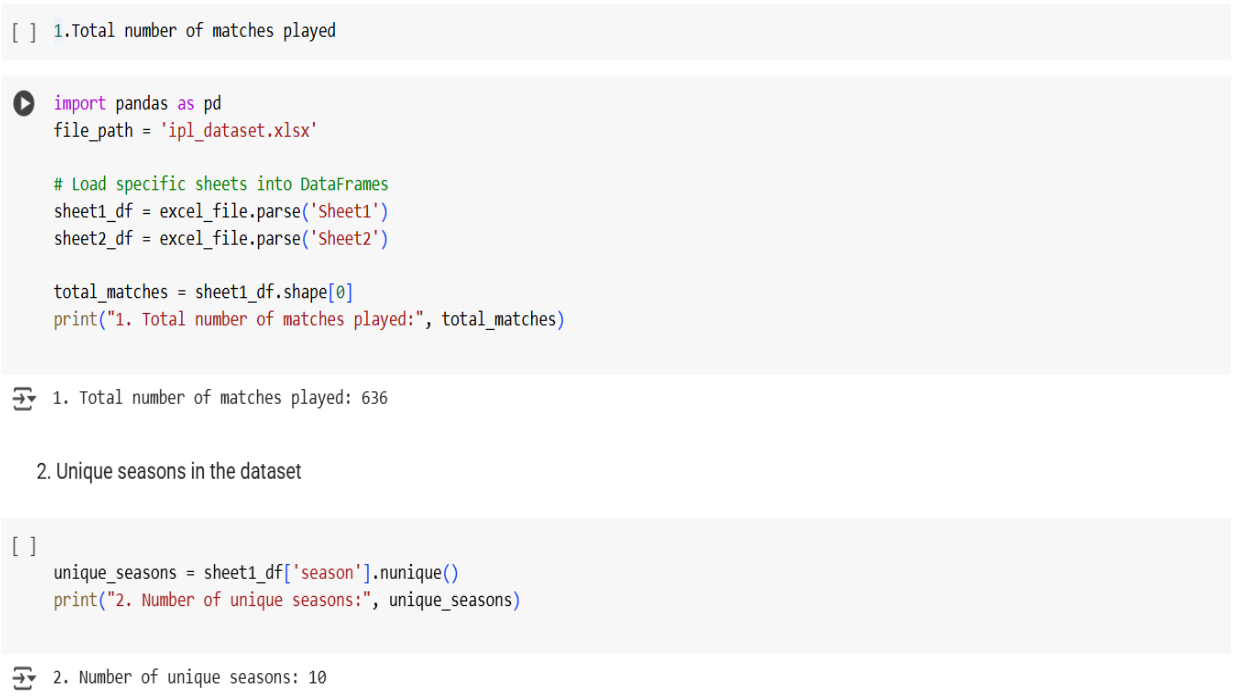
EDS THEORY ASSIGNMENT

NAME : Vikrant Prabhakar Haral

PRN: 202401040332

ROLL NO: CS2-10

Dataset : IPL



[ ] 1.Total number of matches played

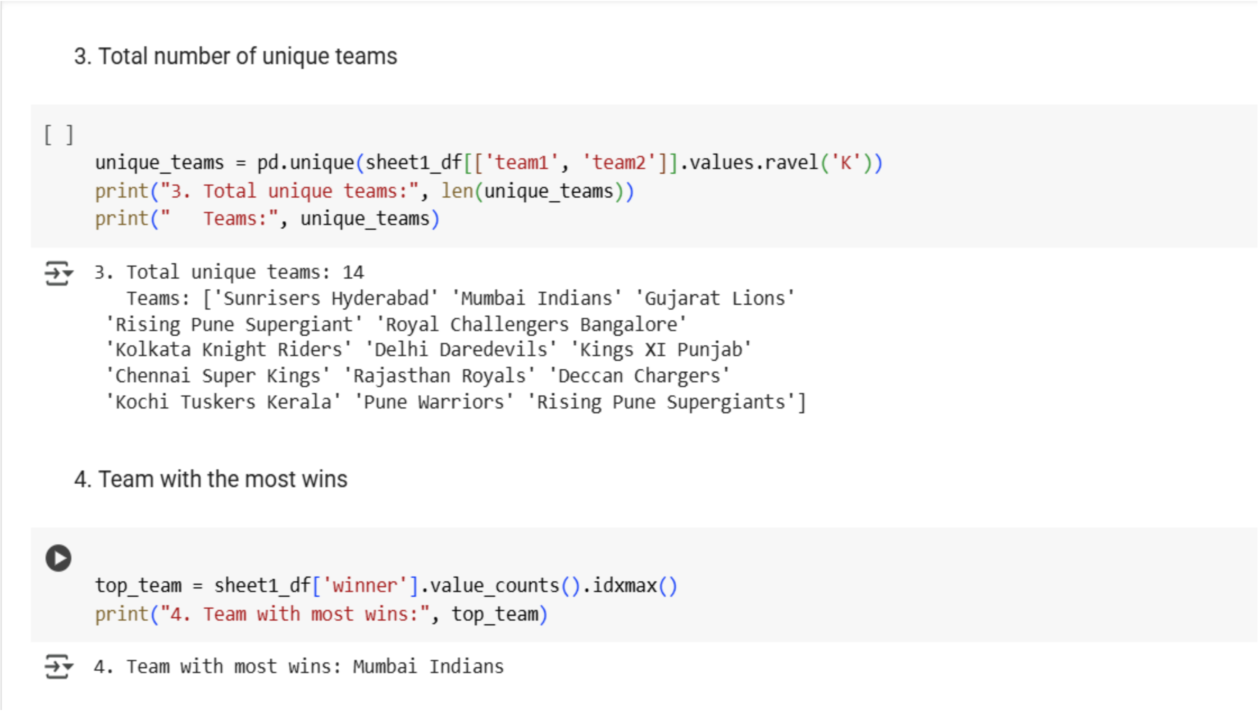
```
import pandas as pd
file_path = 'ipl_dataset.xlsx'

# Load specific sheets into DataFrames
sheet1_df = excel_file.parse('Sheet1')
sheet2_df = excel_file.parse('Sheet2')

total_matches = sheet1_df.shape[0]
print("1. Total number of matches played:", total_matches)
```

```
1. Total number of matches played: 636
```

2. Unique seasons in the dataset

```
unique_seasons = sheet1_df['season'].nunique()
print("2. Number of unique seasons:", unique_seasons)
```

```
2. Number of unique seasons: 10
```



