|  |  |
| --- | --- |
| **Project Code** | Project 5 : Optional |
| **Project Name** | Load and Process IPL Match Data  (This is Enhancement Request on top of Project 1) |

|  |  |  |
| --- | --- | --- |
| **Prepared/Modified by** | **Role** | **Date of Preparation** |
| Charudatta Joshi | AWS Trainer | 31-05-2023 |
| **Reviewed by** | **Role** | **Date of Review** |
|  |  |  |
| **Approved by** | **Role** | **Date of Approval** |
|  |  |  |
| **Circulation List** |  | **Version Number of the template:** |
| **Version Number** | 1.0 | 1.0 |

**Dataset Details:**

**About Dataset**

**This data has ball by ball information for every match played between 2008 and 2022 seasons.**

**Context**:

The Indian Premier League (IPL) is a professional men's Twenty20 cricket league, contested by ten teams based out of ten Indian cities. The league was founded by the Board of Control for Cricket in India (BCCI) in 2007. It is usually held between March and May of every year and has an exclusive window in the ICC Future Tours Programme.

The IPL is the most-attended cricket league in the world and in 2014 was ranked sixth by average attendance among all sports leagues. In 2010, the IPL became the first sporting event in the world to be broadcast live on YouTube. The brand value of the IPL in 2019 was ₹47,500 crores (6.3 billion US dollars), according to Duff & Phelps. According to BCCI, the 2015 IPL season contributed ₹1,150 crores (150 million US dollars) to the GDP of the Indian economy. The 2020 IPL season set a massive viewership record with 31.57 million average impressions and with an overall consumption increase of 23 percent from the 2019 season.

**Tables in Dataset: -**

Table 1: Match Details (File Name : IPL\_Matches\_2008\_2022.csv)



Table 2: Ball\_By\_Ball\_Details (File Name : IPL\_Ball\_by\_Ball\_2008\_2022)



**Abbreviations:**

1. BF: Balls Faced
2. SR: Strike Rate (Batsman: Total Runs/Balls Faced)
3. POS: Position
4. Mat: Matches
5. Inns Innings
6. NO: Not Out
7. HS: Highest Score
8. AVG: Average ( Batsman: Total Runs / Innings)
9. Ov: Overs
10. Econ: Economy Rate of Bowler (Total runs conceded/ Overs balled, Round the half completed over to next whole number)
11. Wkts: Wickets
12. Maid: Maiden Over

**Infrastructure Setup:**

S3 for data store and Glue for data processing

**Mandatory Requirements**

**Requirement No 1: Most Dot Balls Innings**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| POS | Player | Ov | Runs | Wkts | Maid | Dots | SR | Against | Venue | Match Date |
| 1 | Sohail Tanvir | 4 | 10 | 3 | 0 | 18 | 8 | RCB | Sawai Mansingh Stadium | 17-May-08 |
| 2 | Dale Steyn | 4 | 28 | 2 | 0 | 18 | 12 | CSK | Chidambaram | 21-May-08 |
| 3 | Brett Lee | 4 | 9 | 1 | 0 | 17 | 24 | MI | IS Bindra Stadium | 25-Apr-08 |
| 4 | Shaun Pollock | 4 | 9 | 1 | 1 | 17 | 24 | CSK | Wankhede Stadium | 14-May-08 |
| 5 | Irfan Pathan | 4 | 10 | 1 | 0 | 17 | 24 | RR | IS Bindra Stadium | 28-May-08 |
| 6 | Shaun Pollock | 4 | 12 | 3 | 0 | 17 | 8 | KKR | Wankhede Stadium | 16-May-08 |
| 7 | Dale Steyn | 4 | 14 | 1 | 0 | 17 | 24 | MI | M. Chinnaswamy Stadium | 28-May-08 |
| 8 | Shanthakumaran Sreesanth | 4 | 16 | 2 | 0 | 17 | 12 | RCB | M. Chinnaswamy Stadium | 05-May-08 |
| 9 | Shane Watson | 4 | 19 | 1 | 0 | 17 | 24 | CSK | Sawai Mansingh Stadium | 04-May-08 |
| 10 | Vinay Kumar | 4 | 27 | 3 | 0 | 17 | 8 | DEC | Rajiv Gandhi Intl. Cricket Stadium | 25-May-08 |
| 11 | Glenn McGrath | 4 | 12 | 0 | 0 | 16 | 0 | RR | Sawai Mansingh Stadium | 11-May-08 |

**Acceptance Criteria:**

1. Separate JSON file for every season should be created
2. File should list statics of players as shown above
3. File should position players in descending order of dot balls thrown in a single innings
4. Create a output file with the same name as of title (Example: if title is “Good Aws Batch” file name should be “good\_aws\_batch.json”

