Project Report

Problem Statement:

The goal of the project is to identify a cricket team of 11 players who can consistently perform well in T20 matches. The team must be able to score at least 180 runs on average and defend a target of 150 runs. The selection criteria will be based on various performance metrics, such as batting average, strike rate, boundary percentage, and bowling economy, across different roles like openers, middle-order batsmen, finishers, all-rounders, and fast bowlers. The objective is to analyze and filter players based on these defined parameters to form the best possible team.

Purpose of the Data Analysis Project:

This project aims to analyze historical cricket data and apply data-driven filters to select an ideal T20 cricket team. The analysis will identify players' strengths and weaknesses in various roles, enabling the formation of a team that can consistently deliver high-performance outcomes both in scoring runs and defending targets. The data will be filtered using specific criteria that align with the requirements of openers, middle-order players, finishers, all-rounders, and specialist bowlers.

High-Level Algorithm:

1. Openers Selection: (2-3) players

- **Input**: Players' batting statistics such as average, strike rate, innings batted, boundary percentage, and batting position.
- **Criteria**:
    - Batting Average > 30
    - Strike Rate > 140
    - Innings Batted > 3
    - Boundary % > 50
    - Batting Position < 4
- **Output**: A list of potential openers who meet these criteria.

2. Middle Order/Anchors Selection: (2-3) players

- **Input**: Similar to openers but with slightly modified criteria.
- **Criteria**:
    - Batting Average > 40
    - Strike Rate > 125
    - Innings Batted > 3
    - Avg. Balls Faced > 20
    - Batting Position > 2
- **Output**: A list of middle-order players.

3. Finisher/Lower Order Anchor Selection: 2 players

- **Input**: Focuses on lower-order batters and their ability to finish the innings.
- **Criteria**:
    - Batting Average > 25
    - Strike Rate > 130
    - Innings Batted > 3
    - Avg. Balls Faced > 12
    - Batting Position > 4
    - Innings Bowled > 1 (for players who can also bowl)
- **Output**: A list of lower-order batters/finishers.

4. All-rounders Selection: (2-3) players

- **Input**: Players who contribute both with bat and ball.
- **Criteria**:
    - Batting Average > 15
    - Strike Rate > 140
    - Innings Batted > 2
    - Bowling Economy < 7

    - Bowling Strike Rate < 20
    - Innings Bowled > 4
    - Bowling Average < 20
    - Dot Ball % > 40
- **Output**: A list of all-rounders.

5. Specialist Fast Bowlers Selection: (2-3) players

- **Input**: Bowling statistics of fast bowlers.
- **Criteria**:
    - Innings Bowled > 4
    - Bowling Economy < 7
    - Bowling Strike Rate < 16
    - Bowling Average < 20
    - Dot Ball % > 40
- **Output**: A list of specialist fast bowlers.

Final Report Summary for T20 Team Selection Project

1. JavaScript (Web Scraping):

- We utilized JavaScript to perform web scraping and collected raw T20 cricket data in JSON format, which we stored in the t20_json_files. This dataset included detailed match statistics, player performances, and related data points, ensuring we had all the necessary raw data for further analysis.

2. t20_json_files (Raw Data):

- The raw JSON data contained comprehensive details on player statistics such as runs scored, balls faced, boundaries hit, bowling figures (e.g., economy, dot balls), and match-specific details. This raw data was unstructured and required cleaning and standardization for further use.

3. Data Cleaning (Python):

- In the Data Cleaning 20.ipynb notebook, we processed and cleaned the raw JSON data to produce well-structured CSV files, referred to as t20_csv_files (Cleaned Data). Data cleaning involved removing inconsistencies, handling missing values, and ensuring all critical performance metrics (batting average, strike rate, boundary percentage, bowling economy, etc.) were accurately calculated and stored.

4. t20_csv_files (Cleaned Data):

- The cleaned CSV files contained the necessary data fields to support performance analysis and player role filtering. At this stage, the data was structured in a way that aligned with the requirements for role-specific analysis (openers, middle-order players, bowlers, etc.).

5. **Power Query/Excel Transformations:**

- After cleaning, the data was imported into Power Query/Excel for further transformations. We used DAX queries to calculate key performance measures such as:
- Total Runs, Total Innings Batted, Total Innings Dismissed etc.
- These measures were crucial for performing in-depth player evaluations and identifying top performers for various roles.

6. Star Schema:

- We employed a Star Schema model to organize our data for reporting and analysis. This schema was used to structure the data into fact tables and dimension tables.
- The Fact Table contained match and player performance metrics (e.g., runs scored, balls faced, wickets taken, etc.).

7. Cricket.pbix (Power BI Report):

- The Cricket.pbix Power BI report visualized key insights from the data. It highlighted players' performance across matches and roles using calculated measures from the DAX queries.
- The report also utilized Player Selection logic and visualizations to help identify ideal players based on both batting and bowling matchups. Visualizations included performance trends, key comparisons, and player callouts (using Display Text and Color Callout Value) to clearly highlight standout players.