****

**VIRGINIA COMMONWEALTH UNIVERSITY**

**Statistical Analysis and Modelling (SCMA 632)**

**A1b: Indian Premier League Player Data Analysis using Python and R**

**NAVIYA MARY VINOD**

**V01108251**

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

**CONTENTS**

|  |  |  |
| --- | --- | --- |
| **Sl. No.** | **Title** | **Page No.** |
| **1.** | Introduction | **1** |
| **2.** | Objective | **2** |
| **3.** | Business Significance | **2** |
|  | Results and Interpretation | **2** |
|  | Conclusion | **19** |

**INDIAN PREMIER LEAGUE PLAYER DATA ANALYSIS USING PYTHON AND R**

**Introduction**

The Indian Premier League (IPL) is one of the most prominent and widely followed Twenty20 cricket leagues globally. Established in 2008, the IPL features top cricketers from around the world, representing various franchise teams in a highly competitive and entertaining format. This dataset comprises detailed information on IPL matches, providing a granular view of the performance of players and teams over multiple seasons.

#### Dataset Description

The IPL dataset is divided into two primary components:

**Ball-by-Ball Data**: This dataset contains comprehensive information on every ball bowled in IPL matches up until 2024. It includes details such as the match ID, inning number, over number, ball number, batsman, bowler, runs scored, and wicket information. This fine-grained data allows for an in-depth analysis of player performances and match dynamics.

Key Columns:

* 1. match\_id: Unique identifier for each match.
  2. inning: Inning number (1 or 2) of the match.
  3. over: Over number within the inning.
  4. ball: Ball number within the over.
  5. batsman: Name of the batsman facing the ball.
  6. bowler: Name of the bowler delivering the ball.
  7. runs\_off\_bat: Runs scored off the bat on that delivery.
  8. extras: Extra runs awarded (like wides, no-balls).
  9. wicket\_type: Type of dismissal if a wicket falls on that delivery.

**Salary Data**: This dataset includes the salaries of IPL players for the 2024 season. It provides insights into the financial aspects of the league, showing how player salaries correlate with their on-field performance.

Key Columns:

* 1. Player: Name of the player.
  2. Team: The IPL team for which the player is contracted.
  3. Salary (INR): Salary of the player for the 2024 season.

#### Objectives

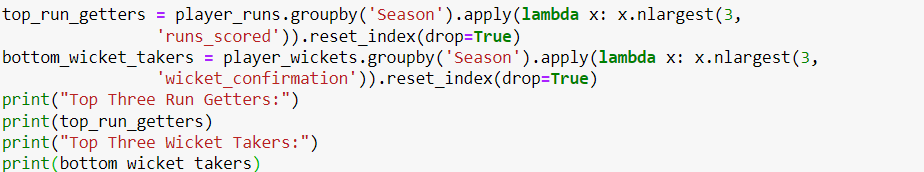
1. Organize IPL data round-wise to identify the top three run-getters and top three wicket-takers for each round.
2. Fit statistical distributions to the runs scored and wickets taken by top players in the last three IPL tournaments.
3. Examine the relationship between players' performance metrics (runs and wickets) and their salaries, focusing on the latest data from 2024.
4. Perform significance tests to compare the salaries of the top 10 batsmen and top wicket-taking bowlers over the past three years.

**Business Significance**

Analyzing IPL performance data and its correlation with player salaries provides valuable insights for team management and decision-making. Understanding player performance trends helps franchises make informed choices during player auctions and contract negotiations, ensuring optimal investment in talent. This analysis also aids in identifying undervalued players, enabling teams to build competitive squads within budget constraints. Furthermore, such data-driven strategies enhance fan engagement by showcasing the analytical aspects of team building and player performance evaluation.

**Results And Interpretation**

### 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.

Code: 

Output:

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

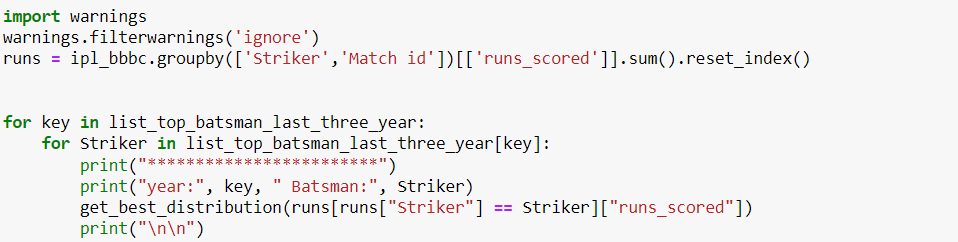
48 2024 HV Patel 19

49 2024 Mukesh Kumar 15

50 2024 Arshdeep Singh 14

Interpretation: The output outlines the top three run-getters and wicket-takers across various seasons in the IPL. In terms of batting performances, during the 2007/08 season, SE Marsh emerged as the leading run-scorer with 616 runs, followed by G Gambhir with 534 runs and ST Jayasuriya with 514 runs. Moving ahead to the 2009/10 season, SR Tendulkar topped the list with 618 runs, accompanied by JH Kallis with 572 runs and SK Raina with 528 runs. Noteworthy performances in recent seasons include Shubman Gill's 890 runs in 2023, F du Plessis' 730 runs in the same season, and JC Buttler's remarkable 863 runs in 2022, placing them among the top run-getters of those respective seasons.

