

```
In [1]: import numpy as np # numerical computing
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import matplotlib.pyplot as plt #visualization
import seaborn as sns #modern visualization
```

```
In [2]: data = pd.read_csv('ContinousDataset.csv')
```

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

Out[3]:

	Unnamed: 0	Scorecard	Team 1	Team 2	Margin	Ground	Match Date	Winner	Host_Count
0	0	ODI # 1	Australia	England	Winner2ndInning	Melbourne	Jan 5, 1971	Australia	Austra
1	1	ODI # 2	England	Australia	Winner2ndInning	Manchester	Aug 24, 1972	England	Engla
2	2	ODI # 3	England	Australia	Winner2ndInning	Lord's	Aug 26, 1972	Australia	Engla
3	3	ODI # 4	England	Australia	Winner2ndInning	Birmingham	Aug 28, 1972	England	Engla
4	4	ODI # 5	New Zealand	Pakistan	Winner1stInning	Christchurch	Feb 11, 1973	New Zealand	New Zeala

```
In [4]: data['Team 1']=='England'
```

Out[4]:

0	False
1	True
2	True
3	True
4	False
...	
7489	False
7490	False
7491	False
7492	False
7493	False

Name: Team 1, Length: 7494, dtype: bool

```
In [5]: data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7494 entries, 0 to 7493
Data columns (total 13 columns):
#   Column          Non-Null Count  Dtype
---
```

```
0  Unnamed: 0      7494 non-null  int64
1  Scorecard      7494 non-null  object
2  Team 1         7494 non-null  object
3  Team 2         7494 non-null  object
4  Margin         7494 non-null  object
5  Ground         7494 non-null  object
6  Match Date     7494 non-null  object
7  Winner         7494 non-null  object
8  Host_Country   7364 non-null  object
9  Venue_Team1    7494 non-null  object
10 Venue_Team2    7494 non-null  object
11 Innings_Team1  7494 non-null  object
12 Innings_Team2  7494 non-null  object
dtypes: int64(1), object(12)
memory usage: 761.2+ KB
```

In [6]:

```
#Checking uniques values in every column
data.nunique()
```

Out[6]:

```
Unnamed: 0      3746
Scorecard      3747
Team 1          23
Team 2          23
Margin           2
Ground         161
Match Date     3231
Winner          23
Host_Country    19
Venue_Team1      3
Venue_Team2      3
Innings_Team1    2
Innings_Team2    2
dtype: int64
```

In [7]:

```
df1=data.apply(lambda x:x[data['Team 1']=='England'])
```

In [8]:

```
df2=data.apply(lambda x:x[data['Team 2']=='England'])
```

In [9]:

```
df1
```

Out[9]:

	Unnamed: 0	Scorecard	Team 1	Team 2	Margin	Ground	Match Date	Winner	Host_Co
1	1	ODI # 2	England	Australia	Winner2ndInning	Manchester	Aug 24, 1972	England	Eng
2	2	ODI # 3	England	Australia	Winner2ndInning	Lord's	Aug 26, 1972	Australia	Eng
3	3	ODI # 4	England	Australia	Winner2ndInning	Birmingham	Aug 28, 1972	England	Eng
5	5	ODI # 6	England	New Zealand	Winner2ndInning	Swansea	Jul 18, 1973	England	Eng

	Unnamed: 0	Scorecard	Team 1	Team 2	Margin	Ground	Match Date	Winner	Host_Co
6	6	ODI # 8	England	West Indies	Winner2ndInning	Leeds	Sep 5, 1973	England	Eng
...
7386	3644	ODI # 3821	England	India	Winner1stInning	Cuttack	Jan 19, 2017	India	
7389	3647	ODI # 3824	England	India	Winner1stInning	Kolkata	Jan 22, 2017	England	
7410	3668	ODI # 3845	England	West Indies	Winner1stInning	North Sound	Mar 3, 2017	England	West I
7413	3671	ODI # 3848	England	West Indies	Winner2ndInning	North Sound	Mar 5, 2017	England	West I
7414	3672	ODI # 3849	England	West Indies	Winner1stInning	Bridgetown	Mar 9, 2017	England	West I

664 rows × 13 columns



In [10]:

```
df2
```

Out[10]:

