IPL PRIMARY DATA ANALYSIS

I have solved all primary analysis data given in PDF on IPL data set

Given Question are following-

1. Top 10 batsmen based on past 3 years total runs scored.

2. Top 10 batsmen based on past 3 years batting average. ( min 60 balls faced in each season)

3. Top 10 batsmen based on past 3 years strike rate (min 60 balls faced in each season)

4. Top 10 bowlers based on past 3 years total wickets taken.

5. Top 10 bowlers based on past 3 years bowling average. ( min 60 balls bowled in each season)

6. Top 10 bowlers based on past 3 years economy rate. ( min 60 balls bowled in each season)

7. Top 5 batsmen based on past 3 years boundary % (fours and sixes).

8. Top 5 bowlers based on past 3 years dot ball %.

9. Top 4 teams based on past 3 years winning %.

10.Top 2 teams with the highest number of wins achieved by chasing targets over the past 3 years.

**1. In this question we want top 10 highest run getter in last 3 season**

* First we use Batting data set we will use it given some data attributes Name ,runs, boundary, s/r, balls face, which position that batted
* I will start from make name taken name column and then use duplicate function to remove duplicate names from column
* Then use SUMIF function to get total runs of different players =SUMIF($E$2:$E$3269,[@batsmanName],$G$2:$G$3269)
* Then we get the runs by the players I am use pivot table it will given the row and value
* Then we use filter to get the top ten payers

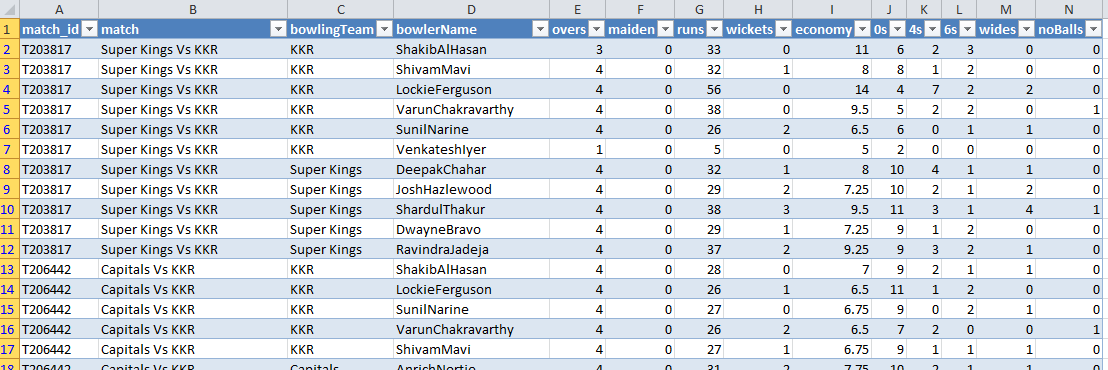
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Bats man Name** | **Average strick rate** | **Ball** | **not out** | **run** |
| Virat Kohli | 105.6668889 | 1090 | 538 | 1385 |
| KL Rahul | 109.0547222 | 1148 | 538 | 1516 |
| Jos Buttler | 115.1452632 | 1027 | 1830 | 1509 |
| Ruturaj Gaikwad | 115.7333333 | 1157 | 1830 | 1593 |
| Shikhar Dhawan | 115.9205 | 1081 | 1830 | 1392 |
| Shubman Gill | 120.899 | 1331 | 1830 | 1851 |
| Fafdu Plessis | 125.1073913 | 1300 | 1830 | 1831 |
| Sanju Samson | 125.9617778 | 902 | 1830 | 1304 |
| Surya kumar Yadav | 138.7589474 | 763 | 538 | 1225 |
| Glenn Maxwell | 150.4987805 | 752 | 1830 | 1214 |

**2. I have to make top 10 list of batsmen best batting average**

* After doing previous task then we get total runs then we have to apply the function to get the how many time player get out in total matches
* After get the average then we have to apply pivot table then taken top 10 best average players

**4.** **Top 10 bowlers based on past 3 years total wickets taken.**

* First use the batting data set it different attributes



* To get the total wickets copy column then use remove duplicate function to remove duplicate name of players
* Then apply SUMIF function for get total no. of wicket by different players use function

=SUMIF($D$2:$D$2437,P2,$H$2:$H$2437)

* Then I am apply pivot table function to get the top10 wicket takers

|  |  |
| --- | --- |
| **Row Labels** | **Sum of total wicket** |
| ArshdeepSingh | 45 |
| AveshKhan | 47 |
| HarshalPatel | 65 |
| KagisoRabada | 45 |
| MohammedShami | 67 |
| RashidKhan | 63 |
| ShardulThakur | 43 |
| TrentBoult | 42 |
| VarunChakravarthy | 44 |
| YuzvendraChahal | 66 |
| **Grand Total** | **527** |

