

### Introduction to Data Management PROJECT REPORT

(Project Semester August-December 2020)

**PROJECT REPORT ON**

## Analyzing Top Chess Players till August 2020

Submitted by

### Shiv Chandra

### 11904848

Programme: Bachelor of Technology Section: KM007

Course Code: INT217

Under the Guidance of **Sandeep Kaur: 23614**

**Assistant Professor**

### Discipline of CSE/IT

**Lovely School of Computer Science & Engineering Lovely Professional University, Phagwara**

# DECLARATION

I, Shiv Chandra, 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: 13/12/2021

**Shiv Chandra**

**Registration No: 11904848**

**Signature:**

# ACKNOWLEDGEMENT

Primarily I'd thank God for being able to complete my project with success. Then I'd like to thank my mentor **Ms. Sandeep Kaur**, whose valuable guidance has been the ones that helped me patch this project and make it full proof success in contribution towards the completion of this project.

Last but not least I'd rather thanks to **Lovely Professional University,** and my parent’s inspiration, who gave me this golden opportunity to learn many new things, to learn another aspect of life.

Shiv Chandra

# CONTENTS:

|  |  |  |
| --- | --- | --- |
| **Sr No.** | **Title** | **Page No.** |
| 1 | Introduction | 5 |
| 2 | Objectives/Scope of the Analysis | 6 |
| 3 | Source of dataset | 7 |
| 4 | ETL Process | 9 |
| 5 | Analysis of dataset | 18 |
| 6 | List of Analysis with results | 29 |
| 7 | Final Dashboard | 34 |
| 8 | Bibliography | 35 |

**INTRODUCTION**

* Data management is important because the data your organization creates is a very valuable resource.
* The last thing you want to do is spend time and resources collecting data and business intelligence, only to lose or misplace that information.
* In that case, you would then have to spend time and resources again to get that same business intelligence you already had.
* And on that data analysis is carried out which show visualization of our problems in efficient way.
* Data Analysis is a process of inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information, informing conclusions, and supporting decision- making.
* This project is based on such data analysis about Top Chess Players till August 2020.
* This dataset contains all old games played in all formats like Standard, Blitz and Rapid. Moreover, it also includes player name, year of birth, federation and fide Id.
* This IMDb dataset contains 10 data fields.

# OBJECTIVES/SCOPE OF ANALYSIS

After analysis of the dataset, the aim of this project is to give answer of given objectives in

an easy way:

* What are the top 5 female players according to standard rating?
* What are the top male players who was born after 1990 according to Blitz rating?
* What are top the players who has GM norm title and also from Russia federation?
* Name the top player who has standard rating more than both of Rapid and Blitz rating?
* What are the top federations who has IM norm title according to standard rating?

# SOURCE OF DATASET

### Source of dataset:

https://www.kaggle.com/vikasojha98/world-top-chess-players-august-2020

The dataset is based on World Top Chess Players till August 2020. The columns included in the dataset are given below:

