In [1]:

```
#import libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [8]:

```
dataset = pd.read_csv("C:\\Users\\admin\\Downloads\\WorldCupMatches.csv")
```

In [9]:

```
dataset.shape
# rows , columns
```

Out[9]:

```
(4572, 20)
```

In [10]:

```
dataset.head()
#for first five records
```

Out[10]:

	Year	Datetime	Stage	Stadium	City	Home Team Name	Home Team Goals	Away Team Goals	Away Team Name	V conditic
0	1930.0	13 Jul 1930 - 15:00	Group 1	Pocitos	Montevideo	France	4.0	1.0	Mexico	
1	1930.0	13 Jul 1930 - 15:00	Group 4	Parque Central	Montevideo	USA	3.0	0.0	Belgium	
2	1930.0	14 Jul 1930 - 12:45	Group 2	Parque Central	Montevideo	Yugoslavia	2.0	1.0	Brazil	
3	1930.0	14 Jul 1930 - 14:50	Group 3	Pocitos	Montevideo	Romania	3.0	1.0	Peru	
4	1930.0	15 Jul 1930 - 16:00	Group 1	Parque Central	Montevideo	Argentina	1.0	0.0	France	

In [11]:

```
dataset.tail()
#last 5 items
```

Out[11]:

	Year	Datetime	Stage	Stadium	City	Home Team Name	Home Team Goals	Away Team Goals	Away Team Name	Win conditions	Attend
4567	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
4568	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
4569	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
4570	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
4571	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	

In [14]:

```
dataset.isnull()
```

Out[14]:

	Year	Datetime	Stage	Stadium	City	Home Team Name	Home Team Goals	Away Team Goals	Away Team Name	Win conditions	Atter
0	False	False	False	False	False	False	False	False	False	False	
1	False	False	False	False	False	False	False	False	False	False	
2	False	False	False	False	False	False	False	False	False	False	
3	False	False	False	False	False	False	False	False	False	False	
4	False	False	False	False	False	False	False	False	False	False	
...	...	...	...	...	...	...	...	...	...	...	
4567	True	True	True	True	True	True	True	True	True	True	
4568	True	True	True	True	True	True	True	True	True	True	
4569	True	True	True	True	True	True	True	True	True	True	
4570	True	True	True	True	True	True	True	True	True	True	
4571	True	True	True	True	True	True	True	True	True	True	

4572 rows × 20 columns

In [15]:

```
dataset.isnull().sum()
```

Out[15]:

```
Year                   3720
Datetime               3720
Stage                  3720
Stadium                3720
City                   3720
Home Team Name         3720
Home Team Goals        3720
Away Team Goals        3720
Away Team Name         3720
Win conditions         3720
Attendance             3722
Half-time Home Goals   3720
Half-time Away Goals   3720
Referee                3720
Assistant 1            3720
Assistant 2            3720
RoundID                3720
MatchID                3720
Home Team Initials     3720
Away Team Initials     3720
dtype: int64
```

In [16]:

```
# dropping all tupples containg NaN
dataset.dropna(inplace=True)
```

In [18]:

```
dataset.isnull().sum()
```

Out[18]:

```
Year                   0
Datetime               0
Stage                  0
Stadium                0
City                   0
Home Team Name         0
Home Team Goals        0
Away Team Goals        0
Away Team Name         0
Win conditions         0
Attendance             0
Half-time Home Goals   0
Half-time Away Goals   0
Referee                0
Assistant 1            0
Assistant 2            0
RoundID                0
MatchID                0
Home Team Initials     0
Away Team Initials     0
dtype: int64
```

In [19]:

```
dataset.shape
```

Out[19]:

```
(850, 20)
```

In [22]:

```
dataset2 = pd.read_csv("C:\\Users\\admin\\Desktop\\WorldCup.csv")
```

In [23]:

```
dataset2.shape
```

Out[23]:

```
(4572, 20)
```

In [24]:

```
dataset2['Year'].mean()
```

Out[24]:

```
1985.0892018779343
```

In [25]:

```
dataset2['Year'].tail()
```

Out[25]:

```
4567   NaN
4568   NaN
4569   NaN
4570   NaN
4571   NaN
Name: Year, dtype: float64
```

In [27]:

```
dataset2['Year'].replace(np.NaN, dataset2['Year'].mean()).tail()
```

Out[27]:

```
4567    1985.089202
4568    1985.089202
4569    1985.089202
4570    1985.089202
4571    1985.089202
Name: Year, dtype: float64
```

In [ ]:

```

```