718750
1974.0         1.342105         1.210526
1978.0         2.078947         0.605263
1982.0         1.865385         0.942308
1986.0         1.423077         1.115385
1990.0         1.288462         0.923077
1994.0         1.596154         1.115385
1998.0         1.531250         1.140625
2002.0         1.390625         1.125000
2006.0         1.343750         0.953125
2010.0         1.187500         1.078125
2014.0         1.217949         1.346154
```

Visualization

Create visual representations of the data.

Bar Charts and Line Graphs

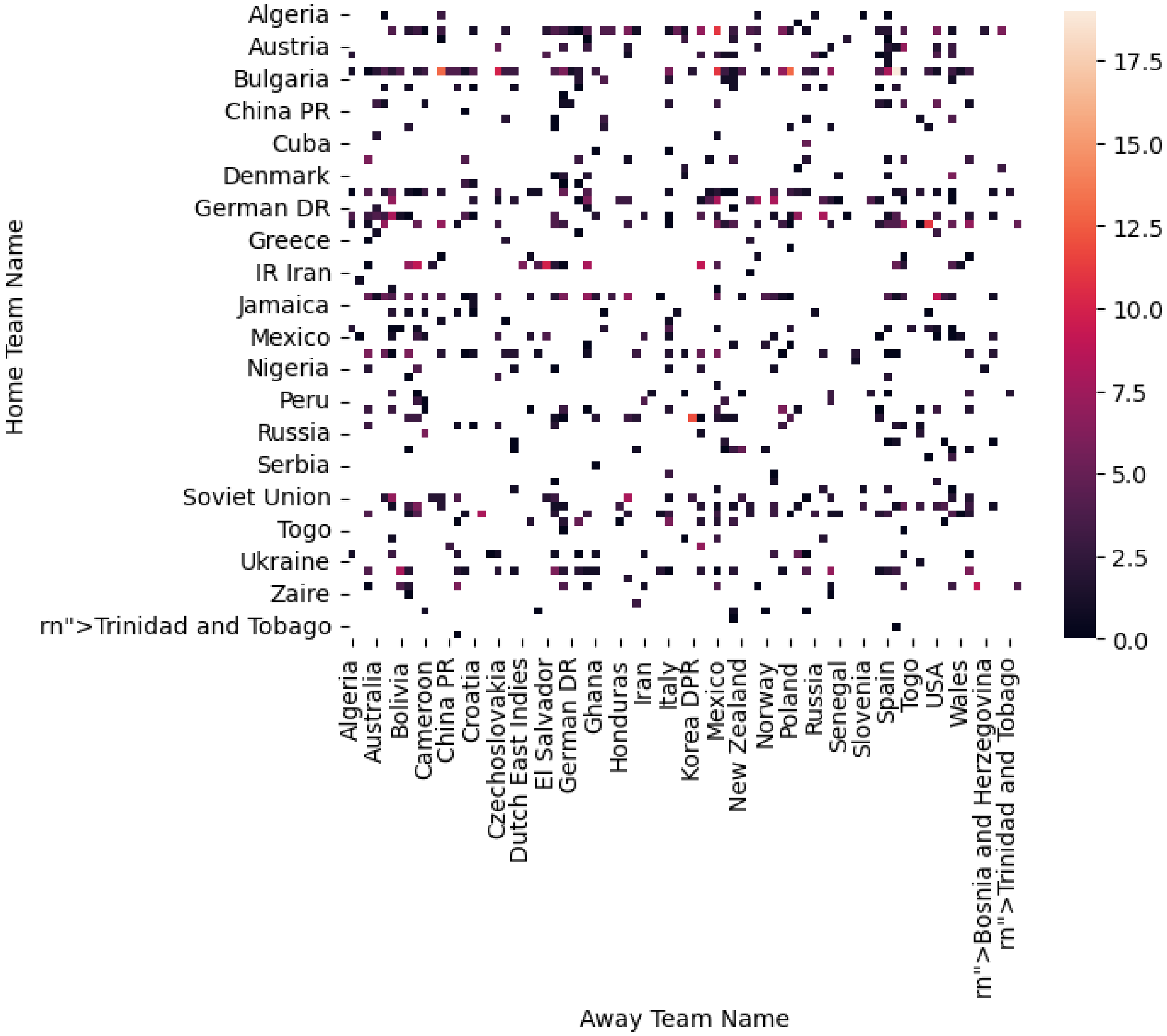
```
import matplotlib.pyplot as plt
import seaborn as sns

# Example: Bar chart for average goals scored by each team
sns.barplot(x='Home Team Name', y='avg_goals_scored',
data=team_performance)
plt.xticks(rotation=90)
plt.show()
```



Heatmaps:

Visualize player movements and key areas of action.

