Assignment No.2

Name:-Aglave Suraj Rajendra

PRN:-202401040060

Div:-CS2

Roll N0.:-08

Sub:-EDS

Dataset: ICC t20 world cup

dataset

Match No.	Date	Venue	1st Team	2nd Team	Stage	Toss Winni	Toss Decisic	First Innings	Fall of wick	Second Inni	Fall of wick	Winners	Method	Won by	Winning Ma	Top Scorer	Highest Sco	Best Bowler	Best Bowler	Best Bo
1st	Saturday,Jur	Grand Prairi	Canada	United State	Group A	United State	Fielding	194	5	197	3	United State	Normal Mai	Wickets	7	Aaron Jones	94	Dilon Heylig	1	
2nd	Sunday,June	Providence	Papua New	Afghanistan	Group C	Afghanistan	Fielding	95	10	101	3	Afghanistan	Normal Mai	Wickets	7	Gulbadin Na	49	Fazalhaq Fa	3	
3rd	Sunday,June	Kensington	Oman	Namibia	Group B	Namibia	Fielding	109	10	109	6	Namibia	Normal Mai	Wickets	4	Jan Frylinck	45	Mehran Kha	3	
4th	Monday,Jur	Nassau Cou	Sri Lanka	South Africa	Group D	Sri Lanka	Batting	77	10	80	4	South Africa	Normal Mai	Wickets	6	Quinton de	20	Anrich Nort	4	
5th	Monday,Jur	Providence	Afghanistan	Uganda	Group C	Uganda	Fielding	183	5	58	10	Afghanistan	Normal Mai	Runs	125	Rahmanulla	76	Fazalhaq Fa	5	
6th	Tuesday,Jur	Kensington	Scotland	England	Group B	Scotland	Batting	90	0	Rain	Rain	Rain	Rain	Rain	Rain	Rain	Rain	Rain	Rain	Rain
7th	Tuesday,Jur	Grand Prairi	Nepal	Netherlands	Group D	Netherlands	Fielding	106	10	109	4	Netherlands	Normal Mai	Wickets	6	Max O'Dow	54	Tim Pringle	3	
8th	Wednesday	Nassau Cou	India	Ireland	Group A	India	Fielding	96	10	97	2	India	Normal Mai	Wickets	8	Rohit Sharm	52	Jasprit Bum	2	
9th	Wednesday	Providence	Papua New	Uganda	Group C	Uganda	Fielding	77	10	78	8	Uganda	Normal Mai	Wickets	3	Riazat Ali Sh	33	Alpesh Ram	17	
10th	Wednesday	Kensington	Australia	Oman	Group B	Oman	Fielding	164	5	125	9	Australia	Normal Mai	Runs	39	Marcus Stoi	67	Marcus Stoi	3	
11th	Thursday,Ju	Grand Prairi	Pakistan	United State	Group A	United State	Fielding	159	7	160	3	United State	Normal Mai	Wickets	1	Monank Pat	50	Nosthush Ke	30	
12th	Thursday,Ju	Kensington	Namibia	Scotland	Group B	Namibia	Batting	155	9	157	5	Scotland	Normal Mai	Wickets	5	Michael Lea	35	Brad Currie	2	
13th	Friday,June	Nassau Cou	Canada	Ireland	Group A	Ireland	Fielding	137	7	125	7	Canada	Normal Mai	Runs	12	Nicholas Kir	49	Jeremy Gore	2	
14th	Friday,June	Providence	Afghanistan	New Zealan	Group C	New Zealan	Fielding	159	6	75	10	Afghanistan	Normal Mai	Runs	84	Rahmanulla	80	Rashid Khar	4	
15th	Friday,June	Grand Prairi	Sri Lanka	Bangladesh	Group D	Bangladesh	Fielding	124	9	125	8	Bangladesh	Normal Mai	Wickets	2	Pathum Niss	47	Rishad Hoss	3	
16th	Saturday,Ju	Nassau Cou	Netherlands	South Africa	Group D	South Africa	Fielding	103	9	106	6	South Africa	Normal Mai	Wickets	4	David Miller	59	Ottneil Baar	4	
17th	Saturday,Ju	Kensington	Australia	England	Group B	England	Fielding	201	7	165	6	Australia	Normal Mai	Runs	36	Jos Buttler	42	Adam Zamp	2	
18th	Saturday,Ju	Providence	West Indies	Uganda	Group C	West Indies	Batting	173	5	39	10	West Indies	Normal Mai	Runs	134	Johnson Cha	44	Akeal Hosei	5	
19th	Sunday,June	Nassau Cou	India	Pakistan	Group A	Pakistan	Fielding	119	10	113	7	India	Normal Mai	Runs	6	Rishabh Par	42	Jasprit Bum	3	
20th	Sunday,June	Sir Vivian Ri	Scotland	Oman	Group B	Oman	Batting	150	7	153	2	Scotland	Normal Mai	Wickets	7	Brandon Mc	31	Chris Greave	1	
21th	Monday,Jur	Nassau Cou	South Africa	Bangladesh	Group D	South Africa	Batting	113	6	109	7	South Africa	Normal Mai	Runs	4	Heinrich Kla	46	Tanzim Has	3	
22th	Tuesday,Jur	Nassau Cou	Canada	Pakistan	Group A	Pakistan	Fielding	106	7	107	3	Pakistan	Normal Mai	Wickets	7	Aaron Johns	52	Mohammac	2	
23th	Tuesday,Jur	Central Brov	Nepal	Sri Lanka	Group D	Rain	Rain	Rain	Rain	Rain	Rain	Rain	Rain	Rain	Rain	Rain	Rain	Rain	Rain
24th	Tuesday,Jur	Sir Vivian Ri	Namibia	Australia	Group B	Australia	Fielding	72	10	74	1	Australia	Normal Mai	Wickets	9	Gerhard Era	36	Adam Zamp	4	
25th	Wednesday	Nassau Cou	United State	India	Group A	India	Fielding	110	8	111	3	India	Normal Mai	Wickets	7	Suryakumar	50	Arshdeep Si	4	
26th	Wednesday	Brian Lara S	West Indies	New Zealan	Group C	New Zealan	Fielding	149	9	136	9	West Indies	Normal Mai	Runs	13	Sherfane Ru	68	Trent Boult	3	
27th	Thursday,Ju	Arnos Vale (	Bangladesh	Netherlands	Group D	Netherlands	Fielding	159	5	134	8	Bangladesh	Normal Mai	Runs	25	Shakib Al Ha	64	Paul van Me	2	
28th	Thursday,Ju	Sir Vivian Ri	Oman	England	Group B	England	Fielding	47	10	50	2	England	Normal Mai	Wickets	8	Jos Buttler	24	Adil Rashid	4	
29th	Thursday,Ju	Brian Lara S	Papua New	Afghanistan	Group C	Afghanistan	Fielding	95	10	101	3	Afghanistan	Normal Mai	Wickets	7	Gulbadin Na	49	Fazalhaq Fa	3	
30th	Friday,June	Central Brov	United State	Ireland	Group A	Rain	Rain	Rain	Rain	Rain	Rain	Rain	Rain	Rain	Rain	Rain	Rain	Rain	Rain
31th	Friday,June	Arnos Vale (	South Africa	Nepal	Group D	Nepal	Fielding	115	7	114	7	South Africa	Normal Mai	Runs	1	Reeza Hend	43	Tabraiz Shar	4	
32th	Friday,June	Brian Lara S	Uganda	New Zealan	Group C	New Zealan	Fielding	40	10	41	1	New Zealan	Normal Mai	Wickets	9	Devon Conv	22	Tim Souther	3	

