National College of Ireland

Project Submission Sheet – 2019/2020

Student Name:	Antonio Caruso
Student ID:	19203608
Programme:	Higher Diploma in Science in Data Analytics **Year:** 2020
Module:	Data Visualization
Lecturer:	Enda Stafford
Submission Due Date:	16/08/2020
Project Title:	An exploratory analysis on Football players' technical features in the videogame FIFA19
Word Count:	2141

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the references section. Students are encouraged to use the Harvard Referencing Standard supplied by the Library. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action. Students may be required to undergo a viva (oral examination) if there is suspicion about the validity of their submitted work.

Signature: ……

Date: 14/08/2020

PLEASE READ THE FOLLOWING INSTRUCTIONS:

1. Please attach a completed copy of this sheet to each project (including multiple copies).
2. Projects should be submitted to your Programme Coordinator.
3. **You must ensure that you retain a HARD COPY of ALL projects**, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer. Please do not bind projects or place in covers unless specifically requested.
4. You must ensure that all projects are submitted to your Programme Coordinator on or before the required submission date. **Late submissions will incur penalties.**
5. All projects must be submitted and passed in order to successfully complete the year. **Any project/assignment not submitted will be marked as a fail.**

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Table of Contents

Table of Figures

1 Project Background

This report presents a series of 7 interactive visualisations created and published through a story with the software Tableau. The topic covered revolves around the world of football and videogames, as the information and insights obtained and demonstrated refer to FIFA 19, the football simulation videogame developed by EA (Electronic Arts Inc., 2020). The choice of this topic lies in an interest of the author of this paper for football and sport videogames in general.

2 Project Objective

The scope of this report is to present interactive visualitations that convey deep insights on relational information between variables across the used datasets, through an exploratory analysis which aims to enhance the readers' engagement with the charts to obtain further insights from the data provided (Kirk, 2019). As aforementioned in the background, the exploratory analysis was conducted on a dataset called "FIFA 19 player database" (Sharma, 2018). The key objective was to provide interactive visualisations on the statistical values about football players across clubs and the main European football leagues available, with a focus on identifying the overall value plus main technical features for each football position.

3 Data pre-processing and analysis

The original FIFA 19 dataset collected consisted of 70 variables and 15,397 observations (Sharma, 2018). The original dataset was subsequently subset into 4 separate datasets containing information on the 6 main European football leagues (Bundesliga, Premier League, La Liga, Serie A, Ligue 1 and Superliga). Each dataset contains information on technical features of football players by their registered professional roles of Goalkeepers, Defenders, Midfielders and Strikers. The four datasets can be considered structured because presented in the form of a excel tables

with defined variables and records (Ward, et al., 2015). Cleansing and filtering processes were applied before commencing the analysis, resulting in the 5 datasets as below;

1st Dataset: FIFA19, presents 8 Variables and 3109 observations

Variable Name	**Data Type**	**Metadata Description**	**Further information**
Player name	Categorical, String	Name of the football player	High Cardinality
Club	Categorical, String	Name of the club	Low cardinality
League	Categorical, String	Name of the League (Bundesliga, Premier League, Ligue 1, La Liga, Serie A TIM, Primeira Liga)	High Cardinality
Position	Categorical, String	The acronym of the position on the pitch as: GK for Goalkeeper, CB for Central Back, RB for Right Back, LB for Left Back, CM for Central Midfielder, CAM for Central Advanced Midfielder, CF for central forward, STR for striker	High Cardinality
Defending	Numerical, discrete, range of values from 0 to 92	Defending value	Low cardinality
Passing	Numerical, discrete, range of values from 0 to 93	Ability to pass the ball	Low Cardinality
Finishing	Numerical, discrete, range of values from 4 to 96	Ability to score a goal	Low Cardinality
Overall Value	Numerical, discrete, range of values from 51 to 95	Player General Value	Low Cardinality

Figure 1: FIFA 2019 Dataset. Source: Sharma (2018)

2nd Goalkeepers dataset: the dataset contains 8 variables and 346 observations

Variable Name	**Data Type**	**Metadata Description**	**Further information**
Goalkeeper name	Categorical, String	Name of the Goalkeeper	High Cardinality
Club	Categorical, String	Name of the club	Low cardinality
League	Categorical, String	Name of the League	High Cardinality
Reflexes	Numerical, discrete, range of values from 55 to 94	GK reaction to catch/save the ball	Low cardinality
Handling	Numerical, discrete, range of values from 49 to 93	How cleanly the GK catches the ball	Low cardinality
Kicking	Numerical, discrete, range of values from 45 to 92	The Goalkeepers' ability to kick the ball	Low Cardinality
Positioning	Numerical, discrete, range of values from 45 to 91	Positioning value in the goal	Low cardinality
Overall	Numerical, discrete, range of values from 45 to 91	Overall Goalkeepers' value	High Cardinality

Figure 2: Goalkeepers Dataset. Source: Sharma (2018)

3rd Defenders dataset: It contains originally 6 variables and 1008 observations

Variable Name	**Data Type**	**Metadata Description**	**Further Information**
Defenders name	Categorical, String	Name of the Defender	High Cardinality
League	Categorical, String	Name of the League	High Cardinality

Position	Categorical, String	Acronym of the exact position in the defense line	High Cardinality
Acceleration	Numerical, discrete, range of values from 30 to 95	Acceleration value	Low Cardinality
Agility	Numerical, discrete, range of values from 28 to 91	Agility Value	Low Cardinality
Interception	Numerical, discrete, range of values from 48 to 93	Interception Value	Low Cardinality
Marking	Numerical, discrete, range of values from 40 to 95	Marking value	Low Cardinality

