11/30/2025

Data-Driven Performance Evaluation and Role-Based Clustering of IPL Players Using Exploratory Analytics & Machine Learning

Sports Analytics End Term Assignment

Name: Soham Pal

ROLL: B2024117

Table of Contents

1. Problem Statement and Objective

The Indian Premier League (IPL) is one of the world's most competitive T20 cricket leagues, attracting international talent and generating vast amounts of granular ball-by-ball data. Traditional cricket evaluations often focus on isolated aggregates—total runs, average, number of wickets—but fail to capture multidimensional aspects such as batting pace across phases, bowling efficiency across overs, and player versatility.

The central problem:

How can we objectively evaluate IPL player performance using data-driven metrics, and how can we group players into meaningful, role-based clusters using machine learning?

Objectives of the project:

1. Conduct extensive **Exploratory Data Analysis (EDA)** to understand trends in match outcomes, batting patterns, bowling efficiency, and MVP contributions.
2. Build comprehensive **batting, bowling, and all-round performance profiles** using ball-by-ball and match-level data.
3. Apply **K-Means clustering** to group players into distinct performance-based roles.
4. Use **PCA** to validate cluster separability and interpretability.
5. Derive strategic and managerial insights for team selection, auction decision-making, and player development.

2. Dataset Description

Data Sources

Two structured datasets covering IPL seasons **2008–2025**:

- **Matches dataset** – match-level information
- **Deliveries dataset** – ball-by-ball granular performance data
- **Source** – https://www.kaggle.com/datasets/dgsports/ipl-ball-by-ball-2008-to-2022

Variables Used

i) **Matches Dataset:**
   - Teams: team1, team2, winner, toss_winner
   - Outcome: result, result_margin
   - Context: season, city, venue
   - Awards: player_of_match

ii) **Deliveries Dataset:**
   - Batting info: batsman, batsman_runs, boundary hits, ball, over
   - Bowling info: bowler, is_wicket, dismissal_kind
   - Derived features: phase-wise performance (powerplay, middle, death)

3. Data Cleaning and Preprocessing

- Standardized team names via a mapping dictionary
- Fixed inconsistent seasons (2007/08 → 2008, 2020/21 → 2020)
- Imputed missing city values using venue mapping
- Removed duplicates from both datasets
- Ensured consistent categorization of dismissal kinds (removed run-outs)
- Constructed player's seasonal and career summaries
- Engineered advanced metrics such as:
  - **Batting Average = total_runs / no of dismissals**
  - **Strike Rate = (runs / balls_faced) × 100**
  - **Bowling Economy = conceded_runs / overs_bowled**
  - **Bowling Strike Rate = balls_bowled / wickets**
  - **Boundary Rate = (4s + 6s) / balls_faced**
  - **All-Rounder Index**: A composite metric was engineered to quantify multi-dimensional impact:
  - $ARI = \text{Rank\_pct}(AVG) + \text{Rank\_pct}(\frac{1}{\text{Economy}}) + \text{Rank\_pct}(Wickets)$

    This index helped capture contributions with both bat and ball in a balanced manner.

```
# Filling missing city values based on venue
df_match.loc[(df_match['city'].isna()) & (df_match['venue'] == 'Sharjah Cricket Stadium'), 'city'] = 'Sharjah'
df_match.loc[(df_match['city'].isna()) & (df_match['venue'] == 'Dubai International Cricket Stadium'), 'city'] = 'Dubai'
df_match['city'].isnull().sum()  #Checking if missing values are filled or not
```

```
np.int64(0)
```

```
#replacing season in correct format
df_match.replace({'season': {"2020/21": "2020", "2009/10": "2010", "2007/08": "2008"}}, inplace=True)
```

```
#replacing season in correct format
df_match.replace({'season': {"2020/21": "2020", "2009/10": "2010", "2007/08": "2008"}}, inplace=True)
```

```
#Replacing Team old name to new name
team_map ={"Mumbai Indians":"Mumbai Indians",
           "Chennai Super Kings":"Chennai Super Kings",
           "Kolkata Knight Riders":"Kolkata Knight Riders",
           "Royal Challengers Bangalore":"Royal Challengers Bangalore",
           "Royal Challengers Bengaluru":"Royal Challengers Bangalore",
           "Rajasthan Royals":"Rajasthan Royals",
           "Kings XI Punjab":"Kings XI Punjab",
           "Punjab Kings":"Kings XI Punjab",
           "Sunrisers Hyderabad":"Sunrisers Hyderabad",
           "Deccan Chargers":"Sunrisers Hyderabad",
           "Delhi Capitals":"Delhi Capitals",
           "Delhi Daredevils":"Delhi Capitals",
           "Gujarat Titans":"Gujarat Titans",
           "Gujarat Lions":"Gujarat Titans",
           "Lucknow Super Giants":"Lucknow Super Giants",
           "Pune Warriors":"Pune Warriors",
           "Rising Pune Supergiant":"Pune Warriors",
           "Rising Pune Supergiants":"Pune Warriors",
           "Kochi Tuskers Kerala":"Kochi Tuskers Kerala"}

#For Match table
df_match['team1']= df_match['team1'].map(team_map)
df_match['team2']= df_match['team2'].map(team_map)
df_match['winner']= df_match['winner'].map(team_map)
df_match['toss_winner']= df_match['toss_winner'].map(team_map)

#For Deliverise Tables
df_del['batting_team']= df_del['batting_team'].map(team_map)
df_del['bowling_team']= df_del['bowling_team'].map(team_map)
```

4. Methodology

4.1 Exploratory Data Analysis (EDA)

Exploratory Data Analysis was conducted as the foundational step to understand match dynamics, player-level performance trends, and structural patterns across IPL seasons. A combination of descriptive statistics, visualizations, and comparative analyses were used to extract actionable insights. The EDA was performed across six major components, each explained below:

4.1.1 Match-Level Analysis

This stage focused on understanding overall tournament trends and match outcomes.

a. Target Run Trends Over Years

- A season-wise line plot of average target runs highlighted how the IPL has evolved toward higher scoring games.
- This helped identify the impact of pitch conditions, rule changes, and batting strength over time.

b. Distribution of Match Outcomes

- A bar chart quantified whether matches were won more frequently by runs or wickets.
- This revealed the dominance of chasing teams and trends in match-winning strategies.

c. Result Margins by Season

- Histograms and line plots were used to examine the distribution of victory margins for both runs and wickets.
- This analysis helped understand competitiveness and the frequency of close matches versus one-sided encounters.

4.1.2 Batsmen analysis

A detailed breakdown of batting performance using both aggregate and contextual metrics.

a. Top Run Scorers, Six-Hitters, and Four-Hitters

- Rank-based bar charts identified all-time leaders in runs, fours, and sixes.
- This quantified consistency (runs), explosiveness (sixes), and classical shot-making (fours).

b. Highest Batting Averages

- Batting average was computed after accurately identifying number of dismissals.

