IFT 598: Data Visualization & Reporting for IT

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

**Project - Phase 3: Dashboard Implementation**

Dr. Asmaa Elbadrawy

December 1, 2021

Section 1: The Dashboard

Graphical user interface, application

Description automatically generated

The dashboard visualizes answers to 10 questions relevant to the FIFA 19 database. Scatter plots, bar graphs, a map, and a tree map were used. Each plot is shown with the relevant filter and scale. Data-ink ratio has been minimized accordingly.

**Link:** https://public.tableau.com/app/profile/neeraj1467/viz/FIFA19Project1/Fifa19Visualizations?publish=yes

Section 2: The Dataset

**FIFA 19 complete player dataset**

**Source:**

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

**Description:**

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.

**Attributes:**

|  |  |  |
| --- | --- | --- |
| 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 |

**Pre-processing:**

The data pre-processing done is as follows:

* The ‘Loaned From’ column had mostly null values. The entire column was dropped from the dataset.
* The columns ‘Value’, ‘Wage’, ‘Release Clause’ were initially in the format ‘€110.5M’, ‘€565K’ and were recognized as strings. These were stripped so that only the numbers remained and were then recognized as such.
* Similarly, the ‘Height’ column was in the format, ‘5’7’, ‘5,8’ and was recognized as strings. The apostrophe ‘’’ was replaced by a period ‘.’ for the numbers to be read as float values.
* Similarly, the ‘Weight’ column was in the format, ‘158lbs’, ‘172lbs’ and was recognized as strings. These were stripped so that only the numbers remained and were then recognized as such.
* The columns ‘LS’, ‘ST’, ‘RS’, ‘LW’, ‘LF’, ‘CF’, ‘LF’, ‘RW’, ‘LAM’, ‘CAM’, ‘RAM’, ‘LM’, ‘LCM’, ‘CM’, ‘RCM’, ‘RM’, ‘LWB’, ‘LDM’, ‘CDM’, ‘RDM’, ‘RWB’, ‘LB’, ‘LCB’, ‘CB’, ‘RCB’, ‘RB’ all were in the format ‘89+2’, ‘93+2’ to show potential growth of the player’s stats. For the purpose of this database, these added numbers were stripped so that only the base numbers (89, 93) remained.

Section 3: Dashboard Users

The prospective users for this dashboard:

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 4: Questions

The list of questions this dashboard will answer are,

1. Does a correlation exist between a player’s age and reaction time?
2. What is each player’s rating in each of the physical statistics (shooting, agility, composure, et cetera)?
3. For every player, is there a relation between their strength, shot power and finishing, given their age?
4. Is a player able to score more freekicks based on his accuracy and the curve he can apply on the ball?
5. Which players are paid the most/least?
6. What is the age distribution of the players?
7. What is the player distribution over the world?
8. Which are the most/least played positions?
9. Which club has the best overall rating, based on the players?
10. Do players from certain nations have better ratings than others?

Section 5: Plots

Chart, scatter chart

Description automatically generated

It answers the question ‘Does a correlation exist between a player’s age and reaction time?’

It measures reactions versus age for each player and shows a trend line.

Table

Description automatically generated

It answers the question ‘What is each player’s rating in each of the physical statistics (shooting, agility, composure, et cetera)?’

It allows to select each attribute and each player, and shows those players’ selected attributes in a stacked bar chart.

Chart, scatter chart

Description automatically generated

It answers the question ‘For every player, is there a relation between their strength, shot power and finishing, given their age?’

It measures shot power vs strength of each player, differentiating also by age.

Chart, scatter chart

Description automatically generated

It answers the question ‘Is a player able to score more freekicks based on his accuracy and the curve he can apply on the ball?’

It measures freekick accuracy versus the curve on the ball for each player, and how finishing is affected by these stats.

Chart, bar chart

Description automatically generated

It answers the question ‘Which players are paid the most/least?’

It shows wages for each player in the form of bar graphs.

Chart, histogram

Description automatically generated

It answers the question ‘What is the age distribution of the players?’

It displays count of all the players at each division. A bell curve is observed.

Map

Description automatically generated

It answers the question ‘What is the player distribution over the world?’

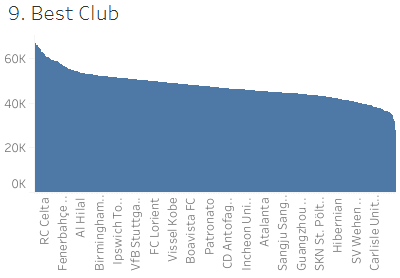
It displays player counts over the entire world.

Chart, bar chart

Description automatically generated

It answers the question ‘Which are the most/least played positions?’

It shows the count of players per position.



It answers the question ‘Which club has the best overall rating, based on the players?’

It shows a histogram with each club according to the sum of the overall rating of its players.

Chart

Description automatically generated

It answers the question ‘Do players from certain nations have better ratings than others?’

It shows the average rating over each country and arranges them according to the best overall rating. It also shows counts per country, given that some countries have more players than others.

Section 6: Interactivity

The interactive features are:

**Measure Values Filter:**

Used to select which of the player’s attributes to be visualized.  
  
**Connected Plots:**

The plot for ‘What is each player’s rating in each of the physical statistics (shooting, agility, composure, et cetera)’.

Value range: 0 – 100.

Name Filter:

Used to select the player for whom attributes will be visualized.

**Connected Plots:**

The plot for ‘What is each player’s rating in each of the physical statistics (shooting, agility, composure, et cetera)’.

Also the plot for ‘Player Wages’

Values: List of all the names of the players. (Single select list). Loaded from Names.

Top N Filter:

Used to select Top N players to be displayed in the visualization.

**Connected Plots:**The plot for ‘Top N Best Player for each position on the field’

Value range : 1 - 10,000

Wage Reference Line:

Used to select wages for players above and below the threshold and shows distinguishing colors.

**Connected Plots:**The plot for ‘Player Wages’

Value range : 1 - 565