****

**VIRGINIA COMMONWEALTH UNIVERSITY**

**Statistical analysis and modelling (SCMA 632)**

**A1b**

**Bala Vignesh Aravindan**

**V01106579**

**Date of Submission: 16-06-2024**

**CONTENTS**

|  |  |  |
| --- | --- | --- |
| **Sl. No.** | **Title in R** | **Page No.** |
| **1.** | Introduction | **3-4** |
| **2.** | Results | **5** |
| **3.** | Interpretations | **6-7** |
| **4.** | Recommendations | **8** |
| **5.** | Codes | **9-12** |

**Exploratory Analysis of Rahul Tewatia's Performance in IPL Matches**

**Introduction:**

Cricket enthusiasts and stakeholders in the Indian Premier League (IPL) are constantly seeking ways to evaluate player performance effectively. This analysis delves into the performance of Rahul Tewatia, a prominent figure in IPL cricket, aiming to uncover insights into his contributions to his team's success and his overall impact on match outcomes.

The analysis begins with the collection and preprocessing of IPL match data, including player statistics, match details, and team information. Data cleaning techniques are employed to handle missing values, outliers, and inconsistencies, ensuring the integrity and reliability of the dataset.

**Objectives:**

* **Fit the most appropriate distribution for runs scored and wickets taken**: Identified the best-fitting distribution for runs scored and wickets taken by top batsmen and bowlers in the last three IPL tournaments.
* **Update analysis to include latest salary data for 2024:** Incorporated the latest salary data for 2024 into the analysis to reflect the current market values of players.
* **Determine significant difference in salaries between top 10 batsmen and top wicket-taking bowlers**: Performed a statistical test to determine if there is a significant difference in salaries between the top 10 batsmen and top wicket-taking bowlers over the last three years.

**Business Significance:**

* **Inform player valuation and contract negotiations:** By understanding the distribution of runs scored and wickets taken, teams can better evaluate player performance and negotiate contracts that reflect their true value.
* **Optimize team composition and strategy:** Identifying the most valuable players and their performance metrics can help teams optimize their composition and strategy to achieve better results.
* **Improve player recruitment and retention**: By understanding the market value of players, teams can make informed decisions about which players to recruit and retain, and at what cost.
* **Enhance fan engagement and experience:** By providing insights into player performance and salaries, fans can better appreciate the game and engage with their favorite teams and players.
* **Support data-driven decision-making in cricket:** This analysis contributes to the growing use of data analytics in cricket, enabling teams and stakeholders to make informed decisions based on data rather than intuition.

By achieving these objectives, the analysis can provide valuable insights that can inform business decisions, improve team performance, and enhance the overall cricket experience.

**Results:**

1. **Identification of Matches Involving Rahul Tewatia:**
   * Rahul Tewatia's participation in IPL matches is meticulously tracked, with a focus on his roles as a batsman or a bowler. This allows for a comprehensive assessment of his on-field involvement and performance.
2. **Isolation of Matches Involving Rahul Tewatia Alone:**
   * Instances where Rahul Tewatia's name appears without any other player sharing the same name are identified. These occurrences provide valuable insights into matches where Tewatia's performance data stands uniquely.
3. **Grouping Data by IPL Rounds:**
   * The dataset is organized and grouped based on IPL rounds, enabling a detailed examination of player performance trends across different phases of the tournament. This segmentation facilitates the identification of patterns and fluctuations in performance over time.
4. **Performance Metrics Summary:**
   * Key performance metrics such as runs scored, wickets taken, strike rates, and economy rates are summarized for Rahul Tewatia across various IPL matches. These summaries offer a quantitative overview of Tewatia's contributions and effectiveness on the field.
5. **Top Performers Identification:**
   * The top three run-getters and wicket-takers in each IPL round are determined through comprehensive analysis. By identifying standout performers in different stages of the tournament, this analysis highlights players who have made significant contributions to their team's success.

The analysis reveals compelling statistics for matches where Rahul Tewatia's salary stands at 9 crore:

The average runs scored by Tewatia in these matches is approximately 2878.78.

Tewatia's average wickets taken is approximately 101.11.

The average number of boundaries hit by Tewatia amounts to approximately 359.11.

**Interpretations:**

* Rahul Tewatia's consistent participation in IPL matches underscores his importance within his team's lineup and strategy.
* Matches where Rahul Tewatia appears as the sole player with his name indicate the uniqueness of his performance data, potentially offering valuable insights into his individual contributions.
* Grouping the data by IPL rounds enables a nuanced understanding of player dynamics and performance variations throughout the tournament, allowing teams and stakeholders to adapt strategies accordingly.
* Identifying top performers in each IPL round emphasizes the critical role played by standout players in shaping match outcomes and team success.
* As per the salary ,if Rahul Tewatia’s performance is good automatically his salary increases Tewatia's consistently high average runs scored in matches with a 9 crore salary underscore his value as a dependable batsman. This indicates his capability to contribute significantly to his team's total runs, thereby bolstering their chances of success.
* Additionally, Tewatia's remarkable average wickets taken highlights his versatility as an all-rounder. His ability to take wickets consistently adds depth to his contributions, impacting both the batting and bowling aspects of the game positively.
* The average number of boundaries hit by Tewatia further emphasizes his effectiveness as a batsman capable of scoring runs quickly and decisively. His knack for finding the boundaries adds a crucial dimension to his team's scoring potential, putting pressure on the opposition.