* Fide Id

 The Id of Players

* Name

 The Name of the Players

* Federation

 The Country of the Players

* Gender

 Male or Female Players

* Year of Birth

 Birth year of the Player

* Standard Rating

 The Classical Chess Rating of Players

* Rapid Rating

 The Rapid Ratings of Players

* Blitz Rating

 Blitz Format Rating of Players

* Inactive Flag

 Completely Missing

* Title

 Highest title achieved by Players

**Sample of dataset with data fields is given below:**

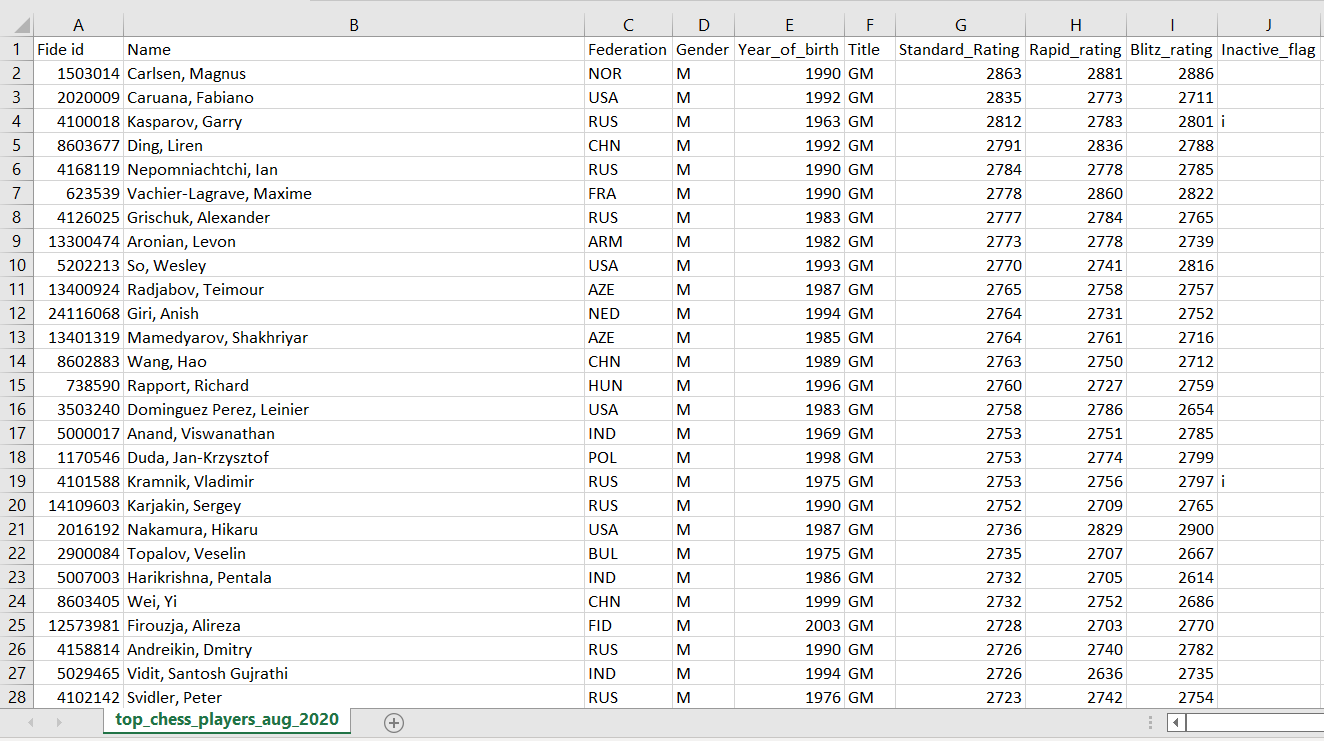
****

Fig.1 Original Dataset

# ETL PROCESS

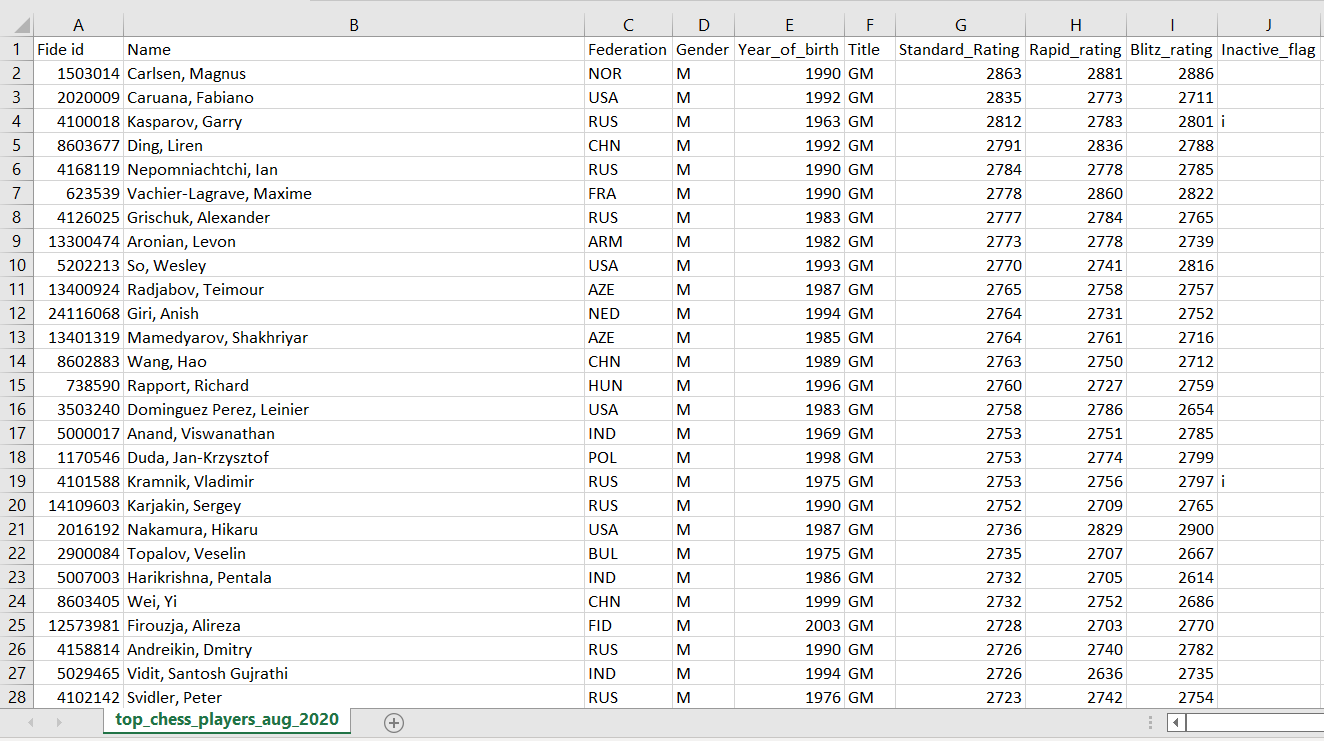
* **ETL** is a process that extracts the data from different source systems, then transforms the data (like applying calculations, concatenations, etc.) and finally loads the data into the Data Warehouse system.
* Full form of ETL is Extract, Transform and Load.
* The triple combination of ETL provides crucial functions that are many times combined into a single application or suite of tools that help in the following areas:
  + Enhances Business Intelligence solutions for decision making.
  + Allows verification of data transformation, aggregation and calculations rules.
  + Allows sample data comparison between source and target system.
  + Helps to improve productivity as it codifies and reuses without additional technical skills.
* ****Initially, the raw dataset was arranged as shown in given picture:

