

Google_Data_Analytics_Capstone

Justin Schubeck

2022-12-22

Introduction

This is a Capstone Project for the Google Data Analytics Professional Certification.

FIFA is a series of association football video games developed and released by Electronic Arts (EA). It will be concluded that the meta in-game attacking tactics will be the following:

- FIFA 2013-2015: Heading
- FIFA 2016-2018: Pace
- FIFA 2019-2021: Long Shots
- FIFA 2022-2023: Tiki Taka

I will be using the 6 phases of the analysis process (Ask, Prepare, Process, Analyse, Share and Act) to help guide my analysis of the dataset.

Phase 1: Ask

- What topic are you exploring?
 - The topic involves a user choosing which player (Cristiano Ronaldo or Lionel Messi) would be best for a specific task in different EA Sports FIFA video game versions.
- What is the problem you are trying to solve?
 - The problem that is trying to be solved would be the decision of which FIFA Ultimate Team player to buy given a certain attacker's play type.
- What metrics will you use to measure your data to achieve your objective?
 - EA Sports FIFA game data will be used to create and compare metrics.
- Who are the stakeholders?
 - The stakeholders are EA Sports FIFA video game players who would be looking to choose which player to buy with in game currency.
- Who is your audience?
 - The audience would be anyone who plays FIFA.
- How can your insights help your client make decisions?
 - The insights from the player data over the years can help the gamer choose which player might be best for specific tasks in football.

Phase 2: Prepare

- Where is your data located?
 - FUTHead (<https://www.futhead.com/>)
- How is the data organized?
 - The data is organized by year and by player.
- Are there issues with bias or credibility in this data?

- The credibility of EA has been questioned by the public in the past to favor Lionel Messi. However, since they also create the gameplay, this should not alter the final decision. * Does your data ROCCC?
- Reliability: The data is reliable because it has come from the same owners, EA Sports, for the whole of the video game.
- Originality: This data set was manually extracted from the FUTHead website, but it can also be found in-game.
- Comprehensiveness: The data is only fully comprehensive for 2013-2023, not 2010-2012.
- Current: The data is up to date.
- Cited: <https://www.futhead.com/> (<https://www.futhead.com/>)
- How are you addressing licensing, privacy, security, and accessibility?
 - This is public data.
- How did you verify the data's integrity?
 - FUTHead is the best source of FIFA game data. It is directly extracted from the game itself.
- How does it help you answer your question?
 - The dataset helps answer the question because it comes directly from the game itself.
- Are there any problems with the data?
 - The data is missing detailed breakdowns for data before 2013, and is missing another single stat before 2017.

Phase 3: Process

- What tools are you choosing and why?
 - I am choosing to use R Cloud and Tableau because the dataset is relatively small. The BigQuery would not be needed for this task.
- Have you ensured your data's integrity?
 - The integrity of the data has been checked for consistency.
- What steps have you taken to ensure that your data is clean?
 - The years 2010-2012 have been removed from the dataset, as they contain very little explainability. Also, the composure statistic of 2017-2023 will not be used since the data was not tracked for the previous years.
- How can you verify that your data is clean and ready to analyze?
 - Ensuring there are no values out of bounds and checking that all years have the same stats will ensure the data is ready to be analyzed.
- Have you documented your cleaning process so you can review and share those results?
 - The process is documented in this file.

I will start by installing and loading the required R packages for my analysis.

```
install.packages('dplyr')
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.2'  
## (as 'lib' is unspecified)
```

```
library('dplyr')
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##   filter, lag
```

```
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

Next, I will import the necessary files onto R.

```
data <- read.csv('data.csv')
head(data)
```

```
##   year statistic_category statistic_name value  player
## 1 2010           General      Overall    93 Ronaldo
## 2 2010           General         Pace    92 Ronaldo
## 3 2010           General    Shooting    89 Ronaldo
## 4 2010           General     Passing    79 Ronaldo
## 5 2010           General   Dribbling    93 Ronaldo
## 6 2010           General   Defending    64 Ronaldo
```

Next, I will clean the datasets. As stated before, the years 2010 to 2012 will be removed, and the Composure statistic will be removed. The new cleaned dataset will be exported to a CSV for Tableau visualization.