Figure 3: Defenders Dataset. Source: Sharma (2018)

4th dataset Midfielders: It presents 4 variables and 1132 observations

Variable Name	**Data Type**	**Metadata Description**	**Further information**
League	Categorical, String	Name of the League	High Cardinality
Nationality	Categorical, String	Nationality of the footballers	High Cardinality
Vision	Numerical, discrete, range of values from 29 to 95	Playing Vision	Low Cardinality
Free Kick	Numerical, discrete, range of values from 24 to 93	The ability to take a free kick	Low Cardinality

Figure 4: Midfielders Dataset. Source: Sharma (2018)

5th dataset Strikers: it presents 10 variables and 280 observations;

Variable Name	Data Type	Metadata Description	Further information
Strikers name	Categorical, String	Name of the Strikers	Low Cardinality
Club	Categorical, String	Name of the club	Low Cardinality
League	Categorical, String	Name of the League	High Cardinality
Nationality	Categorical, String	Nationality of the footballers	High Cardinality
Free Kick	Numerical, discrete, range of values from 22 to 95	The ability to take a free kick	Low Cardinality
Long Shot	Numerical, discrete, range of values from 41 to 95	Ability to perform long shots	Low Cardinality
Volleys	Numerical, discrete, range of values from 36 to 91	Ability to perform volleys	Low Cardinality
Penalties	Numerical, discrete, range of values from 39 to 92	Ability to perform penalties	Low Cardinality
Finishing	Numerical, discrete, range of values from 50 to 96	Ability to score a goal	Low Cardinality
Overall	Numerical, discrete, range of values from 56 to 95	Overall value	Low Cardinality

Figure 5: Strikers Dataset t. Source: (Sharma, 2018)

4 Data Visualisation Report

4.1 First Interactive Visualisation: Dendrogram

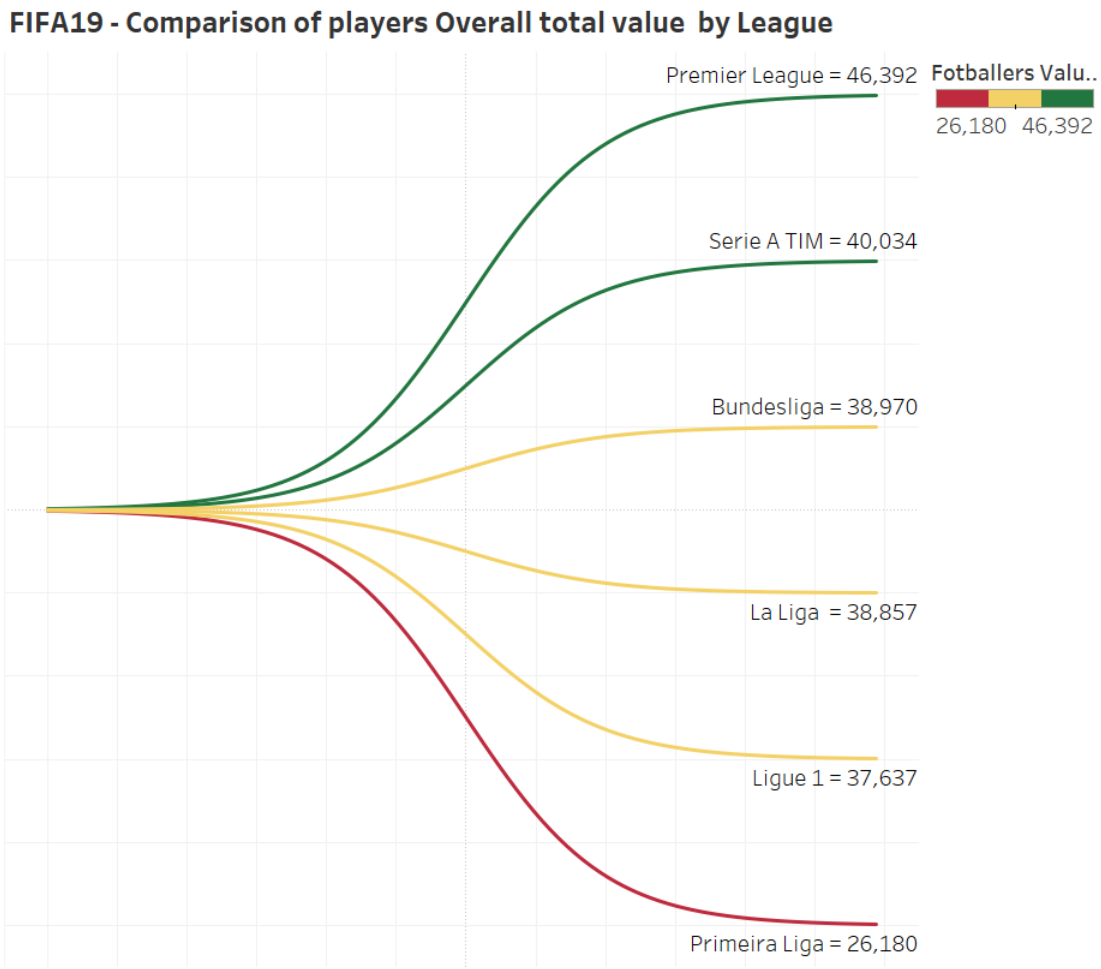

Figure 6: Dendrogram chart "Comparison of Players overall total value by league", own chart. Source: Sharma (2018)

4.1.1 Information and Results

The first visualisation communicates information on summed players overall values across the four main football leagues. It clearly shows how the strongest players on FIFA 2019 play in the English Premier League with a total value for any player of 46,392, followed by the Italian Serie A TIM, The German Bundesliga, the Spanish La Liga and the French Ligue 1. The weakest players play in the Portuguese Primeira Liga, displaying a total value of 26,180, equalling to 28.21% less than the leading Premier League.