**Requirement No 2: Most Runs Over**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| POS | Player | Mat | Inns | NO | Runs | HS | Avg | BF | SR | 100 | 50 | 4s | 6s |
| 1 | Sachin Tendulkar | 15 | 15 | 2 | 618 | 89\* | 47.53 | 466 | 132.61 | 0 | 5 | 86 | 3 |
| 2 | Jacques Kallis | 16 | 16 | 4 | 572 | 89\* | 47.66 | 494 | 115.78 | 0 | 6 | 67 | 9 |
| 3 | Suresh Raina | 16 | 16 | 5 | 520 | 83\* | 47.27 | 364 | 142.85 | 0 | 4 | 45 | 22 |
| 4 | Sourav Ganguly | 14 | 14 | 1 | 493 | 88 | 37.92 | 419 | 117.66 | 0 | 4 | 58 | 15 |
| 5 | Murali Vijay | 15 | 15 | 2 | 458 | 127 | 35.23 | 292 | 156.84 | 1 | 2 | 36 | 26 |
| 6 | Mahela Jayawardena | 13 | 13 | 3 | 439 | 110\* | 43.9 | 298 | 147.31 | 1 | 1 | 55 | 11 |
| 7 | Andrew Symonds | 16 | 16 | 2 | 429 | 54 | 30.64 | 341 | 125.8 | 0 | 4 | 35 | 18 |
| 8 | Saurabh Tiwary | 16 | 15 | 1 | 419 | 61 | 29.92 | 309 | 135.59 | 0 | 3 | 29 | 18 |
| 9 | Rohit Sharma | 16 | 16 | 2 | 404 | 73 | 28.85 | 302 | 133.77 | 0 | 3 | 36 | 14 |
| 10 | Naman Ojha | 14 | 14 | 2 | 377 | 94\* | 31.41 | 285 | 132.28 | 0 | 2 | 42 | 15 |

**Acceptance Criteria:**

1. Separate JSON file for every season should be created
2. File should list statics of players as shown above
3. File should position players in descending order of runs scored in a single innings
4. Create a output file with the same name as of title (Example: if title is “Good Aws Batch” file name should be “good\_aws\_batch.json”

**Requirement No 3: Best Bowling Economy Innings**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| POS | Player | Ov | Runs | Wkts | Dots | Econ | SR | Against | Venue | Match Date |
| 1 | Mohammad Hafeez | 4 | 8 | 1 | 14 | 2 | 24 | DEC | Eden Gardens | 20-Apr-08 |
| 2 | Ishant Sharma | 4 | 9 | 2 | 14 | 2.25 | 12 | RCB | M. Chinnaswamy Stadium | 18-Apr-08 |
| 3 | Brett Lee | 4 | 9 | 1 | 17 | 2.25 | 24 | MI | IS Bindra Stadium | 25-Apr-08 |
| 4 | Shaun Pollock | 4 | 9 | 1 | 17 | 2.25 | 24 | CSK | Wankhede Stadium | 14-May-08 |
| 5 | Dominic Thornely | 3 | 7 | 2 | 12 | 2.33 | 9 | KKR | Wankhede Stadium | 16-May-08 |
| 6 | Sourav Ganguly | 3 | 7 | 1 | 10 | 2.33 | 18 | RCB | Eden Gardens | 08-May-08 |
| 7 | Sohail Tanvir | 4 | 10 | 3 | 18 | 2.5 | 8 | RCB | Sawai Mansingh Stadium | 17-May-08 |
| 8 | Irfan Pathan | 4 | 10 | 1 | 17 | 2.5 | 24 | RR | IS Bindra Stadium | 28-May-08 |
| 9 | Siddharth Trivedi | 3 | 8 | 1 | 10 | 2.66 | 18 | DEC | Sawai Mansingh Stadium | 09-May-08 |
| 10 | Manpreet Gony | 4 | 11 | 2 | 13 | 2.75 | 12 | RCB | Chidambaram | 21-May-08 |
| 11 | MF Maharoof | 4 | 11 | 2 | 10 | 2.75 | 12 | RR | Arun Jaitley Stadium | 19-Apr-08 |

**Acceptance Criteria:**

1. Separate JSON file for every season should be created
2. File should list statics of players as shown above
3. File should position players in ascending order of “Economy Rate” for the single inning
4. Create a output file with the same name as of title (Example: if title is “Good Aws Batch” file name should be “good\_aws\_batch.json”

**Requirement No 4:** **Best Bowling Strike Rate Innings**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| POS | Player | Ov | Runs | Wkts | SR | Against | Venue | Match Date |
| 1 | Laxmi Ratan Shukla | 0.5 | 6 | 3 | 1.66 | DC | Eden Gardens | 13-May-08 |
| 2 | James Hopes | 1 | 2 | 2 | 3 | DC | Arun Jaitley Stadium | 17-May-08 |
| 3 | Sohail Tanvir | 4 | 14 | 6 | 4 | CSK | Sawai Mansingh Stadium | 04-May-08 |
| 4 | Shoaib Akhtar | 3 | 11 | 4 | 4.5 | DC | Eden Gardens | 13-May-08 |
| 5 | Amit Mishra | 4 | 17 | 5 | 4.8 | DEC | Arun Jaitley Stadium | 15-May-08 |
| 6 | Lakshmipathy Balaji | 4 | 24 | 5 | 4.8 | PBKS | Chidambaram | 10-May-08 |
| 7 | Yo Mahesh | 4 | 36 | 4 | 6 | MI | Arun Jaitley Stadium | 24-May-08 |
| 8 | Sohail Tanvir | 4 | 14 | 4 | 6 | MI | Sawai Mansingh Stadium | 26-May-08 |
| 9 | Dilhara Fernando | 4 | 18 | 4 | 6 | RCB | M. Chinnaswamy Stadium | 28-May-08 |
| 10 | Umar Gul | 4 | 23 | 4 | 6 | PBKS | Eden Gardens | 25-May-08 |

**Acceptance Criteria:**

1. Separate JSON file for every season should be created
2. File should list statics of players as shown above
3. File should position players in ascending order of “Strike Rate” for the single inning
4. Create a output file with the same name as of title (Example: if title is “Good Aws Batch” file name should be “good\_aws\_batch.json”