**Centre for Analysis, Research and**

**Visualisation (C-MANAV)**

# STARC -Summer Program 2023

**Spectoclava** – Cricket Matchup Dashboard with

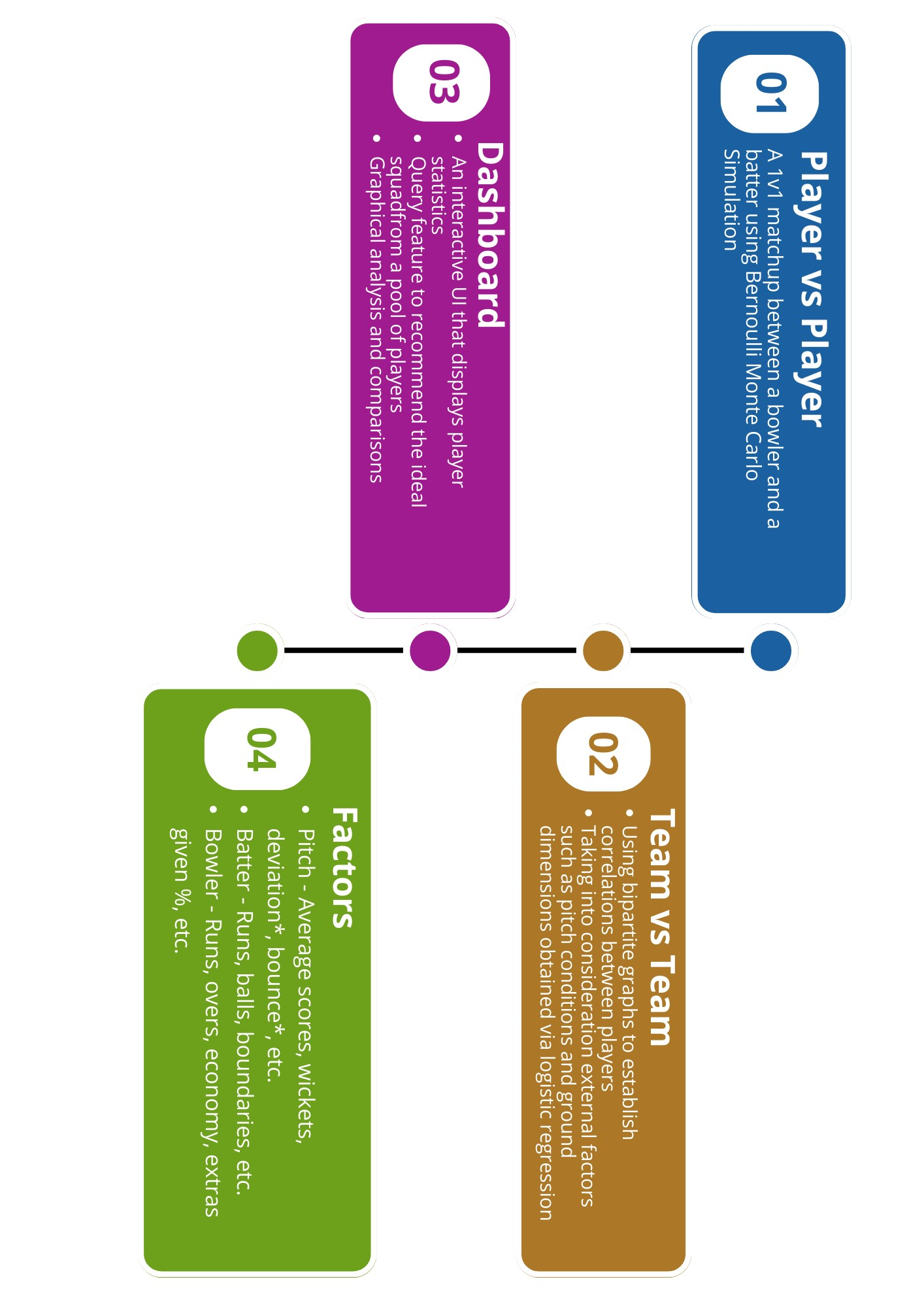
Recommender and Predictor

**Abstract:**

Cricket, a captivating sport with its intricate dynamics, is shaped by a multitude of factors, both known and unknown such as pitch conditions, player performance (recent form), team dynamics, and historical performance data, that play a significant role in determining match outcomes. We aim to explore the various conditions that affect the game of cricket and propose a framework using a dashboard to predict matchups between players and teams.

**Framework**

**Spectoclava – Matchup and Recommender Dashboard**



This framework will aid in making informed decisions to form the most efficient and competitive ‘Playing Eleven’ for tournaments, considering the diverse factors mentioned above.

