IFT 598: Data Visualization & Reporting for IT

(Taha H. Ahmed, Evan Sequeira, Neeraj Patel, Saaransh Rana)

**Project - Phase I: Planning**

Dr. Asmaa Elbadrawy

November 19, 2021

Section 1: Dataset Description

1. Select a dataset of your choice & download it. Explain what this dataset is about.

FIFA 19 complete player dataset

<https://www.kaggle.com/karangadiya/fifa19>

The dataset contains 1000 rows and 89 columns.

The beautiful game of football! And the players who make this sport what it is. Personally, the dataset makes me appreciate the footballers even more by being able to see the incredibly high standards of fitness and performance they need to maintain, which they do, and how they produce phenomenal plays and results! This dataset contains detailed statistical information for every player registered in the latest edition of FIFA 19 (My personal favorite Lionel Messi topping the list) Detailed information such as the player’s age, height, weight, their finishing, heading, shot accuracy, pass accuracy, sprint speed et cetera are all present. These are accurate values since FIFA has licensed rights to using player portfolios, based on past and present performance statistics of the players measured over the football season.

1. For each column, specify its data type: Categorical, Ordinal, Interval or Ratio.
2. For each column, specify its domain; that is the list or range of values that it can take.

|  |  |  |
| --- | --- | --- |
| Column Name | Data Type | Domain (Range of values) |
| F1 | Interval | 0 – 99 |
| ID | Interval | 41 - 232363 |
| Name | Categorical | [L. Messi, Christiano Ronaldo, Neymar Jr, et cetera] |
| Age | Ratio | 19 - 40 |
| Photo | Categorical |  |
| Nationality | Categorical | [Argentina, Portugal, Brazil, et cetera] |
| Flag | Categorical |  |
| Overall | Ratio | 85 – 94 |
| Potential | Ratio | 85 – 94 |
| Club | Categorical | [FC Barcelona, Juventus, Paris Saint-Germain, et cetera] |
| Club Logo | Categorical |  |
| Value | Ratio | €18M - €118M |
| Wage | Ratio | €21K - €565K |
| Special | Ratio | 1473 - 2228 |
| Preferred Foot | Categorical | [Left, Right] |
| International Reputation | Interval | 1 – 5 |
| Weak Foot | Interval | 1 – 5 |
| Skill Moves | Interval | 1 – 5 |
| Work Rate | Categorical | [High/(High/Medium/Low), Medium/(High/Medium/Low)] |
| Body type | Categorical | [Lean, Normal, Stocky] |
| Real Face | Categorical | [Yes, No] |
| Position | Ordinal | [ST, CDM, CAM, LW, et cetera] |
| Jersey Number | Ordinal | 1       - 44 |
| Joined | Interval | (Min = 7/1/2004 & Max = 8/9/2018) |
| Loaned From | Categorical | [FC Barcelona, Real Madrid, Juventus, et cetera] |
| Contract Valid Until | Ratio | 2019 – 2026 |
| Height | Ratio | 5’7 - 6’6 |
| Weight | Ratio | 130lbs – 212lbs |
| LS (Left Striker) | Ratio | 49 – 91 |
| ST (Striker) | Ratio | 49 – 91 |
| RS (Right Striker) | Ratio | 49 – 91 |
| LW (Left Winger) | Ratio | 49 – 92 |
| LF (Left Forward) | Ratio | 48 – 90 |
| CF (Centre Forward) | Ratio | 48 – 93 |
| RF (Right Forward) | Ratio | 48 – 93 |
| RW (Right Winger) | Ratio | 49 – 92 |
| LAM Left Attacking Midfielder) | Ratio | 49 – 93 |
| CAM (Central Attacking Midfielder) | Ratio | 49 – 93 |
| RAM (Right Attacking Midfielder) | Ratio | 49 – 93 |
| LM (Left Midfielder) | Ratio | 53 - 91 |
| LCM (Left Centre Midfielder) | Ratio | 56 – 84 |
| CM (Centre Midfielder) | Ratio | 56 – 84 |
| RCM (Right Centre Midfielder) | Ratio | 56 - 84 |
| RM (Right Midfielder) | Ratio | 53 – 91 |
| LWB (Left Wing Back) | Ratio | 60 – 85 |
| LDM (Left Defensive Midfielder) | Ratio | 74 - 85 |
| CDM | Ratio | 35-90 |
| RDM | Ratio | 33- 88 |
| RWB | Ratio | 34-88 |
| LB | Ratio | 33-87 |
| LCB | Ratio | 32-88 |
| CB | Ratio | 32-87 |
| RCB | Ratio | 32-87 |
| RB | Ratio | 36-74 |
| Crossing | Ratio | 10-93 |
| Finishing | Ratio | 6-95 |
| HeadingAccuracy | Ratio | 8-96 |
| ShortPassing | Ratio | 15-93 |
| Volley | Ratio | 7-90 |
| Dribbling | Ratio | 7-97 |
| Curve | Ratio | 9-93 |
| FKAccuracy | Ratio | 8-94 |
| LongPassing | Ratio | 12-93 |
| BallControl | Ratio | 9-96 |
| Acceleration | Ratio | 15-94 |
| SprintSpeed | Ratio | 16-90 |
| Agility | Ratio | 22-96 |
| Reactions | Ratio | 21-96 |
| Reactions | Ratio | 21 – 96 |
| Balance | Ratio | 16 - 96 |
| ShotPower | Ratio | 2-95 |
| Jumping | Ratio | 15 - 95 |
| Stamina | Ratio | 12-96 |
| Strength | Ratio | 17 - 97 |
| LongShots | Ratio | 3-94 |
| Aggression | Ratio | 11-95 |
| Interceptions | Ratio | 3-92 |
| Positioning | Ratio | 2-95 |
| Vision | Ratio | 10-94 |
| Penalties | Ratio | 5-92 |
| Composure | Ratio | 3-96 |
| Marking | Ratio | 3-94 |
| StandingTackle | Ratio | 2-93 |
| SlidingTackle | Ratio | 3-91 |
| GKDiving | Ratio | 1-90 |
| GKHandling | Ratio | 1-92 |
| GKKicking | Ratio | 1-91 |
| GKPositioning | Ratio | 1-90 |
| GKReflexes | Ratio | 1-94 |
| Release Clause | Ratio | €1.1M - €196.4M |

