Dataset

IPL 2008 to 2022 All Match Dataset

Domain

Sports and Entertainment

Context

IPL is among the most famous cricket league with players coming from worldwide. What makes this series more competitive every year is the drama with every ball and hence this dataset is a complete dataset showing match information and also data for each delivery.

1. The file IPL_Matches_2008_2021.csv contains all details of the match: Venue of the match, Toss Decision, Match Winer, Man of the Match, Squads, etc.
2. The file IPL_Ball_by_Ball_2008_2021.csv contains ball by ball data of each and every match: Batsman on strike, Bowler, Extras, Runs, Wicket Ball, etc.

Attribute Information

IPL_Ball_by_Ball_2008_2022.csv

1. ID — Match ID given by ESPNCricinfo
2. Innings — Inning Number 1 - 1st Inning 2 - 2nd Inning
3. Overs — Over Number 0-19
4. Ballnumber — Ball Number
5. Batter — Batsman on Strike
6. Bowler — Bowler
7. Non-striker — Non-Striker Batsman
8. Extra_type — If the ball is an extra, this will indicate the extra type Possible Values: byes, legbyes, wides, noballs
9. Batsman_run — Run Scored by the batsman for the ball
10. Extras_run — Extra runs for the ball
11. Total_run — Total Runs for the ball
12. Non_boundary — This indicates that the 4 or 6 scored was not via an actual boundary, for example it was all run, or overthrows. 1 if overthrow
13. isWicketDelivery — Is the ball a wicket delivery
14. Player_out — Batsman getting out
15. Kind — Type of wicket taking ball
16. Fielders_involved — Fielders if any involved in the wicket taking ball
17. BattingTeam — Batting Team

IPL_Matches_2008_2022.csv

1. ID — Match ID given by ESPNCricinfo
2. City — City where the game was played
3. Date — Date of the match
4. Season — IPL Season
5. MatchNumber — Match Number for that IPL Season. Also for Qualifiers and Finals, the match number would be Final, Qualifier 2, Qualifier
6. Team1 — Team 1
7. Team2 — Team 2
8. Venue — Stadium
9. TossWinner — Team winning the toss
10. TossDecision — Toss decision: the toss winner either choosing field or bat
11. SuperOver — Did the match had super over
12. WinningTeam — Winner of the match
13. WonBy — Won By runs or wicket
14. Margin — Margin of win
15. Method — Was there D/L method in the match
16. Player_of_Match — Player of the match
17. Team1Players — Team 1 Squad Array value
18. Team2Players — Team 2 Squad Array value
19. Umpire1 — Umpire 1
20. Umpire2 — Umpire 2

Task

Your data-wrangling step must contain basic error checks, missing values checks and their treatment with proper reasoning, outlier checks, and duplicate values checks. You are allowed to explode columns and merge datasets.

Non-Visual Data Analytics: Do both uni-variate and bi-variate analysis and furnish **at least** 5 interesting insights.

And lastly, your analysis must answer the following sample questions:

1. From overs 0-19, which over has the most runs scored in, on average?
2. Top 10 umpires, who were part of the most matches
3. Which team has the highest probability of winning a match given that they had won the toss?
4. Top 10 most successful bowlers in terms of number of wickets taken
5. Calculate for each team the number of playoff matches played.
6. On which venue, the team batting second has the most advantage?
7. Which batsman in the history of IPL has been caught out the most?
8. Which bowler has taken the wicket of the batsman in que no. 7?