4.1.2 Justification of techniques, layout, style, and colours

A Dendrogram Chart was chosen as it shows a composition of hirerarchical data from total values into leagues that can be furtherly decomposed by filtering by Clubs and Players (Kirk, 2019). The hierarchical relationship branches off across multiple dimensions from the root note that identifies a relationships between players values and the leagues they play in.

The dendrogram permits greats interactivity provided by the filtering option to visualise the branches of interest with new data. Each branch was labelled with the main information and the tooltip at the end of the branch provides further information on some players features. A red-green-gold diverging palette was selected with 3 steps to clearly distinguish the two leading leagues in the videogame depicted in green, the three following Leagues ranking with similar values in yellow, and the last ranking Portuguese league with a vivid red. A Multiple Values search bars of Player Name and Club were applied to allow the audience to interact with the dendogram and filter by players' name or club to obtain additional quantitative information on total players values by club or singular players.

4.2 Second Interactive Visualization: Treemap

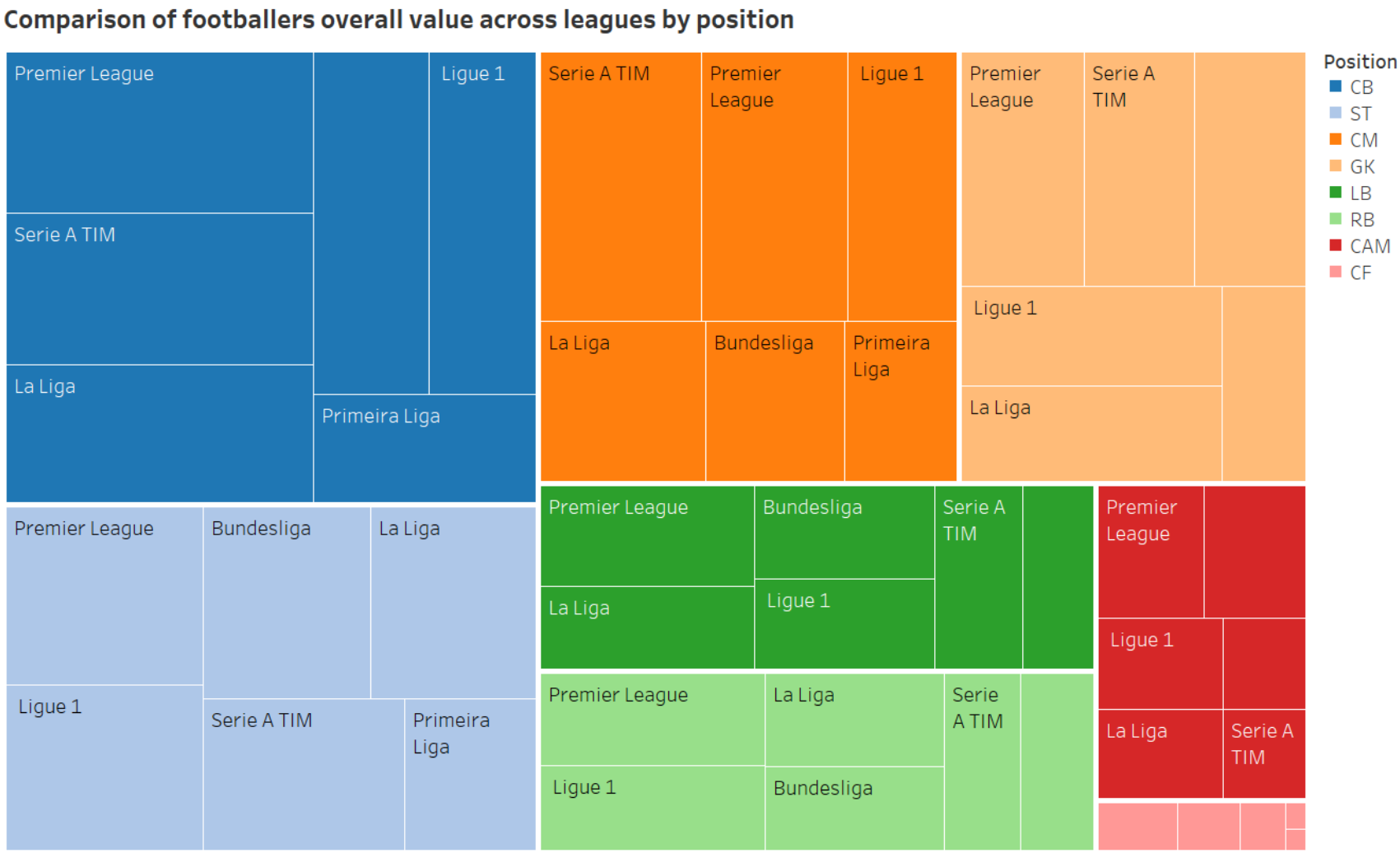

Figure 7: Treemap chart "Comparison of footballers overall value across leagues by position", own chart. Source: Sharma (2018)

4.2.1 Information and Results

The Second visualisation provides detailed information on what previously shown in the dendogram chart by investigating the overall players value across all the leagues by players positions. The treemap shows each rectangular tile representing total players values for each role in each specific league. The coloured legend helps clearly identify the highest valuable positions in each league, in fact It is clearly visibile how the Central Back players have the highest overall value in the football game, with prominence in the Premier League, Serie A TIM and La Liga. To be noted by the size of the rectangulars how CB, ST, CM and GK are the highest valued players in the game, and that the Portuguese Primeira Liga has the lowest players values across all the roles, except for the Central Attacking Midfielders displayed in a dark shade of blue where it ranks ahead of the Italian Serie A. Another relevant poin is the Serie A TIM shows to possess the strongest Central Midfielders, as well as the weak full-backs as it ranks only ahead of Primeira Liga.