Section 2: Prospective Dashboard Users

1. List the prospective users that you will develop the dashboard for, and what do you think they can use this dashboard for.

People participating in Fantasy League Football, where they choose teams and players. Analyzing player performances can lead to informed decisions about which team to choose, what formation they should play in, which player is most likely to play well over the course of the season et cetera.

FIFA 19 players, specifically FUT (FIFA Ultimate Team) which comprises of building a team from various players, taking into account which team they play for, their nationality, their chemistry with each other et cetera.

Bookies, or people that facilitate betting on football matches and their results. Bookies can set various odds for each of the teams and individual players based on this dataset’s analysis.

Football coaches can use this dataset for training metrics. They can review the tactics they used in each season, pertaining to the stats produced, and make changes accordingly for the next season. Some of which might be team formations (4-3-3, 4-2-3-1, 4-4-2, 5-3-2), deciding to play either offensively or defensively, which match they need to win, which match they can afford to rest players et cetera.

The players themselves to measure and observe their metrics. What their strong suits over the season were, where they were lacking and need to improve.

Talent scouts looking for new players to join their team can refer the release clauses of each of the players to see who will be available in the market next season, which player can potentially be bought from other teams, and which can be signed after their release clause ends, as a free agent.

Football analysts observing football trends over a large period of time, which can be used in future football marketing, advertisements, and team promotional events.

This dashboard can be used by both someone who has been watching football since a long time and also by someone who has recently started watching football. It can help them understand the player his strengths and his weaknesses.

Section 3: List of User Requirements & Potential Questions

1. List a comprehensive set of questions that the users might ask about this dataset.

User Requirements:

* Ability to easily compare various statistics.
* Ability to be able to see various metrics in understandable visualizations.
* Able to control how many datapoints are visible for any given visualization.
* Ability to compare player head to heads.
* Ability to see correlations between different attributes.

Questions

1. Is a correlation present between Age of the player and their overall rating?

2. Which club/country has the best average overall rating for players?

3. Is a correlation present between a player’s wage and overall rating?

4. Which club/country players have the highest/lowest wages?

5. Does the player’s work rate influence his Stamina?

6. Is a player’s strength influenced by his body type?

7. Which players have highest/lowest release clauses?

8. Which players are paid the most/least?

9. Do defenders have better Marking/Standing Tackle abilities?

10. Which player is the best rated for each of the positions (RW, CF, ST et cetera)?

11. When did a player join a particular team?

12. When did a player start playing professional football?

13. Who is the tallest player?

14. Who is the shortest player?

15. Who is the lightest player?

16. Who is the heaviest player?

17. Which player plays best at which position?

18. For a particular position, who is the best player?

19. Who is the oldest player?

20. Who is the youngest player?

21. Which nationality has the highest professional football players?

22. Which player has the best overall rating?

23. Who is the most valuable player?

24. Who is the least valuable player?

25. Who has the highest wage?

26. Who has the lowest wage?

27. For a selected player, which foot is his preferred foot to shoot shots?

28. For a selected player, what are his ratings in different positions on the field?

29. Which country has the highest number of players?

30. What's the age distribution of players?

31. What’s the weight distribution over the age of the players?

32. What’s the wages distribution of players?

33. Number of players in different playing position?

34. Work rate of the players ?

35. Player height distribution

36. Highest number of players on the left or right foot?

37. Number of players in different clubs?

38. Is there a correlation between sliding tackle and standing tackle for a player?

39. Is there a correlation between GKHandling and GKReflexes for a player?

40. Is there a correlation between release clause and overall rating?

41. Goal keeping stats of outfield players vs goal keeping stats of GKs?

42. Which is the most/least played position?

43. Is there a correlation between weight and strength?

44. Which club has the highest wages?

45. How many players in each category (weak foot - 1 to 5, skill moves - 1 to 5)

46. Is there correlation between crossing and vision?