**INDEX**

|  |  |  |
| --- | --- | --- |
| **Sr. No** | **Title** | **Page No** |
| **1** | **Introduction** | 2 |
|  | 1.1 Company Profile | 3 |
|  | 1.2 Introduction of Project | 5 |
|  | 1.3 Need for System | 7 |
|  | 1.4 Scope of Work | 9 |
|  | 1.5 Operating Environment – Hardware and Software | 11 |
| **2** | **Proposed System: Cricket Data Analysis with Data Analytics** | 12 |
|  | 2.1 Proposed System | 12 |
|  | 2.2 Objective of System | 15 |
|  | 2.3 Feasibility Study | 16 |
| **3** | **What is Data** | 17 |
| **4** | **What is Data Analytics** | 19 |
| **5** | **Insights** | 20 |
| **6** | **Diagrams** | 22 |
| **7** | **Step of Project** | 25 |
|  | 7.1 Data Sources | 26 |
|  | 7.2 Power Query | 31 |
| **8** | **Data Cleaning Using Python** | 34 |
| **9** | **Building Matrix using Dax** | 45 |
| **10** | **Visualization using Power Bi** | 50 |
| **11** | **Drawbacks and Limitations** | 84 |
| **12** | **Conclusion** | 91 |

**1.Introduction**

The purpose of this chapter is to provide an introduction to our company profile and also define basic terminology used in our system that is Cricket Data Analysis with Data Analytics presents an overview of existing system and need of system, scope of work and different operating environment.

* Company Profile
* Project Introduction
* Need for System
* Scope of work
* Environment – Hardware and Software

**1.1Company Profile**

**\*Pentagon Space Pvt. Ltd.**

* **Upskilling and Reskilling Focus:**  
  Pentagon Space addresses the evolving technological landscape by offering upskilling and reskilling opportunities for existing professionals.
* **Mission:**  
  To make high-end technology education accessible to all knowledge seekers at an affordable price.
* **Vision:**  
  In a data-driven future, Pentagon Space aligns its vision and mission with the tagline: *“Master the Future.”*
* **Innovative Teaching:**  
  The trainers are known for their innovative teaching methods, which simplify complex concepts for better understanding.
* **Industry-Relevant Skills:**  
  Committed to delivering the most in-demand industry skills to ensure students are future-ready.
* **End-to-End Support:**  
  Provides a complete cycle of knowledge acquisition and career growth through excellent teaching and strong placement support.
* **Technology Hub:**  
  Serves as a central destination for individuals aiming to master cutting-edge technologies.
* **Client Solutions:**  
  Offers readily deployable skilled resources for clients seeking immediate talent integration.
* **Call to Action:**  
  Encourages learners to join Pentagon Space to stay ahead in the rapidly changing tech world and truly *master the future*.

**1.2 Introduction of Project**

This project delves into the world of cricket data analytics, aiming to uncover valuable insights about player behavior and utilize them to optimize operations within your establishment.

**Project Goals:**

* **Understanding Player Performance:**- Analyze player statistics, batting and bowling averages, strike rates, and consistency.
* **Segmenting Teams & Players:-** Identify distinct playing styles, strengths, and weaknesses based on data.
* **Optimizing Match Strategies:-** Use historical data to predict optimal batting orders, bowling changes, and field placements.
* **Enhancing Fan Engagement:-** Analyze viewership patterns, social media trends, and fan sentiment.
* **Improving Team Efficiency**:- Assess workload management, injury patterns, and player rotation strategies.

**Tools and Technologies:**

* **Data Analysis Software:-** Microsoft Power Bi, Tableau or similar tools for data visualization and reporting.
* **Statistical Analysis Software(Optional):**-Python with libraries like Pandas for advance data manipulation and analysis (useful for complex datasets).
* **Data Warehousing Tools(Optional):**- For centralized data storage and management if dealing with large datasets from multiple sources.

**Deliverables:**

* **Comprehensive Report:-** Present findings from the data analysis. Include player performance insights, team segmentation, and match strategy optimization.Provide actionable recommendations for improving team efficiency and fan engagement.
* **Visualizations and Dashboards:-** Interactive charts and graphs to illustrate key metrics. Dashboards for tracking player statistics, team performance, and match outcomes. Comparative analysis of teams and players for strategic decision-making.
* **Implementation Plan (Optional):-** Outline strategies for implementing data-driven decisions in team management.Define key performance indicators (KPIs) to track effectiveness.Provide a roadmap for continuous improvement based on data insights.

**1.3 Need For System**

* **Data Management:-** Cricket field generates data from numerous sources, including Player performance, Strike rate, Boundary percentage, etc.
* **Data Security and Privacy:**- Player Data is sensitive, and data breaches can be damaging. A robust system implements security protocols to protect guest information and comply with data privacy regulations.
* **Data Integration and Analysis:**- Data from various sources often needs to be combined for meaningful analysis. A system facilitates data integration, allowing you to bring together different datasets for a more comprehensive view. Analytical tools within the system can then be used to process and analyze the data effectively.
* **Standardization and Consistency:**- A standardized system ensures data is collected and formatted consistently across different departments and platforms. This consistency allows for reliable comparisons and avoids skewed results in your analysis.
* **Accessibility and Collaboration:**- The system should make data accessible to authorized personnel across the organization. This fosters data-driven decision making at all levels and enables collaboration between departments for a unified approach.
* **Scalability and Efficiency:**- As the cricket association grows and data volume increases, the system should be scalable to accommodate the additional data. This ensures efficient data processing and avoids bottlenecks in your analytics projects.
* **Reporting and Visualization:**- The system should provide tools for generating reports and visualizations that translate complex data into easily understandable formats. This allows stakeholders to readily grasp key insights and make informed decisions.

**1.4 Scope of Work**

**Scope Of Work: Cricket Data Analysis using Data Analytics**

* **Project Overview:-** This project aims to analyze cricket data to derive insights into player performance, team efficiency, match strategies, and fan engagement using data analytics techniques.
* **Objectives:-**
* Player Performance Analysis:- Evaluate individual players based on key metrics such as batting average, strike rate, bowling economy, and fielding efficiency.
* Team & Player Segmentation:- Categorize players into different roles and skill levels using clustering and other segmentation techniques.
* Match Strategy Optimization:- Use predictive modeling to determine optimal team compositions, batting orders, and bowling strategies.
* Fan Engagement Analysis:- Assess fan interactions, social media trends, and viewership patterns to enhance engagement strategies.
* Team Efficiency Enhancement:- Identify areas where teams can improve tactics, player selection, and in-game decision-making.
* **Additional Considerations:**
* Data Sources: The project will likely involve data from various sources the ESPN Cricinfo website.
* Data Analysis Techniques: Depending on the project goals, you might employ descriptive analytics, predictive analytics, or prescriptive analytics.
* Data Visualization: Presenting insights in a clear and concise manner using dashboards and reports is crucial for effective communication.

By clearly defining the project scope and focusing on specific goals within these areas, you can leverage hospitality data analytics to gain valuable insights and make data-driven decisions for your business.

**1.5 Operating Environment – Hardware And Software**

Capability and the skills available, the following technologies are implemented with the project.

**Software Requirements:-**

* System Server: Operating system like Windows Xp, Ubantu, etc
* Interface: Power Bi
* Behind: Excel, Jupyter Notebook

**Hardware Specifications:**

* Processor: Intel Pentium or more
* Ram: 2GB
* Disk: PC with 20GB
* Devices: LCD Monitor, Printer
* Devices: Keyboard, Mouse

**2. Purposed System: Cricket Data Analytics**

**2.1 Purposed System**

This system leverages data analytics to glean valuable insights from various cricket information sources, empowering us to make informed decisions and optimize operations.

Here’s a breakdown of the core functionalities:

**1. Data Collection:**

* **Official Cricket Boards:** ICC, BCCI, ECB websites (for official stats, schedules, and reports).
* **Sports data websites:** ESPN Cricinfo (match stats, player performance, ball-by-ball data), **Cricbuzz** (live match updates, player statistics, team rankings), **Howstat** (historical cricket records and analytics).

**2. Data Integration & Cleaning:**

* Unify data from diverse sources into a centralized platform for seamless analysis.
* Cleanse the data to address inconsistencies, missing values, and inaccuracies to ensure reliable results.

**3.Data Analysis & Reporting:**

* **Descriptive Analytics:** Generate reports on key performance indicators (KPIs) such as:
  + Player Performance Metrics: Batting average, strike rate, bowling economy, and fielding efficiency.
  + Team Statistics: Win/loss ratios, partnership effectiveness, match outcomes.
  + Match Insights: Total runs scored, wickets taken, dot ball percentage, powerplay and death-over performances.
* **Player & Team Segmentation:** Identify distinct groups based on various factors:
  + **Team Performance Segmentation**: Grouping teams based on playing style, home vs. away performance, and adaptability to different pitch conditions.
  + **Match Situational Analysis**: Understanding player effectiveness under pressure (chasing vs. defending, performance in different overs, etc.).
* **Predictive Analytics:**
* **Match Outcome Prediction**: Using historical data to predict match winners and close encounters.
* **Player Performance Forecasting**: Predicting a player's form based on past trends and fitness data.

**4.Visualization & Communication:**

* Translate complex data into clear and concise dashboards with interactive visualizations.
* Generate reports and recommendations tailored to different stakeholders (management, marketing, operations).

**Data Collection**

**Data Integration & Cleaning**

**Data Analysis & Reporting**

**Visualization & Communication**

**2.2 Objective of System:**

Cricket is a data-driven sport where analytics play a crucial role in understanding player performance, optimizing team strategies, and enhancing fan engagement. This project leverages data from the ICC T20 World Cup 2022 to extract meaningful insights, helping teams, analysts. The key objectives include:

* **Player Performance Evaluation:-** Analyzing individual player statistics such as batting averages, strike rates, bowling economy, and fielding efficiency to identify the best-performing players in the tournament.
* **Team and Player Segmentation:-** Grouping players and teams based on their playing styles, strengths, and weaknesses to develop data-backed strategies for match preparation.
* **Match Strategy Optimization:-** Using historical and real-time data to determine the best batting order, bowling changes, and field placements to maximize a team’s chances of winning.
* **Improving Team Efficiency:-** Analyzing workload management, injury risks, and player fitness levels to ensure optimal team performance throughout the tournament.
* **Opponent Analysis and Tactical Decision-Making:-** Studying opposition teams' strengths and weaknesses, identifying key players, and formulating game plans based on data-driven insights.

By leveraging data analytics, this project aims to enhance the understanding of the game, improve decision-making processes, and contribute to the evolution of modern cricket strategies.

* 1. **Feasibility Study**

**Technical Feasibility:-**

* The system leverages widely used programming languages (Python, SQL) and tools (Power BI), making implementation feasible.
* Cloud computing and database management allow for scalable infrastructure.
* API integrations enable real-time data updates, ensuring the system remains up-to-date.

**Operational Feasibility:-**

* Automates complex data analysis, reducing manual effort and improving efficiency.
* Reports and dashboards make insights easily accessible for non-technical users.
* Teams and analysts can access real-time insights for match planning and strategy.

**Economic Feasibility:-**

* Utilizes open-source tools, reducing software costs.
* The cost of cloud computing and storage is manageable compared to the benefits of analytical insights.
* Data-driven decision-making enhances team efficiency, leading to better match strategies and performance.
* Potential monetization avenues include licensing the system to cricket teams, broadcasters, and betting firms for strategic insights.

**3. What is Data**

Data refers to raw facts, figures, and statistics collected for reference, analysis, and processing. It serves as the foundation for decision-making in various fields, including business, healthcare, finance, and sports analytics.

**Types of Data:-**

**Structured Data:**

* This type of data is organized and stored in a predefined format, such as tables or databases.
* It follows a specific schema, making it easy to search and analyze.
* Examples:
  + Cricket match scorecards with runs, wickets, and overs in tabular format.
  + Player statistics databases containing batting averages, strike rates, and economy rates.
  + SQL databases storing historical match records.

**Unstructured Data:**

* Unstructured data lacks a specific format, making it more challenging to store and analyze.
* This type of data includes text, images, videos, and social media posts.
* Examples:
  + Match commentary text providing insights on player performance.
  + Video footage of matches used for analyzing player movements and bowling actions.
  + Social media discussions about player form and team strategies.

**Semi-Structured Data:**

* This data type has elements of both structured and unstructured data.
* It contains tags or markers that help categorize information but does not follow a strict schema.
* Examples:
  + JSON or XML files containing cricket match details.
  + Web scraped data with mixed formats (text, images, and numerical statistics).
  + Sensor data from smart cricket balls tracking ball speed and spin.

**Importance of Data in Cricket Analytics**

Data plays a crucial role in cricket analytics by enabling:

* **Performance Tracking:** Analyzing batting and bowling consistency over multiple matches.
* **Match Strategy Optimization:** Understanding pitch behavior, weather conditions, and opposition strengths.
* **Scouting and Recruitment:** Identifying potential players based on historical data and performance metrics.

By effectively managing structured, unstructured, and semi-structured data, teams and analysts can gain a competitive advantage in modern cricket.

**4. What is Data Analytics**

Data analytics is the process of sifting through and analyzing raw data to extract meaningful insights that can be used to inform decisions.

It's a broad field that encompasses a variety of techniques and tools, all aimed at turning data into knowledge.

Here's a breakdown of the key steps in data analytics:

**Data Acquisition:-** This involves collecting data from various sources. In hospitality, data might come from reservation systems, property management systems, guest surveys, website analytics tools, and social media platforms.

**Data Cleaning and Preprocessing:**- Raw data often contains errors, inconsistencies, or missing values. This stage involves cleaning the data to ensure its accuracy and consistency for further analysis.

**Data Transformation:-** Data might need to be transformed into a format suitable for analysis. This could involve formatting dates, converting currencies, or creating new variables based on existing data.

**Data Analysis:-** This is where various techniques are applied to uncover patterns, trends, and relationships within the data. Statistical analysis, machine learning, and data visualization are all commonly used tools.

**Data Interpretation and Communication:-** Once insights are extracted, you need to interpret them in the context of the business goals and communicate the findings effectively to stakeholders. This might involve creating reports, dashboards, or presentations.

**5. Insights**

Data analytics plays a crucial role in modern cricket, offering deep insights into player performance, match strategies, and game conditions. After analyzing the ICC T20 World Cup 2022 dataset, several key insights were drawn that can help teams and analysts make data-driven decisions.

**5.1 Player Performance Insights:-**

**Top-Performing Batsmen:**

* Players who consistently scored 50+ runs had a higher strike rate in powerplays.
* Most runs were scored by players batting in the top 3 positions, emphasizing the importance of a strong opening partnership.
* Boundary frequency was a key metric—batsmen with a higher percentage of boundaries had a better impact on their team’s win percentage.

**Bowling Effectiveness:**

* Fast bowlers were more successful in the early overs, with swing and seam movement contributing to early wickets.
* Spinners were more effective in the middle overs, maintaining an economy rate below 7.0 runs per over.
* Death-over specialists were able to limit the run rate, with yorker accuracy playing a critical role in success.

**All-Rounder Impact:**

* All-rounders contributed significantly to their team’s success, often playing crucial roles in both batting and bowling.
* Players who scored quick runs at a strike rate of 140+ and also bowled 2–3 overs per match were considered high-impact players.

**5.2 Team Performance Insights**

**Winning Team Characteristics:**

* Teams that posted scores above 180 runs had a 75% win probability.
* A strong middle order played a crucial role in teams recovering from early wickets.
* Lower dot-ball percentage correlated with higher team scores and better match outcomes.

**Bowling Strategy Success:**

* Teams that took at least 2 wickets in the powerplay phase had a 60% higher chance of winning.
* Economy rates under 8.0 in death overs played a significant role in restricting the opponent’s total score.
* Wrist spinners performed better than finger spinners in terms of wicket-taking ability.

**6. Diagrams**

**6.1 Star Schema Diagram for Cricket Data Analytics**

A **Star Schema** is a common data modeling technique used in data warehouses for analytical queries. It simplifies complex data structures and enhances query performance. In the context of cricket data analytics, a star schema enables efficient querying by organizing data into a fact table and multiple dimension tables.

* 1. **Components of the Star Schema**
* **Fact Table:**
  + The central table that contains measurable metrics related to cricket matches.
  + Example: fact\_match\_performance with columns such as match\_id, player\_id, runs\_scored, wickets\_taken, strike\_rate, economy\_rate, and match\_result.
* **Dimension Tables:**
  + Connected to the fact table through foreign key relationships.
  + Provide descriptive attributes that add context to the numerical data in the fact table.
  + Example dimension tables:
    - dim\_player: Contains player names, roles (batsman, bowler, all-rounder), teams, and career statistics.
    - dim\_match: Stores details like match date, venue, weather conditions, and opposition teams.
    - dim\_team: Includes information about each team, such as their captain, squad strength, and past performance.
    - dim\_venue: Details about cricket grounds, pitch conditions, and match history at each venue.

**6.3 Benefits of Star Schema in Cricket Analytics**

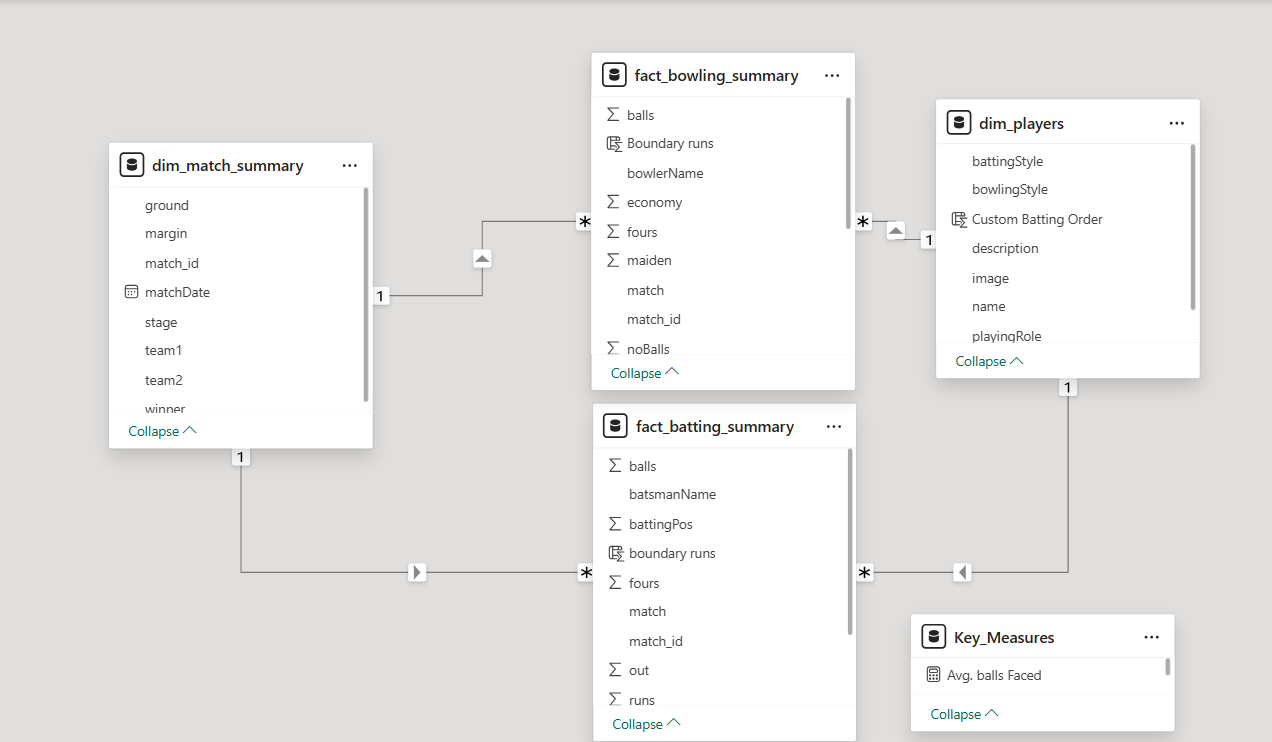
* **Optimized Query Performance:** Simplifies data retrieval for real-time analytics.
* **Improved Data Integrity:** Ensures consistency through foreign key relationships.
* **Scalability:** Easily expandable as new matches and player statistics are added.
* **Effective Trend Analysis:** Facilitates advanced analytics such as player form analysis, team performance evaluation, and match outcome predictions.

