

Comprehensive Analysis of Cricket Dataset: Insights into Player Performances Across Formats (ODI, T20 and Test)








Statistical and Mathematical Methods for Data science

Professor:
Dr. Fahad Samad

Presented By:
Javeria Hassan (23K-8019)
Jaweria Sohail (23K-8050)
Zarnain Maryam - (23K-7702)

Presentation Breakdown

The Presentation is breakdown in the following sections:

-  Abstract
-  Introduction
-  Methodology
-  Dataset Overview
-  Exploratory Data Analysis
-  Probability Analysis
-  Conclusion

Abstract

- ❑ This study involves a comprehensive statistical analysis of a cricket dataset, encompassing ODI, T20, and Test matches. Delving into batting, bowling, and fielding data, we conduct Exploratory Data Analysis (EDA) to gain valuable insights. Through this exploration, we aim to address key questions related to probability, providing a nuanced understanding of player performances and strategic dynamics across different cricket formats.

Introduction

- ❑ **Holistic Exploration:** This report undertakes a comprehensive analysis of cricket batting and fielding data, spanning ODI, T20, and Test formats, to reveal intricate patterns and trends within player performances.
- ❑ **Dynamic Insights:** Rooted in Exploratory Data Analysis (EDA) and probability assessments, our study aims to uncover valuable insights into the nuanced dynamics influencing success on the cricket field.
- ❑ **Beyond Statistics:** Our exploration goes beyond traditional statistics, offering a holistic perspective on how players navigate challenges in diverse formats, from pursuing centuries to strategic fielding intricacies.

Methodology

- ❑ Data Collection
- ❑ Exploratory Data Analysis
- ❑ Statistical Analysis
- ❑ Probability Assessment

Dataset Overview

- ❑ **Dataset Origin:** Provided by the instructor, the cricket dataset is the cornerstone for our analysis, encompassing player performances in diverse cricket formats.
- ❑ **Data Overview:** With 7,638 rows, the ODI, Test, and T20 datasets offer a comprehensive view of player details, match statistics, bowling metrics, and notable achievements.
- ❑ **Cleaning Process:** Meticulous cleaning involved dropping extraneous columns, addressing missing values, transforming career spans, and parsing 'BBI' columns for enhanced analysis.
- ❑ **Quality Assurance:** The refined datasets, free of outliers and missing values, ensure reliability for a thorough and confident exploration of player contributions.

Dataset Overview

- ❑ Summary Statistics
- ❑ Key Metrics: Runs, Batting Average, Wickets, etc.
- ❑ Data Cleaning and Preprocessing

Exploratory Data Analysis - Batting

- ❑ Distribution of Runs, Batting Averages, and Wickets
- ❑ Relationship Between Batting Average and Runs
- ❑ Variation in Strike Rates Among Batsmen

Exploratory Data Analysis

Following meticulous data cleaning, our Exploratory Data Analysis (EDA) kicked off with:

- ❑ Univariate Analysis
- ❑ Statistical Analysis
- ❑ Correlation Analysis
- ❑ Probabilistic Analysis

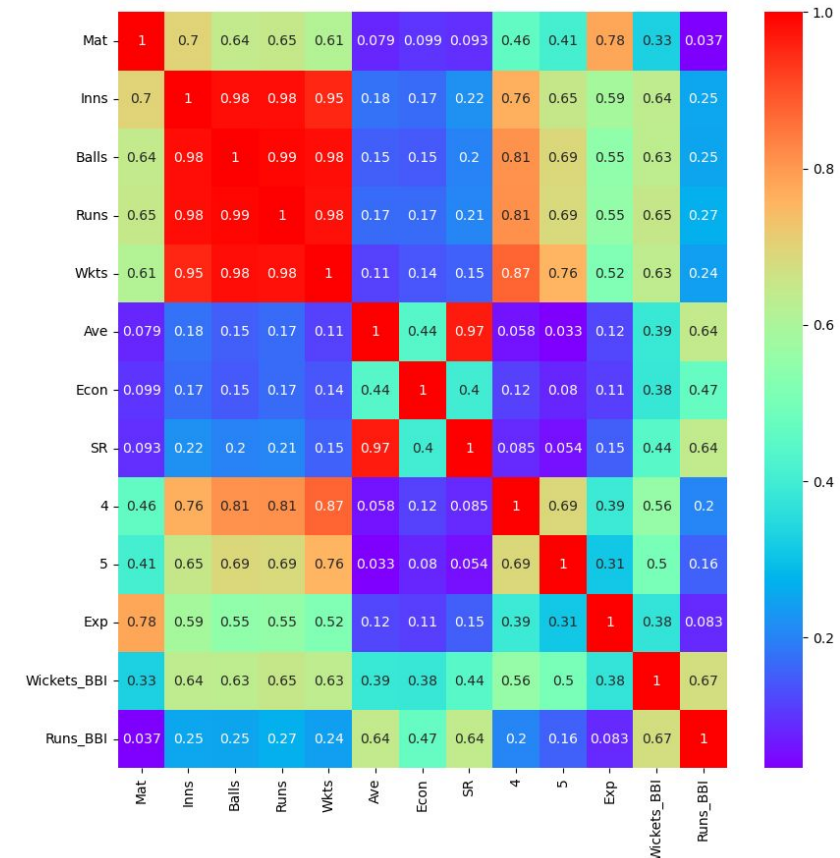
EDA: Statistical Analysis

Our statistical analysis delved into each dataset separately, employing measures such as mean, median, and mode (Fig: 2) to gain a comprehensive understanding of the central tendencies within the cricket datasets.

	Mat	Inns	Balls	Runs	Wkts	Ave	Econ	SR	4	5	Exp	Wickets_BBI	Runs_BBI
count	2582	2582	2582	2582	2582	2582	2582	2582	2582	2582	2582	2582	2582
mean	36.03718	19.431448	865.308675	678.481022	20.933772	25.064105	3.810275	30.136483	0.521301	0.194423	4.099535	1.892719	19.716112
std	58.2743	39.64081	1866.739974	1407.744194	47.728981	27.381531	2.842238	30.988353	1.483518	0.749635	4.424642	1.882986	19.183436
min	1	0	0	0	0	0	0	0	0	0	0	0	0
25%	4	0	0	0	0	0	0	0	0	0	0	0	0
50%	12	4	130	113	3	25.905	4.53	32.25	0	0	3	2	20
75%	41	19	767.25	629.75	18	37.91	5.35	46.6	0	0	7	3	34
max	463	372	18811	13632	534	251	26	234	17	13	23	8	83

EDA: Correlation Analysis (Bowling)

Notably, the highest correlation within the ODI bowling dataset was observed between 'Runs' and 'Balls,' reaching an impressive coefficient of approximately 0.98

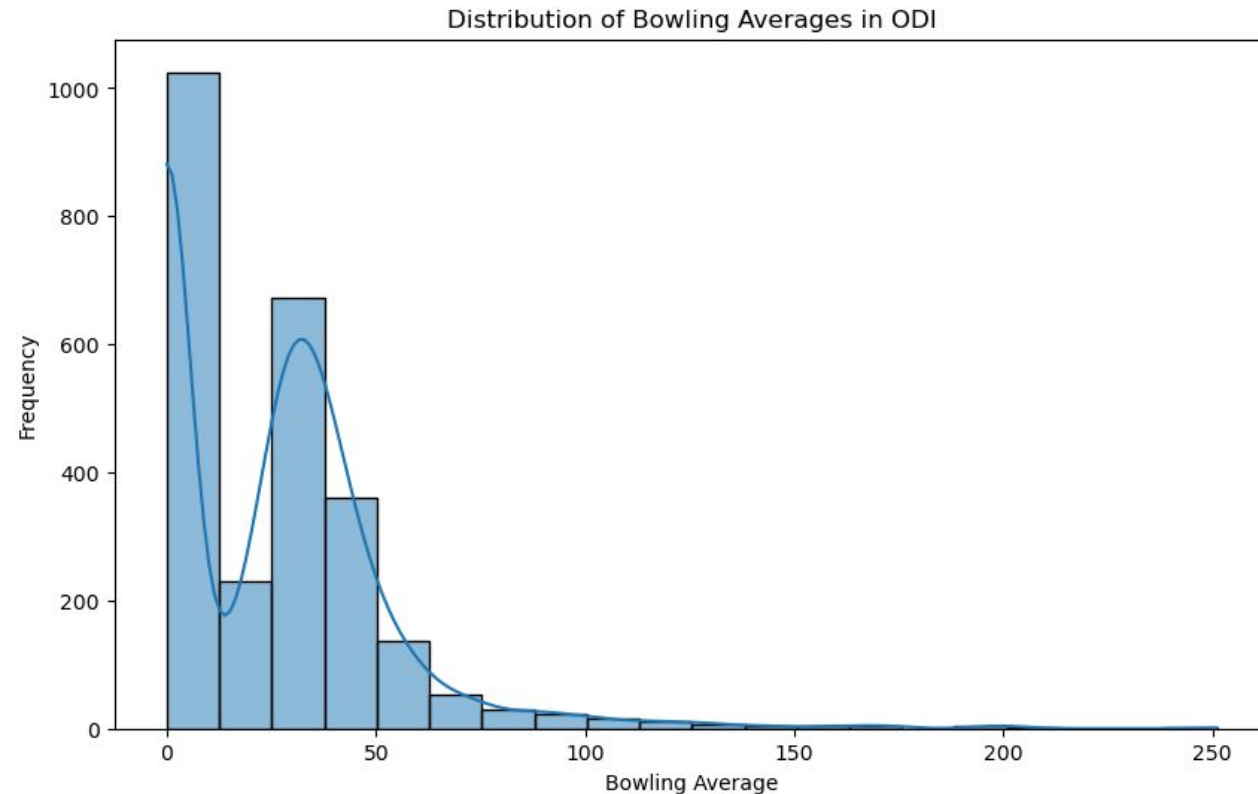


Probability Analysis

- ❑ Probability of Scoring a Century
- ❑ Probability of Achieving a Batting Average Threshold
- ❑ Relationship Between Balls Faced and Century Probability

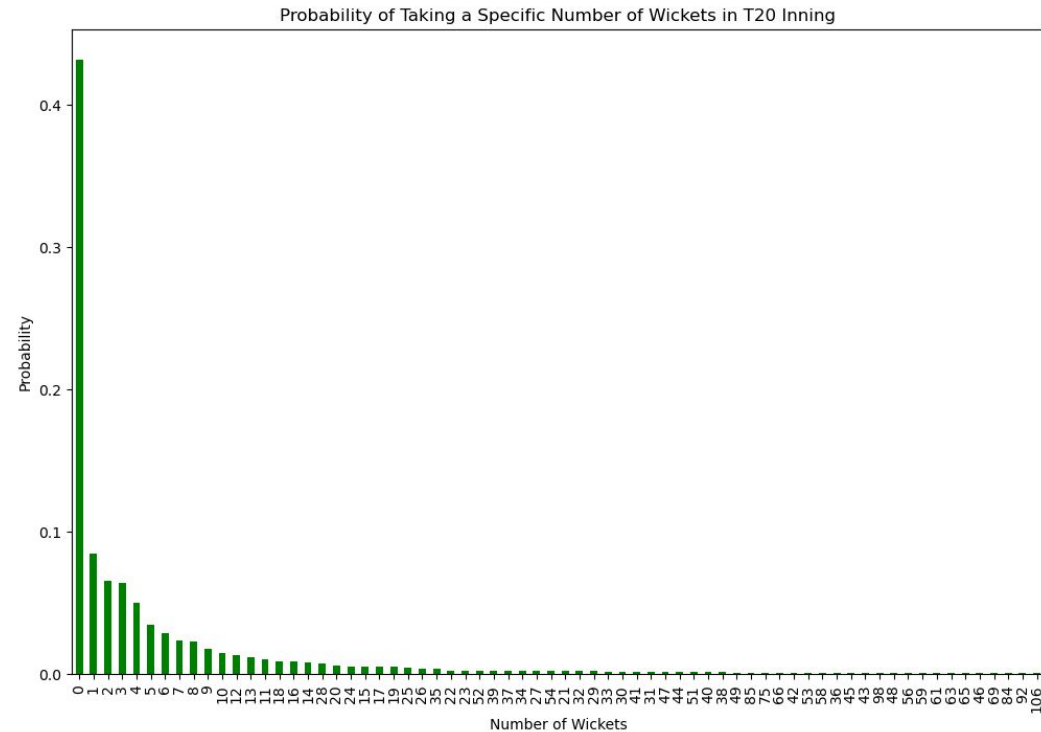
Probability Analysis (Bowling Data)

- ❑ Bowling Average among different Players



Probability Analysis (Bowling Data)

- ❑ Probability of taking specific number of wickets



Probability Analysis (Batting Data)

Player Performance Metrics

- ❑ Calculating Combined Score (Runs + Wickets)
- ❑ Determining the Best Player Performance
- ❑ Weighting Factors for Runs and Wickets

Probability Analysis (Fielding Data)

- ❑ Distribution of Catches and Stumpings
- ❑ High-Performing Fielders
- ❑ Relationship Between Fielding Dismissals and Player Performance

Player Performance Metrics

- ❑ Calculating Combined Score (Runs + Wickets)
- ❑ Determining the Best Player Performance
- ❑ Weighting Factors for Runs and Wickets

Visualizations

- ❑ Graphical Representation of Insights
- ❑ Box Plots, Violin Plots, Scatter Plots
- ❑ Highlights of Key Visuals

Conclusion

Our extensive analysis of cricket datasets across ODI, T20, and Test formats, encompassing meticulous cleaning, exploratory analysis, and probabilistic assessments, has illuminated intricate player performances, revealing crucial insights into distribution patterns, correlations, and nuanced cricket dynamics.