VISUALIZATION CODE SUMMARY

I'll provide a summary of the visualization code and explain the key approaches used:

Data Loading and Cleaning:

- The dataset is loaded using pd.read_csv.
- Missing values are checked using .isnull().sum() and data types are inspected using .info().
- The 'match_date' column is converted to datetime format.

Data Preprocessing:

- Rows with batting statistics are filtered based on specific conditions.
- Asterisks ('*') are removed from the 'runs_scored' column and the column is converted to an integer data type.
- The last 7 matches' average runs scored for each player are calculated and added to a new DataFrame called 'batting_stats'.

Data Merging:

- The 'batting_stats' DataFrame is merged with the original dataset 'df' based on 'player_name' using a left join.

Handling Missing Values:

- A function calculate_bowling_stats is defined to handle missing values in bowling statistics ('wickets', 'catches', 'stumpings', 'runs_conceded').
- Missing values in these columns are replaced with zeros ('0').

Feature Engineering:

- New columns like 'DNB', 'TDNB', 'year', and 'years_of_experience' are created.
- 'DNB' and 'TDNB' values are handled and converted to integers.
- 'year' is extracted from the 'match_date' column.
- 'years_of_experience' is calculated as the difference between the maximum and minimum years for each player and added to 'df'.

Bowling Average Calculation:

- Bowling averages are calculated for each player and added as a new column named 'bowling_average'.
- The 'bowling_average' is calculated as the sum of 'runs_conceded' divided by the sum of 'wickets' for each player.

Data Visualization:

- Various data visualizations are created using seaborn and matplotlib to explore different aspects of the dataset, including runs scored, wickets taken, player experience, and more.

Feature Engineering - Mean and Aggregations:

- Mean runs, wickets, and runs conceded per year for each player are calculated and added as new columns.
- Mean runs, wickets, and runs conceded per opposition for each player are calculated and added as new columns.
- The 'means' DataFrame is merged with 'df'.

Correlation and Heatmap:

- Correlation matrices and heatmaps are created to explore relationships between variables.

Data Export:

- The preprocessed and shuffled DataFrame is exported to a CSV file named 'cricket2--.csv' without including the index column.

This code essentially loads, cleans, preprocesses, and analyzes the cricket dataset. It includes various data transformations, handling of missing values, feature engineering, data visualization, and correlation analysis to gain insights from the data.