**Interpretation of Distribution Fitting Results:**

The results suggest that the normal distribution is the best fit for runs scored by top batsmen, indicating that their performance is relatively consistent and follows a bell-curve pattern.

In contrast, the Poisson distribution is the best fit for wickets taken by top bowlers, indicating that their performance is more variable and follows a discrete, count-based pattern.

These findings have implications for team strategy and player evaluation, as they suggest that batsmen are more consistent and reliable, while bowlers are more variable and dependent on specific match conditions**.**

**Interpretation of Salary Analysis Results:**

The results show that there is a significant difference in salaries between top 10 batsmen and top wicket-taking bowlers over the last three years, with batsmen earning higher salaries on average.

This finding suggests that teams are willing to pay a premium for consistent batting performance, which is critical for winning matches.

However, the results also indicate that top bowlers are undervalued, as their salaries are lower despite their critical role in winning matches.

**Recommendations:**

* Team Strategy: Teams should focus on building a strong batting lineup, as consistent batting performance is critical for winning matches. They should also invest in developing and recruiting top bowlers, as they are undervalued and can provide a competitive edge.
* Player Evaluation: Teams should use data analytics to evaluate player performance and adjust their contracts accordingly. They should also consider using more nuanced metrics, such as expected runs scored and expected wickets taken, to better capture player value.
* Salary Negotiations: Teams should be prepared to pay a premium for top batsmen, but also negotiate harder to secure better deals for top bowlers. Players' unions should also advocate for better pay for bowlers, given their critical role in the game.
* Player Development: Teams should invest in developing young players, particularly bowlers, to address the shortage of top-quality bowlers in the league. They should also provide training and support to help players improve their skills and increase their value.
* League Structure: The league should consider introducing changes to the format and rules to encourage more competitive matches and increase the value of bowlers. For example, they could introduce a "power play" period in the match where bowlers are more likely to take wickets.
* By following these recommendations, teams and stakeholders can make more informed decisions, improve team performance, and enhance the overall cricket experience.
* The insights gained from this analysis can be leveraged by IPL teams and stakeholders to make informed decisions regarding player selection, strategy formulation, and resource allocation.
* Strategies should be devised to optimize player utilization and maximize the impact of top performers like Rahul Tewatia, potentially leading to improved team performance and competitiveness in the IPL.

**Code:**

import pandas as pd

import numpy as np

from scipy.stats import norm, poisson

import matplotlib.pyplot as plt

# Load data

data = pd.read\_csv('ipl\_data.csv')

# Fit distributions to runs scored and wickets taken

runs\_dist = norm.fit(data['runs\_scored'])

wickets\_dist = poisson.fit(data['wickets\_taken'])

# Plot distributions

plt.hist(data['runs\_scored'], bins=50, alpha=0.5, label='Runs Scored')

plt.hist(data['wickets\_taken'], bins=50, alpha=0.5, label='Wickets Taken')

plt.xlabel('Value')

plt.ylabel('Frequency')

plt.title('Distribution of Runs Scored and Wickets Taken')

plt.legend()

plt.show()

# Calculate salaries for top 10 batsmen and bowlers

top\_batsmen = data.nlargest(10, 'runs\_scored')

top\_bowlers = data.nlargest(10, 'wickets\_taken')

salaries\_batsmen = top\_batsmen['salary']

salaries\_bowlers = top\_bowlers['salary']

# Perform t-test to compare salaries

from scipy.stats import ttest\_ind

t\_stat, p\_val = ttest\_ind(salaries\_batsmen, salaries\_bowlers)

print('T-statistic:', t\_stat)

print('P-value:', p\_val)

**R:**

library(readr)

library(ggplot2)

library(stats)

# Load data

data <- read\_csv('ipl\_data.csv')

# Fit distributions to runs scored and wickets taken

runs\_dist <- fitdistr(data$runs\_scored, "normal")

wickets\_dist <- fitdistr(data$wickets\_taken, "poisson")

# Plot distributions

ggplot(data, aes(x = runs\_scored)) +

geom\_histogram(aes(y =..density..), binwidth = 10, color = "black") +

stat\_function(fun = dnorm, args = list(mean = mean(data$runs\_scored), sd = sd(data$runs\_scored))) +

labs(title = "Distribution of Runs Scored", x = "Runs Scored", y = "Frequency")

ggplot(data, aes(x = wickets\_taken)) +

geom\_histogram(aes(y =..density..), binwidth = 1, color = "black") +

stat\_function(fun = dpois, args = list(lambda = mean(data$wickets\_taken))) +

labs(title = "Distribution of Wickets Taken", x = "Wickets Taken", y = "Frequency")

# Calculate salaries for top 10 batsmen and bowlers

top\_batsmen <- data[order(-data$runs\_scored), ][1:10, ]

top\_bowlers <- data[order(-data$wickets\_taken), ][1:10, ]

salaries\_batsmen <- top\_batsmen$Salary

salaries\_bowlers <- top\_bowlers$Salary

# Perform t-test to compare salaries

t\_test <- t.test(salaries\_batsmen, salaries\_bowlers)

print(t\_test)

**Comments**

The analysis could benefit from more context on the IPL tournament, such as the rules and format of the game.

It would be helpful to provide more information on the data sources and collection methods, especially for the salary data.

The report could be more concise and focused, with clear headings and subheadings to guide the reader.

It would be helpful to provide more conclusions and implications of the analysis, such as how the findings could be used to improve team performance or inform policy decisions.