3. Total number of unique teams

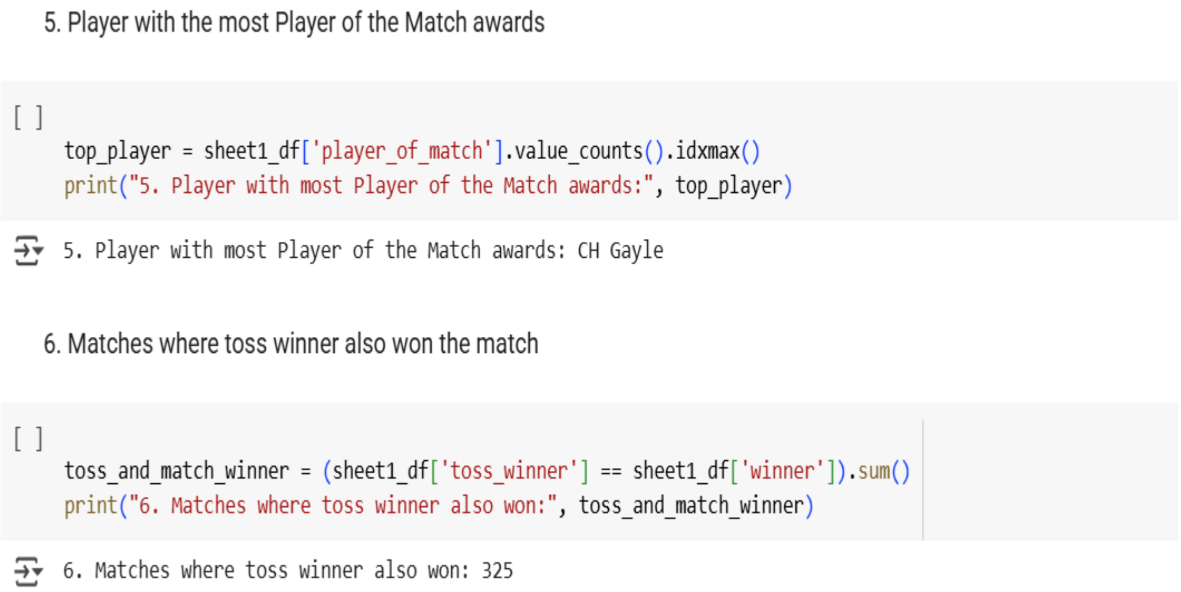
```
unique_teams = pd.unique(sheet1_df[['team1', 'team2']].values.ravel('K'))
print("3. Total unique teams:", len(unique_teams))
print("   Teams:", unique_teams)
```

```
3. Total unique teams: 14
   Teams: ['Sunrisers Hyderabad' 'Mumbai Indians' 'Gujarat Lions'
 'Rising Pune Supergiant' 'Royal Challengers Bangalore'
 'Kolkata Knight Riders' 'Delhi Daredevils' 'Kings XI Punjab'
 'Chennai Super Kings' 'Rajasthan Royals' 'Deccan Chargers'
