Women's and Men's FIFA World Cup Data

I collected this data from the wiki page for Womens FIFA World cup and the wiki page for Mens FIFA World Cup. I chose this to use for my data as Wiki already had built in dataframes on the data of the World Cups that I was looking to analyze and compare. The pandas function also makes it easy to read in the webpage and pull the data in into Jupyter Notebook as we learned in class. I made sure however that both pages were sited from getting the information the FIFA World cup websites before fully deciding to use this information.

The purpose of this analysis is to better understand the different skill level of countries in the Womens and Men FIFA World Cup. Part of this will be looking to see if certain confederations or countries are better during the cup in the Female games or in the Male games. Part will be looking at which gets more attendence or attention from the world at this cup to try an identify if one of the FIFA World ups was more fan engaged than another and why. Finding the why by looking at the final cup teams and different factors to see if that might have been a factor.

The starting point for deciding to analzing the different world cups was because I grew up watching them as well as playing the sport. I have always loved watching the womens world cup more than the mens because I personally just felt it was a little more entertaining to watch plus the US Womens National Team always seemed to perform better than the US Mens National team. This sparked my interest in comparing the two cups to see where the Mens team falls on the list of National teams and where the Womens fall. Comparing the past world cups to see the winners of both sides, the amount of attendence at the games, as well as the Confederations, seeing which league is the best overall.

The set up of the data consist of starting with overall FIFA World Cup stats. Starting with Confederation stats, to individual world cup stats and then getting in to each countries stats. This effort of organizing was trying to help display which Confederation had the best records and then from that looking at which specfic team might have had a bigger contribution in to the overall confederation total.

In [2]:
```
import requests
import pandas as pd
import base64
import json
import urllib
import numpy as np
from scipy import stats
from matplotlib import pyplot as plt
```

```
Intel MKL WARNING: Support of Intel(R) Streaming SIMD Extensions 4.2 (Intel(R) SSE4.2) e
nabled only processors has been deprecated. Intel oneAPI Math Kernel Library 2025.0 will
require Intel(R) Advanced Vector Extensions (Intel(R) AVX) instructions.
Intel MKL WARNING: Support of Intel(R) Streaming SIMD Extensions 4.2 (Intel(R) SSE4.2) e
nabled only processors has been deprecated. Intel oneAPI Math Kernel Library 2025.0 will
require Intel(R) Advanced Vector Extensions (Intel(R) AVX) instructions.
```

In [3]:
```
womens_world = pd.read_html('https://en.wikipedia.org/wiki/FIFA_Women%27s_World_Cup')
```

In [4]:
```
mens_world = pd.read_html('https://en.wikipedia.org/wiki/FIFA_World_Cup')
```

In [5]:
```
##womens_world
```

In [6]:
```
type(womens_world)
```

Out[6]:
```
list
```

Confederation Comparisons

The charts below is showing the different conference that compete in the FIFA World Cup and how many times they have won a cup, or if they have hosted in the past and when. Top is womens, bottom is mens

In [7]:
```
womens_world[3]
```

Out[7]:

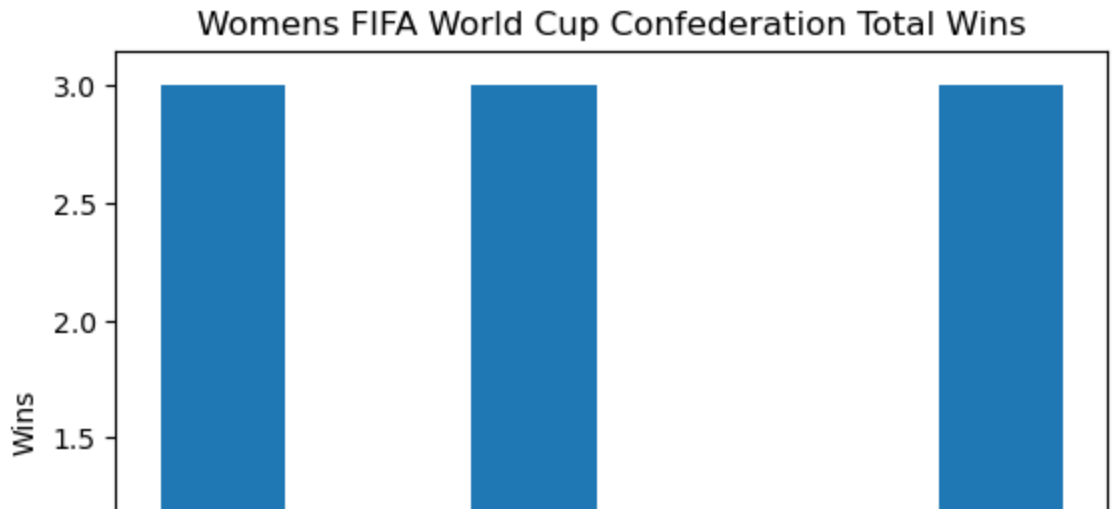
	Confederation	Total	Hosts
0	Asian Football Confederation (AFC)	3	1991: China 2007: China 2023: Australia
1	Confederation of African Football (CAF)	0	NaN
2	Confederation of North, Central America and Ca...	3	1999: United States 2003: United States 2015: ...
3	South American Football Confederation (CONMEBOL)	0	NaN
4	Oceania Football Confederation (OFC)	1	2023: New Zealand
5	Union of European Football Associations (UEFA)	3	1995: Sweden 2011: Germany 2019: France

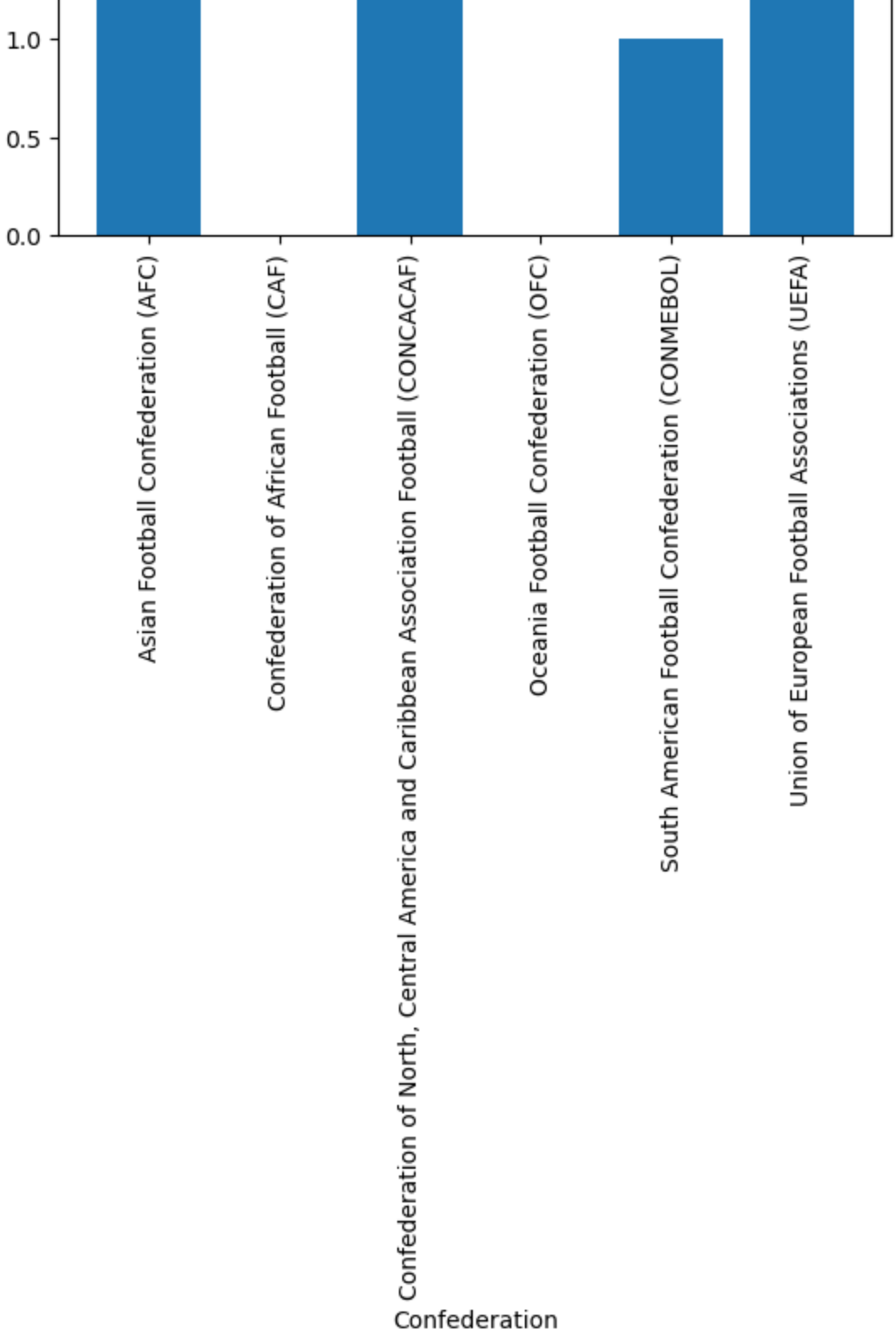
In [8]:
```
womens_conf = womens_world[3].groupby('Confederation').count()['Hosts']
```

In [9]:
```
womens_conf.index
```

Out[9]:
```
Index(['Asian Football Confederation (AFC)',
       'Confederation of African Football (CAF)',
       'Confederation of North, Central America and Caribbean Association Football (CONC
ACAF)',
       'Oceania Football Confederation (OFC)',
       'South American Football Confederation (CONMEBOL)',
       'Union of European Football Associations (UEFA)'],
      dtype='object', name='Confederation')
```

In [10]:
```
fig23, ax23 = plt.subplots()
ax23.bar(x = womens_conf.index, height = womens_world[3]['Total'])

ax23.set_title('Womens FIFA World Cup Confederation Total Wins')
ax23.set_xlabel('Confederation')
ax23.set_ylabel('Wins')
ax23.set_xticks([0,1,2,3,4,5], womens_conf.index, rotation='vertical')
plt.show()
```





