ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING ASSIGNMENT

NAME : SRUSHTI LAXMIKANT BOGALE
CLASS : BCA- AIML
COURSE NAME : ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING
COURSE CODE : AM207E
URN : 2023-B-09062004
TOPIC : DATA VISUALIZATION

TITLE:

“A Data-Driven Review of IPL 2023 Batting Performances"

This project aims to uncover meaningful patterns and insights from the IPL 2023 batting data using data analysis and visualization techniques. By exploring key performance metrics such as total runs, strike rates, boundary counts, batting positions, and team-wise contributions, the project highlights the standout players and teams of the season. Statistical tools like bar charts, histograms, pie charts, and scatter plots are employed to answer specific questions such as: who scored the most runs, which team dominated in six-hitting, and how batting consistency varied across players. The analysis not only identifies Shubman Gill as the season's top performer but also demonstrates how data-driven approaches can enhance the understanding of sports performance.

INTRODUCTION

This project focuses on analyzing the batting performances in the Indian Premier League (IPL) 2023 using Python and data visualization techniques. The dataset includes match-wise batting statistics such as runs scored, balls faced, strike rate, boundaries, batting position, and team contributions.

Through this analysis, we aim to uncover key insights such as the top run-scorers, strike rate trends, consistency of players, and the impact of batting order on performance. Visualization tools like bar graphs, histograms, pie charts, and scatter plots help us interpret the data effectively.

The project highlights standout performers like Shubman Gill, who led in both runs and boundaries, and evaluates team-wise contributions, with Gujarat Titans emerging as the most dominant batting side. This data-driven approach offers a clear and engaging way to understand player and team performance beyond just match results.

METHODOLOGY

The project follows a structured data analysis pipeline using Python. The steps are as follows:

- Data Collection: The dataset, containing per-match batting statistics from IPL 2023, was sourced from kaggle.
- Data Preprocessing: The dataset was cleaned by handling missing values, converting data types (e.g., strike rates and boundaries), and removing inconsistencies.
- Exploratory Data Analysis (EDA): Statistical summaries and visualizations were used to explore and understand trends in the data.
- Data Visualization: Libraries like Matplotlib and Seaborn were used to create visual insights, such as bar charts, histograms, pie charts, line plots, and scatter plots.
- Insight Extraction: Each visualization answered a specific question related to player performance, consistency, or team contributions.

RESULTS AND VISUALIZATION

The analysis yielded several key insights:

- Top Performers: Shubman Gill emerged as the highest run-scorer and most aggressive boundary hitter.
- Strike Rate Trends: Most players had strike rates between 120–140, with some outliers going significantly higher.
- Batting Position Impact: Openers contributed the most runs, while lower-order batsmen had minimal impact.
- Player Consistency: Rohit Sharma and Wriddhiman Saha remained not out the most times; Cameron Green had the highest dismissal count.
- Team Contributions: Gujarat Titans had the highest total runs and sixes, while Delhi Capitals had the lowest in both.
- Comparative Analysis: Shubman Gill outperformed Ruturaj Gaikwad in both consistency and high-scoring knocks.
- Balls Faced vs Strike Rate: Players who faced more balls tended to have more controlled strike rates.

Each of these insights was supported by visual representations like:

- Bar plots for top scorers and team contributions.
- Histograms for strike rate distribution.
- Box plots for comparing player consistency.
- Pie charts for team-wise sixes and run share.
- Line charts for batting position impact.

Conclusion & Future Scope:

Conclusion:

This project successfully leveraged data analytics to gain actionable insights from IPL 2023 batting data. It identified standout players and teams, highlighted performance trends, and showcased the importance of consistency and batting order. Shubman Gill and Gujarat Titans emerged as the leading figures in the season, while the data also shed light on strike rate behaviors and team-wise dynamics.

Future Scope:

- Bowling Analysis: Expanding the project to include bowler performance for a more balanced view.
- Match Outcome Correlation: Analyzing how batting stats impact match results.
- Year-over-Year Comparison: Comparing data across multiple IPL seasons.
- Interactive Dashboards: Creating dynamic dashboards using tools like Power BI or Tableau for real-time insights.
- Machine Learning Models: Predicting player performance or match outcomes based on historical data.