In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

import warnings
warnings.filterwarnings('ignore')
```

In [2]:
```
player_df=pd.read_csv("Fifa WC 2022 Qatar Data Analysis(EDA)\player_stats.csv")
match_df=pd.read_csv("Fifa WC 2022 Qatar Data Analysis(EDA)\Fifa_WC_2022_Match_data.csv"
```

In [3]:
```
match_df.head()
```

Out[3]:

	match_no	day_of_week	date	hour	venue	referee	group	1	2	attenda
0	1	Sun	20-Nov-22	17:00	Al Bayt Stadium	Daniele Orsato	Group A	QATAR	ECUADOR	67
1	2	Mon	21-Nov-22	14:00	Khalifa International Stadium	Raphael Claus	Group B	ENGLAND	IRAN	45
2	3	Mon	21-Nov-22	17:00	Al Thumama Stadium	Wilton Sampaio	Group A	SENEGAL	NETHERLANDS	41
3	4	Mon	21-Nov-22	20:00	Ahmed bin Ali Stadium	Abdulrahman Ibrahim Al Jassim	Group B	UNITED STATES	WALES	43
4	5	Tue	22-Nov-22	11:00	Lusail Iconic Stadium	Slavko Vincic	Group C	ARGENTINA	SAUDI ARABIA	88

5 rows × 59 columns

Informtion about the dataset

In [4]:
```
match_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 64 entries, 0 to 63
Data columns (total 59 columns):
 #   Column                            Non-Null Count  Dtype
---  ------                            --------------  -----
 0   match_no                          64 non-null     int64
 1   day_of_week                       64 non-null     object
 2   date                              64 non-null     object
 3   hour                              64 non-null     object
 4   venue                             64 non-null     object
 5   referee                           64 non-null     object
 6   group                             64 non-null     object
 7   1                                 64 non-null     object
 8   2                                 64 non-null     object
 9   attendance                        64 non-null     int64
 10  1_xg                              64 non-null     float64
 11  2_xg                              64 non-null     float64
 12  1_poss                            64 non-null     int64
 13  2_poss                            64 non-null     int64
 14  1_goals                           64 non-null     int64
 15  2_goals                           64 non-null     int64
 16  score                             64 non-null     object
 17  1_attempts                        64 non-null     int64
 18  2_attempts                        64 non-null     int64
 19  1_conceded                        64 non-null     int64
 20  2_conceded                        64 non-null     int64
 21  1_goal_inside_penalty_area        64 non-null     int64
 22  2_goal_inside_penalty_area        64 non-null     int64
 23  1_goal_outside_penalty_area       64 non-null     int64
 24  2_goal_outside_penalty_area       64 non-null     int64
 25  1_ontarget                        64 non-null     int64
 26  2_ontarget                        64 non-null     int64
 27  1_offtarget                       64 non-null     int64
 28  2_offtarget                       64 non-null     int64
 29  1_attempts_inside_penalty_area    64 non-null     int64
 30  2_attempts_inside_penalty_area    64 non-null     int64
 31  1_attempts_outside_penalty_area   64 non-null     int64
 32  2_attempts_outside_penalty_area   64 non-null     int64
 33  1_yellow_cards                    64 non-null     int64
 34  2_yellow_cards                    64 non-null     int64
 35  1_red_cards                       64 non-null     int64
 36  2_red_cards                       64 non-null     int64
 37  faul_against_1                    64 non-null     int64
 38  faul_against_2                    64 non-null     int64
 39  1_offsides                        64 non-null     int64
 40  2_offsides                        64 non-null     int64
 41  1_passes                          64 non-null     int64
 42  2_passes                          64 non-null     int64
 43  1_passes_compeletd                64 non-null     int64
 44  2_passes_compeletd                64 non-null     int64
 45  1_corners                         64 non-null     int64
 46  2_corners                         64 non-null     int64
 47  1_free_kicks                      64 non-null     int64
 48  2_free_kicks                      64 non-null     int64
 49  1_panelties_scored                64 non-null     int64
 50  2_panelties_scored                64 non-null     int64
 51  1_goal_prevented                  64 non-null     int64
 52  2_goal_prevented                  64 non-null     int64
 53  1_own_goal                        64 non-null     int64
 54  2_own_goal                        64 non-null     int64
 55  1_forced_turnovers                64 non-null     int64
 56  2_forced_turnovers                64 non-null     int64
 57  1_defensive_pressure_applied      64 non-null     int64
 58  2_defensive_pressure_applied      64 non-null     int64
```