The model utilizes a comprehensive approach that incorporates statistical analysis, data mining, and machine learning techniques to analyse and interpret the vast amount of available cricket data. By considering the aforementioned factors, the proposed model seeks to uncover hidden patterns using data manipulation to generate features that contribute to the game's outcome.

This framework provides a reliable foundation for predicting player matchups, enabling teams and selectors to make strategic decisions while selecting the ‘Playing Eleven’. This will help by identifying key strengths and weaknesses of individual players and assessing their performance against specific opponents, teams can maximize their chances of success.

**The Dataset**

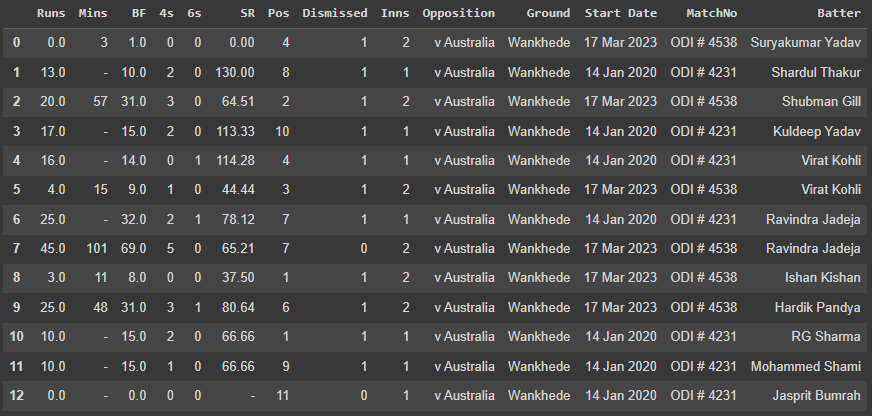
The dataset for each player is scraped from ESPN Cricinfo throughout their career (2000-2023\*) which includes statistics against teams and year-wise statistics. This is done separately for each bowler and batter. The data for each pitch is also scraped and divided into first and second innings. The statistics for each player vs player and team vs team is scraped separately due to the scattered nature of the dataset.

The datasets are cleaned by removing unwanted columns, and the data types of numerical values are converted from strings after removing unwanted characters at play. The raw data is hence converted into the desired format for it to be utilised effectively.

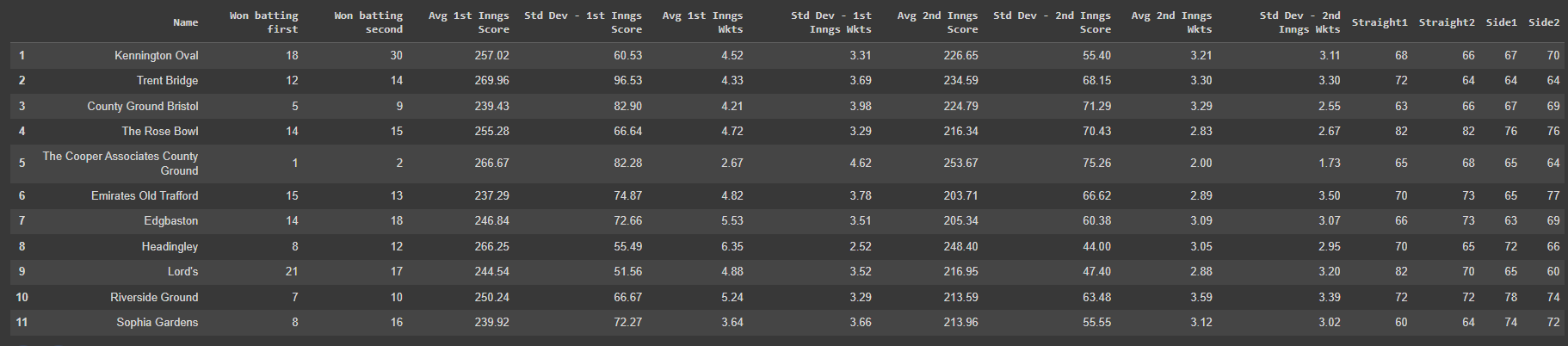
**Bowler Data**

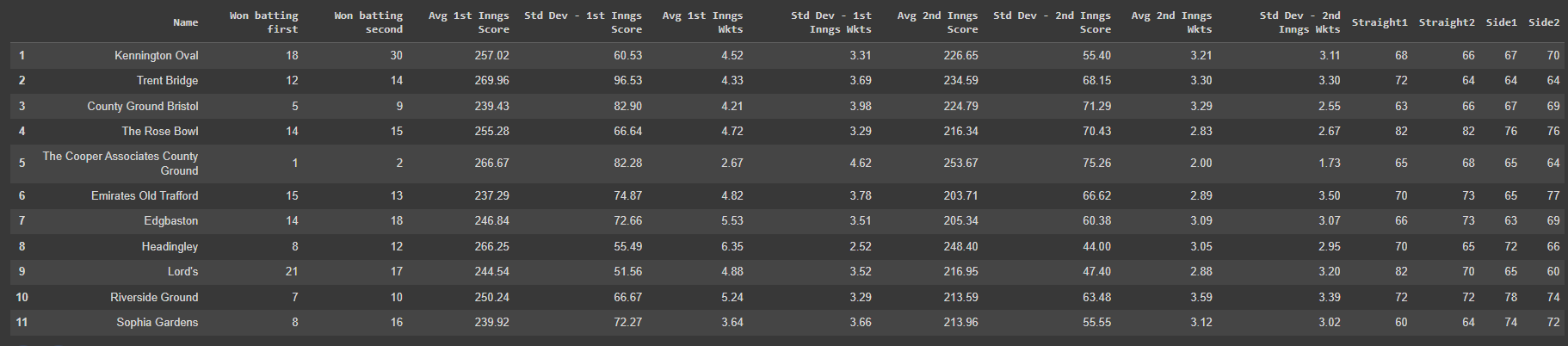
****

**Batter Data**

****

**Pitches Data**

****

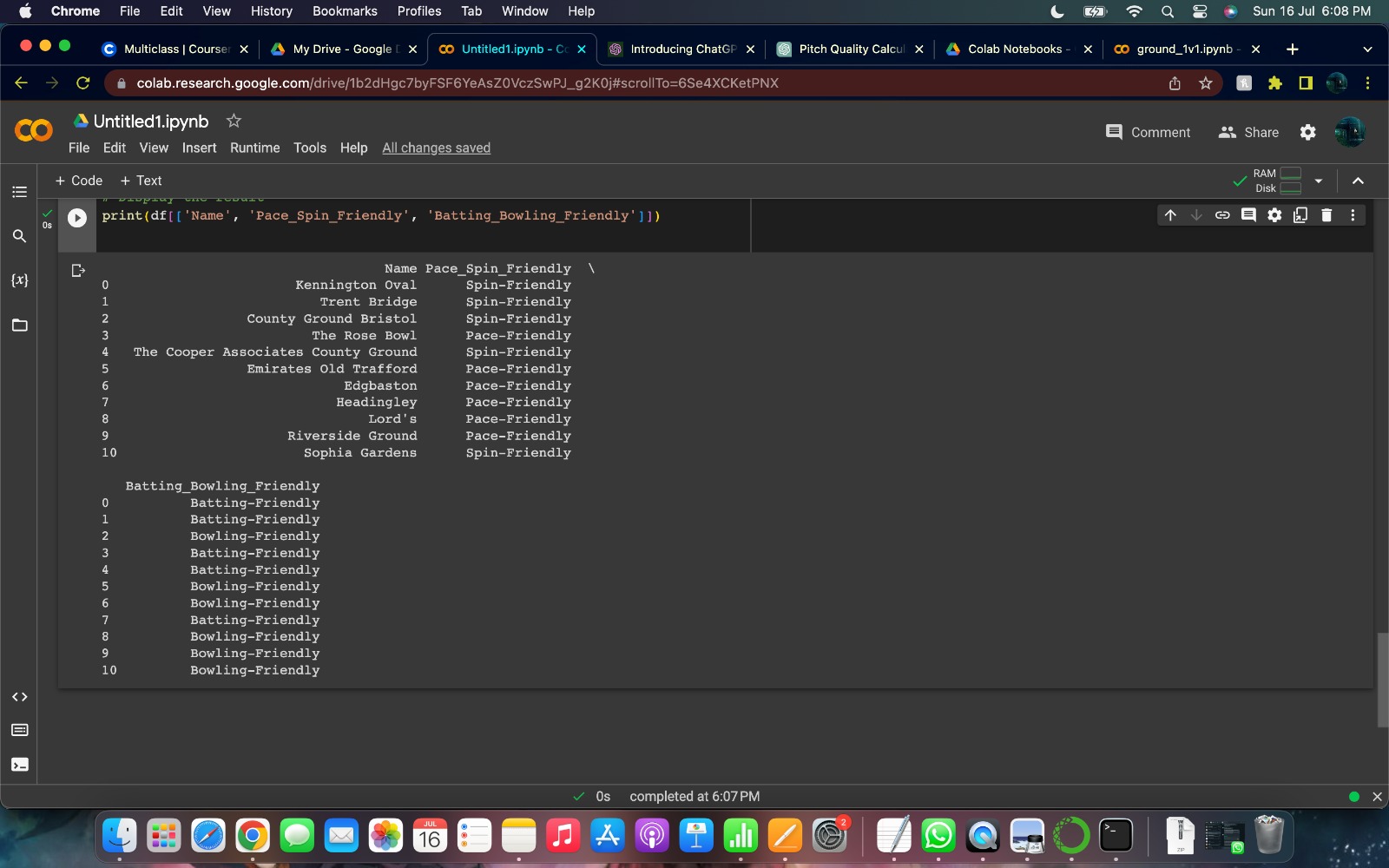
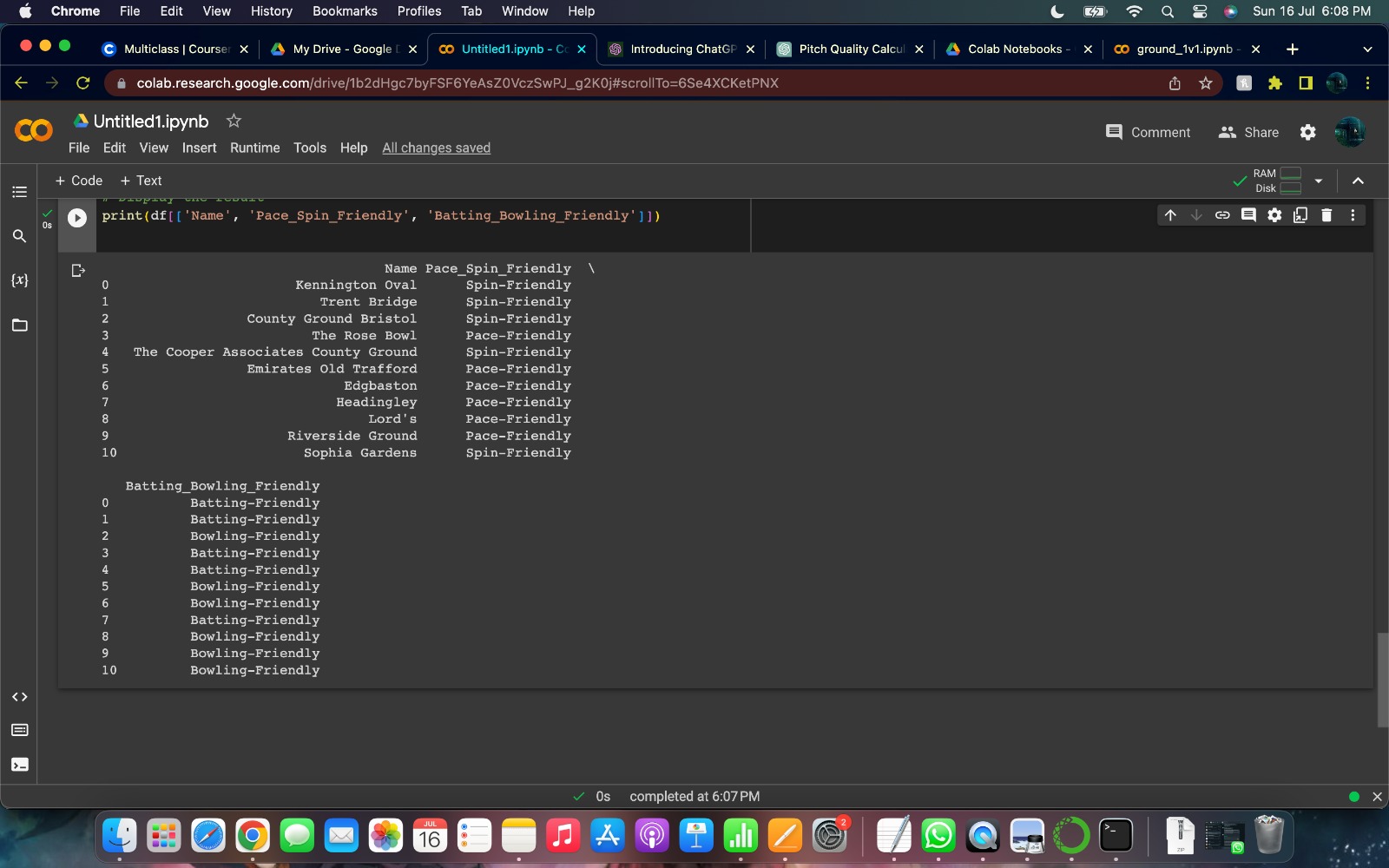
****

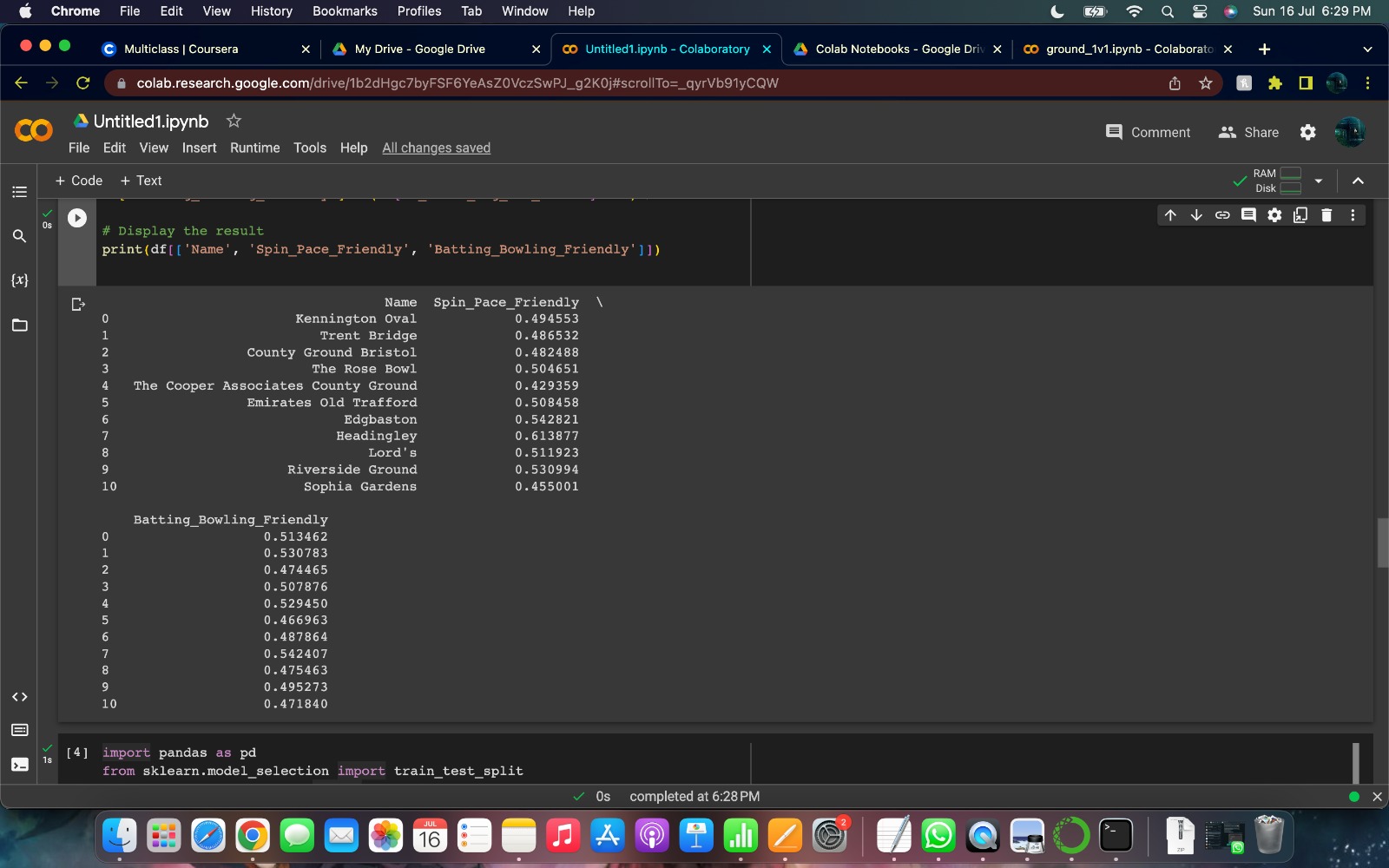
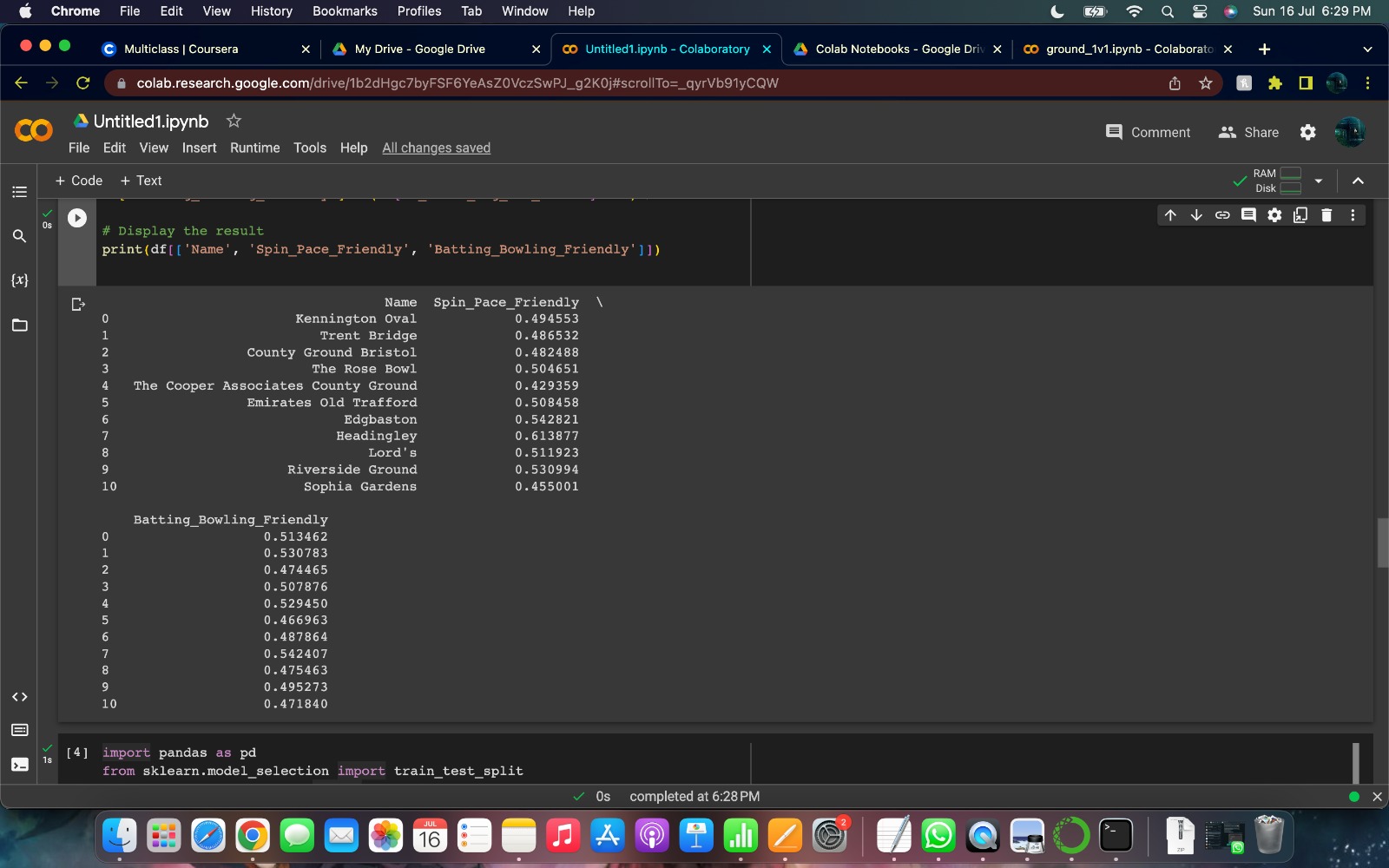
**The Model**

The cleaned data is used to generate features using a machine learning model that affects the outcome of a match. This is done by dividing the initial data into parts such as overall stats and recent stats to get a general idea of the performance of a player, if data for the player does not exist, it is filled by utilising the data for similar players (based on role factor) and how they fair in different matchups.

The ground factors, such as ground dimensions, spin or pace friendly and batting or bowling friendly, are found using a logistic regression model. The obtained features are then used with the player factors to come up with probabilities of various outcomes that might occur in a delivery (the outcomes include: ‘wicket’, ‘0’, ‘1’, ‘2’, ‘3’, ‘4’, ‘6’, ‘extras’).

**Pitch Features**

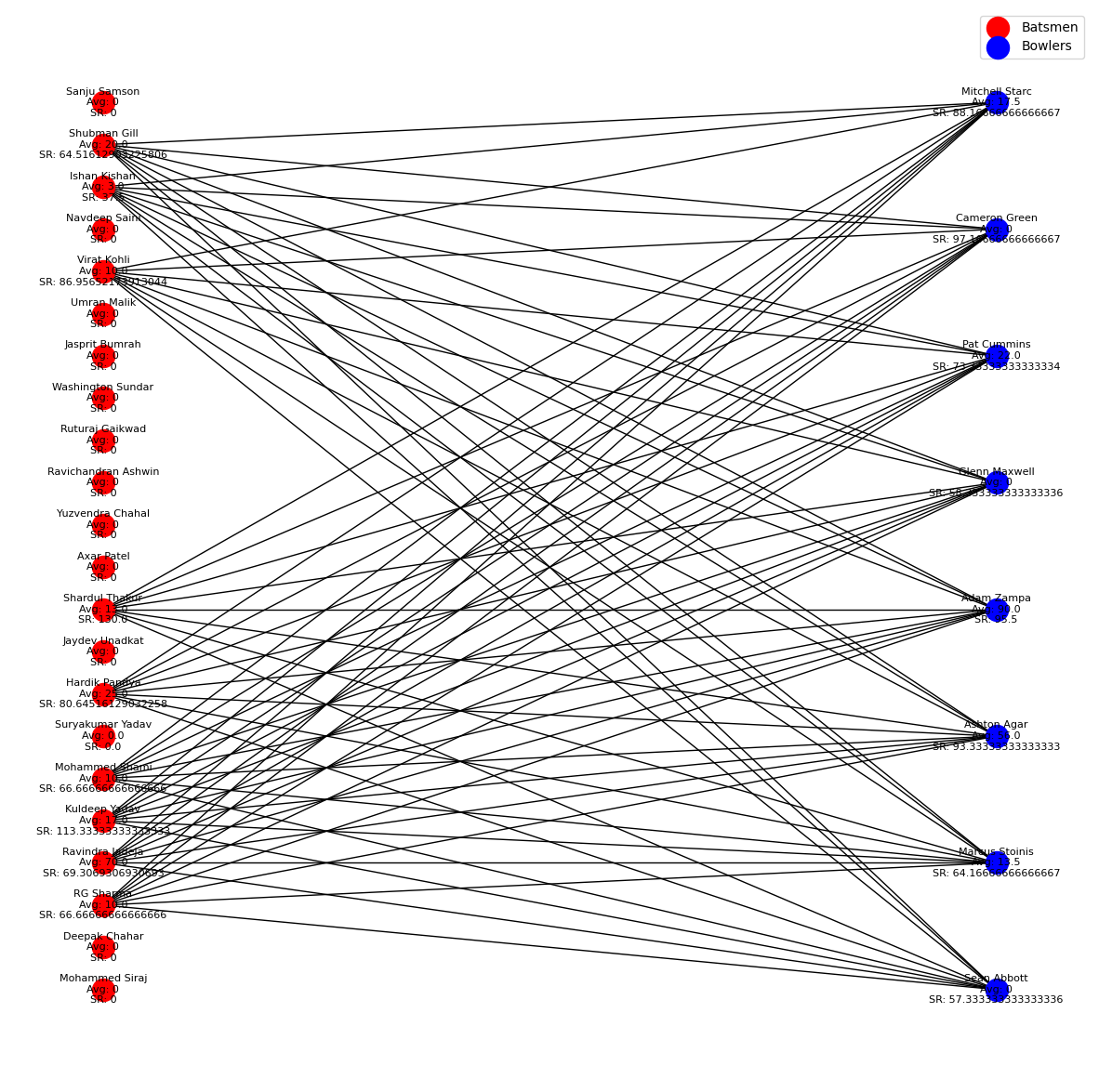




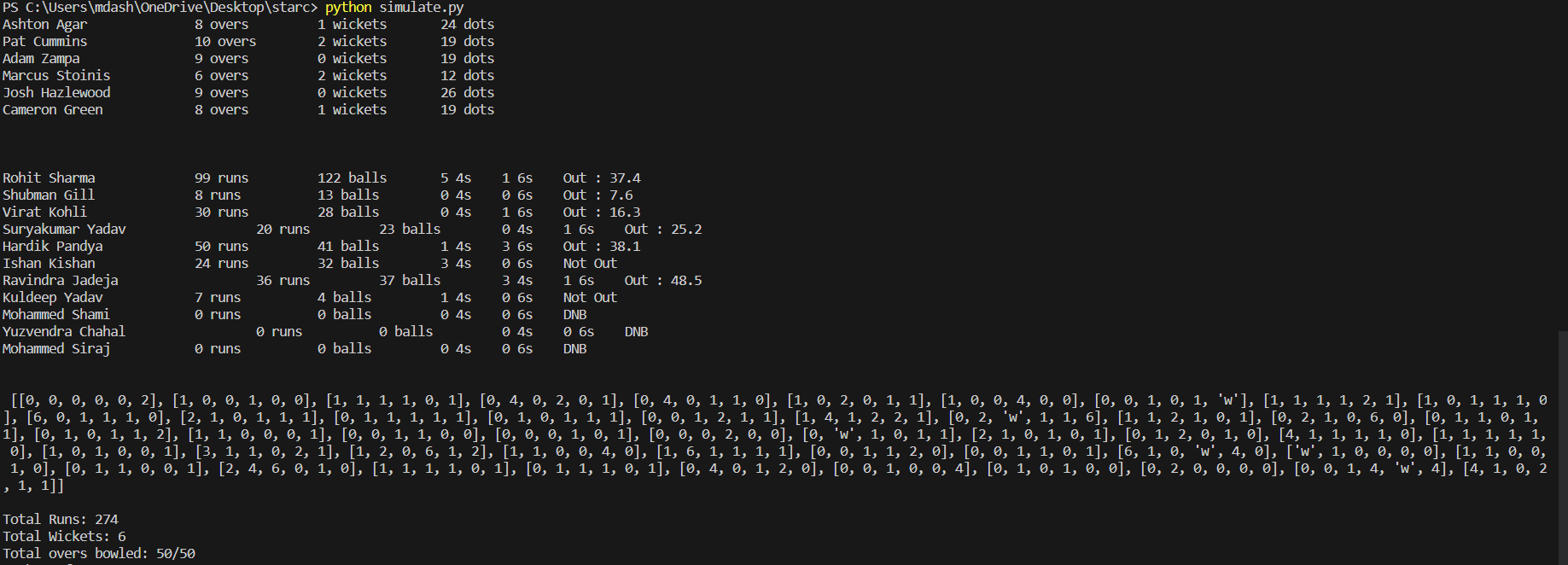
A Bipartite graph is used to store the different probabilities and features between players. The highest probabilities of the suitable features will be used to recommend the top players, role-wise, that are most adaptable to the conditions and can tackle various members of the opponent’s squad to be selected for the team allowing selectors to pick the best ‘Playing Eleven’ for the tournament or series.

The probabilities are further used to perform simulations of Team vs Team matchups. This is achieved using Python implementation to run ball by ball simulations of the match and give the average performance of each bowler and batter that played in the simulated matches.

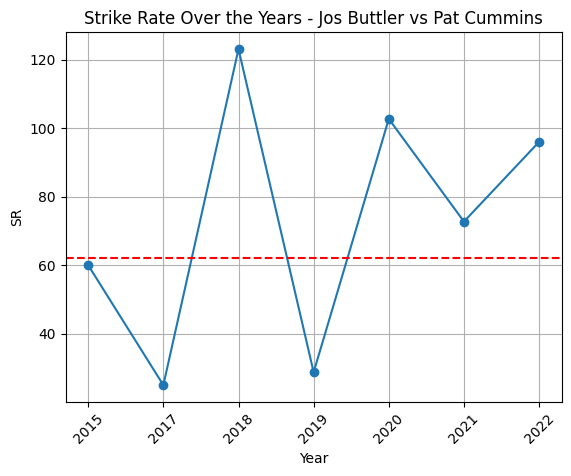
**Bipartite Graph**



**Simulation of an Innings between India and Australia**



These features will be available on a dashboard for all teams and players. It will allow the user to input the squad of a team and will give the recommended team that can be formed based on the multiple parameters used in our model. The dashboard will also allow the user to perform the statistical analysis of a player or perform different matchups of their choice. It can also be used to predict upcoming matches and therefore help in building teams for fantasy leagues or games. Here is an example of the Strike Rate of Jos Buttler against Pat Cummins over the years.



|  |  |  |  |
| --- | --- | --- | --- |
| **Sl. No.** | **Name** | **SRN** | **Individual Contribution** |
| 1) | P Sai Charan | PES2UG21CS364 | * Scrape individual players data including team, from ESPN Cricinfo and organised the data into batters and bowlers. * Design the dashboard for team vs team matchups using NextJS and TailwindCSS * Authentication using Firebase and real-time access to the player statistics (axios and cheerio) |
| 2) | Muhammad Ashar Reza | PES2UG21CS306 | * Scrape grounds at world cup venues data * Clean all the scraped data so it can be utilised. * Implementation of matchup analysis using parameters for team vs team in python using the player vs player analysis. * Implementation of simulator for various probable player vs player and team vs team interactions using python. |
| 3) | Pranit Prasant Pai | PES2UG21CS388 | * Perform graphical analysis and assign various parameters including pitch, form, etc. to each player using Tableau and Power I * Display player vs player matchups based on user input * Prove the working of the model by analysing the ground factors in the 2019 World Cup and confirming using machine learning algorithm. |
| 4) | Pratham R Shetty | PES2UG21CS392 | * Implementation of the logistical regression model in python for ground factors to be calculated. * Implementation of the model in python for the probabilities required. * Implementation of bipartite graph to store probabilities and other factors. * Prove the working of the model by predicting the 2019 World Cup using machine learning model. |

**Software employed:** Python libraries employing machine learning algorithms (Logistical regression), Flask, ReactJS, SQL, Tailwind CSS, Tableau

**Features on Dashboard:**

* Display detailed statistics of players– (overall, year-wise)
* Display ball to ball match details of all matches in ODI (can be extended to other formats)
* Player v player statistics
* Team v team statistics
* Team building to help with team management
* Predict upcoming matches
* Predict various player matchups
* Recommend players for Fantasy cricket games (such as helping in choosing the most valuable player)

Further Scope: Fill gaps in squads in Auctions (with Bid Price)

**Faculty Mentors:** Dr. Bharathi R and Dr. Sandesh B J