	Unnamed: 0	Scorecard	Team 1	Team 2	Margin	Ground	Match Date	Winner	Host_Co
0	0	ODI # 1	Australia	England	Winner2ndInning	Melbourne	Jan 5, 1971	Australia	Al
14	14	ODI # 16	Australia	England	Winner2ndInning	Melbourne	Jan 1, 1975	England	Al
41	41	ODI # 45	Pakistan	England	Winner2ndInning	Sahiwal	Dec 23, 1977	England	P
42	42	ODI # 46	Pakistan	England	Winner2ndInning	Sialkot	Dec 30, 1977	England	P
43	43	ODI # 47	Pakistan	England	Winner1stInning	Lahore	Jan 13, 1978	Pakistan	P
...
7450	3708	ODI # 3889	Pakistan	England	Winner2ndInning	Cardiff	Jun 14, 2017	Pakistan	E
7470	3728	ODI # 3911	West Indies	England	Winner2ndInning	Manchester	Sep 19, 2017	England	E
7473	3731	ODI # 3915	West Indies	England	Winner1stInning	Bristol	Sep 24, 2017	England	E

	Unnamed: 0	Scorecard	Team 1	Team 2	Margin	Ground	Match Date	Winner	Host_C
7474	3732	ODI # 3916	West Indies	England	Winner1stInning	The Oval	Sep 27, 2017	England	E
7476	3734	ODI # 3918	West Indies	England	Winner2ndInning	Southampton	Sep 29, 2017	England	E

664 rows × 13 columns



```
In [11]: frames=[df1,df2]
         result = pd.concat(frames)

In [12]: data = result.reset_index(drop=True)

In [13]: # data['Match Date'].str.split(" ", n=1, expand = True)
         data['Year'] = data['Match Date'].str.split(" ", n=1, expand = True)[1]

In [14]: #data['Year'] = pd.to_datetime(data['Year'])
         data['Year'] = pd.to_numeric(data['Year'])

In [15]: data
```

Out[15]:

	Unnamed: 0	Scorecard	Team 1	Team 2	Margin	Ground	Match Date	Winner	Host_C
0	1	ODI # 2	England	Australia	Winner2ndInning	Manchester	Aug 24, 1972	England	E
1	2	ODI # 3	England	Australia	Winner2ndInning	Lord's	Aug 26, 1972	Australia	E
2	3	ODI # 4	England	Australia	Winner2ndInning	Birmingham	Aug 28, 1972	England	E
3	5	ODI # 6	England	New Zealand	Winner2ndInning	Swansea	Jul 18, 1973	England	E
4	6	ODI # 8	England	West Indies	Winner2ndInning	Leeds	Sep 5, 1973	England	E
...
1323	3708	ODI # 3889	Pakistan	England	Winner2ndInning	Cardiff	Jun 14, 2017	Pakistan	E

	Unnamed: 0	Scorecard	Team 1	Team 2	Margin	Ground	Match Date	Winner	Host_C
1324	3728	ODI # 3911	West Indies	England	Winner2ndInning	Manchester	Sep 19, 2017	England	E
1325	3731	ODI # 3915	West Indies	England	Winner1stInning	Bristol	Sep 24, 2017	England	E
1326	3732	ODI # 3916	West Indies	England	Winner1stInning	The Oval	Sep 27, 2017	England	E
1327	3734	ODI # 3918	West Indies	England	Winner2ndInning	Southampton	Sep 29, 2017	England	E

1328 rows × 14 columns



```
In [16]: # #Selecting three years data from 2015,2016 and 2017
# data = data.apply(Lambda x:x[data['Year']>2014])
# data
```

```
In [17]: data.sort_values(by='Scorecard')
```

	Unnamed: 0	Scorecard	Team 1	Team 2	Margin	Ground	Match Date	Winner	Ho
373	0	ODI # 1	England	Australia	Winner2ndInning	Melbourne	Jan 5, 1971	Australia	
664	0	ODI # 1	Australia	England	Winner2ndInning	Melbourne	Jan 5, 1971	Australia	
127	959	ODI # 1000	England	West Indies	Winner2ndInning	Nottingham	May 24-25, 1995	West Indies	
1082	959	ODI # 1000	West Indies	England	Winner2ndInning	Nottingham	May 24-25, 1995	West Indies	
128	960	ODI # 1001	England	West Indies	Winner1stInning	The Oval	May 26, 1995	England	
...	
1080	922	ODI # 962	Zimbabwe	England	Winner1stInning	Sydney	Dec 15, 1994	Zimbabwe	
126	928	ODI # 968	England	Zimbabwe	Winner1stInning	Brisbane	Jan 7, 1995	England	

