MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE, GWALIOR (M.P.), INDIA

DEEMED UNIVERSITY
(Declared under Distinct Category by Ministry of Education, Government of India)
NAAC ACCREDITED WITH A++ GRADE

DEPARTMENT OF ELECTRONICS ENGINEERING

2023-2024

ASKILL BASED MINI PROJECT

REPORT

"Download the ODI men's cricket match data from kaggel and read detail/information, draw boxplot for any column, find mean for all column. "

BACHELOR OF TECHNOLOGY

IN

ELECTRONICS ENGINEERING

SUBMITTED BY:
RATNESH ASATI
(0901EC221087)

SUBMITTED TO:
Dr. R.JENKIN SUJI
ASSISTANT PROFESSOR
DEPT.OF ELECTRONICS

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE, GWALIOR
(M.P.), INDIA

DEEMED TO BE UNIVERSITY
(Declared under Distinct Category by Ministry of Education, Government of India)
NAAC ACCREDITED WITH A++ GRADE

CERTIFICATE

This is to certify that the mini project report entitled **"Download the ODI men's cricket match data from kaggel and read detail/information, draw boxplot for any column, find mean for a. "** submitted by Ratnesh Asati has been carried out under the guidance of Dr. R.JENKIN SUJI, Department of Electronics Engineering , Madhav Institute of Technology & Science, Gwalior. The project report is approved for submission requirement for Mini Project in 5th semester in Department of Electronics Engineering , from Madhav Institute of Technology & Science, Gwalior (M.P).

SUBMITTED TO:
Dr. R.JENKIN SUJI
ASSISTANT PROFESSOR
DEPT OF ELECTRONICS

SUBMITTED BY:
RATNESH ASATI
(0901EC221087)

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE, GWALIOR (M.P.), INDIA

DEEMED TO BE UNIVERSITY
(Declared under Distinct Category by Ministry of Education, Government of India)
NAAC ACCREDITED WITH A++ GRADE

DECLARATION

I hereby declare that the project work entitled "**Download the ODI men's cricket match data from kaggel and read detail/information, draw boxplot for any column, find mean for all column. "** submitted to the Madhav Institute of Technology & Science Gwalior, is a record of an original work done by me under the guidance of Dr. R.JENKIN SUJI, Department of Electronics Engineering.

The results embodied in this thesis have not been submitted to any other University or Institute for the award of any degree or diploma.

Ratnesh Asati

Date : 2 1 - 1 1 - 2 0 2 4

Place : MITS GWALIOR

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE, GWALIOR (M.P.), INDIA

DEEMED TO BE UNIVERSITY

(Declared under Distinct Category by Ministry of Education, Government of India)

NAAC ACCREDITED WITH A++ GRADE

ACKNOWLEDGEMENT

We would like to express our gratitude towards **Dr. R.JENKIN SUJI**, for her support in accomplishment of our project on **"Download the ODI men's cricket match data from kaggel and read detail/information, draw boxplot for any column, find mean for all column. ".**

I would like to express my sincere appreciation for the opportunity to work on this project. The completion of this project would not have been possible without my own dedication, focus, and the skills I developed throughout the process. It has been a great learning experience.

CONTENT

INTRODUCTION

Cricket, especially One-Day Internationals (ODIs), is a sport that generates immense excitement and an equally massive amount of data. Analyzing this data can provide critical insights into team performances, player statistics, and match outcomes. This project focuses on leveraging a dataset from Kaggle containing information on men's ODI cricket matches. By exploring and analyzing this dataset, we aim to uncover patterns and trends that can be used for strategic planning, fan engagement, and performance improvement.

The project takes a structured approach to understanding the dataset, cleaning and preparing the data for analysis, and applying statistical techniques to derive insights. Through visualization techniques like boxplots, we aim to identify outliers and understand data distributions. Additionally, statistical measures like mean values across different columns will help summarize key trends in the dataset.

Objectives of the Project:

1. Dataset Exploration: Understand the structure, attributes, and content of the dataset to set a foundation for meaningful analysis.
2. Data Preparation: Perform preprocessing tasks like handling missing data, ensuring data consistency, and selecting relevant attributes for analysis.
3. Statistical Analysis: Calculate basic statistical metrics, such as the mean, to summarize numeric columns and interpret trends.
4. Data Visualization: Use boxplots to identify data distribution patterns, detect anomalies, and better understand the dataset.
5. Insight Derivation: Provide interpretations of the results that could be valuable for players, teams, and cricket enthusiasts.
6. Practical Applications: Highlight how data analytics can enhance decision-making in cricket, from strategy formulation to improving team performance.

This project demonstrates the power of data analytics in sports, showcasing how raw data can be transformed into actionable insights. By the end of this analysis, we aim to present findings that not only add to our understanding of ODI cricket but also underline the role of data-driven approaches in modern sports.

CODE AND IMPLEMENTATION

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
file_path = "/content/icc_wc_23_bat.csv"
batting_data = pd.read_csv(file_path)
batting_data.head()
```

	player	dismissal	runs	balls	minutes	4s	6s	strike_rate	team	opponent	innings	match_id
0	Rohit Sharma	b Madushanka	4	2	2	1	0	200.00	India	Sri Lanka	1	0
1	Shubman Gill	c Mendis b Madushanka	92	92	136	11	2	100.00	India	Sri Lanka	1	0
2	Virat Kohli	c Nissanka b Madushanka	88	94	139	11	0	93.61	India	Sri Lanka	1	0
3	Shreyas Iyer	c Theekshana b Madushanka	82	56	84	3	6	146.42	India	Sri Lanka	1	0
4	KL Rahul	c Hemantha b Chameera	21	19	34	2	0	110.52	India	Sri Lanka	1	0

```
# describing the basic columns of the table
b_stats = batting_data[['runs', 'balls', '4s', '6s', 'strike_rate']].describe()
b_stats
```

	runs	balls	4s	6s	strike_rate
count	875.000000	875.000000	875.000000	875.000000	875.000000
mean	26.773714	29.090286	2.558857	0.736000	82.771337
std	31.018840	28.579166	3.148084	1.490633	53.106854
min	0.000000	0.000000	0.000000	0.000000	0.000000
25%	5.000000	8.000000	0.000000	0.000000	50.000000
50%	15.000000	19.000000	1.000000	0.000000	80.430000
75%	39.000000	40.500000	4.000000	1.000000	106.660000
max	201.000000	143.000000	21.000000	11.000000	600.000000

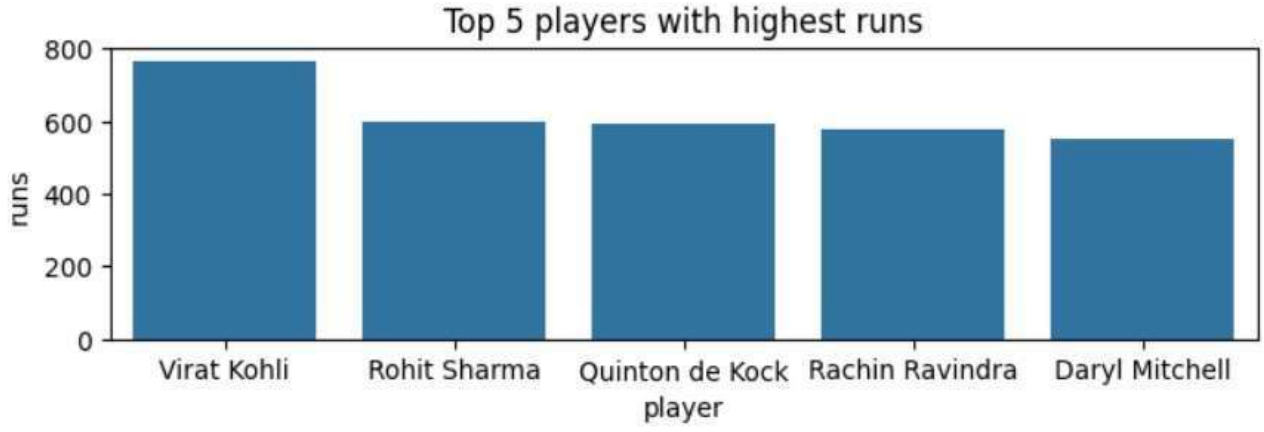
```
most_runs = batting_data.groupby("player")["runs"].sum().sort_values(ascending=False)
most_runs.head(5)
```

player	runs
Virat Kohli	765
Rohit Sharma	597
Quinton de Kock	594
Rachin Ravindra	578
Daryl Mitchell	552

dtype: int64

```
most_runs = batting_data.groupby("player")["runs"].sum().sort_values(ascending=False)
most_runs_plot=most_runs.head(5)
plt.figure(figsize=(8, 2))
sns.barplot(x=most_runs_plot.index, y=most_runs_plot)
plt.title("Top 5 players with highest runs")
```

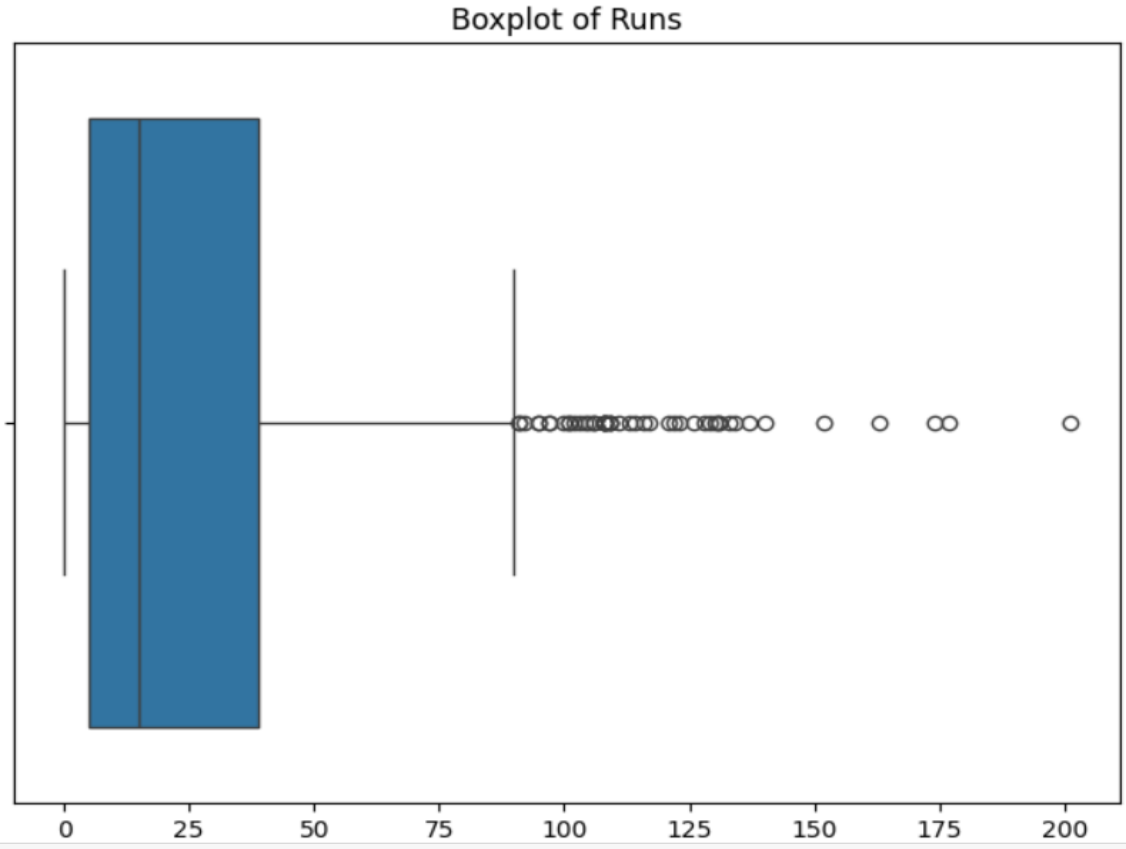
```
Text(0.5, 1.0, 'Top 5 players with highest runs')
```



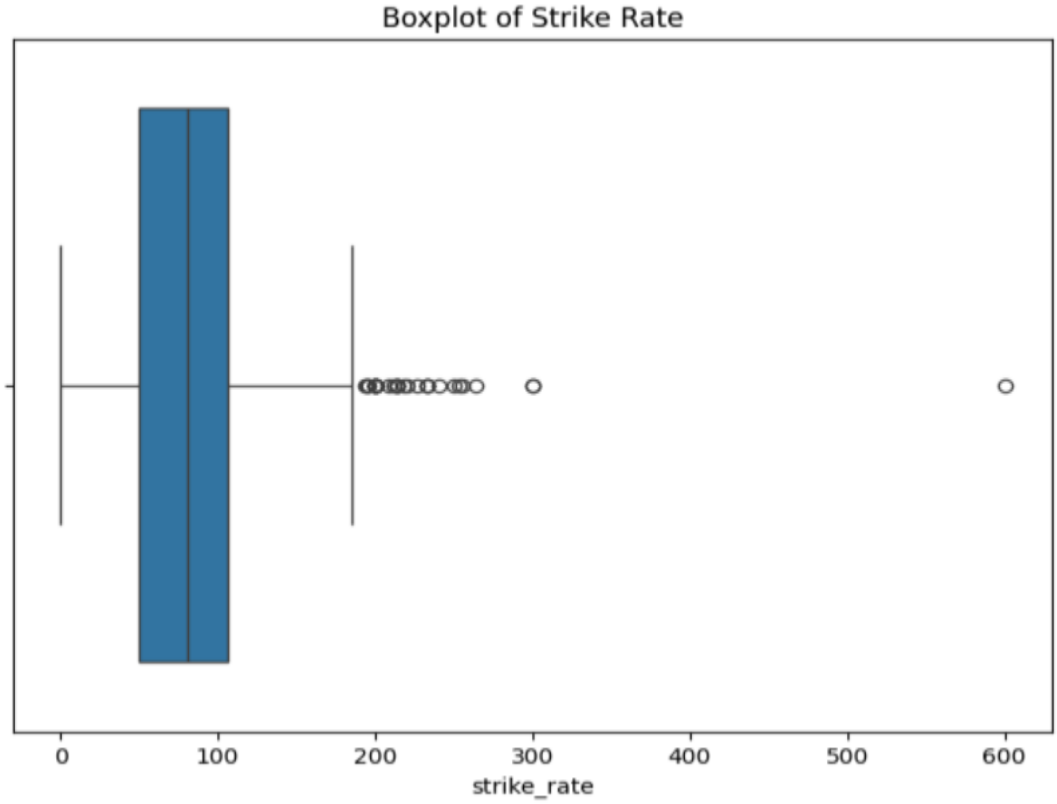

```
# Most runs in an inning by a player
df=batting_data
df['match_id'] = batting_data['team'] + '-' + batting_data['opponent']
idx_max_runs_per_match = df.groupby('match_id')['runs'].idxmax()
most_runs_per_match = df.loc[idx_max_runs_per_match]
most_runs_per_match[most_runs_per_match["team"]=="India"]
```

	player	dismissal	runs	balls	minutes	4s	6s	strike_rate	team	opponent	innings	match_id
838	Rohit Sharma	b Rashid Khan	131	84	111	16	5	155.95	India	Afghanistan	2	India-Afghanistan
641	KL Rahul	not out	97	115	180	8	2	84.34	India	Australia	2	India-Australia
750	Virat Kohli	not out	103	97	113	6	4	106.18	India	Bangladesh	2	India-Bangladesh
31	Rohit Sharma	c Livingstone b Rashid	87	101	162	10	3	86.13	India	England	1	India-England
411	Shreyas Iyer	not out	128	94	145	10	5	136.17	India	Netherlands	1	India-Netherlands
94	Virat Kohli	c Conway b Southee	117	113	149	9	2	103.53	India	New Zealand	1	India-New Zealand
535	Rohit Sharma	c Iftikhar Ahmed b Shaheen Shah Afridi	86	63	91	6	6	136.50	India	Pakistan	2	India-Pakistan
360	Virat Kohli	not out	101	121	195	10	0	83.47	India	South Africa	1	India-South Africa
1	Shubman Gill	c Mendis b Madushanka	92	92	136	11	2	100.00	India	Sri Lanka	1	India-Sri Lanka

```
#boxplot of runs
plt.figure(figsize=(8, 6))
sns.boxplot(x=batting_data['runs'])
plt.title('Boxplot of Runs')
plt.show()
```




```
#boxplot for strike rate
plt.figure(figsize=(8, 6))
sns.boxplot(x=batting_data['strike_rate'])
plt.title('Boxplot of Strike Rate')
plt.show()
```



```python
# Calculate mean for all numeric columns
mean_values = batting_data.mean(numeric_only=True)
print("Mean values for all numeric columns:")
mean_values
```

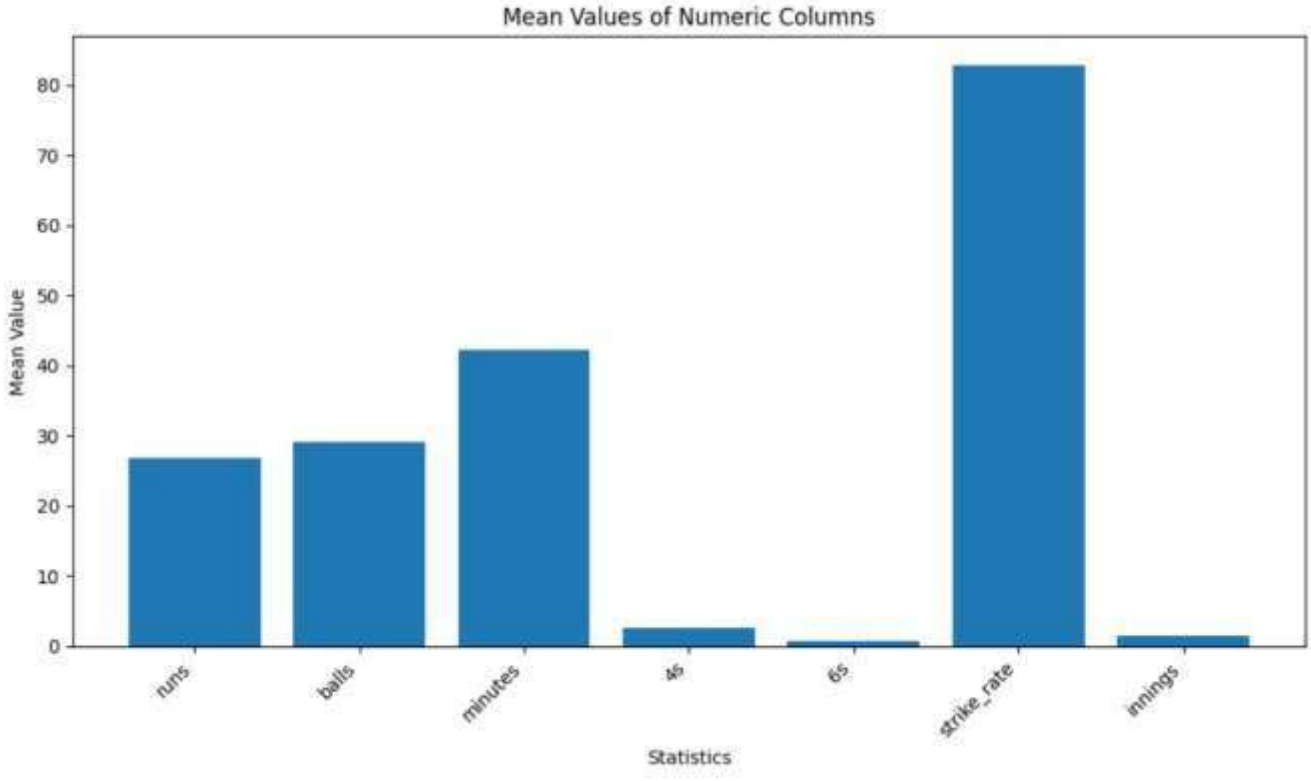
```
Mean values for all numeric columns:
```

	0
runs	26.773714
balls	29.090286
minutes	42.259429
4s	2.558857
6s	0.736000
strike_rate	82.771337
innings	1.459429

dtype: float64

```python
import matplotlib.pyplot as plt

plt.figure(figsize=(10, 6))
plt.bar(mean_values.index, mean_values.values)
plt.xlabel("Statistics")
plt.ylabel("Mean Value")
plt.title("Mean Values of Numeric Columns")
plt.xticks(rotation=45, ha='right')
plt.tight_layout()
plt.show()
```



APPLICATIONS

The analysis of the ICC World Cup 2023 batting dataset has several practical applications in the real world, benefitting players, teams, analysts, and fans. Here's how this data can be effectively used:

1. Performance Analysis

Teams and players can use insights from the data to evaluate performance trends:

-Players can identify strengths (e.g., high strike rates or boundary frequency) and areas for improvement, such as struggles against specific opponents or bowling types.

-Teams can analyze overall team performance, identifying key contributors and areas where strategies need adjustment.

2. Strategic Planning for Matches

Data-driven strategies help teams prepare for future matches:

- Opponent Analysis: Teams can review how players performed against specific opponents to tailor game plans. For instance, focusing on players who consistently score well against a particular opponent's bowling attack.

.

3. Fan Engagement

Data visualization and insights enhance fan experiences:

-Broadcast Enhancements: During live matches, broadcasters can use data to provide real-time comparisons, such as strike rates against spinners versus pacers.

-Fantasy Leagues: Fans participating in fantasy cricket leagues can use this data to pick players based on past performances.

-Storytelling: Analysts and commentators can use data trends to tell compelling stories about players or teams.

4. Historical Insights

Data acts as a repository for understanding trends over time:

-Records and Milestones: Comparing the current World Cup's statistics with past tournaments helps assess the evolution of scoring patterns, strike rates, or boundary frequency.

-Talent Scouting: Upcoming players can be identified based on their standout performances in critical situations.

5. Enhancing Decision-Making with Technology

Advanced analytics powered by this data can integrate into AI tools for predictive insights, such as forecasting match outcomes or projecting a player's career trajectory.

CONCLUSION

This project successfully analyzed batting performances from the ICC World Cup 2023, uncovering key insights into players' runs, strike rates, and boundary-hitting abilities. Through careful data preprocessing and visualization, it demonstrated how data science can enhance performance evaluation, strategic decision-making, and fan engagement in sports.

The analysis highlighted the importance of clean, well-structured data and showcased the growing role of data-driven insights in cricket. While the project achieved its objectives, integrating additional data and advanced techniques in the future could provide a more comprehensive understanding of the game.

This project also served as a valuable learning experience in applying data science methodologies to a real-world dataset. It emphasized the significance of effective data preprocessing, statistical analysis, and visualization in deriving actionable insights. By bridging the gap between raw data and meaningful interpretation, the analysis demonstrated the potential of data-driven approaches to revolutionize sports analytics, not only in cricket but across various domains.