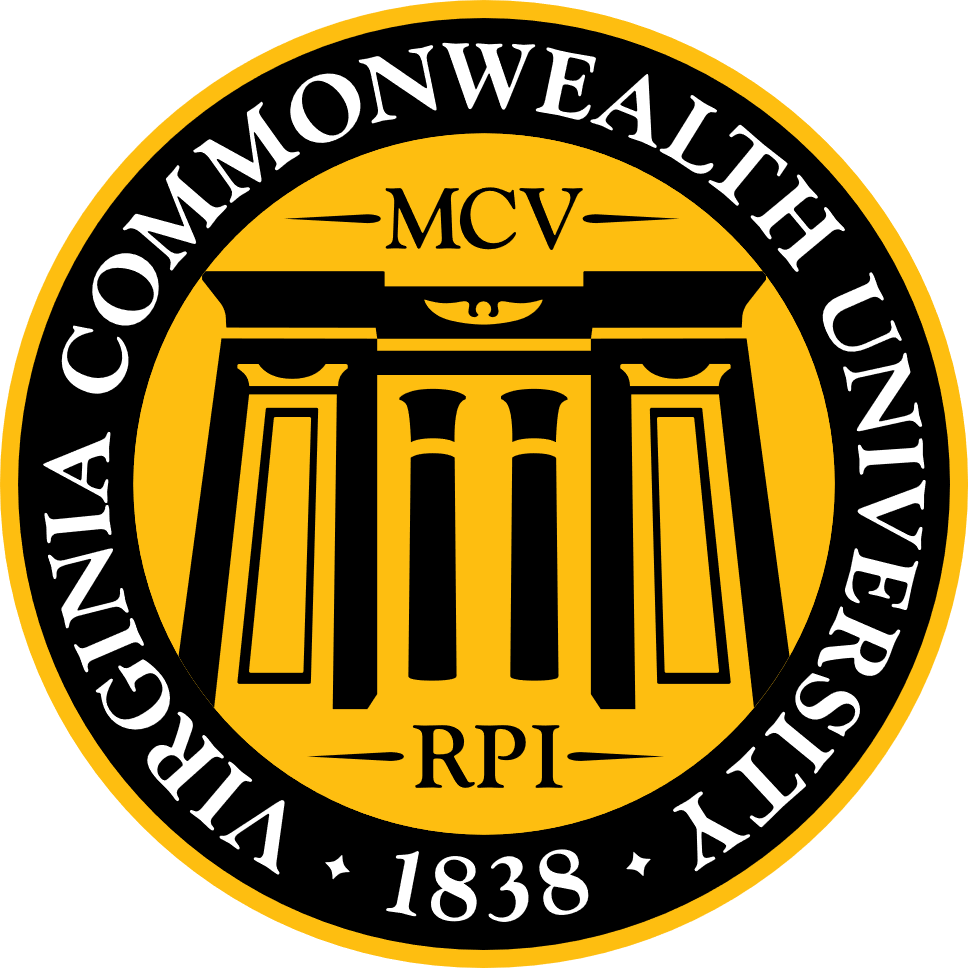
****

**VIRGINIA COMMONWEALTH UNIVERSITY**

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

**A1b: IPL Dataset Analysis of – R/Python (Krunal Pandya)**

**ARVIND SRINIVASAN**

**V01107251**

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

**CONTENTS**

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

**Introduction**

The Indian Premier League (IPL) is a professional Twenty20 cricket league in India, known for its high-profile athletes, competitive matches, and significant financial investments. Player performance is a crucial aspect of the league, influencing team success and individual market value. This report examines the correlation between player performance metrics—specifically, runs scored and wickets taken—and their corresponding salaries for the year 2024. Using detailed ball-by-ball match data and salary information, this study aims to understand how well player compensation aligns with their on-field contributions. Additionally, a case study on Krunal Pandya is included to illustrate specific trends and insights.

**Results**

The analysis was conducted in several key steps, yielding the following results:

1. **Data Preparation:**
   * IPL ball-by-ball data and salary data for 2024 were loaded into the R environment.
   * Total runs scored and wickets taken by each player were calculated for each match.
2. **Data Aggregation:**
   * Performance data were aggregated by year and player, focusing on the 2024 season.
   * A combined performance metric (total performance) was created by summing runs scored and wickets taken.
3. **Fuzzy Name Matching:**
   * To align player names across performance and salary datasets, fuzzy matching techniques were applied using the Jaro-Winkler string distance method.
   * This method ensured that minor discrepancies in player names between datasets were resolved.
4. **Correlation Analysis:**
   * The merged dataset included player salaries and their corresponding total performance.
   * The correlation coefficient between player salaries and total performance for the year 2024 was found to be approximately 0.466.
5. **Specific Player Analysis:**
   * Detailed performance metrics and salary data were analyzed for Krunal Pandya, revealing a total performance score of 79 and a salary of ₹825.

**Interpretations**

The correlation coefficient of 0.466 indicates a moderate positive relationship between player salaries and their total performance. This suggests that while there is a noticeable connection between performance and salary, other factors also significantly influence salary decisions. Below are the key interpretations derived from the analysis:

1. **Moderate Correlation:**
   * A correlation of 0.466 implies that higher salaries are somewhat associated with better on-field performance. However, this relationship is not strong enough to suggest that performance is the sole determinant of salary. Teams likely consider multiple factors when negotiating player contracts.
2. **Performance Metrics:**
   * The aggregated performance metric, combining runs scored and wickets taken, provides a holistic view of a player’s contribution. However, the relative importance of runs versus wickets may vary based on team strategy and game context. Further refinement could involve weighting these components differently.
3. **Market Influences:**
   * The moderate correlation suggests that market dynamics, such as player branding, fan following, and historical performance, play a significant role in determining salaries. For instance, well-known players with a strong fan base may command higher salaries despite average recent performance due to their commercial value and ability to draw crowds.
4. **Case Study - Krunal Pandya:**
   * Krunal Pandya’s analysis revealed a total performance of 79 and a salary of ₹825. This disparity highlights the potential overvaluation of certain players based on factors other than recent on-field performance. Teams may prioritize experience, leadership qualities, or specific skill sets that are not fully captured by the performance metrics used in this analysis.
5. **Other Considerations:**
   * Player fitness, injury history, and versatility (ability to perform in multiple roles) are critical factors influencing salary but are not directly accounted for in the performance metrics. Including such variables could improve the accuracy of salary-performance correlation studies.

**Recommendations**

Based on the findings, several recommendations can be made for IPL teams and stakeholders:

1. **Refine Salary Models:**
   * Develop more sophisticated salary models that incorporate advanced performance metrics, predictive analytics, and market factors. This would ensure a fairer alignment between player compensation and expected contributions.
2. **Holistic Evaluation:**
   * Consider a comprehensive evaluation approach that includes on-field performance, fitness, versatility, leadership qualities, and marketability in salary negotiations. This would provide a more balanced view of a player’s value to the team.
3. **Market Dynamics Understanding:**
   * Acknowledge and strategically leverage market dynamics and player branding. While performance is crucial, commercial value driven by fan engagement and media presence should also be factored into salary decisions.
4. **Continuous Monitoring:**
   * Regularly update performance metrics and salary data to reflect current trends, player development, and emerging talents. This dynamic approach would help in maintaining a responsive and fair salary structure.

**References**

1. Indian Premier League Official Website. (2024). Retrieved from [IPL Official Website](https://www.iplt20.com/)
2. R Documentation. (2024). Retrieved from [R Documentation](https://www.rdocumentation.org/)
3. StringDist Package Documentation. (2024). Retrieved from StringDist Documentation