| ICC_Mens_T20_Worldcup_1_ | + |

```
import pandas as pd
import numpy as np
df = pd.read_csv('/icc t20 world cup.csv')
```

```
# Problem 1: List all unique venues
def unique_venues(df):
    return df['Venue'].unique()

print("Unique Venues:", unique_venues(df))
```

```
Unique Venues: ['Grand Prairie Stadium, Dallas' 'Providence Stadium, Guyana'
 'Kensington Oval, Bridgetown, Barbados'
 'Nassau County International Cricket Stadium, New York'
 'Sir Vivian Richards Stadium, North Sound, Antigua'
 'Central Broward Regional Park Stadium Turf Ground, Lauderhill'
 'Brian Lara Stadium, Tarouba, Trinidad'
 'Arnos Vale Ground, Kingstown, St Vincent'
 'Daren Sammy National Cricket Stadium, Gros Islet, St Lucia']
```

```
# Problem 2: Count of matches played in each stage
def stage_match_count(df):
    return df['Stage'].value_counts()

print("Match Count by Stage:\n", stage_match_count(df))
```

```
# Problem 2: Count of matches played in each stage
def stage_match_count(df):
    return df['Stage'].value_counts()

print("Match Count by Stage:\n", stage_match_count(df))
```

```
Match Count by Stage:
 Stage
Group A          11
Group C          10
Group B          10
Group D          10
Group 1           6
Group 2           5
1st Semi-Final    1
2nd Semi-Final    1
Final             1
Name: count, dtype: int64
```

```
# Problem 3: Team with most match wins
def most_match_wins(df):
    return df['Winners'].value_counts().idxmax()

print("Team with Most Match Wins:", most_match_wins(df))
```

```
Team with Most Match Wins: South Africa
```

```
# Problem 4: Matches won by runs vs wickets
def win_type_count(df):
    return df['Won by'].value_counts()

print("Win Type Count:\n", win_type_count(df))
```

```
Win Type Count:
 Won by
Runs       26
Wickets    25
Rain        4
Name: count, dtype: int64
```

```
# Problem 5: Toss wins by each team
def toss_winners_count(df):
    return df['Toss Winning'].value_counts()

print("Toss Wins by Team:\n", toss_winners_count(df))
```

```
Toss Wins by Team:
 Toss Winning
England                  6
Australia                5
Afghanistan              5
New Zealand              4
India                    4
Namibia                  3
United State Of America  3
South Africa             3
Pakistan                 3
Netherlands              3
Rain                     3
Uganda                   2
Bangladesh               2
West Indies              2
Oman                     2
Nepal                    2
Ireland                  1
Sri Lanka                1
Scotland                 1
Name: count, dtype: int64
```

```python
# Problem 6: How many matches were played in total?
def total_matches(df):
    return len(df)

print("Total Matches Played:", total_matches(df))
```

Total Matches Played: 55

```python
# Problem 7: Count of normal matches
def normal_matches_count(df):
    return df[df['Method'] == 'Normal Match'].shape[0]

print("Number of Normal Matches:", normal_matches_count(df))
```

Number of Normal Matches: 48

```python
# Problem 8: Player of the match awards count
def player_of_match_counts(df):
    return df['Player Of The Match'].value_counts()

print("Player of the Match Counts:\n", player_of_match_counts(df))
```

```
Player of the Match Counts:
 Player Of The Match
Rain                4
Fazalhaq Farooqi    3
Jasprit Bumrah      2
Marcus Stoinis      2
Quinton de Kock     2
Tabraiz Shamsi      2
Adil Rashid         2
Adam Zampa          2
Anrich Nortje       1
David Wiese         1
```

```python
# Problem 9: Top 5 best bowlers
def top_best_bowlers(df):
    return df['Best Bowler'].value_counts().head(5)

print("Top 5 Best Bowlers:\n", top_best_bowlers(df))
```

```
Top 5 Best Bowlers:
 Best Bowler
Rain                4
Fazalhaq Farooqi    3
Rashid Khan         3
Adil Rashid         3
Jasprit Bumrah      2
Name: count, dtype: int64
```

```python
# Problem 10: Variance of Second Innings Scores
def var_second_innings(df):
    scores = pd.to_numeric(df['Second Innings Score'], errors='coerce')
    return np.var(scores)

print("Variance of Second Innings Scores:", var_second_innings(df))
```

