Project Report: Cricket Data Analysis using SQL, Web Scraping, Matplotlib, and Power BI

1. Introduction

The project focuses on analyzing cricket datasets across multiple formats (Test, ODI, T20, and IPL) using SQL for querying, **web scraping** for data collection, **Matplotlib** for statistical visualizations, and **Power BI** for interactive dashboards. The goal is to derive insights into batting performance, bowling efficiency, team strategies, and match outcomes.

2. Objectives

- To scrape and prepare cricket datasets from online sources.
- To query and analyze datasets using SQL.
- To visualize performance statistics using Matplotlib.
- To design **two Power BI dashboards**: one for ICC formats (Test, ODI, T20) and another exclusively for IPL.

3. Tools and Technologies

- **Web Scraping**: For collecting cricket data from web sources.
- **SQL**: For querying and aggregating structured data.
- **Matplotlib**: For statistical graphs and custom charts.
- **Power BI**: For dashboard creation and interactive visualizations.
- **Dataset**: Matches and deliveries data across Test, ODI, T20, and IPL.

4. SQL Queries and Insights

(Same queries as before — match counts, top batsmen, strike rates, wicket takers, economy rates, team wins, toss outcomes, player of match awards, and chasing vs defending wins.)

- These queries provide structured insights into batting, bowling, and match results.

5. Web Scraping

- Used **Python (BeautifulSoup, Pandas, Requests)** to scrape match and player data from cricket websites.
- Cleaned and transformed scraped data into structured tables compatible with SQL and Power BI.

6. Matplotlib Visualizations

- Developed additional visualizations for data validation and trend analysis:
- Run distributions of top batsmen.

• Wickets distribution per bowler.
• Economy rate comparisons.
• Win margin histograms.
• These static plots supplement SQL and Power BI findings.

7. Power BI Dashboards

• **Dashboard 1: ICC Formats (Test, ODI, T20)**
• Match counts comparison across formats.
• Top 10 batsmen and bowlers.
• Strike rate analysis.

• Economy and win records.

• **Dashboard 2: IPL Analysis**

• Batsman strike rate vs runs.
• Bowler economy vs wickets.
• Toss impact on wins.
• Chasing vs defending wins.
• Player of the match leaders.

8. Results & Findings

• Top performers vary by format: T20/IPL highlight strike rates, while Test cricket emphasizes consistency.
• Certain bowlers sustain low economy across formats.
• Toss advantage is notable in IPL but not absolute.
• Chasing proves more successful in IPL compared to defending totals.
• Matplotlib visualizations confirmed statistical distributions supporting Power BI trends.

9. Conclusion

This project demonstrates the integration of **web scraping, SQL, Matplotlib, and Power BI** to analyze cricket data. It highlights key players, bowling strengths, and team strategies across cricket formats, supplemented by two comprehensive dashboards.

10. Future Scope

• Automating web scraping pipelines for real-time data collection.
• Enhancing dashboards with predictive modeling using ML.
• Combining Matplotlib plots with Power BI dashboards for hybrid insights.

Project By: [Your Name]
Tools Used: Web Scraping, SQL, Matplotlib, Power BI
Dashboards: ICC Dashboard & IPL Dashboard