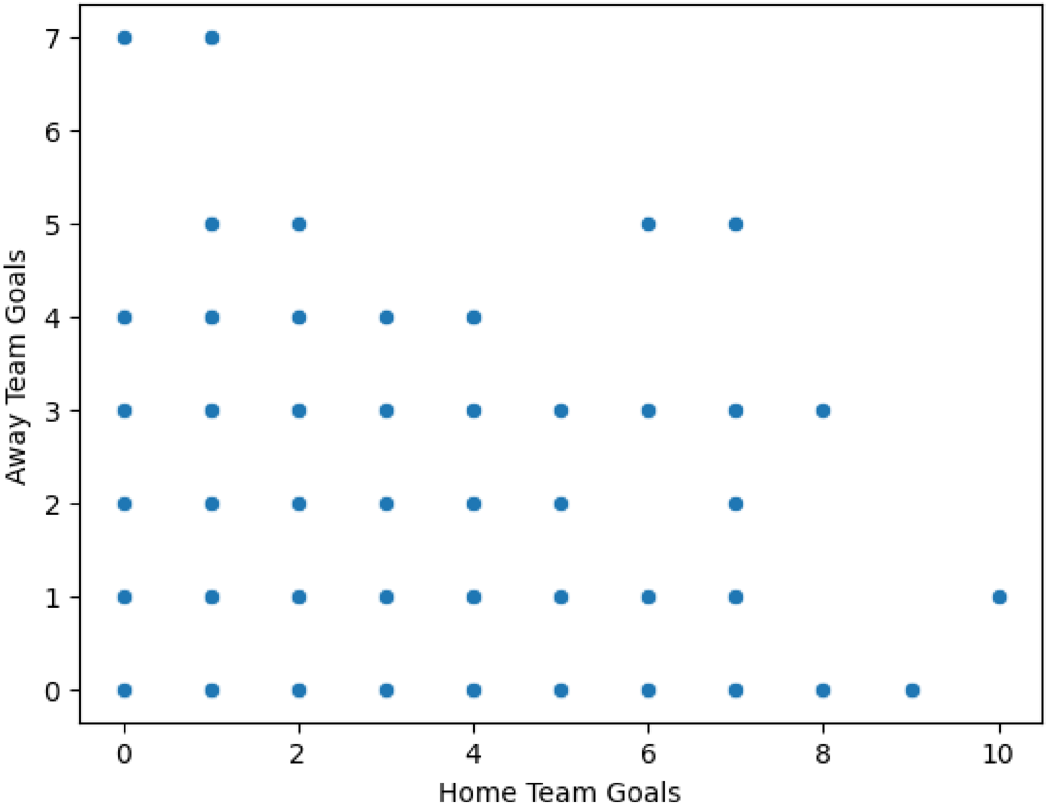
```
# Example: Heatmap of goals scored
sns.heatmap(matches_df.pivot_table(index='Home Team Name',
columns='Away Team Name', values='Home Team Goals', aggfunc='sum'))
plt.show()
```



Scatter Plots:

Show correlations between different metrics and match outcomes.

```
# Example: Scatter plot of goals vs. possession
sns.scatterplot(x='Home Team Goals', y='Away Team Goals',
data=matches_df)
plt.show()
```



Conclusion:

By following the steps outlined above, you can analyze the datasets and uncover key metrics influencing World Cup wins. This project will provide valuable insights into successful strategies and factors contributing to World Cup victories.