	Unnamed: 0	Scorecard	Team 1	Team 2	Margin	Ground	Match Date	Winner	Ho
1081	928	ODI # 968	Zimbabwe	England	Winner1stInning	Brisbane	Jan 7, 1995	England	
759	929	ODI # 969	Australia	England	Winner1stInning	Melbourne	Jan 10, 1995	England	
468	929	ODI # 969	England	Australia	Winner1stInning	Melbourne	Jan 10, 1995	England	

1328 rows × 14 columns



In [18]:

```
#Checking number of duplicate rows
data.Scorecard.duplicated().sum()
```

Out[18]: 664

In [19]:

```
data=data.loc[data.Scorecard.duplicated('first'), :]
```

In [20]:

```
data
```

Out[20]:

	Unnamed: 0	Scorecard	Team 1	Team 2	Margin	Ground	Match Date	Winner	Host_C
664	0	ODI # 1	Australia	England	Winner2ndInning	Melbourne	Jan 5, 1971	Australia	Al
665	14	ODI # 16	Australia	England	Winner2ndInning	Melbourne	Jan 1, 1975	England	Al
666	41	ODI # 45	Pakistan	England	Winner2ndInning	Sahiwal	Dec 23, 1977	England	P
667	42	ODI # 46	Pakistan	England	Winner2ndInning	Sialkot	Dec 30, 1977	England	P
668	43	ODI # 47	Pakistan	England	Winner1stInning	Lahore	Jan 13, 1978	Pakistan	P
...
1323	3708	ODI # 3889	Pakistan	England	Winner2ndInning	Cardiff	Jun 14, 2017	Pakistan	E
1324	3728	ODI # 3911	West Indies	England	Winner2ndInning	Manchester	Sep 19, 2017	England	E
1325	3731	ODI # 3915	West Indies	England	Winner1stInning	Bristol	Sep 24, 2017	England	E

	Unnamed: 0	Scorecard	Team 1	Team 2	Margin	Ground	Match Date	Winner	Host_Co
1326	3732	ODI # 3916	West Indies	England	Winner1stInning	The Oval	Sep 27, 2017	England	E
1327	3734	ODI # 3918	West Indies	England	Winner2ndInning	Southampton	Sep 29, 2017	England	E

664 rows × 14 columns



```
In [21]: #data['Year'].value_counts()
print("Number of Matches Played in these years", data['Year'].value_counts())
```

Number of Matches Played in these years		2007	34
1987	31		
2011	27		
2015	25		
2014	25		
2003	24		
2002	23		
2009	22		
1999	21		
2000	20		
2004	20		
1983	20		
2013	20		
1996	19		
2017	19		
2006	19		
2005	18		
2008	17		
1992	17		
2010	17		
2016	16		
1990	14		
2012	14		
2001	13		
1997	13		
1979	13		
1998	12		
1993	11		
1985	11		
1994	10		
1984	10		
1988	9		
1991	8		
1989	8		
1986	8		
1980	8		
1982	7		
1981	7		
1978	5		
1977	5		
1975	5		
1995	5		

```

1974      4
1972      3
1973      3
1976      3
1971      1
Name: Year, dtype: int64