```
dtypes: float64(2), int64(48), object(9)
memory usage: 29.6+ KB
```

In [5]:
```
match_df.isnull().sum()
```

Out[5]:
```
match_no                              0
day_of_week                           0
date                                  0
hour                                  0
venue                                 0
referee                               0
group                                 0
1                                     0
2                                     0
attendance                            0
1_xg                                  0
2_xg                                  0
1_poss                                0
2_poss                                0
1_goals                               0
2_goals                               0
score                                 0
1_attempts                            0
2_attempts                            0
1_conceded                            0
2_conceded                            0
1_goal_inside_penalty_area            0
2_goal_inside_penalty_area            0
1_goal_outside_penalty_area           0
2_goal_outside_penalty_area           0
1_ontarget                            0
2_ontarget                            0
1_offtarget                           0
2_offtarget                           0
1_attempts_inside_penalty_area        0
2_attempts_inside_penalty_area        0
1_attempts_outside_penalty_area       0
2_attempts_outside_penalty_area       0
1_yellow_cards                        0
2_yellow_cards                        0
1_red_cards                           0
2_red_cards                           0
faul_against_1                        0
faul_against_2                        0
1_offsides                            0
2_offsides                            0
1_passes                              0
2_passes                              0
1_passes_compeletd                    0
2_passes_compeletd                    0
1_corners                             0
2_corners                             0
1_free_kicks                          0
2_free_kicks                          0
1_panelties_scored                    0
2_panelties_scored                    0
1_goal_prevented                      0
2_goal_prevented                      0
1_own_goal                            0
2_own_goal                            0
1_forced_turnovers                    0
2_forced_turnovers                    0
1_defensive_pressure_applied          0
2_defensive_pressure_applied          0
dtype: int64
```

In [6]:
```
# No of matches played on respective venue
match_df['venue'].value_counts()
```

Out[6]:
```
Lusail Iconic Stadium            10
Al Bayt Stadium                   9
Khalifa International Stadium     8
Al Thumama Stadium                8
Education City Stadium            8
Ahmed bin Ali Stadium             7
Stadium 974                       7
Al Janoub Stadium                 7
Name: venue, dtype: int64
```

In [7]:
```
# Match with highest attendence audience
match_df[match_df['attendance']==match_df['attendance'].max()]
```

Out[7]:

	match_no	day_of_week	date	hour	venue	referee	group	1	2	attendance	...	1_p
23	24	Sat	26-Nov-22	20:00	Lusail Iconic Stadium	Daniele Orsato	Group C	ARGENTINA	MEXICO	88966	...	
60	61	Tue	13-Dec-22	20:00	Lusail Iconic Stadium	Daniele Orsato	Semi-Final	ARGENTINA	CROATIA	88966	...	
63	64	Sun	18-Dec-22	16:00	Lusail Iconic Stadium	Szymon Marciniak	Final	ARGENTINA	FRANCE	88966	...	

3 rows × 59 columns

From the above result, all the three highest attendance matches were played on "Lusail Iconic Stadium" with the presence of 88966 spectators.

In [8]:
```
# create "total_goal" feature, by adding 1_goal & 2_goal
match_df['total_goals']=match_df['1_goals']+match_df['2_goals']
```

In [9]:
```
# Highest goal scoring in the match
match_df[match_df['total_goals']==match_df['total_goals'].max()]
```

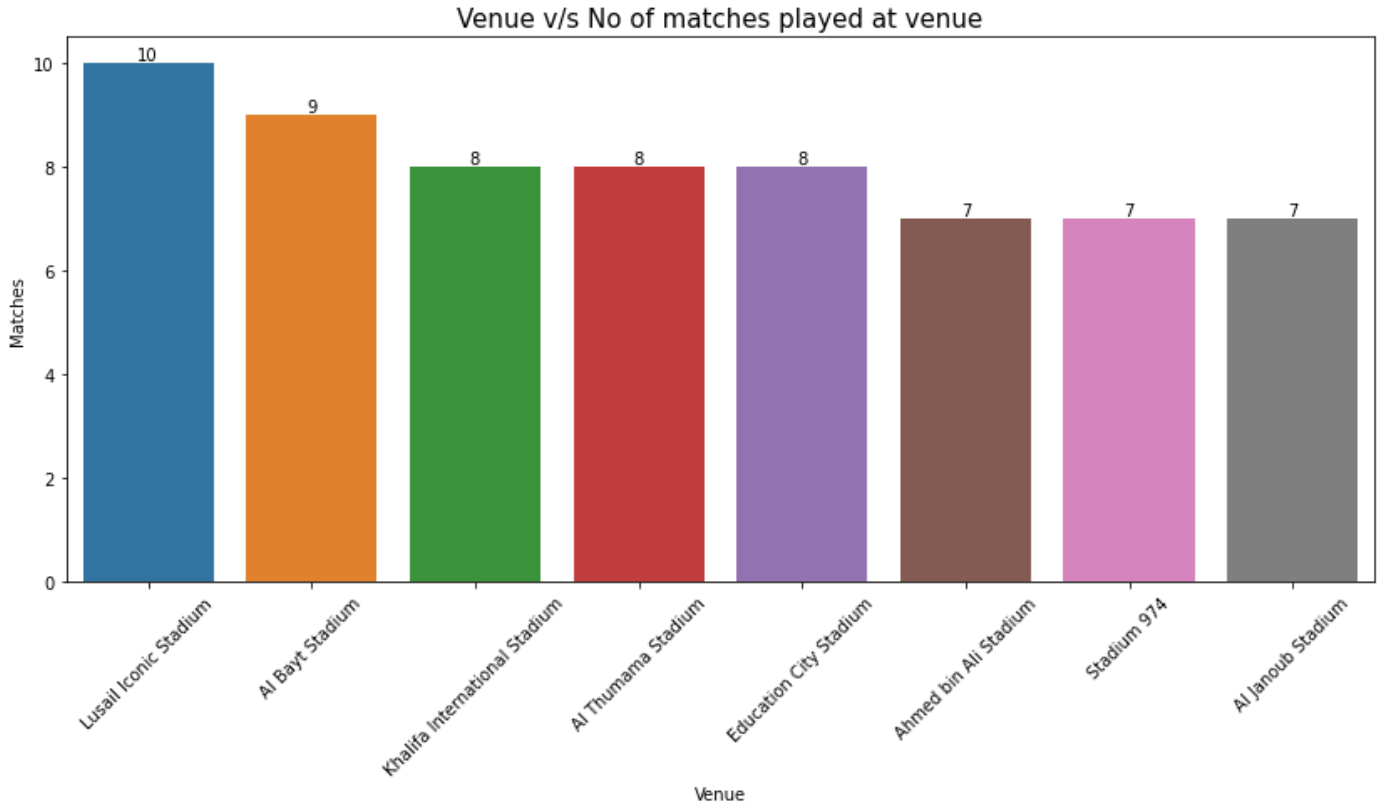
Out[9]:

	match_no	day_of_week	date	hour	venue	referee	group	1	2	attendance	...	2_panel
1	2	Mon	21-Nov-22	14:00	Khalifa International Stadium	Raphael Claus	Group B	ENGLAND	IRAN	45334	...	

1 rows × 60 columns

Match number 2 has highest goals, which was played between ENGLAND and IRAN with 8 goals.

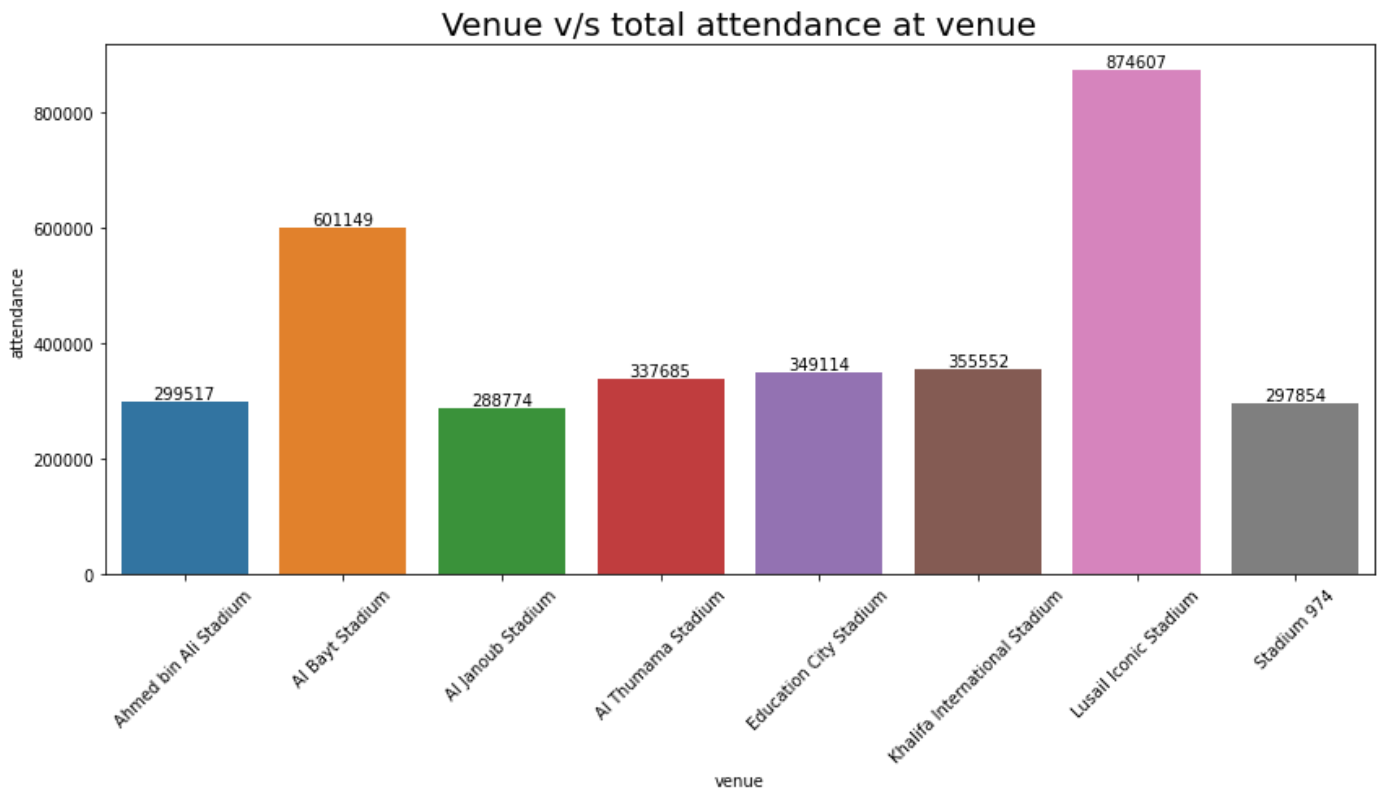
In [10]:
```
# Let's plot Bar graph Venu v/s No of matches played at venue
x=match_df['venue'].value_counts().index
y=match_df['venue'].value_counts().values