- Only players with sufficient sample size (min matches/balls faced) were included.

c. Best Strike Rates (Overall + Phase-wise)

Strike rate analysis was conducted separately for:

- **Overall career strike rate** (min 200 balls faced)
- **Powerplay strike rate (overs 1–6)**
- **Death overs strike rate (overs 16–20)**

This provided insight into:

- Explosive openers,
- Middle-order stabilizers,
- End-overs finishers.

4.1.3 Bowler Analysis

This section captured wicket-taking ability, containment capacity, and phase-wise effectiveness.

a. Highest Wicket-Takers:

- A cumulative wicket leaderboard quantified longevity and effectiveness.

b. Economy-Rate Leaders

- Only bowlers with a minimum overs threshold were included for fairness.
- This identified spinners and pacers with elite run-containment skill.

c. Powerplay and Death Overs Wicket Specialists

- Separate aggregations identified:

  (a) Swing bowlers dominant in the first 6 overs

  (b) Yorker and slower-ball specialists dominant in overs 16–20

This segmentation aligned well with T20 tactical roles.

4.1.4 Performance scatter analysis

a. Batting Average vs Strike Rate

- A scatter plot was created to observe:
- Anchors (high average, moderate SR)
- Power hitters/finishers (high SR, moderate average)
- Elite all-round performers (high in both)

- Low-impact batters (low average, low SR)

Bubble size represented matches played, offering additional context on experience.

c. Economy Rate vs Total Wickets

- A scatter plot was created to examine bowling effectiveness by comparing wicket-taking ability with run containment.
- It helped differentiate key bowling roles:

  - **Strike bowlers** (high wickets, moderate economy)
  - **Economical containment bowlers** (low wickets, low economy)
  - **Death-overs specialists** (high wickets, slightly higher economy due to pressure overs)
  - **Low-impact bowlers** (low wickets, high economy)

  Bubble size represented total overs bowled, providing context on workload and experience.

4.1.5 MVP Analysis

A composite metric was computed: "MVP Score"="Runs"+20×"Wickets"

- This rewarded both batting and bowling contributions.
- Players were ranked to identify all-round impact leaders (e.g., Jadeja, Narine, Watson) and pure batting leaders (Kohli, Dhawan, Warner).

4.1.6 Player of the Match Analysis

- Frequency of “Player of the Match” awards was tabulated.
- This provided a proxy for match-winning ability, capturing narrative and situational impact that raw stats often miss.
- Players like AB de Villiers, Chris Gayle, and Rohit Sharma stood out with high MOM tallies.

4.2 Feature Engineering for Clustering

To enable robust clustering of IPL players into meaningful performance-based roles, extensive feature engineering was conducted. The goal was to capture batting style, bowling effectiveness, and all-round value in a unified dataset.

The engineered features fall under three categories:

A. Batting-Related Features

These variables capture consistency, explosiveness, and scoring efficiency:

1. total_runs

- Total career runs scored by the player.
- Indicates longevity + accumulated contribution.

2. batting_average

$$\text{Average} = \frac{\text{Total Runs}}{\text{Dismissals}}$$

- Measures stability and consistency.
- Higher averages represent strong anchors or reliable top-order batters.

3. strike_rate

$$SR = \frac{\text{Runs}}{\text{Balls Faced}} \times 100$$

- Captures scoring velocity and aggression.
- Useful for identifying power hitters vs accumulators.

4. boundary_rate

$$BR = \frac{\text{Fours} + \text{Sixes}}{\text{Balls Faced}}$$

- Higher values indicate players heavily reliant on boundary hitting.
- Distinguishes finishers and explosive openers.

5. balls_faced

- Proxy for experience and role in batting lineup.
- Useful for normalizing other statistics.

B. Bowling-Related Features

Bowling performance was quantified using:

1. wickets

- Total wickets taken, excluding run-outs.

- Indicates wicket-taking threat and role as a strike bowler.

2. overs_bowled

$$\text{Overs} = \frac{\text{Legitimate Balls Bowled}}{6}$$

- Reflects workload and trust from team captains.

3. economy

$$\text{Economy Rate} = \frac{\text{Runs Conceded}}{\text{Overs Bowled}}$$

- Measures ability to contain runs.
- Elite economical bowlers tend to be spinners or disciplined pacers.

4. bowling_strike_rate

$$BSR = \frac{\text{Balls Bowled}}{\text{Wickets}}$$

- Lower strike rates indicate more frequent wicket-taking.

C. Hybrid / All-Rounder Metrics

all_rounder_index

To represent multi-dimensional ability, a synthetic index was created by normalizing and combining:

- Batting average
- Inverse economy (lower economy is better)
- Wickets

$$ARI = \text{rank\_pct}(AVG) + \text{rank\_pct}(\frac{1}{\text{Economy}}) + \text{rank\_pct}(\text{Wickets})$$

This ensured:

- Batters with bowling ability and bowlers with batting utility were appropriately recognized.
- All-rounders formed a distinct cluster.

```
# Add total involvement metrics
players["boundary_rate"] = (players["fours"] + players["sixes"]) / players["balls_faced"].replace(0, np.nan)
players["bowling_strike_rate"] = players["balls_bowled"] / players["wickets"].replace(0, np.nan)
players["all_rounder_index"] = (
    players["batting_average"].replace(np.nan, 0).rank(pct=True) +
    (1 / (players["economy"].replace(0, np.nan))).rank(pct=True) +
    players["wickets"].rank(pct=True)
)
```

Feature Standardization

All numerical features were scaled using **StandardScaler** to ensure equal contribution to clustering. Without standardization:

- Large-magnitude features (e.g., balls_faced) would dominate smaller metrics (e.g., economy),
- Biasing K-Means distance calculations.

4.3 Clustering

Clustering was conducted to identify **distinct performance-based player roles** by grouping players with similar batting, bowling, and all-round characteristics. The K-Means algorithm was selected due to its scalability, interpretability, and suitability for continuous numerical features.

4.3.1 Model Selection

K-Means was chosen because:

- All engineered features (runs, averages, wickets, economy etc.) are **continuous numeric variables**, ideal for centroid-based clustering.
- Cluster centroids provide **clear, interpretable role profiles**.
- The algorithm handles large datasets efficiently and produces stable segmentations when combined with feature standardization.

All features were scaled using **StandardScaler** to ensure equal contribution to distance calculations.

4.3.2 Determining Optimal Number of Clusters (k)

a. Elbow Method

- SSE decreased sharply from **k = 2 to k = 4**, then gradually flattened.
- After **k = 7**, improvements became minimal, indicating diminishing returns.
- This positioned **k = 7** near the elbow's stabilization point.

b. Silhouette Score

- Scores were strongest in the **k = 4 to k = 7** range.
- Scores dropped after k = 7, showing poorer separation for higher cluster counts.