```

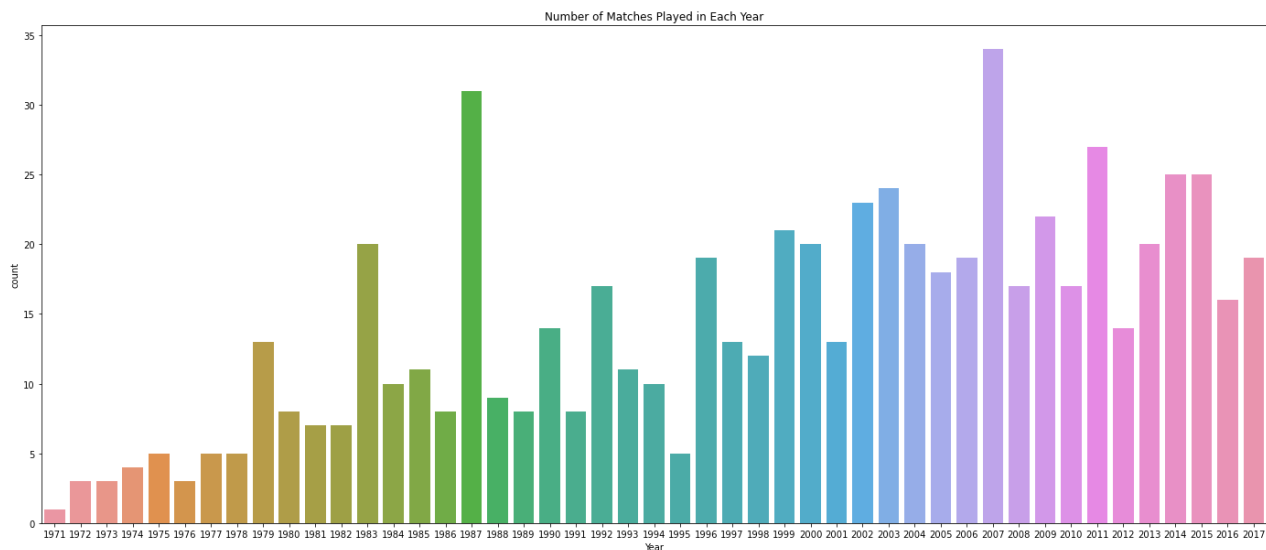
In [22]:

```

#Number of matches played in these Years
plt.figure(figsize=(24,10))
sns.countplot(x='Year', data=data)

plt.title("Number of Matches Played in Each Year")
plt.show()

```



In [23]:

```

#Checking against which team our team won and whole record
print("Number of Wins in these years", data['Winner'].value_counts())

```

```

Number of Wins in these years England      343
Australia      80
India           52
West Indies     42
New Zealand     41
Sri Lanka       34
Pakistan        30
South Africa    29
Zimbabwe        8
Bangladesh      4
Ireland         1
Name: Winner, dtype: int64

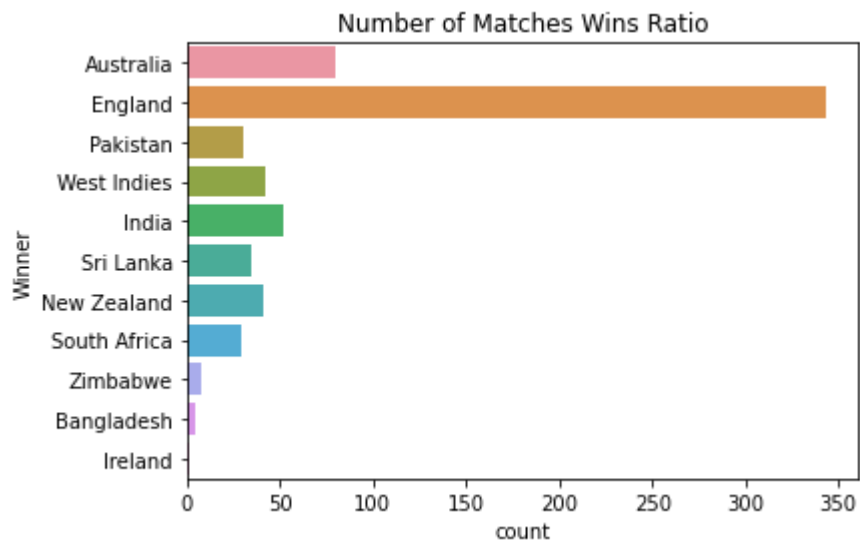
```

In [24]:

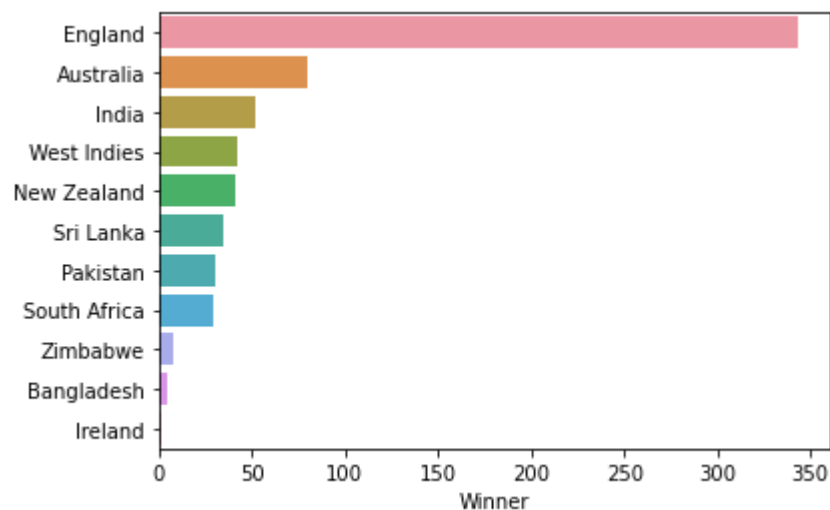
```

#Number of matches played in these Years
sns.countplot(y='Winner', data=data)
plt.title("Number of Matches Wins Ratio")
plt.show()