df=pd.DataFrame({'Venue':x,'Matches':y})

plt.figure(figsize=(14,6))
ax=sns.barplot(data=df,x='Venue',y='Matches')
ax.bar_label(ax.containers[0])
plt.title('Venue v/s No of matches played at venue',fontsize=15)
plt.xticks(rotation=45)
plt.show()
```



In [11]:
```
# total attendance in all the venue of all matches played
match_df.groupby('venue')['attendance'].sum().sort_values(ascending=False)
```

Out[11]:
```
venue
Lusail Iconic Stadium            874607
Al Bayt Stadium                  601149
Khalifa International Stadium    355552
Education City Stadium           349114
Al Thumama Stadium               337685
Ahmed bin Ali Stadium            299517
Stadium 974                      297854
Al Janoub Stadium                288774
Name: attendance, dtype: int64
```

In [12]:
```
# Bar graph Venue v/s total attendance
x=match_df.groupby('venue')['attendance'].sum().index
y=match_df.groupby('venue')['attendance'].sum().values

df=pd.DataFrame({'venue':x,'attendance':y})

plt.figure(figsize=(14,6))
ax=sns.barplot(data=df,x='venue',y='attendance')
ax.bar_label(ax.containers[0])
plt.title('Venue v/s total attendance at venue',fontsize=20)
plt.xticks(rotation=45)
plt.show()
```



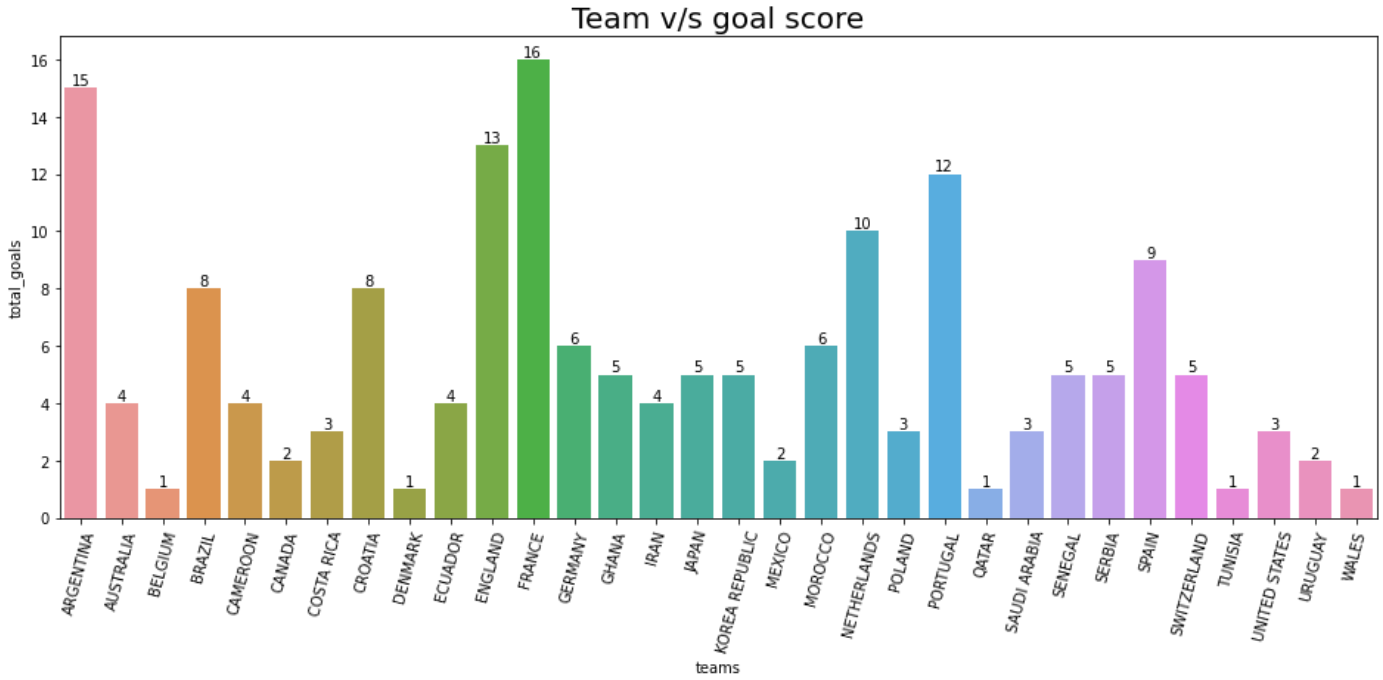
In [13]:
```
# total matches played and goal scored by the team
team_df=pd.DataFrame({'teams':match_df['1'].value_counts().sort_index().index,
                      'total_matches':match_df['1'].value_counts().sort_index().values +
                      'total_goals':match_df.groupby(['1'])['1_goals'].sum().sort_index()

# sort on total_goals
team_df.sort_values(by='total_goals',ascending=False)
```

Out[13]:

	teams	total_matches	total_goals
11	FRANCE	7	16
0	ARGENTINA	7	15
10	ENGLAND	5	13
21	PORTUGAL	5	12
19	NETHERLANDS	5	10
26	SPAIN	4	9
3	BRAZIL	5	8
7	CROATIA	7	8
12	GERMANY	3	6
18	MOROCCO	7	6
15	JAPAN	4	5
27	SWITZERLAND	4	5
25	SERBIA	3	5
24	SENEGAL	4	5
16	KOREA REPUBLIC	4	5
13	GHANA	3	5
14	IRAN	3	4
1	AUSTRALIA	4	4
9	ECUADOR	3	4
4	CAMEROON	3	4
29	UNITED STATES	4	3
6	COSTA RICA	3	3
23	SAUDI ARABIA	3	3
20	POLAND	4	3
5	CANADA	3	2
17	MEXICO	3	2
30	URUGUAY	3	2
22	QATAR	3	1
8	DENMARK	3	1
2	BELGIUM	3	1
28	TUNISIA	3	1
31	WALES	3	1

In [14]:

```
# Team v/s Goal scored
plt.figure(figsize=(16,6))
ax=sns.barplot(data=team_df,x='teams',y='total_goals')
ax.bar_label(ax.containers[0])
plt.title("Team v/s goal score",fontsize=20)
plt.xticks(rotation=75)
plt.show()
```



From the above we can say that France is the highest goal scoring team

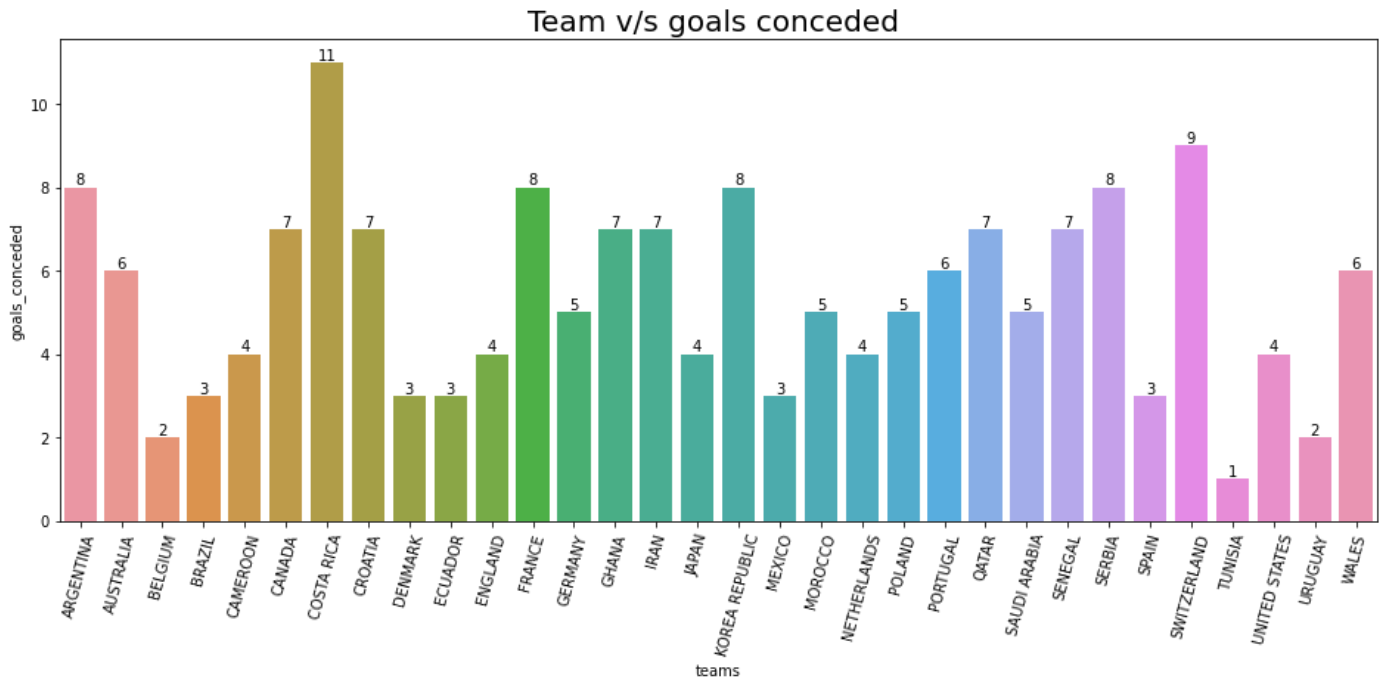
In [15]:
```
# goals conceded by each team
team_df['goals_conceded']=match_df.groupby(['1'])['1_conceded'].sum().sort_index().value