Fig.1.2 Initial Dataset

**Initially Missing Values In Dataset**

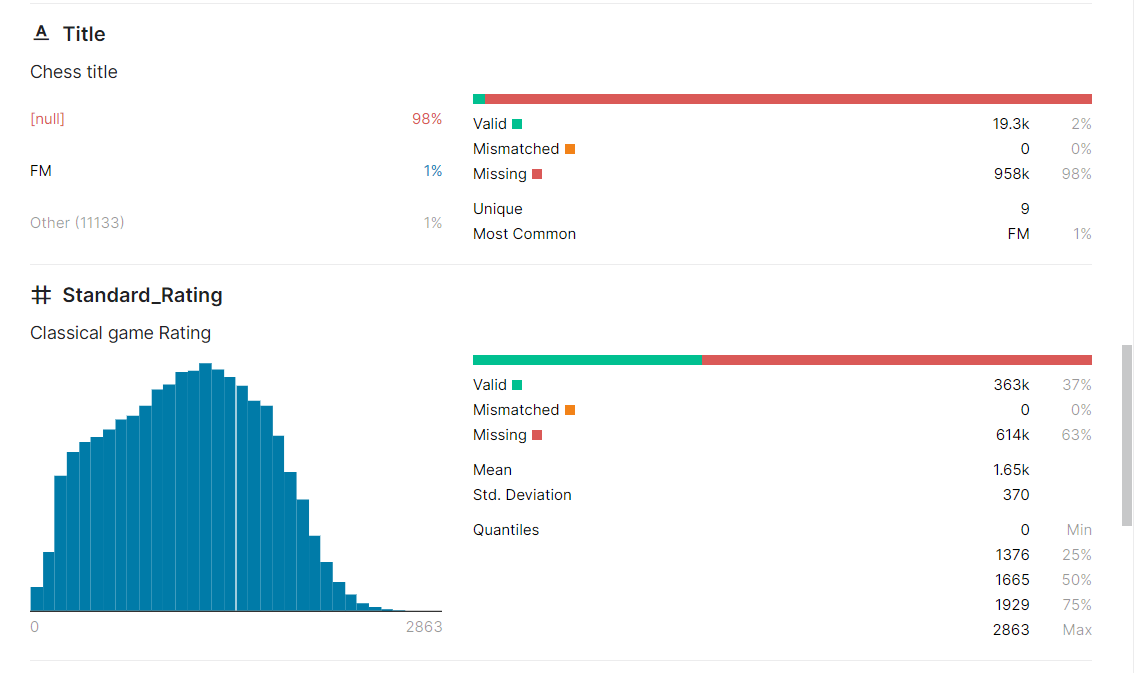
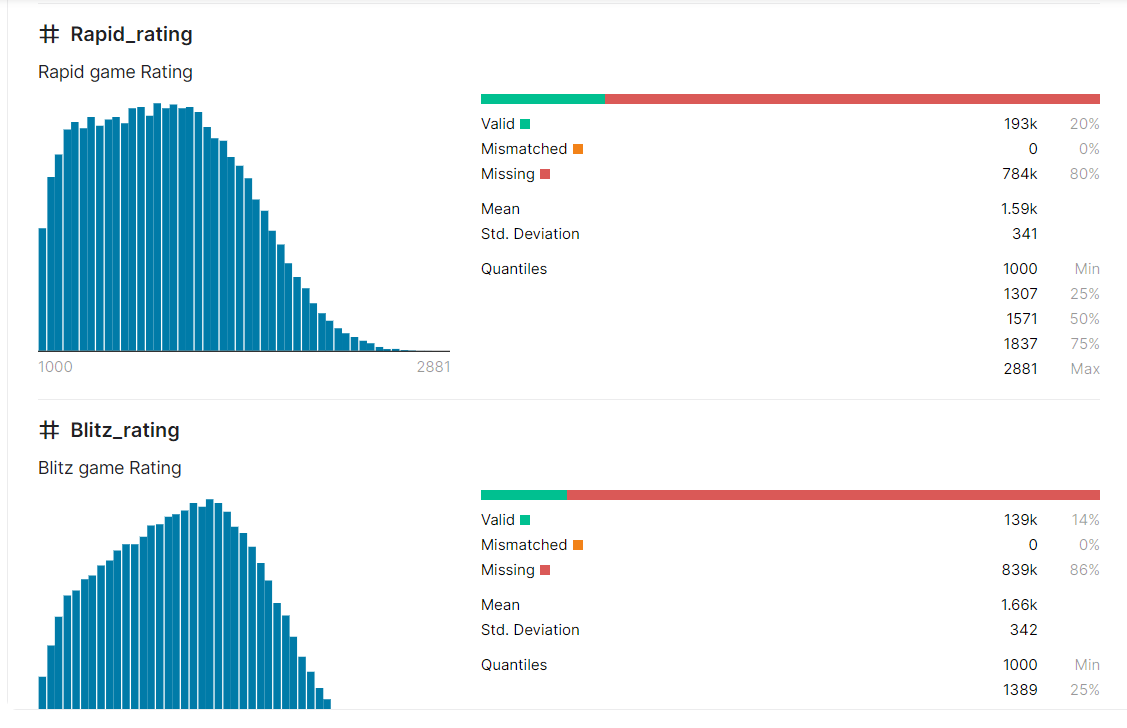
Each columns missing values shown below :-

Fig.1.3. Missing Values

**Disorganized Name Column**

Name column had last name first before first name of each player.

