Capstone Project Submission

Team Member's Name, Email and Contribution:

The Indian Premier League (IPL) is a professional Twenty20 cricket league, contested by eight teams based out of eight different Indian cities/states. In this EDA Project, we were provided with six datasets: players, matches, deliveries, most runs_avg_strikerate, teams and teamwise_home_and_away_win.

As the **First step**, the datasets are made free from NaN values by removing the NaN values and replacing it with other values:

1. **Deliveries df** had ~510000 null values and those are mainly in the player dismissed column and those are removed through filling the na values with zeros as it is more intuitive.
2. **Matches df** had ~1000 entries as null so that they are filled with one hot encoding, and replacing the null values by web surfing since there is no open source API available at this moment.
3. **Players df** had ~350 null values and those are replaced with mean age, and the country column is filled using the same distribution from the players df.
4. **Players Stats df** had ~30 null values which are replaced by calculating the average of the players.
5. **Teams df** had no null values.
6. **Teams home and away wins df** had no null values.

And the second step is the **Exploratory Data Analysis(EDA) part,** where the data is correlated and the trends in the data are discussed. The statistics obtained are as follows:

❖ **By Teams:**
  ➢ Highest Runs
  ➢ Lowest runs
  ➢ Most Wins
  ➢ Least Wins
  ➢ Most Winning Percentage
  ➢ Least Winning Percentage

❖ **By Players:**
  ➢ Most runs
  ➢ Most Sixes
  ➢ Most Fours
  ➢ Highest Strike Rate
  ➢ Highest Average

| **Contributed By:** |
| Akash.K |

| **Please paste the GitHub Repo link.** |

| Github Link:-
https://github.com/AkashKarthikeyan/IPL_T20_Analysis_EDA_Capstone_Project |

| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)** |

| |