# sort on goals_conceded
team_df.loc[:,['teams','total_matches','goals_conceded']].sort_values(by='goals_conceded
```

Out[15]:

	teams	total_matches	goals_conceded
6	COSTA RICA	3	11
27	SWITZERLAND	4	9
0	ARGENTINA	7	8
25	SERBIA	3	8
11	FRANCE	7	8
16	KOREA REPUBLIC	4	8
5	CANADA	3	7
14	IRAN	3	7
24	SENEGAL	4	7
22	QATAR	3	7
7	CROATIA	7	7
13	GHANA	3	7
21	PORTUGAL	5	6
1	AUSTRALIA	4	6
31	WALES	3	6
23	SAUDI ARABIA	3	5
18	MOROCCO	7	5
12	GERMANY	3	5
20	POLAND	4	5
10	ENGLAND	5	4
29	UNITED STATES	4	4
15	JAPAN	4	4
4	CAMEROON	3	4
19	NETHERLANDS	5	4
9	ECUADOR	3	3
8	DENMARK	3	3
3	BRAZIL	5	3
26	SPAIN	4	3
17	MEXICO	3	3
2	BELGIUM	3	2
30	URUGUAY	3	2
28	TUNISIA	3	1

In [16]:

```
# Bar graph Teams v/s goals conceded

plt.figure(figsize=(16,6))
ax=sns.barplot(data=team_df,x='teams',y='goals_conceded')
ax.bar_label(ax.containers[0])
plt.title('Team v/s goals conceded ',fontsize=20)
plt.xticks(rotation=75)
plt.show()
```



From the above we can say that COSTARICA team conceded highest goals

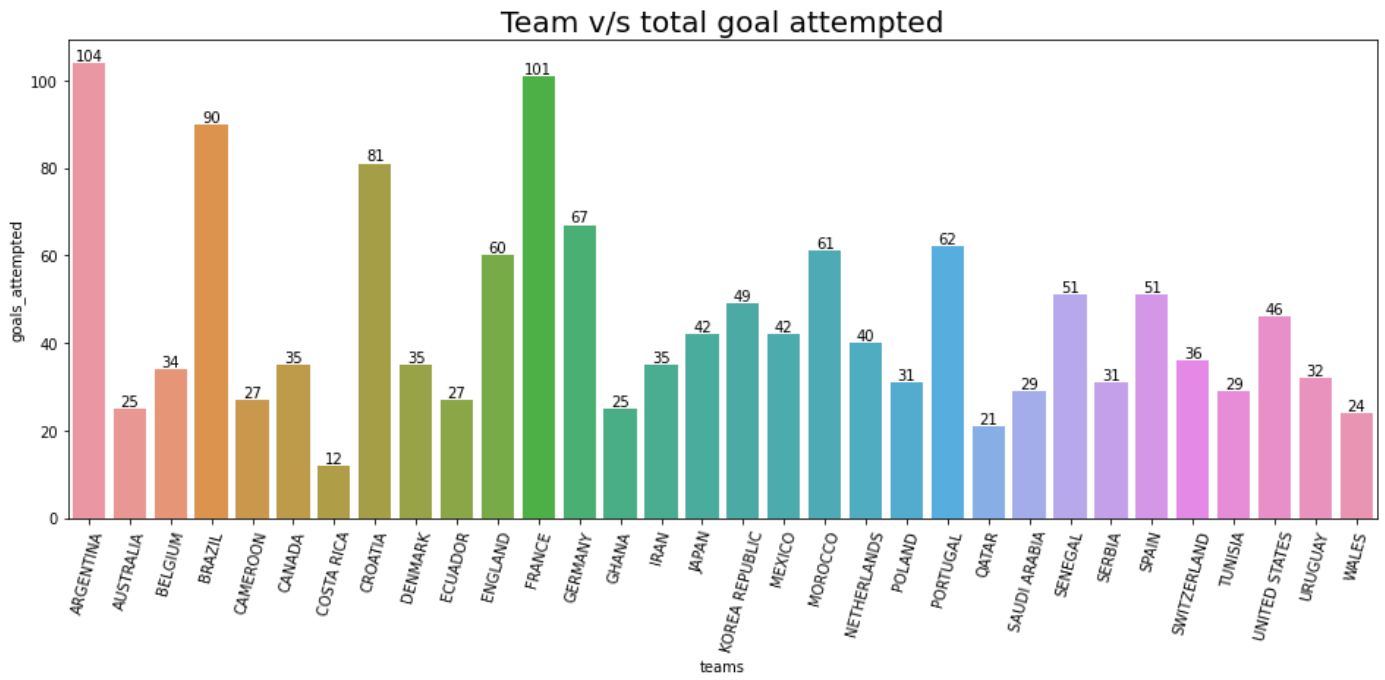
In [17]:
```
# total games played and goals attempted by the team
team_df['goals_attempted']=match_df.groupby(['1'])['1_attempts'].sum().sort_index().valu

# sort on goals_attempted
team_df.loc[:,['teams','total_matches','goals_attempted']].sort_values(by='goals_attempt
```

Out[17]:

	teams	total_matches	goals_attempted
0	ARGENTINA	7	104
11	FRANCE	7	101
3	BRAZIL	5	90
7	CROATIA	7	81
12	GERMANY	3	67
21	PORTUGAL	5	62
18	MOROCCO	7	61
10	ENGLAND	5	60
26	SPAIN	4	51
24	SENEGAL	4	51
16	KOREA REPUBLIC	4	49
29	UNITED STATES	4	46
15	JAPAN	4	42
17	MEXICO	3	42
19	NETHERLANDS	5	40
27	SWITZERLAND	4	36
5	CANADA	3	35
14	IRAN	3	35
8	DENMARK	3	35
2	BELGIUM	3	34
30	URUGUAY	3	32
25	SERBIA	3	31
20	POLAND	4	31
28	TUNISIA	3	29
23	SAUDI ARABIA	3	29
9	ECUADOR	3	27
4	CAMEROON	3	27
1	AUSTRALIA	4	25
13	GHANA	3	25
31	WALES	3	24
22	QATAR	3	21
6	COSTA RICA	3	12

In [18]:

```
# bargraph Teamv/s toal goals attempted
plt.figure(figsize=(16,6))
ax=sns.barplot(data=team_df,x='teams',y='goals_attempted')
ax.bar_label(ax.containers[0])
plt.title('Team v/s total goal attempted',fontsize=20)
plt.xticks(rotation=75)
plt.show()
```



From the above we can say that ARGENTINA team has attempted highest goals

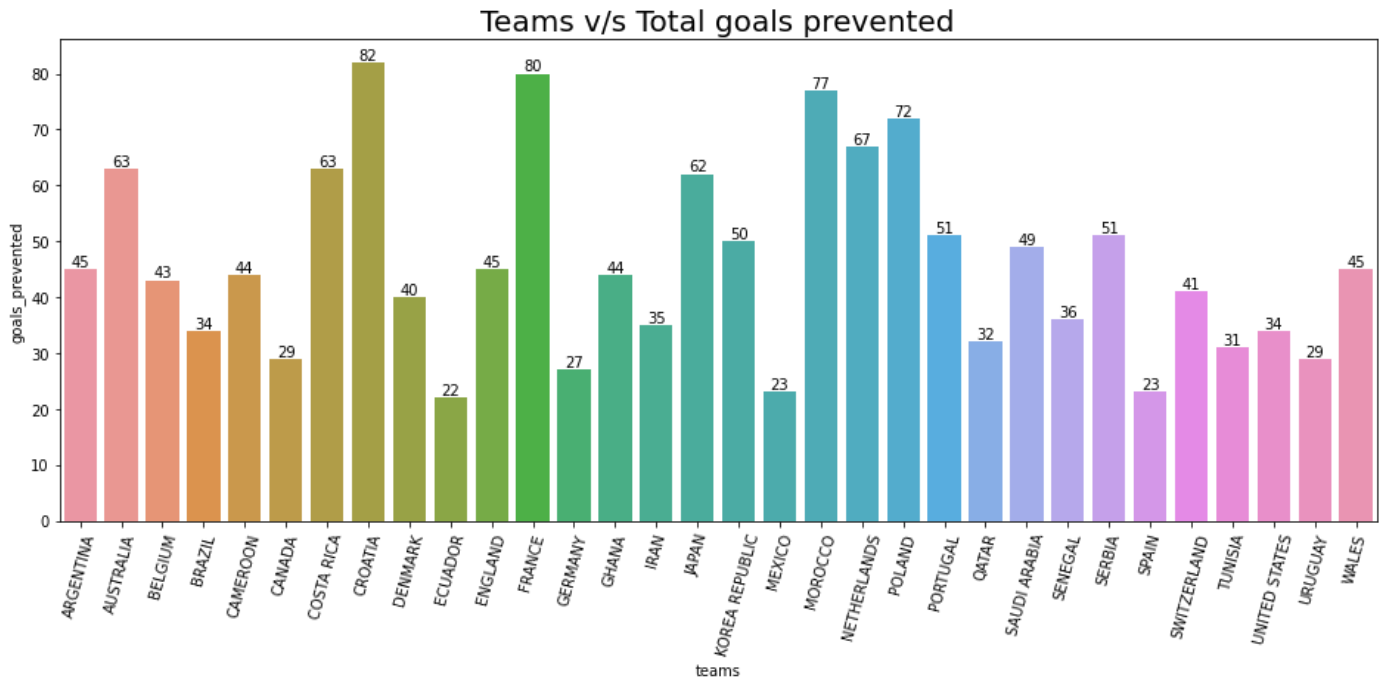
In [19]:
```
# total matches played and goals prevented by the team
team_df['goals_prevented']=match_df.groupby(['1'])['1_goal_prevented'].sum().sort_index(

# sort on goals_prevented
team_df.loc[:,['teams','total_matches','goals_prevented']].sort_values(by='goals_prevent
```

Out[19]:

	teams	total_matches	goals_prevented
7	CROATIA	7	82
11	FRANCE	7	80
18	MOROCCO	7	77
20	POLAND	4	72
19	NETHERLANDS	5	67
6	COSTA RICA	3	63
1	AUSTRALIA	4	63
15	JAPAN	4	62
25	SERBIA	3	51
21	PORTUGAL	5	51
16	KOREA REPUBLIC	4	50
23	SAUDI ARABIA	3	49
0	ARGENTINA	7	45
31	WALES	3	45
10	ENGLAND	5	45
4	CAMEROON	3	44
13	GHANA	3	44
2	BELGIUM	3	43
27	SWITZERLAND	4	41
8	DENMARK	3	40
24	SENEGAL	4	36
14	IRAN	3	35
3	BRAZIL	5	34
29	UNITED STATES	4	34
22	QATAR	3	32
28	TUNISIA	3	31
5	CANADA	3	29
30	URUGUAY	3	29
12	GERMANY	3	27
17	MEXICO	3	23
26	SPAIN	4	23
9	ECUADOR	3	22

In [20]:

```
# Bargrpah Tems v/s goals prevented
plt.figure(figsize=(16,6))
ax=sns.barplot(data=team_df,x='teams',y='goals_prevented')
ax.bar_label(ax.containers[0])
plt.title('Teams v/s Total goals prevented',fontsize=20)
plt.xticks(rotation=75)
plt.show()
```



From the above we can say that CROATIA team has preventd highest goals

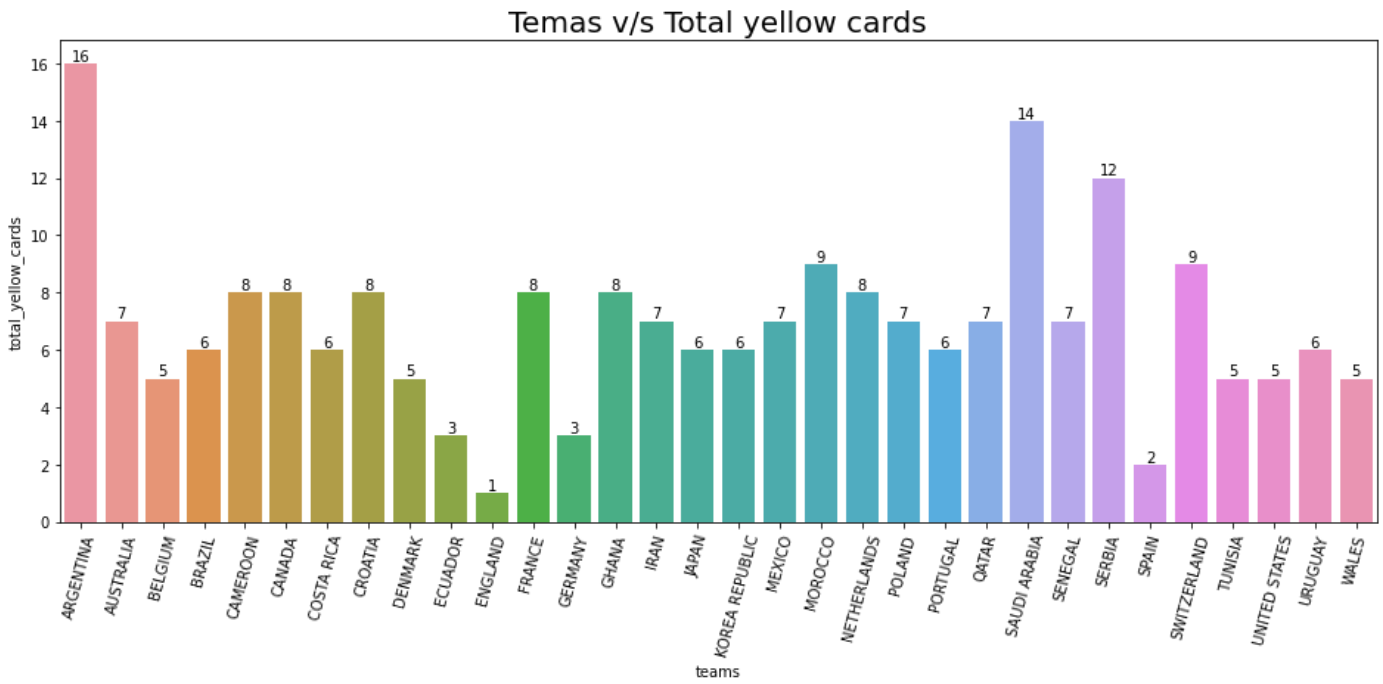
In [21]:
```
# total matches played and yellow cards by the team
team_df['total_yellow_cards']=match_df.groupby(['1'])['1_yellow_cards'].sum().sort_index