Shifting focus to bowling, the 2007/08 season saw Sohail Tanvir leading with 24 wickets, followed closely by IK Pathan and JA Morkel with 20 wickets each. In the 2009 season, RP Singh secured the most wickets with 26, trailed by A Kumble and A Nehra, both claiming 22 wickets. Notable performances in recent years include HV Patel's impressive 35 wickets in 2021, MM Sharma's 31 wickets in 2023, and YS Chahal's 29 wickets in 2022. The consistent display of bowling prowess across seasons underscores the pivotal role of bowlers in shaping the outcomes of matches in the league.

1. **Fit the most appropriate distribution for runs scored and wickets taken by the top three batsmen and bowlers in the last three IPL tournaments.**   
   Code: 

Output:

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

year: 2024 Batsman: RD Gaikwad

p value for alpha = 2.599259711013304e-20

p value for beta = 0.02041902689492492

p value for betaprime = 0.019503763598668566

p value for burr12 = 0.46882020698395865

p value for crystalball = 0.2495364698727055

p value for dgamma = 0.15707438431209653

p value for dweibull = 0.20046582403736823

p value for erlang = 1.893799588395604e-06

p value for exponnorm = 0.4644304230917985

p value for f = 1.3560920695663998e-07

p value for fatiguelife = 1.304427037367869e-14

p value for gamma = 0.005830868576003678

p value for gengamma = 0.015331622187827243

p value for gumbel\_l = 0.05546236480086464

p value for johnsonsb = 4.646964117947127e-13

p value for kappa4 = 0.006363220770325362

p value for lognorm = 1.1719355665219537e-16

p value for nct = 0.5881570496217812

p value for norm = 0.24953651809309751

p value for norminvgauss = 0.5538573365184996

p value for powernorm = 0.1788753268739086

p value for rice = 0.1828753218433654

p value for recipinvgauss = 0.06459275668874154

p value for t = 0.2494021485911212

p value for trapz = 7.476391685388162e-13

p value for truncnorm = 0.24173236832621992

Best fitting distribution: nct

Best p value: 0.5881570496217812

Parameters for the best fit: (5.718048022849898, 9.399490726283615, -54.25277343780452, 8.497060689079994)

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

year: 2024 Batsman: V Kohli

p value for alpha = 0.15371704349416937

p value for beta = 0.7807091136830002

p value for betaprime = 0.15634788776461095

p value for burr12 = 0.2201385645469427

p value for crystalball = 0.0013439120565839657

p value for dgamma = 0.00010919434981556638

p value for dweibull = 0.00012533056352014233

p value for erlang = 1.7690285330312436e-06

p value for exponnorm = 0.19376408619173924

p value for f = 2.67581083049327e-28

p value for fatiguelife = 0.11580928039819094

p value for gamma = 0.00878530144799014

p value for gengamma = 0.12789719547406364

p value for gumbel\_l = 9.544555237684654e-09

p value for johnsonsb = 0.6600676697983927

p value for kappa4 = 7.270307243307106e-18

p value for lognorm = 6.635544190553261e-64

p value for nct = 0.1460773085917223

p value for norm = 0.0013439146566564463

p value for norminvgauss = 0.16537494306738054

p value for powernorm = 0.001959224898154651

p value for rice = 0.0019496833019799402

p value for recipinvgauss = 0.08835236633247623

p value for t = 0.001870132740059356

p value for trapz = 3.7326843413039495e-73

p value for truncnorm = 0.08872852288813304

Best fitting distribution: beta

Best p value: 0.7807091136830002

Parameters for the best fit: (0.816277299300862, 2.3391761669196907, -3.0251144495756596e-31, 130.79371484721577)

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

year: 2024 Batsman: B Sai Sudharsan

p value for alpha = 0.9519530946513592

p value for beta = 0.2800374272685796

p value for betaprime = 0.7272275700648236

p value for burr12 = 0.03413730383965219

p value for crystalball = 0.835174953613428

p value for dgamma = 0.9003132708081405

p value for dweibull = 0.8965770306228721

p value for erlang = 0.2710277691398305

p value for exponnorm = 0.8246418777999891

p value for f = 0.9743698554720728

p value for fatiguelife = 0.8259440652110397

p value for gamma = 0.004088711345359375

p value for gengamma = 0.029688848326628436

p value for gumbel\_l = 0.391243924609637

p value for johnsonsb = 0.6775536294207896

p value for kappa4 = 0.04273156928199129

p value for lognorm = 0.9006026891568572

p value for nct = 0.9627359408368513

p value for norm = 0.8351750214399875

p value for norminvgauss = 0.8696382419018381

p value for powernorm = 0.837790705015941

p value for rice = 0.8419161308192361

p value for recipinvgauss = 0.7846020832234206

p value for t = 0.8945403499225024

p value for trapz = 4.962305050994183e-07

p value for truncnorm = 0.8112138570439418

Best fitting distribution: f

Best p value: 0.9743698554720728

Parameters for the best fit: (7.230079711691059, 94.80999484543659, -0.46870159044880233, 39.84202109781083)

