

IPL Data Analysis Project - Documentation

1. Objective

- Analyze IPL data to extract performance trends.
- Understand factors influencing match outcomes such as toss decisions and win margins.
- Identify top players and effective team strategies using data.

2. Data Source & Collection

- Source: Kaggle IPL Dataset.
- Files Used:
 - matches.csv: Contains match-level information.
 - deliveries.csv: Contains ball-by-ball data.

3. Data Understanding & Preparation

- Merged matches and deliveries datasets for complete match insights.
- Parsed seasons, extracted key fields.
- Handled missing data using SQL queries.

4. Descriptive Statistics

- Matches Analyzed: Over 950 matches.
- Avg Score per Match: Around 160.
- Highest Win Margin: 146 runs.
- Most Common Toss Decision: Field First.

5. Exploratory Data Analysis (EDA)

- Top Batsmen: Identified based on total runs.
- Toss Decisions: Majority choose to field first.
- Win Types: Chasing teams win more frequently.
- Toss vs Match Winner: Weak correlation.
- Win Margins: Most games have close margins.
- Season-wise Trends: Run scoring increasing over seasons.

6. Statistical Analysis

- Correlation Analysis: Weak relation between toss and win.
- Hypothesis Test: Toss decision is not a strong predictor of outcome.

7. Insights & Recommendations

- Teams winning toss often choose to field first.
- Chasing teams tend to win more often—suggests preference for chasing.
- Consistent top scorers drive team success.
- Strategy refinement in powerplay and death overs is key.

8. Tools Used

- Python (Pandas, Seaborn, Matplotlib)
- SQL
- Power BI