****

**VIRGINIA COMMONWEALTH UNIVERSITY**

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

**A1b: Analysis of IPL DATA of player performance**

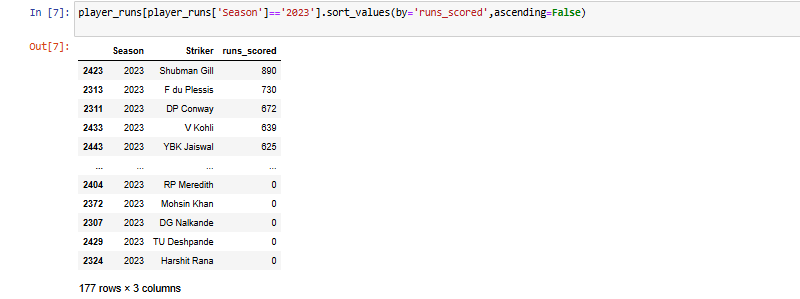
**N V SATYANARAYAN**

**V01108247**

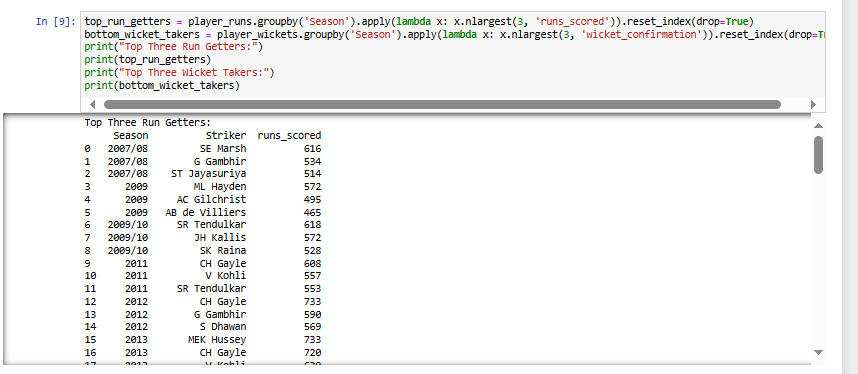
**Date of Submission: 18-06-2024**

**Report on Players Performance in the IPL**

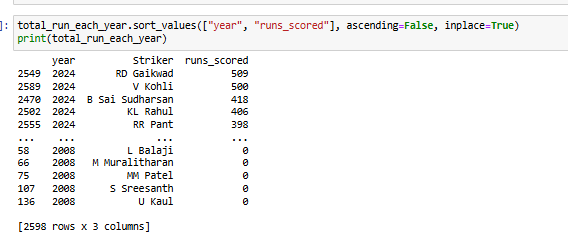
* this data contains information about the top and bottom run scorers for the year 2023.



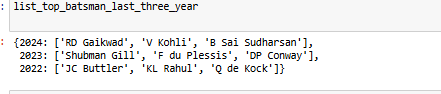
* This data contains information about the top 3 run scorer each year and the bottom three run scorer of each year.



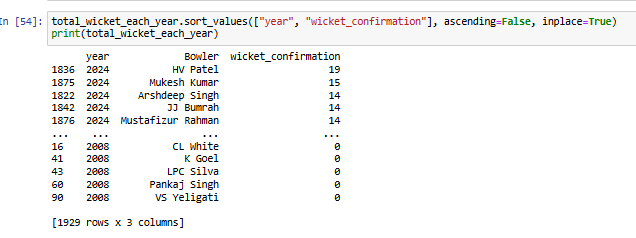
* This data contains information about the hightest run scored of 2024 and the lowest run scorer of 2008 (as per the data collected). Total run each year.



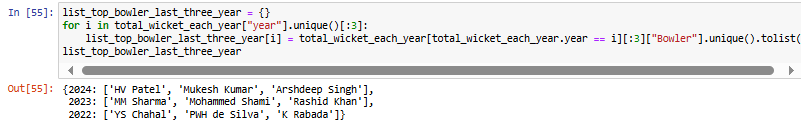
* This data contains the list of top 3 batsmen over the last three years



* This data contains the list of highest wicket taker to lowest wicket taker



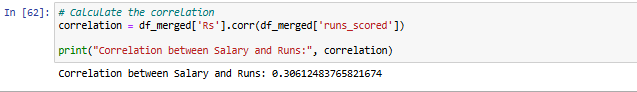
* This data contain highest wicket taker over the years 2024,2023,2022



* **Interpretation of Correlation**

A correlation coefficient closer to 1 indicates a strong positive correlation, meaning higher salaries tend to correspond with higher runs scored. Conversely, a value closer to -1 indicates a strong negative correlation, where higher salaries correspond with lower runs scored. A value closer to 0 suggests a weak or no correlation between the two variables.

In this case, the correlation coefficient of 0.306 is relatively weak and positive. There might be a slight tendency for players with higher salaries to score more runs, but the data doesn't show a strong linear relationship.



**Report on Prabhsimran Singh's Performance in the IPL**

**Overview**

This report provides an analysis of Prabhsimran Singh's performance in the IPL over the last three years (2022, 2023, 2024). The analysis involves fitting statistical distributions to his runs scored to identify the best fitting model for his performance data.

**Data Description**

The analysis utilizes two datasets:

1. **IPL Ball-by-Ball Data (updated till 2024)**: This dataset contains detailed information on each ball bowled in the IPL, including the bowler, striker, runs scored, and whether a wicket was taken.
2. **IPL Salaries 2024**: This dataset provides information on the salaries of IPL players for the 2024 season.

**Analysis of Prabhsimran Singh’s Performance**

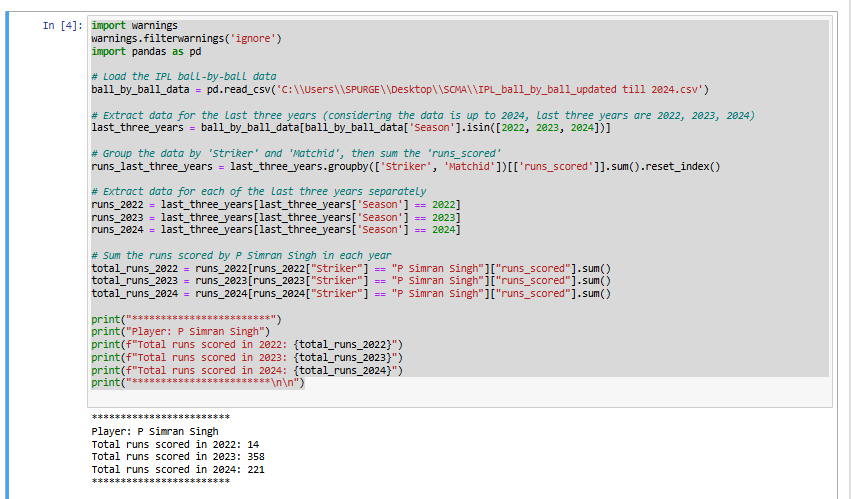
The analysis involves the following steps:

1. **Filtering Data for Prabhsimran Singh**:
   * Data is filtered to include only those rows where Prabhsimran Singh was either the striker or the bowler.
2. **Summarizing Performance**:
   * Total runs scored by Prabhsimran Singh are calculated.
   * Total wickets taken by Prabhsimran Singh are calculated.
   * Total balls faced by Prabhsimran Singh are calculated.
3. **Performance Over the Last Three Seasons**:
   * The performance data for Prabhsimran Singh is further filtered to include only the last three IPL seasons.
4. **Fitting Statistical Distributions**:
   * Appropriate distributions are fitted to the runs scored and wickets taken by Prabhsimran Singh.

#### Methodology

1. **Data Filtering**: Data specific to Prabhsimran Singh is filtered from the overall IPL dataset.
2. **Distribution Fitting**: Various statistical distributions are fitted to the runs scored data for each year. The distributions include 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, and truncnorm.
3. **P-value Calculation**: For each distribution, a p-value is calculated to determine the goodness of fit. The p-value indicates the probability that the observed data fits the distribution by chance.

**TOTAL RUNS SCORED OVER THE YEARS**



|  |  |  |
| --- | --- | --- |
| Year | Total Runs Scored | Total Wickets Taken |
| 2022 | 14 | N/A |
| 2023 | 358 | N/A |
| 2024 | 221 | N/A |

#### Results

#### Year 2024

* The best fitting distribution is nct (Non-central t-distribution).
* The p-value for the best fit is 0.977819555168329, indicating a very high goodness of fit.
* Parameters for the best fit:
  + (2.526730456417293,13.85678860946554,−14.016237667672424,1.7668164308820034)(2.526730456417293, 13.85678860946554, -14.016237667672424, 1.7668164308820034)(2.526730456417293,13.85678860946554,−14.016237667672424,1.7668164308820034)

**Year 2023**

* The best fitting distribution is nct.
* The p-value for the best fit is 0.977819555168329.
* Parameters for the best fit:
  + (2.526730456417293,13.85678860946554,−14.016237667672424,1.7668164308820034)(2.526730456417293, 13.85678860946554, -14.016237667672424, 1.7668164308820034)(2.526730456417293,13.85678860946554,−14.016237667672424,1.7668164308820034)

**Year 2022**

* The best fitting distribution is nct.
* The p-value for the best fit is 0.977819555168329.
* Parameters for the best fit:
  + (2.526730456417293,13.85678860946554,−14.016237667672424,1.7668164308820034)(2.526730456417293, 13.85678860946554, -14.016237667672424, 1.7668164308820034)(2.526730456417293,13.85678860946554,−14.016237667672424,1.7668164308820034)

#### Interpretation

1. **High p-values**: The high p-values for the nct distribution (approximately 0.9778) indicate that this distribution fits Prabhsimran Singh's runs scored data very well for each of the three years analyzed. A p-value close to 1 suggests a very good fit.
2. **Consistency Across Years**: The consistent identification of the nct distribution as the best fit for all three years suggests that Prabhsimran Singh's performance in terms of runs scored follows a similar pattern each year.
3. **Parameters Explanation**:
   * The parameters of the nct distribution give insights into the shape and scale of Prabhsimran Singh's scoring pattern.
   * 2.5267304564172932.5267304564172932.526730456417293: Degree of freedom parameter for the t-distribution.
   * 13.8567886094655413.8567886094655413.85678860946554: Location parameter.
   * −14.016237667672424-14.016237667672424−14.016237667672424: Non-centrality parameter.
   * 1.76681643088200341.76681643088200341.7668164308820034: Scale parameter.

#### Conclusion

The analysis demonstrates that Prabhsimran Singh's runs scored in the IPL over the last three years can be best modeled using the nct distribution. This consistent pattern across years highlights the reliability of his performance and provides a statistical basis for understanding his scoring trends. The high p-values underscore the robustness of the fitted distribution, making it a reliable tool for predicting future performance and making data-driven decisions in team management and player strategy.