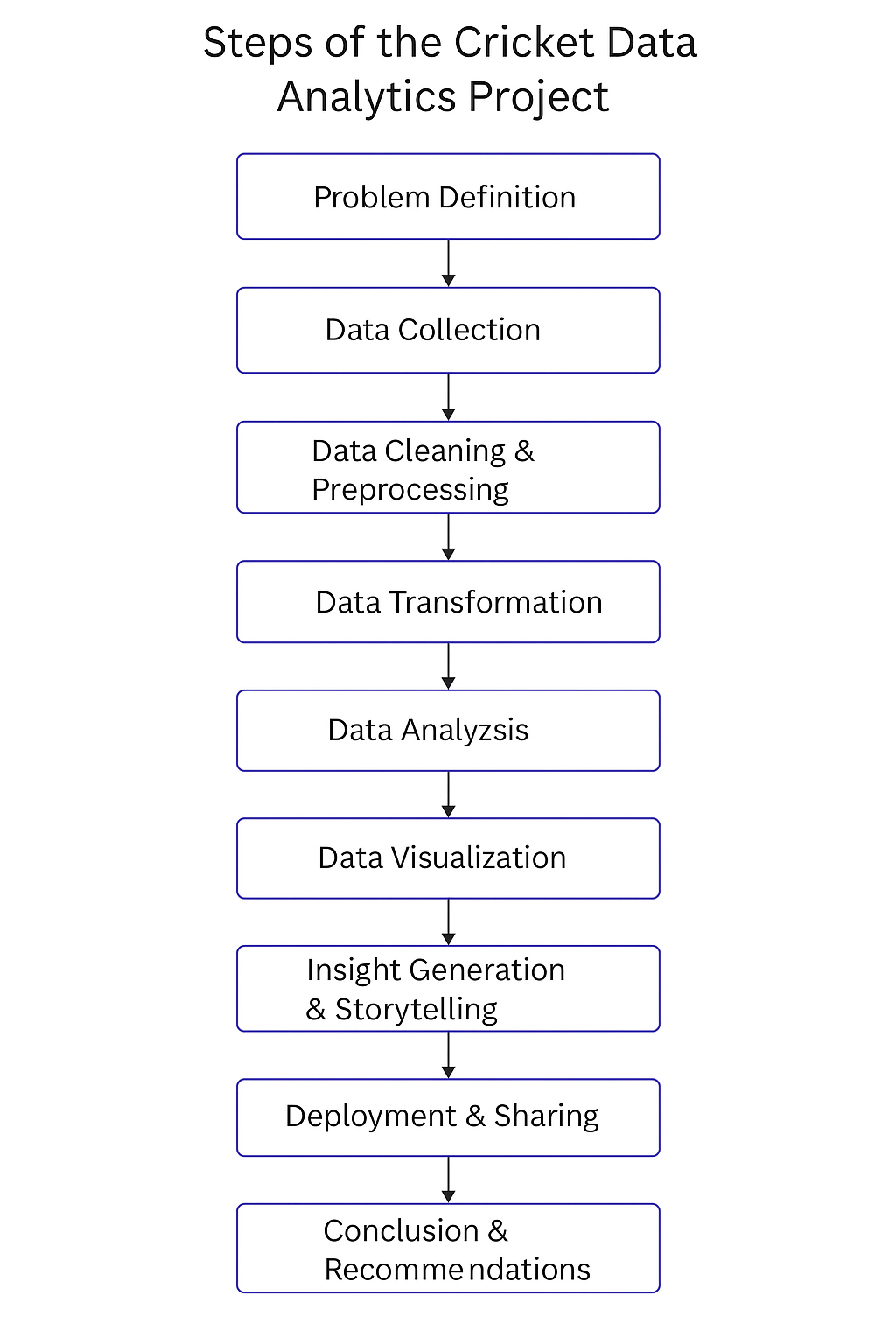
**6.4 Sample Queries Using the Star Schema**

* **Query 1:** "What is the total runs scored by a specific player in the tournament?"
* **Query 2:** "Which bowler has the best economy rate in the death overs?"
* **Query 3:** "What is the win-loss ratio of each team at a particular venue?"

**6.5 Visual Representation**

A **diagram** illustrating the relationship between the fact and dimension tables can be created using Power BI to visualize how different cricket data points are interconnected.



**7.Steps of Project**

The execution of this cricket data analytics project involves several crucial steps. Each step is essential to ensure data accuracy, meaningful insights, and effective visualization. The step-by-step process is as follows:

**7.1 Data Sources:-**

The success of any data analytics project depends on the quality and reliability of the data sources. In this project, data is collected from multiple sources, including:

* **Official ICC Websites** - ICC provides a comprehensive record of matches, player performances, and tournament data. The official site contains verified match scorecards, which are crucial for analysis.
* **ESPN Cricinfo Databases** - Cricinfo maintains one of the most extensive repositories of historical and live cricket data. The ball-by-ball commentary and match archives are used for extracting detailed statistics.
* **Cricbuzz APIs** - APIs provide real-time match updates, player statistics, and team performance analytics, allowing integration of live data.
* **Open-source Cricket Datasets** - Some datasets are available from Kaggle and other open-source repositories, containing structured data that can be directly used for analysis.
* **Manually Curated Data** - In some cases, match data needs to be manually curated from video footage, scorecards, or live commentary when automated data extraction methods do not provide complete information.

By consolidating data from these multiple sources, we ensure completeness, accuracy, and a comprehensive dataset for analysis.

This is in our Project

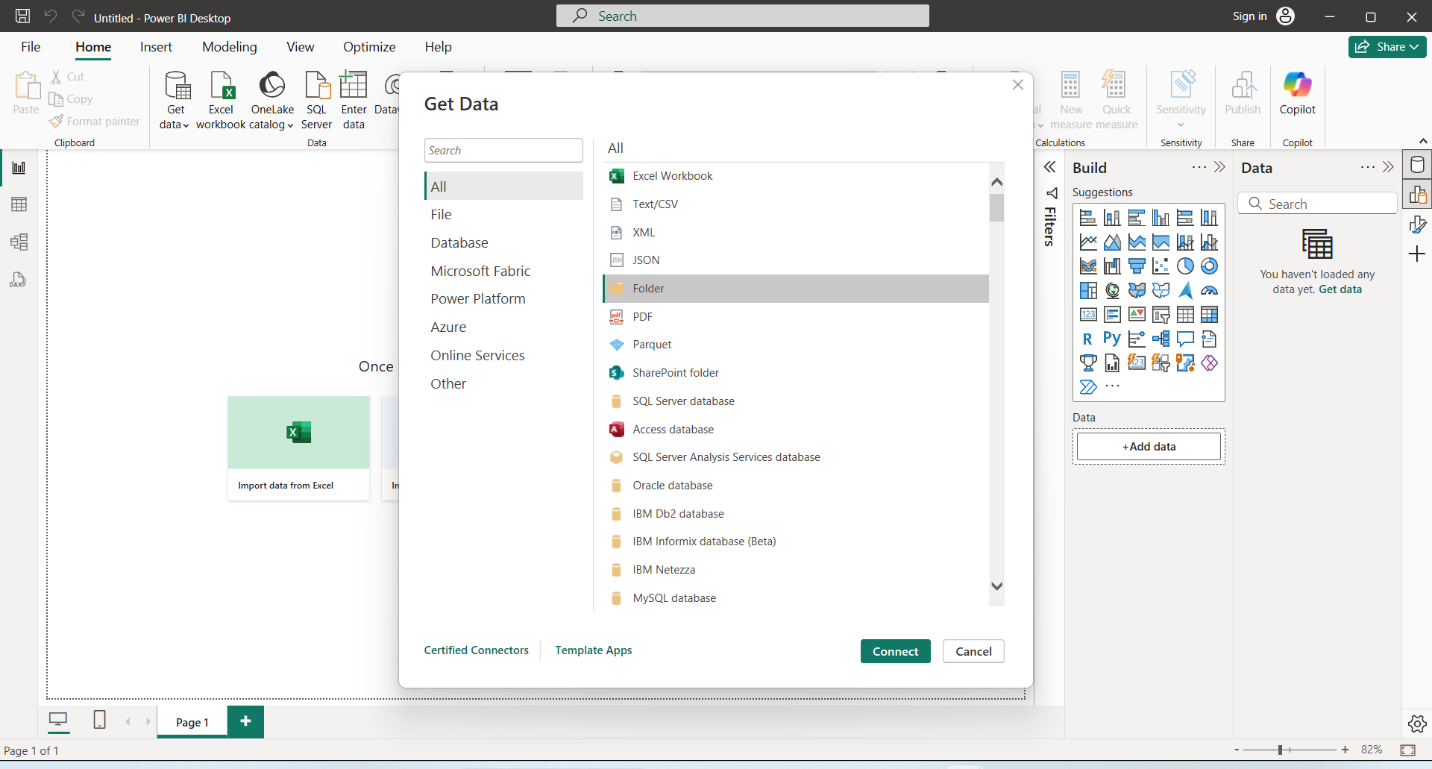
dim\_match\_summary,

dim\_players,

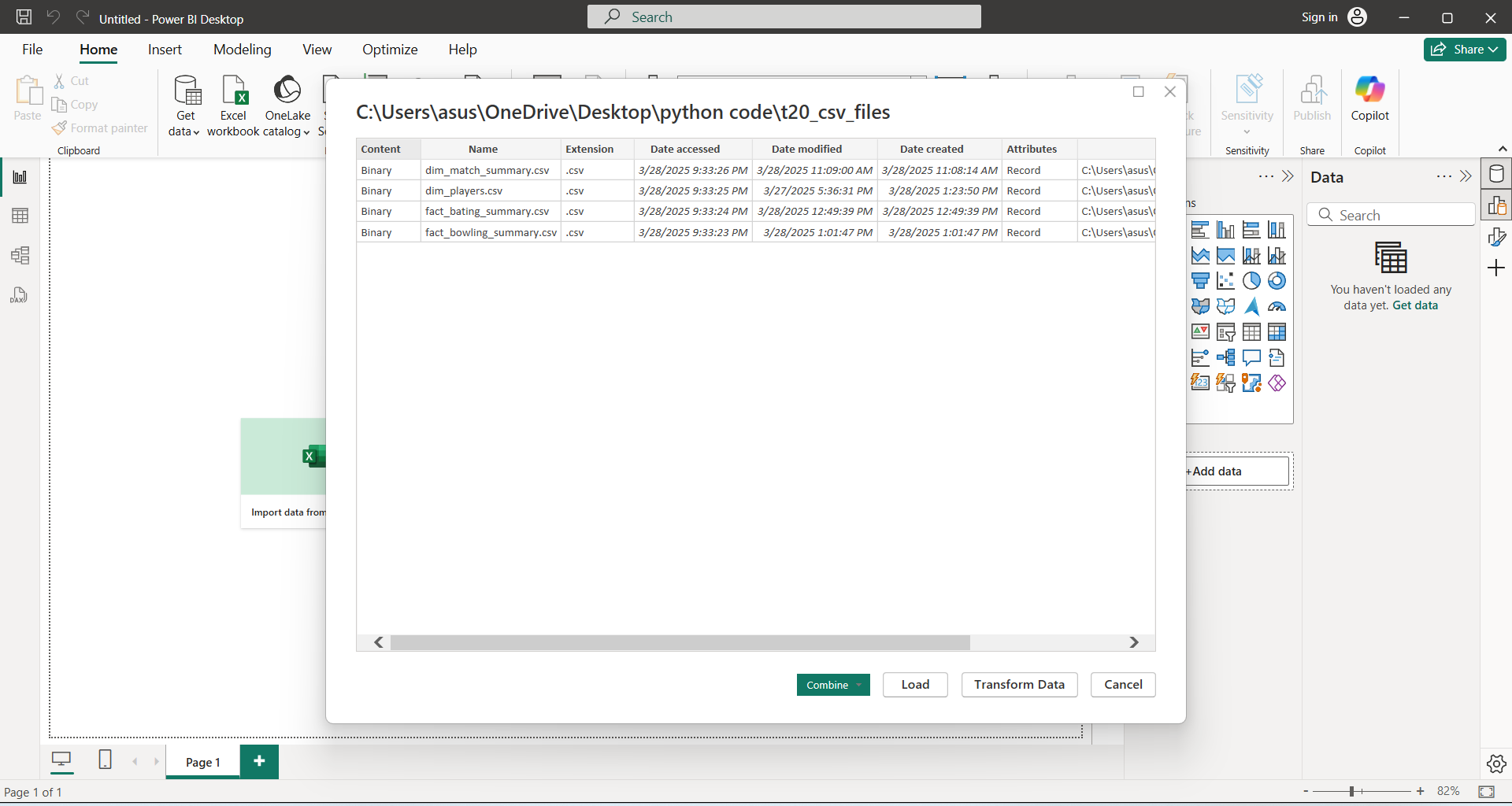
fact\_batting\_summary,

fact\_bowling\_summary

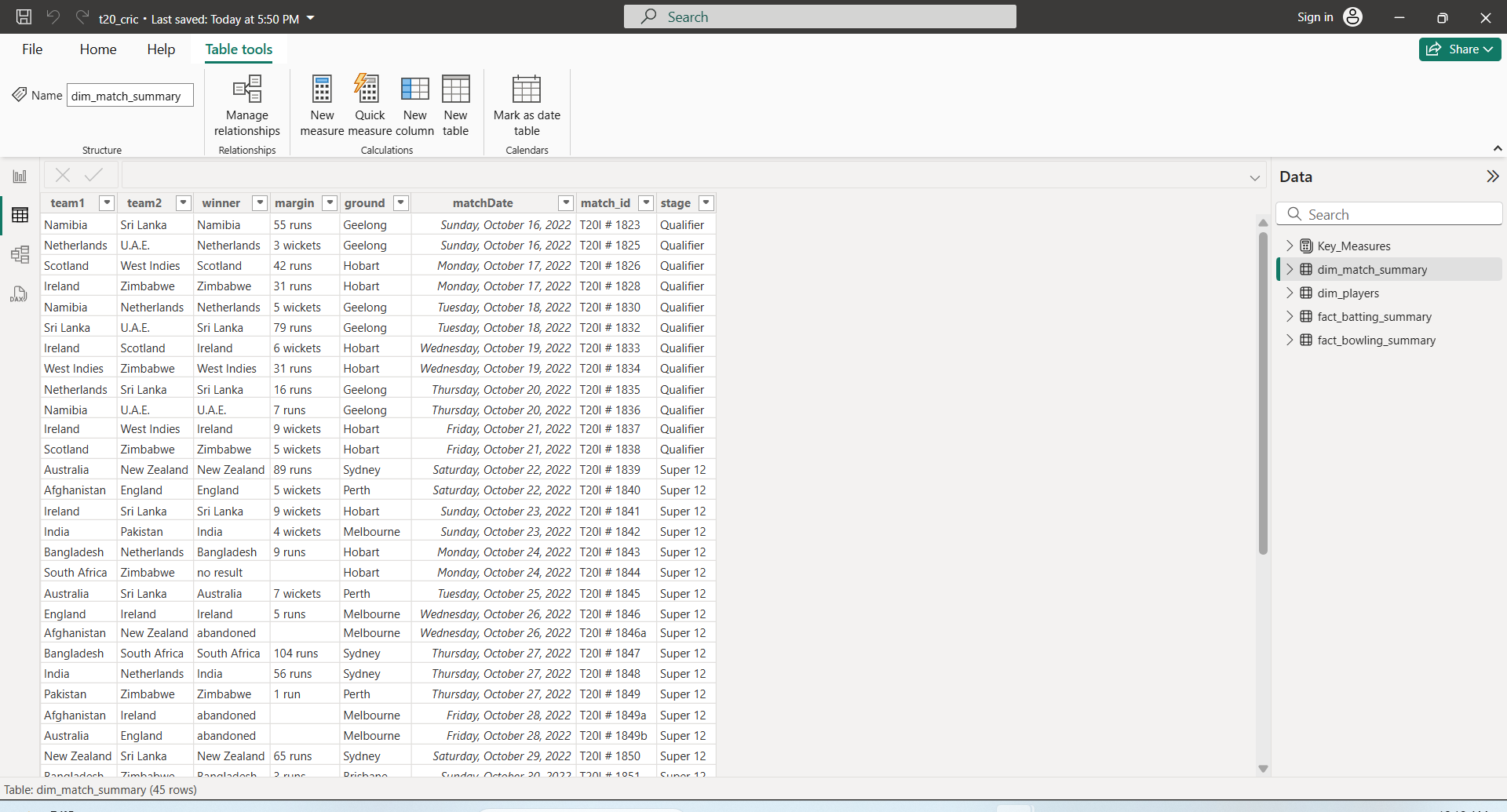
Step 1:- now we have to export a folder that contain the all tables



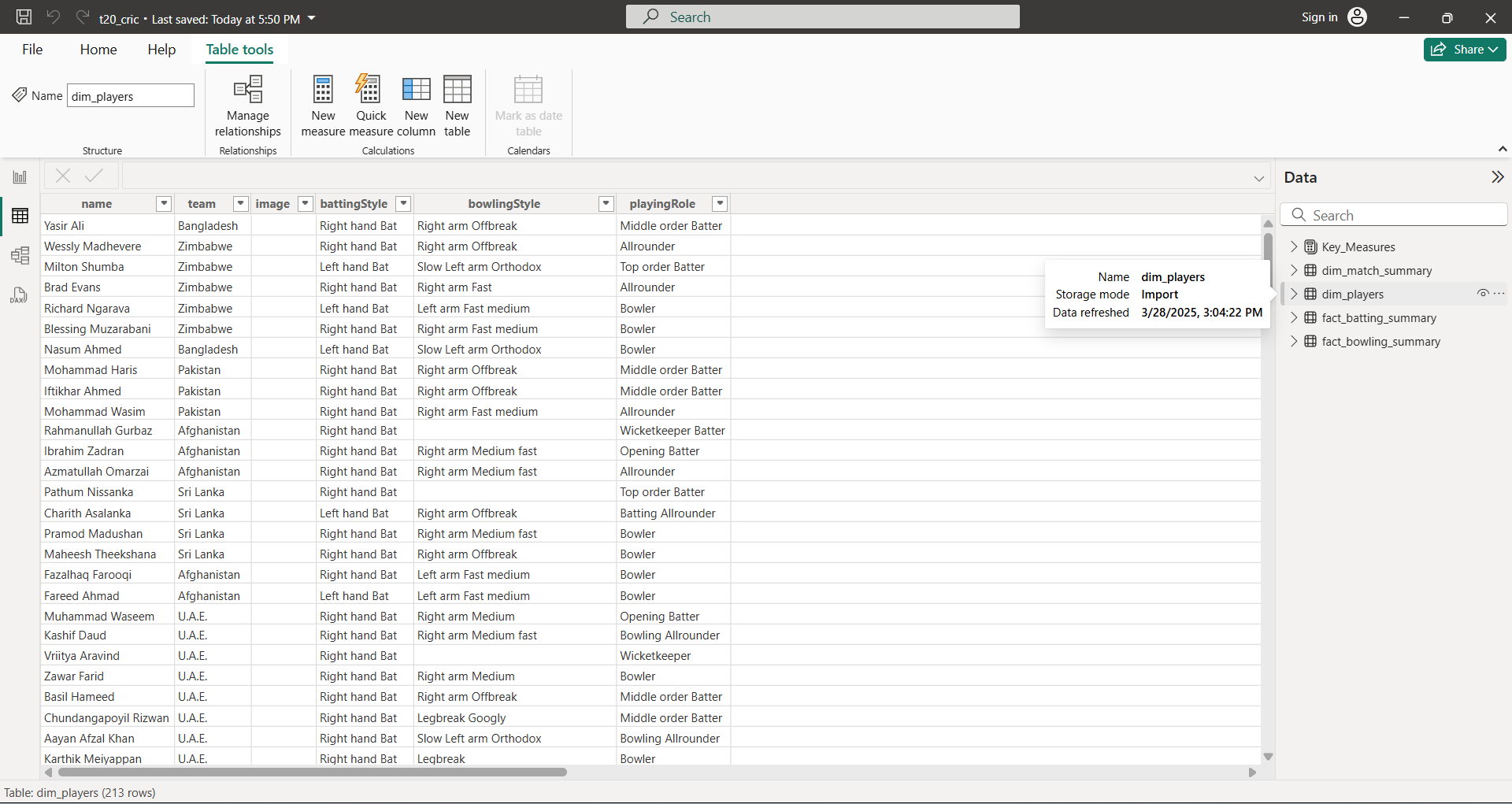
By using the PowerBi, we import the Datasets in project. Click on GET DATA, we can select the folder & click on transform the data & check total datasets are present or not.