In [11]:
```
mens_world[2]
```

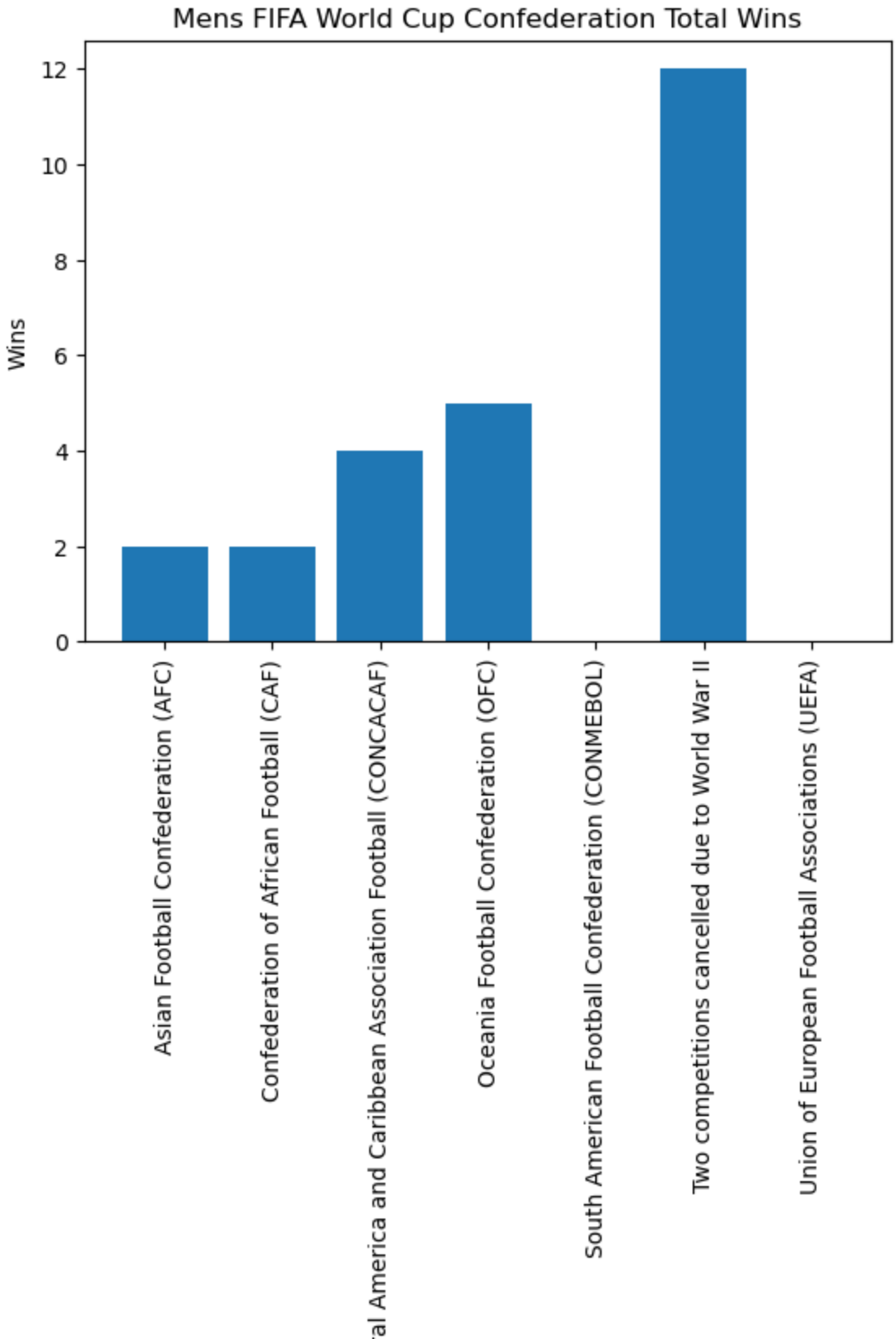
Out[11]:

	Confederation	Total	Hosts
0	Asian Football Confederation (AFC)	2	2002: South Korea, Japan 2022: Qatar 2034: Sau...
1	Confederation of African Football (CAF)	2	2010: South Africa 2030: Morocco
2	Confederation of North, Central America and Ca...	4	1970: Mexico 1986: Mexico 1994: United States ...
3	South American Football Confederation (CONMEBOL)	5	1930: Uruguay 1950: Brazil 1962: Chile 1978: A...
4	Oceania Football Confederation (OFC)	0	NaN
5	Union of European Football Associations (UEFA)	12	1934: Italy 1938: France 1954: Switzerland 195...

6	Two competitions cancelled due to World War II	0	1942 1946

In [12]:
```
mens_conf = mens_world[2].groupby('Confederation').count()['Hosts']
```

In [13]:
```
fig24, ax24 = plt.subplots()
ax24.bar(x = mens_conf.index, height = mens_world[2]['Total'])

ax24.set_title('Mens FIFA World Cup Confederation Total Wins')
ax24.set_xlabel('Confederation')
ax24.set_ylabel('Wins')
ax24.set_xticks([0,1,2,3,4,5,6], mens_conf.index, rotation='vertical')
plt.show()
```



Average Attendance Comparision

The chart below is showcasing the different host countries for the Women's FIFA World Cup, the attendence at each of these events, the venue as well as some games. The chart one below is also showing the same information for the Men's FIFA World Cup as well

In [14]:
```
womens_world[4]
```

Out[14]:

	Year	Hosts	Venues/ Cities	Total attendance †	Matches	Average attendance	Highest attendances		
	Year	Hosts	Venues/ Cities	Total attendance †	Matches	Average attendance	Number	Venue	Game(s)
0	1991	China	6/4	510000	26	18344	65000	Tianhe Stadium, Guangzhou	China PR 4–0 Norway, Opening match
1	1995	Sweden	5/5	112213	26	4316	17158	Råsunda Stadium, Solna	Germany 0–2 Norway, final
2	1999	United States	8/8	1214209	32	37944	90185	Rose Bowl, Pasadena, California	United States 0–0 (5–4p) China PR, final
3	2003	United States	6/6	679664	32	21240	34144	Robert F. Kennedy Memorial Stadium, Washington...	United States 3–1 Sweden, quarter-final
4	2007	China	5/5	1190971	32	37218	55832	Tianjin Olympic Center, Tianjin	China PR 2–0 New Zealand, group stage
5	2011	Germany	9/9	845751	32	26430	73680	Olympiastadion, Berlin	Germany 2–1 Canada, group stage
6	2015	Canada	6/6	1353506	52	26029	54027	BC Place, Vancouver	England 2–1 Canada, quarter-final
7	2019	France	9/9	1131312	52	21756	57900	Parc Olympique Lyonnais, Décines-Charpieu	United States 2–0 Netherlands, final

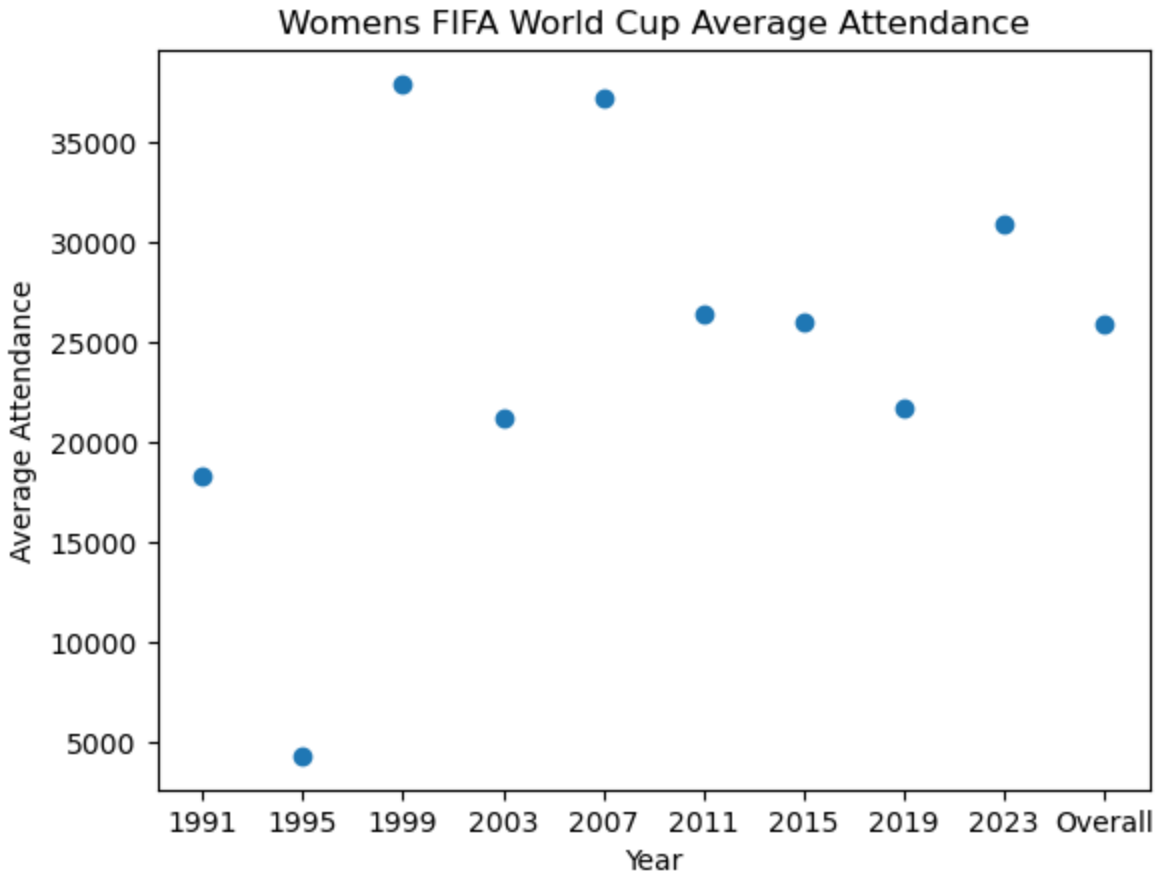
8	2023	Australia New Zealand	10/9	1978274	64	30911	75784	Stadium Australia, Sydney	Five matches, all at Stadium Australia.
9	Overall	Overall	Overall	9015900	348	25908	90185	Rose Bowl, Pasadena (1999)	NaN

Pre work I had to do in order to make the data I wanted to plot work. The scatter plot wasn't liking the "Overall" in the year section so I had to make it not ignore it from the data all together by the lines of code below.