 'Kochi Tuskers Kerala' 'Pune Warriors' 'Rising Pune Supergiants']
```

4. Team with the most wins

```
top_team = sheet1_df['winner'].value_counts().idxmax()
print("4. Team with most wins:", top_team)
```

```
4. Team with most wins: Mumbai Indians
```



5. Player with the most Player of the Match awards

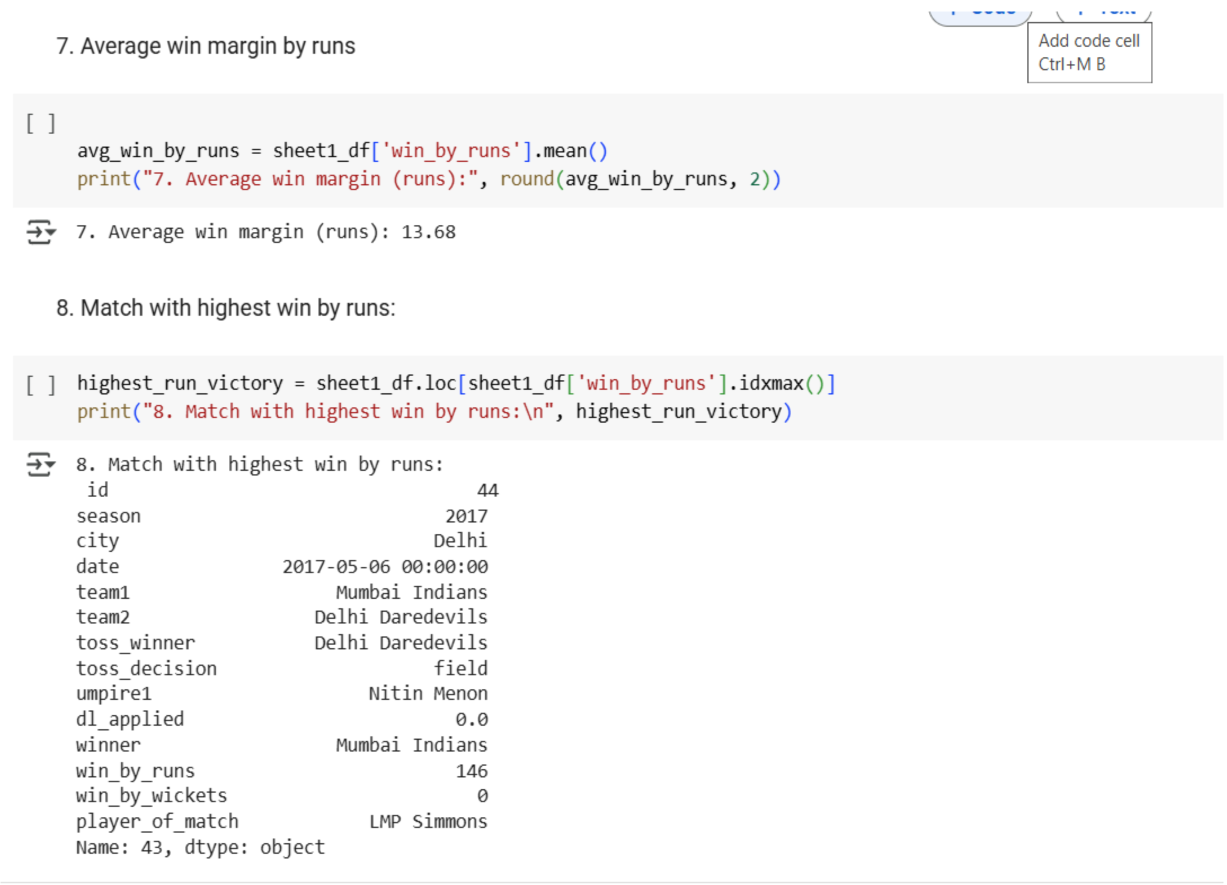
```
[ ]
top_player = sheet1_df['player_of_match'].value_counts().idxmax()
print("5. Player with most Player of the Match awards:", top_player)
```

```
5. Player with most Player of the Match awards: CH Gayle
```

