EDA

Project Title: Exploring the Cricket World Cup: An EDA Journey

Objective:

To use the Cricket World Cup dataset (1975 - Present) to conduct a detailed Exploratory Data Analysis, applying various statistical and data visualization techniques.

Dataset

Cricket_Dataset.zip 2045.2KB

Project Overview:

Students will explore the rich history and dynamics of the Cricket World Cup through data. They will delve into innings, matches, team evolution, venues, umpires, and player styles, utilizing various EDA techniques.

Data Description:

The dataset includes data on innings, matches, teams, venues, umpires, and notable performances like centuries. It spans from the inaugural World Cup in 1975 to the present.

Key EDA Techniques to be Used:

- Statistical Analysis: Understand basic statistics like mean, median, mode, and standard deviation of various metrics (e.g., runs scored, wickets taken).
- Data Visualization: Use plots (bar, line, pie charts), heatmaps, and histograms to visualize distributions and trends.
- Correlation Analysis: Find correlations between different variables like team performance and venue.
- Data Cleaning and Preprocessing: Handle missing data, outliers, and transform variables for analysis.

Project Tasks:

- **Data Acquisition and Preprocessing:**
  - Import the dataset using Python libraries like Pandas.
  - Conduct data cleaning, handling missing values, and outlier detection.
  - Perform initial data preprocessing to format and structure the data for analysis.
- **Exploratory Data Analysis Using Matplotlib and Seaborn:**
  - Use Matplotlib and Seaborn for creating a variety of visualizations to explore the data.
  - Plot histograms, bar charts, and line graphs to understand the distribution of runs, wickets, and player performances over different World Cups.
  - Employ scatter plots to examine relationships between variables such as team performance and match locations.
  - Create heatmaps to visualize correlations between different statistical metrics.
- **In-depth Data Visualization and Analysis:**
  - Innings Visualization: Analyze batting and bowling stats using appropriate plots to showcase trends and outliers.
  - Match Analysis: Visualize match outcomes, duration, and scores over different tournaments.
  - Team Performance: Create visualizations to track the evolution and performance of different teams across World Cups.
  - Venue Analysis: Use geographical plotting (if possible) or bar charts to highlight the impact of venues on match outcomes.
  - Century Analysis: Plot timelines or bar graphs to showcase century scorers and their impact on match results.
  - Umpire Decision Analysis: Investigate any trends or patterns in umpire decisions using categorical plots.
- **Interactive Data Exploration:**
  - (Optional) Incorporate interactive elements using libraries like Plotly to allow dynamic exploration of the data.

Reference Notebook:

exploratory-data-analysis-eda.ipynb 17.7KB