



FOOTBALL ANALYSIS

BUSINESS INTELLIGENCE REPORT

BY : PANKAJ PATIL

CONTENTS

Sr.No	Content	Page Number
1	INTRODUCTION	3
2	DATASET	4
3	DATASET PREPARATION	4-5
4	APPLICATION	6
5	RESULTS	11
6	CONCLUSION	12
7	APPENDIX	13

Introduction

Football is a family of the team sports and most watched sports on the planet. Unlike any other events, this requires lot of planning, taking decisions, selecting players and others. In this particular case, we will focus more on management of particular football team and how with help of the IT tools it can help in better management of the team. So, here we consider, how when the new manager of the football team, will make use of the different statistical analysis of the team, to make better informed choice and buy/sell more players in the transfer market. During summer(June-Aug), most of the managers are confused on making decisions over the transfer market (selling/buying) of the players to make the club most formidable. Specially they need thorough understanding of the market worldwide and player by payer analysis from the different clubs in order to make better choices. In this particular project, we are going to consider about the Real Madrid Football club and Since the departure of the previous manager, the new manager need to know about the football statistics of the previous year and then take informed decisions.

Business Problem:

During the transfer market period(June-Aug) of the football, most of the manager need to have the through analysis of the football market to buy in or exchange few players. There are various factors that need to be analysed in order to take the best decision. Here, quick overview of the process for a new manager goes through to buy new players at the team:

1. Analysis of the previous team Performance
2. Analysis of the Different Players in the team
3. Analysis of the Transfer Market
4. Revenue allocation for the process

In this project, I will focus on solving this decision problem for the managers of the football team and provide a thorough analysis of the entire football team with the help of the various IT tools in the market. Here, I will do analysis of the Real Madrid football club which is a Spanish football team participating in the various leagues of the Europe.

Tools :

In this project, I have used several IT tools for the thorough analysis depending on the nature of the work required for the project.

1. Excel for Data cleaning
2. Trifacta for Data Cleaning and Pre-Processing
3. Tableau for Visual Insights in the form of Stories

Justification of the tool:

Here the dataset, I would be using is the real football data from various football leagues of the Europe. In the most of the cases, I have webscrapped the data from various football websites and Wikipedia pages. In all I have collected more than 10 different datasets related that can help in better overview of the problem. Initially, I will be using the excel to get the web scrapped data. Since, most of the data is web scrapped, a lot of work is needed to be done on cleaning of the data. I will be particularly using, Trafacta for cleaning of data and occasionally R for specific cases. In the end, I plan to use the Tableau when the data is ready for the analysis. With the help of the tableau, I plan to make most out of the data and build interesting stories.

Dataset

Here is the general overview of the dataset and the detailed explanation is given in Data preparation part:

Number of games: 9074

Total number of events: 941,009

List of football Leagues: England, Spain, Germany, Italy, France

Year /season: 2011/2012 to 2016/2017

Data Preparation:

Since this is the real football data, collecting the data was never easy. Also, some of the data were web scrapped from the live commentary of the texts on different football websites. I found one single huge file which as collective data of all the important leagues of the Europe. This dataset contains 9074 games of the footballs clubs in the Germany, Spain, England and Italy in the Europe. It also has stastical analysis of the football bets on it. As I kept working on my project, I needed some other different data related to transfer markets, player stats, club stats to get insights. So, I started web scrapping the data which I needed for the analysis. In total, I have used a total of 16 different datasets in the project.. Below I will provide a brief description of each datasets and the problems I faced in each of them (note: Here all the datasets are the names of the excels files (csv)):

1. Events and Ginf : This is the main dataset which has all the players name, goal records, foul records from 9074 games of the different clubs in the Europe. It also contains the betting history of different websites and the odd of winning/losing the game Since this datasets was huge with almost records of more than 100k , most of the data of the attributes were null values so it did not make sense to use all of them. In some cases, the order of the data was messed up, so I had to use Trifacta to get rid of those data.

2 . Real Madrid: This data is about the current squad of the real madrid team and different information of the players. I have web scrapped this data from the official website of the Real Madrid and Wikipedia. Since after scrapping of the data, it was not arranged properly,so I used trifecta to modify as well as excel to make changes and clean the data.

3. RM2: This data is about the all the matches that Real Madrid team played in the La Liga (one of the Spanish football league). It includes all the wins, losses and scores of the team. I have even web scrapped this data from the Wikipedia page of Real Madrid.

4. RM4: This is a small dataset of the all the cups that Real Madrid team won during the academic year (2017-2018). After web scrapping of the data,I used excel tools to clean and modify the data.

5. RM5: This datasets gives all the list of the strikers at the Real Madrid team and their necessary information to compare different players by in various ways (By goals, by cards..). After web scrapping of the data, I have used excel tools to clean and modify the data.

6. RM6 : This dataset is about the performance of the entire real Madrid team in different aspects of the game such as games played, shots on goal, shots and other. After web scrapping of the data, I have used excel tools to clean and modify the data.

7. RM7 : This datasets contains information of the all midfielders in the Real Madrid team and it allows to assess the performance of the individually. After web scrapping of the data, I have used excel tools to clean and modify the data.

8. RM9: It contains information of the defenders of the Real Madrid team to assess the performance of the players. After web scrapping of the data, I have used excel tools to clean and modify the data.

9. RM10 : This datasets provides information of the shooters(forward players) around all the different teams in the Europe leagues. To limit the size of the datasets, I have only considered the Top 50 players in this category.
After web scrapping of the data, I have used excel and traficta tools to clean and modify the data.

10. RM11 : This datasets provides information of the goal keepers(goalie) around all the different teams in the Europe leagues. To limit the size of the datasets, I have only considered the Top 50 players in this category.
After web scrapping of the data, I have used excel and traficta tools to clean and modify the data.

11. RM13: This datasets provides information of the defenders (back players) around all the different teams in the Europe leagues. To limit the size of the datasets, I have only considered the Top 50 players in this category.
After web scrapping of the data, I have used excel and traficta tools to clean and modify the data.

12. RMtransfer : This dataset contains information of the transfer market of the La Liga League(Spanish league). It contains the budget, expenditures and revenues of all the team for

the transfer market. After web scrapping of the data, I have used excel and traficta tools to clean and modify the data.

13. RMT1 : This dataset contains all the information of the new players joining the Real Madrid team along with their market value as well as transfer fees.

14: RMT2: This dataset contains all the information of the players leaving the Real Madrid team along with their market value as well as transfer fees.

Application

The objective of this whole project was to solve the business problem using the various IT tools. In this project, I have used the Excel and Trifecta for data cleaning. After cleaning of data, the main part was to use this cleaned data to generate the insights from it. In this part, I have used the Tableau software to get meaningful insights from it and answer the main business problem using tableau stories. The main business problem in this project was to analyse the previous year team performances, analyse each player performance and accordingly sign new players with all these insights.

So Here I will discuss how I generated insights and solved the business problems with screen shots from the tableau.

This was the Real Madrid Squad in 2017/18(Fig 1). As we can see, there are Four categories of players(P) namely: Defenders(DF), Forward(FW), Goal Keepers(GK)Midfielders(MF). There is also value of each player assigned (transfer Market column) to better understand the performance.

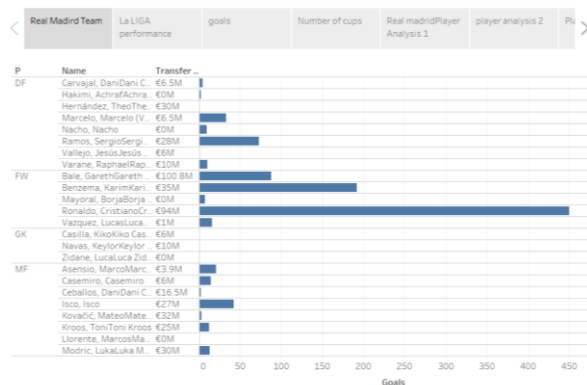


Figure 1

Fig 2 shows the performance of the Real Madrid team against each of its opponent in the La-Liga competition. This insight helps us to understand the tactics, performance of the players to get better assess them.

Real Madrid Team	La LIGA	goals	Number of cups	Real madridPlayer Analysis 1	player analysis 2	Pl
performance						
AT	Date	HT	score			
Atletico	25-Feb-18	Real Madrid	4-0	1		
Atletico Bilbao	18-Apr-18	Real Madrid	1-1	1		
Atletico Madrid	28-Apr-18	Real Madrid	1-1	1		
Barcelona	23-Dec-17	Real Madrid	0-3	1		
Celta Vigo	12-May-18	Real Madrid	0-0	1		
Deportivo La Coruña	21-Jan-18	Real Madrid	7-1	1		
Eibar	22-Oct-17	Real Madrid	3-0	1		
Espanyol	02-Oct-17	Real Madrid	2-0	1		
Getafe	03-Mar-18	Real Madrid	3-1	1		
Girona	18-Mar-18	Real Madrid	4-1	1		
Leganes	18-Apr-18	Real Madrid	2-1	1		
Levante	09-Sep-17	Real Madrid	1-1	1		
Malaga	25-Nov-17	Real Madrid	3-0	1		
Real Betis	20-Sep-17	Real Madrid	0-0	1		
Real Madrid	03-Dec-17	Atletico Bilbao	0-0	1		
	03-Feb-18	Levante	2-2	1		
	08-May-18	Barcelona	2-2	1		
	07-Jan-18	Celta Vigo	2-0	1		
	09-May-18	Sevilla	0-2	1		
	10-May-18	Osasuna	1-2	1		
	14-Oct-17	Getafe	1-2	1		
	15-Apr-18	Malaga	1-0	1		
	17-Sep-17	Real Sociedad	1-3	1		
	28-Feb-18	Real Betis	3-0	1		
	18-Nov-17	Atletico Madrid	0-0	1		
	18-May-18	Villarreal	2-0	1		
	20-Aug-17	Deportivo La Coruña	0-0	1		
	21-Feb-18	Leganes	1-3	1		
	25-Sep-17	Atletico	1-0	1		
	27-Feb-18	Espanyol	1-0	1		
	27-Jan-18	Valencia	1-4	1		
	28-Oct-17	Girona	2-1	1		
	21-Mar-18	Las Palmas	0-0	1		
Real Sociedad	10-Feb-18	Real Madrid	5-0	1		
Sevilla	09-Dec-17	Real Madrid	5-0	1		
Valencia	27-Aug-17	Real Madrid	5-0	1		
Villarreal	19-Jan-18	Real Madrid	0-1	1		

Figure 2

In fig 3: It shows the goals scored by the players of the Real Madrid team. Interesting thing to notice, due variability in data it shows null values as the highest.

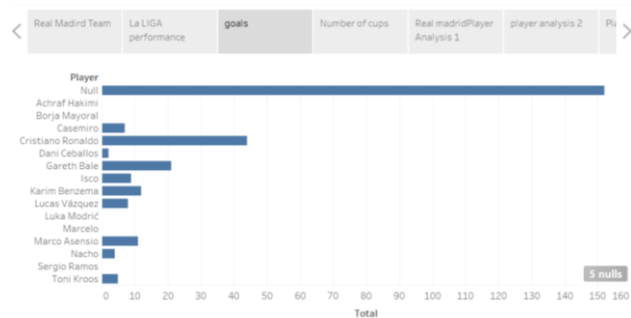


Figure 3

In figure 4(below): It shows the different cups that Real Madrid team has won in the whole year. Interesting things to notice here is, since real Madrid have lost(3rd) in La Liga competition, it would be interesting to see how the team has played in these leagues since there are more than 38 matches played by the team in this competition.

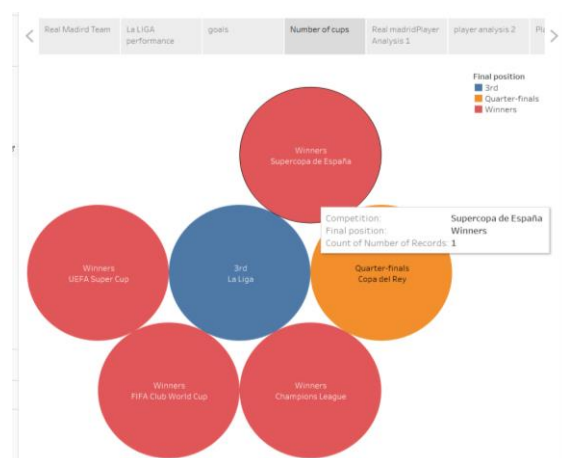


Figure 4

Analysis of the players performance: Here we will do analysis of the entire team in terms of passes, accuracy, number of goals, shots on target, fouls and other factors. This will help to assess the weak player and replace with the new player.

Forwards (figure 5):

In this case, to better assess the players performance, I took various factors into consideration such as described below:

- 1) Assists(A)
- 2) Number of goals(G)
- 3) Games played(GP)
- 4) Shots(S)
- 5) Shots on goal(SG)
- 6) Yellow cards(YC)
- 7) Red cards(RC).



Figure 5

Midfielders: In this case (figure 6), I took the factors into consideration such as crosses, fouls while assessing the performance.

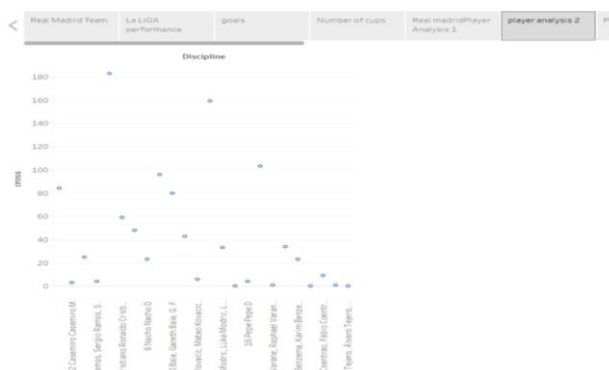


Figure 6

Defenders: In this case (figure 7), I took the factors into consideration such as interceptions, cross, fouls while assessing the performance.

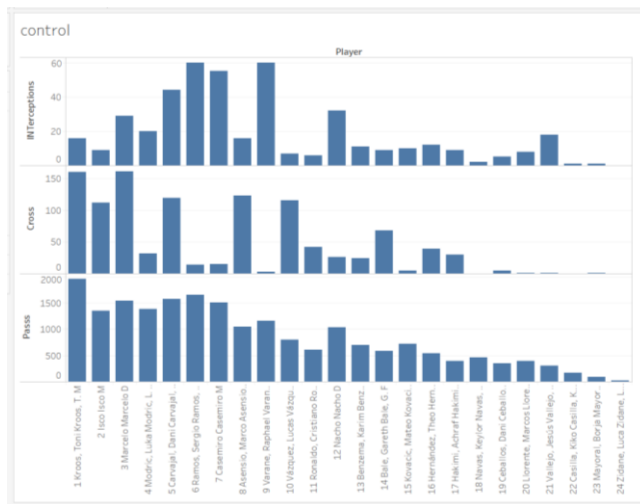


Figure 7

Analysis

Based on these analysis, I have come up with the player rating for the Real Madrid team. This can help in choosing the player and signing new player according to the need of the team.

R	Player	CM	KG	Apps	Miss	Total	Out	Off	Box	Size	Yard	Box	Penalty	Area	Goal
1	Cristiano Ronaldo	33	MIDL,FW	187	83	27	2293	6.6	1.4	0.6	4.6	7.94			
2	Casemiro	26	DMG	185	84	30	2601	1.1	0.7	0.1	0.3	7.49			
3	Gareth Bale	26	MIDCL,FW	185	82	26	1903	3	0.7	0.3	2	7.49			
4	Toni Kroos	26	MID	183	76	27	2268	1.5	1	-	0.5	7.47			
5	Ahmad Hakimi	19	DMG	180	69	9	757	0.3	-	-	0.3	7.34			
6	Luka Modric	33	MID	172	66	26	1976	1.3	0.7	-	0.6	7.29			
7	Iago	28	AM,CL,FW	176	79	30	1781	1.6	0.6	-	1	7.16			
8	Sergio Ramos	30	DMG	184	82	26	2292	1.1	0.1	0.2	0.7	7.09			
9	Marcelo	26	CL	174	80	28	2262	1.2	0.5	-	0.6	7.07			
10	Daniel Carvajal	26	DMG	173	73	25	2223	0.5	0.2	-	0.2	7.06			
11	Marco Asensio	22	AM,CL	182	78	32	1839	1.6	0.7	0.1	1	7.02			
12	Karim Benzema	30	FW	185	81	32	2156	2	0.2	0.2	1.6	7.00			
13	Theo Hernandez	20	DMG	184	82	13	963	0.2	-	-	0.2	6.98			
14	Josko Valjejo	21	DMG	184	79	7	630	-	-	-	-	6.96			
15	Keylor Navas	185	80	27	2430	-	-	-	-	-	-	6.94			
16	Lucas Vázquez	26	AM,FW	173	70	33	1670	1	0.3	0.1	0.6	6.91			
17	Nacho	28	DM,CL	180	76	27	2073	0.3	0.1	-	0.2	6.90			
18	Raphael Varane	26	DMG	191	81	27	2324	0.4	-	-	0.3	6.86			
19	Mateo Kovacic	24	MIDCL	178	78	21	969	0.4	0.3	-	0.1	6.73			
20	Dani Ceballos	21	MIDCL	179	70	12	397	0.6	0.3	-	0.3	6.67			
21	Kiko Casilla	31	GW	190	88	10	960	-	-	-	-	6.53			
22	Borja Mayoral	21	FW	182	74	14	338	1.1	-	0.3	0.6	6.52			
23	Marcos Llorente	28	DMG	184	73	13	540	0.5	0.5	-	-	6.46			
24	Lucas Zidane	183	77	1	80	-	-	-	-	-	-	6.17			

Figure 8

Insights

The following conclusions can be made based on the analysis of the previous visualisations :

1. Team needs more new players in Défense line
2. Team needs a better choice of the Goal keeper.
3. Team need to improve on it strategy.

Real Madrid Characteristics	
+ Strengths	- Weaknesses
<ul style="list-style-type: none"> Attacking set pieces Very Strong Attacking down the wings Very Strong Creating long shot opportunities Very Strong Creating chances using through balls Very Strong Creating scoring chances Very Strong Counter attacks Strong Shooting from direct free kicks Strong Finishing scoring chances Strong Defending set pieces Strong 	<ul style="list-style-type: none"> Avoiding offside Weak Aerial duels Weak Protecting the lead Weak Stopping opponents from creating chances Weak
Real Madrid's Style of Play	
<ul style="list-style-type: none"> Control the game in the opposition's half Possession football Short passes Attempt through balls often Attack through the middle Attacking down the left Rotate their first eleven Non-aggressive 	

Buying a new player

Here are the Options for Defenders (figure 9), these are the players of the all teams participating in the La Liga competition. They are categorised according to the different factors such as blocks, tackles and others.

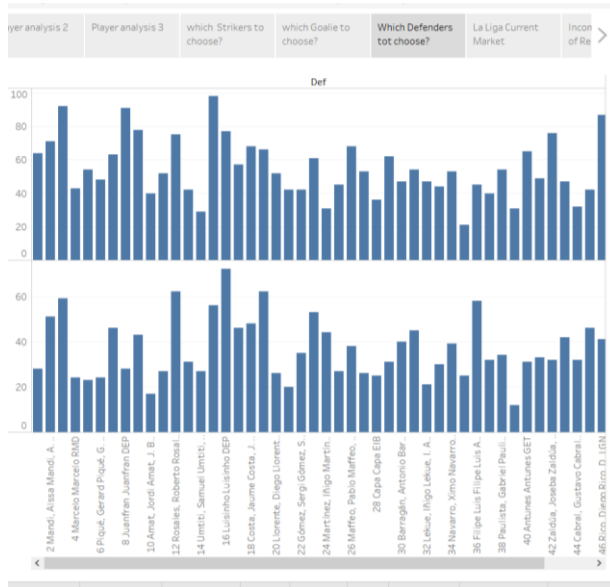


Figure 9

Here are the option for the goalkeepers (figure 10): these are the players of the all teams participating in the La Liga competition. They are categorised according to the different factors such as clean sheets, goals allowed and goals saved.

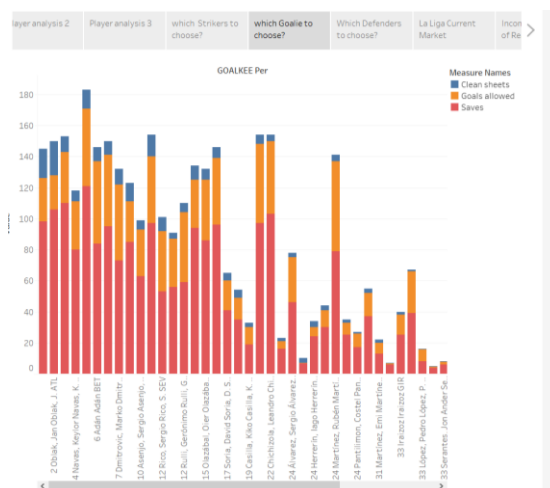


Figure 10

Now, once we have finished with the players performance rating, we need to focus on the transfer market and analyse how much can team spend on its new player to buy from the other team.

Here is the transfer market analysis (figure 11):

Department	value	
Arrivals:	378	1
Balance per club:	£1,854,765	1
Balance per player:	£103,751	1
Departures:	377	1
Expenditures per club:	£38,159,325	1
Expenditures per pl.	£2,122,763	1
Income per club:	£40,014,090	1
Total balance:	£37,095,300	1
Transfer expenses:	£763,186,500	1
Transfer revenues:	£800,281,800	1

Figure 11

Transfer Market Analysis:

Also it needs to be considered, Real Madrid team has expenditures of over 36.54M\$ for buying players from other teams and revenue of 115.86M\$ by selling players last year. This leaves Real Madrid with the amount of 79.20M\$(115.86M\$-36.5M\$) for signing the new players. This budget can be increased but it depends on the expenditures allocated for other resources. In this case, we assume **Real Madrid** team is willing to spend this profit of **79.20M\$** this year to buy new players.

Final Result:

With the help of all the analysis done above, new manager of Real Madrid can think of signing these new players for the upcoming season 2018/19.

SR.NO	PLAYER	POSITION	MARKET VALUE
1	Fernando Pacheco	GK	158K\$
2	Umtiti Samuel	DF	50M\$
3	Mario Vil	DF	500K\$
4	Jan Oblak	GK	70M\$

Conclusion

Evaluate the support provided by the IT tool(s) for solving your problem.

The main objective of this project was to solve the business problem with the help of the IT tools. In my project, the business problem was to assess the performance of the team and sign new players as per the team needs. Initially, I have used Excel and Trifacta for cleaning of the data. I found Trifacta to be extremely helpful when you have large datasets as it pops up with the new suggestion for cleaning and the user experience is even better. For Small datasets, Excel was more efficient since it was easy to make changes in data manually. Once the data is prepared, Tableau has useful features which transforms the data into meaningful stories.

Do you consider, ex-post, that the choice of the tool(s) was appropriate? Explain why.

Yes, the choice of tools, considered for the project was appropriate. As discussed above, Excel and Trifacta are very efficient while cleaning with the raw data. Since most of the data was web scrapped, it was easier to manipulate and clean the data with the help of the excel. Even with Trifacta, Large datasets were easily Manageable. I could have even used R for cleaning data but It would have been a bit time consuming to work in it, when comparing with Excel.

What did you learn from this project?

The project was mainly about solving the business problems with the help of the IT tools. During this project, I came across various new tools to explore the data as well to make meaningful insights. It was good experience working with the large number of datasets and being able to modify the data and make necessary changes as per the requirement. Since, I have used most of the data in this project by scrapping the web data, it was nice experience to learn about the web scrapping and able to manipulate it into clean raw data. Apart from Data cleaning, It was amazing experience to explore the different features present in the Tableau. The idea of able to create the stories with the help of the data fascinated me and gives great insights about the topics.

Sources and appendices:

SR.NO	Dataset	Content	Source
1.	Events	Football records of 9074 games played in Europe	Kaggle
2.	Ginf	Betting History of games and odd of win/losss	Kaggle
3.	Real Madrid(RM)	Current Squad of Real Madrid Team	RM website
4.	RM2	All Matches played by RM in La LIGA competition	Wiki
5.	RM4	History of cups RM won in 2017/18	wiki
6.	RM5	List of all strikers in RM	RM website
7.	RM6	Performance of the RM team	RM website
8.	RM7	Midfielders of the RM team	RM website
9.	RM9	Defenders of the RM team	RM Website

Sr.No	Dataset	Content	Sources
1	RM10	Players around different Leagues	Football websites
2	RM11	Goalkeepers around different leagues	Football websites
3	RM13	Defenders around different leagues	Football websites
4	RMtransfer	Transfer market of the La Liga League	Football Websites
5	RMT1	Incoming players at RM	Football websites
6	RMT2	Outgoing players at RM	Football Websites

Links for the Dataset sources:

1. <https://www.kaggle.com/secareanualin/football-events>

All the remaining datasets were web scrapped from the following websites:

1. <https://www.transfermarkt.com/schnellsuche/ergebnis/schnellsuche?query=jan+oblak&x=0&y=0>

2. <http://www.goal.com/en-us/news/fifa-18-real-madrid-player-ratings-ronaldo-the-highest-rated/vpfvwtk30adm1wgk9fh98u4>

3. <https://www.whoscored.com/Teams/52/Show/Spain-Real-Madrid>

4. <https://www.independent.co.uk/sport/football/premier-league/fantasy-football-scout-tips-gameweek-1-fpl-best-players-a7887121.html>

5. https://en.wikipedia.org/wiki/2017%E2%80%9318_Real_Madrid_C.F._season