**Usecase – 1**

**IPL DATA:**

* **Data till Season 11 (2008 - 2019)**
* **matches.csv - Match by match data**
* **deliveries.csv - Ball by ball data**

**Datasets Description:**

**matches.csv:**

1. **id - unique id for a match**
2. **season – year of the match**
3. **city – city where the match took place**
4. **date – date of the match**
5. **team1 – team batting first**
6. **team2 – team batting second**
7. **toss\_winner – who won the toss**
8. **toss\_decision – bat / field**
9. **result – match result**
10. **dl\_applied – is Duckworth Lewis rule applied**
11. **winner – winner of the match**
12. **win\_by\_runs – win by runs**
13. **Win\_by\_wickets – win by wickets**
14. **player\_of\_match – player of the match**
15. **venue – match venue**
16. **umpire1 – Umpire 1**
17. **umpire2 – Umpire 2**
18. **umpire3 – Umpire 3**

**deliveries.csv:**

1. **match\_id - unique identifier for a match**
2. **inning – Match innings (1st innings / 2nd innings)**
3. **batting\_team – name of the batting team**
4. **bowling\_team – name of the bowling team**
5. **over – current over**
6. **ball – current ball of the over**
7. **batsman – name of the batsman on strike**
8. **non\_striker – name of the batsman on non – striker’s end**
9. **bowler – name of the bowler**
10. **is\_super\_over – Is this a super over(0 or 1)**
11. **wide\_runs – runs given as wide**
12. **bye\_runs – runs given as bye runs**
13. **legbye\_runs – runs given as legbye runs**
14. **noball\_runs – runs given as no-ball**
15. **penalty\_runs – runs given as penalty**
16. **batsman\_runs – runs scored by the batsman**
17. **extra\_runs – total extra runs(wide, bye, no-ball, penalty)**
18. **total\_runs – total runs from the ball (extra runs, batsman runs)**
19. **player\_dismissed – name of the player dismissed(if out)**
20. **dismissal\_kind – how the player was dismissed(if out)**
21. **fielder – fielder involved in the dismissal(if any)**

**From the given csv files above, Solve the below problems using Spark Actions and Transformations (Don’t use dataFrame API):**

1. **Find the number of matches that took place in Hyderabad.**
2. **Find the number matches that Sunrisers won in Hyderabad City.**
3. **Find the number of matches Sunrisers won in each season.**
4. **In which season did Yuvraj Singh has been awarded maximum Player of the match title.**
5. **Which venue hosted maximum matches.**
6. **Find the average win by runs in Eden Gardens in Season 2017.**

**Usecase – 2**

**From the IPL Dataset mentioned above, provide the solutions for below using Actions and Transformations:**

1. **Find out the percentage of each dismissal over all seasons.**
2. **Which stadium is best suited to bat first.**
3. **Which stadium is best suited to bowl first.**