
# Project Report

## Data Visualization

Afonso Quintino – 20220698

Danylo Ivanchyshyn – 20221692

Rodrigo Brigham – 20221607

Vasco Silva – 20221638

NOVA IMS Information Management School

# Introduction

For this project, we were proposed to make a dashboard with several relevant visualizations from data selected by the group. We decided to use a database from the popular game, Football Manager 2023, also known as FM23, which we slightly altered from a dataset we found in Kaggle.

The dashboard produced by our group is available in the provided link and all documentation regarding the preparation of this project can be found in the following Github Repository.

Football Manager 2023 is the most recent version of the worldwide known football simulation video game. The player takes on the role of a football club manager, tasked with handling a club to compete in various leagues and tournaments. The game includes a vast database of real-world football players, coaches, analysts and teams, as well as extensive statistics and information about the sport. The player must make decisions regarding tactics, training, transfers, and finances, among other things, to lead their team to success. However, differently to other football simulation games like FIFA or Pro Evolution Soccer, the users have no direct control over the actions of the field players after the decisions are made. The game is known for its realism, depth, and complexity, and is popular among football fans and enthusiasts. This game has an extremely passionate fan-base that regularly create packages and tools to enhance the experience of all players worldwide. It has been the source of inspiration for many analysts and coaches that are currently getting into the professional world, such as the example of Will Still, a Belgium coach that has gone on an historical 18 game unbeaten run with the Ligue 1 side Reims.

All members of the group have a profound interest in football and have enjoyed at some point playing Football Manager and this inspired us to use this opportunity to create our own tool that is both simple and effective that could be used by anyone that enjoys the game. We decided to use the data set that we found to create a dashboard that allows an analysis of any given league or club that a user might decide to start a game with and help him plan the squad for the seasons ahead.

# Data and Methods

The original dataset was comprised of player information and attribute ratings at the beginning of the 2022-2023 season from all originally available players. The dataset contained 8452 rows and 83 variables which included personal information like the name, age, club, nationality and position of the player; a score from 1-20 in technical attributes like shooting, passing and dribbling, physical attributes like agility, balance and stamina and mental attributes like composure, leadership and determination; a score from 1-20 that a player has, according to his attributes, in each position; and financial information regarding their market value and wage.

Although the dataset was quite vast and complete, there were many technical problems that we had to initially solve in order to then apply data visualization methods and techniques. Some of these were solved manually using Microsoft Excel, whilst others were solved using Python through a Jupyter Notebook that can be found attached to the project documentation.

The first major problem that we found in the data was the lack of information regarding the league that a player was in, although we had information about the club. We manually added the top 11 leagues in the world and removed all other rows that had no league. This led us to only keeping 4640 rows, although it ensured some extra level of data quality. We also renamed some variables to have a more user-friendly name (for instance, the position name 'AML' was changed to 'LW' which represents the left winger position). We also manually corrected an issue

that occurred with players that had the same name, and the duplicates would have a number at the end of the same as 'João Mario1'. We removed these numbers from all names.

Moving on to python, we started by solving an issue regarding the 'Potential Ability' variable. This variable represents the maximum ability that a player can reach throughout a user save game. It is never lower than the current ability, although it may be equal which represents a player that can´t evolve through the seasons. However, for very young players, whose future is still quite uncertain, this variable has a negative value which represent a dynamic potential. Their in-game potential can randomly vary in a given interval according to how a player's game is loaded. To solve this, we added the absolute value of these negative values to a player's current ability and kept it as the maximum potential for a given player.

Another big issue we found in the data was regarding player valuation. We started by rounding the valuation into millions and kept 3 decimal places to have a more tangible and visual number. Although all players have a value, there are some players that a given club may consider 'Not for sale' in-game. This means that you cannot buy the player regardless of your offer or their 'real' transfer value. In the data, these players had a value of 347.975 million euros. To solve this issue, we gave Nan values to all these players and created a simple linear regression that used the 'Current Ability', 'Potential Ability', 'Age' and 'Wage' variables as regressors to calculate a more realistic valuation.

Finally, the variables 'Nationality' and 'Position' were not atomic, something that we wanted to correct. For the nationality, a player could have multiple countries listed as nationalities, but we kept only the first as it usually pertains the nation for which they would most likely play for in the national team. For the 'Position' variables, we used the 10 variables that had a score from 1-20 for the given player in each of those specific positions and selected the one in which he performed the best as his unique position.

In the end, we used the following variables in our visualization dashboard:

| Variable | Description |
|---|---|
| Name | Name of the player. |
| Position | Main position the player occupies on the field. |
| Age | Age of the player. |
| Club | Club that player belongs to |
| League | League of the players club |
| Country | Country of the players league |
| Nationality | Nationality of the player |
| Current Ability | Skill of the player at the beginning of the 2022-2023 season. It is the sum of the ratings of all attributes. |
| Potential Ability | Maximum ability that a player can reach throughout the game |
| Value (in millions) | Rating/Skill of the player in the beginning of the season. |

We implemented the project using Tableau. We encoded the generic positions using colour (Goalkeepers – Red, Defenders – Yellow, Midfielders – Green, Forwards - Blue) and set the filtering to include information automatically only regarding a specific team or league for all visualizations in the dashboard. The following image of our dashboard uses Manchester City as the selected club for the visualizations as it is a club that has diverse set of players that pertain quite well as example for the potential use of this tool.

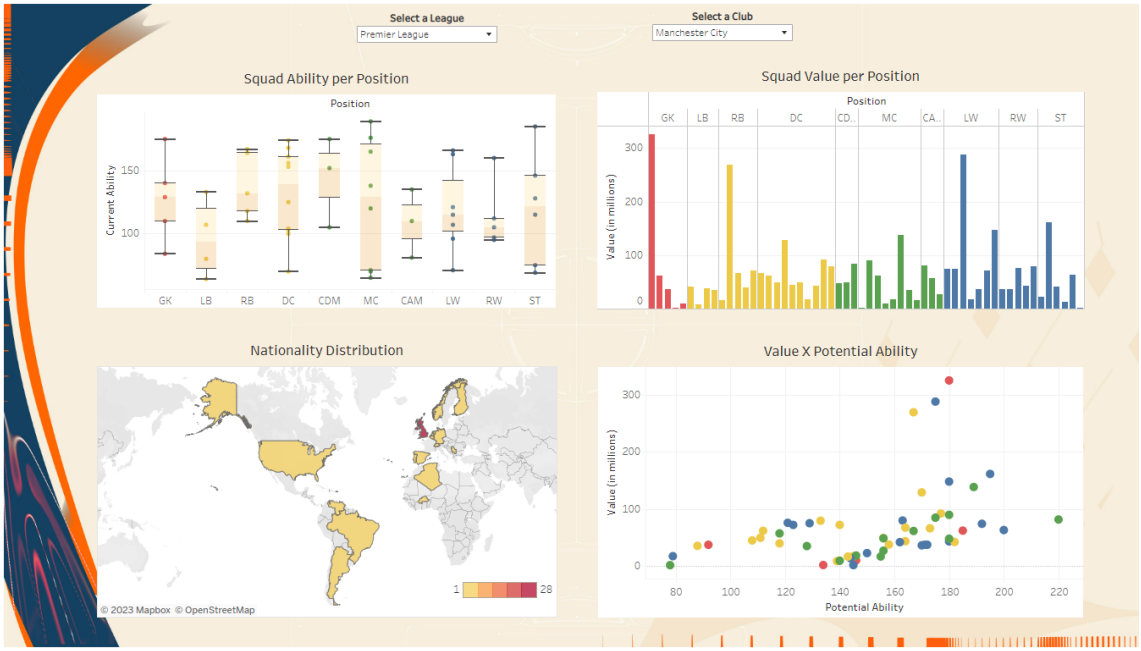


*Figure 1 - Manchester City Dashboard*

# Reading the Visualization

**Visualization 1**: This visualization gives the user a quick overview of the distribution of the team's ability in each position through the use of box plots. This allows a player that has just started a game with a club to overview highly rated areas of the pitch which can then be used to set the most adequate tactic to use onwards. It also gives a view of positions that may be lacking in player depth or quality, as well as positions that might be overloaded with players. In the specific case of Manchester City, we can quickly identify that the LB position might be slightly weaker when compared with the rest of the squad, but they possess, as would be expected from one of the best teams in the world, a fairly balanced distribution of quality players in their roster.

**Visualization 2**: This visualization uses a bar chart with the players value grouped by position. It allows the user to have a quick overview of his players value in comparison with each other. It is grouped by positions which also allows the player to see the areas of the pitch which may be more or less valuable, or that can have some off-loadable high value players. In Manchester City´s club it is quite noticeable the existence of three extremely high value players, but also some very low valuations in positions that have many options and that could be sold in order to trim down the roster slightly.

**Visualization 3:** This visualization allows the player to see the nations from where his players belong to using a world map. IN FM23 there is a layer of player adaptation that regards similar nations. For instance, if you have a core group of players from Brazil, a brazilian signing may adapt more easily to the squad than a different nation, which will reflect in their performances and evolution. Therefore, this is a valuable attribute to have in mind when planning the squad. We used a colour gradient to visually convey the areas of the world where a club has more players from. Manchester City has a diverse group of players from all around the world but mostly from England. This may be due to the high level of their academy which is mostly composed of English youth players that eventually make it into the senior squad.

**Visualization 4:** This visualization uses a scatter plot with the 'Value (in millions)' and 'Potential Ability' variables and allows the user to analyse the distribution of his players value against their maximum future ability. This is extremely useful for squad planning as it allows the manager to find, for instance, high value players that have low potential and might be worth cashing in on a transfer or players that have extremely high potential and a low value and might be worth keeping in the squad for the future.

It is now important to note that all these visualizations interact with each other and should be used together to find relevant insights into the squad. For instance, a user can use the visualization 3 to filter players in the squad by nation and find more granular level insights or on an even more micro level, a user can even see all information regarding just a specific player by clicking in his name in any of the present visualizations. Two or more players from the same team may also be directly compared by selecting them in the visualization. All datapoints in the dashboard are labelled in order to have a user friendly and intuitive experience while navigating.

# Conclusion

We this project, we are confident that we created a useful tool that can be shared with Football Manager players all around the world and be used and understood by anyone regardless of technical background. It successfully achieves its original goal of creating a dashboard that allows early squad planning and insightful analysis of any club.

For future improvements it would be significant to add more leagues if there is the insurance of high data quality and to also have a pipeline that allows for FM23 updates to be regularly placed in our database and therefore keep the dashboard relevant. It could also be interesting to make, in the future, a visualization that allows an intuitive comparison of two or more clubs if a manager is in doubt regarding which one to choose, although the current dashboard allows for the selection/analysis of all teams in a league at the same time. Finally, although it is out of the scope of the course, it would be quite interesting if there is a way that we could create an app that could be downloaded by users and receive their current game save as a database input as there are some user-created downloadable packages that do this for other areas of the game.

# References

**Football Manager 2023 Dataset | Kaggle. (n.d.). Retrieved February 20, 2023, from https://www.kaggle.com/datasets/platinum22/foot-ball-manager-2023-dataset**

**Stade de Reims' Will Still Unbeaten in 12 Matches After Getting Start in Video Game |Bleacher Report. Retrieved April 1 2023, from https://bleacherreport.com/articles/10063532-stade-de-reims-will-still-unbeaten-in-12-matches-after-getting-start-in-video-game**