```

```
In [25]: data_w = data.Winner.value_counts()
sns.barplot(y = data_w.index, x = data_w, orient='h');
```



```
In [26]: data[data['Year']==2015]
```

Unnamed: 0	Scorecard	Team 1	Team 2	Margin	Ground	Match Date	Winner	H
934	3414	ODI # 3578	Australia	England	Winner2ndInning	Sydney	Jan 16, 2015	Australia
935	3423	ODI # 3589	Australia	England	Winner2ndInning	Hobart	Jan 23, 2015	Australia
936	3430	ODI # 3597	Australia	England	Winner1stInning	Perth	Feb 1, 2015	Australia
937	3433	ODI # 3600	Australia	England	Winner1stInning	Melbourne	Feb 14, 2015	Australia
938	3440	ODI # 3607	New Zealand	England	Winner2ndInning	Wellington	Feb 20, 2015	New Zealand

	Unnamed: 0	Scorecard	Team 1	Team 2	Margin	Ground	Match Date	Winner	H
939	3463	ODI # 3630	Bangladesh	England	Winner1stInning	Adelaide	Mar 9, 2015	Bangladesh	
940	3468	ODI # 3635	Afghanistan	England	Winner2ndInning	Sydney	Mar 13, 2015	England	
1289	3420	ODI # 3586	India	England	Winner2ndInning	Brisbane	Jan 20, 2015	England	
1290	3428	ODI # 3595	India	England	Winner2ndInning	Perth	Jan 30, 2015	England	
1291	3444	ODI # 3611	Scotland	England	Winner1stInning	Christchurch	Feb 23, 2015	England	
1292	3452	ODI # 3619	Sri Lanka	England	Winner2ndInning	Wellington	Mar 1, 2015	Sri Lanka	
1293	3485	ODI # 3654	New Zealand	England	Winner1stInning	Birmingham	Jun 9, 2015	England	
1294	3486	ODI # 3655	New Zealand	England	Winner1stInning	The Oval	Jun 12, 2015	New Zealand	
1295	3487	ODI # 3656	New Zealand	England	Winner2ndInning	Southampton	Jun 14, 2015	New Zealand	
1296	3488	ODI # 3657	New Zealand	England	Winner2ndInning	Nottingham	Jun 17, 2015	England	
1297	3490	ODI # 3659	New Zealand	England	Winner2ndInning	Chester	Jun 20, 2015	England	
1298	3511	ODI # 3680	Australia	England	Winner1stInning	Southampton	Sep 3, 2015	Australia	
1299	3512	ODI # 3681	Australia	England	Winner1stInning	Lord's	Sep 5, 2015	Australia	
1300	3513	ODI # 3682	Australia	England	Winner1stInning	Manchester	Sep 8, 2015	England	
1301	3514	ODI # 3683	Australia	England	Winner2ndInning	Leeds	Sep 11, 2015	England	
1302	3515	ODI # 3684	Australia	England	Winner2ndInning	Manchester	Sep 13, 2015	Australia	
1303	3538	ODI # 3707	Pakistan	England	Winner2ndInning	Abu Dhabi	Nov 11, 2015	Pakistan	

	Unnamed: 0	Scorecard	Team 1	Team 2	Margin	Ground	Match Date	Winner	H
1304	3539	ODI # 3708	Pakistan	England	Winner1stInning	Abu Dhabi	Nov 13, 2015	England	
1305	3541	ODI # 3710	Pakistan	England	Winner2ndInning	Sharjah	Nov 17, 2015	England	
1306	3543	ODI # 3712	Pakistan	England	Winner1stInning	Dubai	Nov 20, 2015	England	



In [27]:

data[data['Winner']=='England']

	Unnamed: 0	Scorecard	Team 1	Team 2	Margin	Ground	Match Date	Winner	Host_Cc
665	14	ODI # 16	Australia	England	Winner2ndInning	Melbourne	Jan 1, 1975	England	Au
666	41	ODI # 45	Pakistan	England	Winner2ndInning	Sahiwal	Dec 23, 1977	England	Pa
667	42	ODI # 46	Pakistan	England	Winner2ndInning	Sialkot	Dec 30, 1977	England	Pa
669	53	ODI # 58	Australia	England	Winner2ndInning	Melbourne	Jan 24, 1979	England	Au
672	72	ODI # 77	Australia	England	Winner2ndInning	Melbourne	Dec 8, 1979	England	Au
...
1322	3704	ODI # 3885	Australia	England	Winner1stInning	Birmingham	Jun 10, 2017	England	Er
1324	3728	ODI # 3911	West Indies	England	Winner2ndInning	Manchester	Sep 19, 2017	England	Er
1325	3731	ODI # 3915	West Indies	England	Winner1stInning	Bristol	Sep 24, 2017	England	Er
1326	3732	ODI # 3916	West Indies	England	Winner1stInning	The Oval	Sep 27, 2017	England	Er
1327	3734	ODI # 3918	West Indies	England	Winner2ndInning	Southampton	Sep 29, 2017	England	Er

343 rows × 14 columns

In [28]:

```
result = data.loc[(data['Year'] == 2015) & (data['Winner'] == 'England')]
```

In [29]:

```
result
```

Out[29]:

	Unnamed: 0	Scorecard	Team 1	Team 2	Margin	Ground	Match Date	Winner	Host_
940	3468	ODI # 3635	Afghanistan	England	Winner2ndInning	Sydney	Mar 13, 2015	England	
1289	3420	ODI # 3586	India	England	Winner2ndInning	Brisbane	Jan 20, 2015	England	
1290	3428	ODI # 3595	India	England	Winner2ndInning	Perth	Jan 30, 2015	England	
1291	3444	ODI # 3611	Scotland	England	Winner1stInning	Christchurch	Feb 23, 2015	England	New
1293	3485	ODI # 3654	New Zealand	England	Winner1stInning	Birmingham	Jun 9, 2015	England	
1296	3488	ODI # 3657	New Zealand	England	Winner2ndInning	Nottingham	Jun 17, 2015	England	
1297	3490	ODI # 3659	New Zealand	England	Winner2ndInning	Chester	Jun 20, 2015	England	
1300	3513	ODI # 3682	Australia	England	Winner1stInning	Manchester	Sep 8, 2015	England	
1301	3514	ODI # 3683	Australia	England	Winner2ndInning	Leeds	Sep 11, 2015	England	
1304	3539	ODI # 3708	Pakistan	England	Winner1stInning	Abu Dhabi	Nov 13, 2015	England	
1305	3541	ODI # 3710	Pakistan	England	Winner2ndInning	Sharjah	Nov 17, 2015	England	
1306	3543	ODI # 3712	Pakistan	England	Winner1stInning	Dubai	Nov 20, 2015	England	

In [30]:

```
#Filter dataset for england and australia
dataset_filtered = data[ (data['Team 1']=='Australia') & (data['Team 2']=='England')]
dataset_filtered
```

Out[30]:

	Unnamed: 0	Scorecard	Team 1	Team 2	Margin	Ground	Match Date	Winner	Host_Coi
664	0	ODI # 1	Australia	England	Winner2ndInning	Melbourne	Jan 5, 1971	Australia	Aus
665	14	ODI # 16	Australia	England	Winner2ndInning	Melbourne	Jan 1, 1975	England	Aus
669	53	ODI # 58	Australia	England	Winner2ndInning	Melbourne	Jan 24, 1979	England	Aus
670	54	ODI # 59	Australia	England	Winner2ndInning	Melbourne	Feb 4, 1979	Australia	Aus
671	55	ODI # 60	Australia	England	Winner2ndInning	Melbourne	Feb 7, 1979	Australia	Aus
...
1299	3512	ODI # 3681	Australia	England	Winner1stInning	Lord's	Sep 5, 2015	Australia	Eng
1300	3513	ODI # 3682	Australia	England	Winner1stInning	Manchester	Sep 8, 2015	England	Eng
1301	3514	ODI # 3683	Australia	England	Winner2ndInning	Leeds	Sep 11, 2015	England	Eng
1302	3515	ODI # 3684	Australia	England	Winner2ndInning	Manchester	Sep 13, 2015	Australia	Eng
1322	3704	ODI # 3885	Australia	England	Winner1stInning	Birmingham	Jun 10, 2017	England	Eng

132 rows × 14 columns



In [31]: `dataset_filtered.rename(columns = {'Team 1': 'Team_1', 'Team 2' : 'Team_2'}, inplace=True)`

C:\Users\hassa\miniconda3\envs\general\lib\site-packages\pandas\core\frame.py:5039: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
return super().rename(

In [32]: `# Converting categorical values to encoded values
x = dataset_filtered.drop(['Winner', 'Unnamed: 0', 'Scorecard', 'Match Date', 'Host_Country', 'Innings'])
x`

Out[32]:

	Team_1	Team_2	Margin	Ground	Venue_Team1	Venue_Team2	Innings_Team1	Inn
664	Australia	England	Winner2ndInning	Melbourne	Home	Away	Second	
665	Australia	England	Winner2ndInning	Melbourne	Home	Away	First	

	Team_1	Team_2	Margin	Ground	Venue_Team1	Venue_Team2	Innings_Team1	Inn
669	Australia	England	Winner2ndInning	Melbourne	Home	Away	First	
670	Australia	England	Winner2ndInning	Melbourne	Home	Away	Second	
671	Australia	England	Winner2ndInning	Melbourne	Home	Away	Second	
...
1299	Australia	England	Winner1stInning	Lord's	Away	Home	First	
1300	Australia	England	Winner1stInning	Manchester	Away	Home	Second	
1301	Australia	England	Winner2ndInning	Leeds	Away	Home	First	
1302	Australia	England	Winner2ndInning	Manchester	Away	Home	Second	
1322	Australia	England	Winner1stInning	Birmingham	Away	Home	Second	

132 rows × 9 columns



```
In [33]: y = dataset_filtered['Winner']
y
```

Out[33]: 664 Australia
665 England
669 England
670 Australia
671 Australia
...
1299 Australia
1300 England
1301 England
1302 Australia
1322 England
Name: Winner, Length: 132, dtype: object

```
In [34]: x['Team_1']= x['Team_1'].replace({'Australia':0})
x['Team_2']= x['Team_2'].replace({'England':1})
x_filtered = pd.get_dummies(x)
y_filtered = pd.get_dummies(y,drop_first = True)
```

```
In [35]: print(x_filtered)
```

	Team_1	Team_2	Year	Margin_Winner1stInning	Margin_Winner2ndInning	\
664	0	1	1971	0	1	
665	0	1	1975	0	1	
669	0	1	1979	0	1	
670	0	1	1979	0	1	
671	0	1	1979	0	1	
...
1299	0	1	2015	1	0	
1300	0	1	2015	1	0	
1301	0	1	2015	0	1	
1302	0	1	2015	0	1	
1322	0	1	2017	1	0	

	Ground_Adelaide	Ground_Birmingham	Ground_Brisbane	Ground_Bristol	\
664	0	0	0	0	
665	0	0	0	0	
669	0	0	0	0	
670	0	0	0	0	
671	0	0	0	0	
...	
1299	0	0	0	0	
1300	0	0	0	0	
1301	0	0	0	0	
1302	0	0	0	0	
1322	0	1	0	0	

	Ground_Cardiff	...	Venue_Team1_Away	Venue_Team1_Home	\
664	0	...	0	1	
665	0	...	0	1	
669	0	...	0	1	
670	0	...	0	1	
671	0	...	0	1	
...	
1299	0	...	1	0	
1300	0	...	1	0	
1301	0	...	1	0	
1302	0	...	1	0	
1322	0	...	1	0	

	Venue_Team1_Neutral	Venue_Team2_Away	Venue_Team2_Home	\
664	0	1	0	
665	0	1	0	
669	0	1	0	
670	0	1	0	
671	0	1	0	
...	
1299	0	0	1	
1300	0	0	1	
1301	0	0	1	
1302	0	0	1	
1322	0	0	1	

