Cleaning Process in "FIFA21 Players" Dataset:

"FIFA 21 players" is a very messy and raw dataset of EA Sports' latest installment of their hit FIFA series - FIFA21, from sofifa.com.
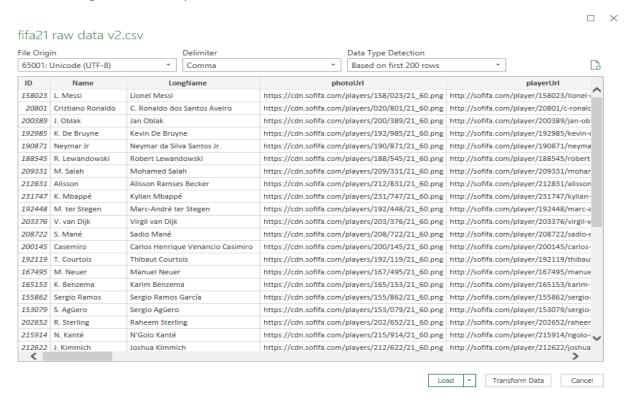
Inspiration

1. Convert the height and weight columns to numerical forms

2. 2. Remove unnecessary line breaks from all columns that contain them.

3. Use the "Joined" column to find out which players have been with the club for more than 10 years!

4. 'Value', 'Wage' and "Release Clause' are string columns. Convert them to numbers. For e.g., "M" in value column is Million, so multiply the row values by 1,000,000, etc.

5. Some columns have 'star' characters. Strip those columns of these stars and make the columns numerical

6. Which players are highly valuable but still underpaid (on low wages)?
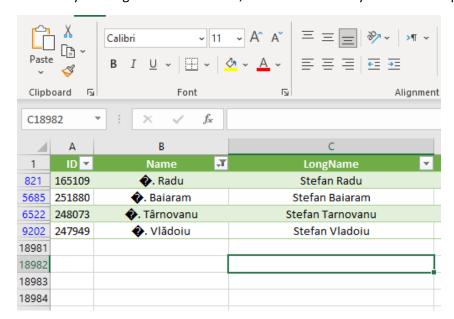
Here is a snapshot of the data:

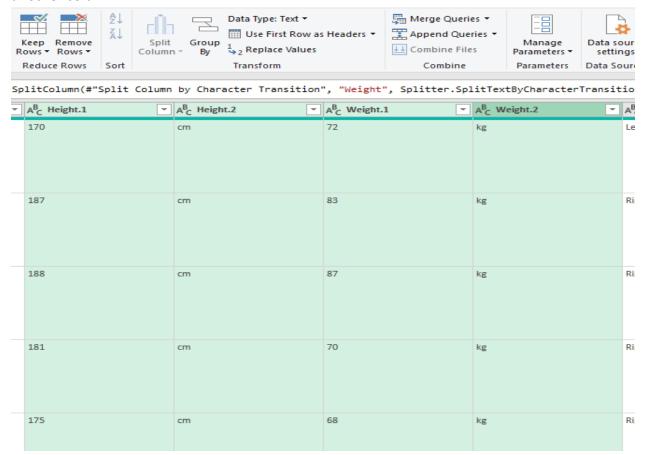| ID	Name	LongName	photoUrl	playerUrl
158023	L. Messi	Lionel Messi	https://cdn.sofifa.com/players/158/023/21_60.png	http://sofifa.com/player/158023/lionel-mes
20801	Cristiano Ronaldo	C. Ronaldo dos Santos Aveiro	https://cdn.sofifa.com/players/020/801/21_60.png	http://sofifa.com/player/20801/c-ronaldo-d
200389	J. Oblak	Jan Oblak	https://cdn.sofifa.com/players/200/389/21_60.png	http://sofifa.com/player/200389/jan-oblak/
192985	K. De Bruyne	Kevin De Bruyne	https://cdn.sofifa.com/players/192/985/21_60.png	http://sofifa.com/player/192985/kevin-de-b
190871	Neymar Jr	Neymar da Silva Santos Jr.	https://cdn.sofifa.com/players/190/871/21_60.png	http://sofifa.com/player/190871/neymar-da
188545	R. Lewandowski	Robert Lewandowski	https://cdn.sofifa.com/players/188/545/21_60.png	http://sofifa.com/player/188545/robert-lew
209331	M. Salah	Mohamed Salah	https://cdn.sofifa.com/players/209/331/21_60.png	http://sofifa.com/player/209331/mohamed
212831	Alisson	Alisson Ramses Becker	https://cdn.sofifa.com/players/212/831/21_60.png	http://sofifa.com/player/212831/alisson-rar
231747	K. MbappÃ©	Kylian MbappÃ©	https://cdn.sofifa.com/players/231/747/21_60.png	http://sofifa.com/player/231747/kylian-mba
192448	M. ter Stegen	Marc-AndrÃ© ter Stegen	https://cdn.sofifa.com/players/192/448/21_60.png	http://sofifa.com/player/192448/marc-andr
203376	V. van Dijk	Virgil van Dijk	https://cdn.sofifa.com/players/203/376/21_60.png	http://sofifa.com/player/203376/virgil-van-
208722	S. ManÃ©	Sadio ManÃ©	https://cdn.sofifa.com/players/208/722/21_60.png	http://sofifa.com/player/208722/sadio-mar
200145	Casemiro	Carlos Henrique Venancio Casimiro	https://cdn.sofifa.com/players/200/145/21_60.png	http://sofifa.com/player/200145/carlos-her
192119	T. Courtois	Thibaut Courtois	https://cdn.sofifa.com/players/192/119/21_60.png	http://sofifa.com/player/192119/thibaut-cc
167495	M. Neuer	Manuel Neuer	https://cdn.sofifa.com/players/167/495/21_60.png	http://sofifa.com/player/167495/manuel-ne
165153	K. Benzema	Karim Benzema	https://cdn.sofifa.com/players/165/153/21_60.png	http://sofifa.com/player/165153/karim-ben
155862	Sergio Ramos	Sergio Ramos GarcÃa	https://cdn.sofifa.com/players/155/862/21_60.png	http://sofifa.com/player/155862/sergio-ram
153079	S. AgÃ¼ero	Sergio AgÃ¼ero	https://cdn.sofifa.com/players/153/079/21_60.png	http://sofifa.com/player/153079/sergio-agu
202652	R. Sterling	Raheem Sterling	https://cdn.sofifa.com/players/202/652/21_60.png	http://sofifa.com/player/202652/raheem-st
215914	N. KantÃ©	N'Golo KantÃ©	https://cdn.sofifa.com/players/215/914/21_60.png	http://sofifa.com/player/215914/ngolo-kan

Data Cleaning Steps:

1. I used Excel Power Query to upload the CSV file and converted it using Unicode (UTF-8) file origin to handle special characters.

2. By filtering the Name column, I deleted the first symbol of these players.

3. In cell H1, Remove the column heading from "↓OVA" to "OVA".
4. I transformed the data into Power Query and used the "Replace Values" feature to remove unwanted newlines, tabs, and non-breaking spaces from all columns and change their types.
5. Changed the values in "OVA", "POT" and "BOV" into percentages by dividing its values by 100 and changing the column type into percent.
6. In "Height" and "Weight" columns I split these columns using split columns feature in Power Query to make their numerical values in a column and their measuring units in another column.
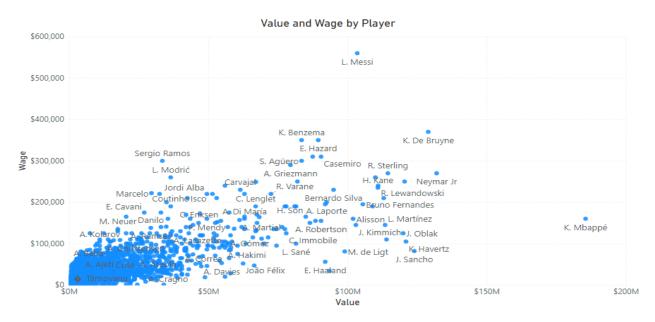

For Weight column I used "IF" function to convert lbm to kg with this formula "=IF([@[Weight.2]]="kg",VALUE([@[Weight.1]]),CONVERT(VALUE([@[Weight.1]]),"lbm","kg"))", then I have used a convertor website https://www.rapidtables.com/convert/length/inch-to-cm.html to convert inches to cm in Height column.
7. Using "Find & Select" feature on the "Home" tab to replace the "M" and "K" characters in the "Value", "Wage" and "Release Clause" columns to their values (M = 1,000,000 and K = 1,000) to make their values numerical.
8. Finally, I used "Find & Select" to remove the "★" in the "W/F", "SM" and "IR" columns.

Which players are highly valuable but still underpaid (with low wages)?

In the scatterplot below, we can see that Kylian Mbappé is the most valuable player at low wages (around $160,000 per week). There are many valuable players with low wages like Sancho, Havertz, Oblak, Arnold and Lautaro Martínez.

Which players have been with the same club for more than 10 years?

The table below shows players who have played for the same club for more than 10 years. It turns out that 40-year-old Japanese goalkeeper "Hitoshi Sogahata" was the most-played player on a team by 23 years.