You can see in below picture: -



Fig.1.4 Disorganized Column

It is Magnus Carlsen **Not** Carlen Magnus and each column

Also contain comma ‘,’ as a separator.

**Tableau Prep Builder**

Tableau Prep Builder is a tool in the Tableau product suite designed to make preparing your data easy and intuitive. Use Tableau Prep Builder to combine, shape, and clean your data for analysis in Tableau.

**Using Tableau Prep Builder**

Start by connecting to your data from a variety of files, servers, or Tableau extracts. Connect to and combine data from multiple data sources. Drag and drop or double-click to bring your tables into the flow pane, and then add flow steps where you can then use familiar operations such as filter, split, rename, pivot, join, union and more to clean and shape your data.

Each step in the process is represented visually in a flow chart that you create and control. Tableau Prep tracks each operation so that you can check your work and make changes at any point in the flow.

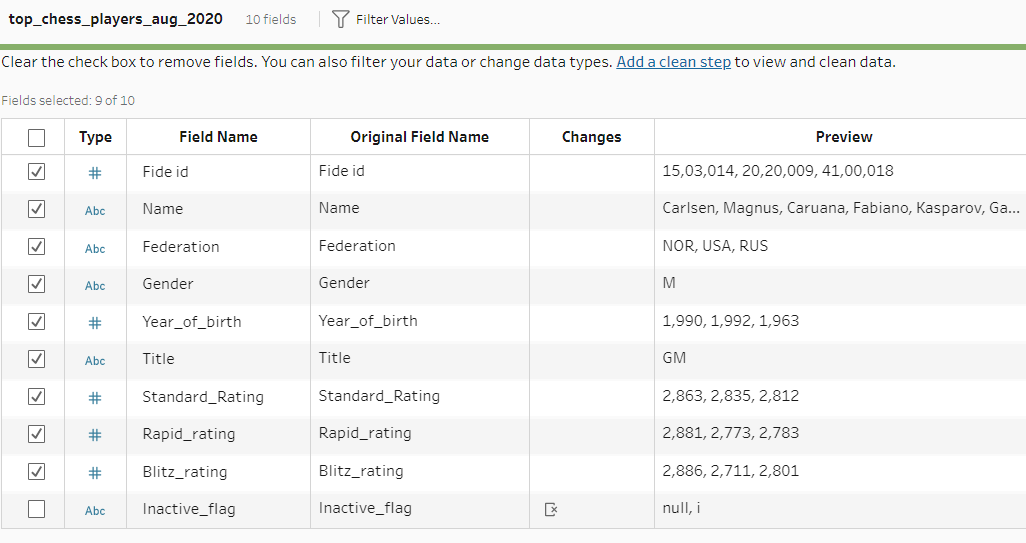
When you are finished with your flow, run it to apply the operations to the entire data set.

Tableau Prep works seamlessly with other Tableau products. At any point in your flow, you can create an extract of your data, publish your data source to Tableau Server or Tableau Online, publish your flow to Tableau Server or Tableau Online to continue editing on the web or refresh your data using a schedule. You can also open Tableau Desktop directly from within Tableau Prep Builder to preview your data.

## Steps taken to clean dataset through Tableau Prep Builder

### Step 1:

* Firstly, connect dataset to tableau prep and uncheck the inactive flag column. This will remove inactive flag column from dataset because this column contains almost missing data.



**Fig.2.1 Connect Dataset**

### Step 2:

* Now click on step to Clean. This will show summary of each column like how many columns contain null values.