```
data_clean <- filter(data, year>2012) %>%
  filter(statistic_name != 'Composure')
write.csv(data_clean, "data_clean.csv", row.names=FALSE)
```

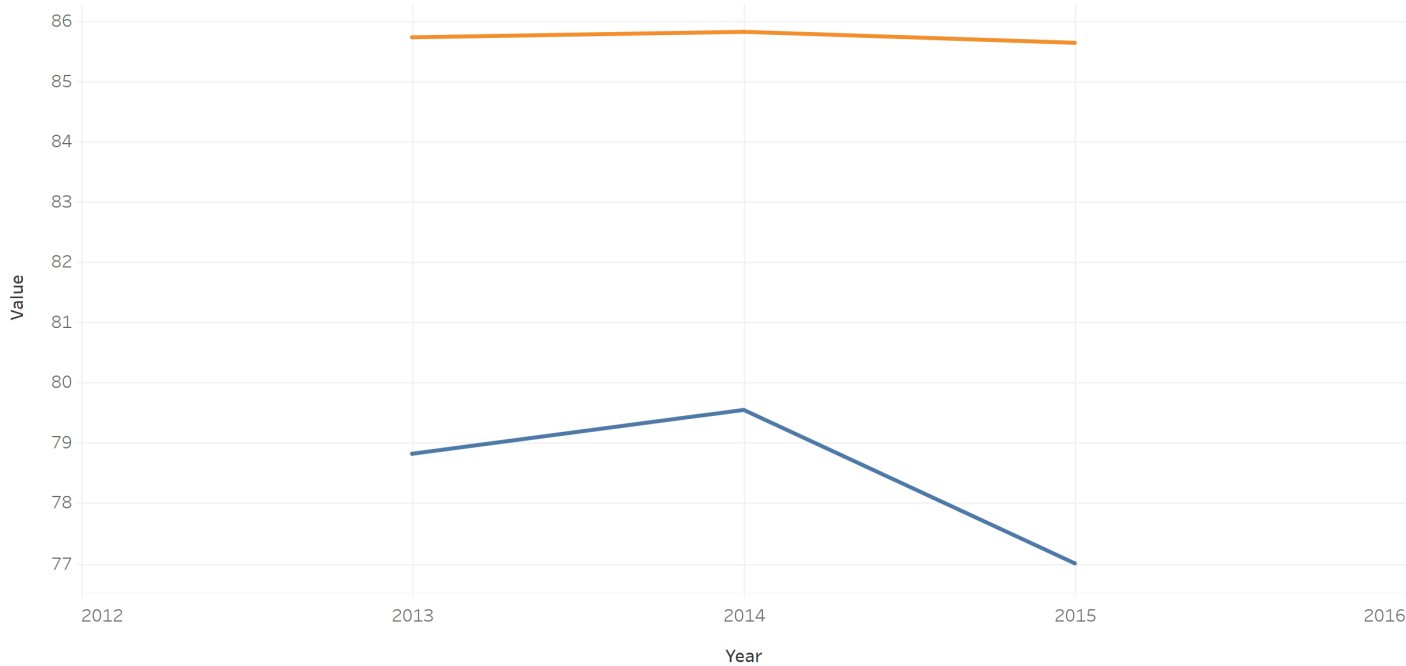
Phase 4: Analyze

- How should you organize your data to perform analysis on it?
 - When collecting the data, I formed a long dataset that contained descriptor columns about the value for each stat in order to filter the data properly when analyzing in Tableau.
- Has your data been properly formatted?
 - Yes, the data was manually collected and formatted as needed.
- What surprises did you discover in the data?
 - I was surprised at how the overall rating of Lionel Messi was higher than Cristiano Ronaldo many years when the individual stats did not seem to warrant the higher overall stat.
- What trends or relationships did you find in the data?
 - I found a relationship between the height and weight of each player and how that reflects on their play styles.
- How will these insights help answer your business questions?
 - These insights alone will not help answer the questions, but they will help explain some of the conclusions from the data.

FIFA 2013-2015: Heading