# sort on total_yellow_cards
team_df.loc[:,['teams','total_matches','total_yellow_cards']].sort_values(by='total_yell
```

Out[21]:

	teams	total_matches	total_yellow_cards
0	ARGENTINA	7	16
23	SAUDI ARABIA	3	14
25	SERBIA	3	12
18	MOROCCO	7	9
27	SWITZERLAND	4	9
4	CAMEROON	3	8
5	CANADA	3	8
7	CROATIA	7	8
19	NETHERLANDS	5	8
11	FRANCE	7	8
13	GHANA	3	8
20	POLAND	4	7
17	MEXICO	3	7
1	AUSTRALIA	4	7
22	QATAR	3	7
14	IRAN	3	7
24	SENEGAL	4	7
21	PORTUGAL	5	6
30	URUGUAY	3	6
16	KOREA REPUBLIC	4	6
15	JAPAN	4	6
6	COSTA RICA	3	6
3	BRAZIL	5	6
8	DENMARK	3	5
28	TUNISIA	3	5
29	UNITED STATES	4	5
2	BELGIUM	3	5
31	WALES	3	5
12	GERMANY	3	3
9	ECUADOR	3	3
26	SPAIN	4	2
10	ENGLAND	5	1

In [22]:

```
# Bargraph Team v/s Total yellow cards
plt.figure(figsize=(16,6))
ax=sns.barplot(data=team_df,x='teams',y='total_yellow_cards')
ax.bar_label(ax.containers[0])
plt.title('Temas v/s Total yellow cards',fontsize=20)
plt.xticks(rotation=75)
plt.show()
```



From the above we can say that ARGENTINA team got highest yellow cards

In [23]:
```
# Teams with red cards
team_df['total_red_cards']=match_df.groupby(['1'])['1_red_cards'].sum().sort_index().val

# sort on total_red_cards
team_df.loc[team_df['total_red_cards']!=0,['teams','total_matches','total_red_cards']].s
```

Out[23]:

	teams	total_matches	total_red_cards
4	CAMEROON	3	1
18	MOROCCO	7	1
31	WALES	3	1

only three teams(WALES, CAMEROON, MOROCCO) got red cards

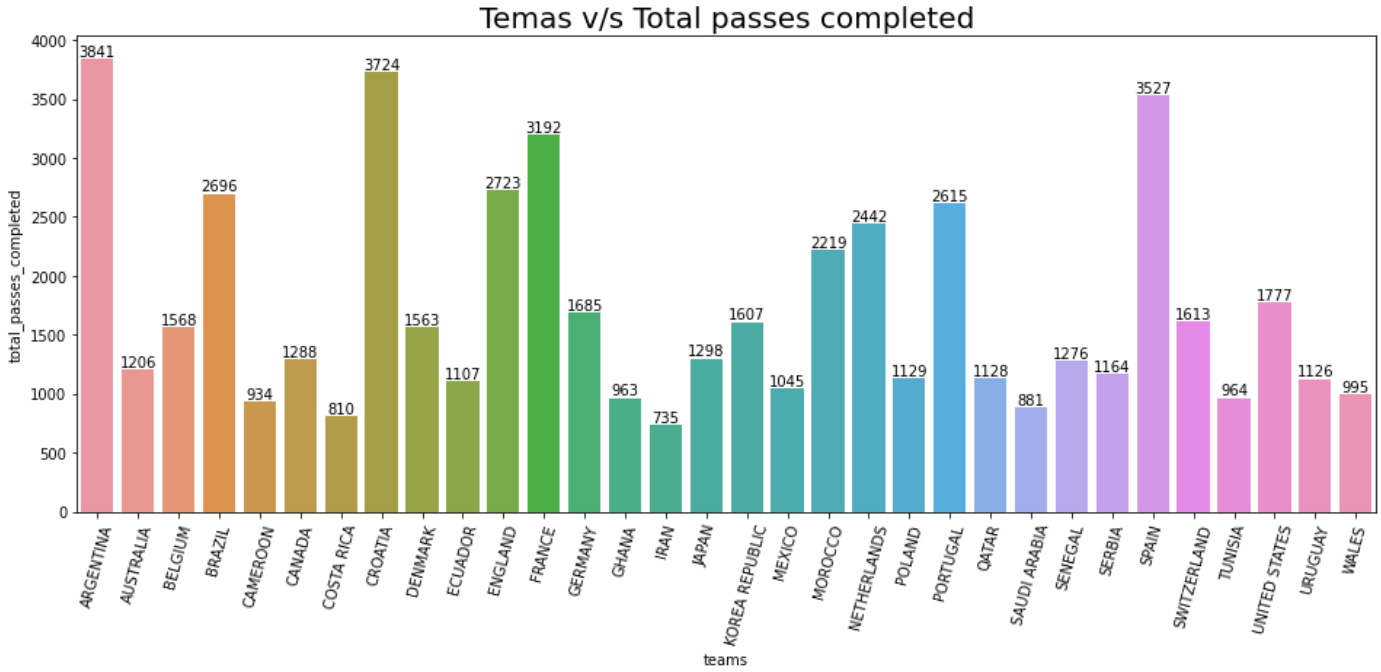
In [24]:
```
# No of matches played and passes completed by the team
team_df['total_passes_completed']=match_df.groupby(['1'])['1_passes_compeletd'].sum().so

# sort on total_passes_completed
team_df.loc[:,['teams','total_matches','total_passes_completed']].sort_values(by='total_
```

Out[24]:

	teams	total_matches	total_passes_completed
0	ARGENTINA	7	3841
7	CROATIA	7	3724
26	SPAIN	4	3527
11	FRANCE	7	3192
10	ENGLAND	5	2723
3	BRAZIL	5	2696
21	PORTUGAL	5	2615
19	NETHERLANDS	5	2442
18	MOROCCO	7	2219
29	UNITED STATES	4	1777
12	GERMANY	3	1685
27	SWITZERLAND	4	1613
16	KOREA REPUBLIC	4	1607
2	BELGIUM	3	1568
8	DENMARK	3	1563
15	JAPAN	4	1298
5	CANADA	3	1288
24	SENEGAL	4	1276
1	AUSTRALIA	4	1206
25	SERBIA	3	1164
20	POLAND	4	1129
22	QATAR	3	1128
30	URUGUAY	3	1126
9	ECUADOR	3	1107
17	MEXICO	3	1045
31	WALES	3	995
28	TUNISIA	3	964
13	GHANA	3	963
4	CAMEROON	3	934
23	SAUDI ARABIA	3	881
6	COSTA RICA	3	810
14	IRAN	3	735

In [25]:

```
# Bargraph Team v/s total passes
plt.figure(figsize=(16,6))
ax=sns.barplot(data=team_df,x='teams',y='total_passes_completed')
ax.bar_label(ax.containers[0])
plt.title('Temas v/s Total passes completed',fontsize=20)
plt.xticks(rotation=75)
plt.show()
```



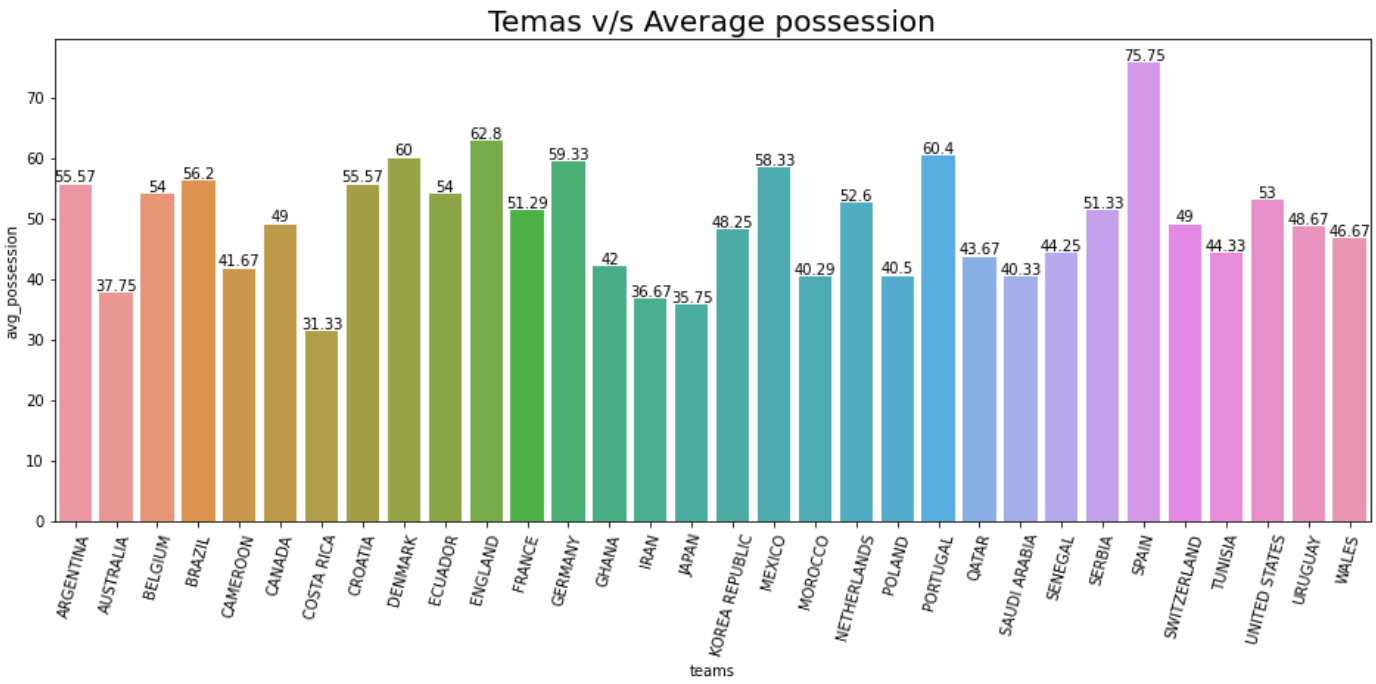
In [26]:
```
# no of matches played and possession by team
team_df['avg_possession']=round((match_df.groupby(['1'])['1_poss'].sum().sort_index().va

# sort on avg_possession
team_df.loc[:,['teams','total_matches','avg_possession']].sort_values(by='avg_possession
```

Out[26]:

	teams	total_matches	avg_possession
26	SPAIN	4	75.75
10	ENGLAND	5	62.80
21	PORTUGAL	5	60.40
8	DENMARK	3	60.00
12	GERMANY	3	59.33
17	MEXICO	3	58.33
3	BRAZIL	5	56.20
0	ARGENTINA	7	55.57
7	CROATIA	7	55.57
2	BELGIUM	3	54.00
9	ECUADOR	3	54.00
29	UNITED STATES	4	53.00
19	NETHERLANDS	5	52.60
25	SERBIA	3	51.33
11	FRANCE	7	51.29
5	CANADA	3	49.00
27	SWITZERLAND	4	49.00
30	URUGUAY	3	48.67
16	KOREA REPUBLIC	4	48.25
31	WALES	3	46.67
28	TUNISIA	3	44.33
24	SENEGAL	4	44.25
22	QATAR	3	43.67
13	GHANA	3	42.00
4	CAMEROON	3	41.67
20	POLAND	4	40.50
23	SAUDI ARABIA	3	40.33
18	MOROCCO	7	40.29
1	AUSTRALIA	4	37.75
14	IRAN	3	36.67
15	JAPAN	4	35.75
6	COSTA RICA	3	31.33

In [27]:

```
# Bargraph Team v/s avg possession

plt.figure(figsize=(16,6))
ax=sns.barplot(data=team_df,x='teams',y='avg_possession')
ax.bar_label(ax.containers[0])
plt.title('Temas v/s Average possession',fontsize=20)
plt.xticks(rotation=75)
plt.show()
```



In [28]:
```
# Teams with own goals
team_df['total_own_goals']=match_df.groupby(['1'])['1_own_goal'].sum().sort_index().valu

# sort on total_own_goals
team_df.loc[team_df['total_own_goals']!=0,['teams','total_matches','total_own_goals']].s
```

Out[28]:

	teams	total_matches	total_own_goals
0	ARGENTINA	7	1
18	MOROCCO	7	1

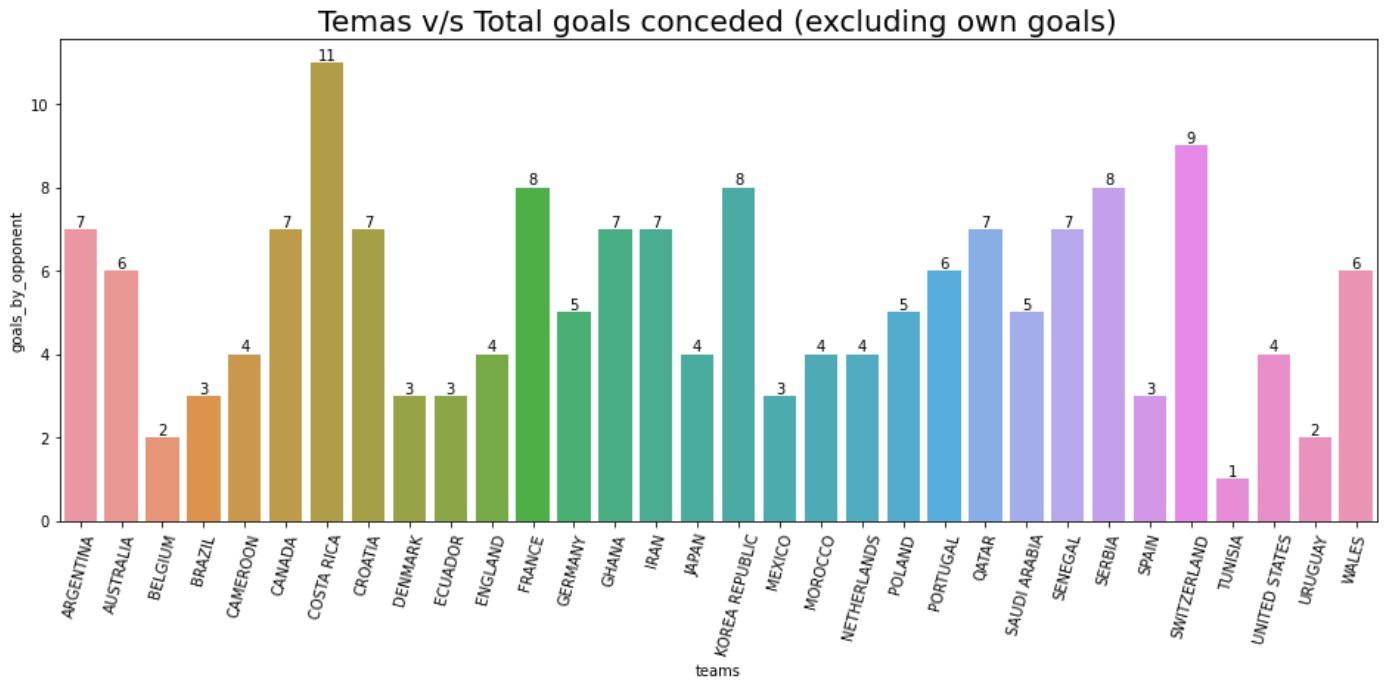
In [29]:
```
# Total games played and goals conceded by the team excluding own goals
team_df['goals_by_opponent']=team_df['goals_conceded']-team_df['total_own_goals']