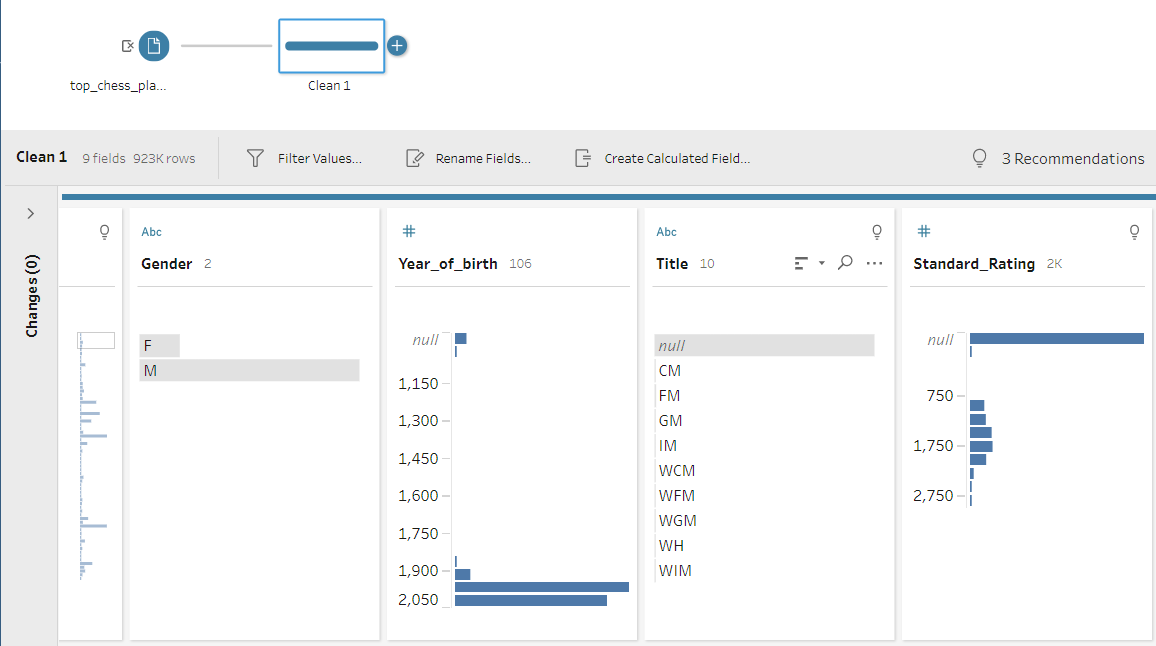


Fig.2.2. Summary of Dataset

### Step 3:

* Now, Right Click on null values and select exclude. This will remove the null values from the dataset. After remove you can see the changes.

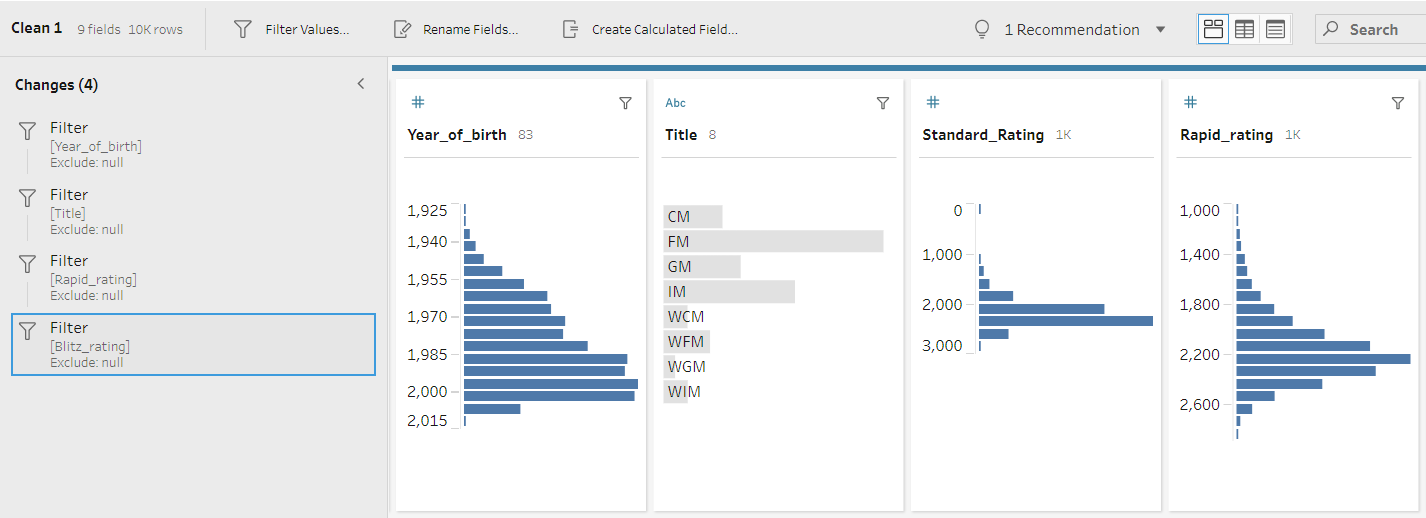


Fig.2.3 Remove Null Values

### Step 4:

* Now, for fixing the name column. Click on three dots in name column and select automatic split option from the split option. This will automatically separate from comma “,”. Now, we have first and last name in two different columns.

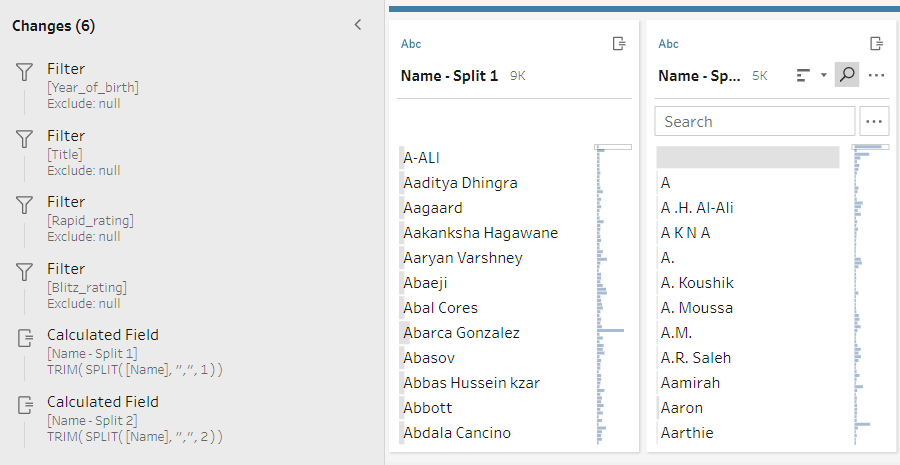


Fig.2.4 Split Name Column

### Step 5:

* Now, Merge the both split columns using calculated field option.
* Use formula: - [Name - Split 2]+" "+[Name - Split 1]
* This will create a new calculated field as we required.