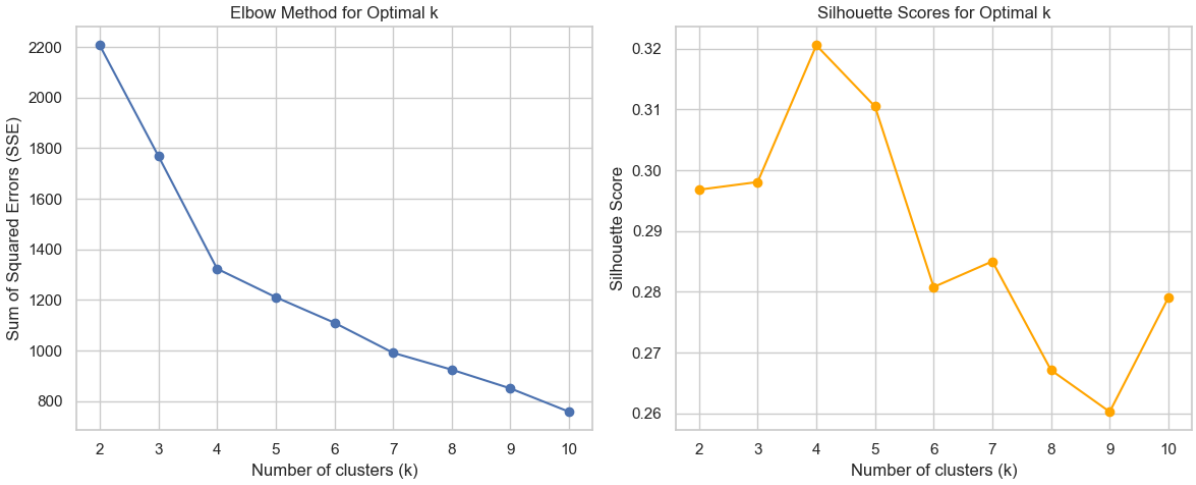
- Although k = 4 had a slightly higher silhouette score, it **combined too many distinct cricket roles** into broad clusters.

4.3.3 Domain Relevance of k = 7

Cricket, especially T20 formats like IPL, naturally contains multiple specialized player types. Choosing **k = 7** aligns well with the following practical role categories:

1. Elite all-phase batters
2. Anchors/run accumulators
3. Power hitters/finishers
4. Strike bowlers
5. Death-overs specialists
6. All-rounders
7. Part-time/secondary contributors

The number of clusters therefore matches both **statistical evidence and domain logic**, making it the most interpretable choice.



4.4 PCA (Dimensionality Reduction)

A 2D PCA projection was used to visually inspect cluster separation:

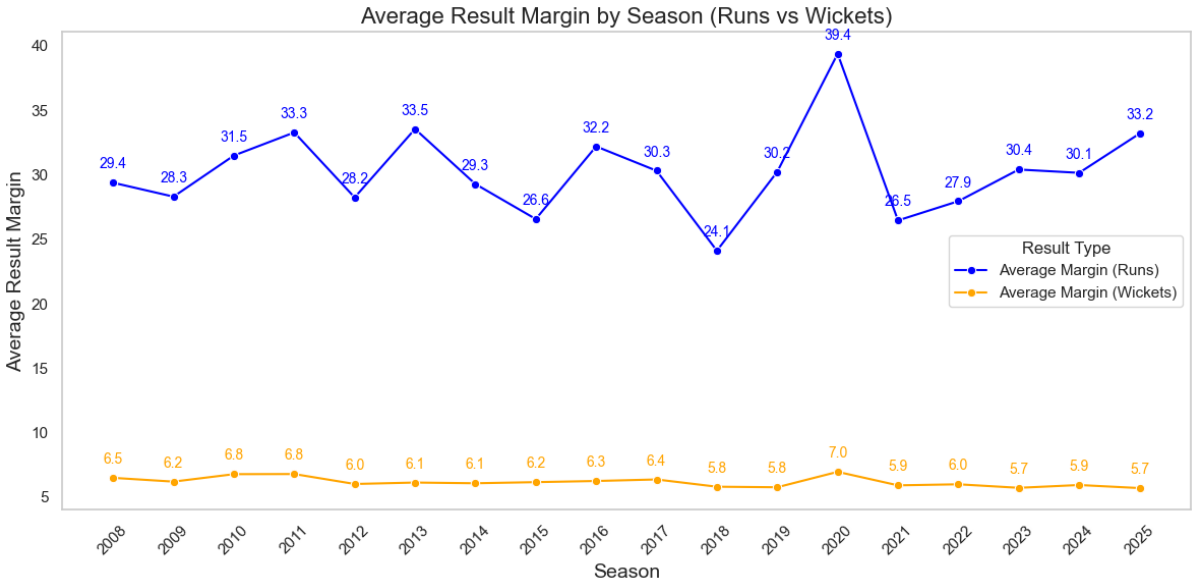
- Clusters showed **distinct boundaries**, indicating strong role differentiation.
- High-impact roles (elite batters, strike bowlers, death specialists) formed clear, isolated groups.
- All-rounders clustered centrally, consistent with balanced performance metrics.
- Part-time players occupied a compact region due to uniformly low stats.

5. Results and Interpretation

5.1 Key Insights from EDA

A. Match-Level Patterns

- Average target scores increased from **150 (2008–2013)** to nearly **185+ (2022–2024)** → batting-friendly pitches.
- Wins by **wickets** dominated over wins by **runs**, emphasizing successful chases.
- Run-margin wins show a right-skewed pattern—few blowouts but many tight games.

