**A logo of a football team

Description automatically generated**

**Project FIFA Moneyball – Real Madrid FC.**

**Team member (yellow team)**

1. Berna Ozer
2. Tarinee Meesters
3. Pablo Corral Robles
4. Pablo Pazos

**Objectives**

1. Finding insights of the dataset
2. Data cleaning and transformation for analysis
3. Applying the statistical and machine learning techniques for fact finding, setting criteria, answering questions on the area of our interests what the best strategies is to select team players for Real Madrid Club
4. Provide solutions and giving recommendation how to select the best players and which players should be replaced based on our criteria.
5. Using data visualization to communicate to audiences.
6. Present your insights in a thoughtful, clear, and accurate way.

**Deliverables**

1. Well-commented Jupyter notebook for our analysis
2. The final dataset after all cleaning and transformations.
3. Repository with project finding, code, and data visualization
4. 5 minutes presentation of our objectives, main findings, and challenges.

**Links to our project**

1. GitHub repository

<https://github.com/TMeesters/project-fifa-moneyball>

1. Trello Board (project tracker)

<https://trello.com/b/DmGu0Prq/fifa-project-kanban>

1. Link to presentation

<https://docs.google.com/presentation/d/12CluylRD2OSSCxhEodF3SApVDhRgyMdY/edit#slide=id.g297d7eedb59_3_8>

**Dataset**

All the necessary files, as well as the dataset, can be found in the following repository: [Mid-bootcamp project -FIFA Moneyball.](https://github.com/ta-data-pt-rmt/project-fifa-moneyball)

**Team Strategies**

1. Using statistical and machine learning techniques to predict the best selection of players for Real Madrid club.
2. Selecting the best player based on their performances and potentials
3. Finding the best possible future players to replace the existing ones
4. Predicting wage for new players

**Data Cleaning & Transformation**

There are many columns that are not related to our analysis. We have selected only relevant data. These are: 'Overall', 'Potential', 'Value', 'Positions Played', 'Best Position', 'Nationality', 'Age', 'Club Name', 'Wage', 'Release Clause', 'Club Position', 'Contract Until'.

Then we create a subset with only the players that play for Real Madrid.

A screenshot of a computer

Description automatically generated

**Data Visualization**

We use heatmap diagram to check correlation in the dataset. This helps us to identify what could be an interesting area for our analysis.

**Criteria to select which players to transfer**

1. Age: players who are older than 30 years old will be transferred.

2. Wage: take wage of player into consideration. And use this data to calculate saving that can be used as a new budget.

A screenshot of a phone

Description automatically generated

**Criteria to select best possible future players for replacement**

1. Performance and potential
2. Age
3. Position
4. Contract end date (taking into consideration after selection of the potential players)

Searching new players in different position. The requirements are;

1. Players must have at least 85 points potential.

2. Players age in between 20-25 years old for LW position, not older than 28 years old for CM position and not older than 25 years old for ST position.

The filtered data is put in a new data frame, and we focus on the top 10 players that match the requirement. Then we create “Transfer index” to check how good the potential of each player is, compared to their release clause.

Based on this index, we select the player that has the highest transfer index to perform further analysis.

A screenshot of a table

Description automatically generated

**Predicting wage**

We standardize the feature data using a standard scaler then reshape the data of the selected players to match the shape of the other data. Then create a Linear Regression model to predict the wage for this player.

**Main Challenges**

1.Time limitation

There are many interesting areas we would like to further analyse but we don’t have enough time. So we can only focus on one main interesting area which is predicting wage of the potential players.

2. Trello Board

Trello Board is a very useful tool to track project status, but this is an unfamiliar tool for the team. We spend quite some time to learn to work with this tool and create nice features. Combining Trello Board and working on the project become overloaded for us.

3.Too many variables

There are many unrelated data. We select only useful data and create a new data frame for the analysis.