**INTRODUCTION TO DATA MANGEMENT PROJECT REPORT**

(Project Semester August-December 2019)

***FIFIA 19***

Submitted by

**Yug Ahuja**

Registration No.- **11714531**

Programme**- B.Tech (CSE)**

Section- **KM094**

Course Code- **INT217**

Under the Guidance of

**Ms. Vasudha**

**Discipline of CSE/IT**

**Lovely School of Computer Science & Engineering**

**Lovely Professional University, Phagwara**

**CERTIFICATE**

This is to certify that Yug Ahuja bearing Registration no. 11714531 has completed INT217 project titled, **“FIFA 19”** under my guidance and supervision. To the best of my knowledge, the present work is the result of his original development, effort and study.

**Signature and Name of the Supervisor**

**Designation of the Supervisor**

**School of Computer Science & Engineering**

Lovely Professional University

Phagwara, Punjab.

Date: 20/11/19

**DECLARATION**

I Yug Ahuja, student of Computer Science & Engineering under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date:20th Nov,2019

Registration No.- 11714531 Name of the student Yug Ahuja

**ACKNOWLEDGEMENT**

I would like to express my special thanks of gratitude to my data management teacher “Ms Vasudha” for her able guidance and support in completing my project. Her suggestions and instructions have served as major contributor towards the successful completion of project.

Lastly, I would like to thank my classmates in resolving queries and problems.

**TABLE OF CONTENT**

1. [Introduction](#introduction) ……………………………………………………………………………………….6
2. [Objectives/Scope of the Analysis](#objective) …………………………………………………………………7
3. [Source of dataset](#source) …………………………………………………………………………………..8
4. [ETL process](#etl) ……………………………………………………………………………………….9-14

5. [Analysis on dataset (for each analysis)](#year) …………………………………………………………….15-25

1. Introduction
2. Analysis results
3. Visualization
4. Results

6. [Bibliography](#biblo) ………………………………………………………………………………………26

[**I****NTRODUCTION**](#toc)

There were times when we were required to analyse large amounts of data and produce easy to read and understand reports. That used to be very hectic work at that times, but pivot tables have made our work easy to maximum extent. Pivot tables allow us to analyse such data and produce reports that meet our business reporting requirements.

I have worked on a project titled as “FIFA 19 player dataset” using pivot table and charts and has made various analysis from a very large data of around 13 Thousand players who were part of football in 2019. A pivot table is a data summarization tool that is used in context of data processing. Pivot table are used to summarize, sort, reorganize, group, count, total or average data stored in a database. It allows its users to transform columns into rows and rows into columns.

Moreover, pivot charts helps to represent the result of analysis in a very catchy manner no matter how large the data just like the FIFA dataset for year 2019 was too big to analyse. A pivot table helps you to extract the significance from a large, detailed data set. It provides n number of features like sorting, filtering etc.In the project analysis like nationality, football club, age, player ID, skills category wise and many more very almost ease with the help of pivot table and charts. Data Analysis is a process of inspecting, transforming and modelling data with goal of discovering useful information, information conclusions and supporting decision making. It has multiple assets and approaches, encompassing diverse techniques under a variety of names, while being used in different business, science and social science domains.

[**OBJECTIVES/SCOPE**](#toc)

**Here are the objectives to be completed via this dashboard**

* Player Profile
* Player Statistics
* Player Id
* Player Skills
* Player Gameplay
* Player IR rating

[**SOURCE OF DATASET**](#toc)

[[https://www.kaggle.com/karangadiya/fifa19](https://www.kaggle.com)](https://www.kaggle.com/karangadiya/fifa19)

This is the source of my data set “FIFA 19”.it has n number of datasets of all fields, sizes, places, departments etc. It is an amazing source of datasets. While other sources of reference were also there like data. World, Wikipedia. **KAGGLE** is an online community of data scientist and machine learner, owned by [Google](https://en.wikipedia.org/wiki/Google). Kaggle allows users to find and publish data sets, explore and build models in a web-based data-science environment, work with other data scientists and machine learning engineers, and enter competitions to solve data science challenges.

[**ETL PROCESS**](#toc)

**ETL** is defined as a **process** that extracts the data from different RDBMS source systems, then transforms the data (like applying calculations, concatenations, etc.) and finally loads the data into the Data Warehouse system. **ETL** full-form is Extract, Transform and Load.

In computing, extract, transform, load is the general procedure of copying data from one or more sources into a destination system which represents the data differently from the source or in a different context than the source.

In this project also ETL Process is used using tableau prep for extraction like from csv to excel then for cleaning purpose like (removal of null values or negative values, redundancy, spelling error, grouping etc). then it transformed data is loaded to for further analysis. this is how etl process plays a very important role data analytics. This is how ten analysis are done on different basis.

1. **Extraction:**  
   The first step of the ETL process is extraction. In this step, data from various source systems is extracted which can be in various formats like relational databases, No SQL, XML and flat files into the staging area. It is important to extract the data from various source systems and store it into the staging area first and not directly into the data warehouse because the extracted data is in various formats and can be corrupted also. Hence loading it directly into the data warehouse may damage it and rollback will be much more difficult. Therefore, this is one of the most important steps of ETL process.
2. **Transformation:**  
   The second step of the ETL process is transformation. In this step, a set of rules or functions are applied on the extracted data to convert it into a single standard format. It may involve following processes/tasks:
   * Filtering – loading only certain attributes into the data warehouse.
   * Cleaning – filling up the NULL values with some default values, mapping U.S.A, United States and America into USA, etc.
   * Joining – joining multiple attributes into one.
   * Splitting – splitting a single attribute into multiple attributes.
   * Sorting – sorting tuples on the basis of some attribute (generally key-attribute).
3. **Loading:**  
   The third and final step of the ETL process is loading. In this step, the transformed data is finally loaded into the data warehouse. Sometimes the data is updated by loading into the data warehouse very frequently and sometimes it is done after longer but regular intervals. The rate and period of loading solely depends on the requirements and varies from system to system.

**For the FIFA 19 dataset ETL process works as follows:**

**Step1.** Extraction

Data set when extracted from Kaggle which was in csv format.