4.2.2 Justification of Charts, techniques, layout, style, and colours

The choice of a Treemap (Kirk, 2019) is due to the necessity of categorizing the leagues as rectangular tiles whose value constituted as part of the whole players value for each position. In fact, The visualization shows a composition of the whole value per role into many components represented in a descending order by each league (Extreme Presentation, 2020). High interactivity is provided by the two filters by Club and Nationality that allow additional quantitative information on each value per league arranged in the treemap. The Tableau Classic 20 palette was applied as the dark and light shade of each colours divide the roles into pairs helping read the map and the arrangement of the rectangulars to quickly glance the highest and lowest valued roles.

4.3 Third interactive Visualisation: Packed Bubble Chart

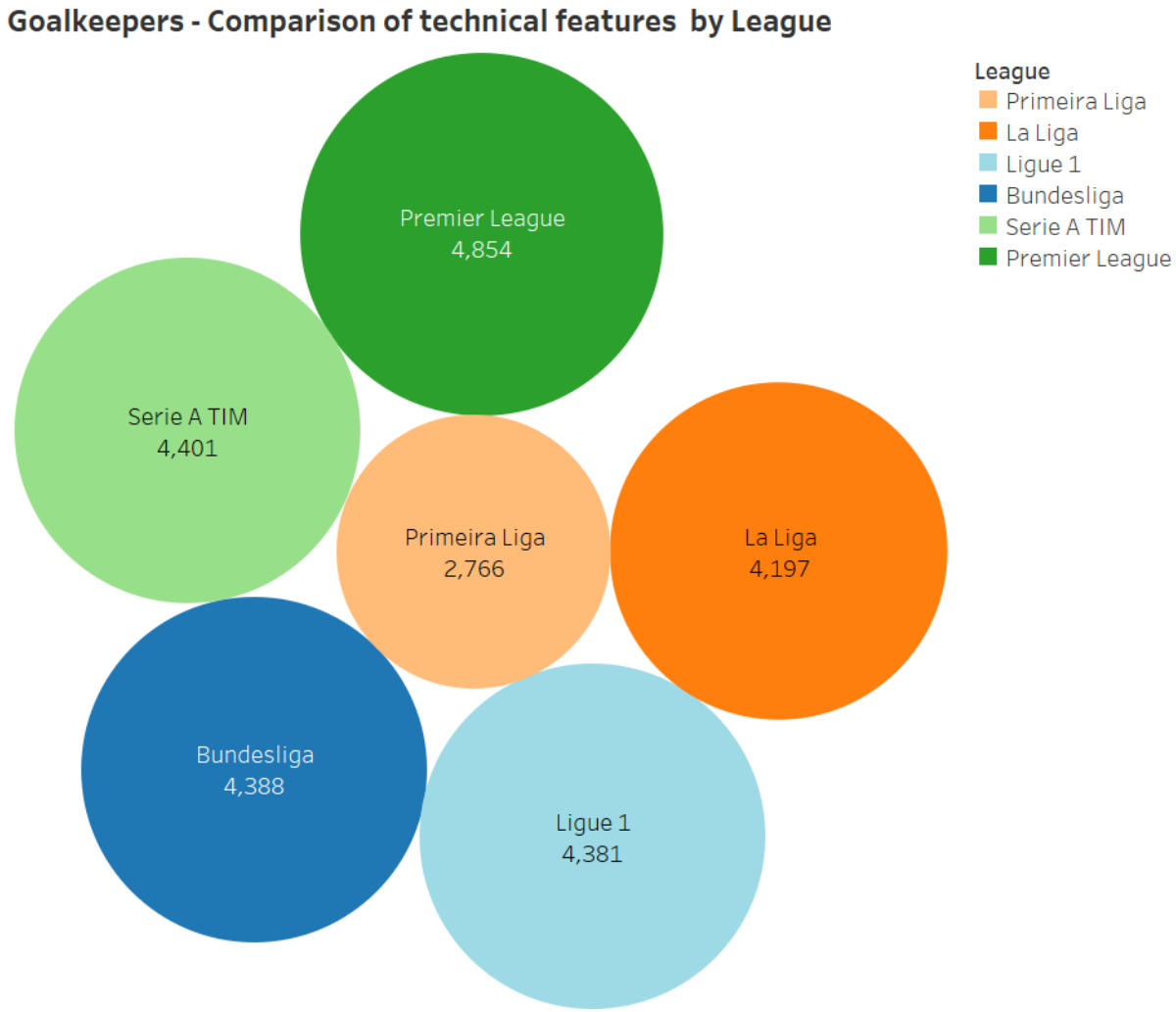

Figure 8: Packed bubble chart "Goalkeepers – Comparison of technical features by league", own chart. Source: Sharma (2018)

4.3.1 Information and Results

The third visualisation communicates information on which league has the strongest goalkeepers by their own overall value, plus some extra information on specific recognisable characteristics. The size of the bubble is given by the overall total values of the goalkeepers playing in a specific league, from which It can be immediately noticed how the strongest goalkeepers play in the Premier League with a total overall value of 4,854, whereas the weakest perform in the Portuguese Primeira Liga with 2,766, almost 57% difference between the two leagues. On the contrary, in the other leagues the goalkeepers perform total overall value in the range between 4,000 and 4,500.

4.3.2 Justification of the Chart, techniques, layout, style, and colours

A Packed Bubble chart was chosen as it provides an instantaneous relationship between two or more variables or (Kirk, 2019). A tableau classic 20 palette was applied with darker and lighter hues of green, blue and orange to give an impact of the leagues in pairs about their values and the difference from each other. The tooltip option provides more information on the goalkeepers' technical features such as handling, kicking, positioning and reflexes. What makes the chart interactive are the filtering multiple values dropdown options, that allows to filter by Goalkeepers' nationalities, names and clubs to determine a change in the values and the size of the bubbles and gain more insights.

4.4 Fourth interactive Visualisation: Side-by-side circle view

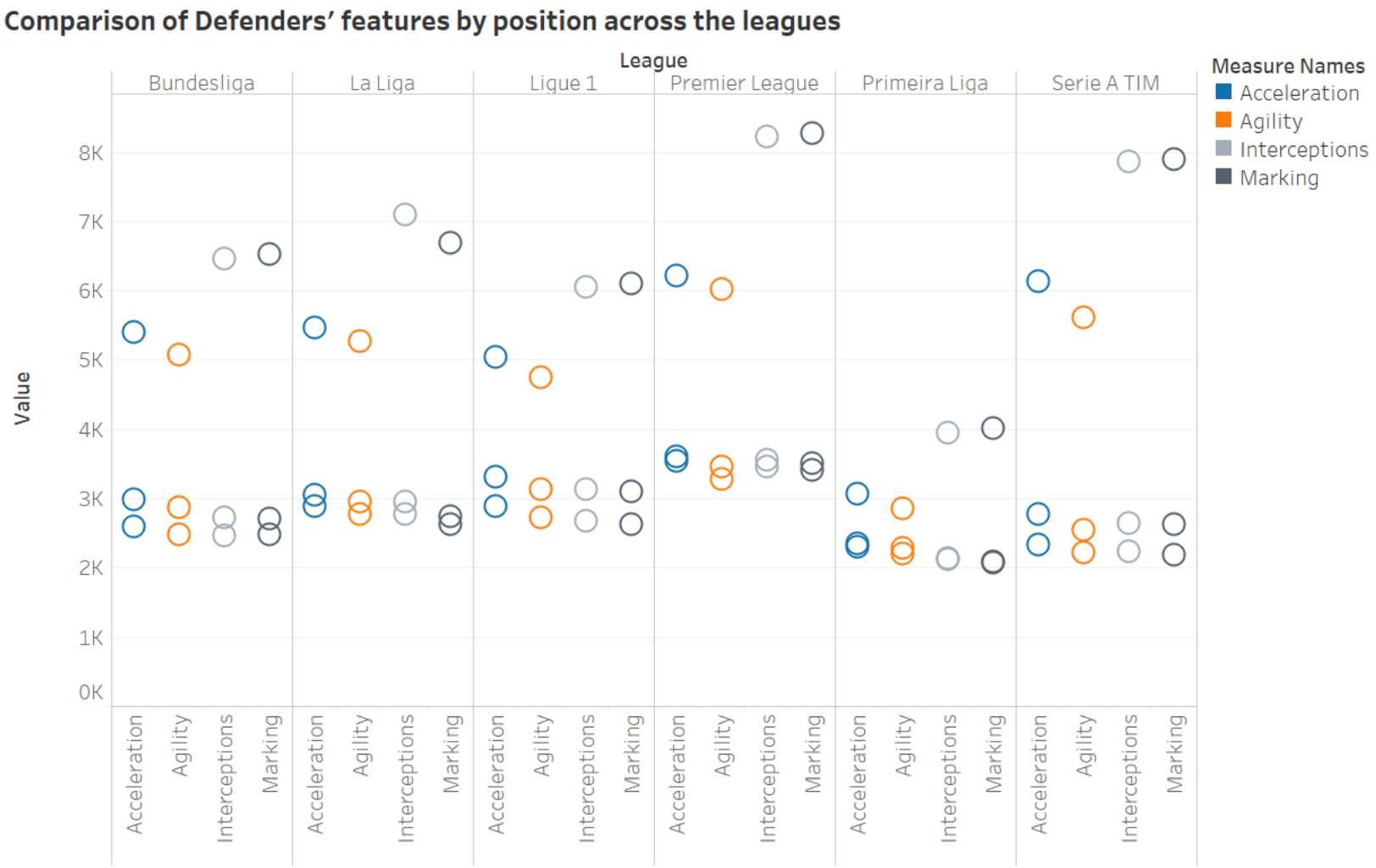

Figure 9: Side-by-side circle view chart "Comparison of Defenders' features by position across the leagues", own creation. Source: Sharma (2018)

4.4.1 Information and Results

The fourth visualisation displays information on four main defenders' technical features of Acceleration, Agility, Interception and Marking by defenders' position on the pitch and across all the leagues. The charts immediately conveys the information that the strongest Central Backs play in Premier League and Serie A TIM, as clearly shown by the highest Interception and Marking compared to the other leagues with around 7-8 thousands total value, and also by Acceleration and Agility in the range of 5,500 and 6,500. For the full-backs the values are consistently lower than central-backs across all the leagues below the treshold of 4,000 , and another interesting fact is that the only leagues where values for full and right backs do not overlap are the Ligue 1 and the Serie A TIM, with the first seeing strongest right backs and the the second better Left-Backs across all the features.

4.4.2 Justification of chart, techniques, layout, style, and colours

The side-by-side circle view chart was chosen as it shows the relationship between four numerical variables by defenders' position, allowing a comparison of defenders' performance across the six leagues (Kirk, 2019). The X axis displays the technical feature with their numerical value appearing on the Y axis where circles were placed. Since the values for Left-backs and Right-backs overlap in chart, A colour blind palette was applied to circle views to help colour-blind people clearly distinguish the values and avoid difficulties (Shaffer, 2016). The interactivity is provided by the multiple values drop-down bars that allow to delve into the data by filtering by Club or Defenders' names.

4.5 Fifth visualization: Interactive bubble plot

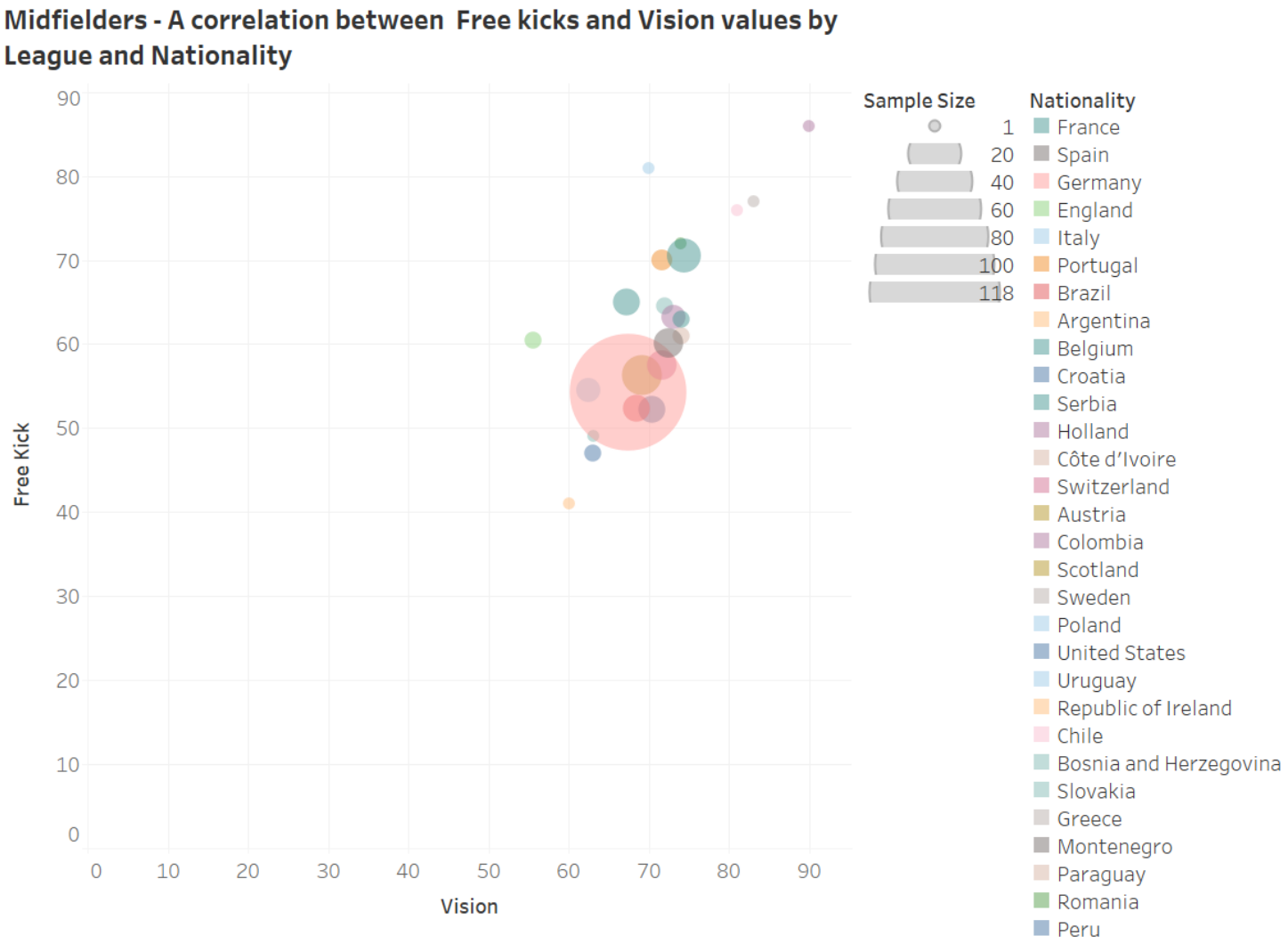

Figure 10: Interactive bubble plot "Midfielders: a correlation between free kick and vision values by league and nationality", own creation. Source: Sharma (2018)

4.5.1 Information and Results

The fifth visualisation is an interactive bubble plot providing information about the most important nationals midfieders possessing the highest free kick and vision values across the leagues. The above example shows the Midfielder stats in Bundesliga. The coloured bubble represent different nationals playing in that league, and the large pink bubble shows a value in the range of 50-70 for German midfielders who are obviously the most numerous nationals to play in Germany. The second strongest represented nation with the highest values in Bundesliga appears to be France, whose midfielders reach a range of 70-74 for Free kick and vision values. Nevertheless, is is visibile the 1 Colombian fotballer only to possess the average highest values between 86-90 for vision and free kick values, followed by a Sweden and Chilean fotballers which stand out particularly for vision skills (83 and 81) and an Italian fotballer ranking at 81 for Free Kick ability. The weakest Nation in the chart is represented by 1 Argentinian in the small yellow bubble footballer who possess an average 41 for free kick and 60 for vision.

4.5.2 Justification of chart, techniques, layout, style, and colours

The interactive bubble plot displays the relationship between the two variables of Average Vision and Average Free kicks for the midfielders on FIFA19 (Kirk, 2019). A third quantitative variable shown in the sample size represented the number of fotballers by nationality which determined the size of the bubbles. The layout shows the average Vision value on the X axis and the average Free Kick values on the Y axis. An automatic colour palette was applied because of the higher number of nationalities than available pallettes to distringuish each with a different colour. Interactity is provided by a slide bar that allows to filter the midfielders by League, to understand which nationals have the highest values of vision and free kick by country in each professional league.

4.6 Sixt visualisation: Sunburst Chart

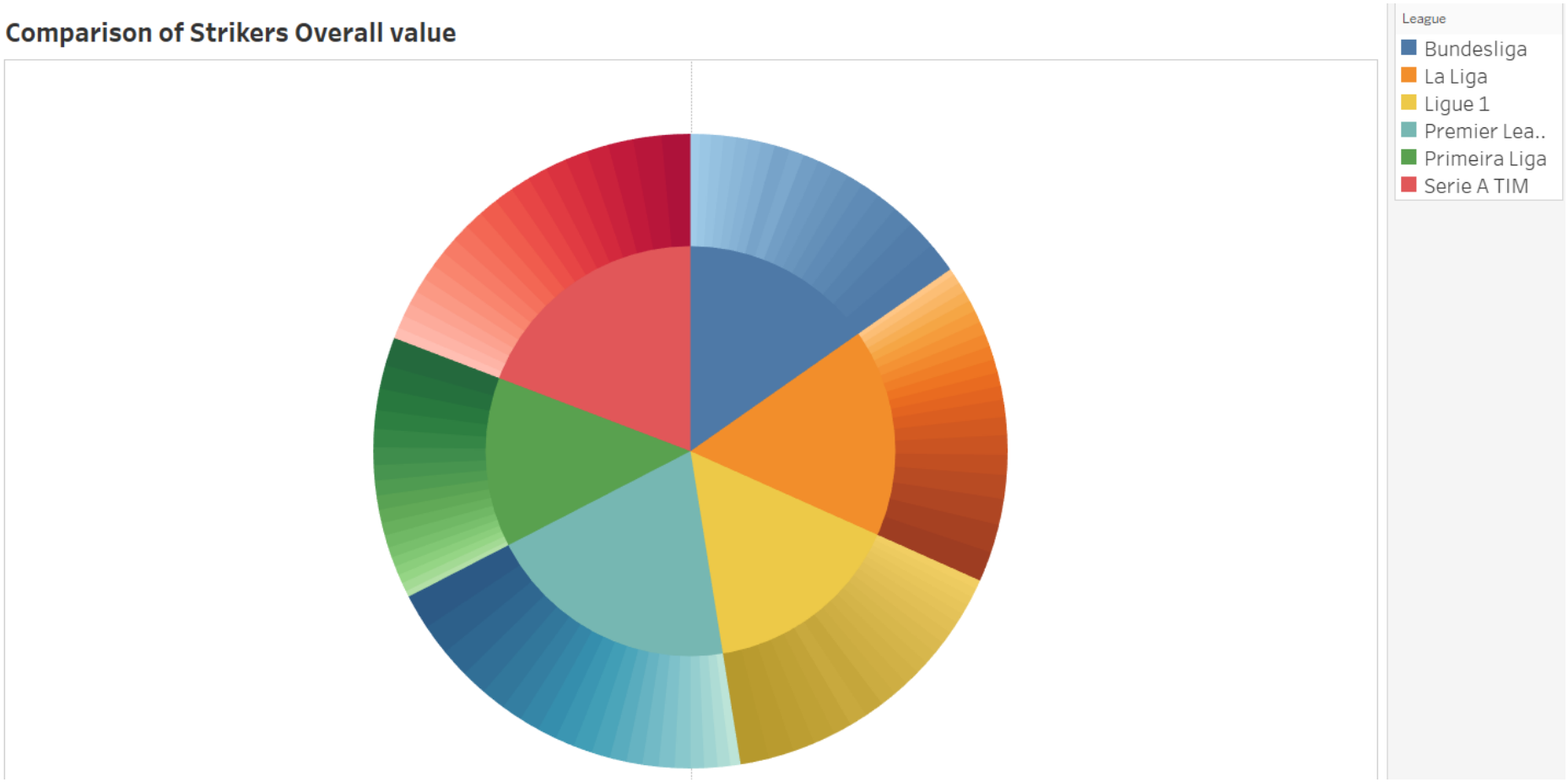

Figure 11: Sunburst chart "Hierarchical Strikers Overall Value". Own creation, source: Sharma (2018)

4.6.1 Information and Results

The sunburst provides information about Strikers on FIFA 2019. The chart begins from the center and spreads out information on the overall total value of the strikers playing in each league, as well as other relevant technical features such as finishing, headers, long shot and shot power. By interacting with the sunburst, it is possible to see the relevant quantitative values applied clockwise in ascending orders, therefore the legend helps understand how the highest valued strikers play in the Premier League.

4.6.2 Justification of chart, techniques, layout, style, and colours

The sunburst was applied as it shows hierarchical distribution of the same variable (the overall strikers value) on two different layers of leagues and clubs as proportional quantitative parts of a whole (Ward, et al., 2015).The first layer shows a 20 medium classic palette applied where the total value of strikers per angle is displayed, whereas the second layer sees a different shades of the colours applied clockwise from a lighter to a darker hues to identify the lowest to the highest value of strikers for each club represented on each radial layout (Kirk, 2019). The tooltip provides the information needed, and filters bars allow to filter by strikers' name and strikers' nationality to gain more information about how determinant certain strikers are for their clubs and the leagues they plain in on FIFA 19.

4.7 Seventh visualisation: Side-by-side Circle view

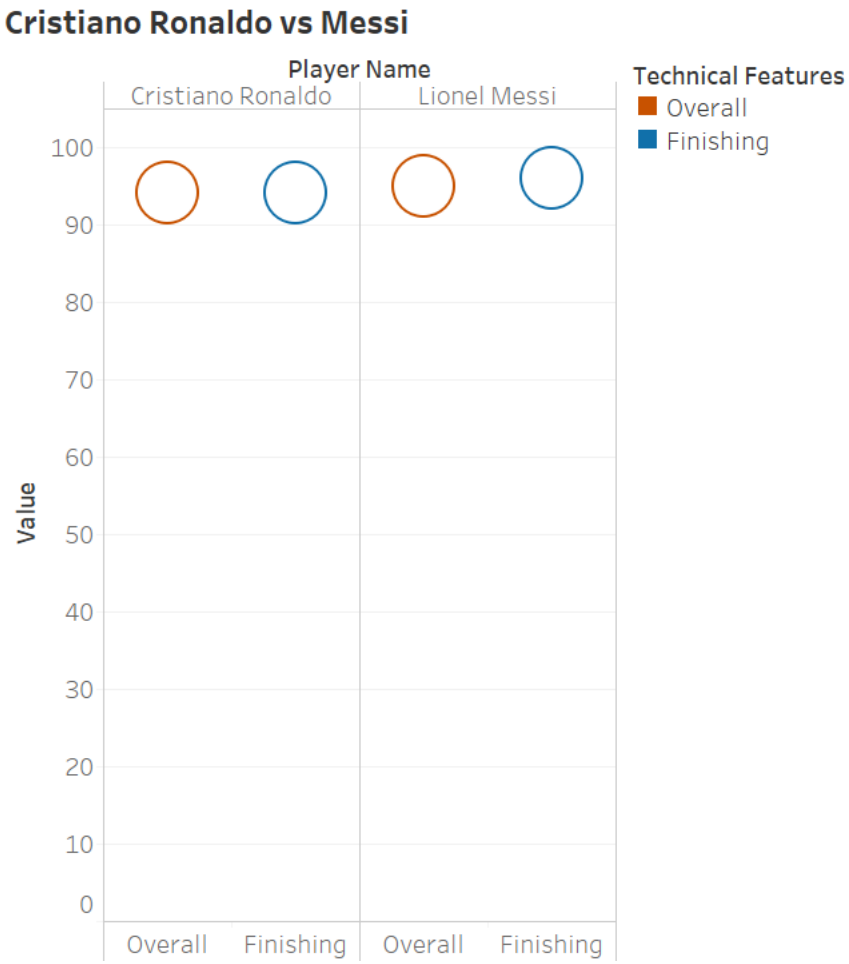

Figure 12: Side by Side Circle view chart "Cristiano Ronaldo VS Messi", own creation. Source: Sharma (2018)

4.8 Information and Results

The final visualisation above displays a comparison between the two most acclaimed strikers in the world, Cristiano Ronaldo and Lionel Messi. The comparison is based on two numerical values such as Finishing and Overall, with Messi resulting to be the best striker in the game with a value of 96 value for it's ability to score and 95 overall, against the 94 for Ronaldo for both the features.

4.9 Justification of chart, techniques, layout, style, and colours

The Circle view chart was chosen as it shows a distribution of two numerical variables for the two players (Kirk, 2019). Although the chart presents two values only initially, A colour blind palette was applied to the circles to distinguish the values (Shaffer, 2016). In fact, by interacting with the charts, more interesting information would arise, for example the better ability of Ronaldo in the heading and penalties, as well as Messi's highest ability to take successful free kicks.

5 Conclusion and Personal Reflections

This project has been a great opportunity to harness the power of interactive visualisations and publish my first project in Tableau. The main challenges encountered in the beginning revolved around the data preparation, as the original dataset contained over 15,000 observations and and a large number of numerical information. Through a data pre-processing I was able to understand the information I wanted to extract from the dataset first, and split them into five different datasets to make the analysis and the creation of visualization smoother and effective. The hardest charts to build were definitely the Dendogram and the Sunburst which required long time and applications, but the overall process was facilitated by a very good and clear dataset which contained lots of relevant and clear information.

6 References

Electronic Arts Inc., 2020. *Electronic Arts.* [Online]
Available at: https://www.ea.com/en-gb
[Accessed 10 August 2020].

Extreme Presentation, 2020. *The Extreme Presentation Method,.* [Online]
Available at: https://extremepresentation.com/
[Accessed 9th July 2020].

Kirk, A., 2019. *Data visualisation : a handbook for data driven design.* 2nd ed. s.l.:Sage.

Shaffer, J., 2016. *5 tips on designing colorblind-friendly visualizations.* [Online]
Available at: https://www.tableau.com/about/blog/2016/4/examining-data-viz-rules-dont-use-red-green-together-53463
[Accessed 9th July 2020].

Sharma, A., 2018. *FIFA 19 Player Database, Complete player database as per FIFA 19.* [Online]
Available at: https://www.kaggle.com/aishwarya1992/fifa-19-player-database/kernels
[Accessed 05 August 2020].

Ward, M., Grinsrein , G. & Keim, D., 2015. *Interactive Data Visualization, Foundations, Techniques, and Applications.* 2nd ed. Boca Raton: Taylor & Francis Group, LLC..

7 Appendix – FIFA 19 Story

Below is the link to the story published on Tableau public titled “FIFA 19 – Insights on players technical features”.

https://public.tableau.com/profile/antonio.caruso4080#!/vizhome/FIFA19-InsightsonPlayersTechnicalFeatures/Story1_1?publish=yes

The story contains five dashboards accurately placed on each story tab, to draw the viewer from general information to more detailed insights on Goalkeepers & Defenders, Midfielders, Strikers, and finally a comparison between Cristiano Ronaldo and Messi.