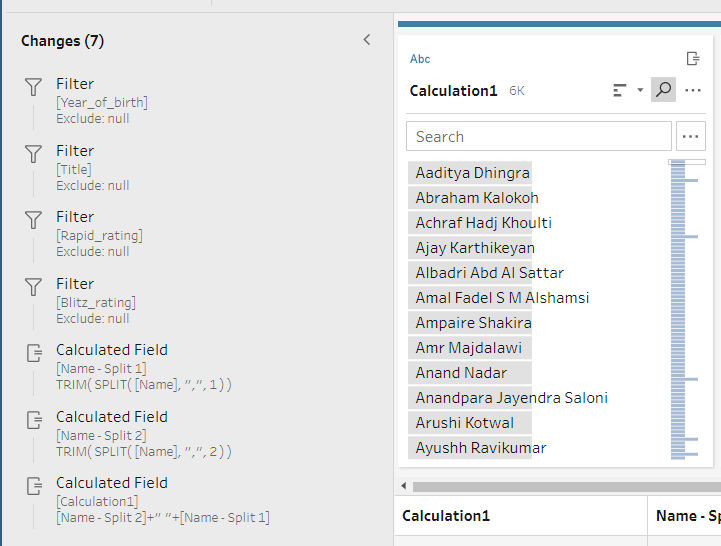


Fig. 2.5 Creating Name column

### Step 6:

* Finally, Remove both splited columns and the original name column also.
* Now, Rename calculation1 as Name

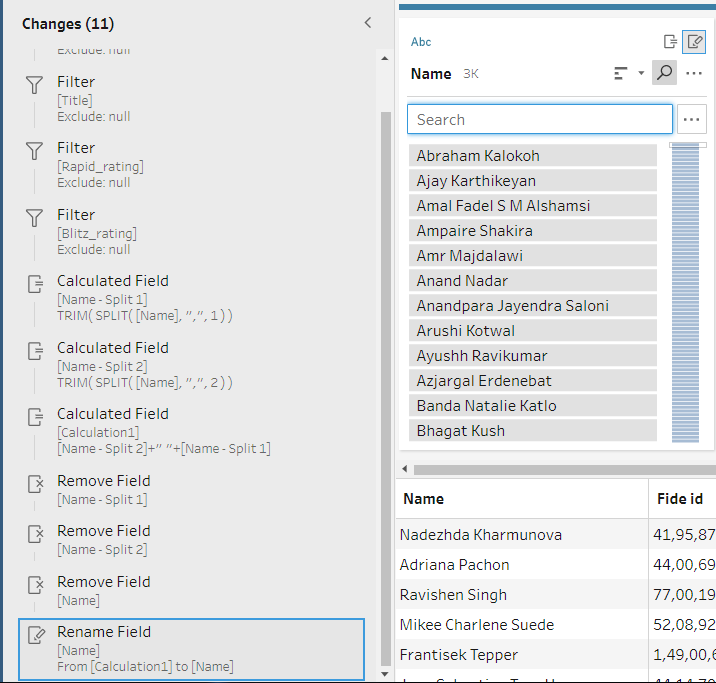


Fig.2.6 Removing Unnecessary Fields

### Step 7:

* Finally, generate new output field as we required and run the flow to make the changes save in new file.