**5. Top 10 bowlers based on past 3 years bowling average. (min 60 balls bowled in each season)**

* In this task we are using the previous data in total wickets
* For bowling average we use total runs conceded by the bowler per wicket
* So we have total wicket then get the total runs conceded use the function =SUMIF($D$2:$D$2437,P2,$G$2:$G$2437)
* After this we apply =R2/Q2 to get the bowling average
* Then we have to apply then total overs thrown by the bowler apply limit using filters on it use filter on it by use function

=SUMIF($D$2:$D$2437,P2,$E$2:$E$2437) and apply filter in overs by 10overs

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Bowler name** | **total wicket** | **Run conced** | **Bowling average** | **overs** |
|  | ShakibAlHasan | 4 | 187 | 46.75 | 26 |
|  | ShivamMavi | 16 | 460 | 28.75 | 54.1 |
|  | LockieFerguson | 26 | 747 | 28.73 | 84.8 |
|  | VarunChakravarthy | 44 | 1209 | 27.48 | 159.4 |
|  | SunilNarine | 36 | 1056 | 29.33 | 160 |
|  | VenkateshIyer | 3 | 115 | 38.33 | 12.3 |
|  | DeepakChahar | 27 | 707 | 26.19 | 84 |
|  | JoshHazlewood | 34 | 746 | 21.94 | 90.3 |
|  | ShardulThakur | 43 | 1220 | 28.37 | 127.9 |
|  | DwayneBravo | 30 | 562 | 18.73 | 67.2 |
|  | RavindraJadeja | 37 | 1014 | 27.41 | 136 |
|  | AnrichNortje | 31 | 768 | 24.77 | 92.4 |
|  | RavichandranAshwin | 33 | 1175 | 35.61 | 156 |

After get this table we are apply pivot table to get the top 10 data

**6. Top 10 bowlers based on past 3 years economy rate. (min 60 balls bowled in each season)**

* In this task we are using the previous data in total wickets or Bowling average.
* For bowling best economy rate calculate by the total run conceded by in total overs
* So we have total wicket then get the total runs conceded use the function =SUMIF($D$2:$D$2437,P2,$G$2:$G$2437)
* Then we have to apply then total overs thrown by the bowler apply limit using filters on it use filter on it by use function

=SUMIF($D$2:$D$2437,P2,$E$2:$E$2437) and apply filter in overs by 10overs

* Then we have to apply the total runs/total overs to get the best economy =R2/T2
* Many data we already show in the previous table

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Bowler name** | **total wicket** | **Run conced** | **Bowling average** | **overs** | **Economy** |
|  | ShakibAlHasan | 4 | 187 | 46.75 | 26 | 7.19 |
|  | ShivamMavi | 16 | 460 | 28.75 | 54.1 | 8.50 |
|  | LockieFerguson | 26 | 747 | 28.73 | 84.8 | 8.81 |
|  | VarunChakravarthy | 44 | 1209 | 27.48 | 159.4 | 7.58 |
|  | SunilNarine | 36 | 1056 | 29.33 | 160 | 6.60 |
|  | VenkateshIyer | 3 | 115 | 38.33 | 12.3 | 9.35 |
|  | DeepakChahar | 27 | 707 | 26.19 | 84 | 8.42 |
|  | JoshHazlewood | 34 | 746 | 21.94 | 90.3 | 8.26 |
|  | ShardulThakur | 43 | 1220 | 28.37 | 127.9 | 9.54 |
|  | DwayneBravo | 30 | 562 | 18.73 | 67.2 | 8.36 |
|  | RavindraJadeja | 37 | 1014 | 27.41 | 136 | 7.46 |
|  | AnrichNortje | 31 | 768 | 24.77 | 92.4 | 8.31 |
|  | RavichandranAshwin | 33 | 1175 | 35.61 | 156 | 7.53 |

* Then we apply the pivot table to get the top 10 data of bowler have best economy

**7. Top 5 batsmen based on past 3 years boundary % (fours and sixes).**

* In this case we are using the batting data set then we already get the in previous questions
* We get the total no of 6&4 by the batters and sum of runs by them in the form of boundary by using the function =(SUMIF($E$2:$E$3269,[@batsmanName],$I$2:$I$2369))\*4+(SUMIF($E$2:$E$3269,[@batsmanName],$J$2:$J$2369))\*6
* Then we apply total runs by BOUNDARY/total runs by batsman by =(Q3/P3)\*100

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
| **batsmanName** | **Ball** | **run** | **Run by boundary** | **Boundary %** |
| PatCummins | 80 | 156 | 124 | 79.48718 |
| DewaldBrevis | 113 | 161 | 122 | 75.7764 |
| AndreRussell | 468 | 745 | 564 | 75.7047 |
| KyleMayers | 246 | 365 | 276 | 75.61644 |
| SunilNarine | 112 | 154 | 116 | 75.32468 |
| YashasviJaiswal | 744 | 1132 | 844 | 74.5583 |
| RashidKhan | 172 | 304 | 224 | 73.68421 |
| EvinLewis | 149 | 224 | 164 | 73.21429 |
| PrabhsimranSingh | 272 | 391 | 284 | 72.63427 |
| MatthewShort | 92 | 117 | 84 | 71.79487 |

**8. Top 5 bowlers based on past 3 years dot ball %.**

* In the case we are using the previous task then we use this function to get the dot balls =SUMIF($D$2:$D$2437,P2,$J$2:$J$2437)
* After that we get the total balls bowled by the bowler by the function multiply total overs by the 6 ball get total balls thrown get % dot balls =T2\*6
* Then we apply the percentage on it and get the values in the table form

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Bowler name** | **total wicket** | **Run conced** | **Bowling average** | **overs** | **Economy** | **Dot balls** | **total Ball boweled** | **Dot ball %** |
| ShakibAlHasan | 4 | 187 | 46.75 | 26 | 7.19 | 50 | 156 | 32.05% |
| ShivamMavi | 16 | 460 | 28.75 | 54.1 | 8.50 | 140 | 324.6 | 43.13% |
| LockieFerguson | 26 | 747 | 28.73 | 84.8 | 8.81 | 208 | 508.8 | 40.88% |
| VarunChakravarthy | 44 | 1209 | 27.48 | 159.4 | 7.58 | 361 | 956.4 | 37.75% |
| SunilNarine | 36 | 1056 | 29.33 | 160 | 6.60 | 360 | 960 | 37.50% |

* Final result after we apply the pivot table in it

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Bowler name** | **total wicket** | **Run conced** | **Bowling average** | **overs** | **Economy** | **Dot balls** | **total Ball boweled** | **Dot ball %** |
| SimarjeetSingh | 4 | 138 | 34.50 | 18 | 7.67 | 57 | 108 | 52.78% |
| MoisesHenriques | 4 | 45 | 11.25 | 10 | 4.50 | 31 | 60 | 51.67% |
| MohsinKhan | 17 | 304 | 17.88 | 43 | 7.07 | 130 | 258 | 50.39% |
| ChrisWoakes | 5 | 82 | 16.40 | 11 | 7.45 | 33 | 66 | 50.00% |
| MarkWood | 11 | 130 | 11.82 | 16 | 8.13 | 48 | 96 | 50.00% |

**9. Top 5 teams with the highest number of wins achieved by chasing targets over the past 3 years**

* .In the data we are using the match summary data to get the this information
* Apply =UNIQUE() function to get the different teams in the teams column.
* And after that we are get the count the total wins by the every teams. By function =COUNTIF(C2:C207,G2)
* Then we apply the counif function to get win while chasing =COUNTIFS(C2:C207,[@Team],D2:D207,"\*\*"&"runs")

|  |  |  |
| --- | --- | --- |
| **Team** | **Total win** | **total win while chasing** |
| Super Kings | 25 | 14 |
| KKR | 21 | 7 |
| RCB | 25 | 14 |
| Mumbai | 20 | 7 |
| Punjab Kings | 19 | 11 |
| Sunrisers | 13 | 4 |
| Capitals | 21 | 7 |
| Royals | 22 | 13 |
| Titans | 23 | 9 |
| Super Giants | 17 | 11 |

* The we apply the pivot table to solve it

|  |  |
| --- | --- |
| **Row Labels** | **Average of total win while chasing** |
| Punjab Kings | 11 |
| RCB | 14 |
| Royals | 13 |
| Super Giants | 11 |
| Super Kings | 14 |
| **Grand Total** | **12.6** |

**10. Top 4 teams based on past 3 years winning %.**

* We are using match summary data to get the winning % of the data
* So we have to get the total no of match played
* Get the data by using function =COUNTIF($A$2:$A$206,G2)+COUNTIF($B$2:$B$206,G2)

Then we have already total win match

* Then apply the win % by using the formula =H2/J2 so that we use the % function in that value then apply in the pivot table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Team** | **Total win** | **total win while chasing** | **total match played by the team** | **Win by%** |
| Super Kings | 25 | 14 | 44 | 56.82% |
| KKR | 21 | 7 | 45 | 46.67% |
| RCB | 25 | 14 | 45 | 55.56% |
| Mumbai | 20 | 7 | 44 | 45.45% |
| Punjab Kings | 19 | 11 | 42 | 45.24% |
| Sunrisers | 13 | 4 | 41 | 31.71% |
| Capitals | 21 | 7 | 43 | 48.84% |
| Royals | 22 | 13 | 45 | 48.89% |
| Titans | 23 | 9 | 32 | 71.88% |
| Super Giants | 17 | 11 | 29 | 58.62% |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Team** | **Total win** | **total win while chasing** | **total match played by the team** | **Win by%** |
| Super Kings | 25 | 14 | 44 | 56.82% |
| Titans | 23 | 9 | 32 | 71.88% |
| Super Giants | 17 | 11 | 29 | 58.62% |

* Top 3 teams wins