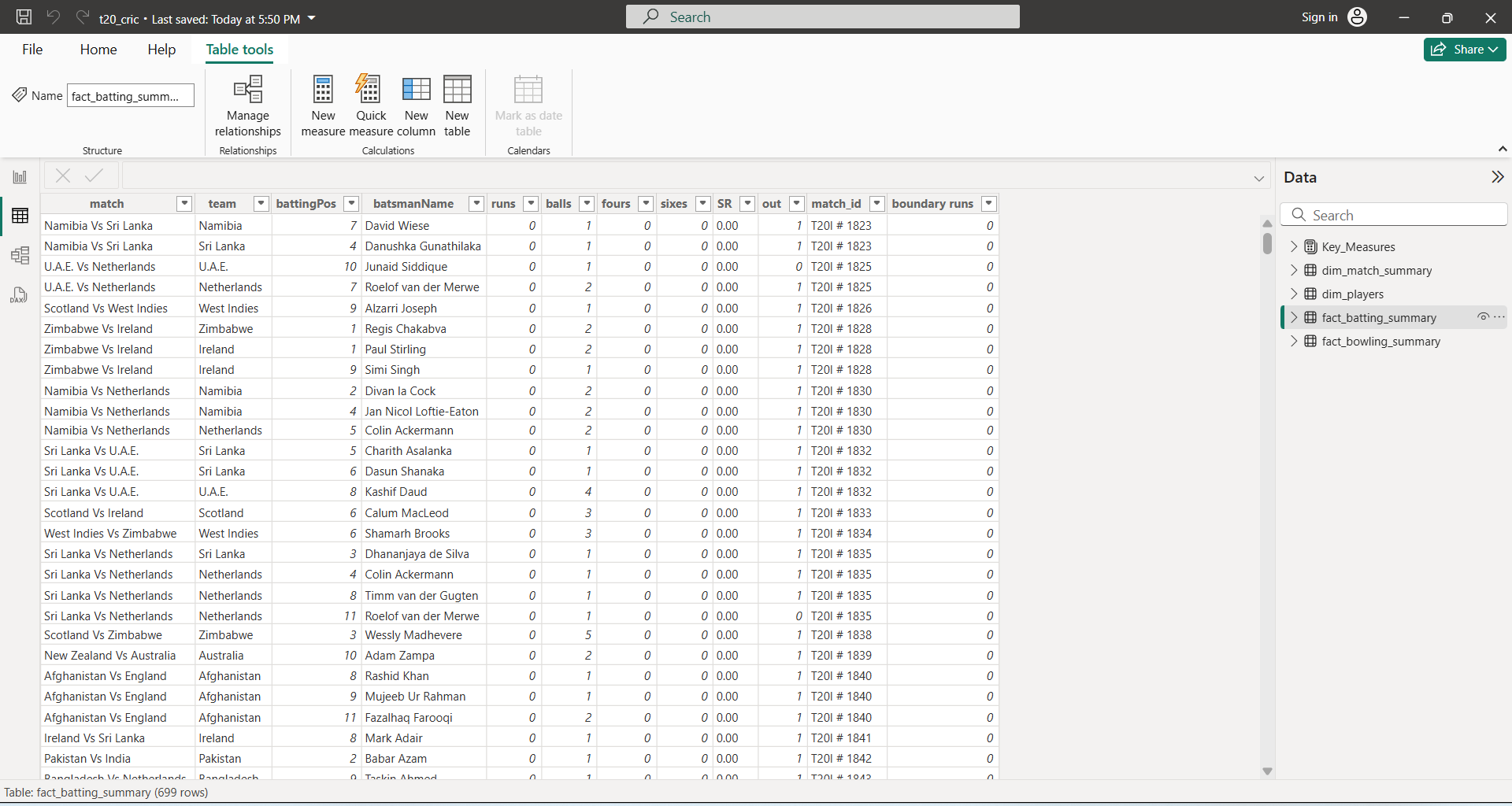
**dim\_match\_summary :-**



**dim\_players:-**



**fact\_batting\_summary:-**



**fact\_bowling\_summary:-**



**7.2 Power Query**

Power Query is an essential tool in Power BI that enables data transformation before analysis. The key steps in Power Query include:

**Data Importing:**

* Bringing in raw data from different sources such as CSV files, APIs, and databases.
* Connecting to live data streams where necessary.

**Data Cleaning:**

* Removing duplicate records to prevent redundancy in analysis.
* Handling missing values using interpolation, mean imputation, or other statistical techniques.
* Standardizing naming conventions for players, teams, and venues.

**Data Transformation:**

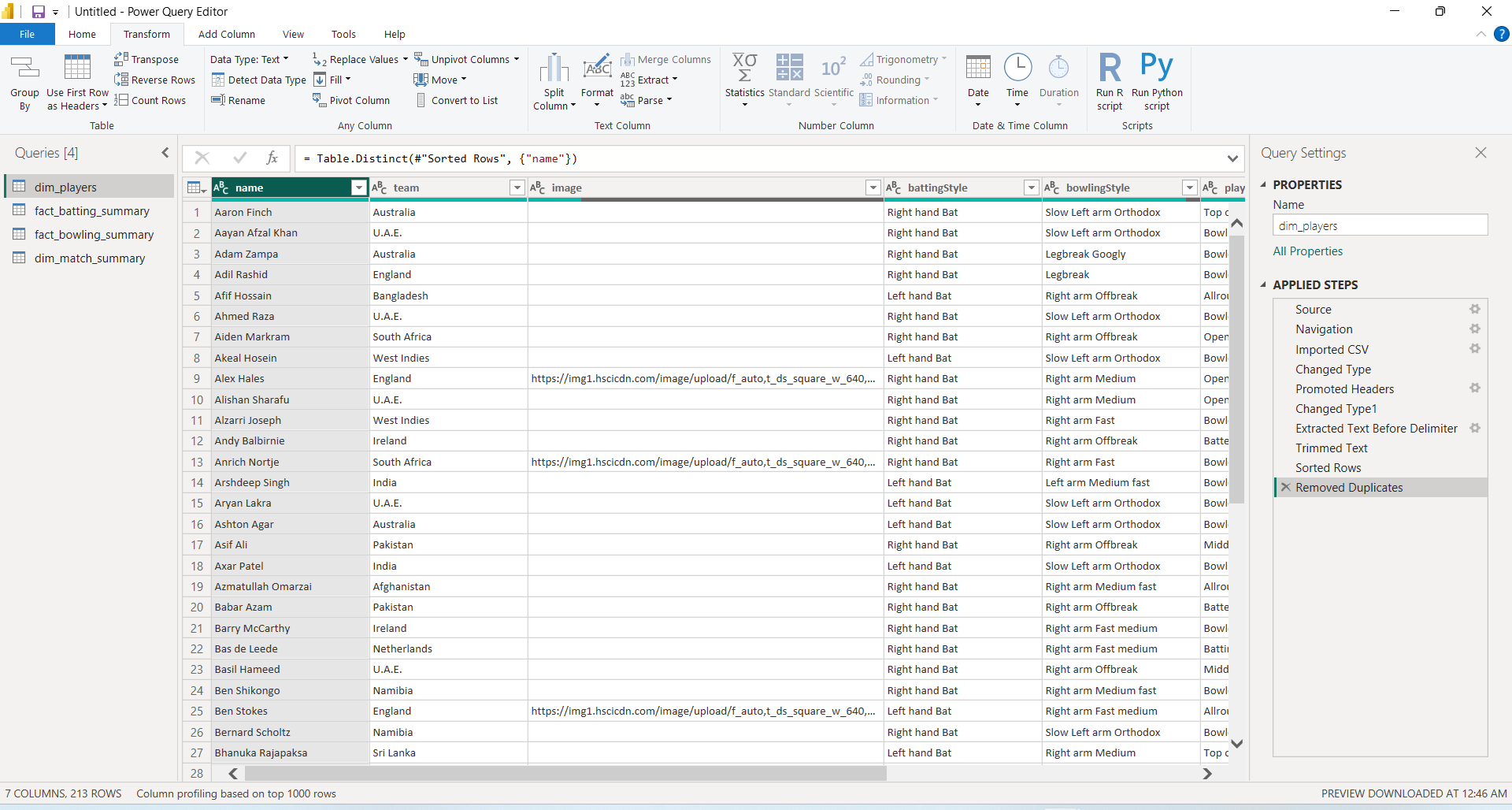
* Merging different datasets to create a single structured dataset.
* Creating calculated columns for metrics such as batting average, economy rate, and strike rate.
* Formatting columns such as date fields for uniformity.

**Data Loading:**

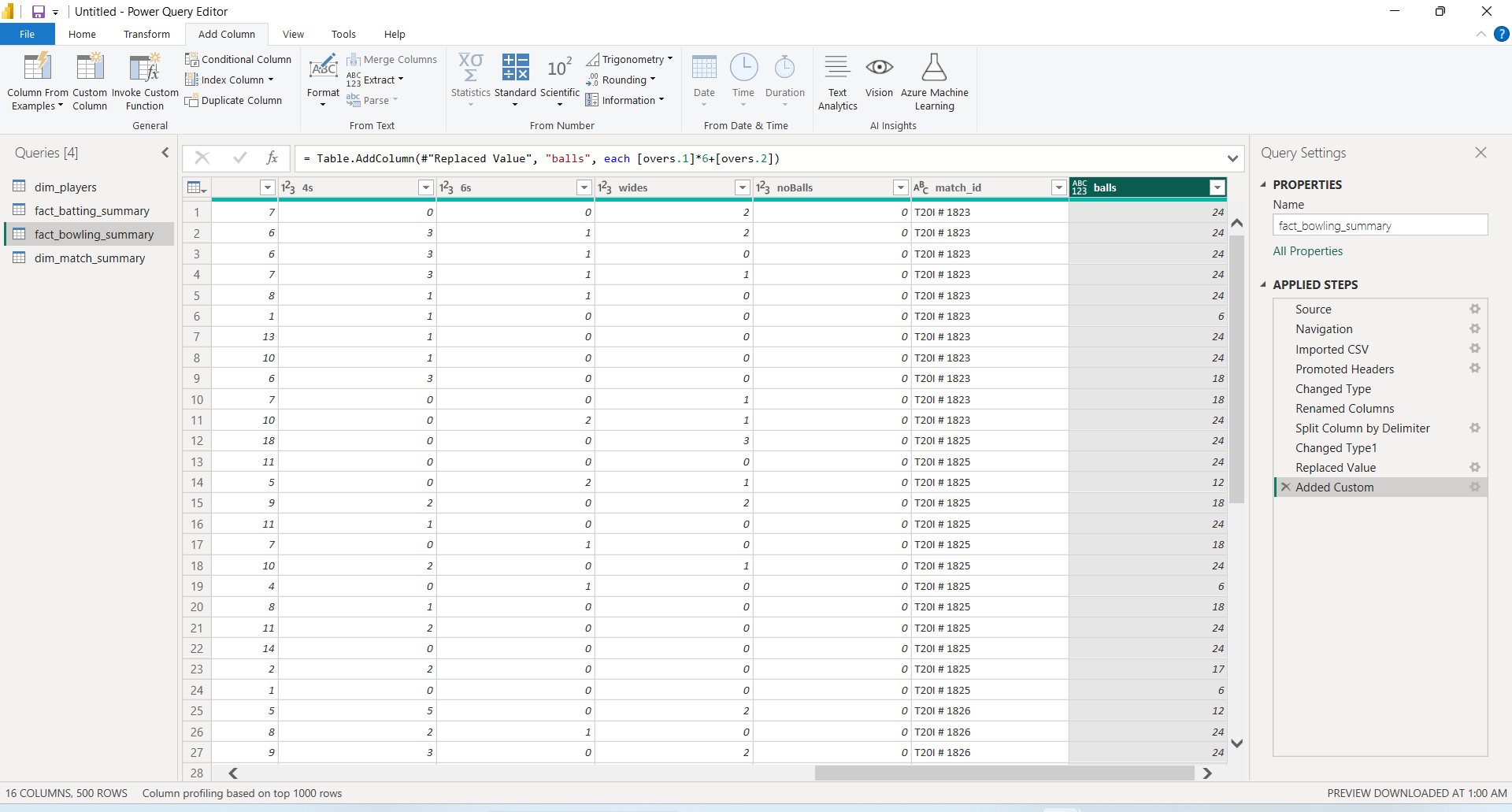
* Once the data is cleaned and transformed, it is loaded into Power BI for further visualization.

Power Query acts as a bridge between raw data and meaningful insights by structuring the data in a format that is easy to analyze

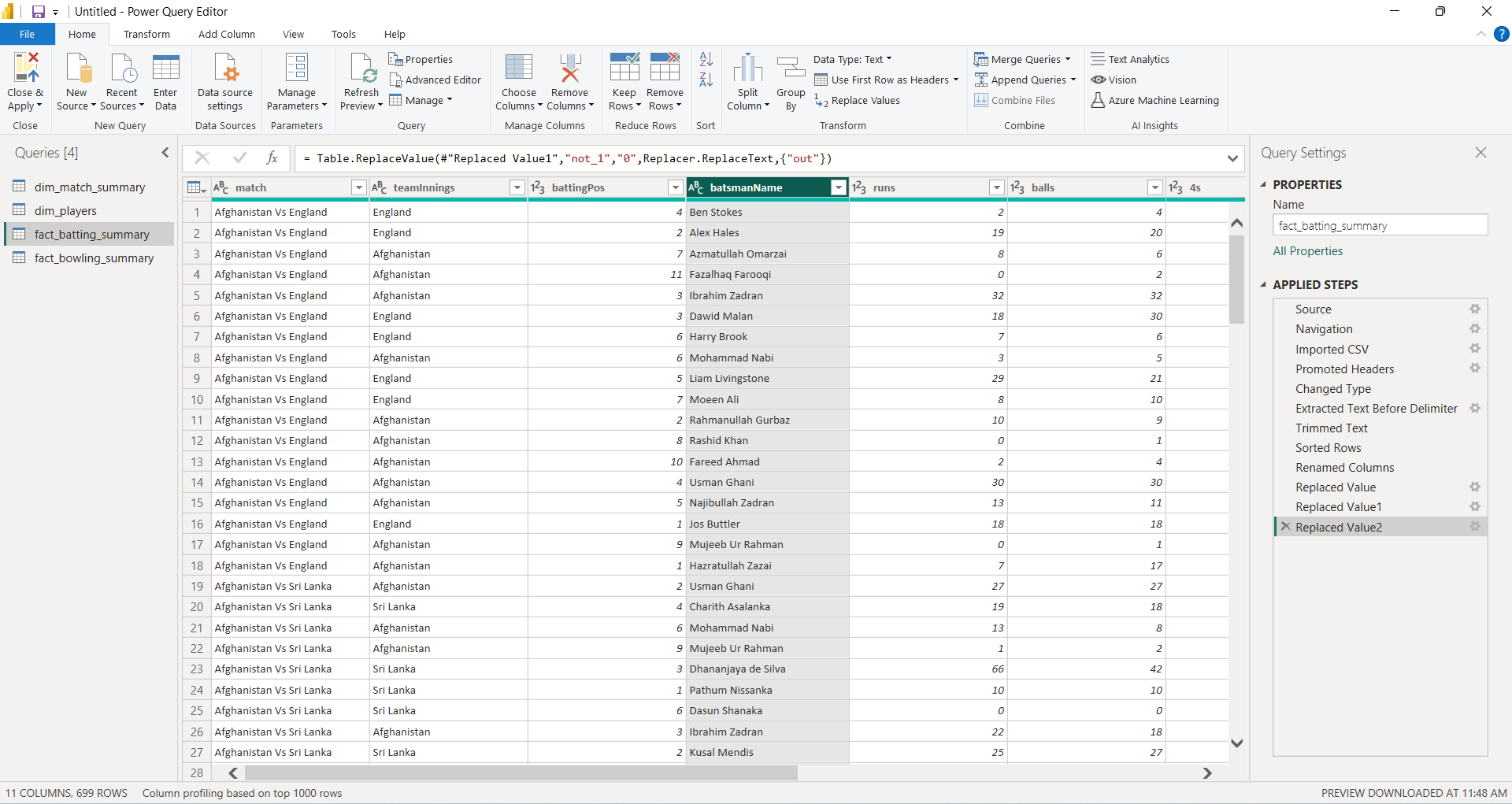
Using Power Bi:-



Here we are creating the new column which shows how much ball a bowling player has done.



Removing the © from player name in fact\_batting\_summary



**8. Data Cleaning Using Python**

Python libraries like Pandas offer powerful tools for handling complex cleaning tasks like identifying and imputing missing values based on patterns or statistical methods.

You can incorporate Python libraries for data cleaning within Cricket Data Analytics project

**1. Libraries Selection:**

**pandas:** pandas is a powerful Python library used for data manipulation and analysis. It provides flexible data structures, mainly DataFrames and Series, which help efficiently handle structured data.

**numpy:** numpy (Numerical Python) is a fundamental Python library used for numerical computations. It provides **fast, memory-efficient** support for large multi-dimensional arrays and matrices, along with a collection of mathematical functions to operate on these structures.

**2. Data Loading and Exploration:**

Use pandas' read\_csv function to import reservation data from CSV files

The official websites of ICC T20 WorldCup 2022.

Explore the data using head(), tail(), info(), and describe() functions to understand data types, identify missing values, and get basic statistics.

**3. Missing Value Handling:**

Identify missing values using isnull() and isna() functions.

Depending on the context and data quality, you can address missing values by:

Dropping rows/columns: Use dropna() if missing values are sparse and unlikely to significantly impact analysis. However, be cautious of data loss.

Filling missing values: Use fillna() to fill missing values with a suitable strategy. This could involve:

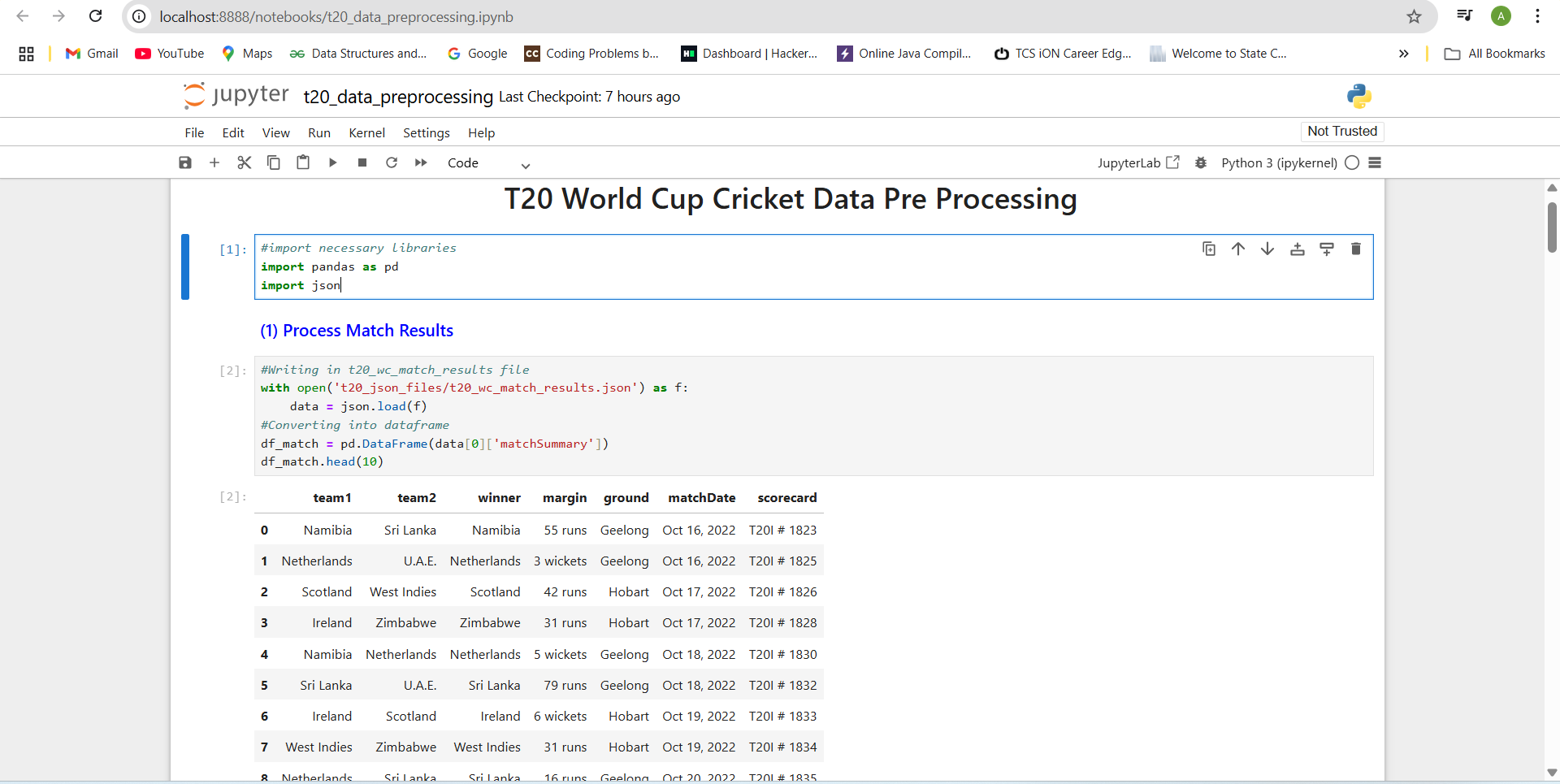
Filling with a constant value (e.g., mean, median) for numerical columns.

