GUVI

DATA SCIENCE JOB FAIR-2024

DATA / BUSINESS ANALYST

Sports Analytics

Objective:

The primary goal of this data analysis project is to evaluate the fitness and performance of cricket players in preparation for an upcoming T20 game. The analysis will focus on understanding each player's current form and historical performance. The insights gained will aid team management in making informed decisions regarding player selection and strategies for the upcoming game.

Data Understanding:

The data includes player details such as full name, date of birth, playing style, teams played for, and career statistics including batting and bowling averages, high scores, wickets, and more. The information is provided in JSON format.

Data Collected From: ESPN

Dataset: Link

Features:

PlayerName: The name of the cricket player.

PersonDetail: Details about the player,playing style, and teams played for.

TopStats: Top-level statistics for each player.

Stats: Detailed career statistics, including batting, bowling, and match-specific details.

Approach:

- Import data from all excel files
- Data Cleaning and Preprocessing:
    - Parse and extract relevant information from JSON formatted data in the columns.
    - Handle missing or inconsistent data.
- Exploratory Data Analysis (EDA):
    - Explore the distribution of player attributes such as age, batting style, and bowling style.
    - Analyze the distribution of key career statistics (runs, wickets) across the dataset.
    - Identify any correlations between player attributes and performance metrics.
- Performance Analysis:
    - Compare the performance of players in different formats.
    - Analyze batting and bowling averages, strike rates, and other relevant metrics.
    - Identify standout performances and consistent performers.
    - Investigate player performance in different conditions (home vs. away, day vs. night matches and so on).
    - Explore performance trends in specific tournaments or series.
    - Create visualizations (bar plots, scatter plots, etc.) to represent key insights.
    - Visualize trends in player performance over time.
    - Present comparisons between players in a visually informative manner.

Submission:

- Provide a well-commented Python file (*.ipynb/*.py) or (*.pbix) containing the complete code for the project, organized into sections for data preprocessing and Analysis.
- Upload the same into github with proper Readme file.
- Presentation on the entire project, including Problem Statement, Tools Used, Approaches and Insights Found.

Evaluation metrics:

- Project evaluation will be done in the live session and have to showcase the approaches done to complete the project
- You are supposed to write a code in a modular fashion (in functional blocks)
- Maintainable: It can be maintained, even as your codebase grows.
- Portable: It works the same in every environment (operating system)
- You have to maintain your code on GitHub.(Mandatory)
- You have to keep your GitHub repo public so that anyone can check yourcode.(Mandatory)
- Proper readme file you have to maintain for any project development(Mandatory)
- Follow the coding standards: https://www.python.org/dev/peps/pep-0008/
- You should include basic workflow and execution of the entire project in the readme file on GitHub

GitHub Repo:

The attached reference document will help you use GitHub effectively. - **Link**