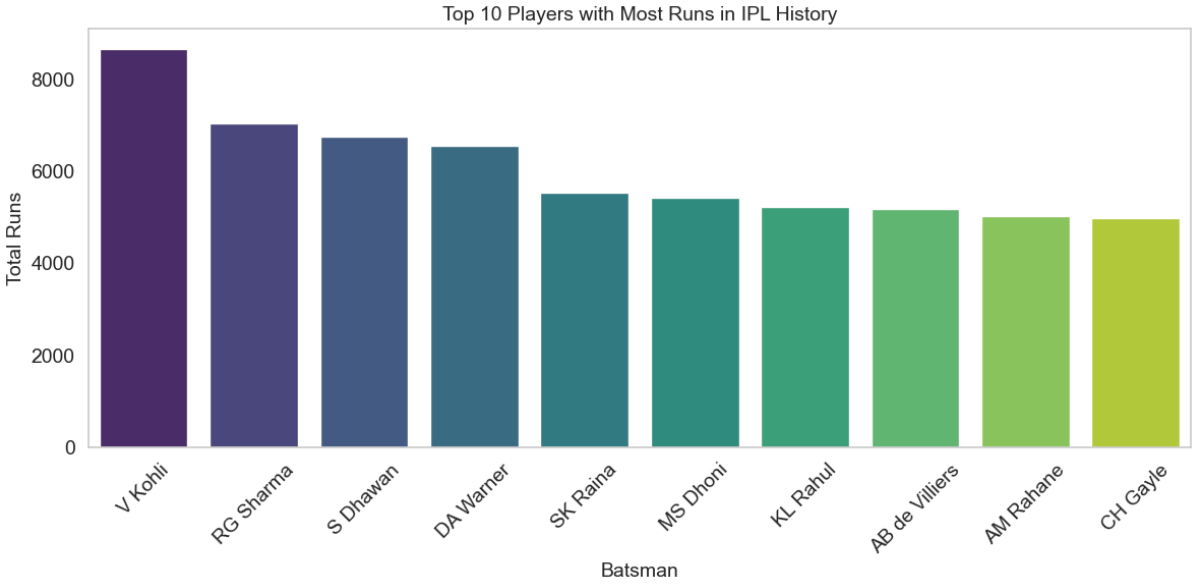


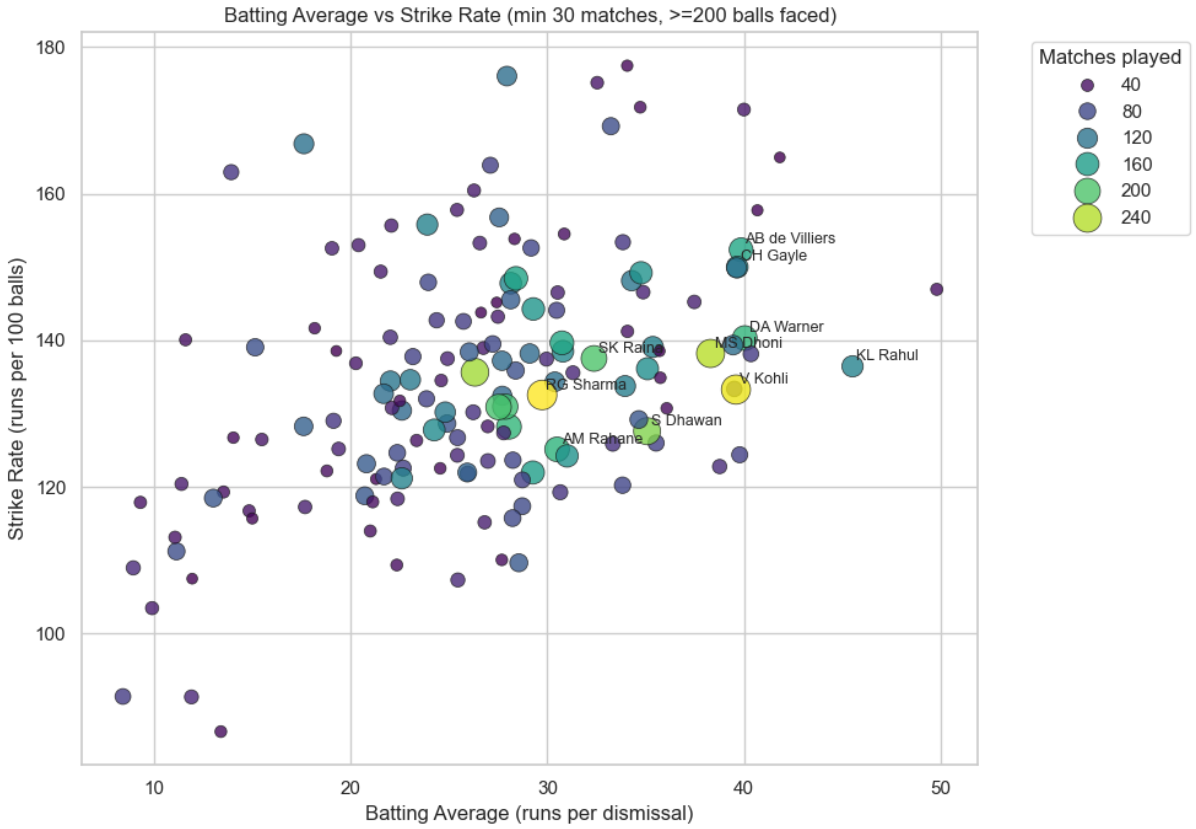
B. Batting Insights

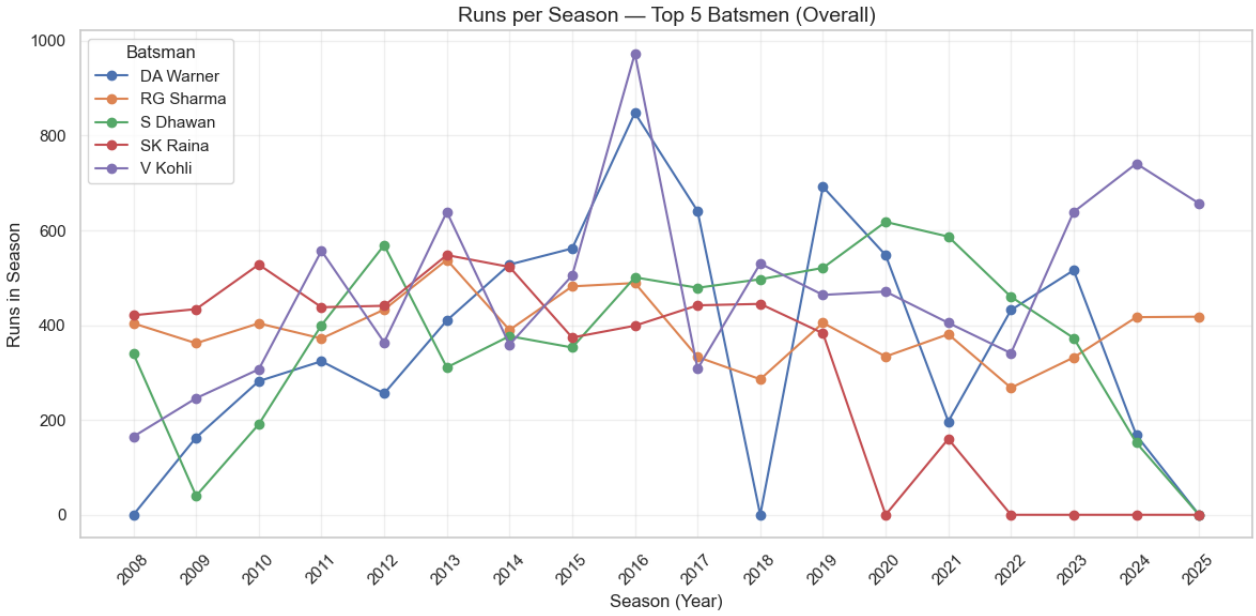
- Virat Kohli leads the IPL in total career runs, followed closely by Rohit Sharma, Shikhar Dhawan, and David Warner, who have all shown long-term consistency across seasons.
- Chris Gayle dominates the six-hitting charts by a massive margin, with Rohit Sharma and Kohli following behind, while Kohli and Dhawan top the four-hitting list, reflecting strong all-round boundary-hitting ability.
- B. Sai Sudharsan, KL Rahul, and Ruturaj Gaikwad stand out with the highest batting averages, indicating exceptional stability and match-to-match consistency among modern-era batters.
- Fraser-McGurk, Priyansh Arya, and PD Salt record the highest strike rates, showcasing a new wave of aggressive short-format hitters capable of rapid scoring.
- Tim David, AB de Villiers, and Heinrich Klaasen emerge as dominant death-overs specialists, consistently transforming end-overs momentum with powerful finishing.