Filling with a specific string (e.g., "Unknown") for categorical columns.

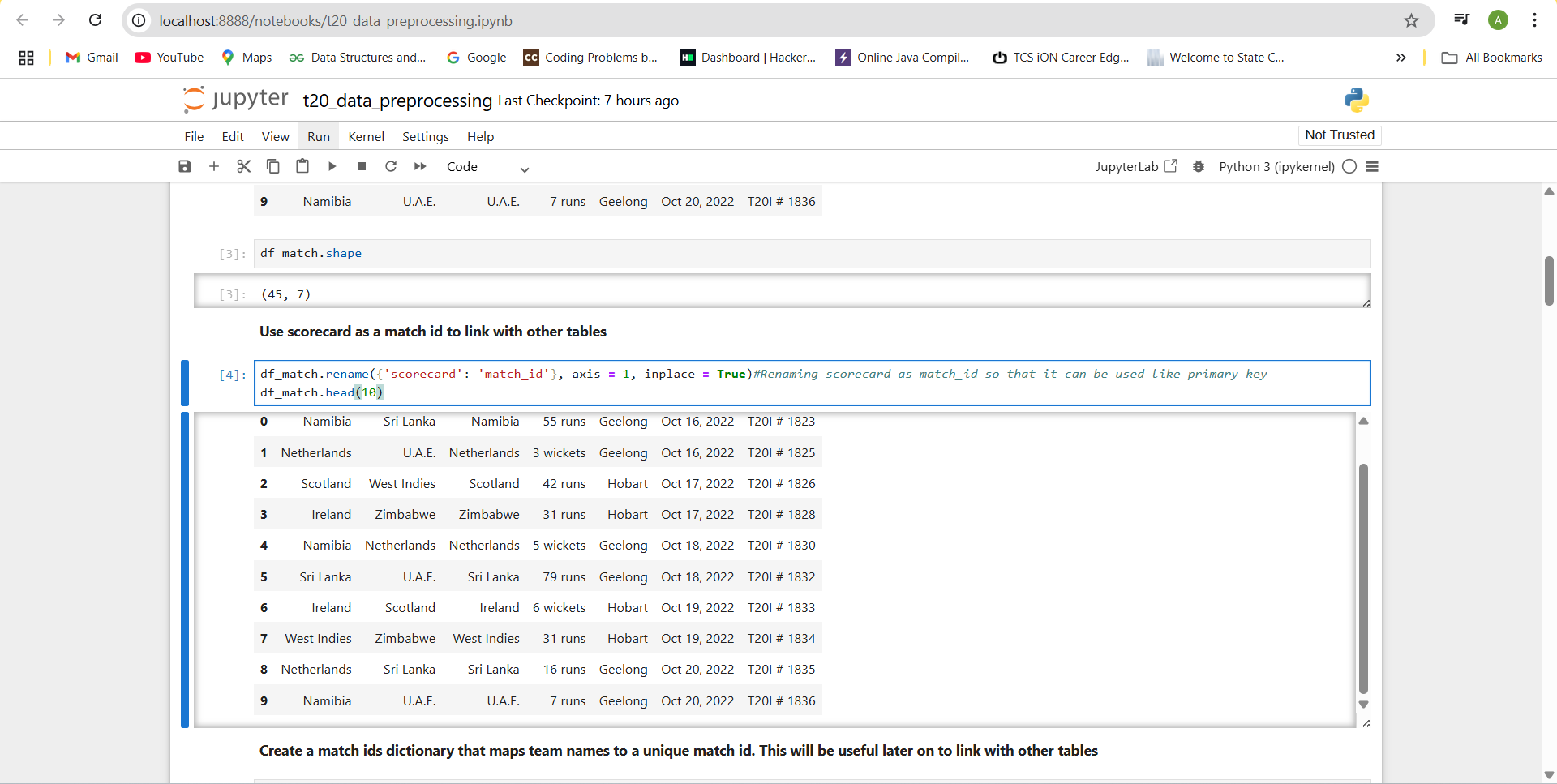
Using interpolation techniques (linear, forward fill) for numerical columns with a time component (e.g., filling missing daily occupancy rates).

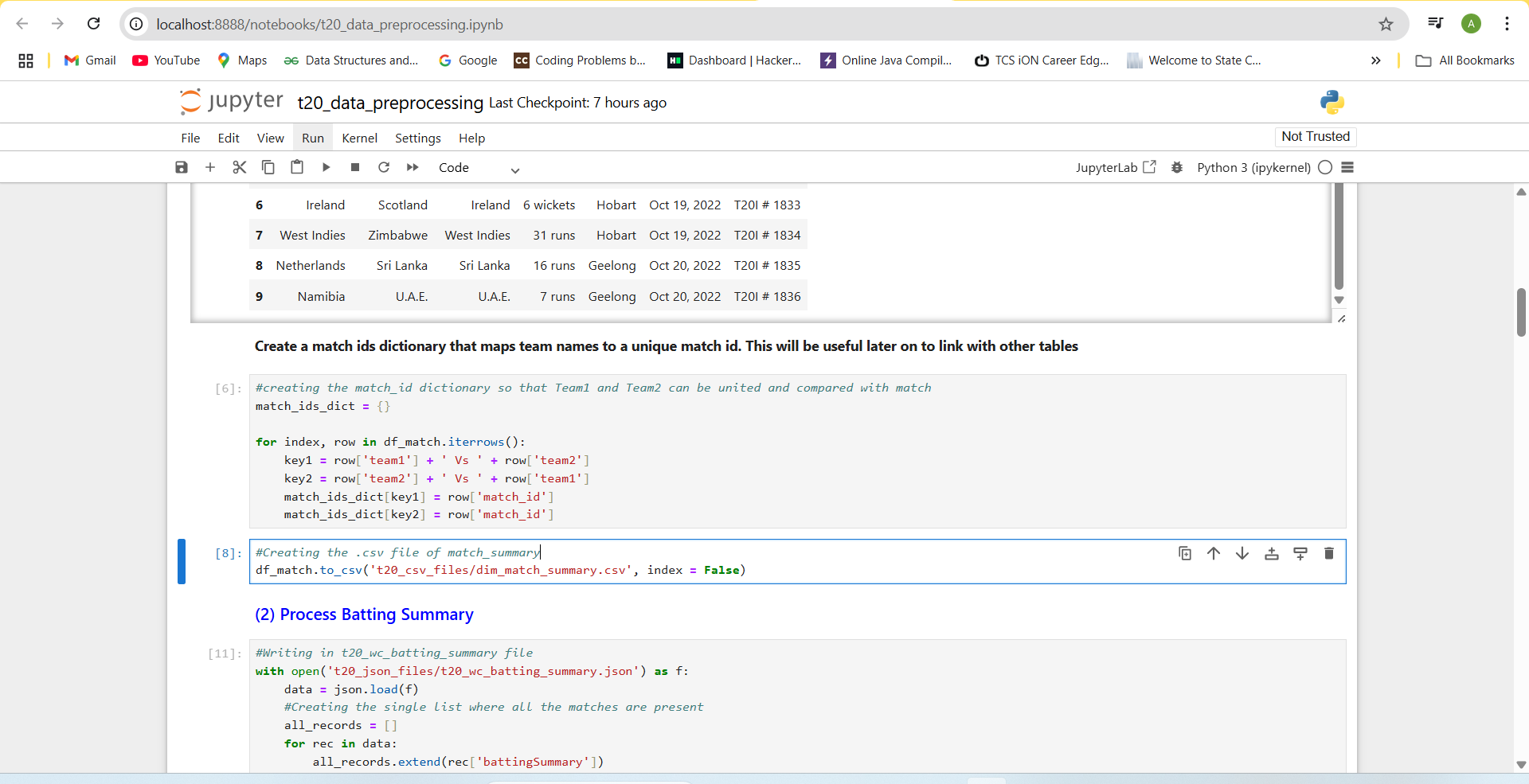
**Output screen for Data Cleaning with explanation in jupyter notebook:**

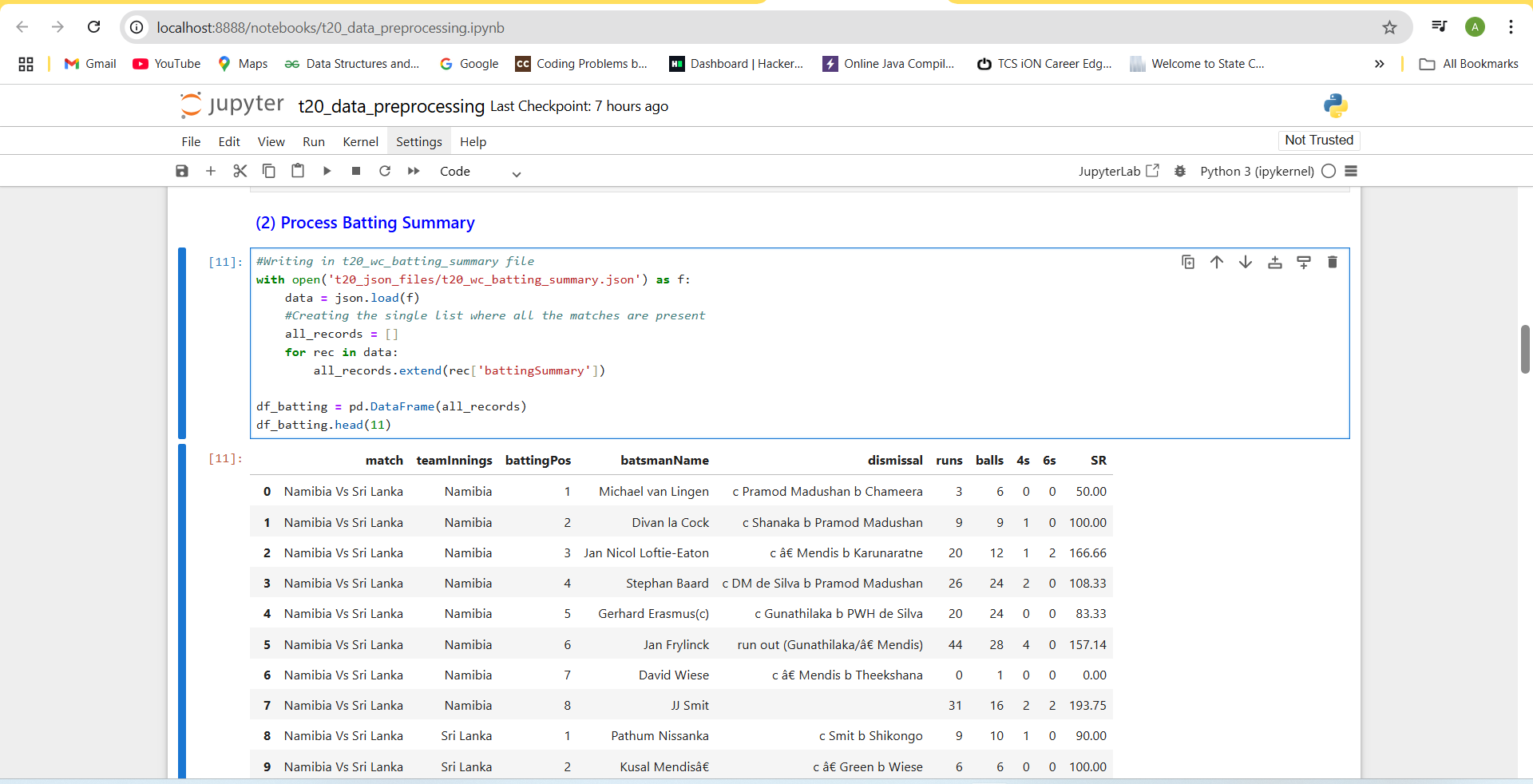
* Importing the necessary libraries in code and Writing in t20\_wc\_match\_results file and converting into dataframe



* Changing the title of score card to match\_id. We are going to use scorecard as match\_id to link with other tables.



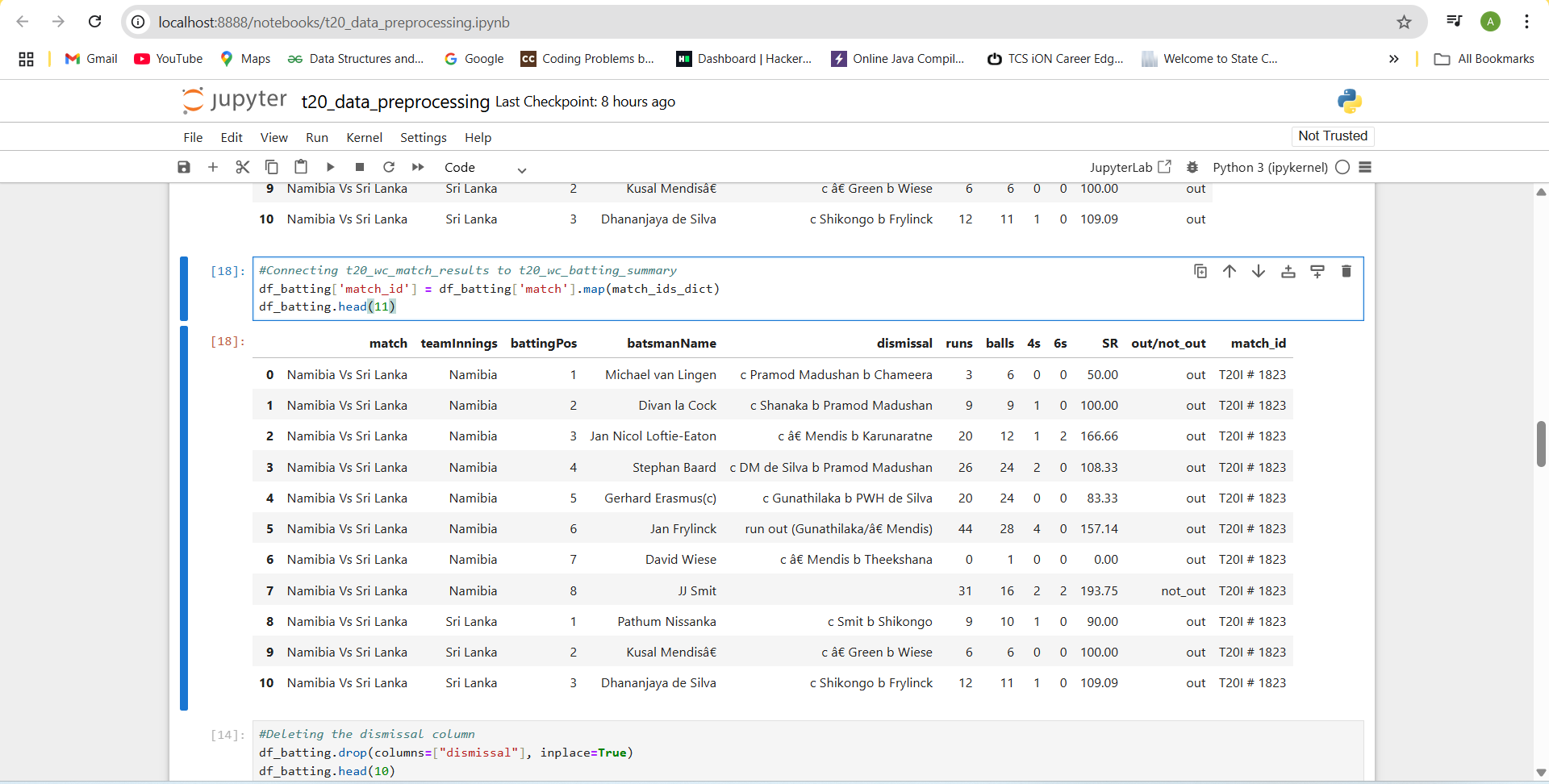
* Create a match ids dictionary that maps team names to a unique match id. This will be useful later on to link with other tables and Creating the .csv file of match\_summary.
* Processing Batting Summary Writing in t20\_wc\_batting\_summary file and converting them into dataframes.



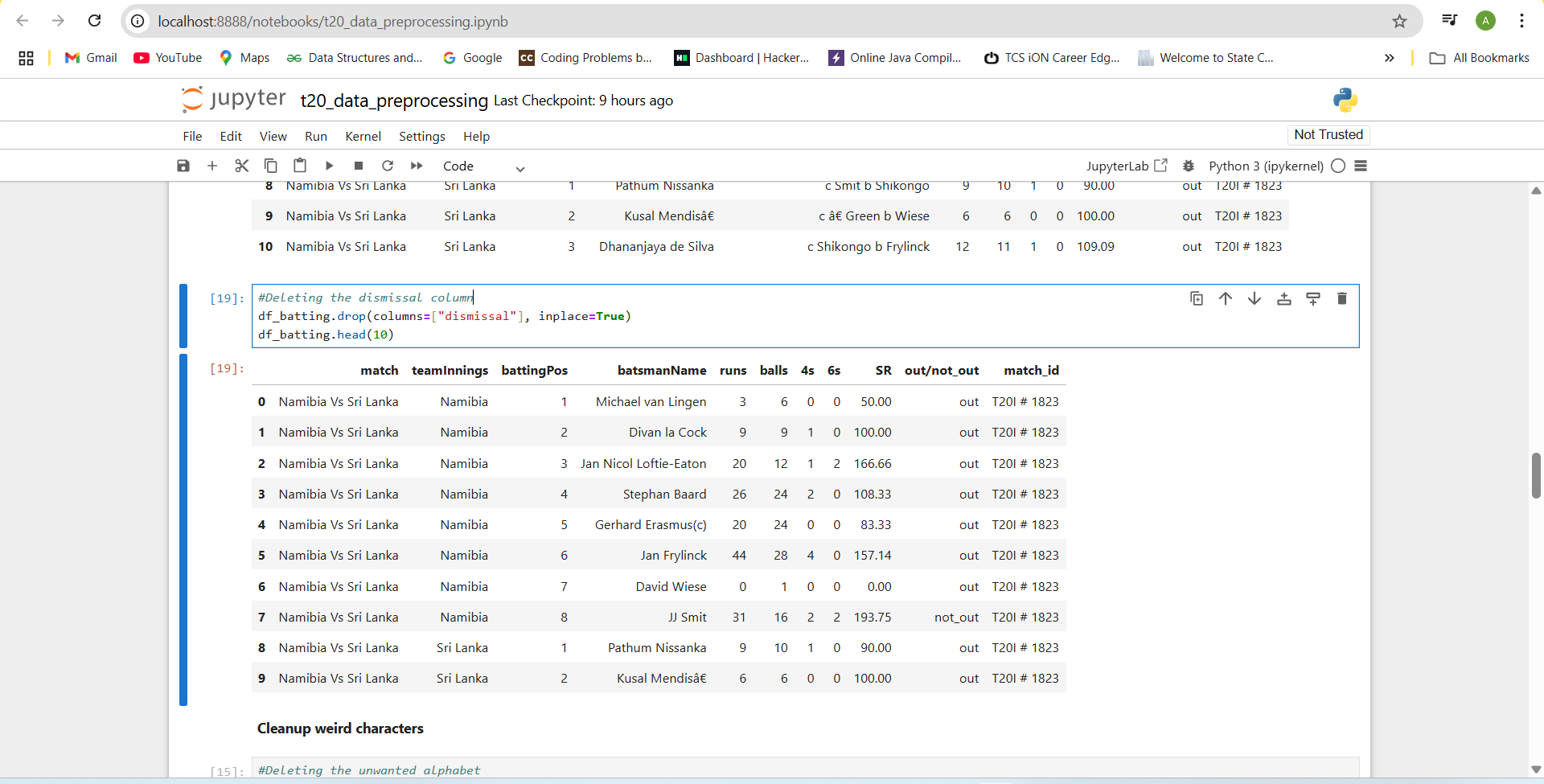
* Creating the out/not out column in the batting summary