Variance of Second Innings Scores: 1403.4886582083811

```python
# Problem 11: All unique toss decisions
def unique_toss_decisions(df):
    return df['Toss Decision'].unique()

print("Unique Toss Decisions:", unique_toss_decisions(df))
```

Unique Toss Decisions: ['Fielding' 'Batting' 'Rain']

```python
# Problem 12: Most frequent 1st team
def most_frequent_first_team(df):
    return df['1st Team'].value_counts().idxmax()

print("Most Frequent 1st Team:", most_frequent_first_team(df))
```

Total Matches Played: 55

```python
# Problem 13: Most frequent 2nd team
def most_frequent_second_team(df):
    return df['2nd Team'].value_counts().idxmax()

print("Most Frequent 2nd Team:", most_frequent_second_team(df))
```

Most Frequent 2nd Team: England

```python
# Problem 14: Number of matches played by India
def india_matches_count(df):
    return df['1st Team'].str.contains('India').sum() + df['2nd Team'].str.contains('India').sum()

print("Number of Matches Played by India:", india_matches_count(df))
```

Number of Matches Played by India: 9

```python
# Problem 15: Matches using D/L method
def dl_method_matches(df):
    return df[df['Method'] == 'D/L Method'].shape[0]

print("Number of Matches Using D/L Method:", dl_method_matches(df))
```