6. Matches where toss winner also won the match

```
[ ]
toss_and_match_winner = (sheet1_df['toss_winner'] == sheet1_df['winner']).sum()
print("6. Matches where toss winner also won:", toss_and_match_winner)
```

```
6. Matches where toss winner also won: 325
```



7. Average win margin by runs

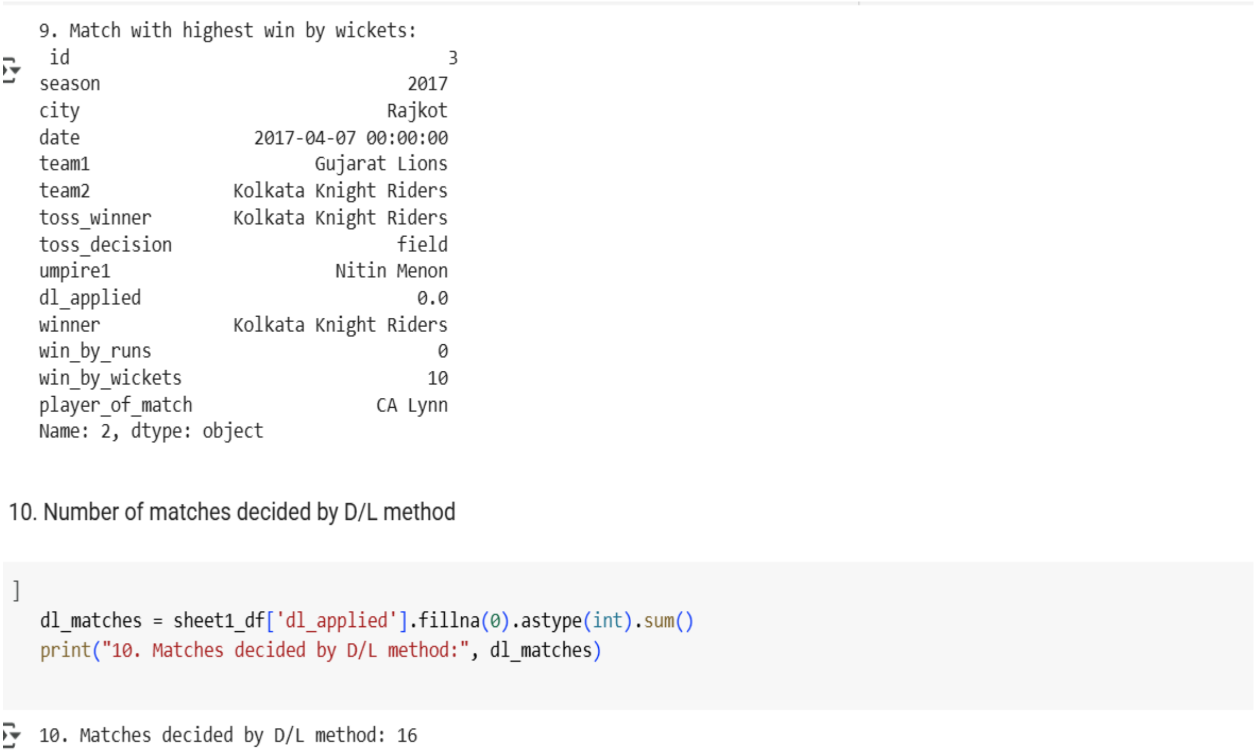
```
[ ]
avg_win_by_runs = sheet1_df['win_by_runs'].mean()
print("7. Average win margin (runs):", round(avg_win_by_runs, 2))
```

```
7. Average win margin (runs): 13.68
```

8. Match with highest win by runs:

```
[ ] highest_run_victory = sheet1_df.loc[sheet1_df['win_by_runs'].idxmax()]
    print("8. Match with highest win by runs:\n", highest_run_victory)
```

```
8. Match with highest win by runs:
 id                                  44
season                             2017
city                              Delhi
date                2017-05-06 00:00:00
team1                    Mumbai Indians
team2                  Delhi Daredevils
toss_winner            Delhi Daredevils
toss_decision                     field
umpire1                     Nitin Menon
dl_applied                          0.0
winner                   Mumbai Indians
win_by_runs                         146
win_by_wickets                        0
player_of_match             LMP Simmons
Name: 43, dtype: object
```

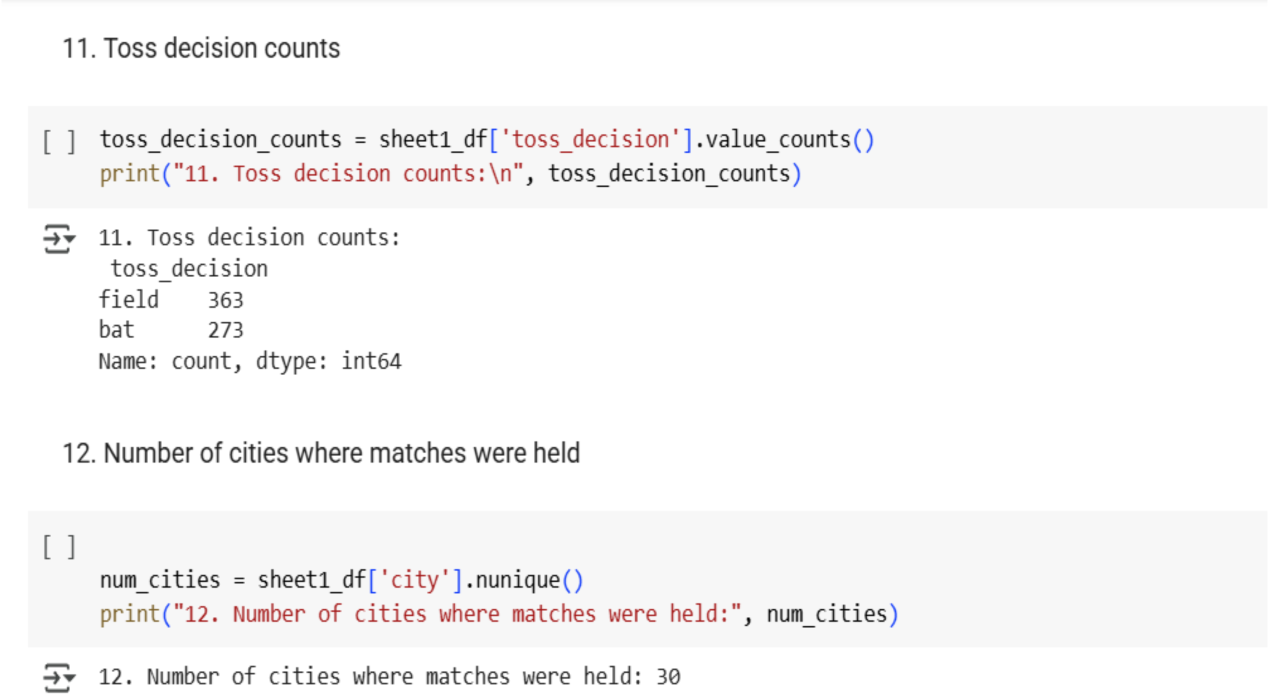


```
9. Match with highest win by wickets:
 id                                   3
season                             2017
city                             Rajkot
date                2017-04-07 00:00:00
team1                     Gujarat Lions
team2             Kolkata Knight Riders
toss_winner       Kolkata Knight Riders
toss_decision                     field
umpire1                     Nitin Menon
dl_applied                          0.0
winner            Kolkata Knight Riders
win_by_runs                           0
win_by_wickets                       10
player_of_match                 CA Lynn
Name: 2, dtype: object
```

10. Number of matches decided by D/L method

```python
dl_matches = sheet1_df['dl_applied'].fillna(0).astype(int).sum()
print("10. Matches decided by D/L method:", dl_matches)
```

```
10. Matches decided by D/L method: 16
```



11. Toss decision counts

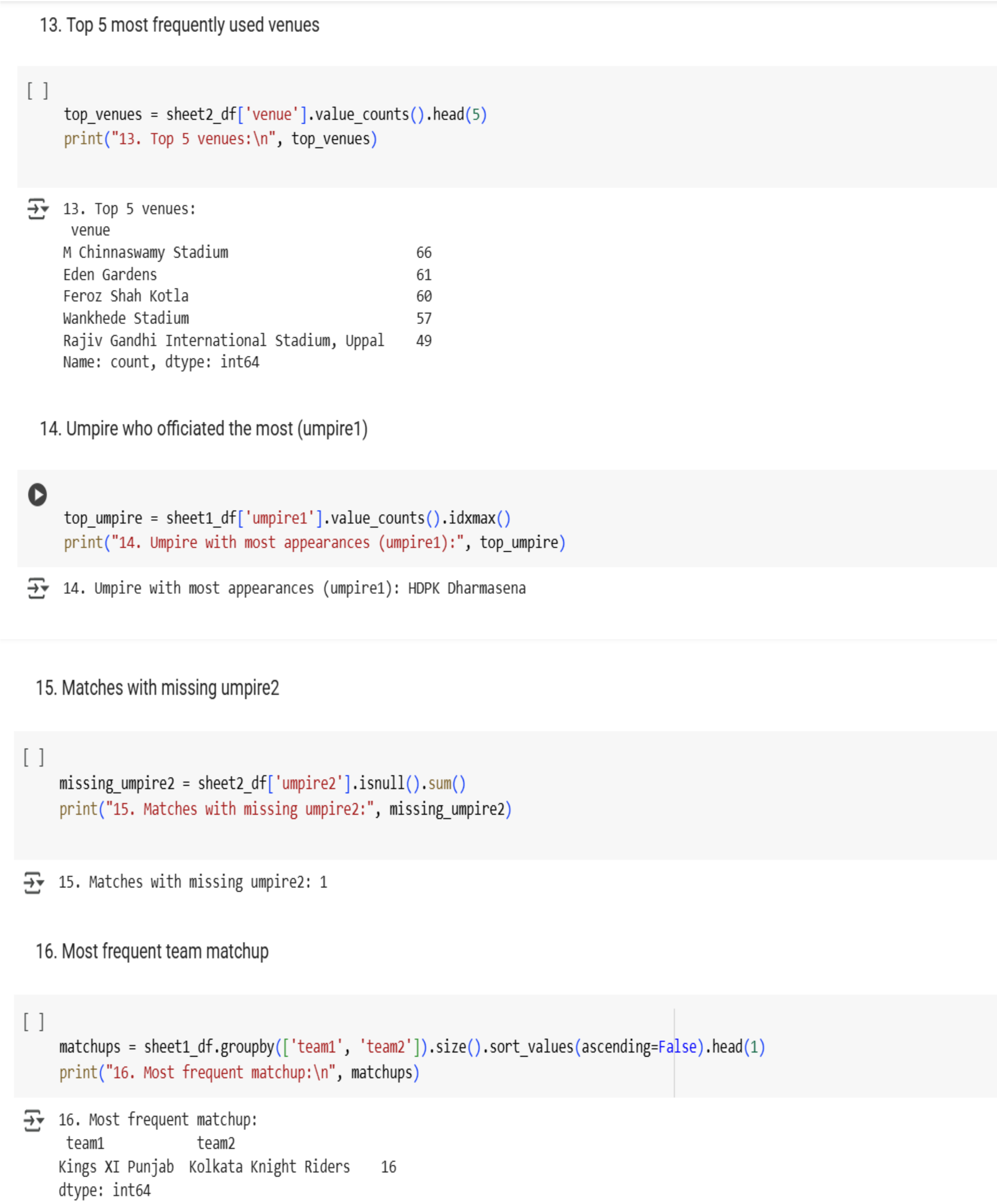
```python
toss_decision_counts = sheet1_df['toss_decision'].value_counts()
print("11. Toss decision counts:\n", toss_decision_counts)
```

```
11. Toss decision counts:
 toss_decision
field    363
bat      273
Name: count, dtype: int64
```

12. Number of cities where matches were held

```python
num_cities = sheet1_df['city'].nunique()
print("12. Number of cities where matches were held:", num_cities)
```

```
12. Number of cities where matches were held: 30
```



13. Top 5 most frequently used venues

```python
top_venues = sheet2_df['venue'].value_counts().head(5)
print("13. Top 5 venues:\n", top_venues)
```

```
13. Top 5 venues:
 venue
M Chinnaswamy Stadium                    66
Eden Gardens                             61
Feroz Shah Kotla                         60
Wankhede Stadium                         57
Rajiv Gandhi International Stadium, Uppal    49
Name: count, dtype: int64
```

14. Umpire who officiated the most (umpire1)

```python
top_umpire = sheet1_df['umpire1'].value_counts().idxmax()
print("14. Umpire with most appearances (umpire1):", top_umpire)
```

```
14. Umpire with most appearances (umpire1): HDPK Dharmasena
```

15. Matches with missing umpire2

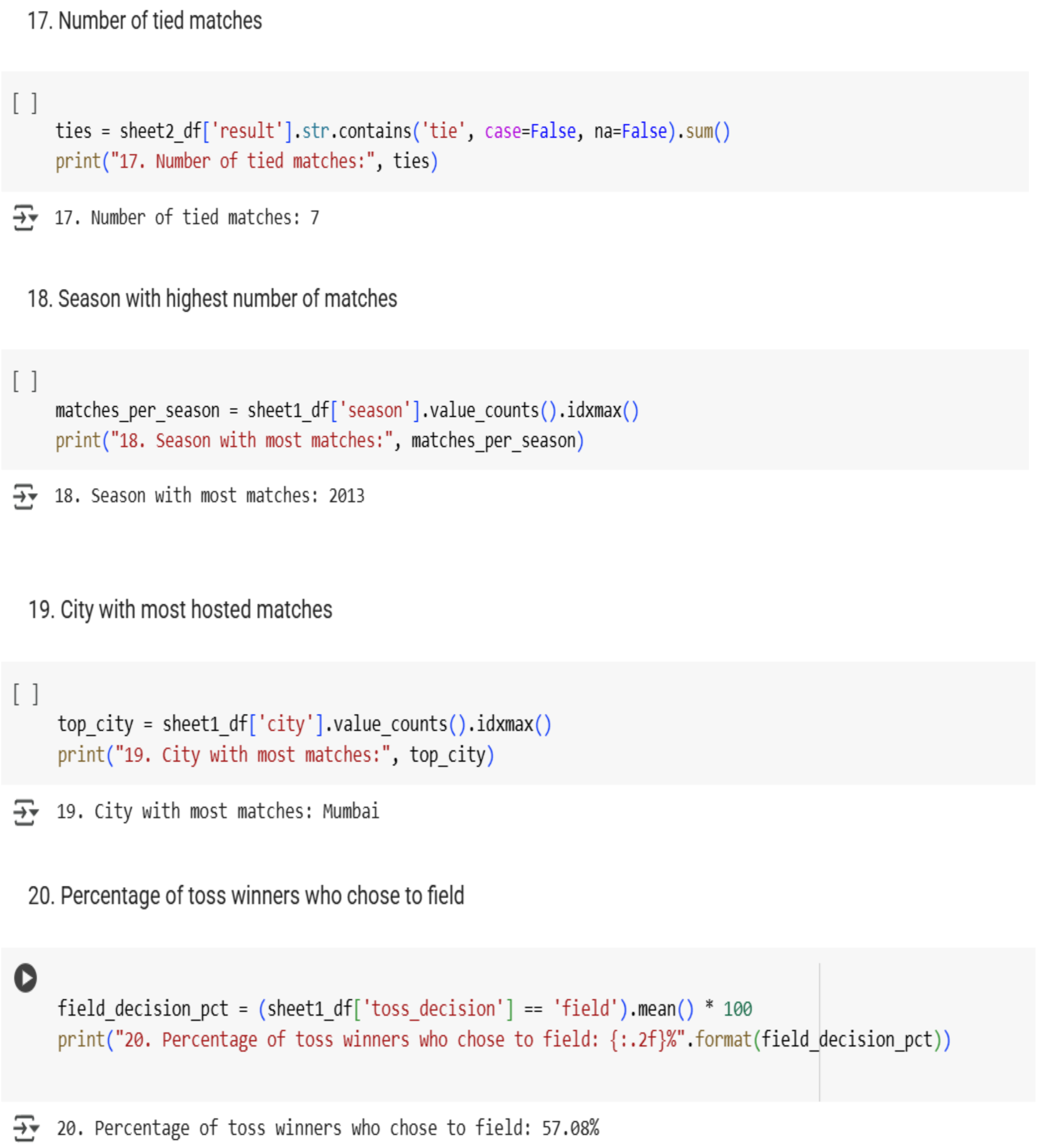
```python
missing_umpire2 = sheet2_df['umpire2'].isnull().sum()
print("15. Matches with missing umpire2:", missing_umpire2)
```

```
15. Matches with missing umpire2: 1
```

16. Most frequent team matchup

```python
matchups = sheet1_df.groupby(['team1', 'team2']).size().sort_values(ascending=False).head(1)
print("16. Most frequent matchup:\n", matchups)
```

```
16. Most frequent matchup:
 team1            team2
Kings XI Punjab  Kolkata Knight Riders    16
dtype: int64
```



17. Number of tied matches

```
ties = sheet2_df['result'].str.contains('tie', case=False, na=False).sum()
print("17. Number of tied matches:", ties)
```

```
17. Number of tied matches: 7
```

18. Season with highest number of matches

```
matches_per_season = sheet1_df['season'].value_counts().idxmax()
print("18. Season with most matches:", matches_per_season)
```

```
18. Season with most matches: 2013
```

19. City with most hosted matches

```
top_city = sheet1_df['city'].value_counts().idxmax()
print("19. City with most matches:", top_city)
```

```
19. City with most matches: Mumbai
```

20. Percentage of toss winners who chose to field

```
field_decision_pct = (sheet1_df['toss_decision'] == 'field').mean() * 100
print("20. Percentage of toss winners who chose to field: {:.2f}%".format(field_decision_pct))
```

```
20. Percentage of toss winners who chose to field: 57.08%
```