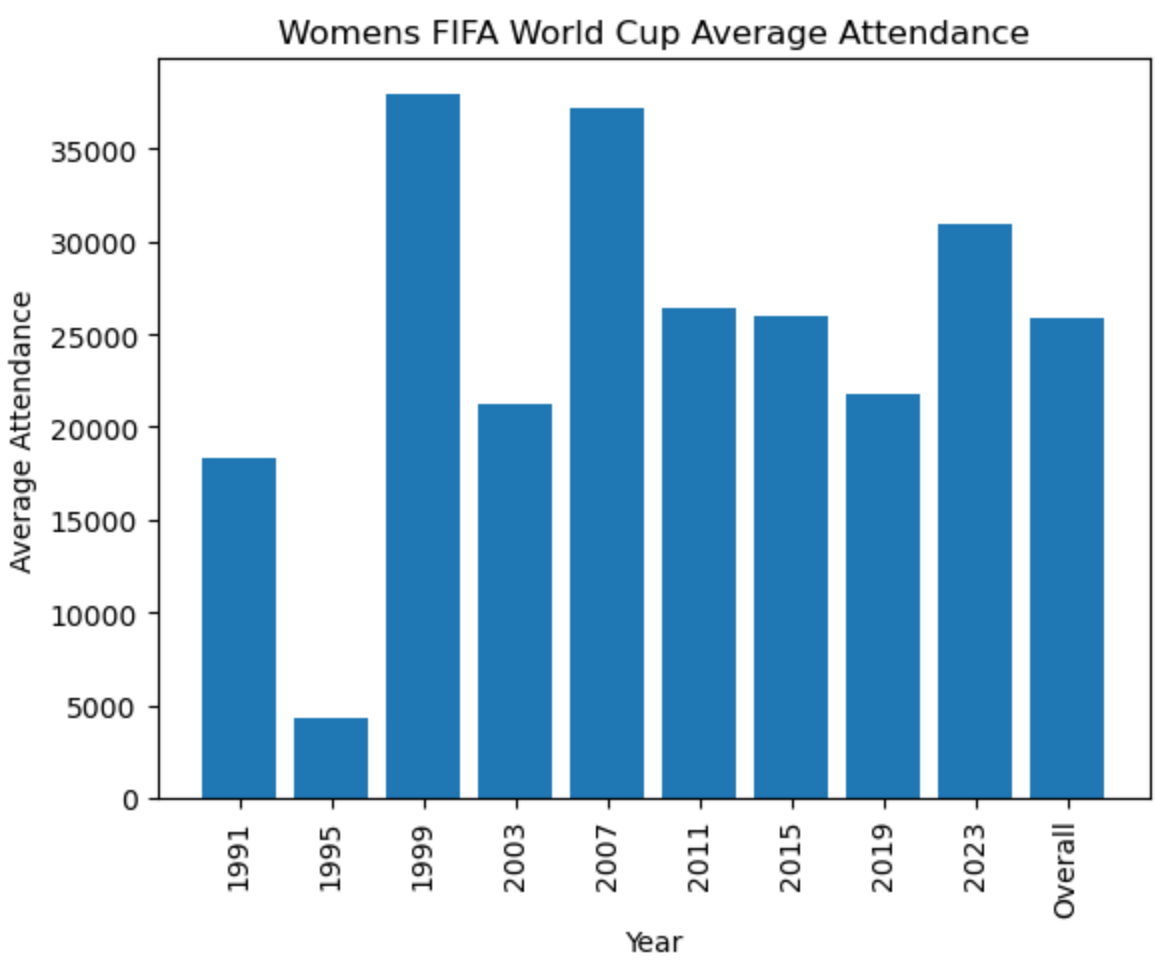
In [15]:
```
womens_world[4].columns = ['Year', 'Hosts', 'Venues/ Cities', 'Total attendance', 'Matc
```

In [16]:
```
test_df = womens_world[4][womens_world[4]['Year'] != 'Overall']
```

In [17]:
```
fig9, ax9 = plt.subplots()
ax9.scatter(x = womens_world[4]['Year'], y = womens_world[4]['Average attendance'])

ax9.set_title('Womens FIFA World Cup Average Attendance')
ax9.set_xlabel('Year')
ax9.set_ylabel('Average Attendance')
plt.show()
```



In [18]:
```
fig10, ax10 = plt.subplots()
ax10.bar(x = womens_world[4]['Year'], height = womens_world[4]['Average attendance'])

ax10.set_title('Womens FIFA World Cup Average Attendance')
ax10.set_xlabel('Year')
ax10.set_ylabel('Average Attendance')
ax10.set_xticks([0,1,2,3,4,5,6,7,8,9], womens_world[4]['Year'], rotation='vertical')
plt.show()
```



In [19]:
```
mens_world[3]
```

Out[19]:

	Year	Hosts	Venues/ Cities	Total attendance †	Matches	Average attendance	Highest attendances ‡		
	Year	Hosts	Venues/ Cities	Total attendance †	Matches	Average attendance	Number	Venue	Game(s)
0	1930	Uruguay	3/1	590549.0	18	32808.0	93000	Estadio Centenario, Montevideo	Uruguay 6–1 Yugoslavia, semi-final
1	1934	Italy	8/8	363000.0	17	21353.0	55000	Stadio Nazionale PNF, Rome	Italy 2–1 Czechoslovakia, final
2	1938	France	10/9	375700.0	18	20872.0	58455	Olympique de Colombes, Paris	France 1–3 Italy, quarter-final
3	1950	Brazil	6/6	1045246.0	22	47511.0	173,850[94]	Maracanã Stadium, Rio de Janeiro	Brazil 1–2 Uruguay, deciding match
4	1954	Switzerland	6/6	768607.0	26	29562.0	63000	Wankdorf Stadium, Bern	West Germany 3–2 Hungary, final
5	1958	Sweden	12/12	819810.0	35	23423.0	50928	Ullevi Stadium, Gothenburg	Brazil 2–0 Soviet Union, group stage
6	1962	Chile	4/4	893172.0	32	27912.0	68679	Estadio Nacional,	Brazil 4–2 Chile, semi-final

								Santiago	
7	1966	England	8/7	1563135.0	32	48848.0	98270	Wembley Stadium, London	England 4–2 West Germany, final
8	1970	Mexico	5/5	1603975.0	32	50124.0	108192	Estadio Azteca, Mexico City	Mexico 1–0 Belgium, group stage
9	1974	West Germany	9/9	1865753.0	38	49099.0	83168	Olympiastadion, Munich	West Germany 1–0 Chile, group stage
10	1978	Argentina	6/5	1545791.0	38	40679.0	71712	Estadio Monumental, Buenos Aires	Italy 1–0 Argentina, group stage
11	1982	Spain	17/14	2109723.0	52	40572.0	95500	Camp Nou, Barcelona	Argentina 0–1 Belgium, Opening match
12	1986	Mexico	12/11	2394031.0	52	46039.0	114600	Estadio Azteca, Mexico City	Mexico 1–1 Paraguay, group stage Argentina 3–2...
13	1990	Italy	12/12	2516215.0	52	48389.0	74765	San Siro, Milan	West Germany 4–1 Yugoslavia, group stage
14	1994	United States	9/9	3587538.0	52	68991.0	94194	Rose Bowl, Pasadena, California	Brazil 0–0 (3–2p) Italy, final
15	1998	France	10/10	2785100.0	64	43517.0	80000	Stade de France, Saint-Denis	Brazil 0–3 France, final
16	2002	South Korea Japan	20/20	2705197.0	64	42269.0	69029	International Stadium, Yokohama, Japan	Brazil 2–0 Germany, final
17	2006	Germany	12/12	3359439.0	64	52491.0	72000	Olympiastadion, Berlin	Germany 1–1 (4–2p) Argentina, quarter-final
18	2010	South Africa	10/9	3178856.0	64	49670.0	84490	Soccer City, Johannesburg	Spain 1–0 Netherlands, final
19	2014	Brazil	12/12	3429873.0	64	53592.0	74738	Maracanã Stadium, Rio de Janeiro	Germany 1–0 Argentina, final
20	2018	Russia	12/11	3031768.0	64	47371.0	78011	Luzhniki Stadium, Moscow	France 4–2 Croatia, final
21	2022	Qatar	8/5	3404252.0	64	53191.0	88966	Lusail Stadium, Qatar	Argentina 3–3 (4–2p) France, final
22	2026	Canada Mexico United States	16/16	NaN	104	NaN	NaN	NaN	NaN

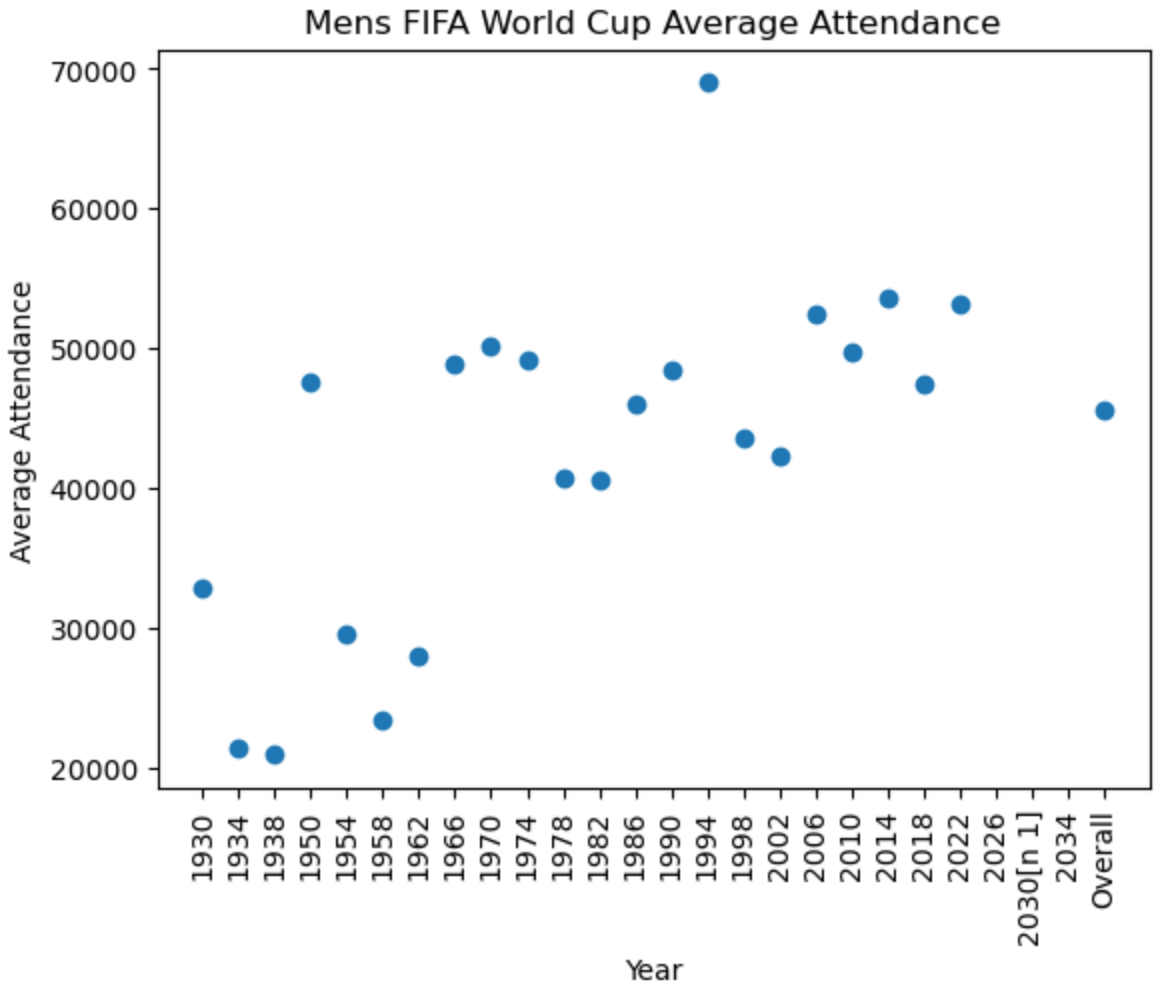
23	2030[n 1]	Morocco Portugal Spain	NaN	NaN	104	NaN	NaN	NaN	NaN
24	2034	Saudi Arabia	NaN	NaN	104	NaN	NaN	NaN	NaN
25	Overall	Overall	Overall	43936730.0	964	45577.0	173,850[94]	Maracanã Stadium, Rio de Janeiro (1950)	NaN

Pre work I had to do in order to make the data I wanted to plot work. The scatter plot wasn't liking the "Overall" in the year section so I had to make it not ignore it from the data all together by the lines of code below.

