

VIRGINIA COMMONWEALTH UNIVERSITY

Statistical analysis and modelling (SCMA 632)

A1b: Preliminary preparation and analysis of data- Descriptive statistics

POORNIMA

V01107763

Date of Submission: 18-06-2024

CONTENTS

|  |  |  |
| --- | --- | --- |
| Sl. No. | Title | Page No. |
| 1. | Introduction | 1 |
| 2. | Objectives | 1 |
| 3. | Business Significance | 1 |
| 4. | Results and analysis | 2-5 |
| 5. | Codes | 6-10 |
| 6. | References |  |

IPL PLAYER ANALYSIS USING R AND PYTHON

Introduction

Analyzing IPL player performance using programming languages like R and Python offers a comprehensive understanding of player dynamics, team strategies, and overall league trends. By leveraging statistical techniques and data visualization tools, analysts can uncover valuable insights that aid in decision-making for team management, fantasy league participants, broadcasters, sponsors, and other stakeholders.

Our objectives include finding the relation between the players performance and the salary, finding appropriate distribution for runs scored and wickets takes etc.

OBJECTIVES

* Arrange the data IPL round-wise and batsman, ball, runs, and wickets per player per match. Indicate the top three run-getters and tow three wicket-takers in each IPL round.
* Fit the most appropriate distribution for runs scored and wickets taken by the top three batsmen and bowlers in the lost three IPL tournaments.
* Find the relationship between a player’s performance and the salary he gets in your data.

BUSINESS SIGNIFICANCE

The focus of this study gives several key areas like **Sports Performance Analysis**, Talent Identification and Development, Fan Engagement and Entertainment, Fantasy Leagues and Wagering, Business and Revenue Generation, Academic Research and Innovation, Professional Development and Career Opportunities. In essence, the significance of studying IPL player analysis extends beyond the realm of sports. It intersects with entertainment, business, technology, academia, and professional development, offering multifaceted opportunities for innovation, engagement, and growth.

A)RESULTS AND INTERPRETATION

1) Arrange the data IPL round-wise and batsman, ball, runs, and wickets per player per match. Indicate the top three run-getters and tow three wicket-takers in each IPL round.

top\_run\_getters = player\_runs.groupby('Season').apply(lambda x: x.nlargest(3, 'runs\_scored')).reset\_index(drop=True)

bottom\_wicket\_takers = player\_wickets.groupby('Season').apply(lambda x: x.nlargest(3, 'wicket\_confirmation')).reset\_index(drop=True)

print("Top Three Run Getters:")

print(top\_run\_getters)

print("Top Three Wicket Takers:")

print(bottom\_wicket\_takers)

Result:

Top Three Run Getters:

Season Striker runs\_scored

0 2007/08 SE Marsh 616

1 2007/08 G Gambhir 534

2 2007/08 ST Jayasuriya 514

3 2009 ML Hayden 572

4 2009 AC Gilchrist 495

5 2009 AB de Villiers 465

6 2009/10 SR Tendulkar 618

7 2009/10 JH Kallis 572

8 2009/10 SK Raina 528

42 2022 JC Buttler 863

43 2022 KL Rahul 616

44 2022 Q de Kock 508

45 2023 Shubman Gill 890

46 2023 F du Plessis 730

47 2023 DP Conway 672

48 2024 RD Gaikwad 509

49 2024 V Kohli 500

50 2024 B Sai Sudharsan 418

Top Three Wicket Takers:

Season Bowler wicket\_confirmation

0 2007/08 Sohail Tanvir 24

1 2007/08 IK Pathan 20

2 2007/08 JA Morkel 20

3 2009 RP Singh 26

4 2009 A Kumble 22

5 2009 A Nehra 22

6 2009/10 PP Ojha 22

7 2009/10 A Mishra 20

8 2009/10 Harbhajan Singh 20

39 2021 HV Patel 35

40 2021 Avesh Khan 27

41 2021 JJ Bumrah 22

42 2022 YS Chahal 29

43 2022 PWH de Silva 27

44 2022 K Rabada 23

45 2023 MM Sharma 31

46 2023 Mohammed Shami 28

47 2023 Rashid Khan 28

48 2024 HV Patel 19

49 2024 Mukesh Kumar 15

1. 2024 Arshdeep Singh 14

INTERPRETATION:

 The data highlights the top-performing batsmen and bowlers over various IPL seasons, showcasing players with the most runs and wickets in each season.

 Shubman Gill and HV Patel have delivered standout performances in 2023 and 2021, respectively.

 There is a mix of both seasoned players like V Kohli and emerging talents like Shubman Gill, indicating the dynamic nature of IPL, where both experience and new talent thrive.

 The recurring names among top performers emphasize consistency and excellence in performance across multiple seasons.

2)Fit the most appropriate distribution for runs scored and wickets taken by the top three batsmen and bowlers in the lost three IPL tournaments.

Code:

import scipy.stats as st

def get\_best\_distribution(data):

dist\_names = ['alpha','beta','betaprime','burr12','crystalball',

'dgamma','dweibull','erlang','exponnorm','f','fatiguelife',

'gamma','gengamma','gumbel\_l','johnsonsb','kappa4',

'lognorm','nct','norm','norminvgauss','powernorm','rice',

'recipinvgauss','t','trapz','truncnorm']

dist\_results = []

params = {}

for dist\_name in dist\_names:

dist = getattr(st, dist\_name)

param = dist.fit(data)

params[dist\_name] = param

# Applying the Kolmogorov-Smirnov test

D, p = st.kstest(data, dist\_name, args=param)

print("p value for "+dist\_name+" = "+str(p))

dist\_results.append((dist\_name, p))

# select the best fitted distribution

best\_dist, best\_p = (max(dist\_results, key=lambda item: item[1]))

# store the name of the best fit and its p value

print("\nBest fitting distribution: "+str(best\_dist))

print("Best p value: "+ str(best\_p))

print("Parameters for the best fit: "+ str(params[best\_dist]))

return best\_dist, best\_p, params[best\_dist]

list\_top\_batsman\_last\_three\_year = {}

for i in total\_run\_each\_year["year"].unique()[:3]:

list\_top\_batsman\_last\_three\_year[i] = total\_run\_each\_year[total\_run\_each\_year.year == i][:3]["Striker"].unique().tolist()

import warnings

warnings.filterwarnings('ignore')

runs = ipl\_bbbc.groupby(['Striker','Match id'])[['runs\_scored']].sum().reset\_index()

for key in list\_top\_batsman\_last\_three\_year:

for Striker in list\_top\_batsman\_last\_three\_year[key]:

print("\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*")

print("year:", key, " Batsman:", Striker)

get\_best\_distribution(runs[runs["Striker"] == Striker]["runs\_scored"])

print("\n\n")

|  |
| --- |
|  |
| Season | Striker | runs\_scored |
| 2423 | 2023 | Shubman Gill | 890 |
| 2313 | 2023 | F du Plessis | 730 |
| 2311 | 2023 | DP Conway | 672 |
| 2433 | 2023 | V Kohli | 639 |
| 2443 | 2023 | YBK Jaiswal | 625 |
| ... | ... | ... | ... |
| 2404 | 2023 | RP Meredith | 0 |
| 2372 | 2023 | Mohsin Khan | 0 |
| 2307 | 2023 | DG Nalkande | 0 |
| 2429 | 2023 | TU Deshpande | 0 |
| 2324 | 2023 | Harshit Rana |  |

INTERPRETATION:

The data and its statistical analysis provide a comprehensive view of player performance in the 2023 IPL season:

* Top Performers: Shubman Gill, F du Plessis, DP Conway, V Kohli, and YBK Jaiswal stand out as the leading run-scorers, indicating their crucial roles in their respective teams.
* Lower Scoring Players: Players with 0 runs likely had different primary roles (e.g., bowlers) and their contributions should be evaluated based on their main responsibilities.
* Statistical Insights: By fitting the run data to statistical distributions, we gain insights into the consistency and variability of each top batsman’s performance, aiding in predictive modeling and strategic planning.

Overall, this detailed analysis helps in understanding player contributions, their scoring patterns, and provides valuable information for team strategies and future performance predictions.

1. 3) Fit the most appropriate distribution for runs scored and wickets taken by the player allotted to you.  
     
   # Initialize the dictionary to store top bowlers for each of the last three years

ls\_livingstone\_bat = {}

# Loop through the unique years in the dataset, limited to the last three years

for i in total\_run\_each\_year["year"].unique()[:3]:

# Filter the dataset to include only records for LS Livingstone in the current year

axar\_patel\_data1 = total\_run\_each\_year[(total\_run\_each\_year["year"] == i) & (total\_run\_each\_year["Striker"] == "LS Livingstone")]

# Get the unique list of years where LS Livingstone appears in the filtered dataset

ls\_livingstone\_bat[i] = axar\_patel\_data1["Striker"].unique().tolist()

# Print the dictionary to verify the results

print(ls\_livingstone\_bat)

Top Three Run Getters:

Season Striker runs\_scored

0 2007/08 SE Marsh 616

1 2007/08 G Gambhir 534

2 2007/08 ST Jayasuriya 514

3 2009 ML Hayden 572

4 2009 AC Gilchrist 495

5 2009 AB de Villiers 465

6 2009/10 SR Tendulkar 618

7 2009/10 JH Kallis 572

8 2009/10 SK Raina 528

9 2011 CH Gayle 608

10 2011 V Kohli 557

11 2011 SR Tendulkar 553

12 2012 CH Gayle 733

13 2012 G Gambhir 590

14 2012 S Dhawan 569

15 2013 MEK Hussey 733

16 2013 CH Gayle 720

17 2013 V Kohli 639

18 2014 RV Uthappa 660

19 2014 DR Smith 566

20 2014 GJ Maxwell 552

21 2015 DA Warner 562

22 2015 AM Rahane 540

23 2015 LMP Simmons 540

24 2016 V Kohli 973

25 2016 DA Warner 848

26 2016 AB de Villiers 687

27 2017 DA Warner 641

28 2017 G Gambhir 498

29 2017 S Dhawan 479

30 2018 KS Williamson 735

31 2018 RR Pant 684

32 2018 KL Rahul 659

33 2019 DA Warner 692

34 2019 KL Rahul 593

35 2019 Q de Kock 529

36 2020/21 KL Rahul 676

37 2020/21 S Dhawan 618

38 2020/21 DA Warner 548

39 2021 RD Gaikwad 635

40 2021 F du Plessis 633

41 2021 KL Rahul 626

42 2022 JC Buttler 863

43 2022 KL Rahul 616

44 2022 Q de Kock 508

45 2023 Shubman Gill 890

46 2023 F du Plessis 730

47 2023 DP Conway 672

48 2024 RD Gaikwad 509

49 2024 V Kohli 500

50 2024 B Sai Sudharsan 418

Top Three Wicket Takers:

Season Bowler wicket\_confirmation

0 2007/08 Sohail Tanvir 24

1 2007/08 IK Pathan 20

2 2007/08 JA Morkel 20

3 2009 RP Singh 26

4 2009 A Kumble 22

5 2009 A Nehra 22

6 2009/10 PP Ojha 22

7 2009/10 A Mishra 20

8 2009/10 Harbhajan Singh 20

9 2011 SL Malinga 30

10 2011 MM Patel 22

11 2011 S Aravind 22

12 2012 M Morkel 30

13 2012 SP Narine 29

14 2012 SL Malinga 25

15 2013 DJ Bravo 34

16 2013 JP Faulkner 33

17 2013 R Vinay Kumar 27

18 2014 MM Sharma 26

19 2014 SP Narine 22

20 2014 B Kumar 21

21 2015 DJ Bravo 28

22 2015 SL Malinga 26

23 2015 A Nehra 25

24 2016 B Kumar 24

25 2016 SR Watson 23

26 2016 YS Chahal 22

27 2017 B Kumar 28

28 2017 JD Unadkat 27

29 2017 JJ Bumrah 23

30 2018 AJ Tye 28

31 2018 S Kaul 24

32 2018 Rashid Khan 23

33 2019 K Rabada 29

34 2019 Imran Tahir 26

35 2019 JJ Bumrah 23

36 2020/21 K Rabada 32

37 2020/21 JJ Bumrah 30

38 2020/21 TA Boult 26

39 2021 HV Patel 35

40 2021 Avesh Khan 27

41 2021 JJ Bumrah 22

42 2022 YS Chahal 29

43 2022 PWH de Silva 27

44 2022 K Rabada 23

45 2023 MM Sharma 31

46 2023 Mohammed Shami 28

47 2023 Rashid Khan 28

48 2024 HV Patel 19

49 2024 Mukesh Kumar 15

50 2024 Arshdeep Singh 14

INTERPRETATION:  
The provided data on top run-getters and wicket-takers in the IPL from 2007/08 to 2024 showcases the dynamic nature of the league. It highlights both consistent performers and emerging talents, reflecting the evolving competitive landscape of the tournament. The IPL continues to be a platform where established stars shine and new players emerge, contributing to the excitement and unpredictability of the game.

4) Find the relationship between a player’s performance and the salary he gets in your data.

# Calculate the correlation

correlation = df\_merged\_runs['Rs'].corr(df\_merged\_runs['runs\_scored'])

print("Correlation between Salary and Runs:", correlation)

Correlation between Salary and Runs: 0.3061248376582168

INTERPRETATION:

The correlation of 0.3061 between Salary and Runs Scored suggests that there is a positive, though moderate, relationship between how much players are paid and their run-scoring performance. While higher-paid players tend to score more runs, the relationship is not strong enough to conclude that salary is a major predictor of runs scored. This indicates that other factors also play significant roles in determining both salary and player performance.

5) total wicket each year by the bowler.

total\_wicket\_each\_year.sort\_values(["year", "wicket\_confirmation"], ascending=False, inplace=True)

print(total\_wicket\_each\_year)

year Bowler wicket\_confirmation

1836 2024 HV Patel 19

1875 2024 Mukesh Kumar 15

1822 2024 Arshdeep Singh 14

1842 2024 JJ Bumrah 14

1876 2024 Mustafizur Rahman 14

... ... ... ...

16 2008 CL White 0

41 2008 K Goel 0

43 2008 LPC Silva 0

60 2008 Pankaj Singh 0

90 2008 VS Yeligati 0

INTERPRETATION:

The data on the top wicket-takers in 2024 highlights the competitive nature of the IPL, showcasing both emerging talents and consistent performers. HV Patel's leading performance with 19 wickets is particularly notable, while other bowlers like Mukesh Kumar, Arshdeep Singh, JJ Bumrah, and Mustafizur Rahman also made significant contributions to their teams. Comparing this to the early seasons, it is evident that the league has evolved, with more players consistently performing at high levels.