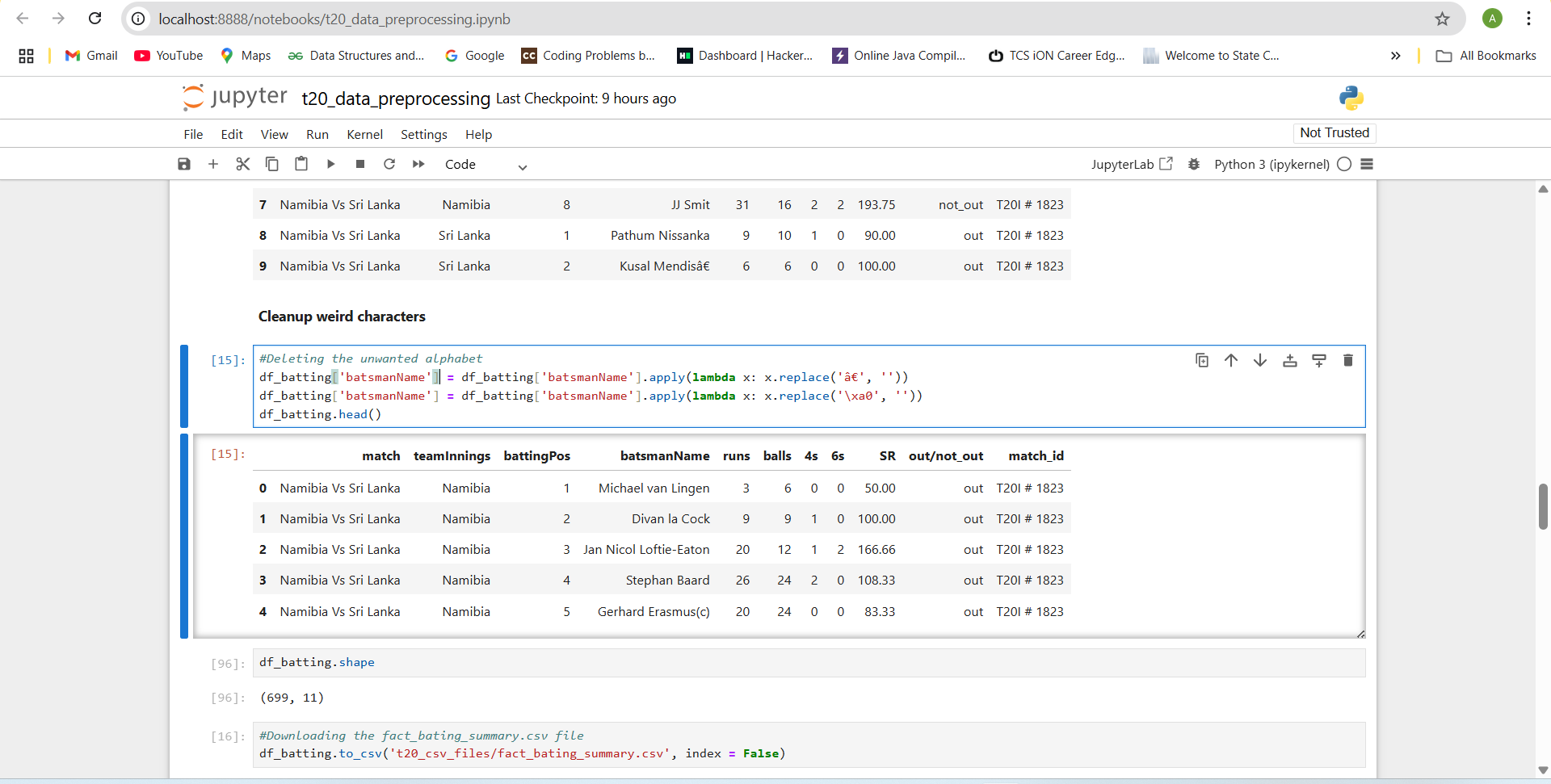
* Connecting t20\_wc\_match\_results to t20\_wc\_batting\_summary



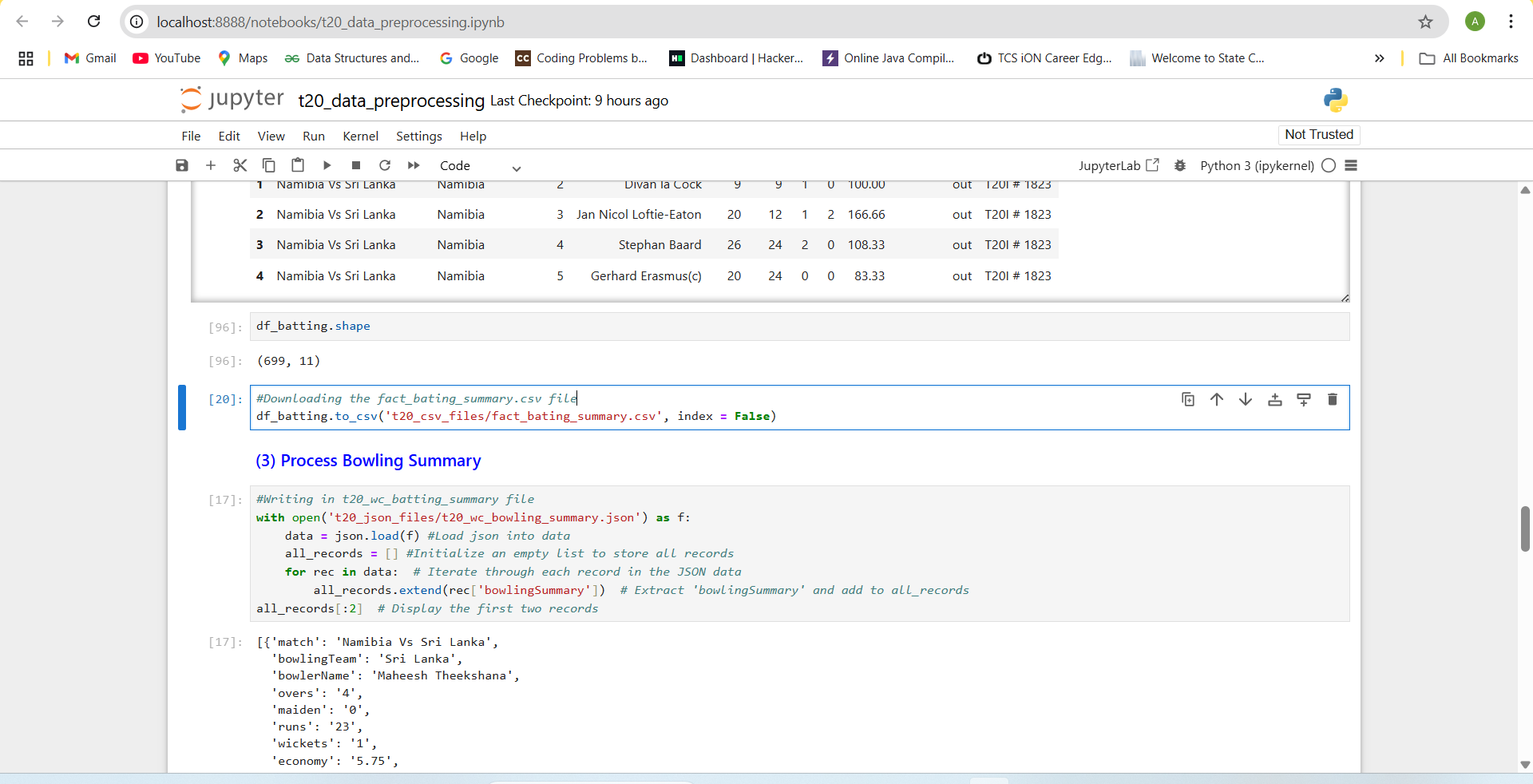
* Deleting unwanted columns from the table.



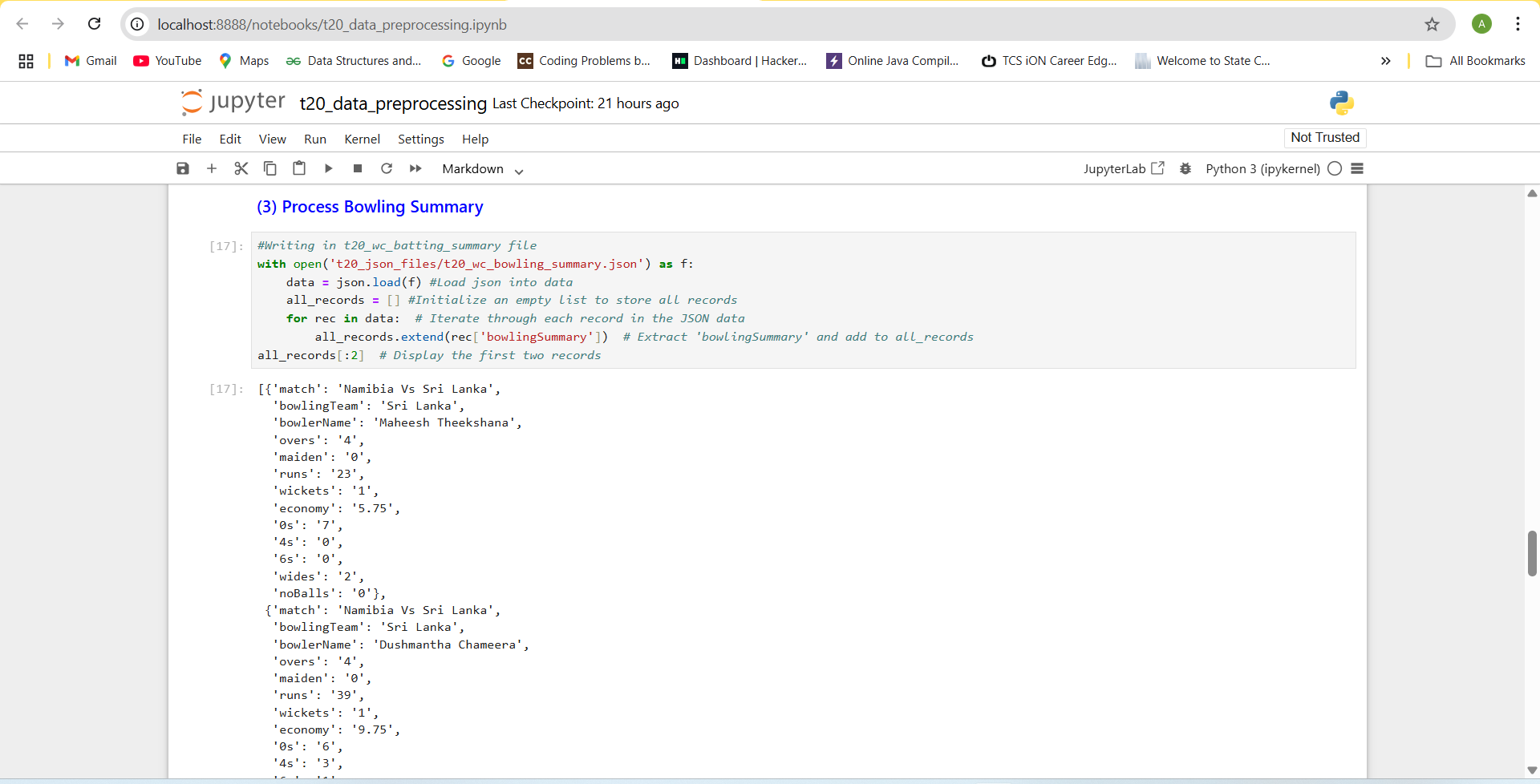
* Cleaning the weird character.



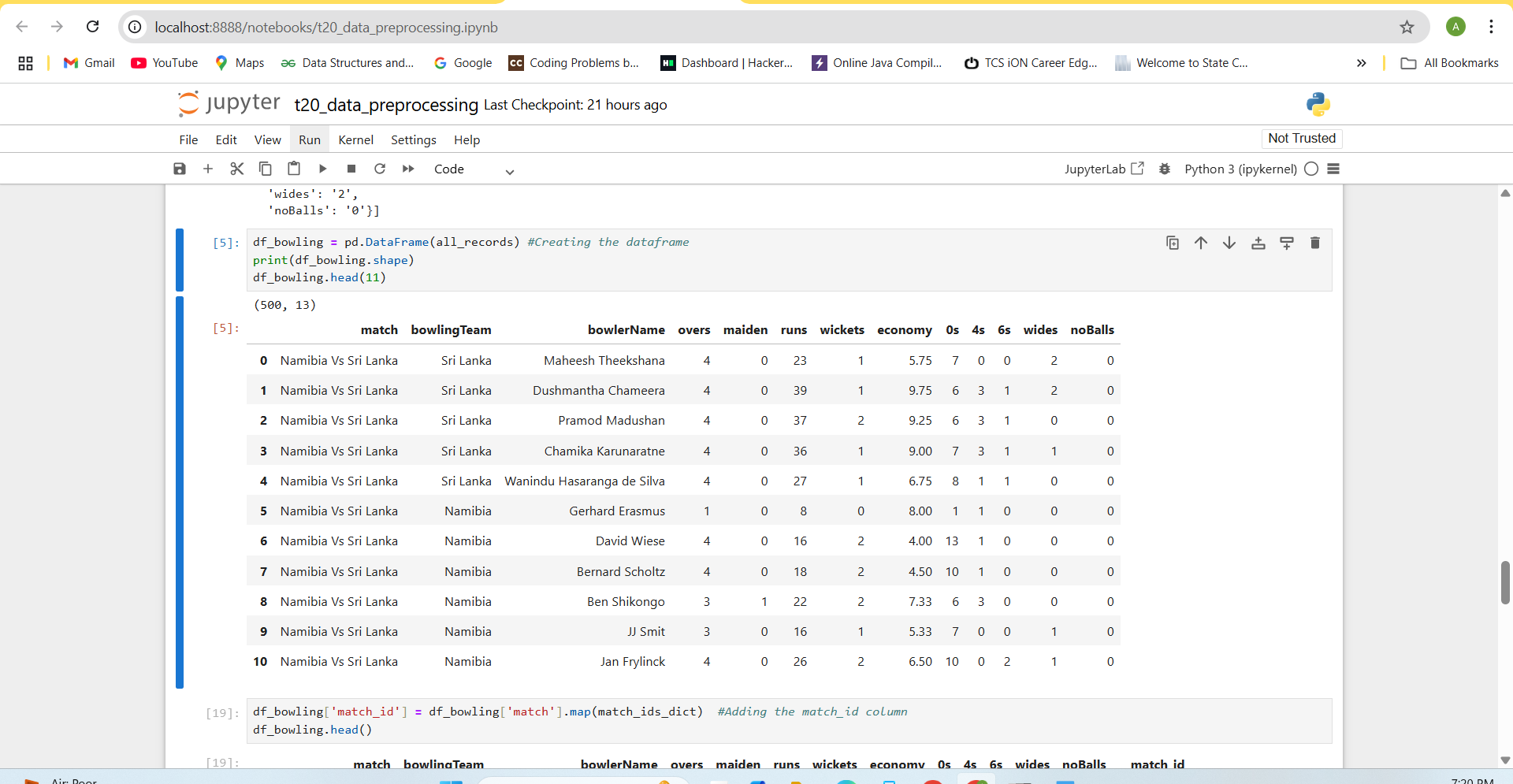
* Downloading the cleaned format of the batting summary.



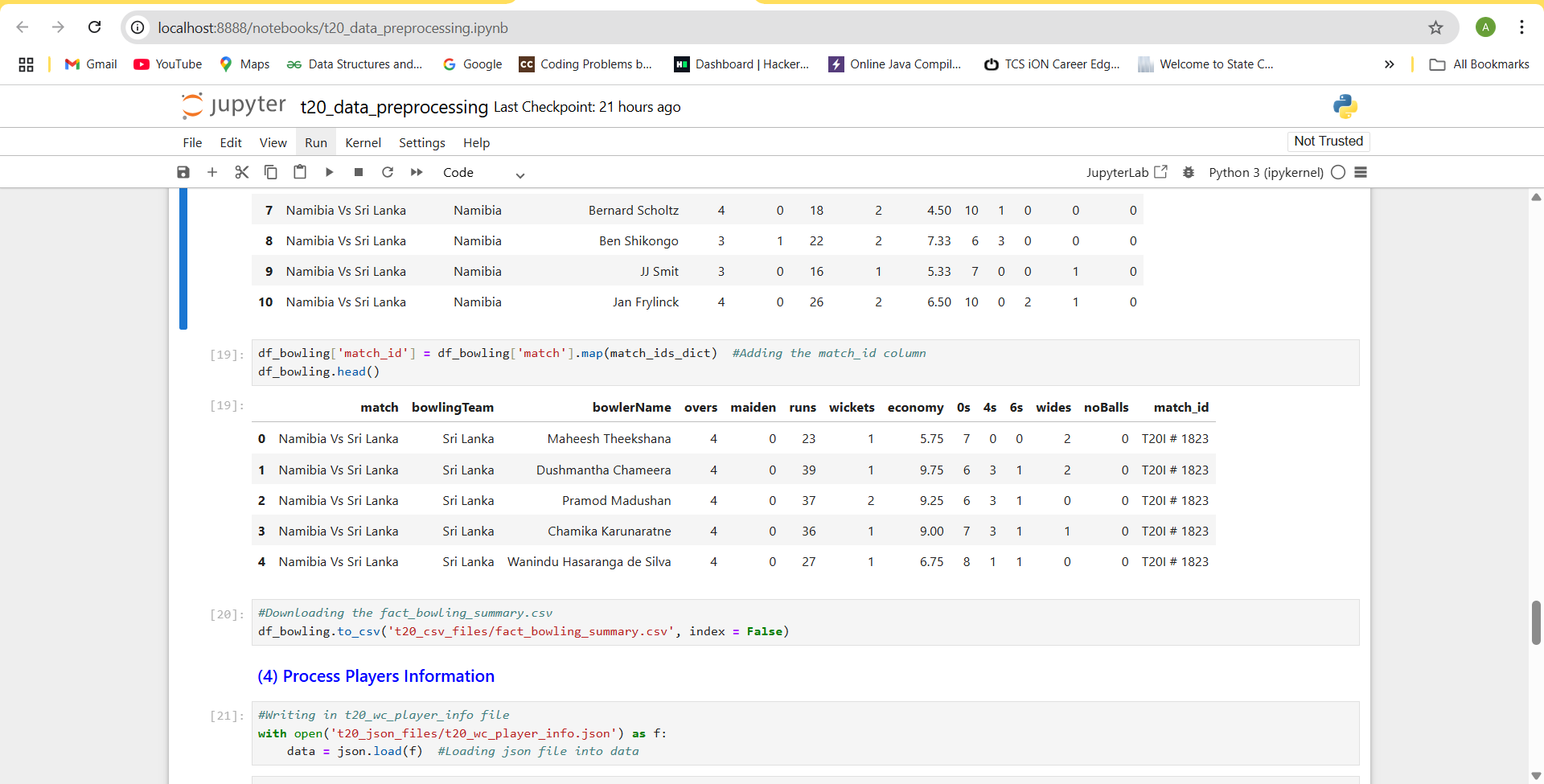
* Writing in the Bowling Summary and printing the first two records.



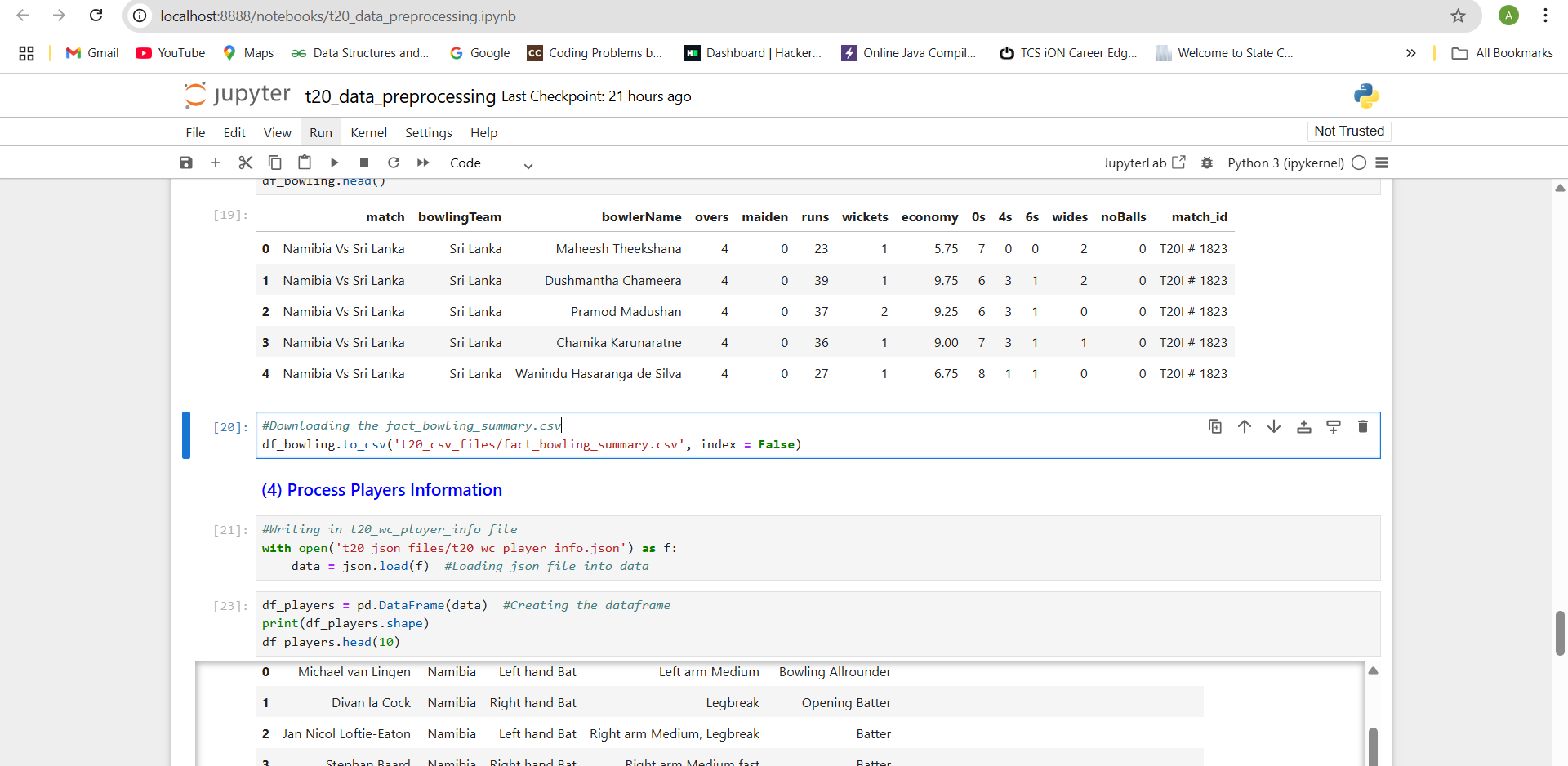
* Creating the Dataframe and printing the first 11 rows.



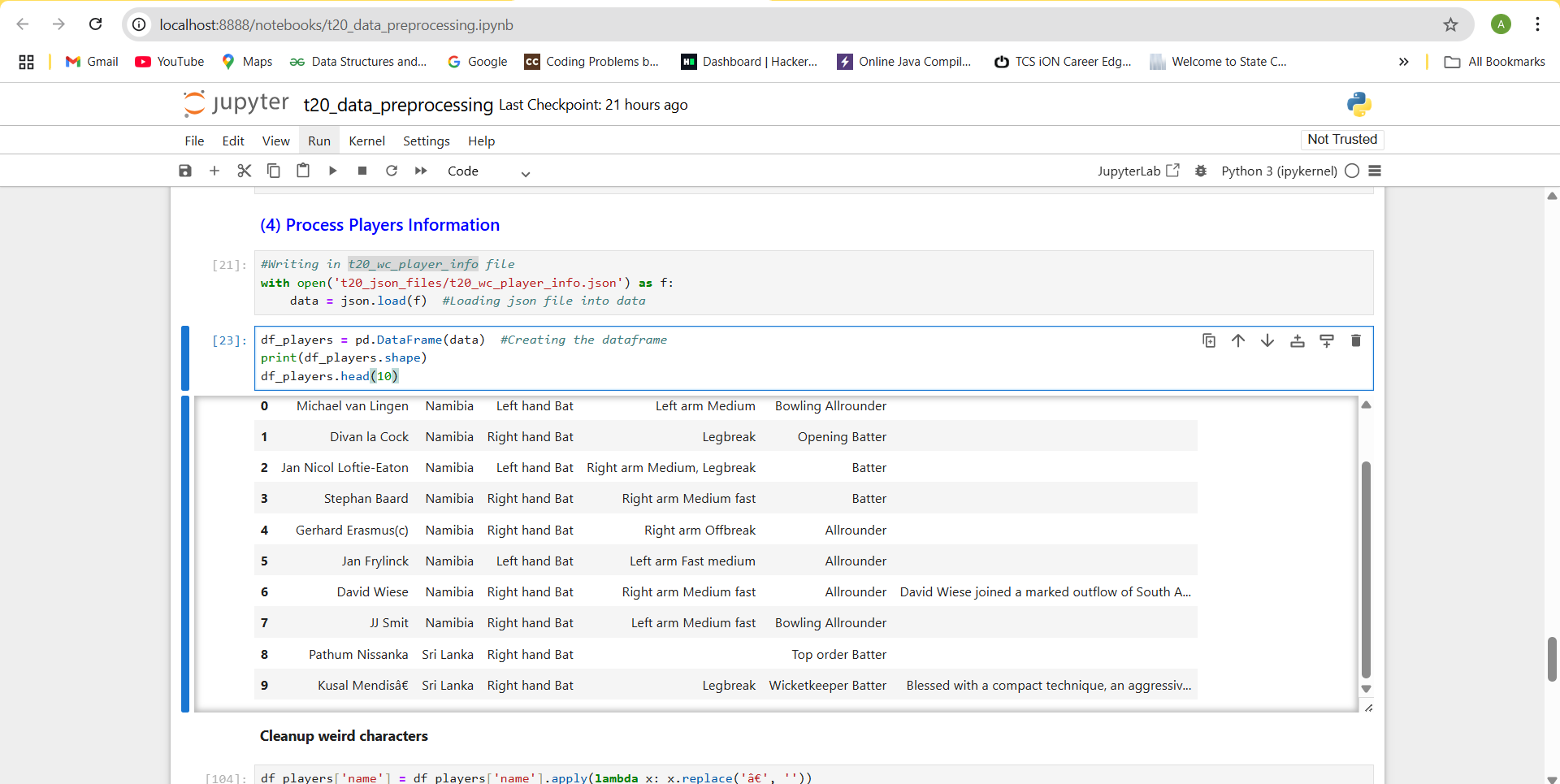
* Adding the match\_id column to the Bowling Summary so we can link with the another column.



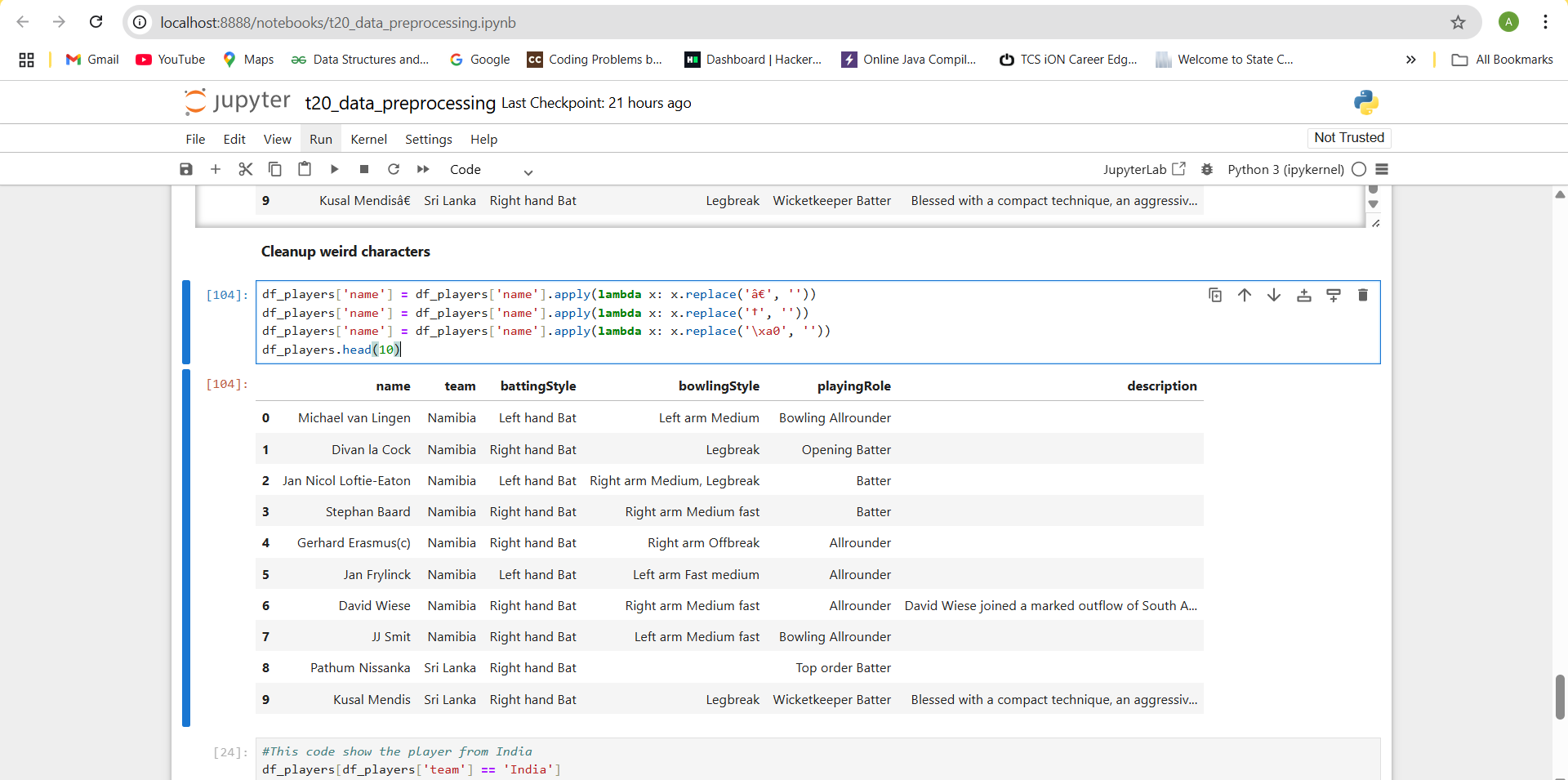
* Downloading the edited Bowling Summary in .csv format.



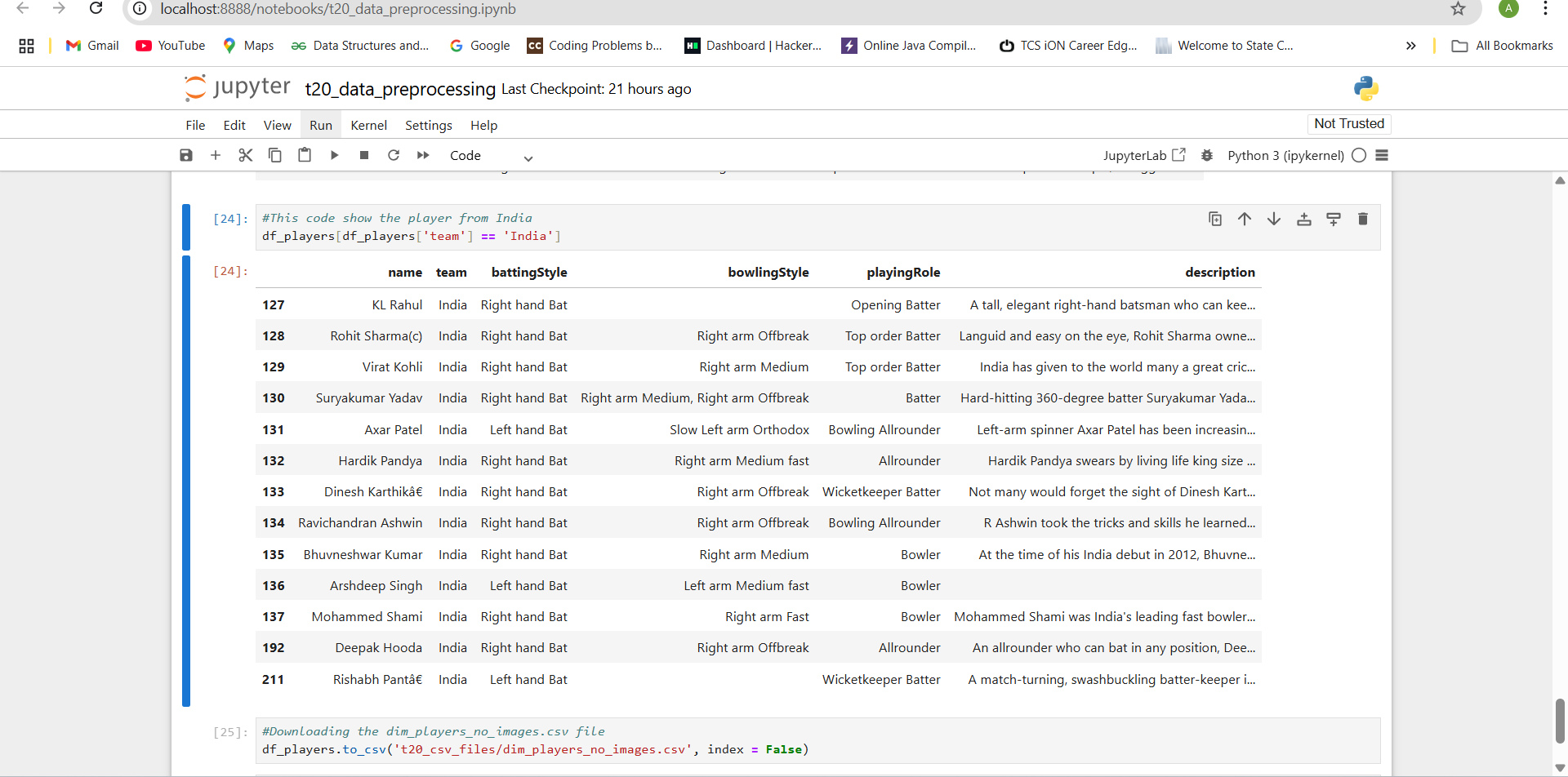
* Loading the t20\_wc\_player\_info file and creating the Dataframe from it.



* Cleaning up wired characters like â€\xa0. From the table.



* This code show the players from India.



* Downloading the dim\_players\_no\_images in .csv format which makes us easy to use it in the Power Bi.



**9. Building Matrix using Dax:-**

Cricket performance analytics relies on a robust set of metrics that capture player performance across batting, bowling, and fielding disciplines. These metrics are implemented using DAX (Data Analysis Expressions) in Power BI to build dynamic dashboards and insights for coaches, analysts, and fans. The following section explains these metrics in a detailed paragraph format, including the logic behind them and the DAX formulas used for calculation.

**Batting Measures:-**

* **Create a Total Runs Measure:-** This measure will tell us which player has scored how much runs in total.

Total Runs = SUM(fact\_batting\_summary[runs])

* **Total Innings Batted:-** Total number of innings a batsman got a chance to bat.

Total Innings Batted = COUNT(fact\_batting\_summary[match\_id])

* **Total Innings Dismissed:-** Indicates how many times the player got out.

SUM(fact\_batting\_summary[out]).

* **Batting Average**:- which shows the average number of runs per dismissal.  
  Batting Avg = DIVIDE([Total Runs], [Total Innings Dismissed])
* **Balls Faced** metric is calculated as:  
  Balls Faced = SUM(fact\_batting\_summary[balls\_faced]),
* **Strike Rate**:  
  Strike Rate = DIVIDE([Total Runs] \* 100, [Balls Faced]).
* **Batting Position**:- Batting position of a player

Batting Position = ROUNDUP(AVERAGE(fact\_batting\_summary[batting\_pos]),0)

* **Boundary Percentage** is derived as:  
  Boundary % = DIVIDE([Boundary Runs], [Total Runs]) \* 100,  
  indicating how much of a batsman’s score comes from boundaries.
* **Avg Balls Faced:-** This tells us the avg balls faced by the batter in an inning.

AVERAGE(fact\_batting\_summary[balls])

**Bowling Measures:-**

Bowling metrics help quantify a bowler’s effectiveness, consistency, and impact.

* **Total wickets taken:-** The Total Wickets measure sums the number of wickets taken:  
  Total Wickets = SUM(fact\_bowling\_summary[wickets]).
* **Total Overs Bowled:-** The Overs Bowled metric aggregates the total number of overs:  
  Overs Bowled = SUM(fact\_bowling\_summary[overs]).
* **Runs Conceded:-** Runs Conceded tracks the total runs given away by the bowler:  
  Runs Conceded = SUM(fact\_bowling\_summary[runs\_conceded]).
* **Bowling Economy:-** Average number of runs conceded in an over. Economy = DIVIDE( [Runs Conceded], ([balls Bowled]/6),0)
* **Bowling Strike Rate:**- Bowling Strike Rate Number of balls bowled per wicket.

Bowling Strike Rate = DIVIDE([balls Bowled], [wickets],0)

* **Bowling Average** is a key indicator of performance, calculated as:  
  Bowling Average = DIVIDE([Runs Conceded], [Total Wickets]),  
  which represents average runs conceded per wicket.
* **Total Innings Bowled:-** Total number of innings bowled by a bowler.

Total Innings Bowled =

DISTINCTCOUNT(fact\_bowling\_summary[match\_id])

* **Dot Ball Percentage:-** Percentage of dot balls bowled by a bowler.

Dot ball % = DIVIDE(SUM(fact\_bowling\_summary[zeros]), SUM(fact\_bowling\_summary[balls]),0)

* **Player Selection:-** To understand if a player is selected or not.

Player Selection = if(ISFILTERED(dim\_player[name]),"1","0")

* **Display Text:-** To display the text of a number player is selected.

Display Text = if([Player Selection] = "1", " " ,"Select Player(s) by clicking the player’s name to see their individual or combined strength.")

* **Color Callout Value:-** To Display the value only when a player is selected.

Color Callout Value = if([Player Selection]="0", "#D0CF1D","#1D1D2E")

**Calculated Column Name:**

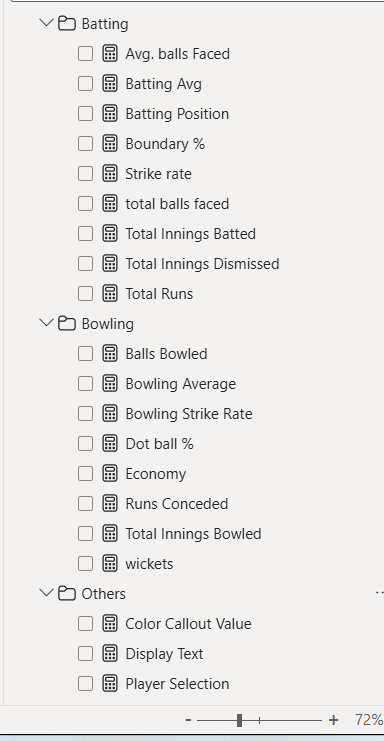
* **Boundary Runs:-**To fix the total number of runs scored by hitting fours and sixes.

Boundary runs = fact\_batting\_summary[fours]\*4 + fact\_batting\_summary[sixes]\*6

* **Boundary Runs Bowling:-** To find total number of runs conceded by bowlers in boundaries.

Boundary runs = fact\_bowling\_summary[fours]\*4 +fact\_bowling\_summary[Sixes]\*6

Here’s a list of measure which we are going to use in this.



**10. Visualization Power Bi**

Using the Power Bi we can create a Dashboard using Power Bi Tool:-

Here’s the data given. It contain data of ICC World Cup T20 2022.

We can create this simple & Nice Clean Dashboard for understanding the records of players.

One best news is that this is changeable according the filter, week, months so easily we can understand the data.

We have created 6 pages each representing different purpose.

Namely:-

1)Power Hitters/Openers

2)Anchors/Middle Order

3)Finisher/Lower Order Anchor

4)All Rounders/ Lower Middle Anchors

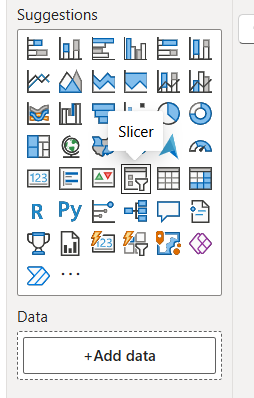
5) Specialist Fast Bowlers / Tail End

6)Final Team

**First Page:-**

**Power Hitters/Openers:-** These are the batsman who are perfect for openers. They have the highest strike rate and consistant gameplay. Using this page we can choose easily who should be our opener.