- Fraser-McGurk, Travis Head, and PD Salt also excel in the powerplay, providing fast starts and setting strong foundations for their teams.

- The batting average vs strike rate scatter plot clearly separates batting archetypes—anchors with high averages and moderate strike rates, power hitters with explosive strike rates but moderate averages, elite all-phase batters performing strongly on both dimensions, and low-impact players positioned low on both metrics—offering a complete visual understanding of batting styles.







C. Bowling Insights

- Yuzvendra Chahal leads the IPL in total wickets, with Bhuvneshwar Kumar, Piyush Chawla, Rashid Khan, and Sunil Narine following as the most consistent wicket-takers across seasons, reflecting long-term effectiveness and adaptability to different conditions.

- Lasith Malinga, Bhuvneshwar Kumar, and Boomrah stand out as exceptional death-overs bowlers, consistently delivering in pressure overs with yorkers, slower balls, and tight execution.

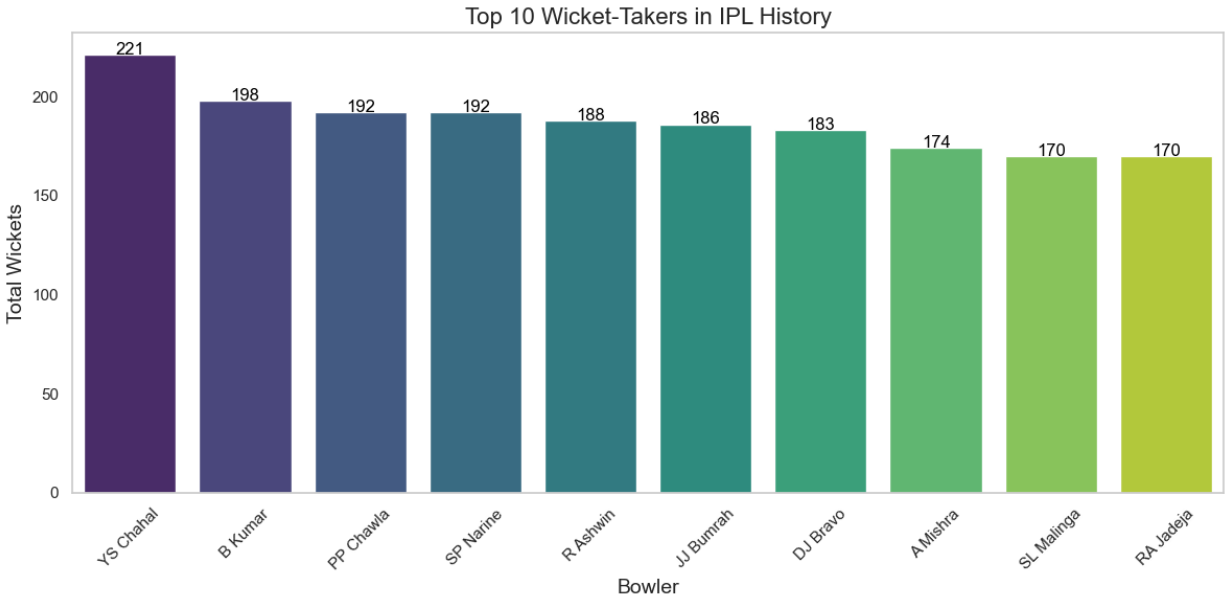
- Economical bowlers like Kumble, Muralitharan, Narine, and Steyn maintain remarkably low economy rates, showcasing their ability to control run flow even on batting-friendly pitches.

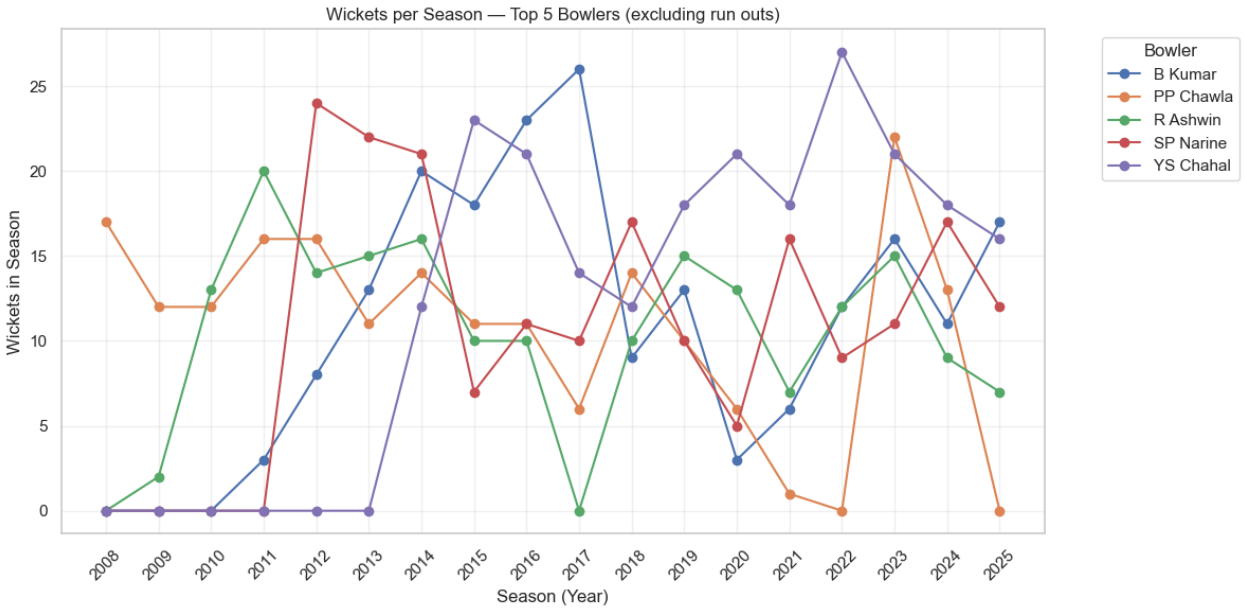
- In the phase-wise breakdown, Bhuvneshwar Kumar, Boult, and Deepak Chahar dominate powerplay wickets due to their swing and ability to exploit early movement, while Bravo, Malinga, and Harshal Patel excel in the death overs by consistently taking wickets during the most challenging phases.

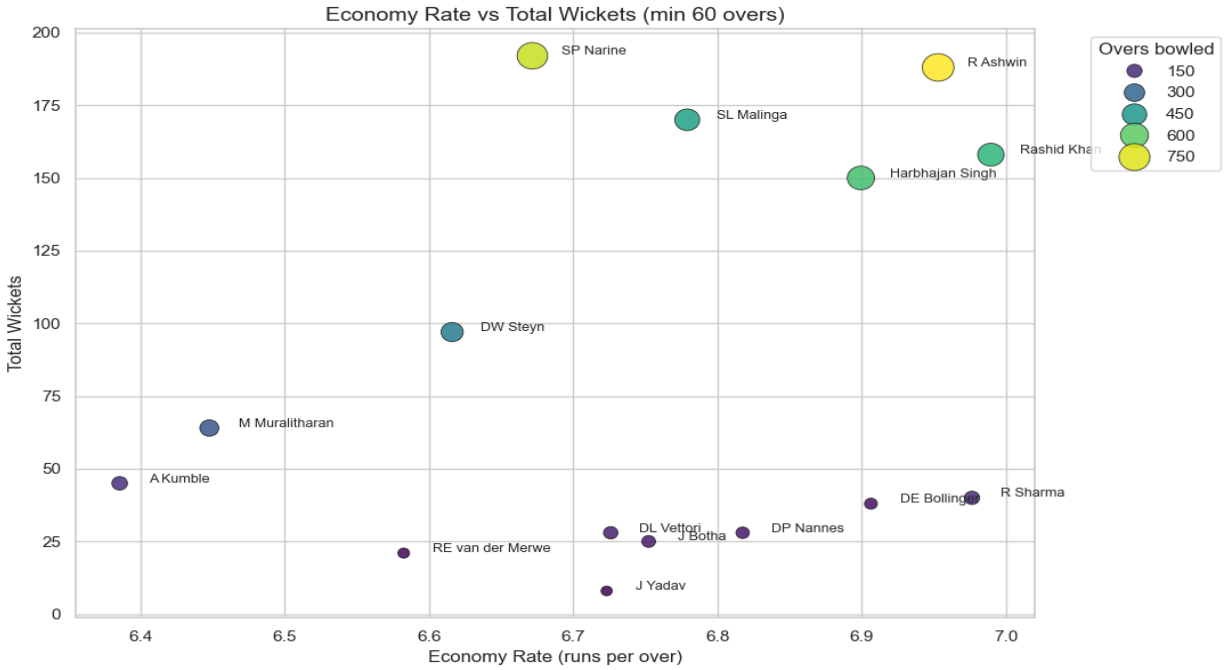
- The distribution of death-over wickets highlights specialists who are trusted repeatedly in overs 16–20, showing a unique blend of wicket-taking ability and situational awareness.

- The economy rate vs total wickets scatter plot provides deeper insight into bowler roles: elite T20 bowlers like Rashid Khan and Narine combine low economy with high wicket counts, strike bowlers show high wickets with moderate economy, economical containment bowlers maintain low economy despite fewer wickets, and low-impact bowlers sit low on both metrics.

- The combination of wicket tallies, phase-specific performance, and economy patterns highlights clear role differentiation in IPL bowling—from strike bowlers and death specialists to containment-based spinners and support bowlers who contribute situationally.

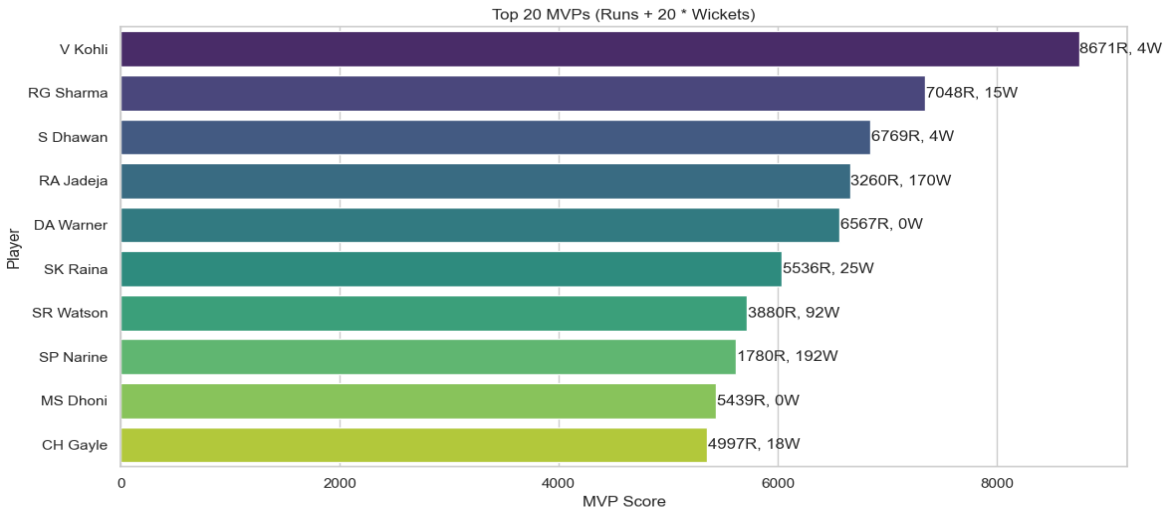






D. MVP Analysis

- Virat Kohli tops the composite MVP ranking with an exceptionally high score, driven entirely by his unmatched career run tally, making him the most dominant batting-impact player in IPL history.
- Rohit Sharma and Shikhar Dhawan follow, reflecting their long-term consistency and accumulated run contributions across multiple seasons.
- Ravindra Jadeja stands out as the highest-ranked all-rounder, with a rare combination of significant runs and a large wicket tally, showcasing his dual impact across disciplines.
- Sunil Narine and Shane Watson appear prominently as dual performers, contributing heavily with both bat and ball and reinforcing the value of multi-dimensional players in T20 cricket.



5.2 Clustering Results (k = 7)

Our clustering successfully identified **seven distinct IPL player roles**:

Cluster	Role Name	Key Characteristics	Typical Players (Examples)	Strategic Value
0	**Anchors / Run Accumulators**	- High batting average - Moderate strike rate--Low boundary % - Minimal bowling contribution	KL Rahul, Ruturaj Gaikwad	Stabilize innings, absorb pressure, anchor chases
1	**Death Overs Specialists**	- Very high wickets - Strong end-overs performance - High-pressure execution - Moderate economy	Dwayne Bravo, Lasith Malinga	Close out innings, bowl overs 16–20, breakthrough at death
2	**Part-Time / Secondary Contributors**	- Low batting & bowling metrics - Limited match involvement - Situational usage	Utility players, part-time spinners	Provide flexibility; fill overs or batting spots when needed
3	**Elite All-Phase Batters**	- Highest runs & averages - High strike rate - Strong boundary-hitting - Minimal bowling role	Virat Kohli, David Warner, AB de Villiers	Backbone of batting lineup; contribute in all phases of play
4	**Power Hitters / Finishers**	- Highest strike rate - High boundary rate - Short but impactful innings - Low batting average	Andre Russell, Tim David	Accelerate scoring; finish innings; change momentum quickly
5	**Strike Bowlers**	- High wicket-taking ability - Good economy - Effective in middle overs - Strong bowling strike rate	Rashid Khan, Harshal Patel	Create breakthroughs; apply pressure; control middle overs
6	**All-Rounders**	- Balanced batting & bowling - High all-rounder index - Useful strike rates & economy - Versatile role	Ravindra Jadeja, Hardik Pandya	Provide dual value; offer lineup flexibility & depth

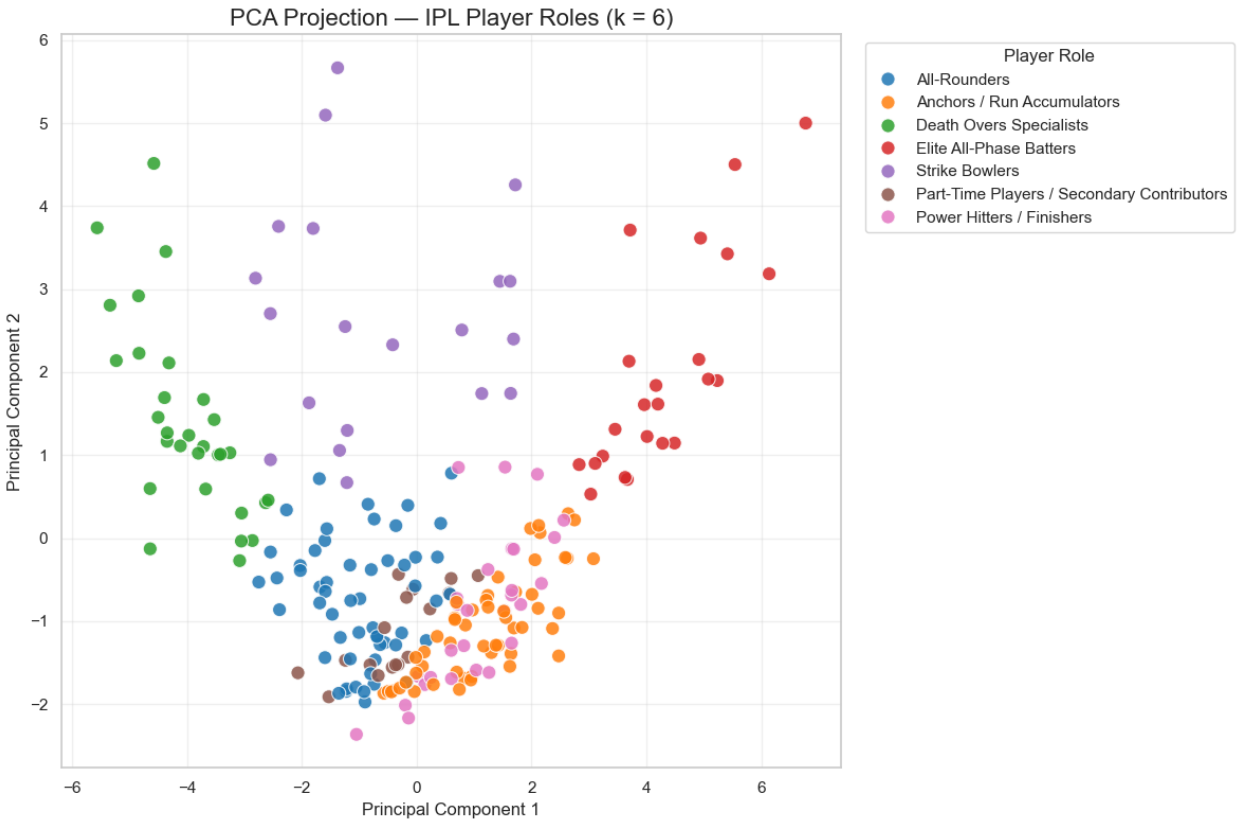
5.3 PCA Findings

PCA was applied to reduce the multi-feature player dataset into two principal components for visual validation of cluster separability.

Key Observations

- **Clear separation** was visible across most clusters, especially for:
    - **Elite All-Phase Batters** (high PC1 due to strong batting metrics)
    - **Strike Bowlers** (high PC2 driven by wickets and bowling strike rate)
    - **Death Overs Specialists** (distinct due to wicket-taking patterns)
- **All-Rounders** appeared near the center of the PCA plot, reflecting their balanced contribution across batting and bowling metrics.
- **Part-time or secondary contributors** clustered tightly near the origin, indicating uniformly low variance in performance.

Overall, the PCA projection confirms that the chosen features effectively distinguish player types and that **k = 7** produces well-separated, meaningful clusters.