It had null values

It had spelling mistakes

It had unnecessary columns

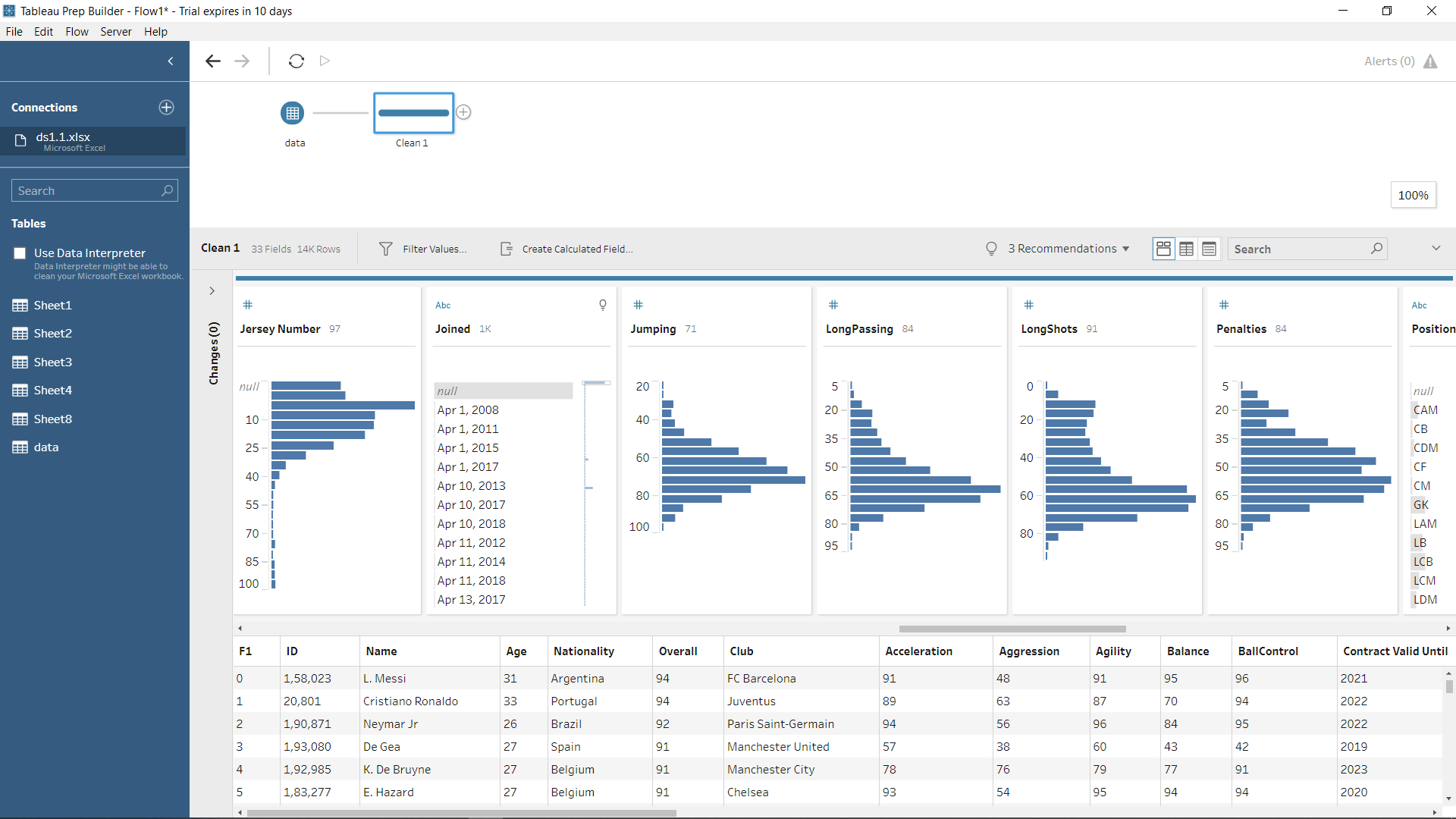
So first the file was extracted

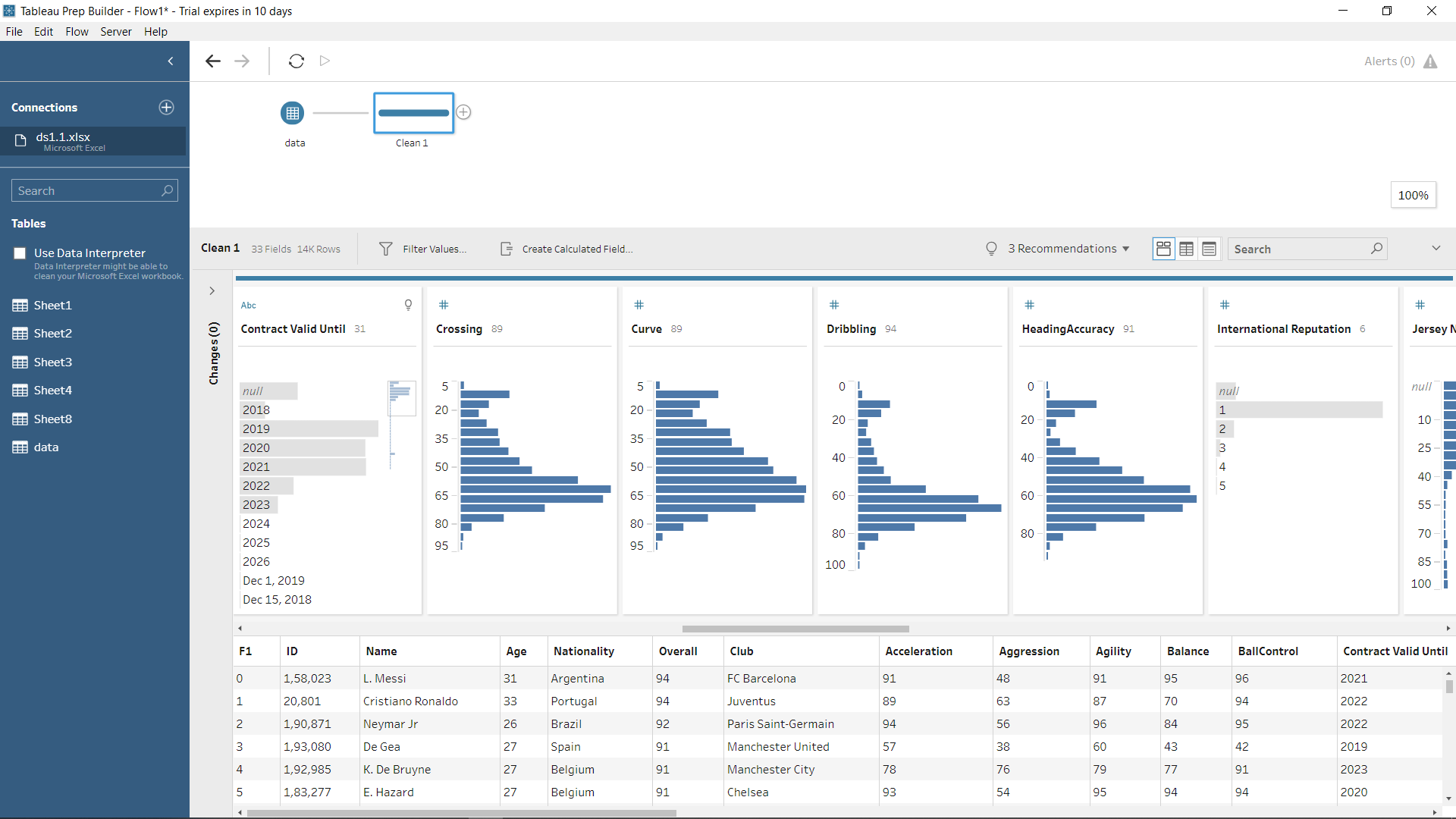
To make certain modifications

To clean the data

To analyse the results

This is how it looked before cleaning





Step 2. Transformation, cleaning has to be done after it is extracted from the source.