	Venue_Team2_Neutral	Innings_Team1_First	Innings_Team1_Second	\
664	0	0	1	
665	0	1	0	
669	0	1	0	
670	0	0	1	
671	0	0	1	
...	
1299	0	1	0	
1300	0	0	1	
1301	0	1	0	
1302	0	0	1	
1322	0	0	1	

	Innings_Team2_First	Innings_Team2_Second
664	1	0
665	0	1
669	0	1
670	1	0
671	1	0
...
1299	0	1

1300	1	0
1301	0	1
1302	1	0
1322	1	0

[132 rows x 38 columns]

In [36]:

```
# Splitting data into training and test set
from sklearn.model_selection import train_test_split

x_train,x_test,y_train,y_test = train_test_split(x_filtered,y_filtered,test_size = 0.3,
```

In [37]:

```
display(x_train.head())
display(y_train.head())
```

	Team_1	Team_2	Year	Margin_Winner1stInning	Margin_Winner2ndInning	Ground_Adelaide	Gro
702	0	1	1985	0	1	0	
860	0	1	2007	1	0	0	
786	0	1	1999	1	0	0	
1225	0	1	2009	0	1	0	
858	0	1	2007	1	0	0	

5 rows x 38 columns

England	
702	0
860	1
786	0
1225	0
858	1

In [38]:

```
# Model building
from xgboost import XGBClassifier

classifier = XGBClassifier(learning_rate=0.25,n_estimators=500,objective='binary:logist

classifier.fit(x_train,y_train)
y_pred = classifier.predict(x_test)
```

C:\Users\hassa\miniconda3\envs\general\lib\site-packages\xgboost\sklearn.py:888: UserWarning: The use of label encoder in XGBClassifier is deprecated and will be removed in a future release. To remove this warning, do the following: 1) Pass option use_label_encoder=False when constructing XGBClassifier object; and 2) Encode your labels (y) as integers starting with 0, i.e. 0, 1, 2, ..., [num_class - 1].

warnings.warn(label_encoder_deprecation_msg, UserWarning)

C:\Users\hassa\miniconda3\envs\general\lib\site-packages\sklearn\preprocessing_label.py:98: DataConversionWarning: A column-vector y was passed when a 1d array was expected.

Please change the shape of y to (n_samples,), for example using ravel().

```
y = column_or_1d(y, warn=True)
```

C:\Users\hassa\miniconda3\envs\general\lib\site-packages\sklearn\preprocessing_label.p
y:133: DataConversionWarning: A column-vector y was passed when a 1d array was expected.
Please change the shape of y to (n_samples,), for example using ravel().

```
y = column_or_1d(y, warn=True)
```

[13:23:59] WARNING: ..\src\learner.cc:1061: Starting in XGBoost 1.3.0, the default evalu
ation metric used with the objective 'binary:logistic' was changed from 'error' to 'logl
oss'. Explicitly set eval_metric if you'd like to restore the old behavior.

In [39]:

```
# Model accuracy
from sklearn.metrics import roc_auc_score
try:
    display(roc_auc_score(y_test,y_pred))
except ValueError:
    pass
display(x_filtered.head())
```

1.0

	Team_1	Team_2	Year	Margin_Winner1stInning	Margin_Winner2ndInning	Ground_Adelaide	Grou
664	0	1	1971	0	1	0	
665	0	1	1975	0	1	0	
669	0	1	1979	0	1	0	
670	0	1	1979	0	1	0	
671	0	1	1979	0	1	0	

5 rows × 38 columns

In [40]:

```
# Filtering data for current venue

x_current = x_filtered[(x_filtered['Team_2']==1) & (x_filtered['Venue_Team1_Home']==1)
                      (x_filtered['Ground_Melbourne']==1)]

display('Melbourne',classifier.predict(x_current).mean())

x_current = x_filtered[(x_filtered['Team_2']==1) & (x_filtered['Venue_Team1_Home']==1)
                      (x_filtered["Ground_Sydney"]==1)]

display('Sydney',classifier.predict(x_current).mean())

x_current = x_filtered[(x_filtered['Team_2']==1) & (x_filtered['Venue_Team1_Home']==1)
                      (x_filtered['Ground_Perth']==1)]

display('Perth',classifier.predict(x_current).mean())
```

'Melbourne'

0.2727272727272727

'Sydney'

0.47619047619047616

'Perth'

0.4

In [41]:

```
# ! jupyter nbconvert --to webpdf --allow-chromium-download England_ODI_Prediction.ipyn
```

In []: