



VIRGINIA COMMONWEALTH UNIVERSITY

Statistical analysis and modelling (SCMA 632)

A1b: Preliminary preparation and analysis of data- Descriptive statistics

KATHRYN SHAJU

V01111327

Date of Submission: 18-06-2024

CONTENTS

Sl. No.	Title	Page No.
1.	Introduction	1
2.	Results & Interpretations	2 -3
3.	Recommendations	3
4.	References	4

Introduction

For this study we have focused on the data set containing IPL salaries of the players as well as IPL ball by ball update (2024), which is a comprehensive dataset including all information of the IPL players from year 2008-2024. This data set has been run through two sophisticated statistical programs, R & Python. This ensures that the results and interpretations of the following is accurate and prove to be immensely helpful for future researchers.

The objectives of this study mainly include arranging the data IPL round-wise and batsman, ball, runs, and wickets per player per match, to indicate the top three run-getters and tow three wicket-takers in each IPL round, fitting the most appropriate distribution for runs scored and wickets taken by the top three batsmen and bowlers in the last three IPL tournaments and finding the relationship between a player's performance and the salary he gets (for this we have focused on player, A.Nortje).

Researching and interpreting this data not only help in recording the different plays during the years, but most importantly this data when cleaned and manipulated to reveal the data you require is very useful in understanding the performance of the player throughout the years allowing potential, team owners to make an informed decision in understanding the strength and weakness of the players. Moreover testing the relationship between the player's salaries and performance also helps in a wider understanding of what exactly could define the limit of salary negotiated for a player.

Results & Interpretations

1. Arranging the Data IPL round-wise and batsman, ball, runs, and wickets per player per match

In both R and Python we run codes to arrange the Data IPL round-wise and batsman, ball, runs, and wickets per player per match, this helps our disjointed data to become more orderly allowing us to navigate more easily. R and Python are both good to clean the data and to do data manipulation resulting in accurate results required.

2. Top three run-getters and tow three wicket-takers in each IPL round

Coding for the top three run-getters and top three wicket-takers in each IPL round, allow us to receive information about the top 3 run scorers and wicket takers in every season, allowing easy understanding of the most competent players.

3. Fit distribution for runs scored and wickets taken by the top three batsmen and bowlers in the last three IPL tournaments

First coding to get the top 3 batsmen and bowlers in the last 3 IPL tournaments we obtain the information, following which we run a fit distribution to ascertain the pattern and characteristics of data which not only helps us in identifying the most suitable statistical model that can be fit for our data set but also the parameters to be considered.

4. Relationship between a player's performance and the salary he gets

For this purpose we have taken Anrich Nortje as the player and try to understand if there is any correlation between his performance and the salary that he receives.

From both R and Python, the results indicate that there is a relationship between him receiving his salary and his performance in the game. Wherein in both software the p value <0.05 .

Recommendations

Researching and interpreting this data not only help in recording the different plays during the years, but most importantly this data when cleaned and manipulated to reveal the data you require is very useful in understanding the performance of the player throughout the years allowing potential, team owners to make an informed decision in understanding the strength and weakness of the players. Moreover testing the relationship between the player's salaries and performance also helps in a wider understanding of what exactly could define the limit of salary negotiated for a player.

References

IPL_ball_by_ball_updated(2024) dataset

https://drive.google.com/file/d/1DwBQEwMecrwj9Vby-_ohHNOonJMwxtWiT/view?usp=sharing

IPL SALARIES (2024) data set

<https://github.com/scma-632/datasets/blob/main/IPL%20SALARIES%202024.xlsx>