

## **Data collection**

In order to create the model, we collected data about 103 defense players – center, right and left backs, that has played in the English premier league in the season of 2019-2020. In other words, all the data refers to the season of 2019-2020. The data was collected from two main websites: [fotmob](#) and [premierleague](#). For each player, we collected data about 24 categories which characterizes him in the best way to our opinion. We believe that based on these categories, every player's market value is determined. Here are our characteristics:

- Age – the player's age
- Height – the player's height
- Country – the native country of the player
- Minutes – total minuets that the player has played in the full season
- Interceptions – total number of passes blocks
- Blocks – total number of shots that the defender has blocked
- Clearances – total number of clearances
- Total number of games without conceiving goals
- Total number of goals conceded when the player was playing on the field
- Total number of duels won by the player
- Total number of duels lost by the player
- Total number of aerial duels won by the player
- Total number of aerial duels lost by the player
- Total number of tackles
- Total number of assists
- Total number of key passes
- Total number of crosses
- Crosses accuracy percentage
- Total number of accurate long balls
- Total number of yellow cards
- Total number of red cards
- Total number of goals
- Total number of passes

In addition to all these categories, we found the current market value of each player. This is the feature that our model is trying to predict, based on the data we collected about the 103 players.