# sort on goals_by_opponent
team_df.loc[:,['teams','total_matches','goals_conceded','goals_by_opponent']].sort_value
```

Out[29]:

	teams	total_matches	goals_conceded	goals_by_opponent
6	COSTA RICA	3	11	11
27	SWITZERLAND	4	9	9
16	KOREA REPUBLIC	4	8	8
11	FRANCE	7	8	8
25	SERBIA	3	8	8
24	SENEGAL	4	7	7
22	QATAR	3	7	7
14	IRAN	3	7	7
13	GHANA	3	7	7
0	ARGENTINA	7	8	7
7	CROATIA	7	7	7
5	CANADA	3	7	7
21	PORTUGAL	5	6	6
1	AUSTRALIA	4	6	6
31	WALES	3	6	6
23	SAUDI ARABIA	3	5	5
20	POLAND	4	5	5
12	GERMANY	3	5	5
29	UNITED STATES	4	4	4
4	CAMEROON	3	4	4
15	JAPAN	4	4	4
19	NETHERLANDS	5	4	4
18	MOROCCO	7	5	4
10	ENGLAND	5	4	4
9	ECUADOR	3	3	3
8	DENMARK	3	3	3
3	BRAZIL	5	3	3
26	SPAIN	4	3	3
17	MEXICO	3	3	3
2	BELGIUM	3	2	2
30	URUGUAY	3	2	2
28	TUNISIA	3	1	1

In [30]:

```
# Bargraph Temas v/s Total goals conceded(excluding own goal)
plt.figure(figsize=(16,6))
ax=sns.barplot(data=team_df,x='teams',y='goals_by_opponent')
ax.bar_label(ax.containers[0])
plt.title('Temas v/s Total goals conceded (excluding own goals)',fontsize=20)
plt.xticks(rotation=75)
plt.show()
```



Player Statistics

In [31]:
```
player_df.head()
```

Out[31]:

	player	position	team	age	club	birth_year	games	games_starts	minutes	minutes_90s	...
0	Aaron Mooy	MF	Australia	32-094	Celtic	1990	4	4	360	4.0	...
1	Aaron Ramsey	MF	Wales	31-357	Nice	1990	3	3	266	3.0	...
2	Abdelhamid Sabiri	MF	Morocco	26-020	Sampdoria	1996	5	2	181	2.0	...
3	Abdelkarim Hassan	DF	Qatar	29-112	Al Sadd SC	1993	3	3	270	3.0	...
4	Abderrazak Hamdallah	FW	Morocco	32-001	Al-Ittihad	1990	4	0	68	0.8	...

5 rows × 31 columns

In [32]:
```
player_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 680 entries, 0 to 679
Data columns (total 31 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   player                    680 non-null    object
 1   position                  680 non-null    object
 2   team                      680 non-null    object
 3   age                       680 non-null    object
 4   club                      679 non-null    object
 5   birth_year                680 non-null    int64
 6   games                     680 non-null    int64
 7   games_starts              680 non-null    int64
 8   minutes                   680 non-null    int64
 9   minutes_90s               680 non-null    float64
 10  goals                     680 non-null    int64
 11  assists                   680 non-null    int64
 12  goals_pens                680 non-null    int64
 13  pens_made                 680 non-null    int64
 14  pens_att                  680 non-null    int64
 15  cards_yellow              680 non-null    int64
 16  cards_red                 680 non-null    int64
 17  goals_per90               680 non-null    float64
 18  assists_per90             680 non-null    float64
 19  goals_assists_per90       680 non-null    float64
 20  goals_pens_per90          680 non-null    float64
 21  goals_assists_pens_per90  680 non-null    float64
 22  xg                        677 non-null    float64
 23  npxg                      677 non-null    float64
 24  xg_assist                 677 non-null    float64
 25  npxg_xg_assist            677 non-null    float64
 26  xg_per90                  675 non-null    float64
 27  xg_assist_per90           675 non-null    float64
 28  xg_xg_assist_per90        675 non-null    float64
 29  npxg_per90                675 non-null    float64
 30  npxg_xg_assist_per90      675 non-null    float64
dtypes: float64(15), int64(11), object(5)
memory usage: 164.8+ KB
```

In [33]:
```
player_df.isna().sum().sort_values(ascending=False)
```

Out[33]:
```
npxg_xg_assist_per90        5
npxg_per90                  5
xg_xg_assist_per90          5
xg_assist_per90             5
xg_per90                    5
npxg_xg_assist              3
xg_assist                   3
npxg                        3
xg                          3
club                        1
cards_red                   0
goals_assists_pens_per90    0
goals_pens_per90            0
goals_assists_per90         0
assists_per90               0
goals_per90                 0
player                      0
position                    0
pens_att                    0
pens_made                   0
goals_pens                  0
assists                     0
goals                       0
minutes_90s                 0
minutes                     0
games_starts                0
games                       0
birth_year                  0
age                         0
team                        0
cards_yellow                0
dtype: int64
```

In [34]:
```
# Top Goal scorer of WC 2022
player_df.loc[player_df['goals']==player_df['goals'].max(),['player','team','birth_year'
```

Out[34]:

	player	team	birth_year	games	club	goals
377	Kylian Mbappé	France	1998	7	Paris S-G	8

In [35]:
```
# Top 5 Goal scorer of WC 2022
player_df.loc[:,['player','team','birth_year','games','club','goals']].sort_values(by='g
```

Out[35]:

	player	team	birth_year	games	club	goals
377	Kylian Mbappé	France	1998	7	Paris S-G	8
388	Lionel Messi	Argentina	1987	7	Paris S-G	7
330	Julián Álvarez	Argentina	2000	7	Manchester City	4
502	Olivier Giroud	France	1986	6	Milan	4
672	Álvaro Morata	Spain	1992	4	Atlético Madrid	3

In [36]:
```
# Top Goal assist of WC 2022
player_df.loc[player_df['assists']==player_df['assists'].max(),['player','team','club','
```

Out[36]:

	player	team	club	assists	goals
388	Lionel Messi	Argentina	Paris S-G	3	7
103	Bruno Fernandes	Portugal	Manchester Utd	3	2
240	Harry Kane	England	Tottenham	3	2
273	Ivan Perišić	Croatia	Tottenham	3	1
63	Antoine Griezmann	France	Atlético Madrid	3	0

In [37]:
```
# Top 5 players with yellow cards holder
player_df.loc[:,['player','games','club','assists','cards_yellow']].sort_values(by='card
```

Out[37]:

	player	games	club	assists	cards_yellow
407	Marcos Acuña	6	Sevilla	0	3
227	Gonzalo Montiel	4	Sevilla	0	3
603	Strahinja Pavlović	3	RB Salzburg	0	2
366	Ko Itakura	3	M'Gladbach	1	2
276	Jackson Irvine	4	St. Pauli	0	2

In [ ]:

In [ ]: