**Cricket Data Analysis (Python | Power BI) - October 2024**

**Problem Statement**

The goal of this project is to identify the best 11 players based on T20 World Cup data, with specific performance criteria:

* The team must average at least 180 runs scored.
* The team must also be able to defend an average of 180 runs.

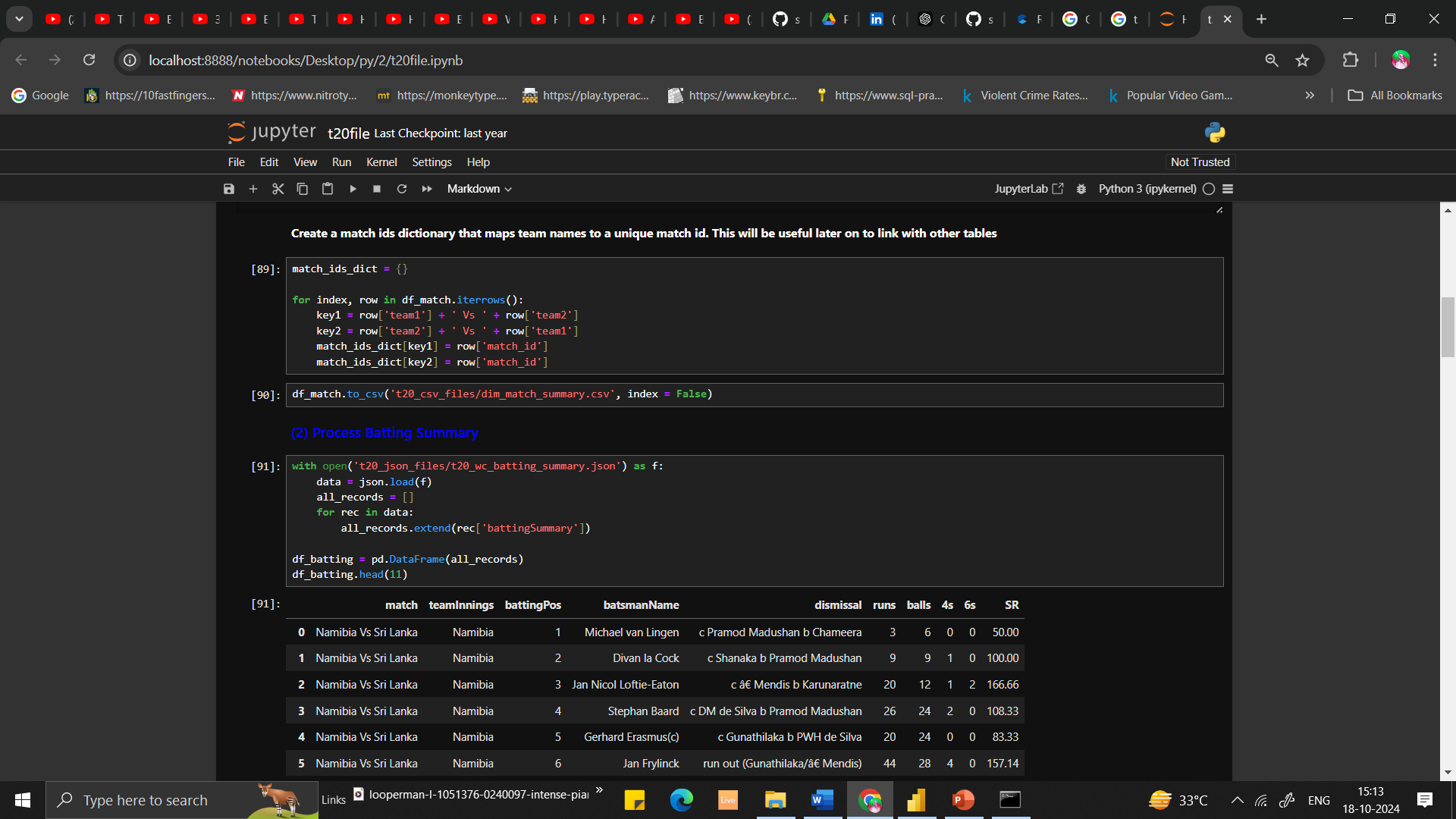
To achieve this, players were categorized into distinct roles: Openers, Batsmen, Bowlers, and All-Rounders. Each category has unique metrics to evaluate player performance. The overall selection process involves a summation algorithm that ranks players based on their individual metrics.

**Methodology**

**Data Collection**

* **Web Scraping:** Data was collected from the ESPN Sports website using a web scraper developed in Python. The scraper extracted relevant data and delivered it in JSON format.