Here 3 cleaning is done

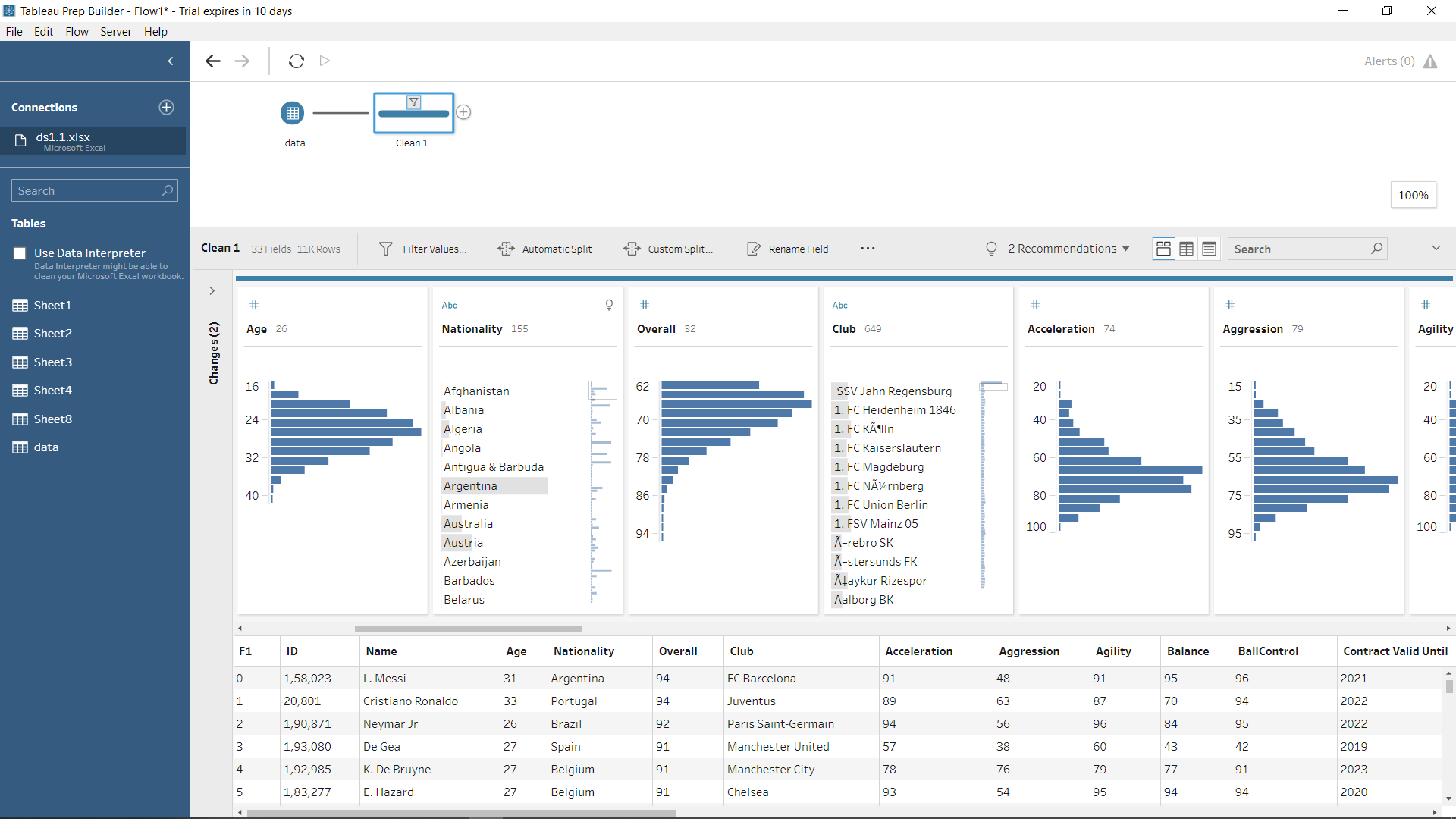
In first it removed unwanted columns

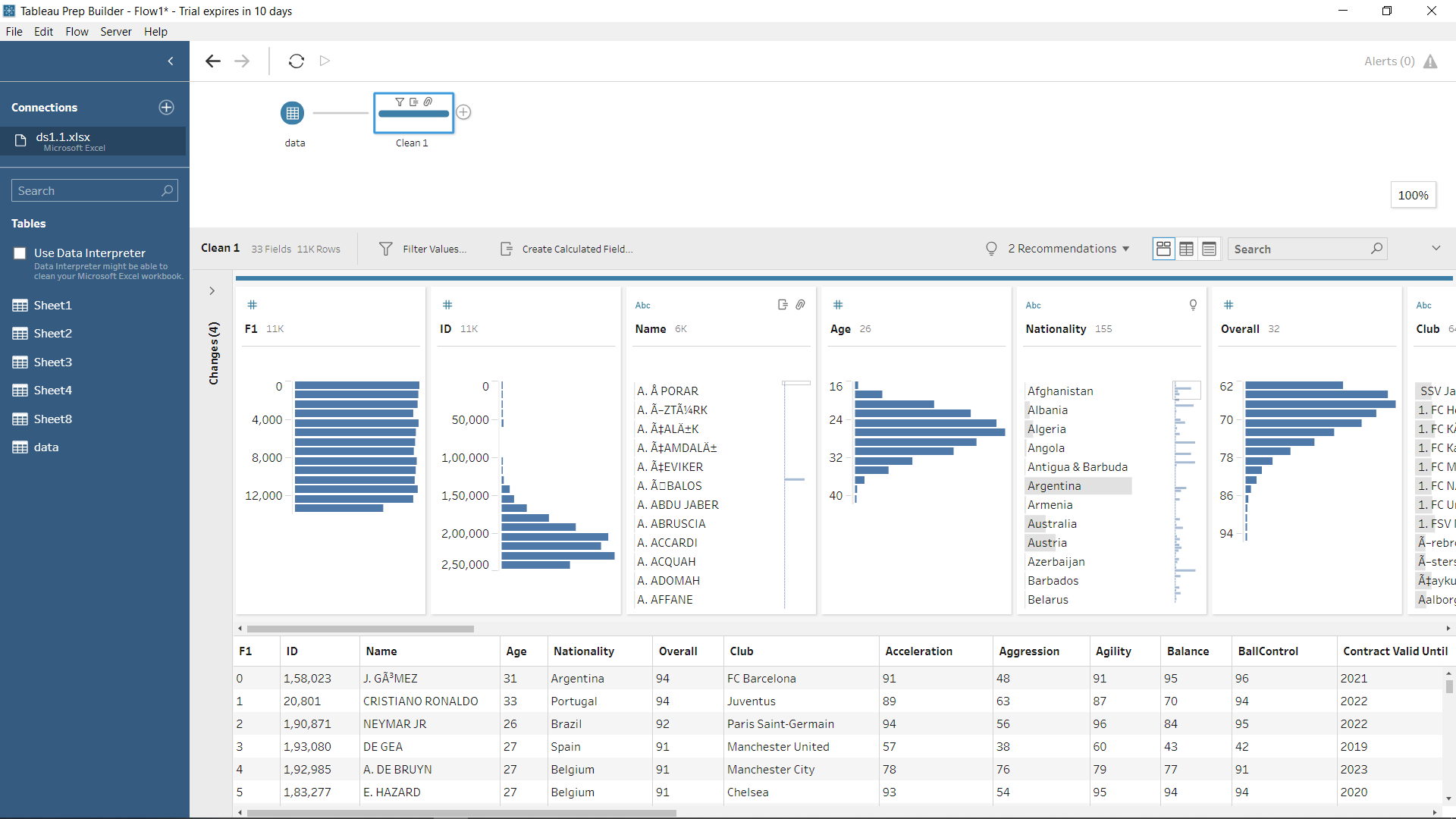
In second it removed null values and entries

In third it corrected spelling mistakes.

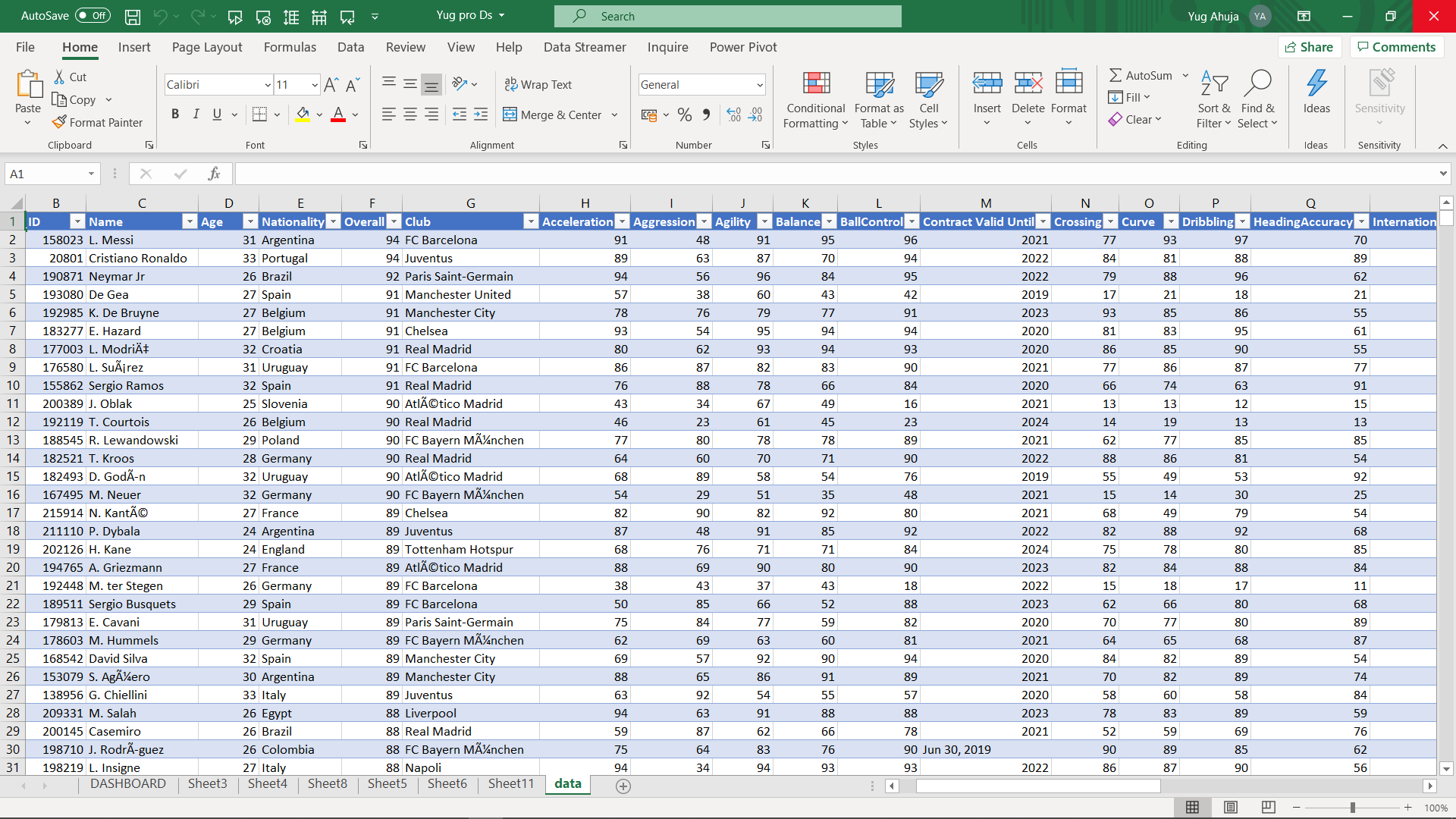
Removal of the unnecessary column (location) as it has no use in analysis. For this we just selected the entire column and pressed delete button from keyboard hence deleted the unwanted column from the table.

2. Removal of null values which is very important part of cleaning.





After all the cleaning and modifications are done dataset looks like:

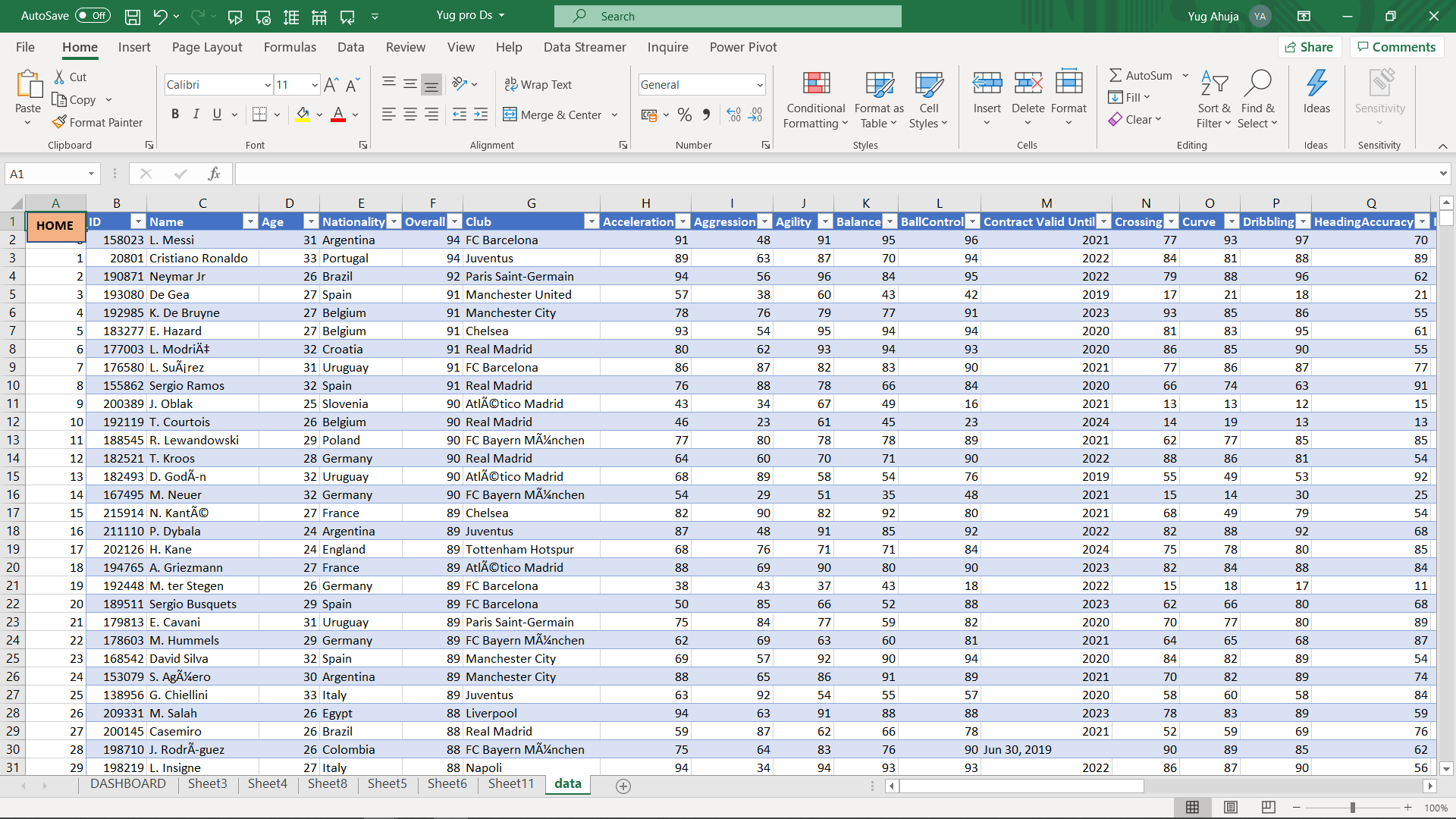


**Step 3.** After transformation data is loaded and is worked further on analysis

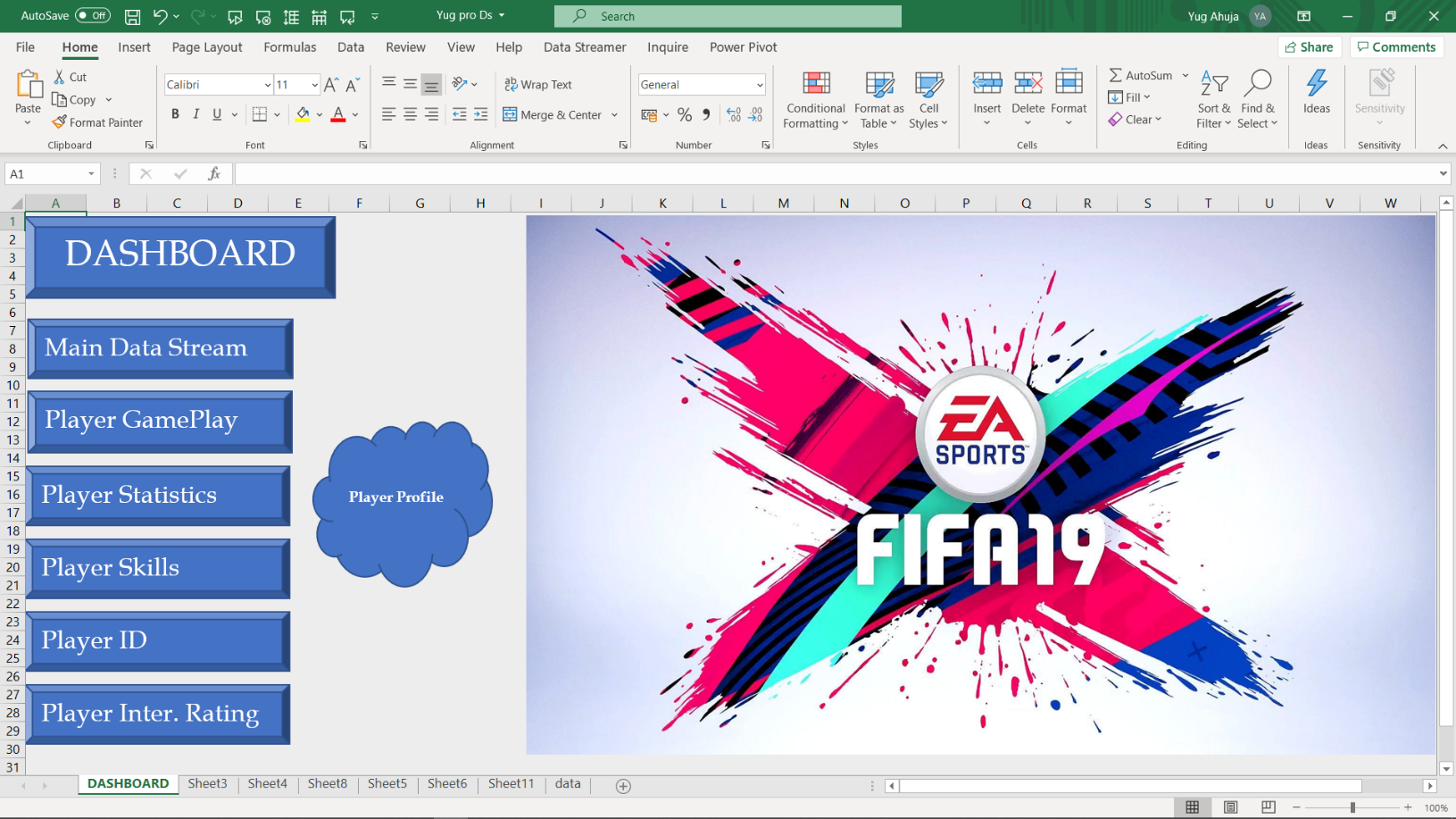
Here further few analysis are done using pivot table and charts in most effective manner.

* Player Profile
* Player Statistics
* Player Id
* Player Skills
* Player Gameplay
* Player IR rating

**Snapshot of final cleaned Main data stream.**



**Snapshot of DASHBOARD**



1. **PLAYER PROFILE**

**Introduction**

Under this sheet information about the nationality, football club, age, jersey number and joining date of the particular players are analysed using pivot and slicers.

**Specific Requirements, functions and formulas**

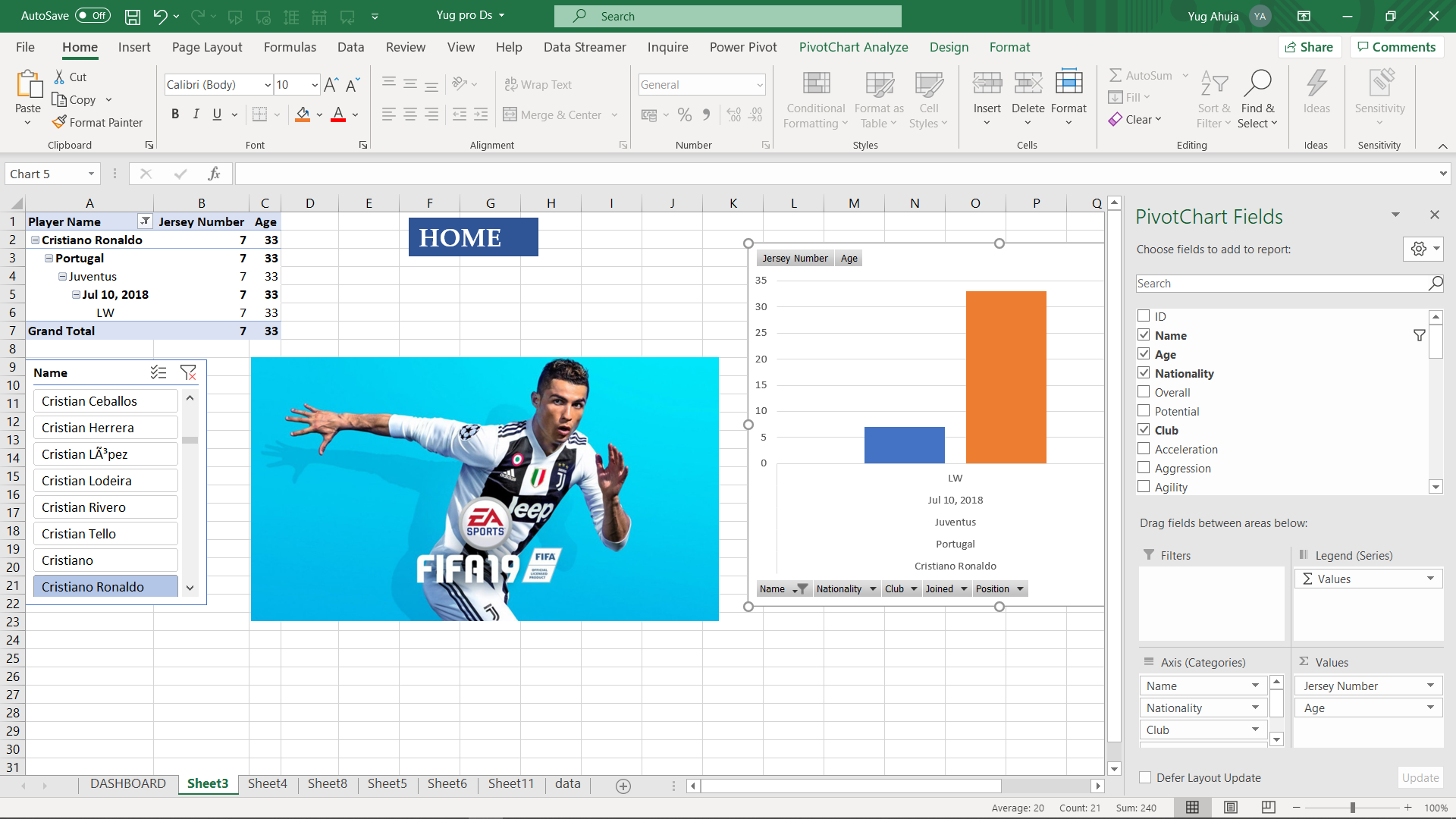
No such specific requirements, functions or formulae is used as in dataset.

**Analysis results**

|  |  |  |
| --- | --- | --- |
| **Player Name** | **Jersey Number** | **Age** |
| **Cristiano Ronaldo** | **7** | **33** |
| **Portugal** | **7** | **33** |
| Juventus | 7 | 33 |
| **Jul 10, 2018** | **7** | **33** |
| LW | 7 | 33 |
| **Grand Total** | **7** | **33** |
|  |  |  |

**Pivoting is used to analyse the data and using slicers we are able to locate information for each player.**

**Visualization**



1. **PLAYER STATISTICS**

**Introduction**

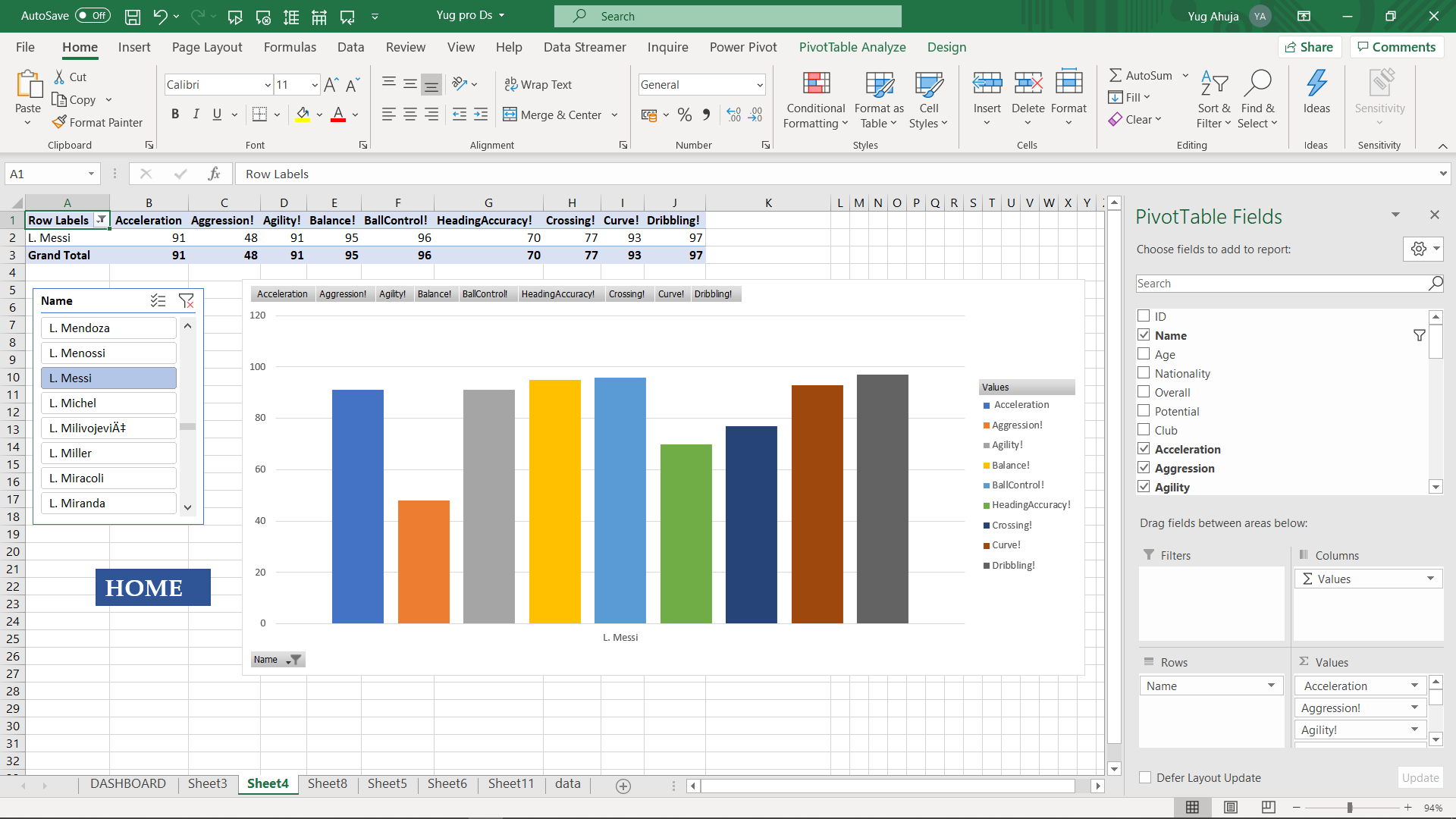
Under this all the players were analysed according to their performance in the field.

Eg:- Acceleration, Aggression, Agility, Balance, Ball Control, Heading Accuracy, Crossing, Curve, Dribbling.

**Analysis results**

The performance of the players was analysed with ease. In above chart analysis of player Messi is performed using slicer and pivot charts.

**Visualization**



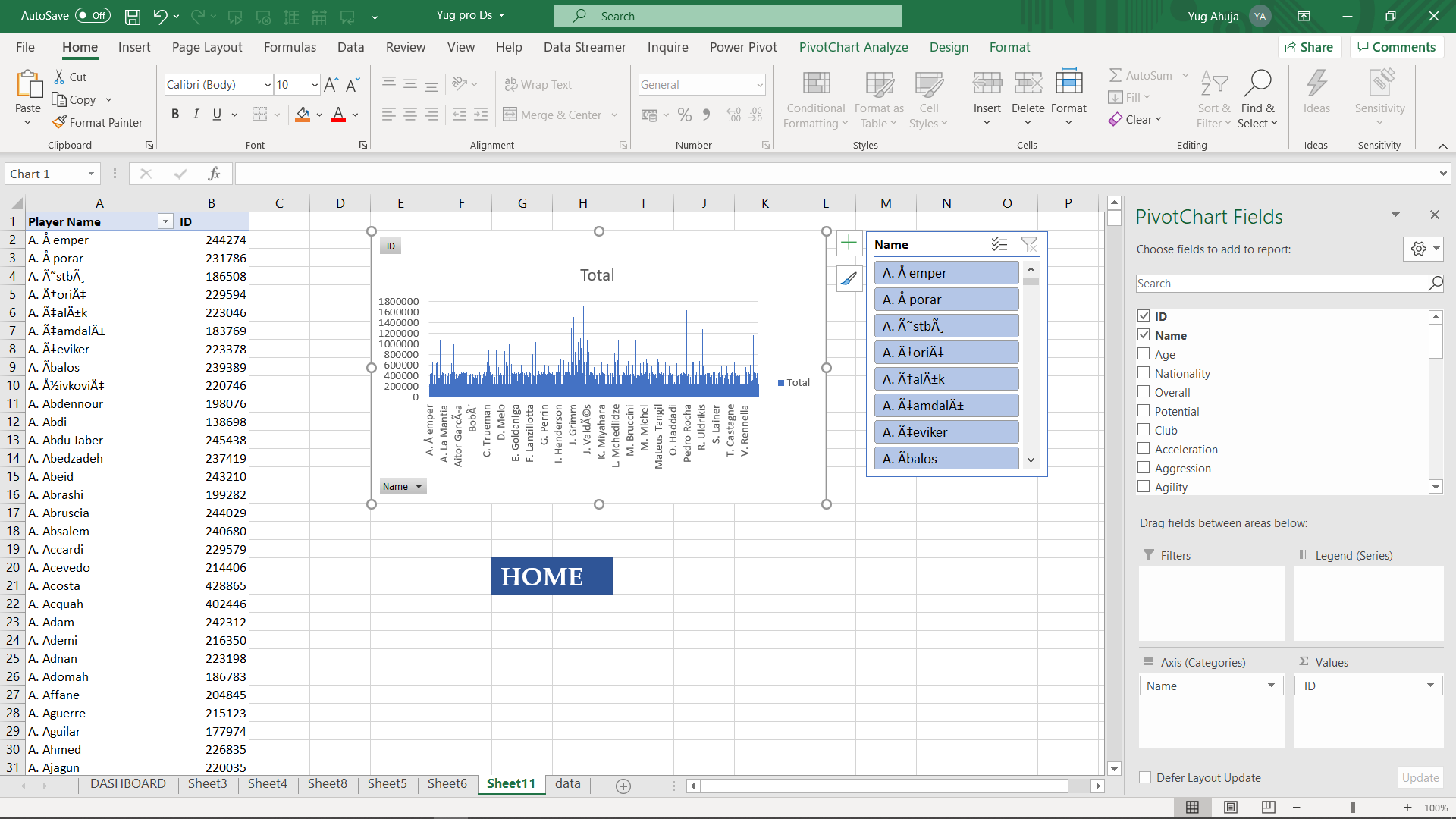
1. [**PLAYER**](#toc) **ID**

**Introduction**

Under this analysis each player was provided with unique player ID in order to differentiate between the rest of the players.

**Analysis results**

**Visualization**



1. **PLAYER SKILLS**

**Introduction**

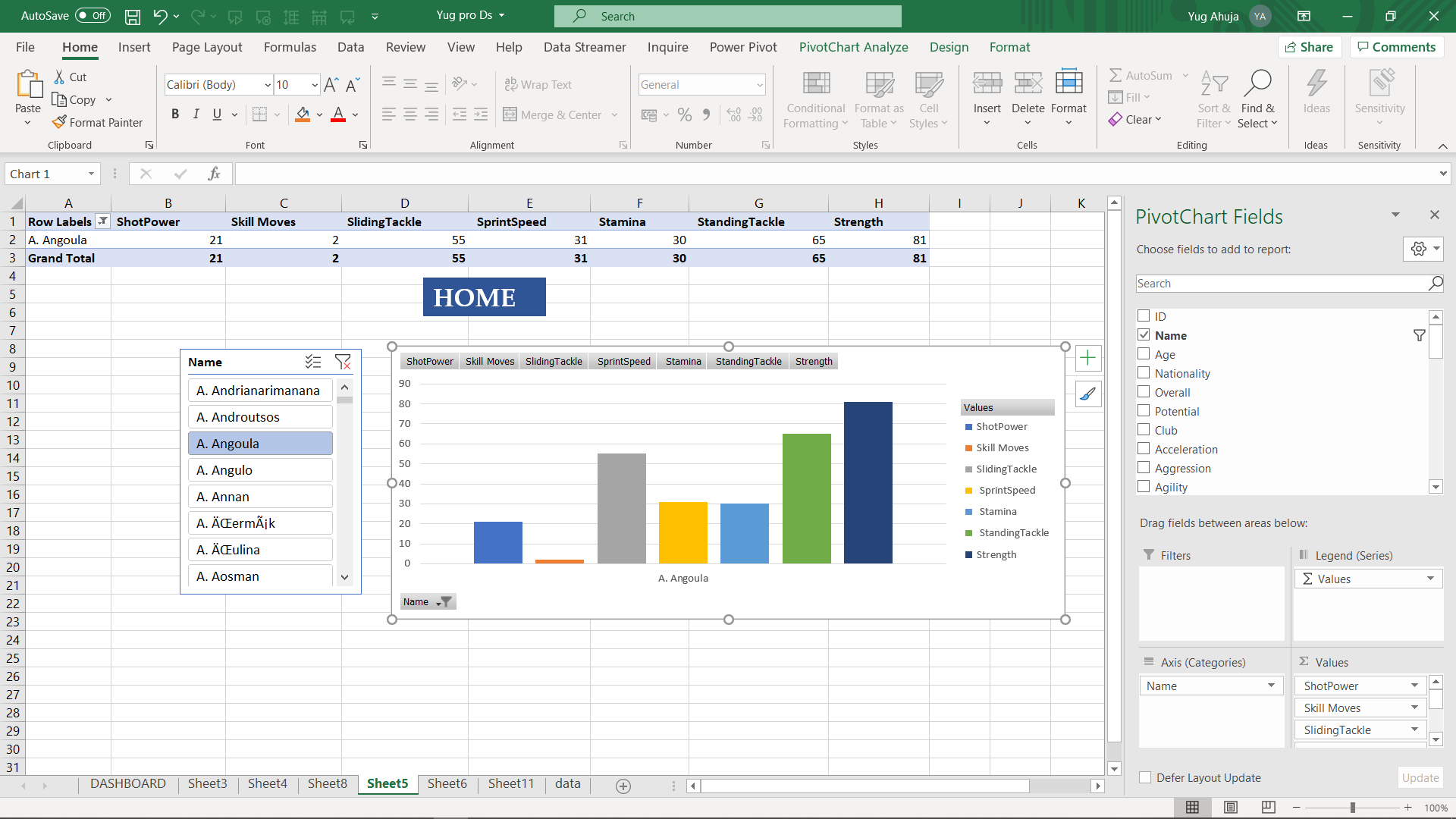
The players were analysed on the basis of their skill level.

**Analysis results**

The players were analysed on the basis of the following factors:

* Shot power
* Skill moves
* Sliding tackle
* Sprint speed
* Stamina
* Standing tackle
* Strength

**Visualization**



1. **PLAYER GAMEPLAY**

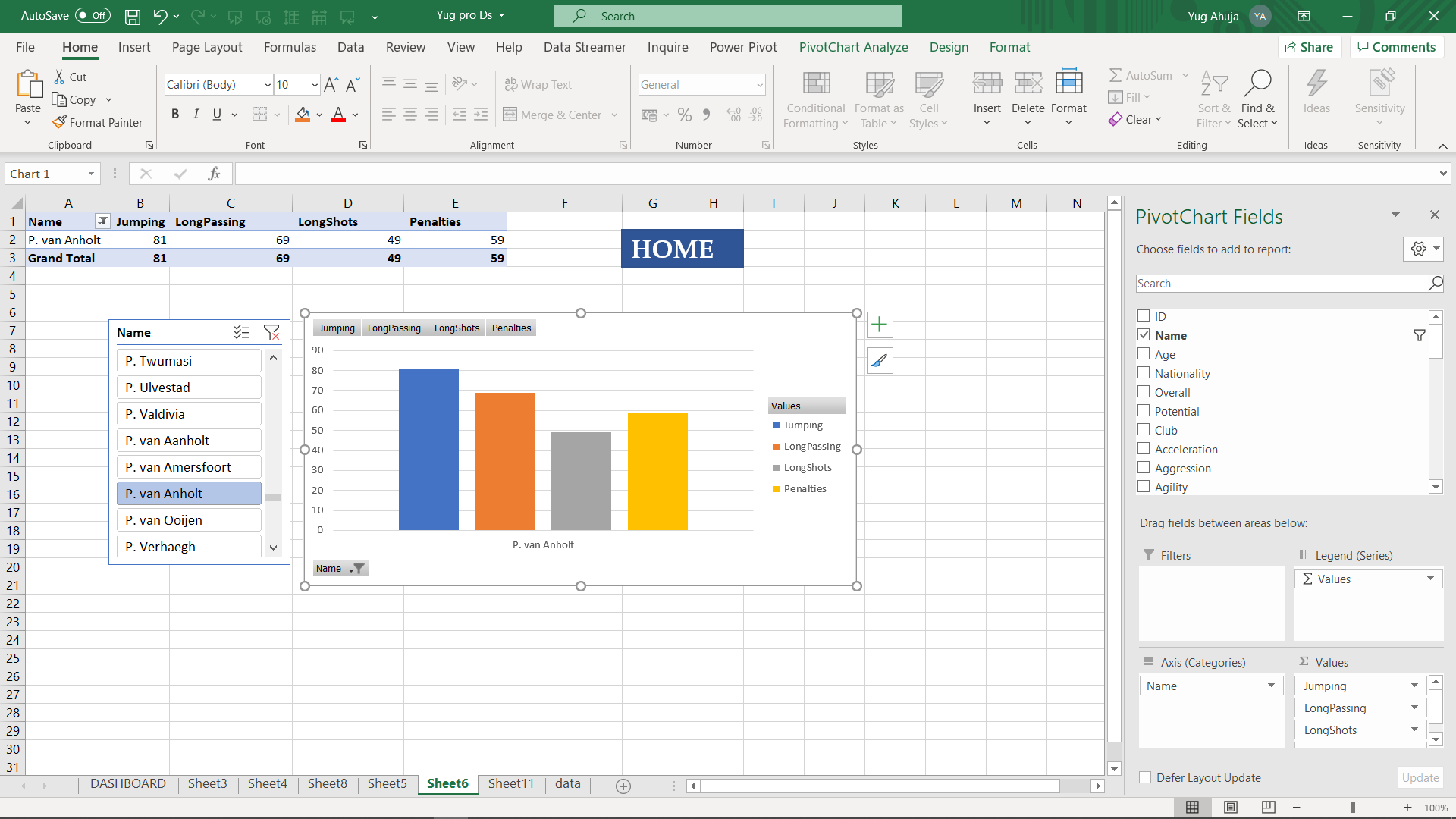
**Introduction**

Under this the entire gameplay of the players was analysed based on the following factors:

* Jumping
* Long passing
* Penalties
* Long shots

**Analysis results**

**Visualization**



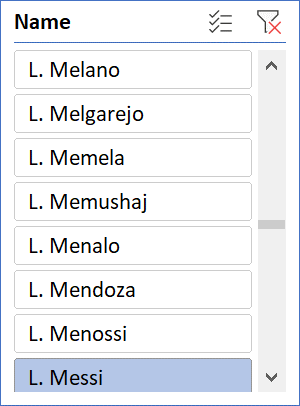
1. **[PLAYER](#toc) IR RATING**

**Introduction**

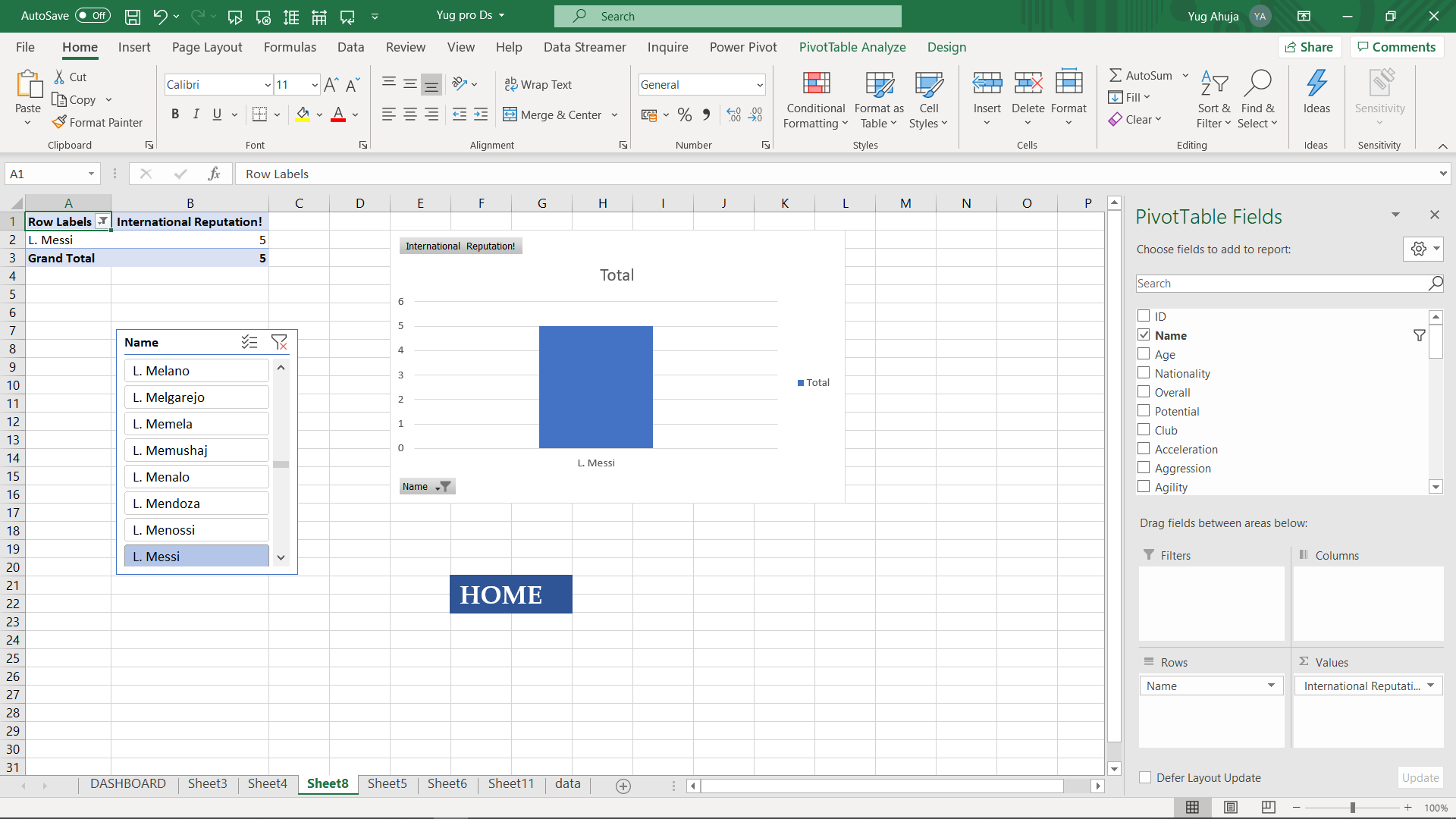
All the players in the data set were rated internationally on the scale of 1-5 on this basis of their performance and personality and their social life.

**Analysis results**

|  |  |
| --- | --- |
| **Row Labels** | **International Reputation!** |
| L. Messi | 5 |
| **Grand Total** | **5** |

****

**Visualization**



**BENEFITS OF PIVOT**

A **pivot table** is a data summarization tool that is **used** in the context of data processing. **Pivot tables** are **used** to summarize, sort, reorganize, group, count, total or average data stored in a database. It allows its users to transform columns into rows and rows into columns. a **Pivot Table** is a tool built into Excel that allows you to summarize large quantities of data quickly and easily. Given an input **table** with tens, hundreds, or even thousands of rows, **Pivot Tables** allow you to extract answers to a series of basic questions about your data with minimal effort. Use a **pivot table** to build a list of unique values. Because **pivot tables** summarize data, they can be **used** to find unique values in a field. This is a good way to quickly see all the values that appear in a field and also find typos, and other inconsistencies.  pivot table is a tool in data summation that is common in many business software. It is utilized to collect the summary of a specific data set in a compressed technique. It is a very useful tool in consolidating a large quantity of data that is contained in Microsoft Excel. They let the user make a faster organization and drawing of conclusions from data being collected. Pivot table consists of columns, rows, pages, and data fields. These can be moved around and it assists in expanding, isolating, summarizing, and grouping the specific data. And all of these can be accomplished in real time.

**Pivot Table Slicers** are a visual filter in the form of an interactive button. There are several cool things that you can do with **Pivot Table Slicers**, like customize them, filter them, connect them to multiple **Pivot Tables** plus much more!

It can create instant data

It makes data analyses easier

It summarises data easily

It helps to give quick overview

It is very fast and easy to work with.

**LEARNING OUTCOMES**

I have learnt a lot from this dashboard project using excel. Outcomes are as follows:

1. Learnt to use pivot technique
2. Learnt to analyse data efficiently which is the main task of data scientists.
3. Pivot charts are very helpful in data representation
4. Also, learnt to manage large datasets.
5. Learnt to make dynamic dashboards.
6. Slicers use make task so much easier.
7. Create, modify, and format PivotTables

[**BIBLOGRAPHY**](#toc)

[**www.kaggle.com**](http://www.kaggle.com)

[**www.data.world.com**](http://www.data.world.com)

**www.wikipedia.com**