Dive into the heart of the exhilarating IPL 2024 season with this comprehensive dataset, offering a treasure trove of statistics on the stellar performances of cricket's brightest stars. With meticulous records spanning across various facets of the game, this dataset encapsulates the essence of each player's journey throughout the tournament.

Attributes included in the dataset:

Year: The year of the IPL season, indicating 2008 to 2024 in this case.
Player_Name: Names of the players showcasing their prowess on the cricket field.
Matches_Batted: The number of matches in which the player batted.
Not_Outs: Number of times the player remained not out while batting.
Runs_Scored: Total runs scored by the player throughout the season.
Highest_Score: Player's highest individual score in a single match.
Batting_Average: The average runs scored per dismissal.
Balls_Faced: Total number of balls faced by the player while batting.
Batting_Strike_Rate: The rate at which the player scores runs per 100 balls faced.
Centuries: Number of centuries scored by the player.
Half_Centuries: Number of half-centuries scored by the player.
Fours: Total number of boundaries (4 runs) hit by the player.
Sixes: Total number of sixes (6 runs) hit by the player.
Catches_Taken: Number of catches taken by the player in the field.
Stumpings: Number of times the player effected a stumping as a wicketkeeper.
Matches_Bowled: The number of matches in which the player bowled.
Balls_Bowled: Total number of balls bowled by the player.
Runs_Conceded: Total runs conceded by the player while bowling.
Wickets_Taken: Number of wickets taken by the player.
Best_Bowling_Match: Player's best bowling performance in a single match.
Bowling_Average: The average runs conceded per wicket taken.
Economy_Rate: The average number of runs conceded per over bowled.
Bowling_Strike_Rate: The rate at which the player takes wickets per ball bowled.
Four_Wicket_Hauls: Number of times the player took four wickets in an inning.
Five_Wicket_Hauls: Number of times the player took five wickets or more in an inning.

Data Cleaning & Preprocessing

1. Identify and handle any missing values in the dataset.
2. Are there any inconsistencies in the dataset (e.g., mismatched data types, outliers)?
3. Convert all numerical columns to appropriate data types for analysis.
4. Normalize or standardize relevant numerical features for better comparison.
5. How would you handle data if there are duplicate records?

Exploratory Data Analysis (EDA)

6. What is the distribution of the **Batting Average** for all players?
7. Which player has the highest **Batting Strike Rate** across all seasons?
8. What is the trend in **Runs Scored** for each player over different years?
9. Identify the top 3 players with the most **Centuries**.
10. Which player has the highest **Bowling Economy Rate**?

Feature Engineering

11. Create a new feature called **"Boundary Percentage"**, which calculates the percentage of runs scored from fours and sixes.
12. Create a feature called **"Batting Impact Score"** based on runs, strike rate, and average.
13. Generate a metric to evaluate a player's **all-round performance** considering both batting and bowling stats.
14. Categorize players into **Aggressive, Balanced, and Defensive** based on their strike rate.
15. Compute the **year-over-year growth** in performance metrics for each player.

Visualization & Insights

16. Create a bar chart to show the **Runs Scored vs. Balls Faced** for all players.
17. Generate a heatmap showing the correlation between different numerical features.
18. Plot a trend line showing the **performance of a single player** over multiple years.
19. Create a pie chart for the **distribution of centuries and half-centuries** across all players.
20. Develop a visualization to compare **Bowling Average vs. Bowling Strike Rate**.

Python-Based Queries (If the data is stored in a relational database)

21. Write a Python query to get the **top 5 players** with the highest **Batting Average**.
22. Write a Python query to find out which player has played the **most matches**.
23. Retrieve all players who have scored more than **500 runs in a season**.
24. Identify players who have taken at least **one five-wicket haul**.
25. Write a query to calculate the **total number of boundaries** hit by each player.