**Data Cleaning**

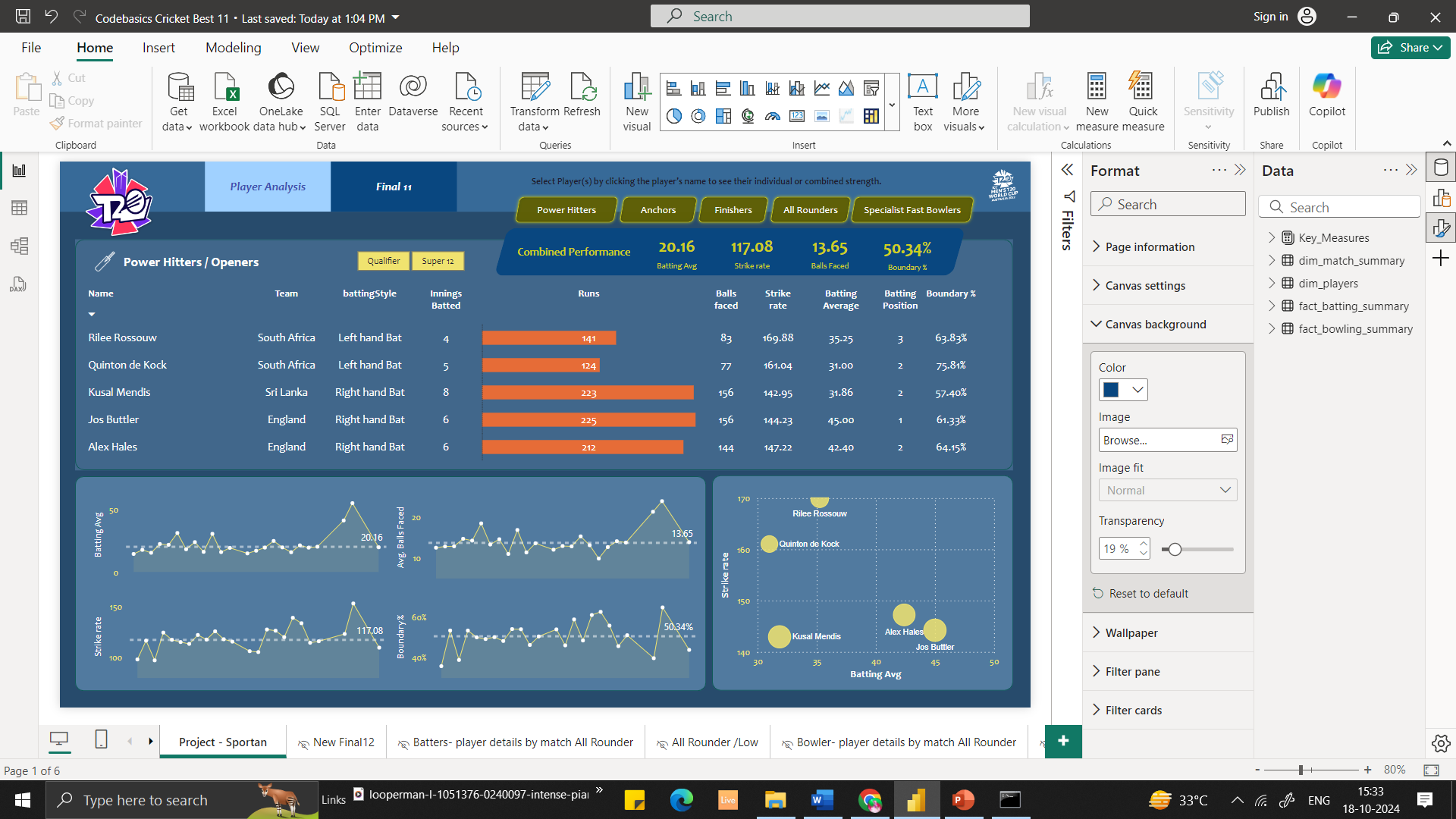
* The JSON files were imported into Python, where data cleaning and transformation were performed using the Pandas library. This included:
  + Removing special characters.
  + Merging datasets of match results with player statistics.
* After cleaning, the data was saved as a CSV file for further analysis in Power BI.