In [20]:
```
mens_world[3].columns = ['Year', 'Hosts', 'Venues/ Cities', 'Total attendance', 'Matches
```

In [21]:
```
mens_df = mens_world[3][mens_world[3]['Year'] != 'Overall']
```

In [22]:
```
fig11, ax11 = plt.subplots()
ax11.scatter(x = mens_world[3]['Year'], y = mens_world[3]['Average attendance'])

ax11.set_title('Mens FIFA World Cup Average Attendance')
ax11.set_xlabel('Year')
ax11.set_ylabel('Average Attendance')
ax11.set_xticks([0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25], m
plt.show()
```

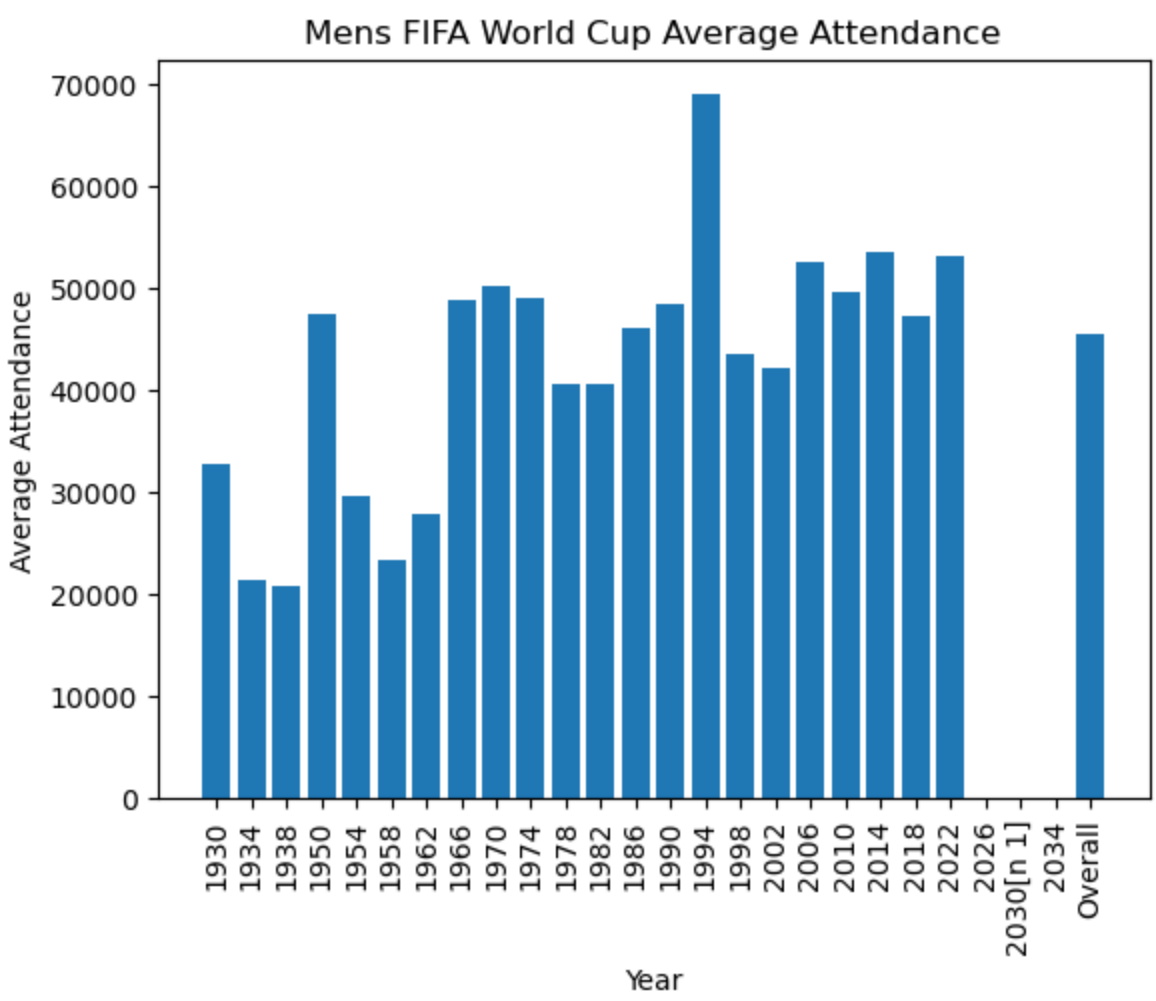


Taking the NaN out of the mens data chart

In [23]:
```
mens_world[3]['Total attendance'].dropna()
```

Out[23]:
```
0       590549.0
1       363000.0
2       375700.0
3      1045246.0
4       768607.0
5       819810.0
6       893172.0
7      1563135.0
8      1603975.0
9      1865753.0
10     1545791.0
11     2109723.0
12     2394031.0
13     2516215.0
14     3587538.0
15     2785100.0
16     2705197.0
17     3359439.0
18     3178856.0
19     3429873.0
20     3031768.0
21     3404252.0
25    43936730.0
Name: Total attendance, dtype: float64
```

In [50]:
```
fig26, ax26 = plt.subplots()
ax26.bar(x = mens_world[3]['Year'], height = mens_world[3]['Average attendance'])

ax26.set_title('Mens FIFA World Cup Average Attendance')
ax26.set_xlabel('Year')
ax26.set_ylabel('Average Attendance')
ax26.set_xticks([0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25], m
plt.show()
```



From comparing the average attendence of the Women's and Men's World Cup in a histagram, we can se that the mens average attendance is higher than the the women's. The average highest the mens has ever been was when it was held in the US in California at 68991 in the year 1994. The average highest the women's has ever been is 3991 and once again it was held in the United States in 1999. I feel that this shows that the area the Cup is held does matter for attendence, however as the I do also think from looking at the Total Attendence, that just as the Cup continues and grows, the Total Attendence has also been increasing no matter what.

Comparision of Champions and Runner ups

The chart below is showcasing the winners and runnerups of the Women's FIFA World Cup, as well as the score of the final game and third place game. It also list the number of teams that where at each tournment throughout the years which allows to see how the tournment has grown from when it first started. The chart right below is showing the same information for the Men's FIFA World Cup as well.

In [25]:
```
womens_world[5]
```

Out[25]:

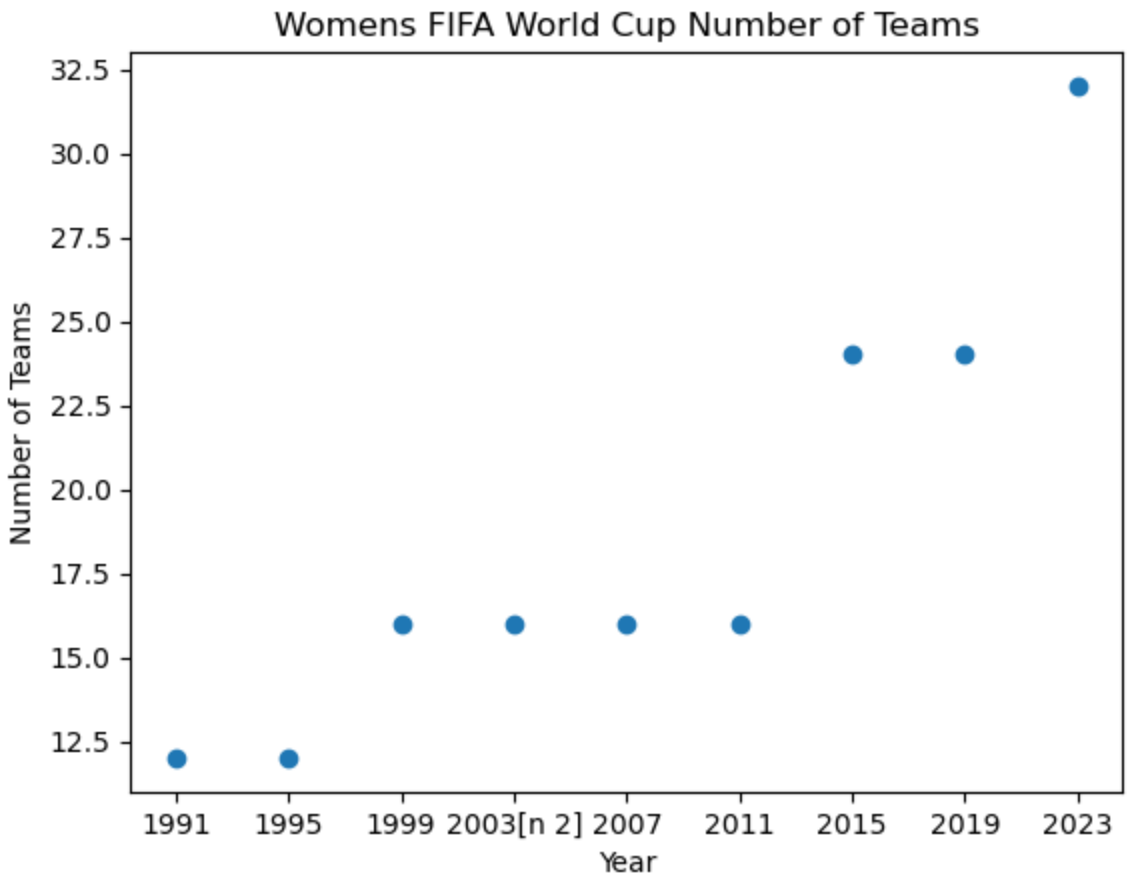
	Ed.	Year	Hosts	Final			Third-place playoff		No. of teams
	Ed.	Year	Hosts	Champions	Score	Runners-up	Third place	Score	Fourth place	No. of teams
0	1	1991	China	United States	2–1 Tianhe Stadium,	Norway	Sweden	4–0 Provincial Stadium,	Germany	12

					Guangzhou			Guangzhou		
1	2	1995	Sweden	Norway	2–0 Råsunda Stadium, Solna	Germany	United States	2–0 Strömvallen, Gävle	China	12
2	3	1999	United States	United States	0–0 (a.e.t.) (5–4 p) Rose Bowl, Pasadena	China	Brazil	0–0[n 1] (5–4 p) Rose Bowl, Pasadena	Norway	16
3	4	2003[n 2]	United States	Germany	2–1 (a.e.t.) Home Depot Center, Carson	Sweden	United States	3–1 Home Depot Center, Carson	Canada	16
4	5	2007	China	Germany	2–0 Hongkou Stadium, Shanghai	Brazil	United States	4–1 Hongkou Stadium, Shanghai	Norway	16
5	6	2011	Germany	Japan	2–2 (a.e.t.) (3–1 p) Commerzbank-Arena, Frankfurt	United States	Sweden	2–1 Rhein-Neckar-Arena, Sinsheim	France	16
6	7	2015	Canada	United States	5–2 BC Place, Vancouver	Japan	England	1–0 (a.e.t.) Commonwealth Stadium, Edmonton	Germany	24
7	8	2019	France	United States	2–0 Parc Olympique Lyonnais, Lyon	Netherlands	Sweden	2–1 Allianz Riviera, Nice	England	24
8	9	2023	Australia New Zealand	Spain	1–0 Stadium Australia, Sydney	England	Sweden	2–0 Lang Park, Brisbane	Australia	32

Below I am changing the column names again in order to call a single column from the chart in an easier way.