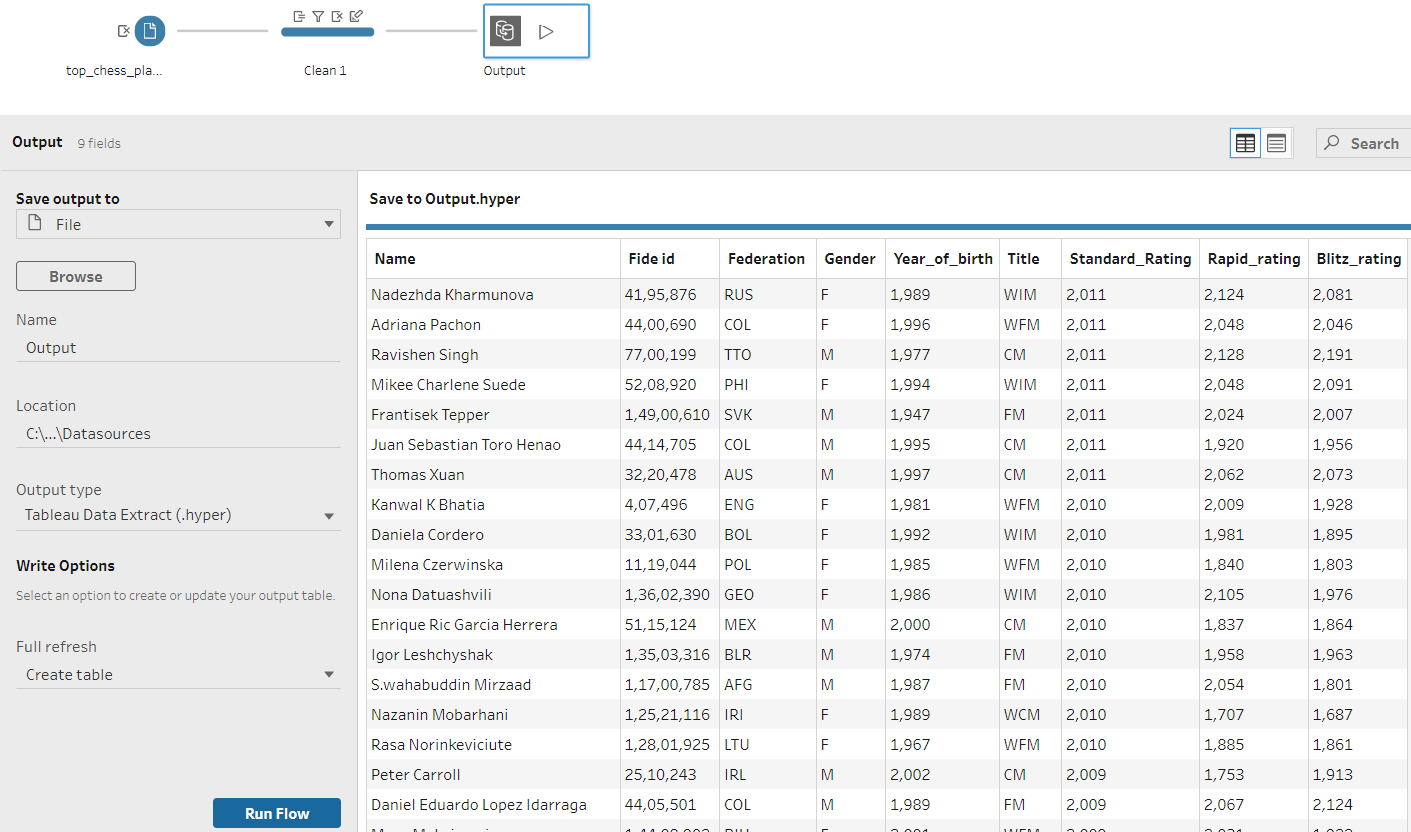


Fig.2.7 Saving Changes

**Finally, after cleaning the data, the final dataset sample is shown below:**

****

Fig.2.8. Cleaned Dataset

# Analysis on dataset

### Top Female Players according to standard rating

### Introduction

* + By performing this analysis, we will get top 10 female player from highest to lowest standard rating.

### Description

* + The analysis is based on name, gender and standard rating.

### Specific requirements, functions and formulas

* + Pivot table is used for the analysis.
  + Max or Min function is used in pivot table for the Standard Rating in the pivot table
  + In Filter Gender column is used in the pivot table