**Dashboard Creation**

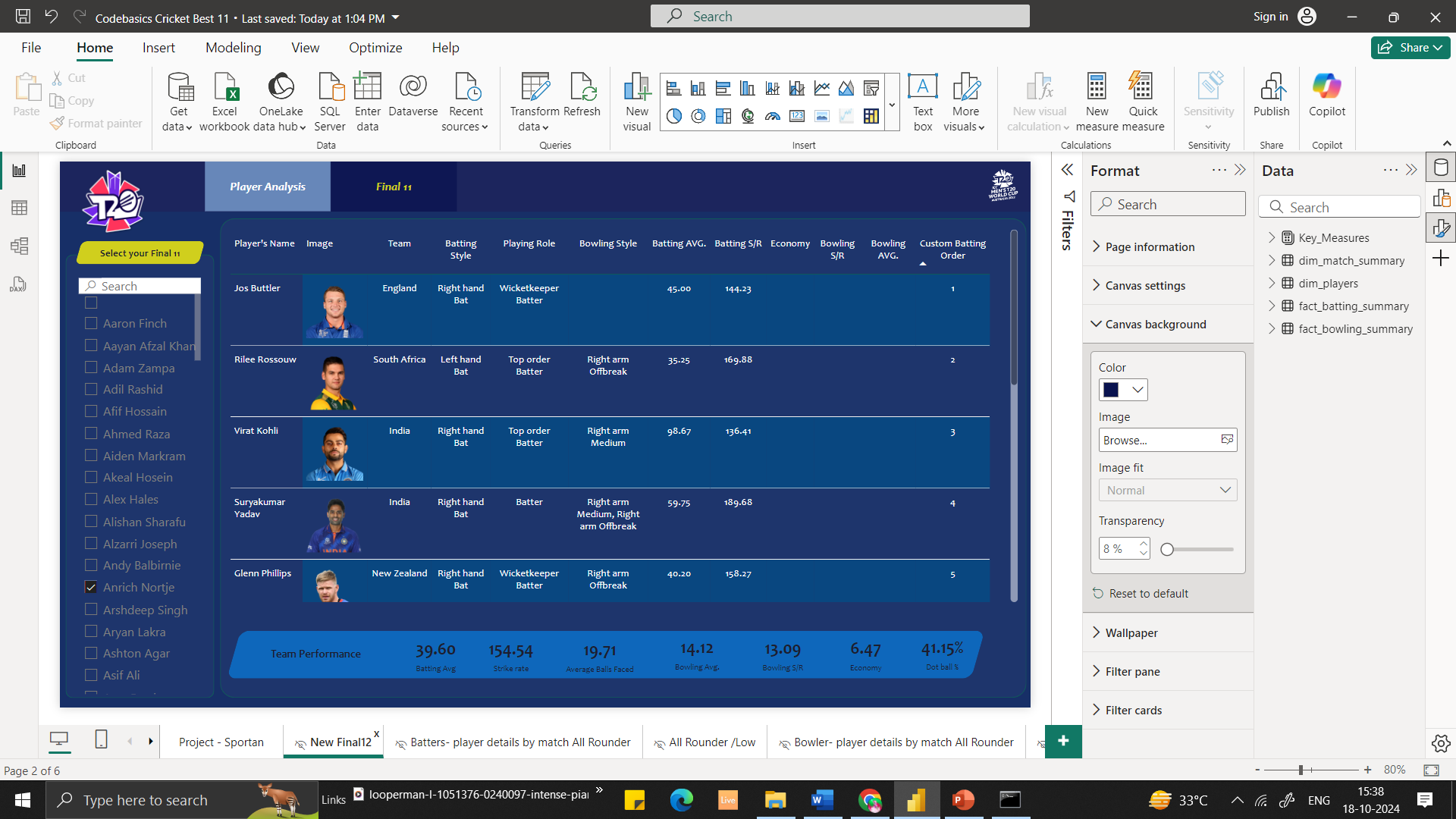
* Before creating the dashboard in Power BI, additional data cleaning and transformation were conducted using Power Query. This included:
  + Trimming and formatting data.
  + Removing duplicates.
  + Creating custom columns and measures for total runs, total matches, strike rate, boundary percentage, etc.

**Dashboard Insights**

* **Dashboard 1:**
  + A table displaying player information with slicers for easy filtering.
  + Graphs illustrating batting average, strike rate, and boundary percentage for each player.
  + A scatter chart to visualize the relationship between bowling strike rate and player ratings, highlighting the importance of lower bowling strike rates for better performance.

 **Dashboard 2:**

* + A table presenting the final selection of the best 11 players.
  + Player selection was based on metrics such as runs scored per over, wickets taken per two overs, high batting strike rates, and effective bowling.

 **Project Learnings**

* Gained experience in web scraping using Beautiful Soup 4 and other Python libraries.
* Improved Python skills for data cleaning and data manipulation using Pandas.
* Enhanced understanding of Power BI for data visualization and dashboard creation.

**Conclusion**

This project successfully identifies the best 11 players for a T20 team based on comprehensive data analysis and visualization techniques. The methodology outlined provides a framework for future analyses in cricket and other sports domains.

4o mini