In [26]:
```
womens_world[5].columns = ['Ed.', 'Year', 'Host', 'Final Champions', 'Score', 'Runners-u
```

In [27]:
```
fig18, ax18 = plt.subplots()
ax18.scatter(x = womens_world[5]['Year'], y = womens_world[5]['No. of teams'])

ax18.set_title('Womens FIFA World Cup Number of Teams')
ax18.set_xlabel('Year')
ax18.set_ylabel('Number of Teams')
plt.show()
```



This chart is showcasing the growth of the teams in the Women's World Cup. Starting out with 12 teams for two years before increasing gradually to the most recent Cup when for the first time it have 32 teams. If looking down at the mens plot, we can see that the womens still have less teams in the Cup than the mens, but they haven't been as many World Cup's yet for the womens. At nine cups in for the men's they were still only at 16 teams in the games. So the women are showing a faster growth of teams than the mens.

In [28]:
```
mens_world[4]
```

Out[28]:

	Ed.	Year	Host	Final			Third-place play-off	
	Ed.	Year	Host	Champion	Score	Runner-up	Third	Score	Fourth
0	1	1930	Uruguay	Uruguay	4–2 Estadio Centenario, Montevideo	Argentina	United States	– [n 2]	Yugoslavia
1	2	1934	Italy	Italy	2–1 (a.e.t.) Stadio Nazionale PNF, Rome	Czechoslovakia	Germany	3–2 Stadio Giorgio Ascarelli, Naples	Austria
2	3	1938	France	Italy	4–2 Stade de Colombes, Paris	Hungary	Brazil	4–2 Parc Lescure, Bordeaux	Sweden
3	–	1942	(Not held because of World War II)	(Not held because of World War II)	(Not held because of World War II)	(Not held because of World War II)	(Not held because of World War II)	(Not held because of World War II)	(Not held because of World War II)
4	–	1946	(Not held because of	(Not held because of	(Not held because of World	(Not held because of	(Not held because of	(Not held because of	(Not held because of

			World War II)	World War II)	War II)	World War II)	World War II)	World War II)	World War II)
5	4	1950	Brazil	Uruguay	2–1 [n 3] Maracanã, Rio de Janeiro	Brazil	Sweden	3–1 [n 3] Pacaembu, São Paulo	Spain
6	5	1954	Switzerland	West Germany	3–2 Wankdorfstadion, Bern	Hungary	Austria	3–1 Hardturm, Zürich	Uruguay
7	6	1958	Sweden	Brazil	5–2 Råsundastadion, Solna	Sweden	France	6–3 Ullevi, Gothenburg	West Germany
8	7	1962	Chile	Brazil	3–1 Estadio Nacional, Santiago	Czechoslovakia	Chile	1–0 Estadio Nacional, Santiago	Yugoslavia
9	8	1966	England	England	4–2 (a.e.t.) Wembley Stadium, London	West Germany	Portugal	2–1 Wembley Stadium, London	Soviet Union
10	9	1970	Mexico	Brazil	4–1 Estadio Azteca, Mexico City	Italy	West Germany	1–0 Estadio Azteca, Mexico City	Uruguay
11	10	1974	West Germany	West Germany	2–1 Olympiastadion, Munich	Netherlands	Poland	1–0 Olympiastadion, Munich	Brazil
12	11	1978	Argentina	Argentina	3–1 (a.e.t.) Monumental de Núñez, Buenos Aires	Netherlands	Brazil	2–1 Monumental de Núñez, Buenos Aires	Italy
13	12	1982	Spain	Italy	3–1 Santiago Bernabéu, Madrid	West Germany	Poland	3–2 Estadio José Rico Pérez, Alicante	France
14	13	1986	Mexico	Argentina	3–2 Estadio Azteca, Mexico City	West Germany	France	4–2 (a.e.t.) Estadio Cuauhtémoc, Puebla	Belgium
15	14	1990	Italy	West Germany	1–0 Stadio Olimpico, Rome	Argentina	Italy	2–1 Stadio San Nicola, Bari	England
16	15	1994	United States	Brazil	0–0 (a.e.t.) (3–2 p) Rose Bowl, Pasadena	Italy	Sweden	4–0 Rose Bowl, Pasadena	Bulgaria
17	16	1998	France	France	3–0 Stade de France, Saint-Denis	Brazil	Croatia	2–1 Parc des Princes, Paris	Netherlands
18	17	2002	South Korea Japan	Brazil	2–0 International Stadium, Yokohama	Germany	Turkey	3–2 Daegu Stadium, Daegu	South Korea
19	18	2006	Germany	Italy	1–1 (a.e.t.) (5–3 p) Olympiastadion, Berlin	France	Germany	3–1 Gottlieb-Daimler-Stadion, Stuttgart	Portugal
20	19	2010	South Africa	Spain	1–0 (a.e.t.) Soccer City, Johannesburg	Netherlands	Germany	3–2 Nelson Mandela Bay	Uruguay

									Stadium, Port Elizabeth
21	20	2014	Brazil	Germany	1–0 (a.e.t.) Maracanã, Rio de Janeiro	Argentina	Netherlands	3–0 Estádio Nacional, Brasília	Brazil
22	21	2018	Russia	France	4–2 Luzhniki Stadium, Moscow	Croatia	Belgium	2–0 Krestovsky Stadium, Saint Petersburg	England
23	22	2022	Qatar	Argentina	3–3 (a.e.t.) (4–2 p) Lusail Stadium, Lusail	France	Croatia	2–1 Khalifa International Stadium, Al Rayyan	Morocco
24	23	2026	Canada Mexico United States	NaN	NaN	NaN	NaN	NaN	NaN
25	24	2030[n 1]	Morocco Portugal Spain	NaN	NaN	NaN	NaN	NaN	NaN
26	25	2034	Saudi Arabia	NaN	NaN	NaN	NaN	NaN	NaN

Pre work I had to do in order to make the data I wanted to plot work. The scatter plot wasn't liking the "(Not helod because of World War II)" in the year section so I had to make it not ignore it from the data all together by the lines of code below.

I also had to change the column names of the chart above to make it easier in call which column I wanted to plot.

In [29]:
```
mens_world[4].columns = ['Ed.', 'Year', 'Host', 'Final Champions', 'Score', 'Runners-up'
```

Checking to see how to go about getting "(Not held because of World War II") to not show up in the chart

In [31]:
```
type(mens_world[4]['Final Champions'].iloc[24])
```

Out[31]:
```
float
```

Dropping the NaN in the chart

In [56]:
```
mens_world[4]['Final Champions'].dropna()
```

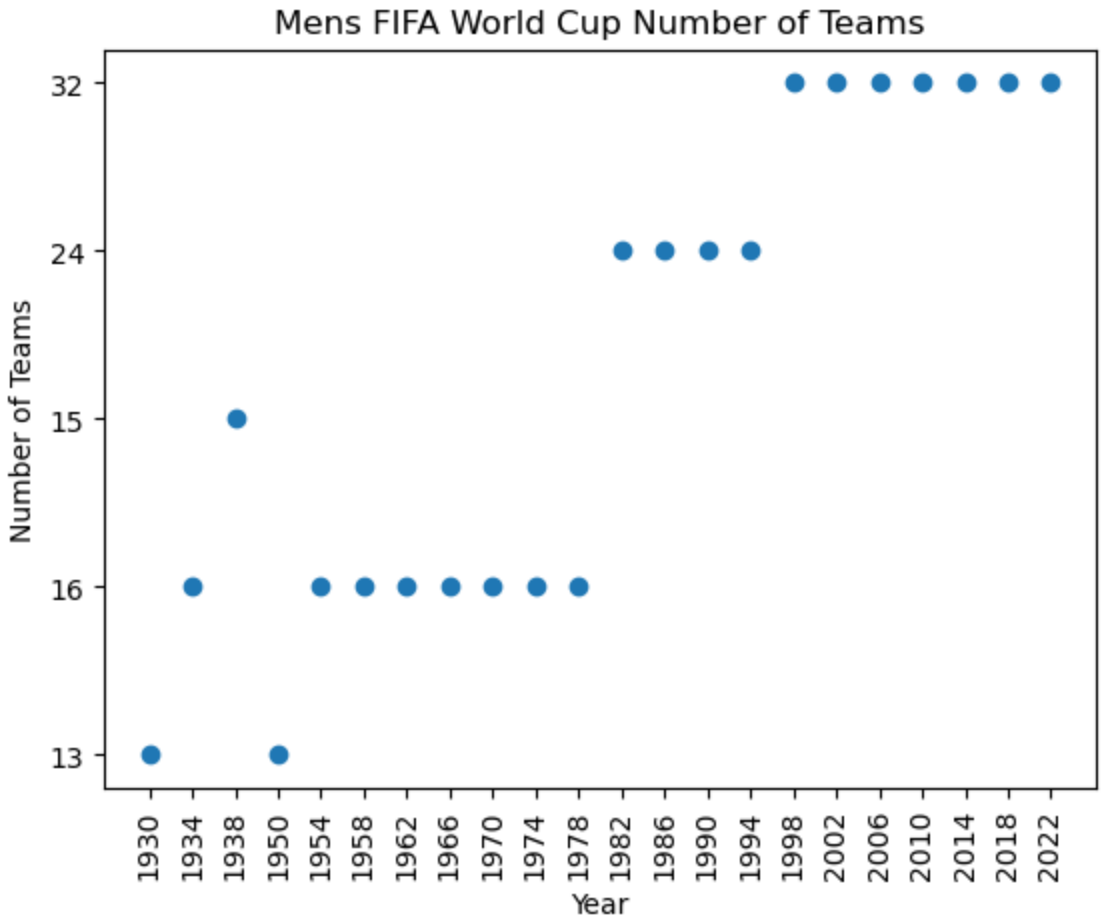
Out[56]:
```
0                                Uruguay
1                                  Italy
2                                  Italy
3     (Not held because of World War II)
4     (Not held because of World War II)
5                                Uruguay
6                           West Germany
7                                 Brazil
8                                 Brazil
9                                England
10                                Brazil
11                          West Germany
12                             Argentina
13                                 Italy
14                             Argentina
```

```
15                          West Germany
16                                Brazil
17                                France
18                                Brazil
19                                 Italy
20                                 Spain
21                               Germany
22                                France
23                             Argentina
Name: Final Champions, dtype: object
```

In [64]:
```
mens_world[4] = mens_world[4].dropna()
```

In [65]:
```
mens_world[4] = mens_world[4][mens_world[4]['Final Champions'] != '(Not held because of
```

In [66]:
```
fig19, ax19 = plt.subplots()
ax19.scatter(x = mens_world[4]['Year'], y = mens_world[4]['No. of teams'])

ax19.set_title('Mens FIFA World Cup Number of Teams')
ax19.set_xlabel('Year')
ax19.set_ylabel('Number of Teams')
ax19.set_xticks([0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21], mens_world[4]
plt.show()
```



The chart above is showcasing the how the number of teams has grow throughout the years for the Mens World Cup. There is an error in this plot for the two dots that line with "(Not held because of World War II)". These plots shouldn't be there as they aren't actually showcasing the correct information. With this being said though, there was still a drop in teams right after war, right before the war the teams was at 15 but after the war it jumped back down to 13 where it first started. This could have been from countries still recovering from the war and not having as many people focusing on the sport as before. However the

following years it was held it did jump back up to being at 16 which it was at two world cups before the war. From this point the cup's teams has just been increasing in size all the way to 48.

