Data Premier League

Data Visualization and Prediction

Scenario: The 2024 Franchise Cricket Championship Prediction Challenge

Get ready for the ultimate challenge in cricket analysis – the "2024 Franchise Cricket Championship Prediction Challenge"! The cricketing world is abuzz with anticipation as ten top-tier teams are set to battle it out in a round-robin format. Your task is to predict and answer the burning question on everyone's minds:

"Who will win the 2024 Cricket Season ?"

Problem Description:

Utilize your analytical skills to analyze historical data, player performances, team dynamics, and any other relevant factors to make an informed prediction about the winner of the 2024 Cricket Season. Your submission should include:

- **Data Exploration and Cleaning:**
    1. Perform data cleaning on the dataset to handle any missing or inconsistent values.
    2. Investigate historical performance data of all ten teams in the previous seasons.
    3. Identify standout players, recent form and any other factors that influence the result.
- **Statistical Analysis:**
    1. Conduct statistical analyses to compare the batting and bowling averages of each team.
    2. Explore any statistical insights that could influence the championship outcome.
- **Data Visualization:**
    1. Create informative visualizations showcasing trends and patterns in the data.
    2. Visualize head-to-head performance statistics between the ten teams.
    3. Employ creative visualizations to enhance your predictive insights.
- **Prediction and Rationale:**
    1. Make a bold prediction on the team that will clinch the 2024 Cricket Championship.
    2. Provide a detailed rationale supporting your prediction, citing specific data points and analyses that shaped your decision.

- **Documentation**
    1. Prepare a comprehensive documentation summarizing your findings and methodology. Highlight key insights and approaches you took. Your documentation should be clear and concise.

In this coding competition, the dataset is organized into 10 folders, each representing a cricket team. Within each team folder, there are three subfolders for batsmen, bowlers, and all-rounders. The batsmen subfolder contains Excel files with multiple sheets, each representing the dataset of an individual batsman. The columns in the batsmen dataset include:

- **Match:** Match number.
- **Innings:** Innings played.
- **Date**: Date of the match.
- **M/Inns**: Match/Innings (1^{st} or 2^{nd})
- **Posn:** Batting position.
- **Versus:** Opponent team.
- **Ground:** Venue of the match.
- **How Dismissed**: Mode of dismissal.
- **Runs**: Runs scored.
- **B/F:** Balls faced.
- **S/R:** Strike rate.
- **Aggr:** Aggregate runs.
- **Avg:** Batting average.
- **S/RC:** Cumulative strike rate.

Sheet Name:

<batsman name>

In the bowler subfolder, the dataset includes the following columns:

- **Match:** Match number.
- **Date:** Date of the match.
- **M/Inns:** Match/Innings (1^{st} or 2^{nd})
- **Versus:** Opponent team.
- **Ground:** Venue of the match.
- **Batsman Dismissed:** Batsman dismissed by the bowler.

- **Overs:** Overs bowled.
- **Wickets:** Wickets taken.
- **S/R**: Strike rate of the bowler.
- **E/R:** Economy rate of the bowler.
- **Wkts:** Total wickets taken.
- **Avg:** Bowling average.

Sheet Name:

<bowler name>

For all-rounders, there are two separate Excel files, one for batting and one for bowling. In the batting and bowling dataset, the columns are similar to the batsmen and bowlers dataset.

Sheet name:

<allrounder name> a : Batting Data

<allrounder name> b : Bowling Data

Problem Statement: **"Who will win the 2024 Cricket Season ?"**

Participants are invited to employ their imagination, prediction skills, and visualization techniques with the given dataset. As you explore the intricacies of the cricket data through the questions provided, envision and predict compelling insights. Let your creativity and analytical prowess shine as you navigate through the diverse dimensions of the cricketing world encapsulated in this dataset.

1. **Innings Distribution:**
   How is the distribution of innings for players in the dataset? Are they more inclined towards top-order, middle-order, or lower-order positions?
2. **Versus Opponent Analysis:**
   Which opponent has proven to be the most challenging for the players in terms of average runs scored and strike rate?
3. **Ground Impact:**
   Does the batting performance vary significantly based on the ground? Identify grounds where players tend to perform exceptionally well or struggle.

4. **Dismissal Patterns:**

What are the most common modes of dismissal among the players? Analyse the distribution of dismissals (e.g., caught, bowled, LBW).

5. Display the **significance of each factor** (eg: ground, innings, versus etc …) that affect the player performance.
6. Hypothetically **swap two players** from different teams and analyse the potential impact on the winning percentage of both teams. Consider player strengths, team dynamics, and historical performances.
7. Compare the **winning percentage** of all teams against every other team

   a) considering the ground factor

   b) without considering the ground factor

8. **Consistency in Runs Accumulation:**
   Analyse the consistency of players in accumulating runs. Identify players with a consistent run-scoring pattern and those with more variable performances.
9. **Positional Impact on Performance:**
   How does the batting position impact the average runs scored and strike rate? Are certain positions associated with higher or lower performance?
10. **Comparison of Strike Rates:**
    Compare the overall strike rates of different players. Identify players with consistently high strike rates and those with lower rates.
11. **Innings-wise Performance:**
    Analyse the performance of players in the 1st and 2nd innings. Are there notable differences in runs scored and strike rates?
12. **Create the best batting line-up;**
    a) For any one team against every other team
    b) Of 2024 season
13. Who will get **orange cap (highest run scorer) and purple cap (leading wicket taker)** in 2024 season?
14. **Impact of Player Removal:**
    Explore the potential impact on a team if a key player is removed. Analyse changes in batting, bowling, and fielding dynamics of each team.
15. **Likely Struggling Team in Upcoming Season:**
    Analyse historical data to predict which team is likely to face challenges and struggle the most in the upcoming season.
16. **Batsman-Specific Dismissals:**
    Explore which batsmen are frequently dismissed by specific bowlers. Identify the batsmen who struggle the most against particular bowlers.