MCT Format

This is a sample format for your MCT. You have to perform exploratory data analysis (EDA) on a provided dataset using Python libraries such as NumPy, Pandas, Statistics, Matplotlib, and Seaborn. You will answer questions and complete tasks related to understanding, cleaning, exploring, and visualizing the data.

1. **Import Libraries:** Begin by importing the necessary libraries: NumPy, Pandas, Statistics, Matplotlib, Seaborn, etc.
2. **Load Data:** Read the CSV file into a Pandas DataFrame.
3. **Data Cleaning:**
    - Check for missing values and handle them appropriately (e.g., dropping rows/columns, imputation).
    - Explore the data types of each column and ensure they are correct.
    - Identify and handle any outliers or anomalies.
4. **Data Understanding:**
    - Describe the dataset: number of rows and columns, data types, etc.
    - Summarize the central tendency and dispersion of numerical features (e.g., mean, median, standard deviation).
    - Explore the distribution of categorical features.
5. **Data Exploration and Visualization:**
    - Create histograms to visualize the distribution of numerical features.
    - Visualize the distribution of categorical features (e.g., 'Category' and 'City') using bar charts or pie charts to understand the frequency of each category.
    - Use boxplots to compare distributions across different groups or categories.
    - Explore relationships between variables using scatter plots or correlation matrices.
    - Generate bar charts or pie charts to visualize categorical data.
    - Employ Seaborn to create more advanced visualizations like pairplots, heatmaps, etc.
6. **Insights and Conclusion:**
    - Summarize key findings from the EDA.
    - Discuss any interesting patterns or relationships discovered.
    - Draw conclusions and propose potential avenues for further analysis.
7. **Record Explanation Video:**
    - Record a 5 to 10-minute video summarizing the project's key concepts and insights

Questions:

1. Analyze the distribution of total runs across different overs. Are there specific phases of the innings (e.g., powerplay, middle overs, death overs) where scoring rates tend to be higher or lower?
2. Explore the relationship between the batsman_runs and the type of dismissal (dismissal_kind). Do certain dismissal types tend to occur after batsmen have scored a significant number of runs?
3. Compare the average number of total runs scored per over by different batting_teams. Do certain teams have consistently higher or lower scoring rates?
4. Analyze the frequency and distribution of extras (extra_runs) conceded by bowling_teams across different overs. Are there specific phases of the game where teams tend to give away more extras?
5. Explore the relationship between batsman_runs and the number of balls faced. Do certain batsmen score runs at a faster pace than others?
6. Analyze the distribution of dismissal_kind across different bowling_teams. Do certain teams have a higher proportion of specific dismissal types (e.g., more catches, more bowled dismissals)?
7. Explore the relationship between the over and the number of wickets (player_dismissed) taken. Are there specific overs where bowlers tend to be more successful in taking wickets?
8. Analyze the performance of individual bowlers in terms of their economy rate (average runs conceded per over) against different batting_teams. Do certain bowlers have more success against specific teams?
9. Explore the relationship between the type of dismissal_kind and the fielder involved. Do certain fielders specialize in particular types of dismissals (e.g., catches, run-outs)?
10. Compare the bowling strike rate (average balls bowled per wicket) of different bowlers across different seasons. Are there any noticeable trends or changes in bowling effectiveness over time?
11. Analyze the distribution of runs scored by batsman across different batting positions (e.g., opening batsmen, middle-order batsmen). Do certain batting positions tend to score more runs?
12. Explore the relationship between the batsman and the bowler in terms of dismissal frequency. Do certain batsmen struggle against specific bowlers?
13. Analyze the runs scored by different pairs of batsman and non_striker to identify successful partnerships. Are there specific partnerships that consistently contribute significantly to the team's total score?
14. Compare the strike rate (runs scored per 100 balls) of individual batsmen against different bowling_teams. Do certain batsmen perform better against specific teams?
15. Analyze the frequency of boundaries (fours and sixes) hit by different batsman. Do certain batsmen have a more aggressive batting style with a higher proportion of boundaries?

16. Analyze the average total runs scored per match across different seasons. Has the overall scoring rate in IPL matches changed over time?
17. Explore the trends in dismissal types (dismissal_kind) across different seasons. Are there any noticeable shifts in how batsmen are getting out over time?
18. Analyze the impact of the inning (first or second) on the total runs scored. Do teams tend to score differently depending on whether they bat first or second?
19. Explore the relationship between the match id and the number of wide_runs bowled. Are there specific matches with an unusually high number of wides?
20. Analyze the distribution of noball_runs across different seasons. Has the frequency of no-balls changed over time?