Programming Competition - Hackathon

DBDA Aug-2019 Batch 15th Jan 20

Cricket Data Analysis

- The problem mentioned below revolves around cricket match dataset. The dataset contains 2 files which contains following details.
  - deliveries.csv
    - MATCH_ID
    - INNING
    - BATTING_TEAM
    - BOWLING_TEAM
    - OVER
    - BALL
    - BATSMAN
    - BOWLER
    - WIDE_RUNS
    - BYE_RUNS
    - LEGBYE_RUNS
    - NOBALL_RUNS
    - PENALTY_RUNS
    - BATSMAN_RUNS
    - EXTRA_RUNS
    - TOTAL_RUNS
  - matches.csv
    - MATCH_ID
    - SEASON
    - CITY
    - DATE
    - TEAM1
    - TEAM2
    - TOSS_WINNER
    - TOSS_DECISION
    - RESULT
    - WINNER
- The dataset is uploaded in git repository.

Sunbeam Institute of Information Technology, Pune & Karad

Programming Competition - Hackathon

DBDA Aug-2019 Batch 15th Jan 20

Solve the below questions.

1. Top 4 teams which elected to field first after winning toss in the year 2016 and 2017.
   - Output Expected

YEAR	TEAM	COUNT
2016		
2017		

2. List total number of fours, sixes, total score with respect to team and year.
   - Output Expected

YEAR	TEAM_NAME	FOURS_COUNT	SIXES_COUNT	TOTAL_SCORE

3. Top 10 best economy rate bowler with respect to year who bowled at least 10 overs (LEGBYE_RUNS and BYE_RUNS should not be considered for Total Runs Given by a bowler).
   - Economy = (Total Runs Given/Overs bowled)
   - Output Expected

YEAR	PLAYER	ECONOMY

4. Find the team name which has Highest Net Run Rate with respect to each year.
   - Net Run Rate = (Total Runs Scored / Total Overs Faced) – (Total Runs Conceded /Total Overs Bowled)

5. Find the team name which has Highest Net Run Rate in month of April (including all years).
   - Net Run Rate = (Total Runs Scored / Total Overs Faced) – (Total Runs Conceded /Total Overs Bowled)

- Technologies & tools
  - It is expected to solve each of the above problem using
    - MySQL or Hive
    - Java 8 (using Java 8 streams)
    - Python 3 (using collections)
    - Python 3 (using numpy & pandas)
    - Use matplotlib to display results in appropriate charts for each problem statement.
  - IDE: Pycharm or VScode or Eclipse or any other IDE.

- **Evaluation**
  - For each group "1" mark will be given for completion of all problem statements in all specified languages.
  - Obviously result should be same in all technologies.
- **Guidelines**
  - Students can refer and use documentations, classwork demos and online tutorials.
  - Students are expected to make group (of 4 students) as per project technologies.
  - Students are advised to use GIT for collaboration.

---------- ALL THE BEST ----------