Assignment 7 – EDA

Import the attached CSV files (IPL\_Dataset.csv) and answer the following questions:

Q1. What is the maximum number of matches played by an individual player in a season? Print the player name along with the number of matched played.

Q2. Top 2 players with maximum Average who have scored atleast 2 half centuries ?

Q3. Create 2 new columns based on Player name. First column will have first name and second column will have last name. Eg: for the player Shikhar Dhawan, Shikhar will be the first name and Dhawan will be the last name.

Q4. Create a new column (Cleaned\_Highest\_score) based on Highest score variable. Remove the Asterik(\*) mark and convert the data type into INT.

Q5. Print the total number of centuries scored in the entire season.

Q6. Print all the player names whose strike rate is less than the average strike rate of all players in entire season. Print the player name, his strike rate and average strike rate.

Q7. Please check the correlation between the features and create a heat map.

Q8. Check the list of players who has an average greater than 50 as well strike rate above 120. Print player name, average and strike rate.

Q9. Please check the list of players who has an average greater than 40 and balls faced above 100. Print player name, average and balls faced.

Q10. Players who scored atleast one century in this season. Create visualization.

Q11. Players who scored atleast 4 half centuries in this season.

Q12. Check the list of players who hit more than 45 boundaries and more than 10 sixes in this season.

Q13. Plot a histogram of number of matches played in a season by players.

Q14. Plot the histogram of balls faced by players.

Q15. Top 10 players with most runs in a season.

Q16. Print the players who played the match but didn't get the batting.

Q17. Create a new column to show the percentage of total runs scored in 4s and 6s. Then print the top 5 players with maximum percentage.

Q18. Print the players with top 5 Not out percentages (Not Out percentage can be calculated as number of Not outs divided by Innings).

Q19. Create visualization of top 10 players with highest number of sixes.

Q20. Scatter plot of runs scored by a player v/s balls faced in a season. Then find the relationship between these 2 variables.

