T20 World Cup 2022 : Building A Winning Team Through Data Analysis

By

Vrajesh Somani

221370680021

Internal Guide: Prof. Raj Hakani

Designation: Assistant Professor

In the subject of

Mini Project (1628004)

Post Graduate Diploma in Data Science (Semester 2)

June, 2023

Graduate School of Engineering and Technology, Gujarat Technological University

Nr.Vishwakarma Government Engineering College, Visat Three Roads, Visat - Gandhinagar Highway, Chandkheda, Ahmedabad - 382424, Gujarat

CERTIFICATE

This is to certify that project work embodied in this report titled "T20 World Cup 2022 : Building A Winning Team Through Data Analysis" was carried out by Vrajesh Somani (221370680021) at GTU – Graduate School of Engineering and Technology (Institute code - 137) for partial fulfilment of Mini Project – 1628004 in Post Graduate Diploma in Data Science to be awarded by Gujarat Technological University. This project work has been carried out under my guidance and supervision and it is up to my satisfaction.

Date:

Place: Ahmedabad

Signature and Name of Guide	**Signature and Name of Director**
Prof. Raj Hakani	**Prof. (Dr.) S. D. Panchal**
Assistant Professor	**Professor and Director**
GTU – GSET	**GTU – GSET**

TABLE OF CONTENTS

List Of Figures

List Of Tables

ABSTRACT

The T20 World Cup is a highly competitive cricket tournament where teams strive to build a winning combination to secure victory. In recent years, data analysis has emerged as a valuable tool for teams in their quest to create a successful team strategy.

This research report focuses on the use of data analysis in building a winning team for the T20 World Cup 2022. It explores the various data points and metrics that teams can analyze to evaluate player performance, identify key strengths and weaknesses, and make informed decisions regarding team selection and strategy.

Through the application of statistical analysis and advanced data techniques, this report delves into the significance of data-driven decision-making in team composition. It examines the impact of batting and bowling statistics, fielding efficiency, player form, and situational performance on team success.

Furthermore, the report investigates the utilization of machine learning algorithms to identify patterns, trends, and insights from historical data. It explores the use of predictive analytics to anticipate future performance and make proactive decisions regarding player selection, batting orders, bowling rotations, and tactical adjustments during matches.

By harnessing the power of data analysis, teams can gain a competitive edge in team selection, strategy formulation, and performance optimization. The report delves into the challenges and opportunities associated with data analysis, including data availability, quality, and interpretation, as well as the integration of subjective assessments and expert knowledge into the decision-making process.

Ultimately, this research report provides valuable insights into the application of data analysis in building a winning team for the T20 World Cup 2022. By leveraging data-driven insights, teams can make informed decisions, maximize player potential, and enhance their chances of success in this highly anticipated tournament.

1. Introduction

The T20 World Cup is a prestigious international cricket tournament that showcases the best teams from around the world competing in the thrilling T20 format. In the pursuit of victory, teams are increasingly turning to data analysis to gain a competitive edge in team selection, strategy formulation, and performance optimization.

This report focuses on the utilization of data analysis using Python and Power BI in building a winning team for the T20 World Cup 2022. The combination of Python's robust data analysis libraries and Power BI's powerful visualization capabilities provides teams with the tools to extract meaningful insights from complex cricket data and make informed decisions.

With Python, teams can leverage a wide range of libraries such as Pandas, NumPy to perform comprehensive data analysis. These libraries enable teams to explore batting and bowling statistics, analyze player performance metrics, and uncover patterns and trends in historical data. Python's versatility also allows for advanced statistical analysis and the development of machine learning models to predict player performance and optimize team strategies.

In conjunction with Python, Power BI offers a user-friendly platform to visualize and present the analyzed data in an intuitive and impactful manner. Teams can create interactive dashboards and reports, incorporating charts, graphs, and tables that enable easy interpretation of the data. Power BI's ability to connect with multiple data sources ensures that teams can integrate their Python-generated insights seamlessly into the visualization process.

By combining the analytical capabilities of Python and the visualization power of Power BI, teams can gain a comprehensive understanding of player performance, identify key strengths and weaknesses, and make data-driven decisions when building their winning team for the T20 World Cup 2022.

Throughout this report, we will explore the methodologies and techniques involved in utilizing Python and Power BI for data analysis in the context of the T20 World Cup. We will delve into the process of collecting and preparing cricket data, performing statistical analysis, developing predictive models, and visualizing the insights using Power BI.

By harnessing the potential of data analysis using Python and Power BI, teams can enhance their decision-making processes, optimize team performance, and increase their chances of success in the fast-paced and fiercely competitive T20 World Cup 2022.

2. Identification of Broad Area of Problem

Analysing player performance can be a critical component in creating the best team for a T20 Cricket World Cup tournament.

- Identify Key Players: Player analysis can help identify the most talented and consistent performers in the team, who are likely to make a significant impact on the tournament. These key players can be given specific roles and responsibilities, such as opening the batting, leading the bowling attack, or fielding in key positions.

- Optimize Strategies: Player analysis can help in developing specific strategies for different situations. For example, if a particular player has a strong record against spin bowling, the team management may choose to open the batting with a spin bowler to unsettle the opposition.

- Performance Improvement: Player analysis can help in identifying areas for improvement in a player's technique or fitness. By working on these areas, players can become more effective and consistent performers.

- Visualization and Insights Communication: Presenting the analyzed data and insights in a clear and visually appealing manner is essential for effective decision-making. Using Power BI, the challenge lies in creating interactive and intuitive dashboards, charts, and reports that allow stakeholders, coaches, and team management to easily interpret and understand the insights derived from the data analysis.

By addressing these problem areas, the project aims to utilize data analysis techniques, Python programming, and Power BI visualization to optimize team performance, make informed decisions, and increase the chances of building a winning team for the T20 World Cup 2022.

3. Dataset Description

To successfully create the project, you will need a comprehensive dataset that encompasses various aspects of player performance, match data, and team statistics. Here is a description of the dataset that would be ideal for this project:

3.1 Player Data

This dataset should include player profiles, including their names, ages, batting and bowling styles, playing roles, and career statistics such as batting average, strike rate, bowling average, economy rate, and fielding statistics.

3.2 Team Information

This dataset would include information about all the teams participating in the tournament, such as team name, player names, player statistics, and team rankings.

3.3 Match Schedule

This dataset would include information about the schedule of all the matches in the tournament, including the date, time, and location of each match.

3.4 Match Results

This dataset would include information about the results of each match, such as the winning team, the margin of victory, and the player of the match.

3.5 Venue Information

This dataset would include information about the different venues where the matches are being played, such as the name of the stadium, its location, seating capacity, and other relevant details.

By utilizing a well-curated and comprehensive dataset, you can perform meaningful data analysis, develop predictive models, and create informative visualizations that will aid in building a winning team strategy for the T20 World Cup 2022.

4. Data Pre-Processing

Pre-processing the web-scraped data from the ESPN site for project involves several steps to clean and prepare the data for analysis. Here are some key pre-processing steps:

4.1 Data Cleaning

Check for missing values, inconsistent formatting, and irrelevant information in the scraped data. Remove any unnecessary columns or rows that do not contribute to the analysis. Handle missing values by either imputing them or removing the respective data points, depending on the impact of missing data on the analysis.

4.2 Data Integration

If you have scraped data from multiple web pages, combine the data into a single dataset. Ensure that the columns are consistent across different web pages and that the data is properly aligned.

4.3 Data Transformation

Convert data types as needed. For example, convert text-based columns to numeric if they represent numerical values. Format dates and timestamps consistently for analysis. Normalize or scale the data if required to bring all variables to a similar scale.

4.4 Feature Engineering

Create new features or derive additional insights from the existing data. For example, calculate batting strike rate, bowling economy rate, or fielding efficiency based on the available data. These new features can provide valuable insights for analysis and decision-making.

4.5 Data Aggregation

Aggregate data at the appropriate level for analysis. For instance, calculate player statistics across matches or tournaments by grouping data based on player names or unique identifiers. Aggregate team-level statistics by combining individual player performances.

4.6 Handling Outliers

Identify and handle outliers, which are data points that deviate significantly from the rest of the data. Decide whether to remove outliers or transform them based on their impact on the analysis and the nature of the data.

4.7 Data Normalization

Normalize the data if necessary to eliminate any biases due to differences in scales or units. Normalization ensures that all variables are treated equally during analysis.

4.8 Data Formatting

Format the dataset in a structured manner, ensuring clear column headers and consistent naming conventions. This will facilitate easier data analysis and visualization using Python and Power BI.

By performing these pre-processing steps, you can ensure that the web-scraped data is clean, consistent, and ready for analysis. This will enable you to derive meaningful insights and build a winning team strategy for the T20 World Cup 2022 using Python and Power BI.

Figure 1 Openers Criteria

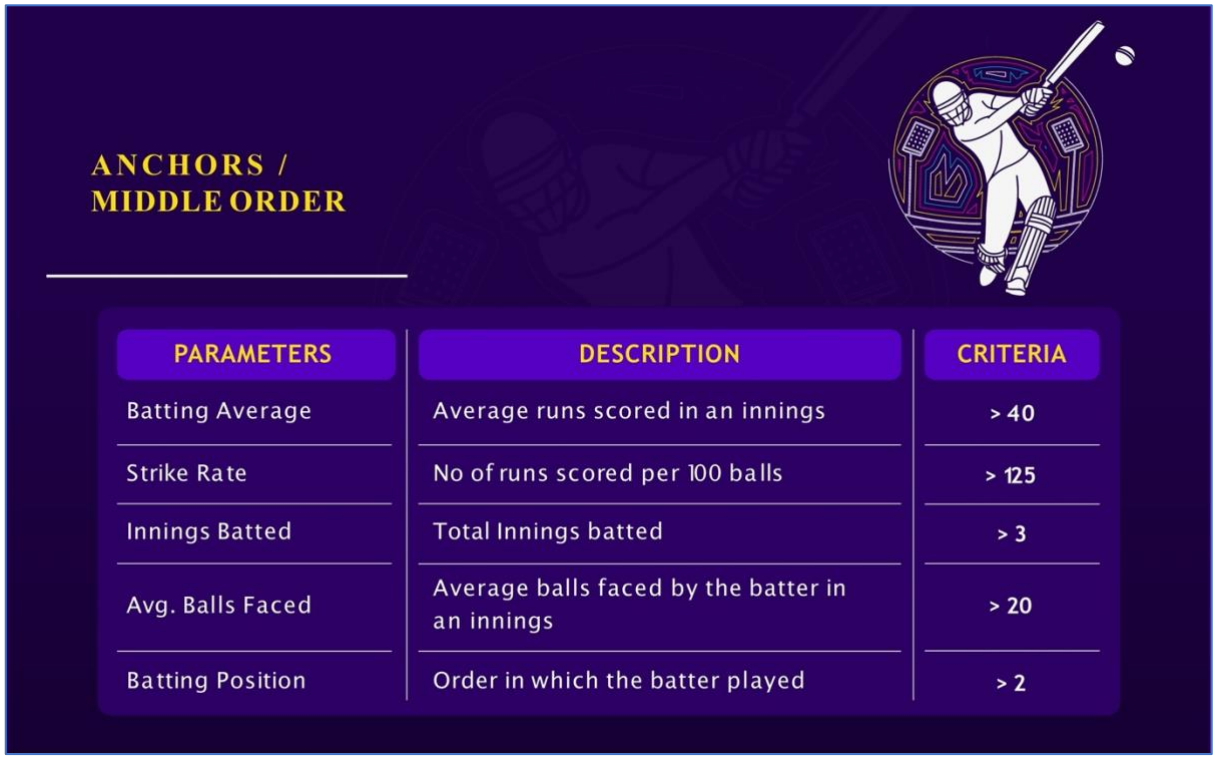
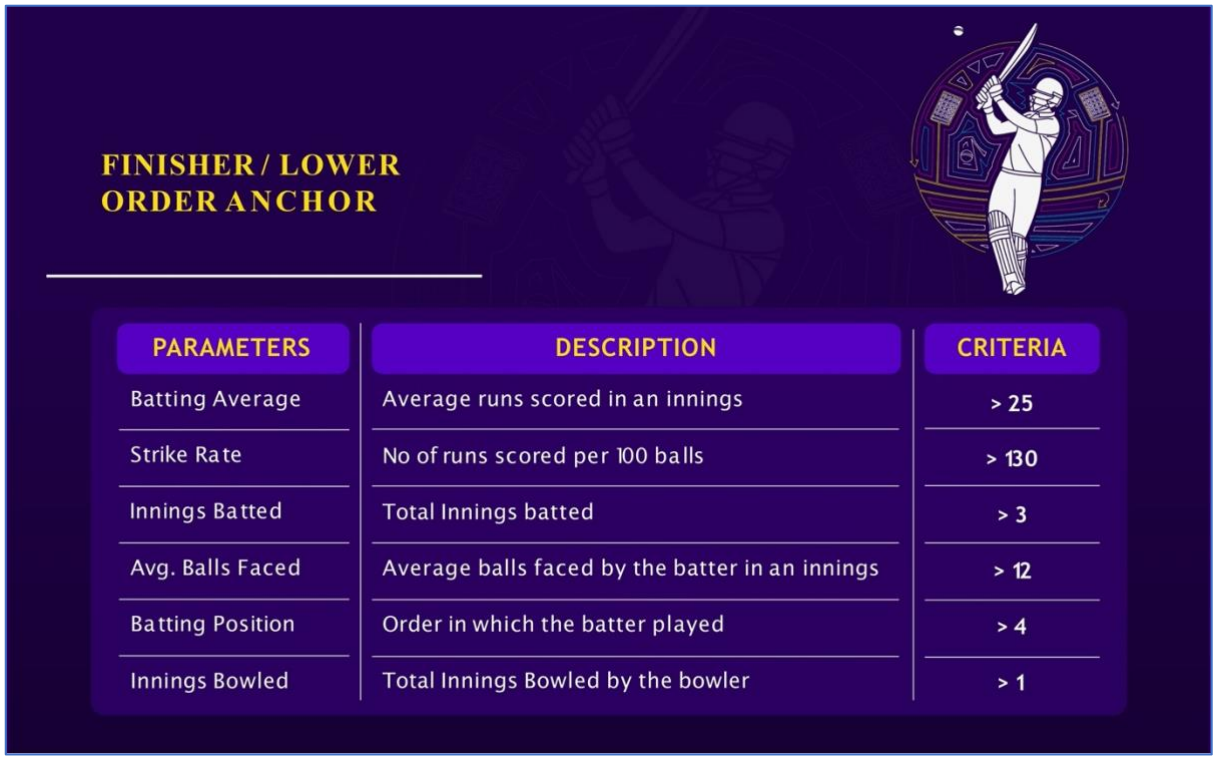
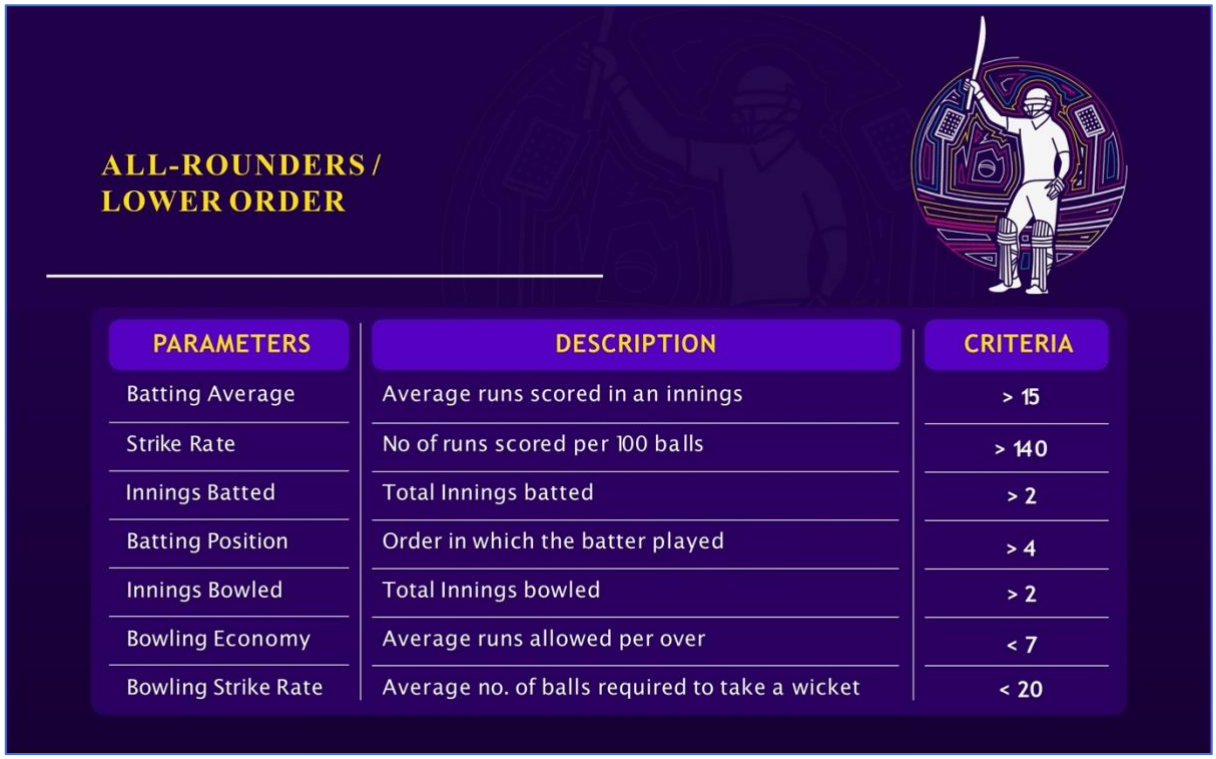


Figure 2 Middle Order Criteria



PARAMETERS	DESCRIPTION	CRITERIA
Batting Average	Average runs scored in an innings	> 25
Strike Rate	No of runs scored per 100 balls	> 130
Innings Batted	Total Innings batted	> 3
Avg. Balls Faced	Average balls faced by the batter in an innings	> 12
Batting Position	Order in which the batter played	> 4
Innings Bowled	Total Innings Bowled by the bowler	> 1

Figure 3 Finisher Criteria



PARAMETERS	DESCRIPTION	CRITERIA
Batting Average	Average runs scored in an innings	> 15
Strike Rate	No of runs scored per 100 balls	> 140
Innings Batted	Total Innings batted	> 2
Batting Position	Order in which the batter played	> 4
Innings Bowled	Total Innings bowled	> 2
Bowling Economy	Average runs allowed per over	< 7
Bowling Strike Rate	Average no. of balls required to take a wicket	< 20

Figure 4 All Rounders Criteria

Figure 5 Fast Bowlers Criteria

5. ML Model Development / Statistical Methods

5.1 Player Performance Analysis

Perform exploratory data analysis (EDA) to understand the distribution and patterns of player performance metrics such as batting average, strike rate, bowling average, and economy rate. Apply regression analysis to identify factors that significantly influence player performance, such as age, playing role, or batting position.

5.2 Match Strategy Analysis

Perform match scenario analysis to understand the impact of various strategies, such as batting order changes, fielding positions, or bowling tactics, on match outcomes. Apply simulation techniques to simulate match scenarios and assess the likelihood of success based on historical data and learned models.

5.3 Visualization and Dashboarding

Utilize Power BI to create interactive dashboards and reports to present the insights derived from the ML models and statistical analyses. Visualize player performance trends, team composition, and match strategy recommendations through informative charts, graphs, and tables.

By combining these ML and statistical approaches, you can gain valuable insights into player performance, optimize team composition, and make data-driven decisions to build a winning team.

Measures:

Sno	Measures	Description / Purpose	DAX FORMULA	TABLE
1	Total Runs	Total number of runs scored by the batsman	Total Runs = SUM(fact_batting_summary[runs])	fact_batting_summary
2	Total Innings Batted	Total number of innings a batsman got a chance to bat	Total Innings Batted = COUNT(fact_batting_summary[match_id])	fact_batting_summary
3	Total Innings Dismissed	To find the number of innings batsman got out	SUM(fact_batting_summary[out])	fact_batting_summary
4	**Batting Average**	**Average runs scored in an innings**	**Batting Avg = DIVIDE([Total Runs],[Total Innings Dismissed],0)**	**fact_batting_summary**
5	Total balls Faced	Total number of balls faced by the batsman	total balls faced = SUM(fact_batting_summary[balls])	fact_batting_summary
6	**Strike Rate**	**No of runs scored per 100 balls**	**Strike rate = DIVIDE([Total Runs],[total balls faced],0)*100**	**fact_batting_summary**
7	Batting Position	Batting position of a player	Batting Position = ROUNDUP(AVERAGE(fact_batting_summary[batting_pos]),0)	fact_batting_summary
8	Boundary %	Percentage of boundaries scored by the Batsman	Boundary % = DIVIDE(SUM(fact_batting_summary[Boundary runs]),[Total Runs],0)	fact_batting_summary
9	Avg. balls Faced	Average balls faced by the batter in an innings	AVERAGE(fact_batting_summary[balls])	fact_batting_summary
10	Wickets	Total number of wickets taken by a bowler	wickets = SUM(fact_bowling_summary[wickets])	fact_bowling_summary
11	balls Bowled	Total number of balls bowled by the bowler	balls Bowled = SUM(fact_bowling_summary[balls])	fact_bowling_summary
12	Runs Conceded	Total runs conceded by the bowler	Runs Conceded = SUM(fact_bowling_summary[runs])	fact_bowling_summary
13	**Bowling Economy**	**Average number of runs conceded in an over**	**Economy = DIVIDE([Runs Conceded], ([balls Bowled]/6),0)**	**fact_bowling_summary**
14	**Bowling Strike Rate**	**Number of balls bowled per wicket**	**Bowling Strike Rate = DIVIDE([balls Bowled], [wickets],0)**	**fact_bowling_summary**
15	Bowling Average	No. of runs allowed per wicket	Bowling Average = DIVIDE([Runs Conceded],[wickets],0)	fact_bowling_summary
16	Total Innings Bowled	Total number of innings bowled by a bowler	Total Innings Bowled = DISTINCTCOUNT(fact_bowling_summary[match_id])	fact_bowling_summary
17	Dot Ball %	Percentage of dot balls bowled by a bowler	Dot ball % = DIVIDE(SUM(fact_bowling_summary[zeros]), SUM(fact_bowling_summary[balls]),0)	fact_bowling_summary
18	Player Selection	To understand if a player is selected or not	Player Selection = if(ISFILTERED(dim_player[name]),"1","0")	
19	Display Text	To display a text of no 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.")	
20	Color Callout Value	To display a value only when a player is selected	Color Callout Value = if([Player Selection]="0", "#D0CF1D","#1D1D2E")	

Table 5.1 DAX Measures

```
bat = pd.read_csv("t20_csv_files/fact_bating_summary.csv") #read csv file
run1 = bat.groupby('batsmanName')['runs'].sum() #groupby batsman and calculate total runs
run1 = run1.sort_values(ascending=False).head(11) # sort the runs in descending order

plt.rcParams['figure.figsize']=(13,6) #graph size
plt.rcParams['xtick.major.pad'] = 7 #batsman name away from x-axis
fig,ax = plt.subplots() #subplot

plt.bar(run1.index, run1.values, color = colors)# create the bar chart
plt.xlabel('Batsman Name') #set the x-axis label
plt.xticks(rotation=90)

ax.spines['top'].set_visible(False) # to remove top border
ax.spines['right'].set_visible(False) # to remove right border

for patch,color,lb in zip(ax.patches,colors,ax.get_xticklabels()):
    x=patch.get_x()
    y=patch.get_height()
    a=y + 11
    ax.text(x,a,y,color=color,rotation=0,fontsize=14,fontweight='bold')

plt.ylabel('Total Runs')# set the y-axis label
plt.title('Top 11 Players with the Highest Runs')# set the chart title
plt.show()# show the plot
```

Figure 6 Code For Top 11 Batsman

```
bat = pd.read_csv("t20_csv_files/fact_bating_summary.csv") #read csv file
run2 = bat.groupby('teamInnings')['runs'].sum() #groupby team and calculate total runs
run2 = run2.sort_values(ascending=False)# sort the runs in descending order

plt.rcParams['figure.figsize']=(13,6) #graph size
plt.rcParams['xtick.major.pad'] = 7 #batsman name away from x-axis
fig,ax = plt.subplots() #subplot

plt.bar(run2.index, run2.values, color = colors) # create the bar chart
plt.xlabel('Batting Team') #set the x-axis label
plt.xticks(rotation=45)

ax.spines['top'].set_visible(False) # to remove top border
ax.spines['right'].set_visible(False) # to remove right border

for patch,color,lb in zip(ax.patches,colors,ax.get_xticklabels()):
    x=patch.get_x()
    y=patch.get_height()
    a=y + 29
    ax.text(x,a,y,color=color,rotation=0,fontsize=14,fontweight='bold')

plt.ylabel('Total Runs') # set the y-axis label
plt.title('Total Runs by Batting Team') # set the chart title
plt.show() # show the plot
```

Figure 7 Code For Runs Scored By Team

```
bowl = pd.read_csv("t20_csv_files/fact_bowling_summary.csv") #read csv file
wicket1 = bowl.groupby('bowlerName')['wickets'].sum() #groupby bowler and calculate total wickets
wicket1 = wicket1.sort_values(ascending=False).head(11)# sort the wickets in descending order and select top 11

plt.rcParams['figure.figsize']=(13,6) #graph size
plt.rcParams['xtick.major.pad'] = 7 #bowler name away from x-axis
fig,ax = plt.subplots() #subplot

plt.bar(wicket1.index, wicket1.values, color = colors)# create the bar chart
plt.xlabel('Bowler Name') #set the x-axis label
plt.xticks(rotation=90)

ax.spines['top'].set_visible(False) # to remove top border
ax.spines['right'].set_visible(False) # to remove right border

for patch,color,lb in zip(ax.patches,colors,ax.get_xticklabels()):
    x=patch.get_x()
    y=patch.get_height()
    a=y + 0.4
    ax.text(x,a,y,color=color,rotation=0,fontsize=14,fontweight='bold')

plt.ylabel('Total Wickets')# set the y-axis label
plt.title('Top 11 Players with the Highest Wickets')# set the chart title
plt.show()# show the plot
```

Figure 8 Code For Top 11 Bowlers

```
bowl = pd.read_csv("t20_csv_files/fact_bowling_summary.csv") #read csv file
wicket2 = bowl.groupby('bowlingTeam')['wickets'].sum() #groupby team and calculate total wickets
wicket2 = wicket2.sort_values(ascending=False)# sort the wickets in descending order

plt.rcParams['figure.figsize']=(13,6) #graph size
plt.rcParams['xtick.major.pad'] = 7 #team name away from x-axis
fig,ax = plt.subplots() #subplot

plt.bar(wicket2.index, wicket2.values, color = colors)# create the bar chart
plt.xlabel('Bowling Team')#set the x-axis label
plt.xticks(rotation=45)

ax.spines['top'].set_visible(False) # to remove top border
ax.spines['right'].set_visible(False) # to remove right border

for patch,color,lb in zip(ax.patches,colors,ax.get_xticklabels()):
    x=patch.get_x()
    y=patch.get_height()
    a=y + 1.5
    ax.text(x,a,y,color=color,rotation=0,fontsize=14,fontweight='bold')

plt.ylabel('Total Wickets')# set the y-axis label
plt.title('Total Wickets by Bowling Team')# set the chart title
plt.show()# show the plot
```

Figure 9 Code For Wickets Taken By Team

6. Parameters Tuning and Importance / Power BI Dashboard

Parameter tuning refers to the process of selecting the optimal values for the parameters of a machine learning algorithm or model. These parameters control the behaviour and performance of the algorithm, and tuning them can significantly impact the results obtained.

6.1 Performance Optimization

Parameter tuning can improve the performance of a machine learning model. Different parameter settings can lead to different levels of accuracy, precision, recall, or other performance metrics.

6.2 Overfitting and Underfitting Prevention

Parameters in machine learning algorithms often control the complexity or flexibility of the model. If the parameters are not properly set, the model may overfit or underfit the data. Overfitting occurs when the model becomes too complex and captures noise or irrelevant patterns in the training data, leading to poor generalization on new data. Underfitting, on the other hand, happens when the model is too simplistic to capture the underlying patterns in the data.

6.3 Grid Search and Random Search

Grid search and random search are commonly used techniques for hyperparameter tuning. Grid search exhaustively searches through a predefined set of hyperparameters, while random search randomly samples hyperparameters from specified distributions. These methods help in finding the best combination of hyperparameters that maximize model performance.

6.4 Importance of Parameter Tuning

Parameter tuning helps improve the model's predictive accuracy and generalization by optimizing the hyperparameters. It ensures that the model is not underfitting or overfitting the data. Optimal parameter tuning can lead to better model performance, higher accuracy, and more reliable predictions for team selection, match strategy, and player performance assessment.

Power BI offers powerful visualization capabilities and features that can enhance the analysis and presentation of project. Here are some key Power BI visualization dashboards and DAX (Data Analysis Expressions) features that one can utilize:

6.5 Interactive Dashboards

Create interactive dashboards to provide an overview of team performance, player statistics, and match outcomes. Use visualizations such as line charts, bar charts, or heat maps to display trends and comparisons.

Incorporate slicers and filters to allow users to dynamically explore the data and drill down into specific aspects, such as individual player performance or team rankings.

6.6 Player Performance Visualizations

Use visualizations like scatter plots or radar charts to compare player performance across multiple metrics, such as batting average, strike rate, and bowling economy rate. Add tooltips to provide detailed information on each data point.

Implement KPI (Key Performance Indicator) visualizations to track and highlight player performance against predefined targets or benchmarks.

6.7 DAX Calculations and Measures

Utilize DAX functions to create calculated columns and measures that perform complex calculations and aggregations. For example, calculate batting average, strike rate, or net run rate using DAX formulas.

Implement time intelligence functions to enable period-over-period comparisons, or rolling averages for player or team performance metrics.

7. Data and Results Visualization

7.1 Exploratory Data Analysis (EDA)

Visualizing the data through various charts, graphs, and plots helps in understanding the underlying patterns, trends, and distributions. EDA allows you to identify outliers, explore relationships between variables, and gain insights into the data. Technique such as histograms can be used to visualize different aspects of the data.

7.2 Player Performance Visualization

Create visualizations that showcase player performance metrics such as batting averages, strike rates, bowling averages, and economy rates. Visualizations like bar charts, line plots, or radar charts can be used to compare and rank players based on their performance. These visualizations help in identifying top-performing players, assessing strengths and weaknesses, and making informed decisions regarding team composition.

7.3 Power BI Dashboarding

Utilize Power BI to create interactive dashboards and reports that consolidate all the relevant visualizations and insights. Power BI allows you to connect to your data sources, integrate different visualizations, and create dynamic filters and slicers. This enables stakeholders, coaches, and team management to explore the data, gain actionable insights, and make informed decisions.

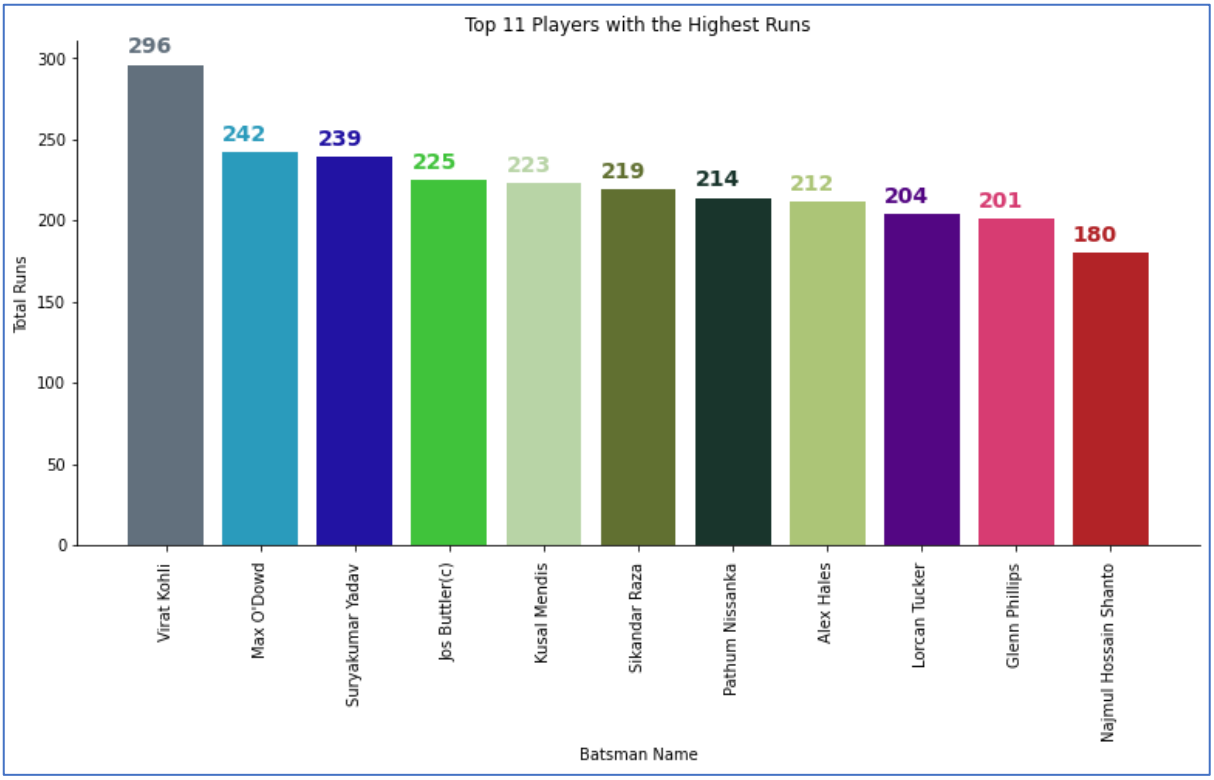


Figure 10 Top 11 Batsman

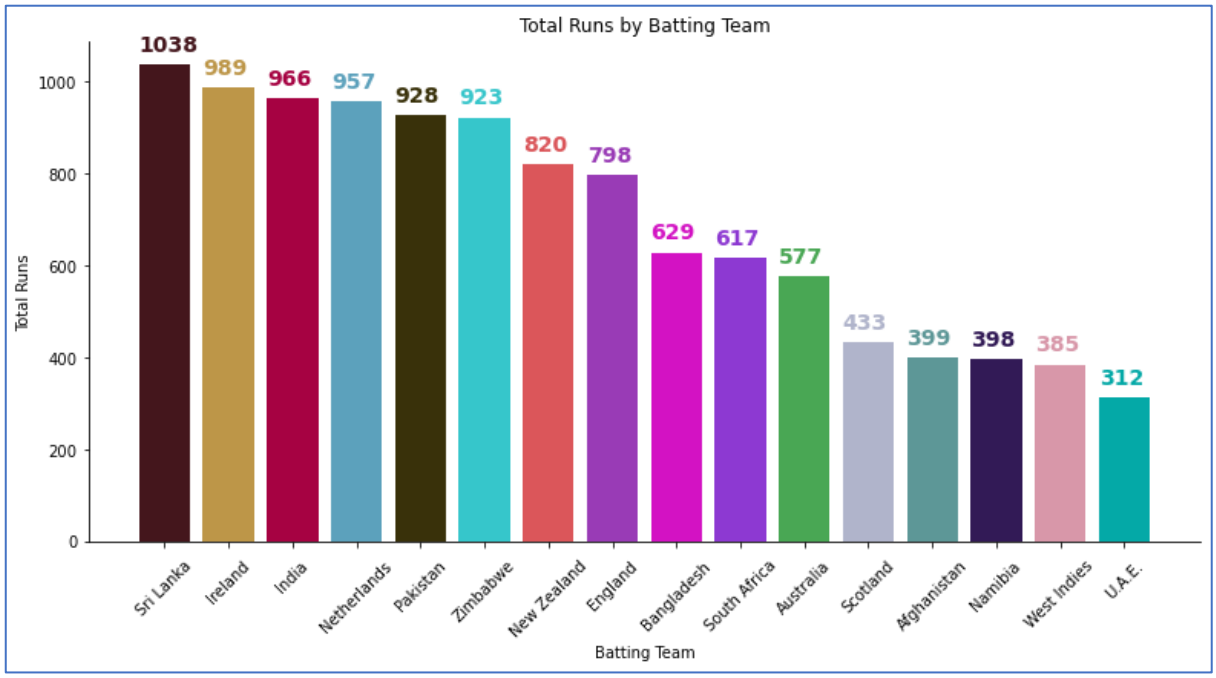


Figure 11 Runs Scored By Team

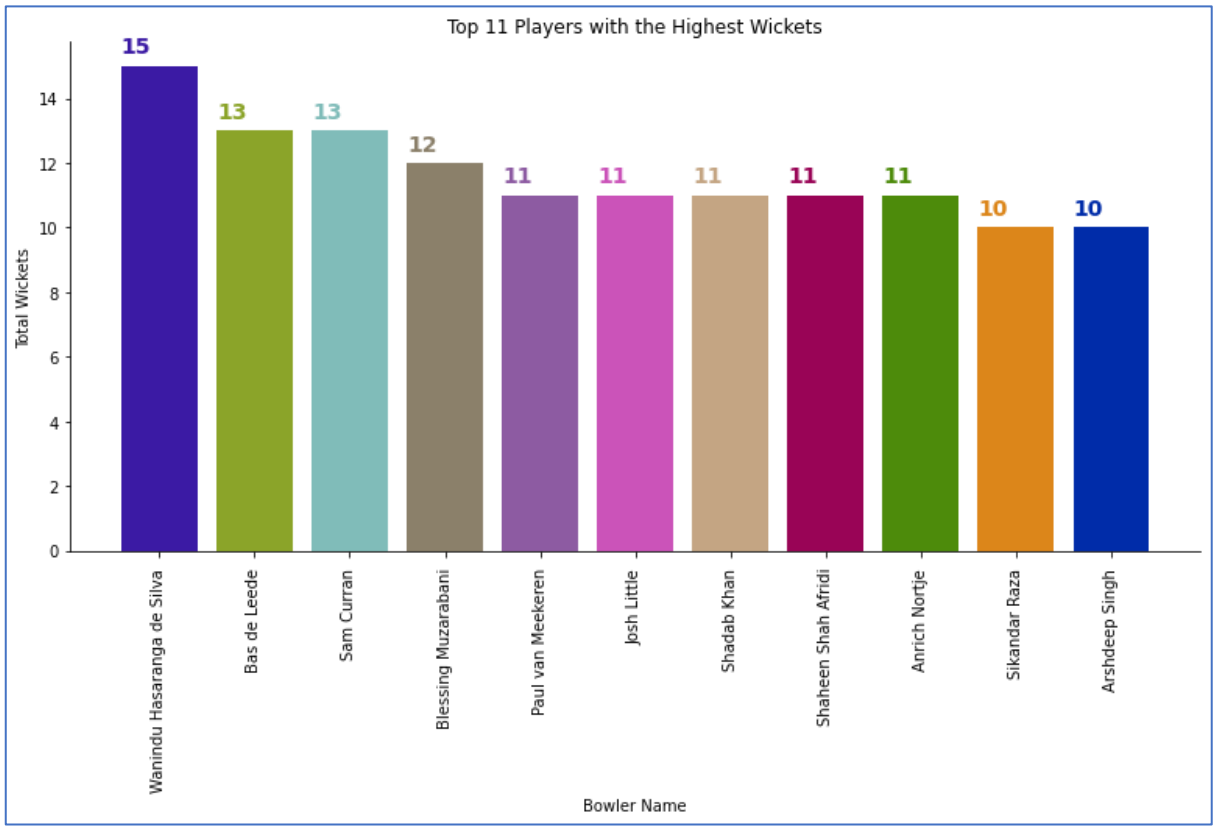


Figure 12 Code Top 11 Bowlers

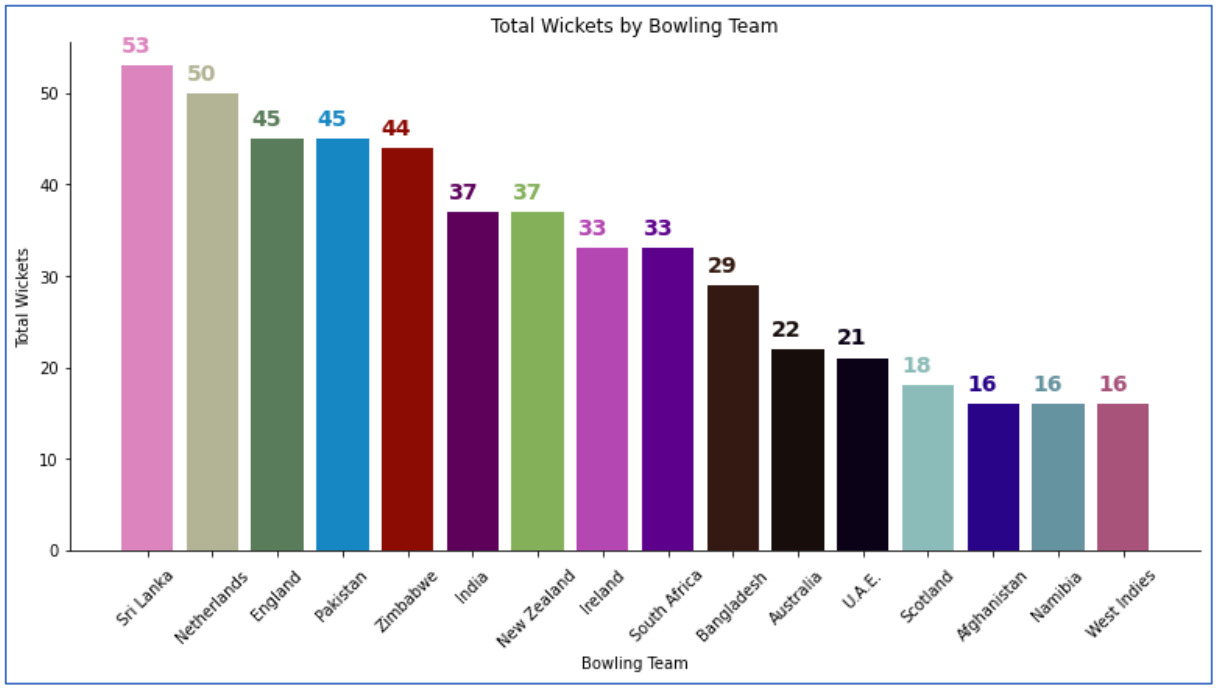


Figure 13 Wickets Taken By Team

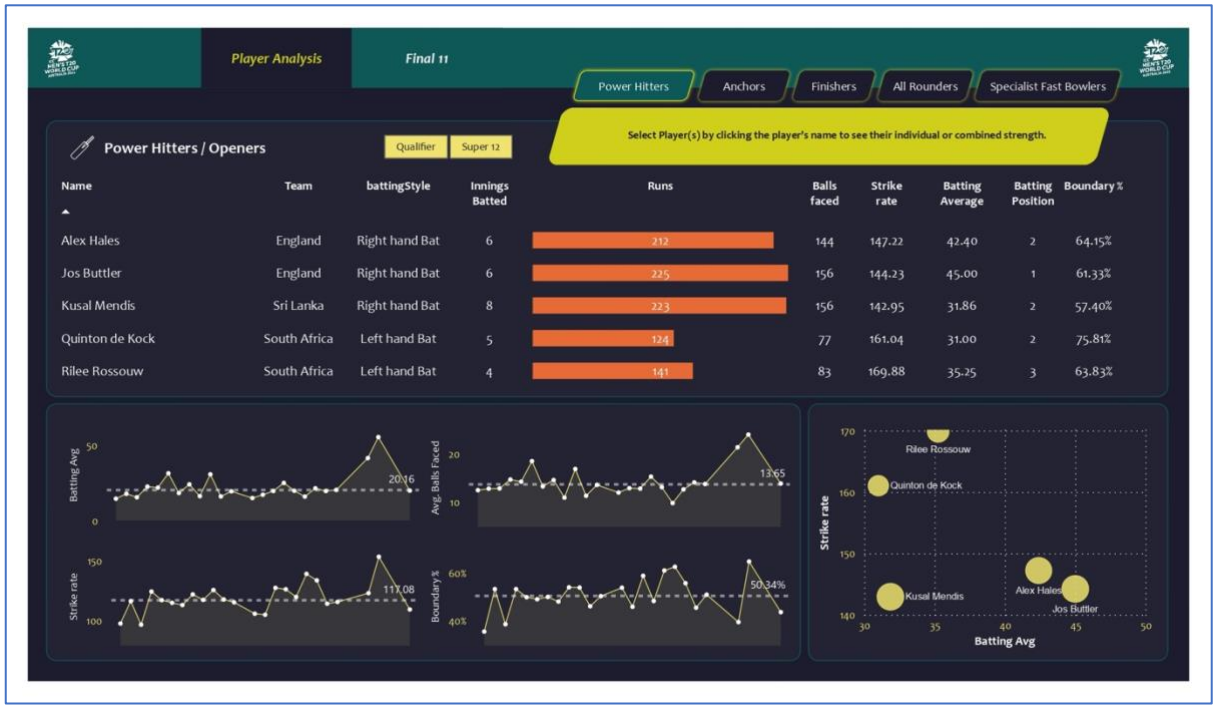


Figure 14 Power Hitters

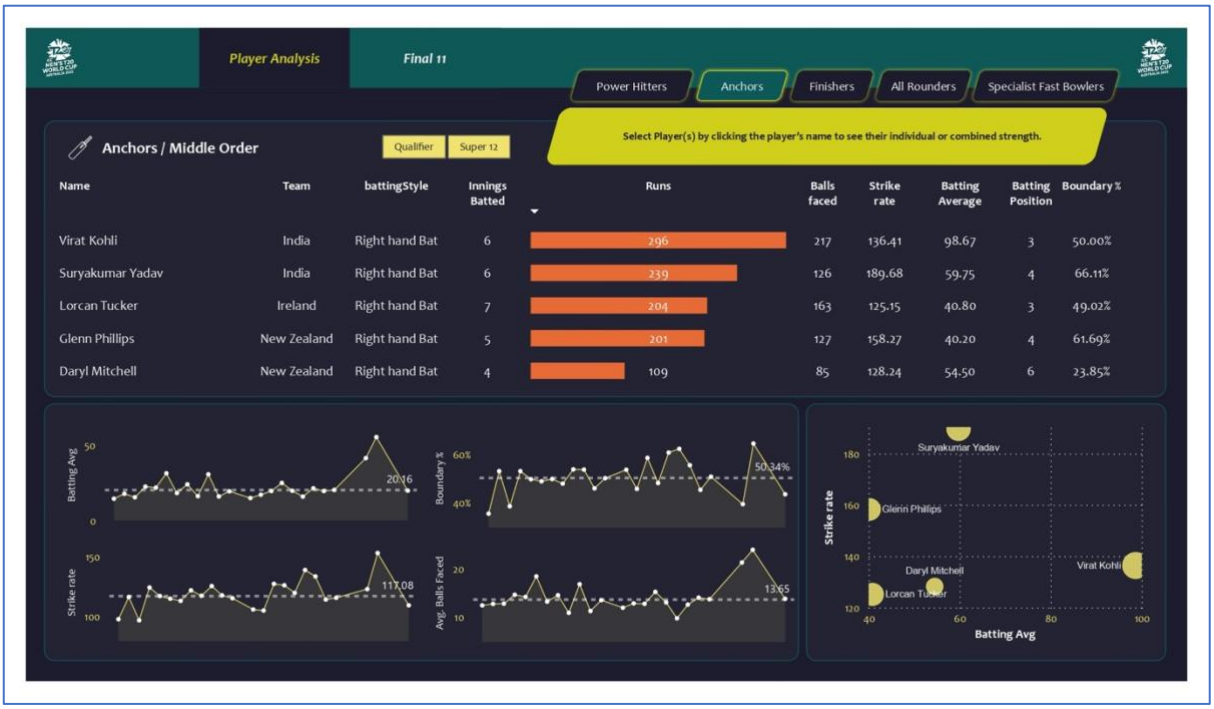


Figure 15 Anchors

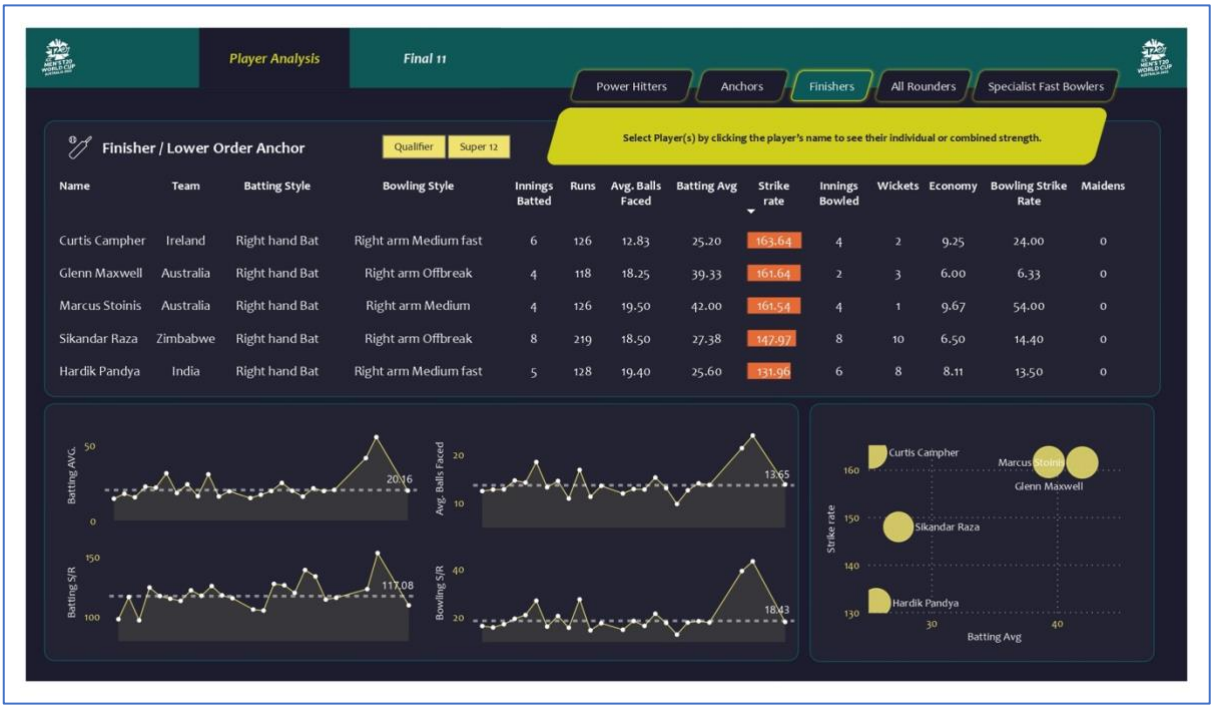


Figure 16 Finishers

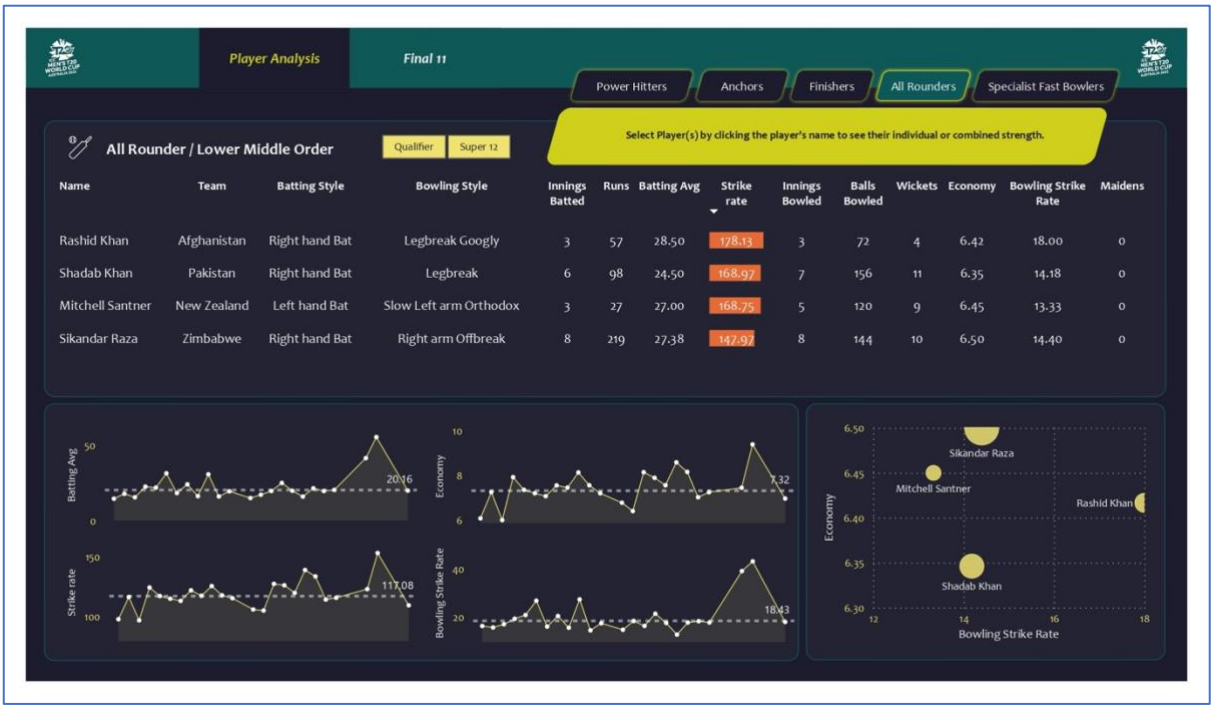


Figure 17 All Rounders

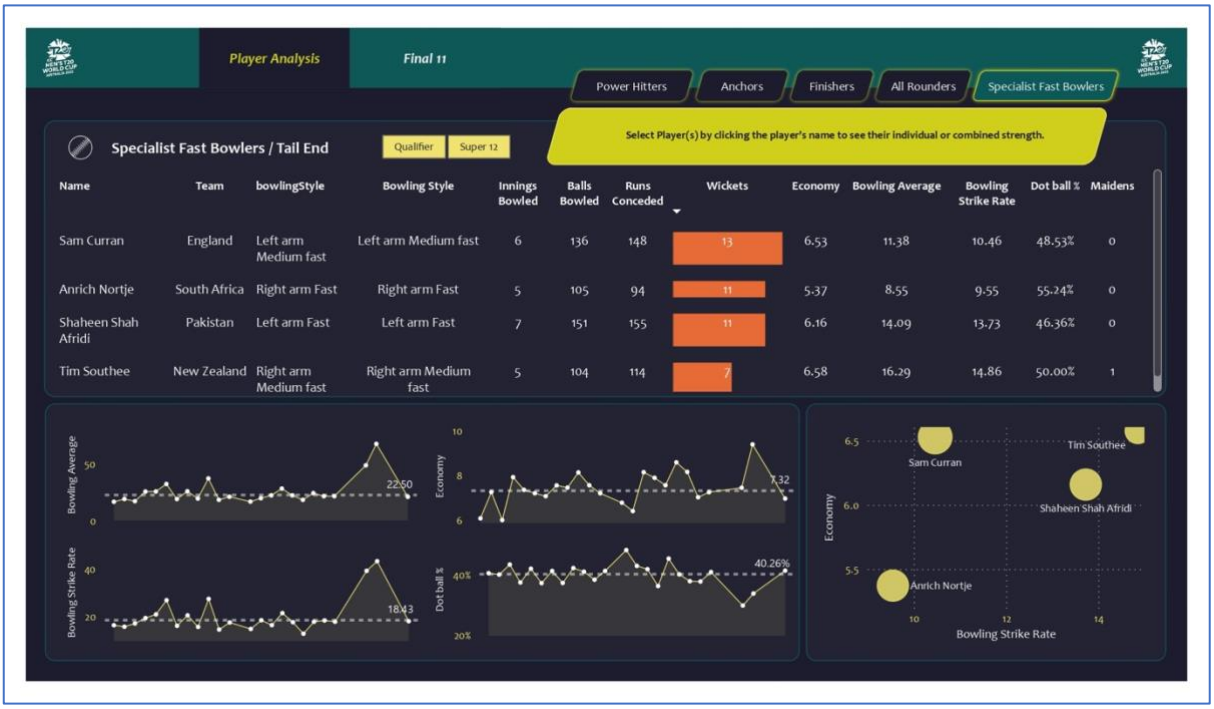


Figure 18 Fast Bowlers

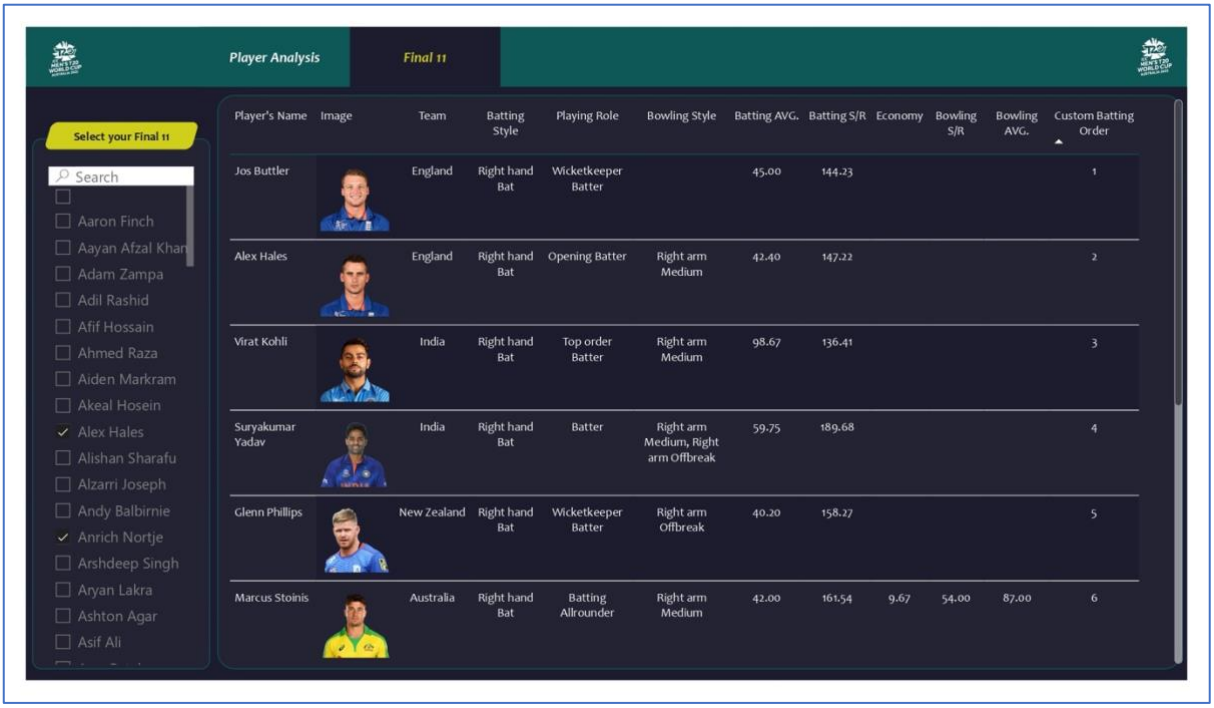


Figure 19 Final 11

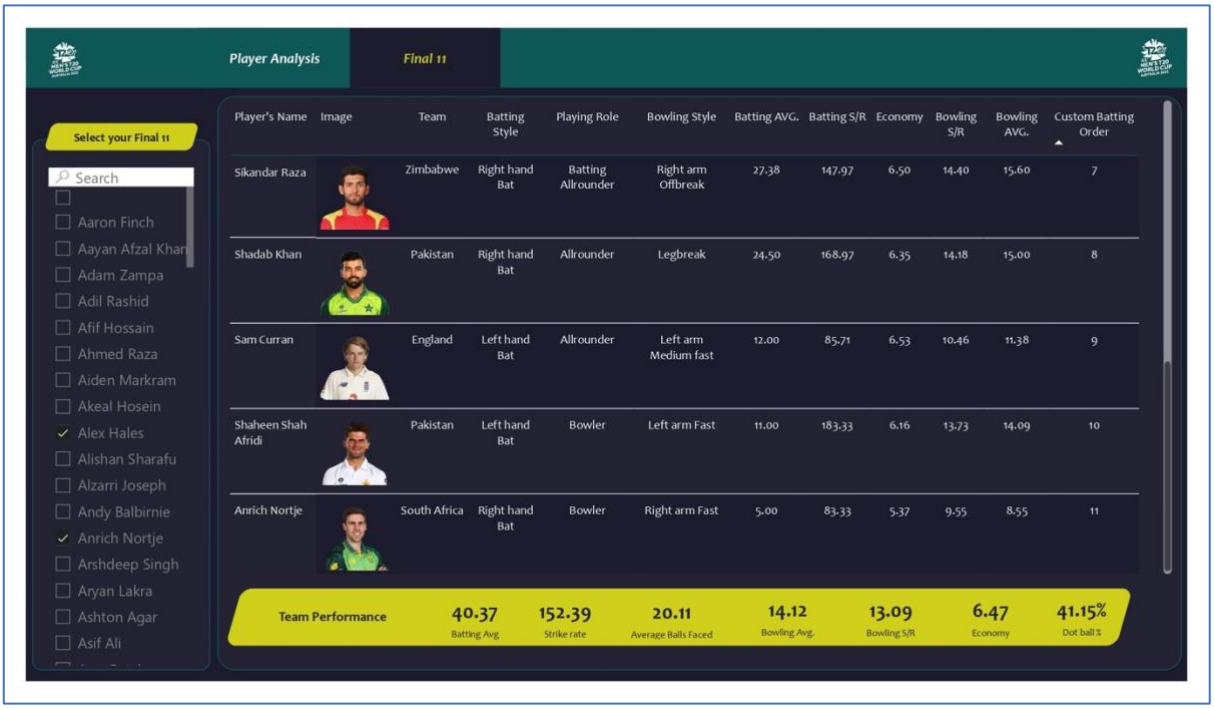


Figure 20 Final 11 (cont.)

7.4 Results

By comparing your predicted team with the ICC Team of the Tournament, you can gain valuable insights into the accuracy of your predictions and the performance of players in the T20 World Cup 2022. This comparison helps you evaluate your analysis methods, refine your player selection criteria, and enhance your ability to build winning teams in future cricket tournaments.

Position	Name - Team
1	Jos Buttler (England)
2	Alex Hales (England)
3	Virat Kohli (India)
4	Suryakumar Yadav (India)
5	Glenn Philips (New Zealand)
6	Marcus Stoinis (Australia)
7	Sikandar Raza (Zimbawe)
8	Shadab Khan (Pakistan)
9	Sam Curran (England)
10	Anrich Nortje (South Africa)
11	Shaheen Shah Afridi (Pakistan)

Table 7.1 Predicted Team Using Analysis

Figure 21 Upstox Team Of The Tournament

8. Usefulness and Limitation of Work

8.1 Usefulness of the Work

The work presented in this project holds several potential benefits and usefulness in the context of Cricket. Some of the key areas where this work can be valuable are:

- Talent Identification and Development: By analysing players' performance in the Cricket World Cup, it becomes easier to identify young talents with exceptional skills. This information can guide talent scouts and coaches in nurturing and developing promising players, thereby ensuring a strong pipeline of talent for the future. This contributes to the growth and sustainability of the sport and provides opportunities for aspiring cricketers to excel.

- Performance Evaluation: Analysing player performance in the Cricket World Cup can provide valuable insights into individual players' skills, strengths, weaknesses, and overall contributions to the team. This analysis can help selectors, coaches, and cricket enthusiasts make informed decisions about player selection, team strategies, and future talent development. It ensures that the best players are identified and given opportunities, leading to improved team performance.

- Fan Engagement and Experience: Cricket is immensely popular, and fans are highly invested in the performances of their favourite players and teams. Analysing player data and providing meaningful insights enhances the overall fan experience. It allows fans to understand the nuances of the game better, appreciate the skills of individual players, and engage in informed discussions and debates. It also enables cricket enthusiasts to make predictions, participate in fantasy leagues.

- Research and Innovation: Cricket data analysis opens up avenues for research and innovation in sports analytics. Researchers, statisticians, and data scientists can explore novel approaches, develop predictive models, and uncover new patterns and trends in player performance.

- Sports Science: Analysing players' performance data can help sports scientists and analysts develop new and innovative methods for training and improving athletes' performance. This can have wider applications beyond cricket and benefit other sports and athletes as well.

8.2 Limitations of the Work

While this project offers potential benefits, it is important to acknowledge and consider the limitations that may impact the findings and conclusions. Here are some potential limitations to be aware of:

Data Quality and Availability: The quality and availability of the web-scraped data from the ESPN site may vary, potentially leading to inconsistencies, missing values, or inaccuracies. The reliability and completeness of the data can impact the accuracy and reliability of the analysis.

External Factors: The analysis may not account for external factors that can significantly influence player and team performance, such as injuries, team dynamics, or changes in playing conditions. These factors can have a significant impact on the outcomes and may not be fully captured by the available data.

Assumptions and Simplifications: The analysis and models developed for team selection, match strategy, and performance prediction are based on certain assumptions and simplifications. These assumptions may not fully reflect the complexity and nuances of cricket, potentially leading to limitations in the accuracy and applicability of the results.

Generalizability: The findings and recommendations derived from this project may be specific to the T20 World Cup 2022 dataset and may not be directly applicable to other tournaments, formats, or cricket leagues. Caution should be exercised when generalizing the results beyond the scope of the analyzed dataset.

Future Performance Uncertainty: While data analysis and modeling techniques can provide insights based on historical data, they cannot guarantee future performance. The dynamic nature of cricket and the uncertainties associated with player form, injuries, and other factors make it challenging to accurately predict future outcomes.

It is crucial to recognize these limitations and communicate them transparently in the project report. This allows readers and stakeholders to interpret the results appropriately and make informed decisions based on the strengths and weaknesses of the analysis.

9. Timeline Chart

This timeline chart outlines the major milestones and tasks completed throughout the course of the project:

9.1 Data Collection and Exploration (Week 1-2):

- Gathered data from reliable sources.
- Explored the dataset to understand its structure, variables, and any data pre-processing requirements.

9.2 Data Pre-Processing (Week 3-4):

- Performed data cleaning to handle missing values, outliers, and inconsistencies.
- Conducted feature engineering to extract relevant features for model development.
- Applied scaling and normalization techniques to ensure consistent data representation.

9.3 ML Model Development / Statistical Methods (Week 5-7):

- Developed and implemented the chosen model using different Python libraries.
- Evaluated the model's performance using appropriate evaluation metrics.

9.4 Parameters Tuning and Importance / Power BI Dashboard (Week 8-9):

- Conducted parameter tuning to optimize the model's performance.
- Assessed the importance of different features in the model using feature importance techniques.
- Utilize DAX functions to create calculated columns and measures that perform complex calculations and aggregations. For example, calculate batting average, strike rate, or net run rate using DAX formulas.

9.5 Data and Results Visualization (Week 10-11):

- Visualized the data to identify trends and patterns.
- Power BI Dashboard building.
- Created visualizations to showcase the model's performance and highlight any discrepancies or limitations.

9.6 Usefulness and Limitation of Work (Week 12-13):

- Evaluated the usefulness and potential applications of the analysis of players and teams which can help sports scientists and analysts develop new and innovative methods for training and improving athletes' performance.
- Discussed the limitations and challenges associated with the model and its predictions.

9.7 Report and Presentation Preparation (Week 14-15):

- Compiled all the findings, analyses, and visualizations into a comprehensive final report.
- Created a clear and concise presentation to communicate the project's objectives, methodology, and results.
- Prepared for presentation to share the project outcomes, insights, and recommendations.

9.8 Final Review (Week 16):

- Completed the timeline chart and incorporated it into the project report.
- Conducted a final review of the project report to ensure all sections are complete and cohesive.
- Made any necessary revisions or additions based on feedback and recommendations.

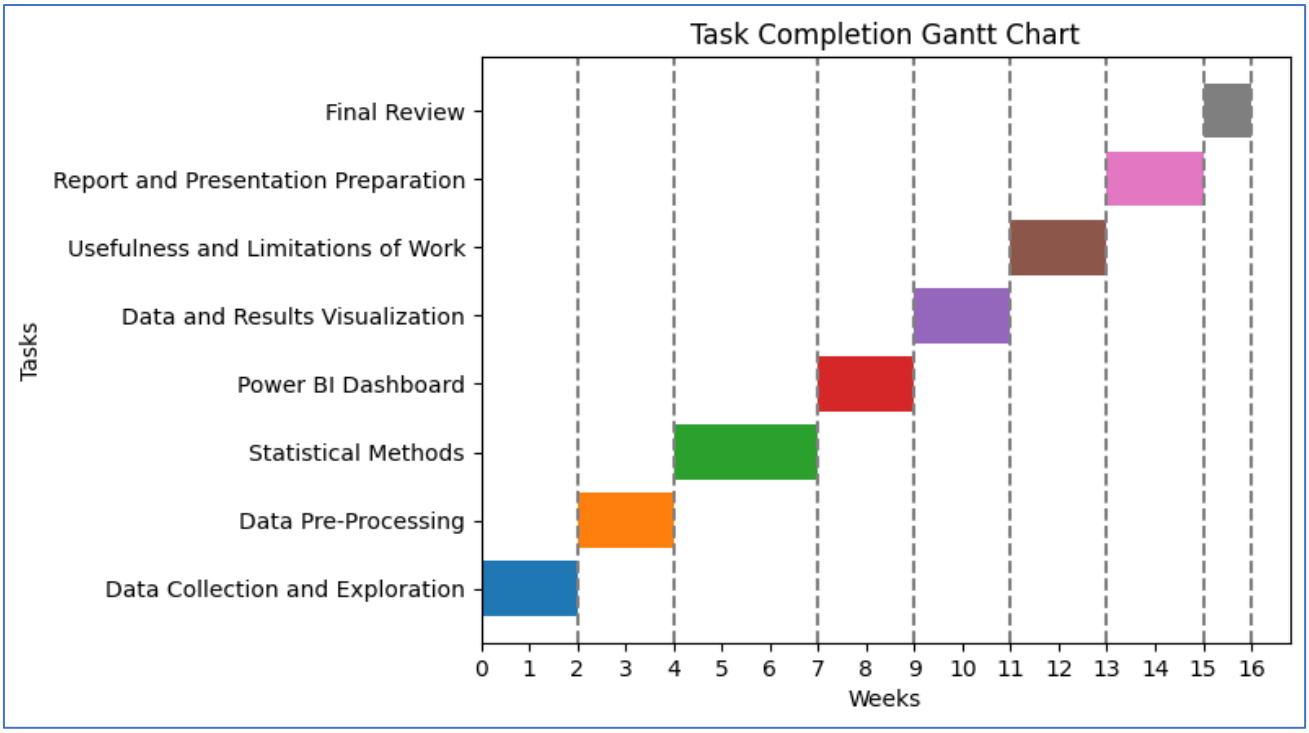


Figure 22 Timeline Chart

The timeline chart provides a visual representation of the project's progression and highlights the key activities completed within each phase. It demonstrates the sequential nature of the project and the time allocated to different tasks.

Bibliography and References

Bibliography:

1. Chander, S., & Vamsi, P. (2021). Cricket Match Outcome Prediction Using Machine Learning Algorithms. International Journal of Computer Sciences and Engineering, 9(5), 495-502.

2. Dutta, S., Mukherjee, S., & Roy, R. (2019). Performance Analysis of Cricketers Using Data Mining Techniques. International Journal of Data Science and Analysis, 5(2), 30-37.

3. Power BI documentation. Retrieved from https://docs.microsoft.com/en-us/power-bi/

4. Python documentation. Retrieved from https://docs.python.org/

References:

1. ESPN – Cricinfo
   https://www.espncricinfo.com/series/icc-men-s-t20-world-cup-2022-23-1298134
   For finding the data of T20 World Cup 2022

2. Bright Data
   https://brightdata.com/cp/start
   Web scrapping the data

3. Gupta, A., & Goyal, N. (2021). Sports analytics: Data-driven decision-making in cricket. Decision, 48(2), 121-140.

4. ChatGPT
   https://chat.openai.com/chat
   For help related coding and description.

5. Github
   https://github.com/Vrajesh2/2022-World-Cup-Analysis
   To find all the project related files