Pre work I had to do in order to plot the Final Champions and how many they won in to a bar plot to get a better visual of home many championships each country has won. I had to do this for the men and for the women.

In [35]:
```
womens_final = womens_world[5].groupby('Final Champions').count()['Year']
```

In [36]:
```
womens_final.index
```

Out[36]:
```
Index(['Germany', 'Japan', 'Norway', 'Spain', 'United States'], dtype='object', name='Final Champions')
```

In [67]:
```
mens_final = mens_world[4].groupby('Final Champions').count()['Year']
```

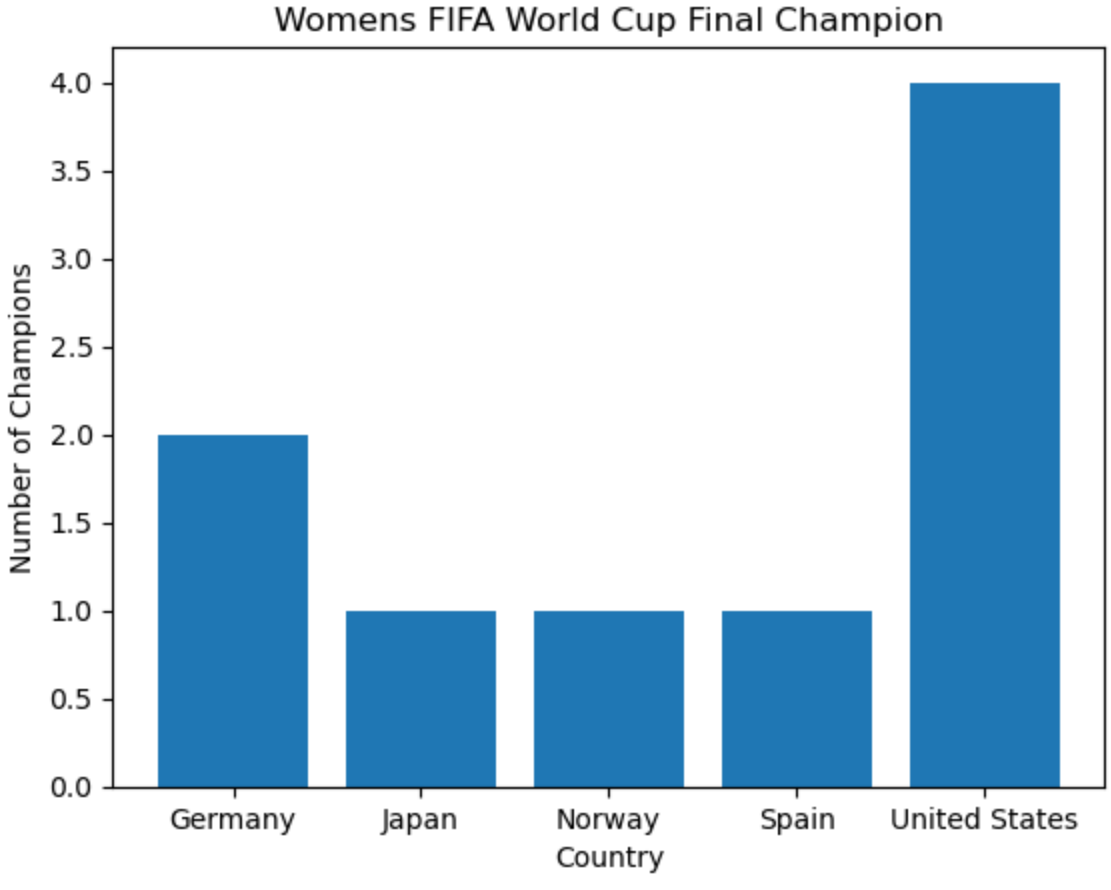
In [68]:
```
mens_final.index
```

Out[68]:
```
Index(['Argentina', 'Brazil', 'England', 'France', 'Germany', 'Italy', 'Spain',
       'Uruguay', 'West Germany'],
      dtype='object', name='Final Champions')
```

In [69]:
```
mens_final.values
```

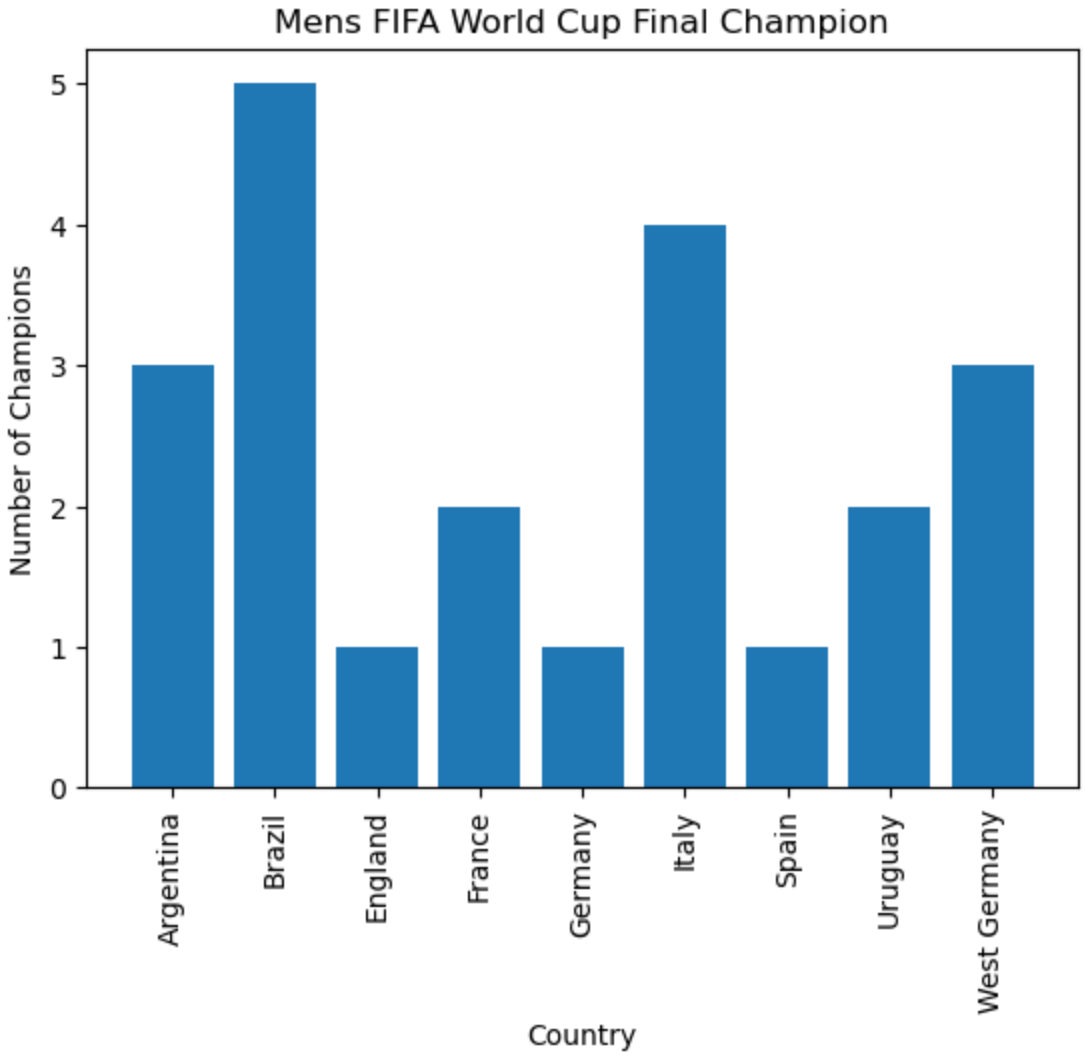
Out[69]:
```
array([3, 5, 1, 2, 1, 4, 1, 2, 3])
```

In [40]:
```
fig22, ax22 = plt.subplots()
ax22.bar(x = womens_final.index, height = womens_final.values)

ax22.set_title('Womens FIFA World Cup Final Champion')
ax22.set_xlabel('Country')
ax22.set_ylabel('Number of Champions')
plt.show()
```



From the chart above, we can see that the United States Women's National team has won the most World Cup's. What we can also see from the set is that the only main winners of the World Cup for the women have been between the countries listed in the plot, Germany, Japan, Norway, Spain, and United States.

In [70]:
```
fig21, ax21 = plt.subplots()
ax21.bar(x = mens_final.index, height = mens_final.values)

ax21.set_title('Mens FIFA World Cup Final Champion')
ax21.set_xlabel('Country')
ax21.set_ylabel('Number of Champions')
ax21.set_xticks([0,1,2,3,4,5,6,7,8], mens_final.index, rotation='vertical')
plt.show()
```



Unlike the women's we can see that more teams have one at least one World Cup throughout the years of the tournment. Brazil with the leading about of whens at 5 and Italy next with 4. We can also see that the US Men's National team as never won a World Cup Title.

Comparisions of countries and who has won the most tournments

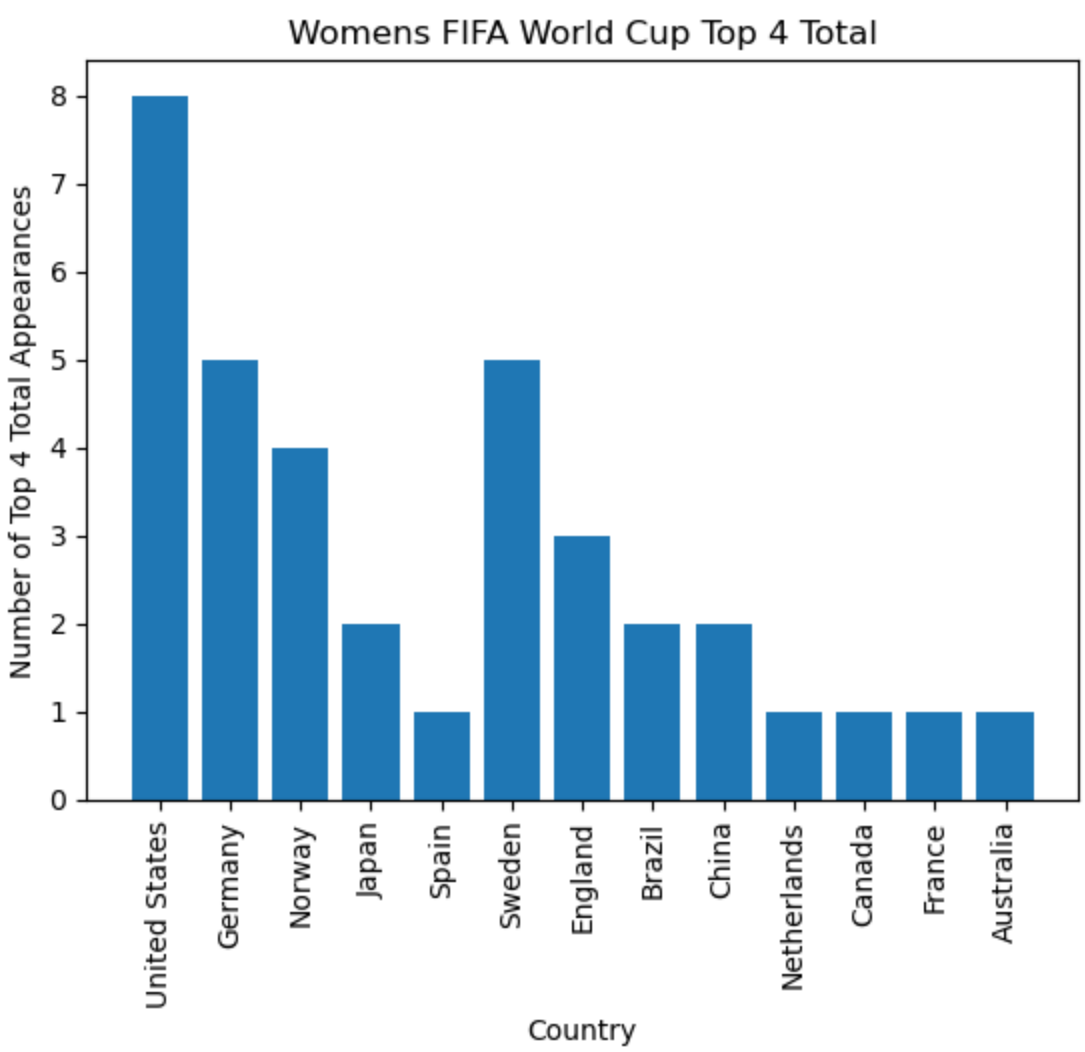
The charts below is showing how many times each country has shown up in the top four and what year they did from most Titles to least Titles . What I am using the charts below for however is the last column, "Top 4 total" to see how many times the countries have showed up in the top four in the tournment.