Number of Matches Using D/L Method: 0

```python
# Problem 16: Calculate the mean of First Innings Score
def find_mean_first_innings(df):
    scores = pd.to_numeric(df['First Innings Score'], errors='coerce')
    scores = scores.dropna().values  # Remove NaN
    return np.mean(scores)

print("Mean of First Innings Score:", find_mean_first_innings(df))
```

```
Mean of First Innings Score: 133.8653846153846
```

```python
# Problem 17: Find the maximum Winning Margin
def find_max_winning_margin(df):
    margins = pd.to_numeric(df['Winning Margin'], errors='coerce').dropna().values
    return np.max(margins)

print("Max Winning Margin:", find_max_winning_margin(df))
```

```
Max Winning Margin: 134.0
```

```python
# Problem 18: Count matches won by more than 50 runs or wickets
def count_high_margin_wins(df):
    margins = pd.to_numeric(df['Winning Margin'], errors='coerce').dropna().values
    return np.sum(margins > 50)

print("Matches won by margin > 50:", count_high_margin_wins(df))
```

```
Matches won by margin > 50: 6
```

```python
# Problem 19: Find the minimum First Innings Score
def find_min_first_innings(df):
    scores = pd.to_numeric(df['First Innings Score'], errors='coerce').dropna().values
    return np.min(scores)

print("Min First Innings Score:", find_min_first_innings(df))
```

```
Min First Innings Score: 40.0
```

```python
# Problem 20: Calculate the standard deviation of First Innings Scores
def std_first_innings(df):
    scores = pd.to_numeric(df['First Innings Score'], errors='coerce').dropna().values
    return np.std(scores)

print("Standard Deviation of First Innings Score:", std_first_innings(df))
```

```
Standard Deviation of First Innings Score: 43.586796186174865
```

```python
# Problem 21: Count the number of matches where First Innings Score > 150
def count_scores_above_150(df):
    scores = pd.to_numeric(df['First Innings Score'], errors='coerce').dropna().values
    return np.sum(scores > 150)

print("Scores above 150:", count_scores_above_150(df))
```

```
Scores above 150: 19
```

```python
# Problem 22: List all unique values of Winning Margin
def unique_winning_margins(df):
    margins = pd.to_numeric(df['Winning Margin'], errors='coerce').dropna().values
    return np.unique(margins)

print("Unique Winning Margins:", unique_winning_margins(df))
```

```
Unique Winning Margins: [  1.   2.   3.   4.   5.   6.   7.   8.   9.  10.  12.  13.  18.  21.
  24.  25.  28.  36.  39.  41.  47.  50.  68.  83.  84. 104. 125. 134.]
```

```python
# Problem 23: Calculate the median of First Innings Scores
def median_first_innings(df):
    scores = pd.to_numeric(df['First Innings Score'], errors='coerce').dropna().values
    return np.median(scores)

print("Median First Innings Score:", median_first_innings(df))
```

```
Median First Innings Score: 131.5
```

```python
# Problem 24: Check if there are any duplicate First Innings Scores
def has_duplicate_scores(df):
    scores = pd.to_numeric(df['First Innings Score'], errors='coerce').dropna().values
    return len(scores) != len(np.unique(scores))

print("Has Duplicate Scores:", has_duplicate_scores(df))
```

```
Has Duplicate Scores: True
```

```python
# Problem 25: Count how many First Innings Scores are between 100 and 160
def count_scores_in_range(df):
    scores = pd.to_numeric(df['First Innings Score'], errors='coerce').dropna().values
    return np.sum((scores >= 100) & (scores <= 160))

print("Scores between 100 and 160:", count_scores_in_range(df))
```

```
Scores between 100 and 160: 26
```

```python
# Problem 26: Find the index of the match with the highest Winning Margin
def index_max_margin(df):
    margins = pd.to_numeric(df['Winning Margin'], errors='coerce').dropna().values
    return np.argmax(margins)

print("Index of Highest Winning Margin:", index_max_margin(df))
```

```
Index of Highest Winning Margin: 16
```

```python
# Problem 27: Create a boolean array showing if each Winning Margin is greater than 5
def margin_above_5_boolean(df):
    margins = pd.to_numeric(df['Winning Margin'], errors='coerce').dropna().values
    return margins > 5

print("Margins > 5:", margin_above_5_boolean(df))
```

```
Margins > 5: [ True  True False  True  True  True  True False  True False False  True
  True False False  True  True  True  True False  True  True  True  True
  True  True  True False  True  True  True False False  True  True  True
  True  True  True False  True  True  True False False  True  True  True
  True  True  True  True  True  True  True  True  True False  True  True
  True  True  True]
```

```python
# Problem 28: Normalize the Winning Margins using Min-Max normalization
def normalize_margins(df):
    margins = pd.to_numeric(df['Winning Margin'], errors='coerce').dropna().values
    return (margins - np.min(margins)) / (np.max(margins) - np.min(margins))

print("Normalized Margins:", normalize_margins(df))
```

```
Normalized Margins: [0.04511278 0.04511278 0.02255639 0.03759398 0.93233083 0.03759398
 0.05263158 0.01503759 0.28571429 0.         0.03007519 0.08270677
 0.62406015 0.0075188  0.02255639 0.26315789 1.         0.03759398
 0.04511278 0.02255639 0.04511278 0.06015038 0.04511278 0.09022556
 0.18045113 0.05263158 0.04511278 0.         0.06015038 0.30075188
 0.03007519 0.01503759 0.15037594 0.61654135 0.04511278 0.77443609
 0.12781955 0.05263158 0.34586466 0.20300752 0.04511278 0.06015038
 0.36842105 0.15037594 0.06766917 0.01503759 0.17293233 0.05263158
 0.06015038 0.5037594  0.04511278]
```

```python
# Problem 29: Calculate average First Innings Score for matches with Winning Margin > 50
def avg_score_high_margin(df):
    scores = pd.to_numeric(df['First Innings Score'], errors='coerce').dropna().values
    margins = pd.to_numeric(df['Winning Margin'], errors='coerce').dropna().values
    min_len = min(len(scores), len(margins))
    return np.mean(scores[:min_len][margins[:min_len] > 50])

print("Average Score for High Margin Matches:", avg_score_high_margin(df))
```

```
Average Score for High Margin Matches: 126.83333333333333
```

```python
# Problem 30: Sort the First Innings Scores in ascending order
def sort_first_innings_scores(df):
    scores = pd.to_numeric(df['First Innings Score'], errors='coerce').dropna().values
    return np.sort(scores)

print("Sorted First Innings Scores:", sort_first_innings_scores(df))
```

```
Sorted First Innings Scores: [ 40.  47.  56.  72.  77.  77.  78.  90.  95.  95.  96. 103. 106. 106.
```