* This page consist of various tools like slicer, chart, line chart, etc.
* In Power Bl, a slicer is a visual filter that you can place directly on a report page to control which data is displayed in other visualizations on the same page.
* It allows users to interactively explore the data and see how changes in one variable affect the others.
* Here's a breakdown of the functionalities and benefits of slicers in Power BI:
* Interactive Filtering: Users can click on specific values within a slicer to filter the data shown in charts, graphs, and other visuals on the report page. This allows for quick and easy exploration of trends and patterns within the data.
* Improved User Experience: Slicers provide a user-friendly way to filter data without requiring users to navigate through menus or dropdowns. This makes it easier for people to interact with the report and gain insights.
* Focus on Specific Data: Slicers can help focus the user's attention on specific subsets of data. For example, you could create a slicer for hotel location and then analyze revenue trends for a particular location.
* Highlighting Important Filters: By placing frequently used filters as slicers on the report canvas, you can ensure they are always readily accessible to users.
* Clarity on Applied Filters: Unlike dropdown filters, slicers visually show the currently selected filters, providing users with a clear understanding of how the data is being filtered.
* Here are some common types of slicers you can use in Power BI :-
* Dropdown Slicers: These allow users to select single or multiple values from a list.
* List Slicers: Similar to dropdown slicers, but display all available values in a vertical list.
* The Dig Can shown On next Page



Here, These slicers are designed to let users filter the data displayed in other visuals on the report page.

First Slicer: Filter By Name

This slicer allows users to filter the data by name.

The slicer includes the following players options:

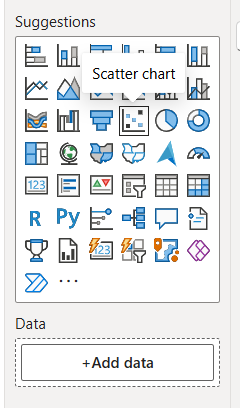
* Virat Kohli
* Rilee Rossouw
* Jos Buttler
* Suryakumar Yadav

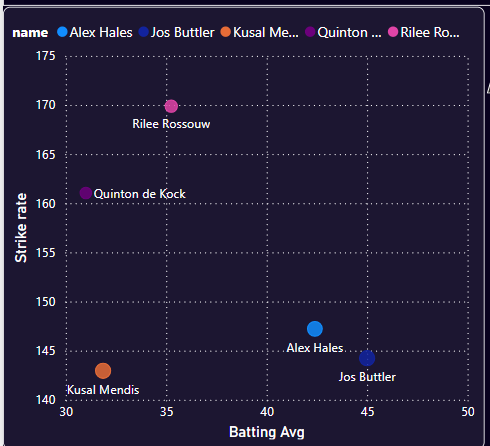
Users can select one or more players by clicking on their names in the slicer.

**SCATTER CHART**

* A **scatter plot** is a graphical representation used in data analytics to examine the relationship between two numerical variables. It consists of a series of points, where each point represents an individual observation in a dataset. These points are plotted on a two-dimensional graph, with one variable displayed on the **X-axis** (horizontal) and the other on the **Y-axis** (vertical).
* Scatter plots are particularly effective in identifying patterns, trends, and correlations between variables, as well as spotting anomalies or outliers that deviate from general behavior.
* In Power BI, scatter plots are an interactive and insightful way to explore relationships within a dataset. Users can drag and drop fields to define the X and Y axes and use additional options to enhance the chart. For example, the **Details** field allows each point to be uniquely identified (such as by a player name or customer ID), while the **Legend** groups points by categories (like teams or regions), and the **Size** parameter adjusts the size of the data points based on another numeric field.
* Moreover, the **Play Axis** in Power BI allows the creation of animated scatter plots over a time series, helping users see how data trends evolve over time.
* Scatter plots are widely used in various domains. In **cricket analytics**, for instance, a scatter plot might show the relationship between a player’s **batting average** (X-axis) and **strike rate** (Y-axis), with bubble size representing the number of matches played and colors indicating different teams.
* This visual helps in understanding which players are more consistent or aggressive and how they compare across teams.
* In business scenarios, scatter plots can help understand how sales figures relate to customer satisfaction or how advertising spend affects revenue.

Overall, scatter plots are a fundamental tool in exploratory data analysis, offering clear, intuitive insights into complex datasets by visually presenting relationships between key metrics.

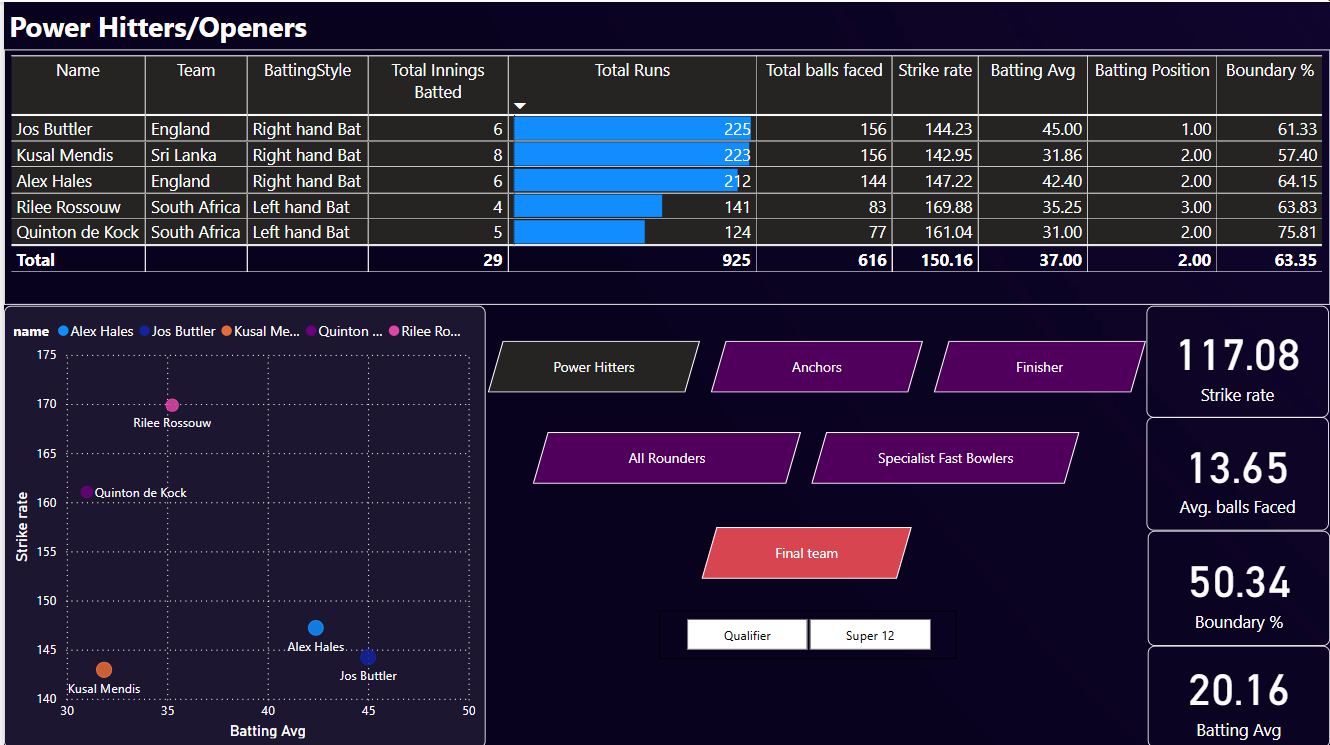




**Table:-**

* In a Cricket Analytics dashboard built using Power BI, the **Table visual** plays a vital role in presenting detailed, structured, and easy-to-interpret data.
* It resembles an Excel spreadsheet and is particularly useful for displaying granular-level information such as player statistics, match summaries, or comparative analysis between teams.
* For instance, a table can list out key player metrics like player names, matches played, runs scored, wickets taken, strike rate, and economy rate.
* This format allows viewers to compare performance side by side and derive meaningful insights with clarity.
* Moreover, tables are commonly used to display match-by-match summaries, where columns may include match date, competing teams, venue, result, top scorer, and player of the match.
* Such tables help in tracking tournament progress and evaluating team performances over time. Bowling analysis tables often include details like overs bowled, runs conceded, wickets taken, dot balls, and economy rate, offering valuable insights into bowling efficiency.
* Similarly, batting performance can be broken down by match or over, helping analysts understand a player’s consistency and impact under different conditions.
* Tables in Power BI support features such as sorting, conditional formatting, and slicer-based filtering, which enhance their interactivity and usefulness. For example, users can sort a player performance table by highest runs or filter by a specific team to view individual contributions.
* Conditional formatting allows highlighting of top performers, such as highlighting the best economy rates in green or the lowest strike rates in red, helping analysts visually identify trends without manual inspection.
* Overall, the table visual is indispensable in cricket dashboards for delivering clear, raw, and reliable insights. It provides users with the depth of data needed for player scouting, match reviews, and strategic planning. While other visuals like charts and graphs are excellent for trends and summaries, tables are best for delivering factual, record-level details in a visually structured format.

The screenshot of the first page i.e. **Power Hitters/Openers** is given on next page.



This Power BI dashboard segment delivers a high-impact, interactive view of Power Hitters and Openers in the ICC T20 World Cup 2022. By combining detailed tabular data with insightful scatter plots, dynamic role-based filtering, and instant KPI summaries, it empowers decision-makers to drill into player performance quickly. Buttons enhance usability by simplifying navigation between player categories and match stages, making it not just a static report but a live decision-support tool for analysts, coaches, and enthusiasts alike.

**Buttons:**

In **Power BI**, **buttons** are interactive elements that allow you to **enhance user navigation and interactivity** in reports and dashboards. They play a key role in making reports more dynamic, allowing users to explore data more efficiently. Let’s break it down with detailed explanations and examples.

The Buttons in power Bi is used to navigate form one page to another. There are six buttons on this pages namely

i]Power Hitters

ii]Anchors

iii]Finishers

iv]All Rounders

v] Special fast bowlers

vi] Final team

In the Page no one i.e. Power Hitters the page contains various Data visualization like tables, chart, button and card.

These buttons are likely linked with **Bookmarks and Page Navigation** features to toggle between datasets or filtered visuals. For instance, clicking “Anchors” would show players with higher batting averages and lower strike rates, while “Specialist Fast Bowlers” might switch the visuals to bowling metrics like economy rate and wickets.

This button-driven interface improves **user interactivity**, letting coaches or analysts filter players by their match role with a single click.

Power Hitters page contains of names of Four players. There is data given of all the players like for example Name, Team, Batting Style, Total innings batted, Total runs, Total balls faced by players, Players Strike Rate, Batting Average, Batting Position, Boundary Percentage.

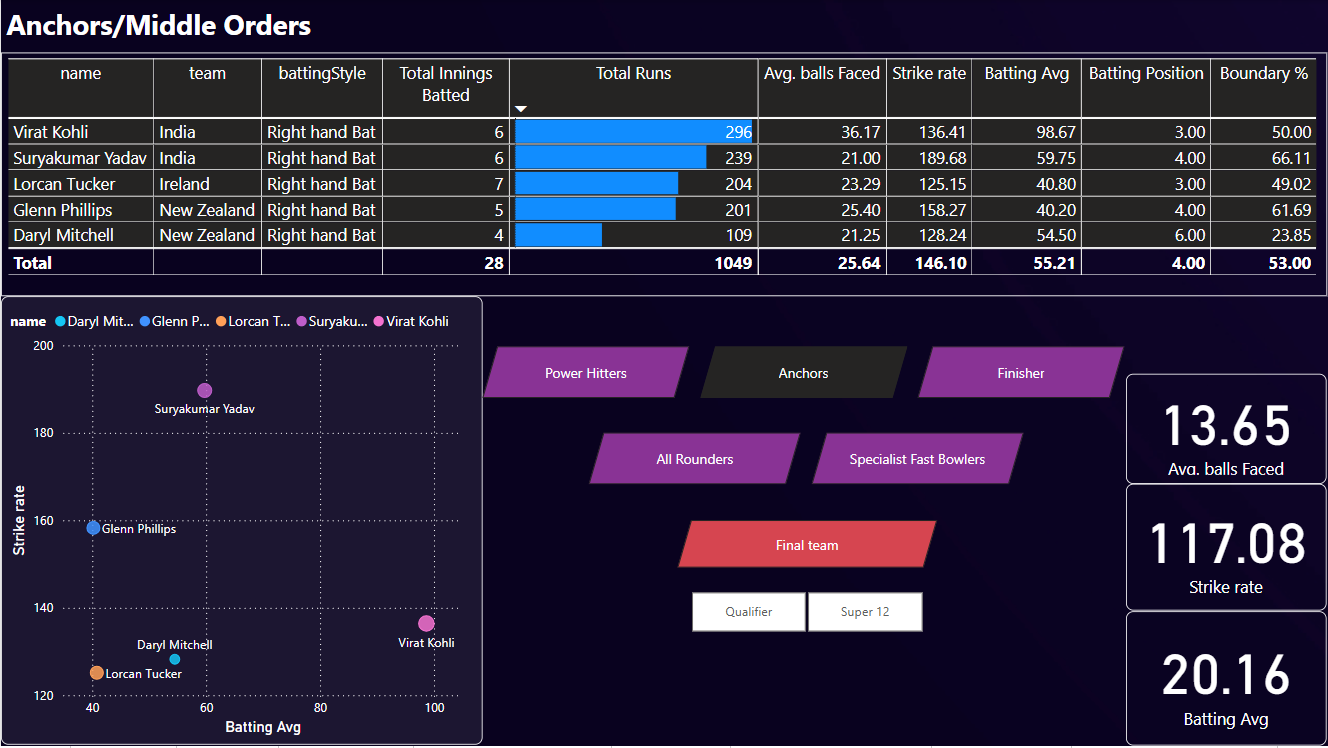
There are cards also which show the strike rate of all the selected players, Average Balls Faced, Boundary Percentage, Batting Average.

This report segment focuses on identifying and evaluating the **top-performing power hitters and opening batsmen** in the ICC T20 World Cup 2022, using a Power BI dashboard to visualize and interpret key batting metrics. The analysis is based on player performance data including innings played, runs scored, strike rate, batting average, balls faced, boundary percentage, and batting position. This visually interactive dashboard aids in talent recognition, match planning, and strategic player selection.

**Insights:**

* **Alex Hales** holds the highest strike rate (147.22) among consistent top-order performers, while **Rilee Rossouw** tops all with a strike rate of **169.88**, indicating his aggressive play style.
* **Quinton de Kock** has the **highest boundary percentage (75.81%)**, making him a key boundary hitter.
* **Jos Buttler** combines consistency with power, scoring the highest total runs (225) with a strong average of 45.00, making him the most reliable opener.

**Second Page**

****

This Power BI dashboard segment focuses on evaluating middle-order batsmen who serve as the backbone of the batting lineup—often referred to as Anchors. These players typically provide stability, build innings, and rotate the strike while accelerating when needed. The dashboard highlights key metrics such as total runs, strike rate, average, balls faced, and boundary percentage.

| **Player** | **Team** | **Runs** | **Strike Rate** | **Batting Avg** | **Boundary %** |
| --- | --- | --- | --- | --- | --- |
| **Virat Kohli** | India | 296 | 136.41 | **98.67** | 50.00% |
| **Suryakumar Yadav** | India | 239 | **189.68** | 59.75 | **66.11%** |
| **Lorcan Tucker** | Ireland | 204 | 125.15 | 40.80 | 49.02% |
| **Glenn Phillips** | New Zealand | 201 | 158.27 | 40.20 | 61.69% |
| **Daryl Mitchell** | New Zealand | 109 | 128.24 | 54.50 | 23.85% |

**Key Takeaways:**

* **Virat Kohli** leads in **total runs (296)** and boasts an exceptional **average (98.67)**, showcasing his consistency and dependability.
* **Suryakumar Yadav** combines an **aggressive strike rate (189.68)** with a strong average, making him a **high-impact middle-order striker**.
* **Glenn Phillips** and **Lorcan Tucker** contribute solid runs with moderate averages, playing a role in maintaining the scoring rate in the middle overs.
* **Daryl Mitchell**, despite fewer runs, holds a decent average (54.50) indicating potential as a reliable contributor.

**Scatter Plot – Strike Rate vs Batting Average**

The scatter chart visual maps players based on **batting average** and **strike rate**, enabling instant identification of high performers:

* **Suryakumar Yadav** is positioned as a standout high-strike-rate batsman with strong average.
* **Virat Kohli**, while not as aggressive, reflects unmatched consistency.
* The remaining players occupy mid-range positions with moderate averages and strike rates, indicating more supporting roles.

**Interactive Navigation Buttons**

This dashboard retains button-based role filtering—**Power Hitters**, **Anchors**, **Finishers**, etc.—allowing users to explore different player segments. The **“Qualifier” and “Super 12”** buttons likely serve as filters for match phases.

**KPI Cards (Static View)**

Although the values on the right appear static or generic across pages, they indicate average metrics:

* **Avg. Balls Faced**: 13.65
* **Strike Rate**: 117.08
* **Batting Average**: 20.16

These are likely page-level summaries and may not reflect the current "Anchors" group accurately unless dynamically linked.

**Conclusion**

This dashboard effectively visualizes **middle-order batting performance**, identifying key players who balance scoring with innings-building. **Kohli and Suryakumar** emerge as standout performers with unique strengths—consistency and aggression respectively. The clean layout, supported by charts and buttons, facilitates strategic selection and comparative performance analysis, aiding team management in middle-order planning.

**Finisher**

****

**Section Focus: Finisher and Lower-Order Batsmen**

This section of the Power BI dashboard focuses on evaluating the performance of finishers and lower-order anchors—players who play crucial roles in the final overs of an innings. These cricketers often come in under pressure and are tasked with finishing the game strongly or stabilizing a collapsing innings. The analysis combines both batting and bowling metrics, making it highly relevant for assessing all-rounder capabilities.

**Table Visual – Player Performance Breakdown**

| Player | Team | Runs | Batting Avg | Strike Rate | Wickets | Bowling Econ | Bowling SR |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Curtis Campher | Ireland | 126 | 25.20 | 163.64 | 2 | 9.25 | 24.00 |
| Glenn Maxwell | Australia | 118 | 39.33 | 161.64 | 3 | 6.00 | 6.33 |
| Marcus Stoinis | Australia | 126 | 42.00 | 161.54 | 1 | 9.67 | 54.00 |
| Sikandar Raza | Zimbabwe | 219 | 27.38 | 147.97 | 10 | 6.50 | 14.40 |
| Hardik Pandya | India | 128 | 25.60 | 131.96 | 8 | 8.11 | 13.50 |

**Key Insights:**

* Curtis Campher tops the strike rate chart (163.64), reflecting strong finishing ability.
* Marcus Stoinis leads with a batting average of 42.00, showcasing consistency in the lower order.
* Sikandar Raza and Hardik Pandya provide the best all-round contributions, with Raza taking the most wickets (10) and maintaining a solid strike rate.
* Glenn Maxwell is a highly efficient bowler with the best bowling economy (6.00) and strike rate (6.33), complementing his explosive batting.

**Scatter Plot – Strike Rate vs Batting Average**

The scatter chart positions finishers based on their strike rate and batting average:

* Stoinis and Maxwell stand out as players offering both quick scoring and average stability.
* Curtis Campher is shown with a high strike rate but a slightly lower batting average.
* Hardik Pandya, although consistent, has the lowest strike rate among the group.

This visual supports quick decision-making by identifying ideal finishers based on aggressiveness and reliability.

**Interactive Buttons and Filters**

Just like other pages, this dashboard includes navigation buttons (Power Hitters, Anchors, Finishers, etc.) and match phase filters (Qualifier, Super 12), enabling focused analysis by player role and tournament stage.

KPI Cards (Possibly Static)

* Batting Average: 20.16
* Strike Rate: 117.08
* Bowling Strike Rate: 18.43
* Economy: 7.32

These values may represent overall or static indicators for benchmarking but aren’t dynamic to the current segment unless linked.

**Conclusion**

This dashboard presents a strong comparative view of finishers and lower-order anchors, combining both batting explosiveness and bowling impact. Players like Stoinis, Maxwell, and Raza demonstrate high all-round value. This report helps team selectors and analysts evaluate which players can best close out matches or support during the death overs with both bat and ball.

**All Rounders/ Lower Middle Order**

****

**Section Focus: Dual-role Players (Batting & Bowling)**

This section provides insights into the **performance of key all-rounders**, who serve as both reliable lower-middle order batters and important bowling assets. These players are crucial in balancing a team by contributing in both innings.