Heading consisted of 11 individual statistics: Heading, Reactions, Physical, Crossing, Curve, Aggression, Jumping, Strength, Finishing, Positioning, and Shot Power. Ronaldo is distinctly better in 6 of these stats over all three years, and Messi 3 respectively. Over all these stats, Ronaldo has a clear gap for the Heading meta of about 7 points on average, as can be seen

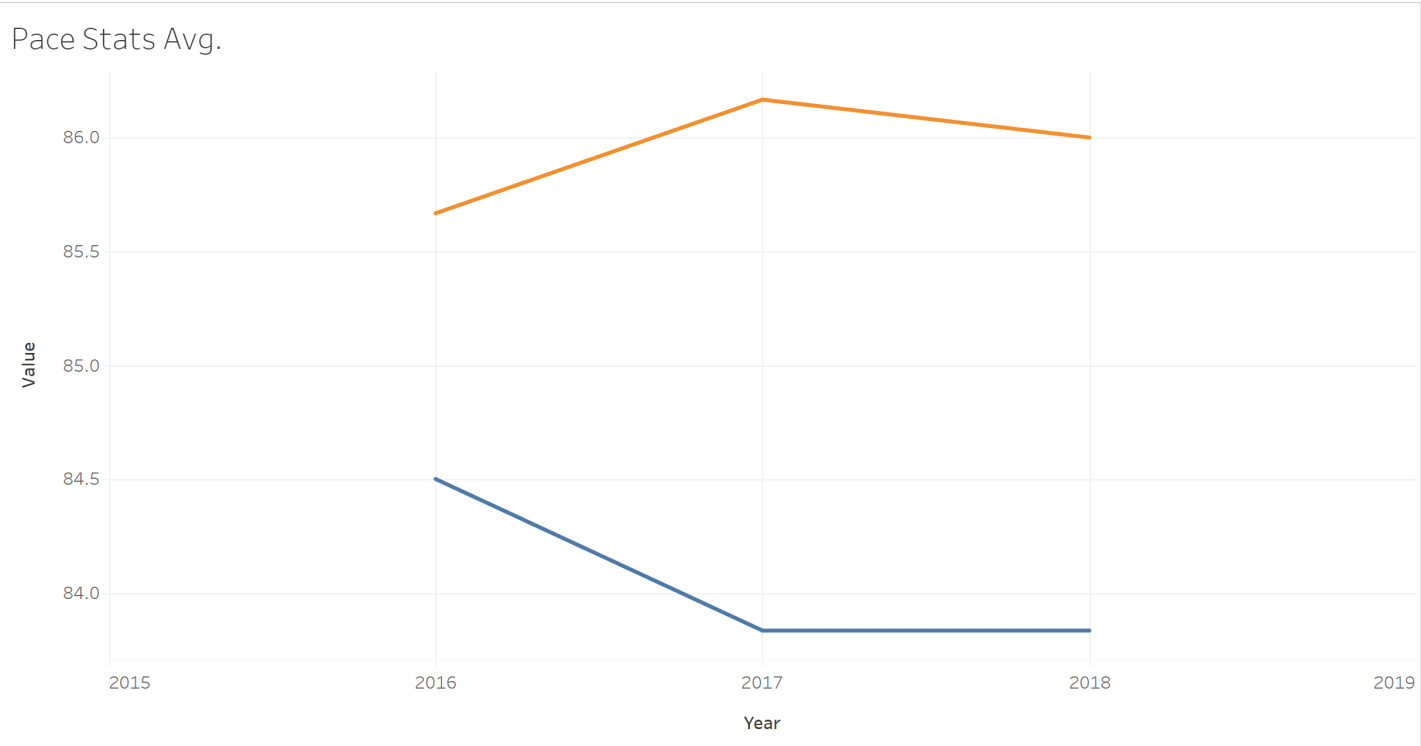
Heading Stats Avg.



FIFA 2016-2018: Pace

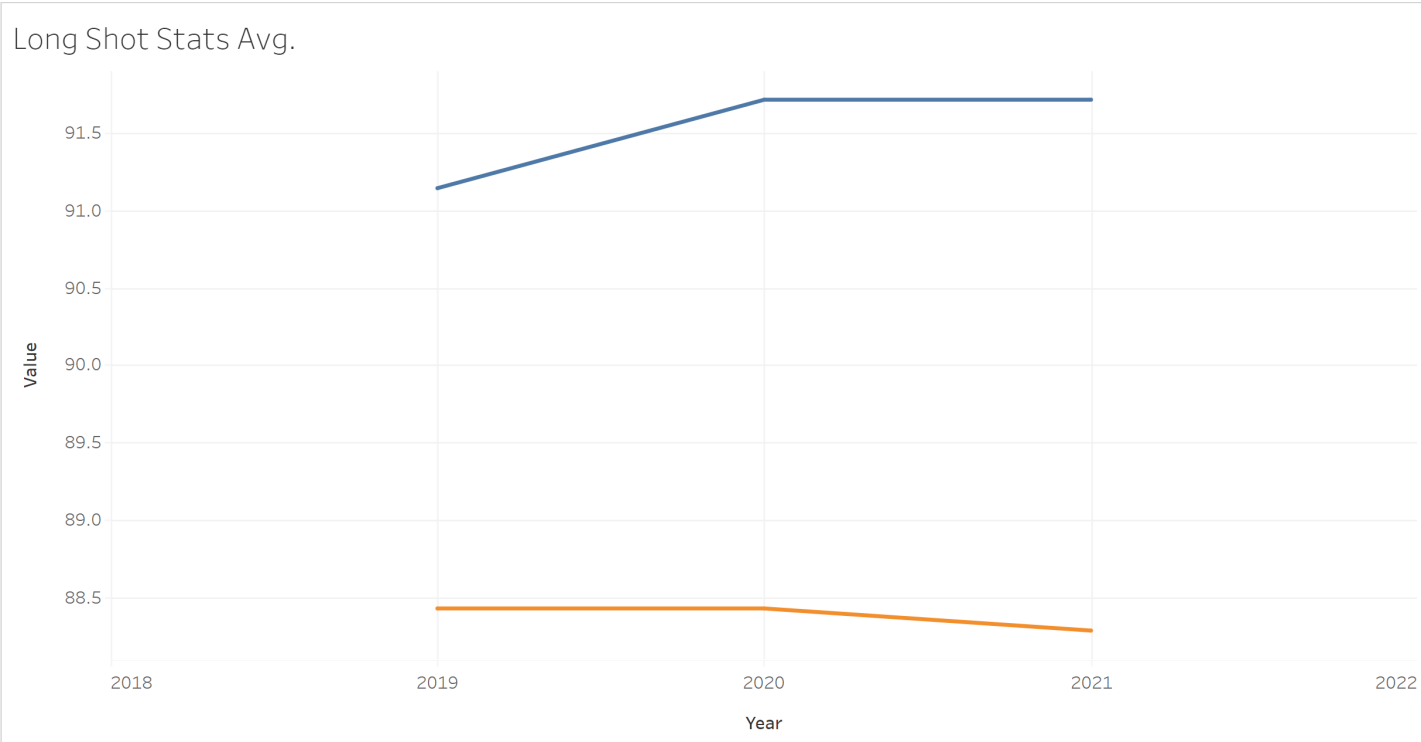
Heading consisted of 12 individual statistics: Agility, Balance, Ball Control, Pace, Shooting, Acceleration, Sprint Speed, Aggression, Stamina, Strength, Finishing, and Positioning. Ronaldo is distinctly better in 7 of these stats over all three years, and Messi 4 respectively. Over all these stats, Ronaldo has a small gap for the Pace meta of

about 2 points on average, as can be seen



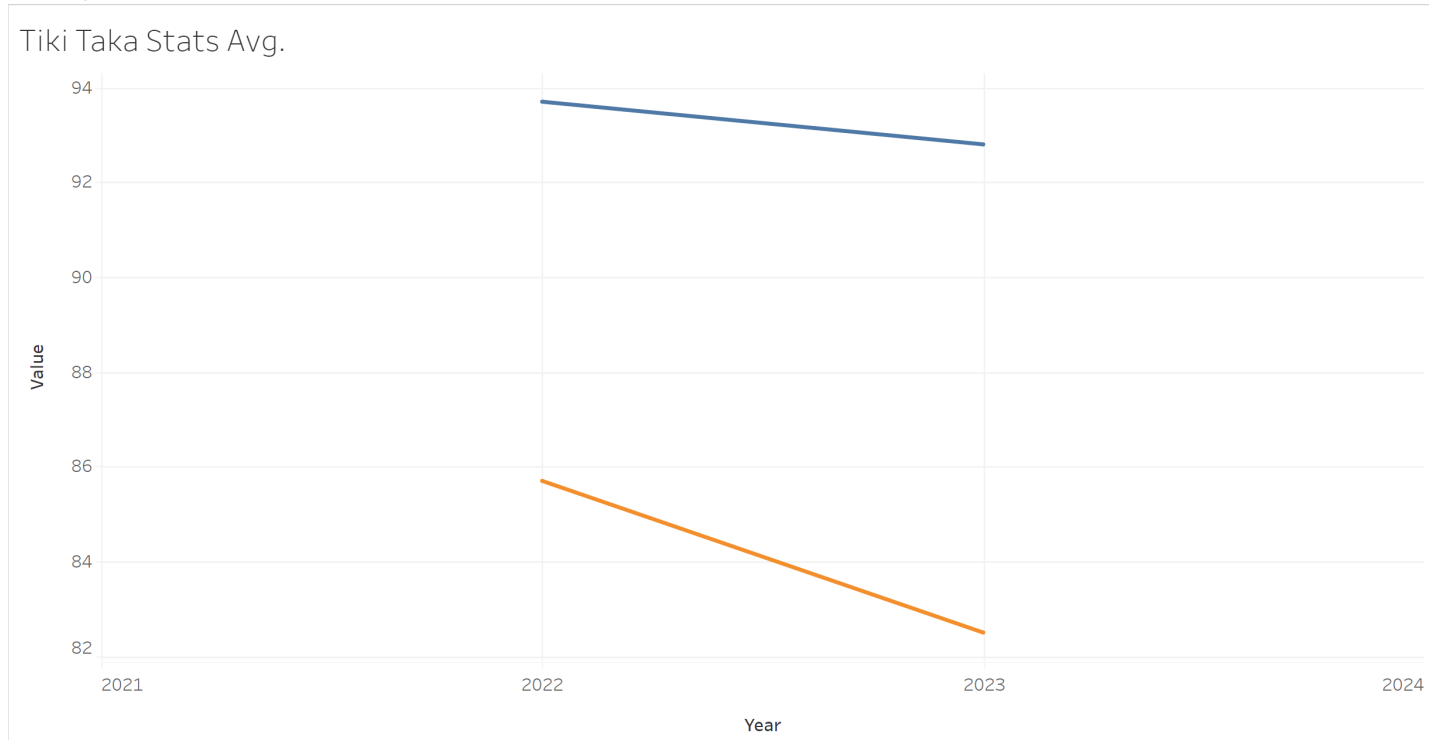
FIFA 2019-2021: Long Shots

Heading consisted of 7 individual statistics: Shooting, Curve, Free Kick, Finishing, Long Shots, Shot Power, and Volleys. Ronaldo is distinctly better in 2 of these stats over all three years, and Messi 4 respectively. Over all these stats, Messi has a small gap for the Long Shots meta of about 3 points on average, as can be seen



FIFA 2022-2023: Tiki Taka

Heading consisted of 10 individual statistics: Agility, Balance, Ball Control, Dribbling, Reactions, Dribbling, Passing, Short Passing, Vision, and Positioning. Ronaldo is distinctly better in 2 of these stats over the two years, and Messi 8 respectively. Over all these stats, Messi has a clear gap for the Tiki Taka meta of about 9 points on average, as can be seen



Phase 5: Share

- Were you able to answer the business question?
 - The business question was answered, as a predicted better FIFA Ultimate Team card for each year's meta attacking tactic was chosen based on data.
- What story does your data tell?
 - The story tells that a larger player like Ronaldo that is 6'2" and about 187lbs does better than a 5'7" 148lbs Messi the physical attacking tactics of the game like heading and sprinting. A shorter player with more agility and closer to the ground can weave around players to do great short passing and snap a quick shot from distance.
- How do your findings relate to your original question?
 - The findings relate to the original question in theory, but it would be up to how the video game is coded to see if the results are replicated.
- Who is your audience? What is the best way to communicate with them?
 - The audience is anyone who plays the EA Sports FIFA video games. The best way to communicate with them would be on the internet, where the FIFA community has a strong presence.
- Can data visualization help you share your findings?
 - Yes, the individual stat comparisons between the players can be found as well.
- Is your presentation accessible to your audience?
 - Tableau (https://public.tableau.com/views/Google-Data-Analytics/Data?:language=en-US&:display_count=n&:origin=viz_share_link)

Phase 6: Act

- What is your final conclusion based on your analysis?
 - The final conclusion is that for an attacking player who fits the overpowered metas of the game, Ronaldo would be better for Heading and Pace, while Messi would be better for Long Shots and Tiki Taka.
- How could your team apply your insights?
 - The team could apply the insights by testing the players in each version of FIFA to see if the results are seen by gamers in the community to agree with the data.
- What next steps would you or your stakeholders take based on your findings?
 - The next steps would be to try the players out in the respective play styles in the game to see community feedback and experiences.