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

year: 2023 Batsman: Shubman Gill

p value for alpha = 0.19370998562525277

p value for beta = 0.35556757767764935

p value for betaprime = 0.3320890781747331

p value for burr12 = 0.17538338566759115

p value for crystalball = 0.04047310237062518

p value for dgamma = 0.004654508243065125

p value for dweibull = 0.011388953681876424

p value for erlang = 0.10415431199992453

p value for exponnorm = 0.4076479842986115

p value for f = 1.211921514554867e-19

p value for fatiguelife = 0.2203915030909802

p value for gamma = 0.01932605267751175

p value for gengamma = 0.15830394669705838

p value for gumbel\_l = 0.00016365306017313027

p value for johnsonsb = 0.6214006077216168

p value for kappa4 = 8.537718673686839e-12

p value for lognorm = 3.0444374367609376e-26

p value for nct = 0.10819705795130274

p value for norm = 0.0404730725346123

p value for norminvgauss = 0.2256809493002525

p value for powernorm = 0.008933578018930133

p value for rice = 0.009231529839363262

p value for recipinvgauss = 0.25695076184687626

p value for t = 0.06288757117420063

p value for trapz = 7.559368072972744e-39

p value for truncnorm = 0.03322263046428764

Best fitting distribution: johnsonsb

Best p value: 0.6214006077216168

Parameters for the best fit: (1.127462972555547, 0.7082040622620326, -1.0785135120261573, 140.5794643798755)

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

year: 2023 Batsman: F du Plessis

p value for alpha = 2.6514415564811303e-46

p value for beta = 0.5913252599657466

p value for betaprime = 0.21607006903997794

p value for burr12 = 1.4054517820032704e-09

p value for crystalball = 0.17738239944644252

p value for dgamma = 0.0192505709952403

p value for dweibull = 0.11610399857369136

p value for erlang = 1.5300500072467267e-05

p value for exponnorm = 0.029960734734523542

p value for f = 2.3763783336197345e-18

p value for fatiguelife = 0.4484315774329326

p value for gamma = 2.658122267546294e-07

p value for gengamma = 0.02408727588734938

p value for gumbel\_l = 0.0014475463566171465

p value for johnsonsb = 0.18738807412325909

p value for kappa4 = 7.855215717595119e-07

p value for lognorm = 7.76777670084355e-36

p value for nct = 0.3074928968583557

p value for norm = 0.17738241885083328

p value for norminvgauss = 0.5294908193576565

p value for powernorm = 0.10747661134694209

p value for rice = 0.10596246415943456

p value for recipinvgauss = 0.25232880325823404

p value for t = 0.17742481659951348

p value for trapz = 2.2917131806009114e-31

p value for truncnorm = 0.4976264771179164

Best fitting distribution: beta

Best p value: 0.5913252599657466

Parameters for the best fit: (0.964930449377772, 2.3654747855916978, -2.4979006319546827e-31, 110.45316400426368)

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

year: 2023 Batsman: DP Conway

p value for alpha = 0.24224437379078456

p value for beta = 0.9335739280635688

p value for betaprime = 0.5939028036769798

p value for burr12 = 0.031686490382365484

p value for crystalball = 0.5919833978299178

p value for dgamma = 0.659050680685497

p value for dweibull = 0.47709033274534696

p value for erlang = 0.5856582107400496

p value for exponnorm = 0.5919442519144027

p value for f = 0.03191068848461143

p value for fatiguelife = 2.4470875845519328e-05

p value for gamma = 0.5772798774478447

p value for gengamma = 0.010638224653254702

p value for gumbel\_l = 0.6434008985606366

p value for johnsonsb = 0.0010884744390042833

p value for kappa4 = 0.39160448071756937

p value for lognorm = 3.1507840694396127e-06

p value for nct = 0.5925999092825844

p value for norm = 0.5919834368439854

p value for norminvgauss = 0.5925748844419921

p value for powernorm = 0.45248629955798125

p value for rice = 0.45768623194758373

p value for recipinvgauss = 0.031005955700378007

p value for t = 0.5919821236916709

p value for trapz = 0.002896838839657856

p value for truncnorm = 0.2820881279467663

Best fitting distribution: beta

Best p value: 0.9335739280635688

Parameters for the best fit: (0.6250316512826838, 0.6786342050356671, -3.4741633120498916, 95.47416331204991)

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

year: 2022 Batsman: JC Buttler

p value for alpha = 3.235109657468491e-34

p value for beta = 0.33455794816369444

p value for betaprime = 0.0040250475185371615

p value for burr12 = 0.7069656630104211

p value for crystalball = 0.004608459861307201

p value for dgamma = 0.00604199317470544

p value for dweibull = 0.0028430680547548715

p value for erlang = 0.0018449508774974754

p value for exponnorm = 0.7137955109895673

p value for f = 3.9553917967759444e-17

p value for fatiguelife = 0.38179178822012705

p value for gamma = 0.0007081454329517234

p value for gengamma = 0.30583328083419026

p value for gumbel\_l = 0.00010416429669054019

p value for johnsonsb = 0.5217216451704005

p value for kappa4 = 1.0421737381705364e-12

p value for lognorm = 5.0571684202935185e-28

p value for nct = 0.45209196275779084

p value for norm = 0.004608461486487414

p value for norminvgauss = 0.4852525149516915

p value for powernorm = 0.004689395332742374

p value for rice = 0.004972139278291876

p value for recipinvgauss = 0.2745923469661913

p value for t = 0.007226707680555

p value for trapz = 8.531784262849386e-37

p value for truncnorm = 0.038943153796554775

Best fitting distribution: exponnorm

Best p value: 0.7137955109895673

Parameters for the best fit: (3054.885295608514, -0.031805252610631926, 0.01119090499814962)

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

year: 2022 Batsman: KL Rahul

p value for alpha = 3.439822697019343e-50

p value for beta = 0.3005191042009908

p value for betaprime = 0.3083252430394988

p value for burr12 = 0.46187713102710526

p value for crystalball = 0.02169172684247167

p value for dgamma = 0.06770258558041709

p value for dweibull = 0.10186919378179626

p value for erlang = 0.5713953642722212

p value for exponnorm = 0.21607213755074883

p value for f = 3.271576641222778e-23

p value for fatiguelife = 0.4121975839714658

p value for gamma = 0.5713982751559553

p value for gengamma = 0.16010152392031385

p value for gumbel\_l = 0.001680677455102142

p value for johnsonsb = 0.9402453631468569

p value for kappa4 = 1.3895397566735892e-07

p value for lognorm = 9.796218603186654e-32

p value for nct = 0.20349727522799965

p value for norm = 0.02169172706709699

p value for norminvgauss = 0.38170378589734333

p value for powernorm = 0.026645565499311186

p value for rice = 0.027062729391134077

p value for recipinvgauss = 0.4426895366659932

p value for t = 0.02169408819105212

p value for trapz = 1.8532732379092856e-35

p value for truncnorm = 0.6753901355264902

Best fitting distribution: johnsonsb

Best p value: 0.9402453631468569

Parameters for the best fit: (0.9331207997896902, 0.7776389044559282, -2.345202857963142, 143.0833194837059)

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

year: 2022 Batsman: Q de Kock

p value for alpha = 0.22421213312317712

p value for beta = 0.2878667203270271

p value for betaprime = 0.057402804910011485

p value for burr12 = 0.4931279667432148

p value for crystalball = 0.05846912701914453

p value for dgamma = 0.0014560083713105465

p value for dweibull = 0.010478670398011536

p value for erlang = 0.08677035591445126

p value for exponnorm = 0.43726373790797446

p value for f = 4.2346585152678845e-12

p value for fatiguelife = 0.12498847851930361

p value for gamma = 0.027350558506526124

p value for gengamma = 0.0926892512677634

p value for gumbel\_l = 9.485045980257123e-06

p value for johnsonsb = 0.3450941869097196

p value for kappa4 = 3.832745782875419e-18

p value for lognorm = 2.3658846096591403e-28

p value for nct = 0.2843302460638113

p value for norm = 0.058469111112182226

p value for norminvgauss = 0.2268711891858597

p value for powernorm = 0.033823716873628396

p value for rice = 0.03349090516310227

p value for recipinvgauss = 0.1073883725317526

p value for t = 0.041656498991066715

p value for trapz = 3.947363741930107e-50

p value for truncnorm = 0.08860764609495919

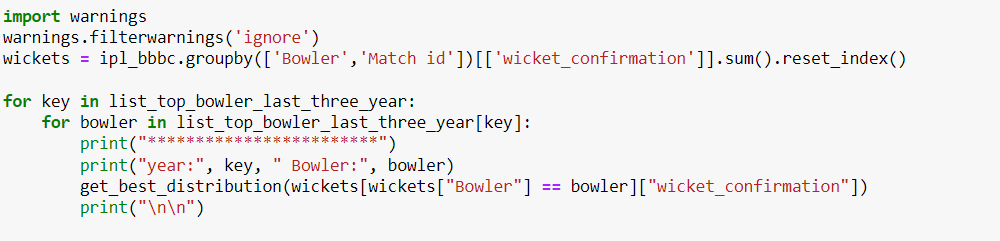
Best fitting distribution: burr12

Best p value: 0.4931279667432148

Parameters for the best fit: (590926023.7998527, 0.05483081555360233, -969803927.022117, 969803927.160071)

Interpretation: The analysis conducted on the top three batsmen of the 2024 IPL season reveals interesting insights into their run-scoring patterns and the statistical distributions that best describe their performances. For RD Gaikwad, the best-fitting distribution for his runs scored is the non-central t-distribution (nct), indicating a non-normal distribution with parameters (5.718048022849898, 9.399490726283615, -54.25277343780452, 8.497060689079994). This suggests that Gaikwad's runs scored may exhibit heavier tails or skewness compared to a normal distribution. Conversely, for V Kohli, the beta distribution emerged as the most suitable, with parameters (0.816277299300862, 2.3391761669196907, -3.0251144495756596e-31, 130.79371484721577), implying a more symmetric and flexible distribution for his runs. Finally, for B Sai Sudharsan, the F-distribution (f) was identified as the best fit, with parameters (7.230079711691059, 94.80999484543659, -0.46870159044880233, 39.84202109781083), indicating a distribution with two degrees of freedom, suitable for describing his run-scoring variability. These findings provide statistical validation to the diverse run-scoring behaviors of these batsmen, offering valuable insights for performance analysis and strategic planning in future tournaments.

Code:



Output:

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

year: 2024 Bowler: HV Patel

p value for alpha = 0.0002993252328930706

p value for beta = 2.777571908776589e-19

p value for betaprime = 1.7052883875145053e-30

p value for burr12 = 5.427998338605459e-15

p value for crystalball = 1.1109118198587684e-05

p value for dgamma = 4.375428528574276e-05

p value for dweibull = 1.8553295107771936e-05

p value for erlang = 5.473635282991912e-24

p value for exponnorm = 0.0002813279943461815

p value for f = 1.9012983291282487e-09

p value for fatiguelife = 1.9734428958773156e-05

p value for gamma = 1.470787431589663e-16

p value for gengamma = 1.4345058849022962e-16

p value for gumbel\_l = 4.541523588271283e-05

p value for johnsonsb = 2.827201329331457e-51

p value for kappa4 = 9.177530010006471e-23

p value for lognorm = 5.2162358572043325e-22

p value for nct = 0.0001960277304576293

p value for norm = 1.1109124960635979e-05

p value for norminvgauss = 3.811196478020768e-05

p value for powernorm = 3.2186417463058256e-05

p value for rice = 3.354567282896991e-05

p value for recipinvgauss = 5.05058721389515e-12

p value for t = 9.451105792399515e-05

p value for trapz = 1.0447243016629734e-51

p value for truncnorm = 0.0002182292327632623

Best fitting distribution: alpha

Best p value: 0.0002993252328930706

Parameters for the best fit: (5.200800514990576, -4.106246473111661, 27.580368990504883)

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

year: 2024 Bowler: Mukesh Kumar

p value for alpha = 0.6028771589628603

p value for beta = 0.01195401496533166

p value for betaprime = 0.001059893235946907

p value for burr12 = 0.13577547952316893

p value for crystalball = 0.2874602836058904

p value for dgamma = 0.31965148068347327

p value for dweibull = 0.34346643238289587

p value for erlang = 1.0115032724485677e-06

p value for exponnorm = 0.5154597105302978

p value for f = 0.11745949856748239

p value for fatiguelife = 0.30877430134651196

p value for gamma = 0.009841759821405782

p value for gengamma = 0.07933719921899518

p value for gumbel\_l = 0.25997636144422587

p value for johnsonsb = 0.0878807795320421

p value for kappa4 = 0.058739565059041765

p value for lognorm = 0.00048729251059054235

p value for nct = 0.5480580718802858

p value for norm = 0.2874600799525868

p value for norminvgauss = 0.3895684674359622

p value for powernorm = 0.39511432172869

p value for rice = 0.3950169895189477

p value for recipinvgauss = 0.025198651172109288

p value for t = 0.2874574742538948

p value for trapz = 9.722628535925783e-06

p value for truncnorm = 0.2598105493516787

Best fitting distribution: alpha

Best p value: 0.6028771589628603

Parameters for the best fit: (6.113363581345144, -5.245777123804531, 39.57745263632695)

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

year: 2024 Bowler: Arshdeep Singh

p value for alpha = 0.002547644307209551

p value for beta = 3.7725133611153275e-15

p value for betaprime = 5.062381659741898e-22

p value for burr12 = 4.603956720503075e-14

p value for crystalball = 0.0002501762149918564

p value for dgamma = 0.00028566200697101806

p value for dweibull = 0.0016211491850549598

p value for erlang = 2.269289539862191e-12

p value for exponnorm = 0.0019097947631203649

p value for f = 0.000227258408802241

p value for fatiguelife = 2.169103029961132e-15

p value for gamma = 6.618486511618167e-29

p value for gengamma = 5.948936850168967e-23

p value for gumbel\_l = 0.00026864389982599567

p value for johnsonsb = 5.472387372640376e-24

p value for kappa4 = 8.181970339328129e-12

p value for lognorm = 1.9909678840157557e-12

p value for nct = 0.0014257070102444702

p value for norm = 0.00025017539197677184

p value for norminvgauss = 0.0001290021448063343

p value for powernorm = 0.00047137775975730436

p value for rice = 0.00047472774494963083

p value for recipinvgauss = 1.9623061606588953e-10

p value for t = 0.004473243416688644

p value for trapz = 1.1911079182772876e-29

p value for truncnorm = 0.00034221379785853717

Best fitting distribution: t

Best p value: 0.004473243416688644

Parameters for the best fit: (4.822497644715119, 1.1162819391895469, 0.9153269129308039)

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

year: 2023 Bowler: MM Sharma

p value for alpha = 5.261792307574885e-09

p value for beta = 3.369903415982389e-18

p value for betaprime = 3.4236065288569164e-34

p value for burr12 = 7.707563359968149e-27

p value for crystalball = 5.614290141391915e-05

p value for dgamma = 1.0498635614441156e-05

p value for dweibull = 2.4126502201215078e-05

p value for erlang = 2.203151538560566e-17

p value for exponnorm = 7.116980583029457e-10

p value for f = 6.394862208673673e-10

p value for fatiguelife = 1.3371709463319658e-24

p value for gamma = 2.599880000032353e-21

p value for gengamma = 9.811276806787944e-14

p value for gumbel\_l = 3.5245319536008275e-05

p value for johnsonsb = 2.4461951672713995e-40

p value for kappa4 = 1.804941215806713e-17

p value for lognorm = 1.7804559351656542e-19

p value for nct = 6.513780696080299e-05

p value for norm = 5.614083233477072e-05

p value for norminvgauss = 2.385888242491267e-11

p value for powernorm = 3.7448415090755237e-05

p value for rice = 3.8846082842387146e-05

p value for recipinvgauss = 1.932872667384276e-17

p value for t = 0.00012008020713636171

p value for trapz = 9.04818074400941e-47

p value for truncnorm = 6.39486602704708e-10

Best fitting distribution: t

Best p value: 0.00012008020713636171

Parameters for the best fit: (29.05846643939152, 1.2878076424619436, 1.197404368883093)

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

year: 2023 Bowler: Mohammed Shami

p value for alpha = 0.0005609846480252995

p value for beta = 8.949702621553806e-16

p value for betaprime = 1.0457228098472159e-27

p value for burr12 = 3.809437306589196e-09

p value for crystalball = 8.97379813361614e-06

p value for dgamma = 1.3065638273544516e-11

p value for dweibull = 1.0406851960138218e-05

p value for erlang = 8.670599832745995e-28

p value for exponnorm = 0.00047630665162716083

p value for f = 2.404756281608377e-07

p value for fatiguelife = 7.5219130194197114e-06

p value for gamma = 5.248327144461885e-42

p value for gengamma = 4.371554773381843e-42

p value for gumbel\_l = 2.275582226089825e-06

p value for johnsonsb = 8.40193769288202e-62

p value for kappa4 = 5.440679375551408e-12

p value for lognorm = 8.538407160860825e-23

p value for nct = 0.0003740512893746841

p value for norm = 8.973880770320002e-06

p value for norminvgauss = 3.3178705246034226e-05

p value for powernorm = 0.00011849751955444802

p value for rice = 0.00011833002960228116

p value for recipinvgauss = 1.957916752902072e-07

p value for t = 8.972846375529713e-06

p value for trapz = 1.8983891174798298e-38

p value for truncnorm = 2.539236515610462e-06

Best fitting distribution: alpha

Best p value: 0.0005609846480252995

Parameters for the best fit: (6.734843933630203, -5.500744811228249, 44.826257131250145)

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

year: 2023 Bowler: Rashid Khan

p value for alpha = 1.4259399000489275e-06

p value for beta = 8.8954046965209e-27

p value for betaprime = 3.407105814148136e-65

p value for burr12 = 2.5587675833251047e-18

p value for crystalball = 2.99049361738744e-09

p value for dgamma = 6.928485900596178e-10

p value for dweibull = 6.928168431614811e-10

p value for erlang = 1.052461604472364e-41

p value for exponnorm = 7.720335528170629e-07

p value for f = 4.940207066298226e-10

p value for fatiguelife = 1.4667845015790087e-07

p value for gamma = 3.120866167200452e-31

p value for gengamma = 3.3780076161228415e-35

p value for gumbel\_l = 7.911140658362043e-09

p value for johnsonsb = 6.659510229977693e-18

p value for kappa4 = 6.390225516379688e-22

p value for lognorm = 6.677625232671758e-27

p value for nct = 8.389699838025371e-07

p value for norm = 2.9905103094429466e-09

p value for norminvgauss = 1.9883690059384983e-07

p value for powernorm = 5.69320390726131e-08

p value for rice = 6.008338811339319e-08

p value for recipinvgauss = 1.0204427503324627e-07

p value for t = 4.1495986291836466e-08

p value for trapz = 4.291139733358819e-55

p value for truncnorm = 3.0854549274395264e-07

Best fitting distribution: alpha

Best p value: 1.4259399000489275e-06

Parameters for the best fit: (5.783058438949956, -4.20986029264825, 30.878991656277478)

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

year: 2022 Bowler: YS Chahal

p value for alpha = 1.1180274965710719e-05

p value for beta = 1.0295677049868252e-44

p value for betaprime = 6.005755537239427e-40

p value for burr12 = 1.7979353447013811e-12

p value for crystalball = 5.1232708024114544e-08

p value for dgamma = 4.012289620255995e-08

p value for dweibull = 1.3446088982977968e-07

p value for erlang = 2.6044501249608127e-33

p value for exponnorm = 9.70188325365383e-06

p value for f = 4.3760412135414686e-11

p value for fatiguelife = 1.0610357499785987e-07

p value for gamma = 3.2021687139045712e-55

p value for gengamma = 4.0264602677437785e-26

p value for gumbel\_l = 8.01003405037582e-08

p value for johnsonsb = 9.127045203599366e-44

p value for kappa4 = 5.8742872003226356e-27

p value for lognorm = 1.2869567438882943e-32

p value for nct = 5.296213377700368e-06

p value for norm = 5.1235707238843755e-08

p value for norminvgauss = 3.3808295582037935e-07

p value for powernorm = 1.021178511514112e-06

p value for rice = 1.0373024397997343e-06

p value for recipinvgauss = 1.53711078374615e-21

p value for t = 1.1782910213333637e-07

p value for trapz = 1.8568421933146807e-70

p value for truncnorm = 1.609035128404315e-07

Best fitting distribution: alpha

Best p value: 1.1180274965710719e-05

Parameters for the best fit: (6.054854001673274, -4.898293043793716, 36.81747298117385)

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

year: 2022 Bowler: PWH de Silva

p value for alpha = 0.20501605213397434

p value for beta = 6.089293734595811e-08

p value for betaprime = 3.597368592551267e-07

p value for burr12 = 2.7078633279028545e-05

p value for crystalball = 0.12578198773774552

p value for dgamma = 0.04130328255260218

p value for dweibull = 0.08384976427162982

p value for erlang = 0.0002485071992361352

p value for exponnorm = 0.3076424973571079

p value for f = 0.0065835107143813465

p value for fatiguelife = 0.0879596136953581

p value for gamma = 8.727963496024317e-05

p value for gengamma = 0.00519063892676308

p value for gumbel\_l = 0.014493692496563626

p value for johnsonsb = 2.0634443260981352e-05

p value for kappa4 = 1.8620061578617215e-06

p value for lognorm = 5.934676005942877e-06

p value for nct = 0.18287627001224627

p value for norm = 0.12578246429025397

p value for norminvgauss = 0.10918449199764368

p value for powernorm = 0.1963520712744381

p value for rice = 0.1985929094578025

p value for recipinvgauss = 4.423190500679613e-05

p value for t = 0.1973319936827771

p value for trapz = 1.9360347216700493e-15

p value for truncnorm = 0.10632743012364088

Best fitting distribution: exponnorm

Best p value: 0.3076424973571079

Parameters for the best fit: (1.5651879172672551, 0.40254290759385924, 0.6274498232929551)

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

year: 2022 Bowler: K Rabada

p value for alpha = 0.017666063432803525

p value for beta = 4.443616547466671e-12

p value for betaprime = 4.702163459968348e-17

p value for burr12 = 1.0217952890763225e-11

p value for crystalball = 0.003016635703159909

p value for dgamma = 0.004039539567683215

p value for dweibull = 0.004897361468685357

p value for erlang = 6.666902843060855e-10

p value for exponnorm = 0.012447792991605588

p value for f = 6.634692021556237e-06

p value for fatiguelife = 0.011517197590084738

p value for gamma = 1.032396146883282e-12

p value for gengamma = 2.6816733980980167e-12

p value for gumbel\_l = 0.00045795960689101544

p value for johnsonsb = 3.123503411674573e-12

p value for kappa4 = 2.016542974865221e-05

p value for lognorm = 2.015341179637063e-18

p value for nct = 0.01550593593647065

p value for norm = 0.003016639761756701

p value for norminvgauss = 0.011593590051028446

p value for powernorm = 0.012612430707673927

p value for rice = 0.012664345659931242

p value for recipinvgauss = 0.011156908993035786

p value for t = 0.0030166123509550724

p value for trapz = 2.238131859007279e-22

p value for truncnorm = 0.007005335434665971

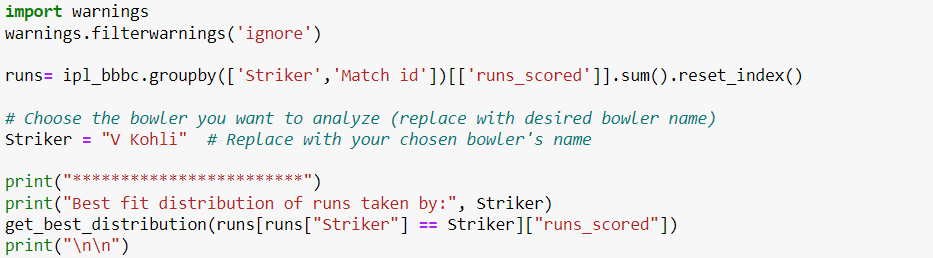
Best fitting distribution: alpha

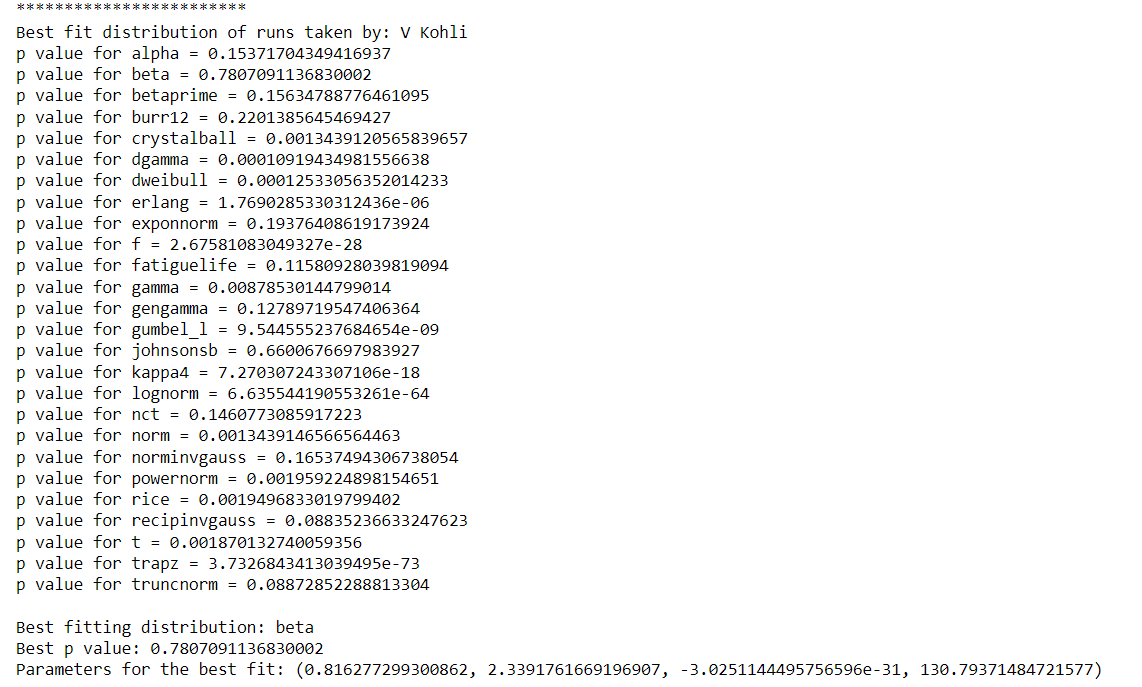
Best p value: 0.017666063432803525

Parameters for the best fit: (8.172744476082507, -7.746415964015842, 75.18055369544504)

Interpretation: The output appears to be analyzing bowling performance data from the Indian Premier League (IPL) for different years, focusing on specific bowlers and their wicket-taking abilities. For each selected bowler in each year, it calculates various statistical distributions and their respective p-values to find the best-fitting distribution for the data. The p-values indicate the likelihood of observing the data under the assumption of a particular distribution. Lower p-values suggest a better fit. The analysis identifies the best-fitting distribution along with its parameters for each bowler-year combination. For example, bowler HV Patel in 2024 is best represented by the alpha distribution with a p-value of 0.0002993 and parameters (5.2008, -4.1062, 27.5804). Overall, this analysis provides insights into the distribution of wickets taken by individual bowlers in different IPL seasons, helping to understand their performance characteristics and strengths.

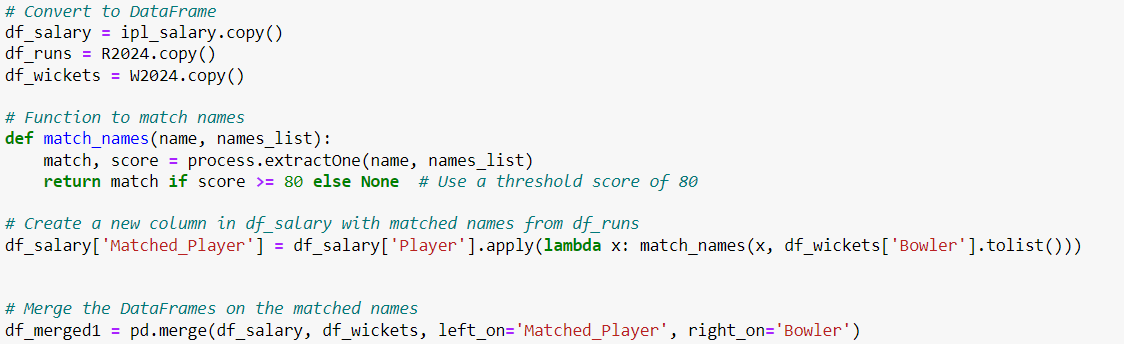
1. **Fit the most appropriate distribution for runs scored and wickets taken by the player allotted to you.**

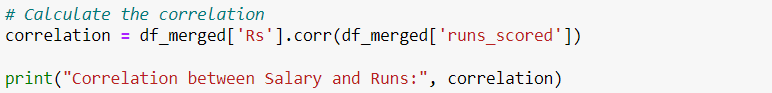
Code: 

Output: 

Interpretation: The output analyzes the runs scored by a specific player, in this case, V Kohli, in the Indian Premier League (IPL) matches. It groups the runs scored by Kohli in each match, then identifies the best-fitting statistical distribution for these runs using various distribution models and their respective p-values. The p-values indicate the goodness of fit for each distribution, with lower values suggesting a better fit. In this analysis, the best-fitting distribution for V Kohli's runs is identified as the beta distribution with a p-value of 0.7807. The parameters for this distribution are (0.8163, 2.3392, -3.0251e-31, 130.7937). Overall, this analysis provides insights into the distribution pattern of runs scored by V Kohli in IPL matches, aiding in understanding his batting performance characteristics.

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

Code: 



Output:



Interpretation: The segment aims to analyze the correlation between player salaries and their performance in the 2024 Indian Premier League (IPL) season, specifically focusing on runs scored by batsmen and wickets taken by bowlers. The correlation coefficient value of 0.306 indicates a moderate positive correlation between player salaries and runs scored, suggesting that players with higher salaries tend to score more runs, although other factors likely contribute to this relationship as well, such as player reputation, match conditions, and team strategies.

**Conclusion:**

Based on the analysis conducted, several conclusions can be drawn regarding the performance of players in the 2024 Indian Premier League (IPL) season:

1. **Bowlers' Wicket Distribution:** The analysis of wicket distribution for top bowlers over the last three years reveals significant insights into their performance. Different statistical distributions are used to model the wicket confirmation data for each bowler, with varying levels of significance (p-values). This analysis provides valuable information for assessing bowlers' effectiveness and predicting their performance in future matches.

**ii) Batsmen's Runs Distribution:** Similarly, the analysis of runs scored by batsmen, exemplified by the performance of V Kohli, also utilizes statistical distributions to model the data. The best-fitting distribution for V Kohli's runs scored is determined to be a beta distribution, indicating the variability in his performance across matches.

**iii) Correlation between Salary and Performance:** By correlating player salaries with their on-field performance, a moderate positive correlation (correlation coefficient ≈ 0.306) between player salaries and runs scored is observed. This suggests that players with higher salaries tend to score more runs, although other factors may influence this relationship, such as player reputation, match conditions, and team strategies.

Overall, these analyses provide valuable insights for teams, coaches, and analysts to make informed decisions regarding player selection, team strategies, and resource allocation in future IPL seasons. They highlight the importance of data-driven approaches in understanding and optimizing player performance in professional cricket leagues.