**SPRINT 1**

There are five dataset games, appearances, lineup, events and players. I have merged them using the common columns and then treated them for null and duplicates. The dictionary for the final dataset is given below.

The Data Dictionary for the ‘Football\_Data\_Merged\_Cleaned’ file is:

1. **game\_id**: Unique identifier for each football match.
2. **competition\_id\_x**: Identifier for the competition type of the game.
3. **season**: Season in which the game took place.
4. **round**: The round or stage of the competition for the game.
5. **date\_x**: Date when the game occurred.
6. **home\_club\_goals**: Number of goals scored by the home team.
7. **away\_club\_goals**: Number of goals scored by the away team.
8. **home\_club\_position**: League position of the home team before the game.
9. **away\_club\_position**: League position of the away team before the game.
10. **home\_club\_manager\_name**: Name of the home team's manager.
11. **away\_club\_manager\_name**: Name of the away team's manager.
12. **stadium**: Stadium where the game was played.
13. **attendance**: Number of spectators who attended the game.
14. **referee**: Referee officiating the game.
15. **home\_club\_name**: Name of the home football club.
16. **away\_club\_name**: Name of the away football club.
17. **aggregate**: Aggregate score summary of the game.
18. **competition\_type**: Type of competition (e.g., domestic league, cup).
19. **game\_event\_id**: Unique identifier for events during the game.
20. **date\_y**: Date of the specific game event.
21. **minute**: Minute in the game when the event occurred.
22. **type\_x**: Type of event (e.g., goals, cards, substitutions).
23. **player\_id\_x**: Unique identifier for the player involved in the event.
24. **description**: Description of the game event.
25. **player\_in\_id**: Player substituted into the game.
26. **player\_assist\_id**: Player who assisted the event (if applicable).
27. **name**: Name of the player involved in the event.
28. **last\_season**: Last season the player played professionally.
29. **current\_club\_id**: Identifier for the player’s current club.
30. **player\_code**: Unique code for identifying the player.
31. **country\_of\_birth**: Player's country of birth.
32. **date\_of\_birth**: Player's date of birth.
33. **sub\_position**: Player's sub-position on the field.
34. **position\_x**: Player's main playing position.
35. **foot**: Dominant foot of the player (e.g., right, left).
36. **height\_in\_cm**: Player's height in centimeters.
37. **market\_value\_in\_eur**: Player's market value in Euros.
38. **highest\_market\_value\_in\_eur**: Player's highest recorded market value in Euros.
39. **contract\_expiration\_date**: Date when the player’s contract expires.
40. **agent\_name**: Name of the player’s agent.
41. **appearance\_id**: Unique identifier for player appearances.
42. **player\_id\_y**: Alternate unique identifier for the player.
43. **date**: Date when the player’s performance was recorded.
44. **player\_name\_x**: Name of the player recorded for the event.
45. **competition\_id\_y**: Identifier for the competition during the appearance.
46. **yellow\_cards**: Number of yellow cards received by the player.
47. **red\_cards**: Number of red cards received by the player.
48. **goals**: Number of goals scored by the player.
49. **assists**: Number of assists made by the player.
50. **minutes\_played**: Total minutes played by the player in the game.

**SPRINT 2**

The **business objectives** for each focus area and the corresponding **business questions** are given below.

**1) Performance Analysis**

**Business Objectives:**

* Evaluate the overall performance of players in terms of goals, assists, yellow/red cards, and minutes played.
* Identify top-performing players for specific matches and seasons.

**Business Questions:**

* Which players scored the most goals in each competition?
* Who were the top players with the most assists during the season?
* What is the average number of goals scored per player across all matches?
* How many yellow and red cards were given to each player in a season?
* Which player played the most minutes during the season?

**2) Player Profile and Market Value**

**Business Objectives:**

* Understand the market value trends of players.
* Examine the relationship between player performance and market value.

**Business Questions:**

* How does the player’s market value correlate with their performance (goals, assists, etc.)?
* Which players have the highest market values and how do they perform in comparison to others?
* What are the changes in a player’s market value over different seasons?

**3) Team Comparison**

**Business Objectives:**

* Compare teams’ performance metrics (goals scored, league position, and overall statistics).
* Identify winning and losing patterns between home and away teams.

**Business Questions:**

* Which team scored the most goals in a given season?
* How do home and away teams compare in terms of goals scored and goals conceded?
* What is the correlation between a team’s league position and their performance (goals, wins, etc.)?
* How do teams perform in different rounds of the competition?

**4) Attendance and Stadium Analysis**

**Business Objectives:**

* Analyze the impact of stadium location and crowd size on team performance.
* Investigate trends in attendance across different rounds and seasons.

**Business Questions:**

* What is the average attendance in different stadiums for the season?
* How does the home team’s performance correlate with stadium attendance?
* Which stadiums have the highest average attendance?

**5) Referee Analysis**

**Business Objectives:**

* Examine how referees influence match outcomes (yellow/red cards, penalties).
* Identify patterns in referee decisions across competitions.

**Business Questions:**

* How many yellow and red cards were given by each referee in the season?
* Which referees officiated the most high-profile matches (with close scores or many goals)?
* Are there any patterns in referee decisions that favor certain teams or players?

**6) Substitution Patterns**

**Business Objectives:**

* Understand substitution strategies and their impact on match performance.
* Analyze the timing of substitutions and their correlation with team performance.

**Business Questions:**

* At which game minute are most substitutions made?
* How do player substitutions affect match results (goals scored, assists, etc.)?
* Which players were substituted in most frequently?

**7) Events Analysis**

**Business Objectives:**

* Analyze the types of in-game events (goals, assists, cards) and their occurrence.
* Examine key events that have the highest impact on match outcomes.

**Business Questions:**

* What percentage of goals were scored by substitutes in comparison to starters?
* How many yellow and red cards were handed out in each round of the competition?
* Which events (goals, assists, cards) had the most impact on the outcome of matches?

**8) Competition Analysis**

**Business Objectives:**

* Assess the performance of teams across different competitions (domestic league, cup).
* Identify the most competitive and successful teams in each type of competition.

**Business Questions:**

* Which teams performed best in each competition type (domestic league, cup)?
* What is the win rate of teams in different rounds of the competition?
* How do teams perform in knockout rounds compared to group stages?

**9) Player Attributes and Demographics**

**Business Objectives:**

* Understand how player demographics (age, height, country of birth, etc.) influence performance.
* Analyze the dominance of certain player characteristics (foot, position, etc.) in the competition.

**Business Questions:**

* Do younger players perform better than older players in terms of goals and assists?
* How does a player’s height affect their performance in various positions (e.g., defenders vs. forwards)?
* Which player demographics (country, foot preference) are most common in high-performing teams?

**10) Contract Management**

**Business Objectives:**

* Track player contracts and contract expiration dates to manage team roster effectively.
* Analyze the relationship between contract expiration dates and player performance or transfer activities.

**Business Questions:**

* Which players are nearing the end of their contracts, and how does this affect their performance?
* Are players with upcoming contract expirations more motivated to perform better?
* What is the average contract length for players in high-performing teams?