****

**VIRGINIA COMMONWEALTH UNIVERSITY**

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

**A1b: Analysis of IPL dataset - Python**

**Manoranjan Mohankumar**

**V01107254**

**Date of Submission: 18-06-2024**

**CONTENTS**

|  |  |  |
| --- | --- | --- |
| **Sl. No.** | **Title** | **Page No.** |
| **1.** | Introduction | **1** |
| **2.** | Results | **2** |
| **3.** | Interpretations | **3** |
| **4.** | Recommendations | **5** |
| **5.** | Codes | **-** |
| **6.** | References | **6** |

**Introduction**

The Indian Premier League (IPL) is one of the most popular and competitive cricket leagues in the world, attracting top cricketing talent from across the globe. Player salaries in the IPL are a significant aspect, often reflecting a player's perceived value and expected contribution to their team. Understanding the relationship between player salaries and their actual on-field performance is crucial for franchises to make informed decisions regarding player acquisitions, retention, and salary negotiations.

This report aims to analyze the correlation between player salaries and their performance during the 2024 IPL season. The primary performance metrics considered in this analysis include the total runs scored by batsmen and the total wickets taken by bowlers. By examining these metrics, we aim to determine whether higher salaries are justified by higher performance levels, thereby providing insights into the effectiveness of salary allocations.

The datasets used for this analysis include:

1. IPL Ball-by-Ball Data: This dataset contains detailed information on each ball bowled in the IPL matches up to the 2024 season, including runs scored and wickets taken.
2. IPL Salary Data for 2024: This dataset includes the salaries of IPL players for the year 2024, along with their international and iconic status.

The analysis involves grouping the data by season, innings, players (both batsmen and bowlers), and calculating aggregate performance metrics such as total runs scored and wickets taken. Additionally, statistical tests are performed to identify the best-fitting distributions for player performance data. The correlation between salaries and performance metrics is then computed to understand the strength and nature of their relationship.

By analyzing these correlations and distributions, the report seeks to provide actionable insights for IPL franchises, helping them optimize their player investment strategies for enhanced team performance and competitive advantage.

**Results**

1. Top Three Run Getters and Wicket Takers:
   * Top Run Getters:
     + Shubman Gill: 890 runs
     + F du Plessis: 730 runs
     + DP Conway: 672 runs
   * Top Wicket Takers:
     + HV Patel: 19 wickets
     + Mukesh Kumar: 15 wickets
     + Arshdeep Singh: 14 wickets
2. Best Distribution Fit for Run Scored by Top Batsmen:
   * RD Gaikwad: Best fitting distribution - NCT (p-value: 0.588)
   * V Kohli: Best fitting distribution - Beta (p-value: 0.781)
   * B Sai Sudharsan: Best fitting distribution - F (p-value: 0.974)
3. Best Distribution Fit for Wickets Taken by Top Bowlers:
   * HV Patel: Best fitting distribution - Alpha (p-value: 0.0003)
   * Mukesh Kumar: Best fitting distribution - Alpha (p-value: 0.603)
   * Arshdeep Singh: Best fitting distribution - T (p-value: 0.0045)
4. Correlation Analysis:
   * Salary and Runs Correlation: The correlation between salary and runs scored is 0.306, indicating a moderate positive relationship.
   * Salary and Total Performance Correlation: The correlation between salary and total performance (runs + wickets) is 0.360, indicating a stronger positive relationship.
5. Specific Analysis for Surya Kumar Yadav:
   * Total Performance (Runs + Wickets) in 2024: 176
   * Salary in 2024: 800 lakh INR

**Interpretations**

1. Top Performers:
   * The identification of top run-scorers and wicket-takers in the 2024 IPL season reveals the players who had the most significant impact on their respective teams. For instance, Shubman Gill leading with 890 runs and HV Patel with 19 wickets indicates their consistent performance and crucial role in their teams' successes.
   * These top performers are likely to be the focal points in matches and attract higher salaries due to their contributions.
2. Distribution Fit:
   * The fitting of various statistical distributions to the performance data of top batsmen and bowlers provides insights into the variability and patterns in player performance.
     + For example, RD Gaikwad’s performance best fits an NCT (Non-Central t-distribution), indicating his performance may have significant variance but is generally higher.
     + V Kohli’s performance fitting a Beta distribution suggests a more predictable and consistent range of scores.
   * Understanding these distributions helps in predicting future performances and making data-driven decisions about player investments.
3. Correlation Insights:
   * The correlation between player salaries and their performance metrics provides a quantitative measure of how well the salaries align with actual performance.
     + A correlation of 0.306 between salary and runs scored indicates a moderate positive relationship, suggesting that higher-paid players tend to score more runs, but the relationship is not very strong.
     + The stronger correlation of 0.360 between salary and total performance (combining runs and wickets) suggests that considering both batting and bowling performances provides a better alignment with player salaries.
   * These correlations imply that while higher salaries generally correspond to better performances, there are other factors at play, and not all high-performing players receive high salaries.
4. Surya Kumar Yadav Analysis:
   * The specific analysis of Surya Kumar Yadav shows a total performance of 176 runs and wickets combined, with a substantial salary of 800 lakh INR.
   * This data point highlights that high salaries are sometimes attributed to players with a proven track record and high potential, even if their performance in a single season may vary.
   * It underscores the importance of considering both current season performance and overall career performance when determining player salaries.
5. Overall Implications:
   * The analysis indicates that IPL franchises tend to invest heavily in players who have demonstrated high performance, but there is still room for improving the alignment between salaries and performance.
   * Factors such as player marketability, fan following, and leadership qualities also likely influence salary decisions, beyond on-field performance alone.
   * By refining their salary strategies based on performance data and distribution fits, franchises can ensure more effective and justified player investments, leading to better team performance and financial efficiency.

**Recommendations**

1. Investment in Key Players:
   * Franchises should continue investing in top-performing players, as higher salaries correlate with better performance, ensuring a competitive edge.
2. Player Performance Monitoring:
   * Regular analysis of player performance and salary alignment can help franchises optimize their budget allocation and player retention strategies.
3. Data-Driven Decisions:
   * Utilize statistical models and distribution fits to predict player performance and make informed decisions regarding player selection and salary negotiations.
4. Focus on Emerging Talent:
   * Identify and invest in emerging players who show potential for high performance, ensuring a balanced and future-ready team composition.

**References:**

IPL\_Ball\_by\_Ball dataset

IPL\_Salaries dataset

ChatGPT