

EDS Assignment 1

Name – Mayur Kapgate

Batch - D2

Roll No. – 428

PRN No. – 202201040065

Question - Find the statistical Analysis of World Cup Winner dataset (**Refer Data Set 3**)

- Find the country who has won the least world Cup
- Find the venue played most world cup matches
- Who is the two finalist played in 1983
- Find the total number of world cup matched played
- Clean the dataset if record is empty with 0 values or delete incomplete data row
- Find the final match hosted by “England”
- Find the all the ‘Years’ Hosted by England in world cup

csv File –

Year	Host	Venue for Final	Team-1	Team-2	Winner	Margin
1975	England	Lord's	WI	Aus	WI	17 runs
1979	England	Lord's	WI	Eng	WI	92 runs
1983	England	Lord's	Ind	WI	Ind	43 runs
1987	India	Kolkata	Aus	Eng	Aus	7 runs
1992	Australia, New Zealand	Melbourne	Pak	Eng	Pak	22 runs
1996	India, Pakistan, Sri Lanka	Lahore (Gdffi)	Aus	SL	SL	7 wickets
1999	England	Lord's	Pak	Aus	Aus	8 wickets
2003	South Africa	Wanderers	Aus	Ind	Aus	125 runs
2007	West Indies	Bridgetown	Aus	SL	Aus	53 runs
2011	India, Pakistan, Sri Lanka, Bangladesh	Wankhede	SL	Ind	Ind	6 wickets
2015	Australia, New Zealand	Melbourne	NZ	Aus	Aus	7 wickets
2019	India, Pakistan, Sri Lanka	Lahore (Gdffi)	Aus	SL	SL	7 wickets

Code –

```

import pandas as pd

# Read the dataset into a DataFrame
df = pd.read_csv('/content/dataset3 (1).csv')

# a. Find the country that has won the least World Cups
least_wins_country = df['Winner'].value_counts().idxmin()
print("Country with the least World Cup wins:", least_wins_country)

# b. Find the venue that has hosted the most World Cup matches
most_matches_venue = df['Venue for Final'].value_counts().idxmax()
print("Venue with the most World Cup matches:", most_matches_venue)

# c. Find the two finalists that played in 1983
finalists_1983 = df.loc[df['Year'] == 1983, ['Team-1', 'Team-2']]
print("Finalists in 1983:", finalists_1983.values.flatten())

# d. Find the total number of World Cup matches played
total_matches = len(df)
print("Total number of World Cup matches played:", total_matches)

# e. Clean the dataset by replacing empty records with 0 values or
deleting incomplete rows
df.fillna(0, inplace=True) # Replace empty records with 0
df.dropna(inplace=True)    # Delete incomplete rows

# f. Find the final match hosted by 'England'
final_match_england = df.loc[df['Host'] == 'England', 'Year'].max()
print("Final match hosted by England:", final_match_england)

# g. Find all the years hosted by England in the World Cup
years_hosted_england = df.loc[df['Host'] == 'England', 'Year'].unique()
print("Years hosted by England in the World Cup:",
years_hosted_england)

```

Output –

```

Country with the least World Cup wins: Pak
Venue with the most World Cup matches: Lord's
Finalists in 1983: ['Ind' 'WI ']
Total number of World Cup matches played: 12
Final match hosted by England: 1999
Years hosted by England in the World Cup: [1975 1979 1983 1999]

```