In [42]:
```
womens_world[6]
```

Out[42]:

	Team	Title(s)	Runners-up	Third place	Fourth place	Top 4 total
0	United States	4 (1991, 1999*, 2015, 2019)	1 (2011)	3 (1995, 2003*, 2007)	NaN	8
1	Germany	2 (2003, 2007)	1 (1995)	NaN	2 (1991, 2015)	5
2	Norway	1 (1995)	1 (1991)	NaN	2 (1999, 2007)	4
3	Japan	1 (2011)	1 (2015)	NaN	NaN	2
4	Spain	1 (2023)	NaN	NaN	NaN	1
5	Sweden	NaN	1 (2003)	4 (1991, 2011, 2019, 2023)	NaN	5
6	England	NaN	1 (2023)	1 (2015)	1 (2019)	3
7	Brazil	NaN	1 (2007)	1 (1999)	NaN	2
8	China	NaN	1 (1999)	NaN	1 (1995)	2
9	Netherlands	NaN	1 (2019)	NaN	NaN	1
10	Canada	NaN	NaN	NaN	1 (2003)	1
11	France	NaN	NaN	NaN	1 (2011)	1
12	Australia	NaN	NaN	NaN	1 (2023*)	1

In [43]:
```
fig25, ax25 = plt.subplots()
ax25.bar(x = womens_world[6]['Team'], height = womens_world[6]['Top 4 total'])

ax25.set_title('Womens FIFA World Cup Top 4 Total')
ax25.set_xlabel('Country')
ax25.set_ylabel('Number of Top 4 Total Appearances')
ax25.set_xticks([0,1,2,3,4,5,6,7,8,9,10,11,12], womens_world[6]['Team'], rotation='verti
plt.show()
```



The bar plot I feel like did the best job of displaying the information of the top 4 in the best way, allowing you to clearly see how many times each country has made it. Once again, we can see the United States has made the most showings at eight followed by Germany and Sweden, which are two of the top champions behind the US as well.

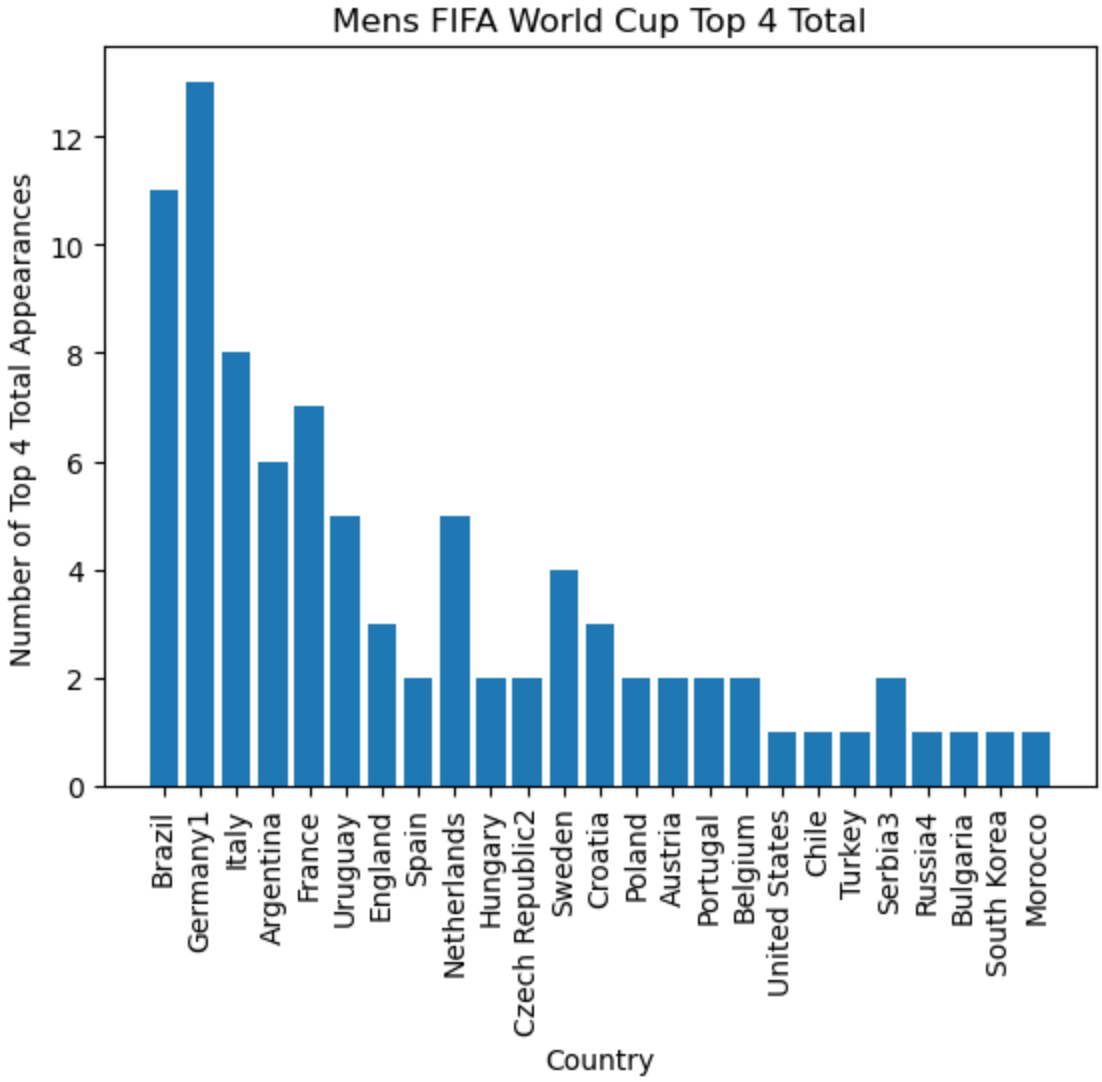
In [44]:
```
mens_world[5]
```

Out[44]:

	Team	Titles	Runners-up	Third place	Fourth place	Top 4 total
0	Brazil	5 (1958, 1962, 1970, 1994, 2002)	2 (1950*, 1998)	2 (1938, 1978)	2 (1974, 2014*)	11
1	Germany1	4 (1954, 1974*, 1990, 2014)	4 (1966, 1982, 1986, 2002)	4 (1934, 1970, 2006*, 2010)	1 (1958)	13
2	Italy	4 (1934*, 1938, 1982, 2006)	2 (1970, 1994)	1 (1990*)	1 (1978)	8
3	Argentina	3 (1978*, 1986, 2022)	3 (1930, 1990, 2014)	NaN	NaN	6
4	France	2 (1998*, 2018)	2 (2006, 2022)	2 (1958, 1986)	1 (1982)	7
5	Uruguay	2 (1930*, 1950)	NaN	NaN	3 (1954, 1970, 2010)	5
6	England	1 (1966*)	NaN	NaN	2 (1990, 2018)	3
7	Spain	1 (2010)	NaN	NaN	1 (1950)	2
8	Netherlands	NaN	3 (1974, 1978, 2010)	1 (2014)	1 (1998)	5

9	Hungary	NaN	2 (1938, 1954)	NaN	NaN	2
10	Czech Republic2	NaN	2 (1934, 1962)	NaN	NaN	2
11	Sweden	NaN	1 (1958*)	2 (1950, 1994)	1 (1938)	4
12	Croatia	NaN	1 (2018)	2 (1998, 2022)	NaN	3
13	Poland	NaN	NaN	2 (1974, 1982)	NaN	2
14	Austria	NaN	NaN	1 (1954)	1 (1934)	2
15	Portugal	NaN	NaN	1 (1966)	1 (2006)	2
16	Belgium	NaN	NaN	1 (2018)	1 (1986)	2
17	United States	NaN	NaN	1 (1930)	NaN	1
18	Chile	NaN	NaN	1 (1962*)	NaN	1
19	Turkey	NaN	NaN	1 (2002)	NaN	1
20	Serbia3	NaN	NaN	NaN	2 (1930, 1962)	2
21	Russia4	NaN	NaN	NaN	1 (1966)	1
22	Bulgaria	NaN	NaN	NaN	1 (1994)	1
23	South Korea	NaN	NaN	NaN	1 (2002*)	1
24	Morocco	NaN	NaN	NaN	1 (2022)	1

In [45]:
```
fig8, ax8 = plt.subplots()
ax8.bar(x = mens_world[5]['Team'], height = mens_world[5]['Top 4 total'])

ax8.set_title('Mens FIFA World Cup Top 4 Total')
ax8.set_xlabel('Country')
ax8.set_ylabel('Number of Top 4 Total Appearances')
ax8.set_xticks([0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24], mens_
plt.show()
```



The bar plot I feel like did the best job of displaying the information of the top 4 in the best way, allowing you to clearly see how many times each country has made it. For this chart, we do final see the US Mens team make the list from there first appearence at the very first games when they made third place that year. The other semi surpising part about this chart is that Germany is higher than Brazil in the most appearances even though Brazil has the most Championship wins and the highest goals scored. I do think this might be from Brazil recent struggles in the cup while Germany is still playing strong.

Comparision of which country has had the most goals

The chart below is for the number of goals each country has scored in the FIFA World Cup ranked from most to least. By looking at these charts we are able to see who are the more active countries when it comes to goal scoring as well as is another way to look at who a better or rising country might be in the cup.

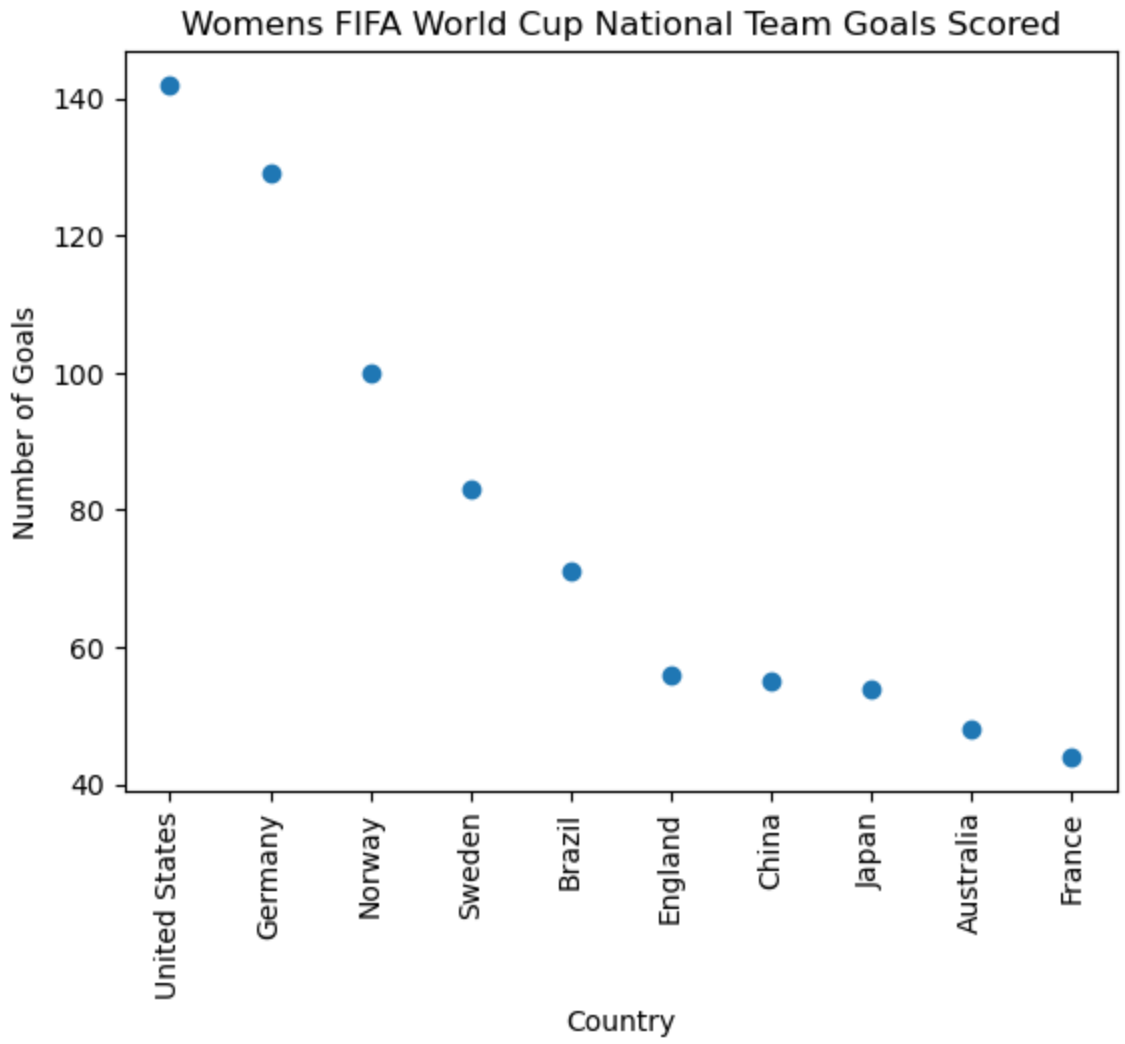
In [46]:
```
womens_world[9]
```

Out[46]:

	Rank	Country	Goals scored
0	1	United States	142
1	2	Germany	129
2	3	Norway	100
3	4	Sweden	83
4	5	Brazil	71

5	6	England	56
6	7	China	55
7	8	Japan	54
8	9	Australia	48
9	10	France	44

In [47]:
```
fig16, ax16 = plt.subplots()
ax16.scatter(x = womens_world[9]['Country'], y = womens_world[9]['Goals scored'])

ax16.set_title('Womens FIFA World Cup National Team Goals Scored')
ax16.set_xlabel('Country')
ax16.set_ylabel('Number of Goals')
ax16.set_xticks([0,1,2,3,4,5,6,7,8,9], womens_world[9]['Country'], rotation='vertical')
plt.show()
```



From this scatter plot, we are able to see that the United States is again at the top with the most goals scored. This fits with the idea of most wins equaling more goal scores. This could partly contribe to the fact that the farther you get in a tournment the more you have chances and opportunities to score compared to getting out after the group stages. We can also see that Germany is number two at the most goals scored, which once again has had previous victories in the cup as shown in the plots above they have won 5 World Cups. Norway being at third for this shows the rising talent that I have meantioned as they don't have as many wins, but in recent world cups they have been becoming a bigger competitor.

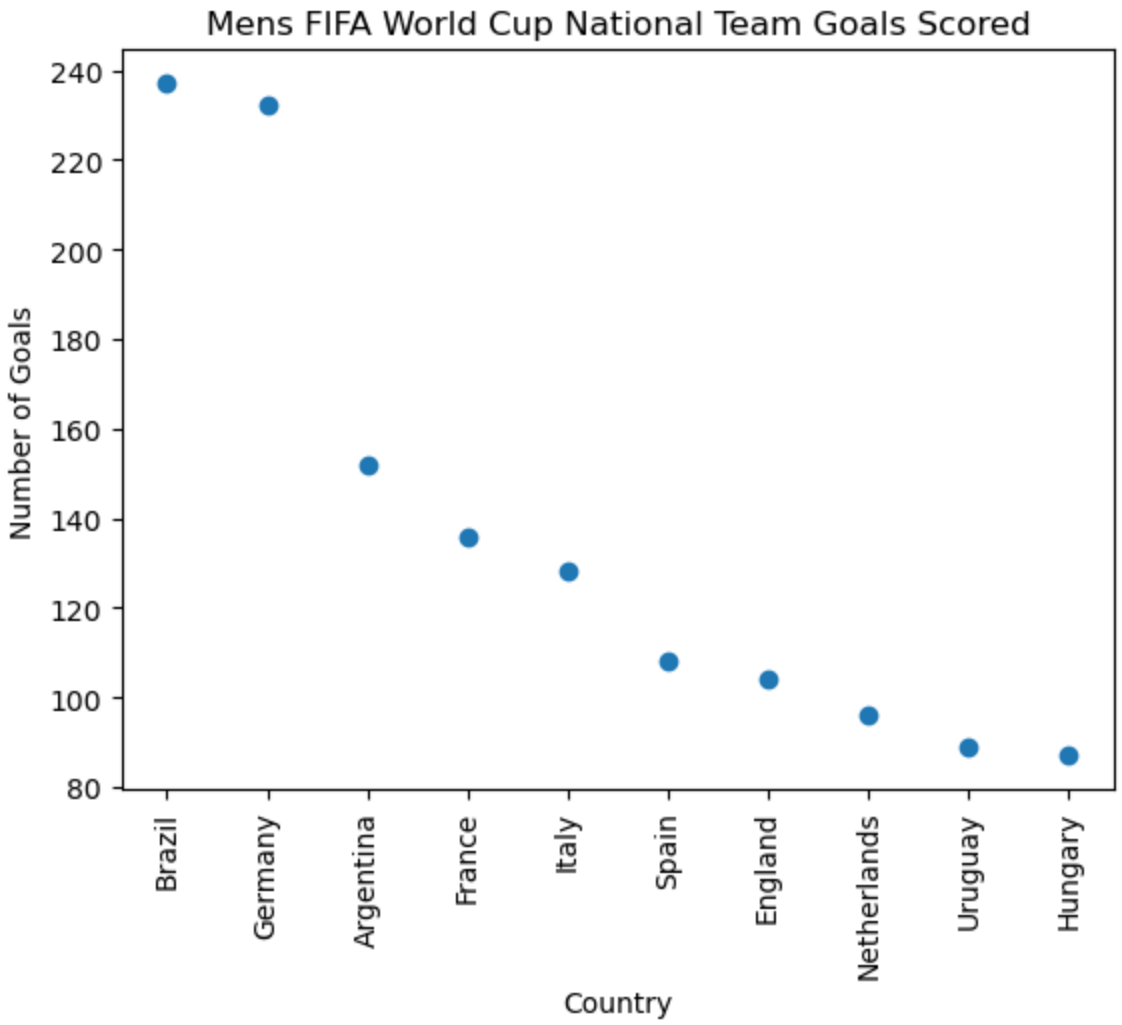
In [48]:
```
mens_world[8]
```

Out[48]:

	Rank	National team	Goals scored
0	1	Brazil	237
1	2	Germany	232
2	3	Argentina	152
3	4	France	136
4	5	Italy	128
5	6	Spain	108
6	7	England	104
7	8	Netherlands	96
8	9	Uruguay	89
9	10	Hungary	87

In [49]:

```
fig17, ax17 = plt.subplots()
ax17.scatter(x = mens_world[8]['National team'], y = mens_world[8]['Goals scored'])

ax17.set_title('Mens FIFA World Cup National Team Goals Scored')
ax17.set_xlabel('Country')
ax17.set_ylabel('Number of Goals')
ax17.set_xticks([0,1,2,3,4,5,6,7,8,9], mens_world[8]['National team'], rotation='vertica
plt.show()
```



From this scatter plot, we are able to see that Brazil is at the top with the most goals scored. This fits with

the idea of most wins equaling more goal scores. Once again, this could partly contribe to the fact that the farther you get in a tournment the more you have chances and opportunities to score compared to getting out after the group stages. We can also see that Germany is number two at the most goals scored. This would like it goes against my orginal thinking because when looking at the chart above, Germany has only won one World Cup, however, West Germany has won four. This plot here I believe has combined the total. If I am won though, and it is just Germany goals, then it does again go against was I was orginally thinking of how the most World Cup wins equals the most goals.

What I learned

I have learned that the US Women's team is a lot better than the US Men's team, that when lookinf for a top competition for the Men's Cup you have to look at either Germany, Brazil, Italy or others. One of the limitations I had is not having enough data on how the US Mens did in the Cups, I do know that there was one year they didn't even qualify for it, so having information like that would have been interesting to look at as well. I also feel that the information limits knowing exactly where current talent in the Cup lands as the US Women's didn't perform very well in the 2023 Cup but they still look like they are the best in their league from these charts and plots. I feel that a new question would be wanting to look more into the rising stars in the Cups and what is causing them to now start rising instead of earlier.

Conclusion

Something the audience can take away from this report is, when comparing the US Women's and US Men's National team, the womens have repeated success in the Cup while the men have one had notable success in the very first games from the knowledge of these charts and plots. Another point you can take away is that the Men's World Cup does tend to draw more attention to it's games than the womens cup does. Lastly, another point to take away is that I do believe at one point location of the Cup matter for the amount of attendence, however now in recent year due to either the Cup getting more popular or travel getting easier, the attendence has been high no matter where it has been held.

In [ ]: