**Capstone Project Submission**

**Instructions:**

i) Please fill all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)** |
| **Problem Statement:**  We want to find out:-   1. The Number of 6's Scored in a season 2. Total runs scored 3. Number of half centuries by each batsman 4. Number of centuries by each batsman 5. Comparison between the teams 6. Batsman who scored the most number of dot balls. 7. Most common dismissal type in IPL 8. Bowler having more than 100 wickets in IPL 9. No of matches each season 10. No of matches on a particular Venue 11. No of matches played by each team 12. Number of wins per team 13. Top players of the match 14. How lucky are the toss wining team 15. Batsman Analysis 16. The Number of 4's Scored in a season   **Approach:** So what can we expect from the multiple IPL datasets? It contains 6 datasets:   1. matches.csv 2. Players.xlsx 3. Deliveries.csv 4. Teams.csv 5. most\_runs\_average\_strikerate.csv 6. teamwise\_home\_and\_away.csv   First we need to explore each dataset and check each columns. Replace missing values in all the datasets and drop few columns too which were not apart of our analysis. We need to check for the unique values in particular column and if needed replace values with similar meaning it with a specific unique name.  From players dataset we can calculate most common bating style and bowling style. We can also get information about country from which most player are coming.  From teams dataset we can get names of teams that playing IPL and from matches dataset we can get information about winning team so we can calculate and compare matches won by teams.  From matches data set we get information about city that hosted most number of matches and we can also calculate number of matches played per IPL season. Matches dataset also contains data about player that won ‘player of match’ award so we can calculate who won the award most number of times. Matches dataset contains information about toss winning, match winning and toss decision. So we can check if winning a toss affects result of winning the match and we can also calculate what decision is made most frequently when a team won the toss.  Delivery dataset contains information about each ball delivery, we can calculate number of 4 and 6 runs from each delivery and group them according to seasons to count number of boundaries per season. We can also calculate each players run individually then we can calculate number of centuries and half centuries made by each player. We can calculate number of dot balls played by each player and check who played most of it.  Delivery data set also contain information for player dismissal so we can calculate who took most wickets and what is most common type of dismissal is in IPL.  From most run average strike rate dataset we can calculate we can calculate most run made by batsman, average run rate and strike rate of top batsmen.  From teamwise home and away dataset we can get each teams’ win in home town and at other venues.  **Conclusion:**   1. The number of matches each season is lies between 50 to 80 matches. 2. The most number of Matches were hosted in Mumbai and the least were hosted in Venue Bloemfontein. 3. The most number of players were from India and the least number of players were playing from Netherland. 4. After winning the toss most of the teams choose to Field. 5. The most number of Matches were won by Mumbai Indians with around 109 wins. 6. In team comparison we observed that Kolkata Knight Riders lose most of the matches against Mumbai Indians loosing more than 19 matches. 7. The batsman who played most number of dot balls is Virat Kohli. 8. The most Dismissal type in IPL is caught out. 9. The most Half Centuries was scored by D. A. Warner. 10. The Player who won most of the Man of The Match was Chris Gayle 11. The average number of Four's and Six's per season lies somewhere around 1800 and 850 respectively. 12. The most full centuries was scored by Chris Gayle. 13. The most common bowling style is right-arm medium. 14. Most of the batsman are right-handers 15. V. Sehwag have the highest strike rate. 16. S.L. Malinga have the highest number of wickets in the IPL |
| **Please paste the drive link to your deliverables folder. Ensure that this folder consists of the project Colab notebook, project presentation and video.** |
|  |