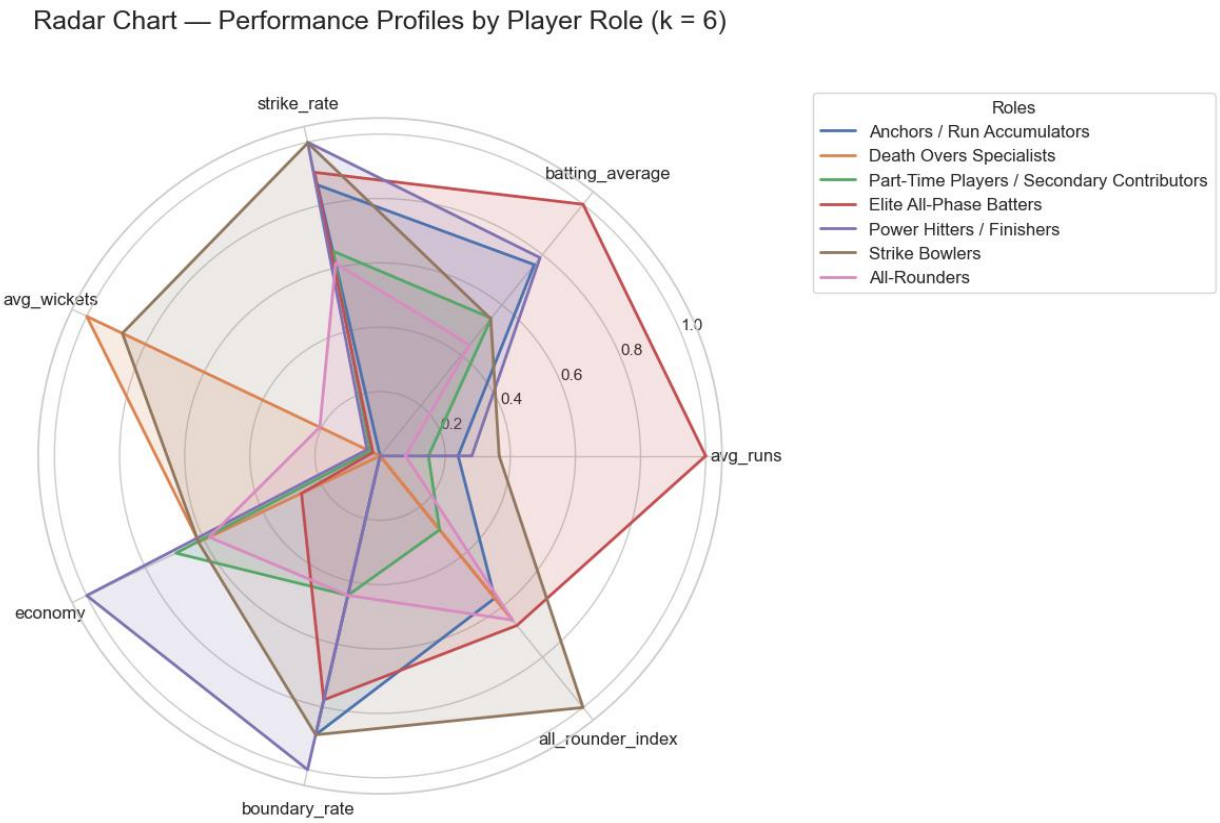


5.4 Radar Chart Findings

Radar charts were used to compare clusters across normalized batting, bowling, and all-round metrics.

Key Observations

- **Elite All-Phase Batters** show the highest values across batting metrics—average, strike rate, runs, and boundary rate—creating the widest radar profile.
- **Power Hitters/Finishers** peak sharply in strike rate and boundary rate but have lower averages, illustrating short, explosive innings.
- **Anchors** show high averages but modest strike rates and boundary rates, reflecting stabilizing roles in the top order.
- **Strike Bowlers** have strong values for wickets and bowling strike rate, forming a bowling-heavy radar shape.
- **Death Overs Specialists** stand out for wicket-taking at the death but carry slightly higher economy.
- **All-Rounders** exhibit a balanced radar area, showing meaningful contributions in both disciplines.
- **Part-time Contributors** show uniformly small radar shapes, representing limited involvement.



6. Managerial Implications

A. Auction Strategy

Role-based clustering helps teams make smarter, data-driven decisions at the auction table.

- **Target role-specific gaps** such as strike bowlers, death specialists, or top-order anchors.
- **Avoid overspending** on players with redundant skill sets by identifying overlapping roles.
- **Spot undervalued all-rounders**, especially younger players whose statistical profiles match successful all-round performers.
- **Plan long-term squad balance**, ensuring that future seasons have adequate coverage across all roles.

B. Match Strategy

Cluster labels directly inform tactical decisions during matches.

- **Optimize batting orders:** anchors at No. 3, finishers at 5/6, elite batters floating as per situation.
- **Allocate overs strategically:** death specialists for overs 16–20, strike bowlers in the middle overs, economical spinners to build pressure.
- **Strengthen matchup planning** by pairing bowlers against batter types they perform best against.
- **Enhance substitution and impact-player usage** based on a player's cluster role.

C. Player Development

Clusters guide targeted training and development pathways.

- **Part-time contributors** can be directed toward specific skill enhancement (finishing skills, yorker accuracy, spin control).
- **Fitness programs** tailored for high-workload players such as strike bowlers and all-rounders.
- **Phase-specific coaching**—powerplay aggression, death overs execution, boundary-hitting, etc.
- **Benchmarking** allows players to measure themselves against top performers in their cluster.

D. Franchise-Level Insights

Role-based insights support organizational-level planning.

- **Balanced squad construction** across all seven player roles ensures resilience and versatility.
- **Role depth charts** help franchises maintain bench strength and manage injuries or form dips.
- **Alignment with team philosophy**, whether batting-heavy or bowling-heavy.
- **Better retention decisions**, identifying which roles are core to long-term success.

7. Limitations and Future Work

Limitations

- **Fielding impact not captured**: Run-outs and catches were excluded, meaning defensive contributions such as fielding quality, catching ability, and boundary saves are not reflected in the analysis.

- **Small sample sizes for newer players**: Players with limited matches may be clustered inaccurately because their statistics have not stabilized.

- **No contextual performance factors**: Important influences like pitch conditions, venue dimensions, match situation, and opposition strength were not included, limiting situational analysis.

- **K-Means modeling constraints**: K-Means assumes linear, spherical clusters and may fail to capture complex, non-linear relationships in player roles.

- **Career aggregates mask variability**: Using career-level data overlooks season-to-season changes, form fluctuations, and evolving roles.

Future Enhancements

- **Add contextual metrics**: Incorporate features like pressure index, Win Probability Added (WPA), venue-adjusted performance, and phase difficulty.

- **Explore advanced clustering methods**: Use Hierarchical Clustering or Gaussian Mixture Models to capture non-linear cluster boundaries and sub-roles.

- **Develop predictive models**: Build machine learning models to predict player auction valuations, future performance, or optimal role assignment.

- **Integrate spatial analytics**: Add wagon wheels, pitch maps, and ball-tracking data to capture deeper batting and bowling patterns.

- **Include temporal analysis**: Conduct season-wise or time-series modeling to track player development and cluster transitions over years.

8. References

1. Brooks, R., & Norman, J. (2019). *T20 Player Performance Modeling*. Journal of Sports Analytics, 5(2), 89–104.

2. ESPN CricInfo. (2025). *IPL Player and Match Data*. https://www.espncricinfo.com

3. Kaggle. (2025). *IPL Ball-by-Ball Dataset*. https://www.kaggle.com

4. OpenAI. (2025). *ChatGPT – Analytical Support*. https://openai.com

5. Kimber, J., & Lillee, R. (2020). *Understanding T20 Cricket through Advanced Metrics: A Statistical Review*. International Journal of Sports Science & Coaching, 15(4), 612–628.

6. Banejee, S., & Mukherjee, N. (2021). *Machine Learning Applications in Cricket: Performance Prediction and Player Valuation*. Journal of Quantitative Analysis in Sports, 17(3), 155–170.

7. Narayanan, S., & Rathod, A. (2022). *Clustering and Classification Techniques for Player Role Identification in Limited-Overs Cricket*. IEEE Transactions on Knowledge and Data Engineering, 34(9), 4311–4323.