Text Preprocessing Practice Question on IPL Data

1. Team Name Normalization

1 Some team names appear in different textual formats (e.g., case differences or extra spaces). Clean the team1 and team2 columns by:

2. Converting text to lowercase

3. Removing leading/trailing spaces

4. Verifying unique team names after cleaning

2. City Name Standardization

Standardize the city column by:

1)Replacing missing city names with "Unknown"

2)Converting all city names to title case

3)Counting matches played in each city

3. Toss Decision Text Analysis

Analyze the toss_decision column:

1)Extract unique decisions

2)Count how many times each decision was taken

3)Visualize the frequency using a bar chart

4, Winner Name Extraction

From the winner column:

1)Identify and remove rows where the match result was "No Result" or "Tie"

2)Count how many matches each team won after cleaning text values

Text Preprocessing Practice Question on IPL Data

5. Player of the Match Text Frequency

Perform text analysis on player of the match:

1)Remove null values

2)Find the top 10 most frequent player names

3)Plot the results using a Seaborn bar plot

6. Venue Tokenization

Count how many matches were played in each venue and plot a bar chart for the top 10.

7. Umpire Name Cleaning

Clean umpire columns (umpire1, umpire2, umpire3) by:

1)Replacing missing values with "Not Assigned"

2)Removing duplicate umpire names per match

3)Finding the most frequently officiating umpire

8. Create a new text column match_summary by combining:

team1, team2, winner, and season

Example: "MI vs CSK – MI won in 2019"

Display sample summaries.

9. Result Type Text Analysis

Analyze the result column:

Identify different textual result types

Count their occurrences

Text Preprocessing Practice Question on IPL Data

Visualize the distribution using a count plot

10. Toss Winner vs Match Winner (Text Matching)

Compare text values in toss_winner and winner:

Create a boolean column indicating whether the toss winner also won the match

Visualize the comparison using a bar chart

11. Winner points Graph

X - team name Y- No. Of win