**Table Visual – Player Performance Summary**

| **Player** | **Team** | **Bat Avg** | **Strike Rate** | **Wickets** | **Economy** | **Bowling SR** |
| --- | --- | --- | --- | --- | --- | --- |
| **Mitchell Santner** | New Zealand | 27.00 | 168.75 | 9 | 6.45 | **13.33** |
| **Rashid Khan** | Afghanistan | 28.50 | **178.13** | 4 | 6.42 | 18.00 |
| **Shadab Khan** | Pakistan | 24.50 | 168.97 | 6 | **6.35** | 14.18 |
| **Sikandar Raza** | Zimbabwe | 27.38 | 147.97 | **10** | 6.50 | 14.40 |

**Key Insights:**

* **Rashid Khan** holds the **highest strike rate (178.13)**, offering excellent lower-order acceleration.
* **Sikandar Raza** emerges as a top all-rounder with **10 wickets**, a good batting average (27.38), and consistent economy.
* **Mitchell Santner** showcases the best **bowling strike rate (13.33)** among the four, taking wickets frequently and economically.
* **Shadab Khan** offers a balanced performance with high batting strike rate (168.97) and **best economy rate (6.35)**.

**Scatter Plot – Bowling Strike Rate vs Economy**

This scatter plot gives a quick visual on bowling efficiency:

* **Shadab Khan** stands out in the lower-left zone, indicating **best economy with low strike rate**—ideal for controlling runs while taking wickets.
* **Rashid Khan** shows a higher strike rate but stays economical.
* All players cluster closely between **economy rates of 6.35–6.50**, indicating disciplined bowling performances from all-rounders.

**Interactive Navigation**

Navigation buttons (Power Hitters, Anchors, Finishers, etc.) allow easy cross-comparison of player types. The “Qualifier” and “Super 12” filters help analyze performance by match stages.

**KPI Cards (Static)**

* **Batting Average**: 20.16
* **Strike Rate**: 117.08
* **Bowling Strike Rate**: 18.43
* **Economy**: 7.32

These cards likely reflect aggregate benchmarks for comparison, though the featured all-rounders consistently **outperform these values**.

**Conclusion**

This dashboard highlights the effectiveness of all-rounders in contributing across both innings. Players like **Rashid Khan** and **Sikandar Raza** offer game-changing capabilities, while **Santner** and **Shadab** provide stability and tactical flexibility. These insights are essential for team composition, especially in knockout stages where depth in batting and bowling matters most.

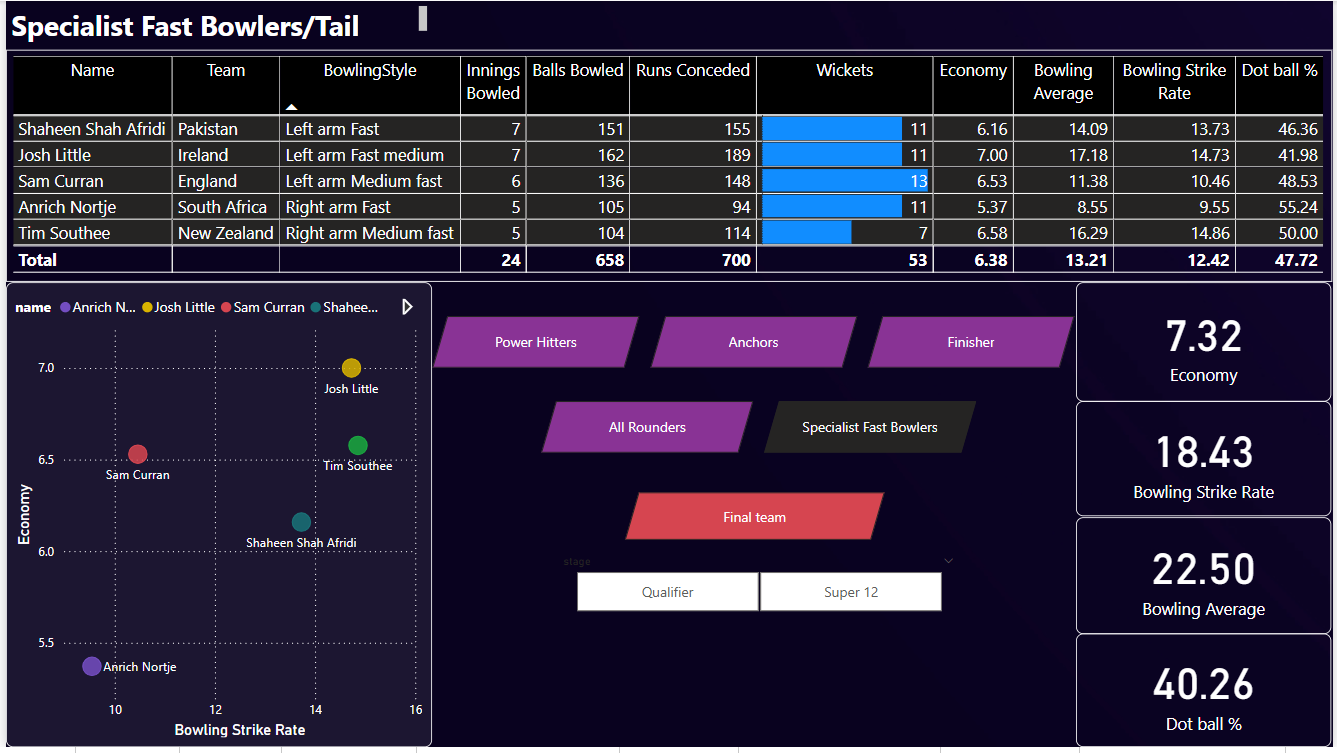
**Specialist Fast Bowlers/Tail End**

**Introduction**

The performance of fast bowlers is often a decisive factor in T20 cricket. In short-format games, where every delivery counts, specialist fast bowlers play a pivotal role in shaping match outcomes. The Power BI dashboard on *Specialist Fast Bowlers/Tail* brings together key performance indicators that allow us to analyze, compare, and evaluate some of the top fast bowlers in the ICC Men’s T20 World Cup 2022.

This dashboard offers a consolidated view of bowling performances across multiple metrics such as economy rate, bowling average, strike rate, dot ball percentage, and wicket tallies. The inclusion of visual tools such as scatter plots and performance categories helps identify outliers and trends that might otherwise be lost in tabular data.

The goal of this report is to analyze and interpret this visualization thoroughly, draw out meaningful insights, compare individual performances, and understand the broader implications for team strategies, selection, and match tactics.



**Dataset Overview and Dashboard Structure**

The dashboard presents data for five specialist fast bowlers:

* Shaheen Shah Afridi (Pakistan)
* Josh Little (Ireland)
* Sam Curran (England)
* Anrich Nortje (South Africa)
* Tim Southee (New Zealand)

Each bowler's data includes the following fields:

* **Innings Bowled**
* **Balls Bowled**
* **Runs Conceded**
* **Wickets Taken**
* **Economy Rate**
* **Bowling Average**
* **Bowling Strike Rate**
* **Dot Ball Percentage**

The layout includes:

* A table summarizing individual and total statistics.
* A scatter plot comparing Economy Rate vs. Bowling Strike Rate.
* Highlighted KPIs (Key Performance Indicators) on the right for average Economy, Strike Rate, Bowling Average, and Dot Ball %.
* Category selections: Power Hitters, Anchors, Finishers, All-Rounders, Specialist Fast Bowlers.
* A tournament stage filter: Qualifier and Super 12.

**Statistical Analysis of Bowlers**

**Total and Aggregate Performance**

From the table, we observe:

* **Total Innings Bowled**: 24
* **Total Balls Bowled**: 658
* **Total Runs Conceded**: 700
* **Total Wickets Taken**: 53
* **Average Economy Rate**: 6.38
* **Average Bowling Average**: 13.21
* **Average Bowling Strike Rate**: 12.42
* **Dot Ball Percentage**: 47.72%

These figures point toward an effective and efficient group of fast bowlers. An average economy of 6.38 is impressive in T20 cricket, where run rates tend to soar. The combined strike rate of 12.42 implies a wicket nearly every two overs, which is highly valuable in restricting the opposition's scoring.

**Individual Performance Analysis**

**Anrich Nortje (South Africa)**

* **Innings**: 5 | **Balls**: 105 | **Runs**: 94 | **Wickets**: 11
* **Economy**: 5.37 | **Average**: 8.55 | **Strike Rate**: 9.55 | **Dot %**: 55.24%

Nortje stands out as the top performer in nearly all bowling metrics. With the lowest economy and best strike rate, he appears to be the most effective wicket-taking option while conceding the fewest runs.

**Sam Curran (England)**

* **Innings**: 6 | **Balls**: 136 | **Runs**: 148 | **Wickets**: 13
* **Economy**: 6.53 | **Average**: 11.38 | **Strike Rate**: 10.46 | **Dot %**: 48.53%

Curran is the leading wicket-taker among the group, with 13 dismissals. While his economy is slightly higher than Nortje's, his consistent wicket-taking ability makes him a strike weapon, especially in death overs.

**Shaheen Shah Afridi (Pakistan)**

* **Innings**: 7 | **Balls**: 151 | **Runs**: 155 | **Wickets**: 11
* **Economy**: 6.16 | **Average**: 14.09 | **Strike Rate**: 13.73 | **Dot %**: 46.36%

Afridi had a strong performance, with a respectable strike rate and economy. His ability to deliver breakthroughs in powerplays is a crucial asset.

**Josh Little (Ireland)**

* **Innings**: 7 | **Balls**: 162 | **Runs**: 189 | **Wickets**: 11
* **Economy**: 7.00 | **Average**: 17.18 | **Strike Rate**: 14.73 | **Dot %**: 41.98%

Little's economy is the highest in this cohort, but his performance must be viewed relative to the strength of Ireland as a team. Despite that, his 11 wickets show his effectiveness.

**Tim Southee (New Zealand)**

* **Innings**: 5 | **Balls**: 104 | **Runs**: 114 | **Wickets**: 7
* **Economy**: 6.58 | **Average**: 16.29 | **Strike Rate**: 14.86 | **Dot %**: 50.00%

Southee displayed strong control with a solid dot ball percentage and decent economy. His experience in pressure situations contributes to his value.

**Visual Analysis – Scatter Plot**

The scatter plot visualizes the relationship between **Economy Rate** (Y-axis) and **Bowling Strike Rate** (X-axis). Here’s what the plot reveals:

* **Bottom Left Quadrant (Best Zone)**: Anrich Nortje occupies this zone – low economy and strike rate – indicating maximum efficiency.
* **Mid-Zone**: Shaheen Afridi and Sam Curran balance economy and strike rate, suggesting reliability and impact.
* **Top Right Quadrant (High Economy, High Strike Rate)**: Josh Little is in this zone, showing slightly less efficiency.
* **Southee** is placed near the center – indicating balanced, though not dominant, performance.

This visual helps analysts quickly identify optimal bowlers and those needing strategic reassignment.

**KPI Interpretation**

The KPI cards display:

* **Average Economy**: 7.32
* **Average Bowling Strike Rate**: 18.43
* **Average Bowling Average**: 22.50
* **Average Dot Ball Percentage**: 40.26%

These numbers are likely dynamically changing based on filters (e.g., match stages), providing instant updates for team planning. Compared to the values in the main table, these metrics might reflect a broader pool, or a filtered dataset.

**Role Categorization and Team Planning**

The category boxes (e.g., Power Hitters, Anchors, Specialist Fast Bowlers) reflect a data model that classifies players by roles. This segmentation is useful for:

* **Balancing team composition**
* **Targeted strategy creation** (e.g., pairing fast bowlers with finishers for death overs)
* **Identifying replacements based on role-specific KPIs**

“Specialist Fast Bowlers” is highlighted here, isolating this subgroup for focused analysis. The “Final Team” box shows how bowlers integrate into the overall team-building strategy.

**Tournament Stage Filter – Strategic Evaluation**

The stage filter (Qualifier vs. Super 12) allows users to analyze how bowler performance varies across match pressure. Performance consistency across stages is a marker of elite ability.

For instance:

* **Did a bowler perform better against top teams in Super 12?**
* **Was someone a surprise performer in the qualifiers?**

These are questions teams can explore using this slicer.

**Tactical Implications and Strategic Insights**

**Bowler Utilization**

Analyzing overs bowled per innings can reveal when each bowler is used. For example:

* **Powerplay specialists**: Shaheen Afridi
* **Middle overs containment**: Tim Southee
* **Death over execution**: Sam Curran, Anrich Nortje

**Dot Ball Influence**

Dot balls not only restrict runs but build pressure leading to wickets. Nortje's 55.24% dot rate is exceptional. Combined with a strike rate of 9.55, he is a lethal combination.

**Economy vs. Impact**

Sometimes, high economy is tolerable if it comes with breakthroughs. Curran is a case in point. Team management must decide whether to prioritize containment or wickets based on match situations.

**The Final 11**



This Page allows you to select the Final 11 players easily. This page contain the list of all players that were in T20 2022. Also Shows the following points of the players combined.

* **Batting Average**
* **Strike Rate**
* **Average Balls Faced**
* **Bowling Average**
* **Dot Balls Percentage**
* **Economy**
* **Bowling Strike Rate**

The back button is also given in order to go to previous page.

**11.Drawbacks and Limitation**

While data analytics in cricket offers substantial benefits, like improved performance evaluation, strategy development, and fan engagement, it also comes with several limitations and drawbacks that must be considered. These limitations stem from technical, operational, analytical, and ethical aspects of building and maintaining such a system. In this section, we delve deep into the various challenges and constraints faced in an end-to-end cricket data analytics project using Python and Power BI.

**11.1 Data Source Limitations**

**Dependency on Third-party Websites:-**

One of the primary limitations of this project is the reliance on third-party websites like ESPNcricinfo or the ICC’s official site for data collection. These platforms are not officially designed for API-based data consumption, and hence, scraping their data becomes a fragile and often unsustainable solution.

* **Unstructured HTML Content**: Websites are designed for human readability, not for automated data extraction.
* **Frequent Structural Changes**: Even a minor change in the HTML structure can break scraping scripts.
* **Legal/Ethical Constraints**: Automated scraping may violate the terms of service of some websites, exposing the project to legal risks.

**Incomplete and Inconsistent Data**

Web scraping often leads to data inconsistencies, missing entries, or gaps in historical records.

* Some datasets may lack detailed ball-by-ball information.
* Data might not always be updated promptly, especially during live matches.
* Key metrics like player fitness, pitch reports, and weather conditions are often excluded.

**11.2 Data Quality Issues**

**Missing and Null Values**

A common challenge in scraped datasets is dealing with missing or null values. If these are not handled properly, they can lead to inaccurate aggregations and misleading insights.

* Missing innings scores, dropped catches, or misrecorded player names can skew statistics.
* Null values in key performance indicators (like strike rate or economy rate) can disrupt dashboards and analytics.

**Inconsistent Naming Conventions**

Players' names may be spelled differently across various sources, or there may be conflicts in match records due to regional naming styles.

* “Rohit Sharma” vs “R. Sharma” or “Sharma, Rohit” can be interpreted as different entities unless cleaned thoroughly.
* Inconsistent team abbreviations can result in incorrect joins and aggregations.

**11.3 Technical Limitations**

**Limited Real-time Analysis**

While Power BI supports scheduled refreshes, it is not ideal for real-time analytics, especially when the data is being fetched through Python scripts and manual scraping.

* Real-time match analysis (like a live scoreboard or player stats update) cannot be implemented efficiently.
* Delay in data refresh can mislead users relying on up-to-date performance metrics.

**Performance Bottlenecks**

Handling large datasets with high granularity, such as ball-by-ball match data for multiple tournaments, can lead to performance degradation.

* Python scripts may slow down with large CSV or JSON files.
* Power BI reports with complex DAX measures and high cardinality visuals may lag, especially on machines with limited RAM.

**Tool Compatibility**

The system requires compatibility between different tools—Python for data processing and Power BI for visualization. Bridging these environments may involve manual data transfers, version mismatches, and formatting issues.

* Power BI may not interpret certain data types or field structures created in Python.
* Exporting large datasets from Python to CSV or Excel and then importing to Power BI increases friction and risks data loss.

**11.4 Analytical Limitations**

**Bias in Data Interpretation**

Analytics systems depend heavily on how the data is prepared and interpreted. Biases can enter during feature selection, aggregation, or visual representation.

* Highlighting only top-performers may overshadow consistent middle-order contributors.
* Over-reliance on averages can hide variability and context-dependent performance.

**Lack of Contextual Data**

Data analytics focuses on numbers but often lacks contextual understanding—such as player injuries, emotional pressure, or captaincy changes.

* A low score may be due to a challenging pitch, not poor form.
* Bowling economy may vary due to match situations rather than player skill.

**11.5 Functional Limitations**

**Manual Data Update Processes**

Since the data pipeline is not fully automated end-to-end, many processes require human intervention:

* Running scraping scripts.
* Cleaning and formatting CSVs.
* Manually importing into Power BI for visualization.

This manual handling increases chances of errors and makes the system inefficient for ongoing or large-scale use.

**Limited Interactivity for End Users**

While Power BI offers some interactivity through slicers and filters, the end user cannot modify backend data or logic. The insights are limited to what has been pre-defined by the developer.

* Users cannot request “custom queries” on-the-fly.
* Advanced filtering or AI-driven insights are not available unless explicitly designed.

**11.6 Maintenance and Scalability Challenges**

**Code and Dashboard Maintenance**

As the project evolves, maintaining the scraping scripts, Python code, and Power BI dashboards becomes a significant challenge.

* Dashboards may break if the dataset structure changes.
* Scripts need constant updating to account for new tournaments or players.

**Scalability Issues**

Expanding the project to include multiple formats (Test, ODI, T20), tournaments, or additional data layers (like sentiment analysis or fantasy points) requires substantial reengineering.

* Existing Power BI data models may not support multi-format data without performance loss.
* More detailed data requires higher compute resources for both Python processing and Power BI rendering.

**11.7 Security and Privacy Concerns**

While not extremely sensitive, sports data analytics may touch on player profiles, injury reports, or coaching strategies which could be confidential.

* Improper handling or unauthorized sharing of data could lead to reputational or legal issues.
* Web scraping from unofficial sources may expose the system to security breaches.

**11.8 User Training and Adoption**

A sophisticated analytics dashboard, even when built perfectly, requires user understanding and training.

* Coaches or selectors may not be familiar with reading visual dashboards.
* There’s a risk that decisions are made based on misinterpreted graphs or KPIs.

**11.9 Visualization Constraints**

Power BI, although a powerful BI tool, does have limitations:

* Custom visuals are limited compared to dedicated front-end frameworks like D3.js or Tableau extensions.
* Interactive features are dependent on DAX complexity and data model performance.
* Mobile optimization of Power BI dashboards remains a challenge.

**11.10 External Dependency Risks**

This project’s success is tightly coupled with several third-party tools and platforms:

* Python library updates may break code compatibility.
* Power BI version changes can alter visual behavior or refresh settings.
* Web sources shutting down or becoming paywalled could kill the data pipeline.

**11.11 Domain-Specific Challenges**

Cricket as a sport is inherently complex. Analytical modeling has to consider a large variety of rules, match conditions, and unpredictable variables:

* Rain interruptions, Duckworth-Lewis method, and super overs introduce statistical complications.
* Player roles (e.g., all-rounders) make it difficult to classify performance on simple axes.
* Historic data lacks granularity needed for deep ML or AI models.

**12.Conclusion**

The **Cricket Data Analytics Project** represents a comprehensive, end-to-end solution designed to transform raw cricket data into actionable insights using modern data science techniques and visualization tools. By leveraging the combined power of Python for data processing and Power BI for business intelligence, this project not only showcases technical proficiency but also illustrates the growing significance of data-driven decision-making in the domain of sports, particularly cricket.

Cricket is no longer just a game of instinct; it's a game of numbers, precision, and strategy. This project demonstrates how analytics can bridge the gap between subjective judgment and objective decision-making. Through systematic data collection, rigorous cleaning, intelligent transformation, and insightful visualization, the system supports analysts, coaches, selectors, and even fans in understanding performance at a granular level.Key insights such as player efficiency in different match scenarios, team performance trends, venue-based patterns, and individual contributions under pressure help stakeholders make more informed and accurate decisions.

In conclusion, the Cricket Data Analytics Project stands as a testament to how data science and business intelligence tools can revolutionize traditional domains like sports. It embodies the future of cricket management, where every ball, shot, and player movement can be captured, analyzed, and used to drive smarter decisions.

Cricket, once known for its unpredictability and nuance, is now being redefined through analytics—and this project is a step in that direction.