IPL DATA ANALYSIS

Group Members

Nilay Sankhe - 24

Yash Pimple - 20

INTRODUCTION

- The Indian Premier League (IPL) is one of the world's most-watched cricket leagues, known for its high-voltage matches and entertainment.

- With hundreds of matches played every season, analyzing IPL data can provide valuable insights regarding:
  - Top-performing teams and players.
  - Winning patterns based on toss, venue, or team performance.
  - Impact of winning toss on match outcomes.

- The aim of this project is to analyze past IPL seasons and extract meaningful insights from the data that can help understand match patterns, player performance, and team strategies.

OBJECTIVES OF THE PROJECT

The main objectives of conducting this IPL Data Analysis are:

1. Analyze Team Performance:
- Identify which team has been the most consistent performer in different seasons.
- Find which team performs better while batting first or chasing targets.

2. Analyze Player Performance:
- Find the top run-scorers and top wicket-takers in IPL history.
- Identify the players who have won the maximum Player of the Match awards.

3. Impact of Toss on Match Outcome:
- Check if winning the toss has any correlation with winning the match.
- Understand whether batting first or second has a higher chance of winning.

4. Venue Analysis:
- Analyze how match outcomes vary based on different stadiums or venues.
- Identify home ground advantages for specific teams.

5. Predictive Insights:
- Generate actionable insights for future IPL matches based on historical data.
- Build a base understanding of factors influencing match outcomes.

DATA COLLECTION

- The data was collected from publicly available sources like:
    - **Kaggle Datasets.**
    - **Official IPL Website (iplt20.com).**
    - **Cricket API or Web Scraping Methods**.
- The dataset primarily contained:
    - Match Details: Team names, match date, venue, winner, result, toss winner, etc.
    - Player Performance Data: Runs scored, wickets taken, strike rates, batting average, etc.
    - Venue Details: Stadium, city, toss decisions, home ground advantage.
- The data was stored in CSV (Comma Separated Values) format, which was later loaded into Python Jupyter Notebook/Colab for analysis.

DATA CLEANING AND PREPROCESSING

The collected raw data had several inconsistencies that needed to be cleaned before analysis.

Steps Taken for Data Cleaning:

1. **Handling Missing Values:**
   - Several fields like player names, match results, or scores had missing values.
   - Used Pandas to fill or drop null values.

2. **Removing Duplicates:**
   - Some matches or player data were repeated in the dataset.
   - Used the drop_duplicates() function to remove them.

3. **Standardizing Data:**
   - Some player names were spelled differently (e.g., *M.S. Dhoni, MS Dhoni*).
   - Standardized them using string operations.

4. **Date-Time Conversion:**
   - Converted match date from string to DateTime format for time-based analysis.

5. **Data Type Conversion:**
   - Converted numerical columns like scores, wickets, and match results to integers/floats for easy analysis.

EXPLORATORY DATA ANALYSIS (EDA)

Exploratory Data Analysis (EDA) is the process of analyzing and visualizing data to uncover key insights and patterns.

A. Team Performance:

• Analyzed the win percentage of teams.

• Found that Mumbai Indians (MI) and Chennai Super Kings (CSK) have the highest winning percentages.

B. Toss Impact:

• Analyzed if winning the toss increases the chance of winning.

• Found that teams winning the toss and opting to chase had a slightly higher winning rate.

C. Player Performance:

• Identified top run-scorers like Virat Kohli, Rohit Sharma and top wicket-takers like Jasprit Bumrah, Rashid Khan.

• Recognized players with the most Player of the Match awards.

D. Venue Analysis:

• Found that home ground advantage plays a crucial role in match outcomes.

• Teams like CSK (at Chennai) and MI (at Wankhede) had higher winning percentages.

Thankyou
Any Questions ?