“Detecting best handball players”

Santiago Caballero

**Introduction & summary**

Sport’s statistics is a quite new field in this part of the world. In Uruguay, player’s identification and selection is done “manually” based on comments, perceptions, but not on truly numbers, statistics, or other mathematical source.

In this context, the purpose of this analysis is to contribute to professionalize this activity by identifying key characteristics in handball players that may be of interest for each club in its scouting.

**Background – Context**

At present, there is few statistics data on sports in Uruguay and is not easy to obtain. In addition to this, there are measurement errors and data is frequently not updated, which directly affects any conclusion driven from it. However, having data is better than not having it, as it helps to emigrate from decisions based on perception to decisions made based on statistics. In this analysis, the aim is to provide an amateur club with information about the new players that are training in their principal and Under-23 team, and how they could possible meet the present and future club’s needs.

**Data**

The dataframe analyzed is provided by a handball amateur club and includes data on new players training in its principal and Under23 teams. This a small database, but the main objective is to provide a sample on how to detect patterns that may be used in a much broader data, which could be applied either in other clubs or sports.

The raw data contains the following variables:

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Unnamed: 0** | **Unnamed: 1** | **Unnamed: 2** | **Edad** | **Unnamed: 4** | **Unnamed: 5** | **Unnamed: 6** | **Unnamed: 7** | **Unnamed: 8** | **Unnamed: 9** | **Comunicado** | **Respondio** |
| **Description** | Player’s name | Phone | Date of birth | Age | Previous experience | Position in the field | Left handed or right handed | First team or U-23 | Physical description | Why se came | Received a notification | Answer to this notificaction |
| **Type** | object | object | datetime64[ns] | float64 | object | object | object | object | object | object | object | object |
| **Count** | 17 | 15 | 16 | 16 | 16 | 16 | 16 | 16 | 12 | 12 | 15 | 10 |

**Methodology**

First, the last four columns are dropped as they are not important for the analysis. Then, the rest of the columns are renamed following PEP 8 (snake case), and finally, those lines which contain NaN values are dropped from the dataframe, resulting as follows:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **name** | **contact** | **date\_of\_birth** | **age** | **background** | **position** | **left\_right** | **category** |
| **Description** | Player’s name | Phone | Date of birth | Age | Previous experience | Position in the field | Left handed or right handed | Training with the first or U-23 team |
| **Type** | object | object | datetime64[ns] | float64 | object | object | object | object |
| **Count** | 15 | 15 | 15 | 15 | 15 | 15 | 15 | 15 |

**Results**

There may be different analysis that can be made in this dataframe, depending on the club’s needs, but the first variable considered is age, to determine if the players are training with the correct team. With this in mind, a histogram is presented with the information on players’ ages:

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

This could also be reflected in this plot, when adding the variable ‘category’ in the y axis:

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

This figure reflects that there are some players training with the U23 team while they should be training with the principal team, showing an administrative problem within the club that is not letting to see its actual potential.

Another analysis made be made by classifying if the players is left or right-handed. A scatter plot is shown as follows on this topic:

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

This means that only 2 out of 15 of the new players are left-handed, therefore the club should check if they need more left-handed players.

Finally, an analysis of the player position is presented. Typically, pivots are most searched and best paid in handball than the rest of the players. Therefore, for the club is important to know how many new pivots are training in its teams. After filtering by position, we have the following result:

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

And then after filtering by position = ‘pivot’ and age<23 there were no results, meaning that there are no new young pivots training in its teams.

**Conclusions**

The main conclusion of the analysis is to show how data analysis data may help this club to better understand the characteristics of new players who are recently training at the club, and if those players match with the club present and future needs. First, an administrative problem is shown as some players aged above 23 are training in the Under-23 team. Second, most of the new players are right-handed, meaning that it will be harder to cover left-handed positions. Finally, very few new players are pivot and none of them are aged below 23 years, which means that the club will need pivots in the future.

**Anecdotes**

An anecdote of the analysis is that in the original data, there is a column called “physical description” which describes each player in a colloquial way. This column was eliminated as nowadays, with the use of cell phones, this could be substituted by a picture of the player.

Finally, a new analysis proposed is to expand this to more clubs or other sports (football, basketball) in order to have a complete knowledge of handball players in the country, this being useful not only for clubs, but also to federations.