### Analysis results

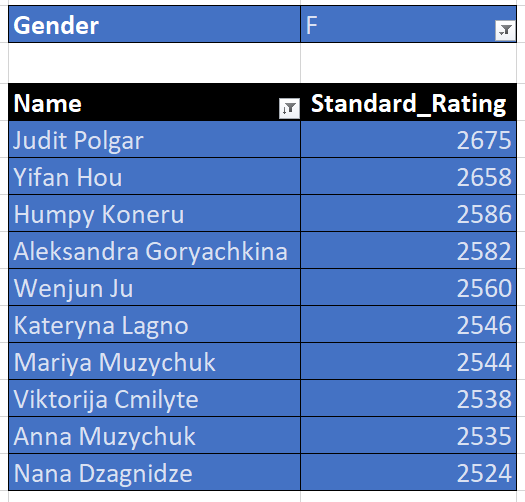


Table. 1.1 Top Female Players in Standard Game

### Visualization

* + When female is selected from filter

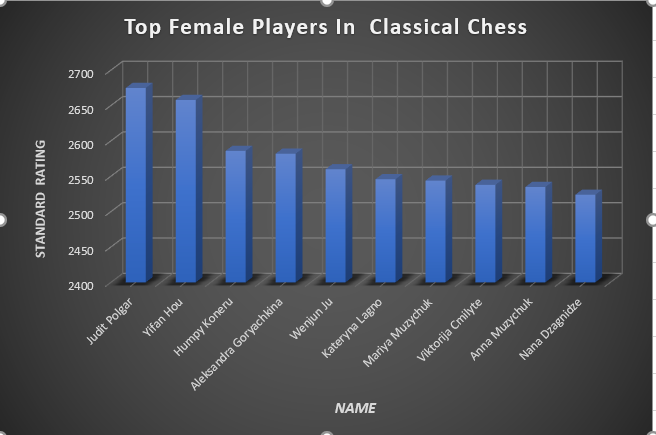


Fig.3.1 Top Female Players In Standard Game Chart

* + When male is selected from filter option

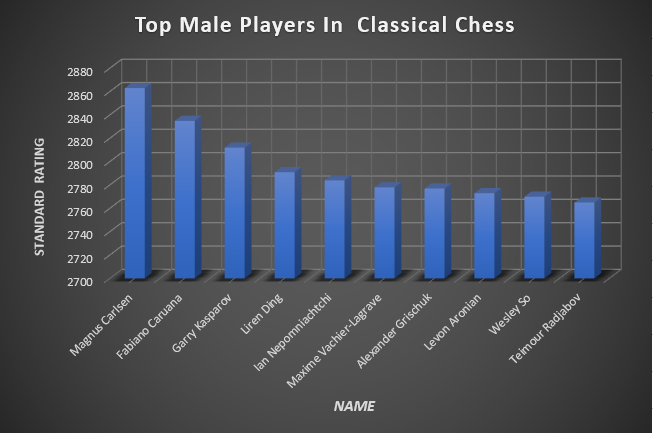


Fig.3.2 Top Male Player in Standard Game Chart

1. **Top Male Players who are born after 1990 in Blitz game**

### Introduction

* + By performing this analysis, we will get the male players who are born after 1990 in blitz game

### Description

* + The analysis based on the name, gender, blitz rating, year of birth

of the dataset

### Specific requirements, functions and formulas

* + Pivot table is used for the analysis.
  + Min function is used in pivot table for the Blitz rating in the pivot table
  + In filter year of birth and gender column is used

### Analysis results

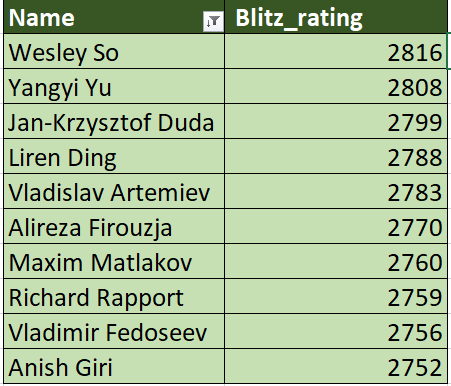


Table 1.2 Top Male in Blitz who born after 1990

**Visualization**

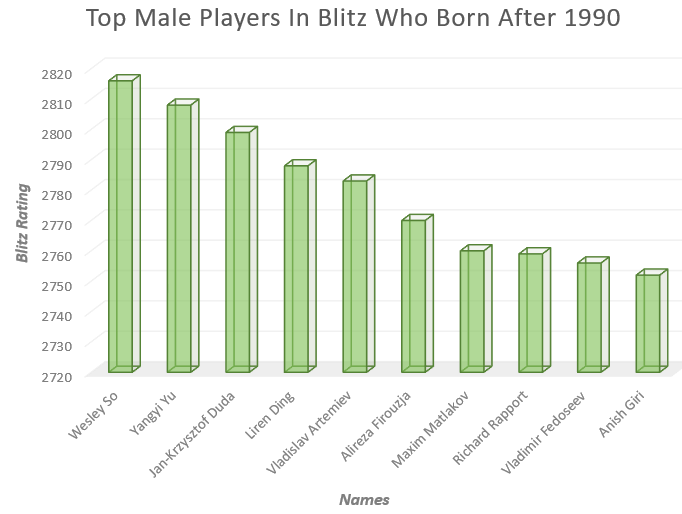
****

Fig 3.3 Top Male in Blitz Game

## Top Grand Masters from Russia Federation

### Introduction

* + By performing this analysis, we will get the top grand masters from Russia

### Description

* + The analysis based on the title, name, standard rating and federation

### Specific requirements, functions and formulas

* + Pivot table is used for the analysis.
  + Standard Rating is used in columns and name is used in rows of the dataset
  + In filter federation and title is used.

### Analysis results

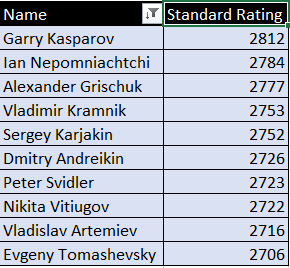
****

Table 1.3 Top GM from Russia

**Visualization**

****

Fig. 3.4 Top GM from Russia

## Top International Masters federations in Classical Game

### Introduction

* + By performing this analysis, we will get top federations in classical game of having IM title

### Description

* + The analysis is based on title, standard rating and federation of the dataset
  + Filter is used to show the title

### Specific requirements, functions and formulas

* + Pivot table is used for the analysis.
  + Average standard rating function is used in pivot table for name of the title

**Analysis results**

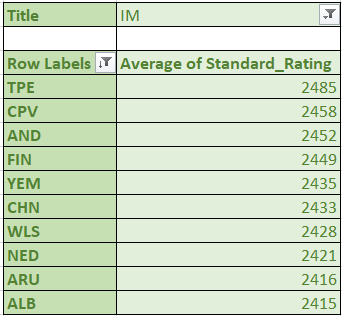


Table 1.4 Top Federations Internation Masters

**Visualization**

****

Fig 4.1 Top IM Federation

## Top Players having more standard rating than both of Blitz and Rapid rating

### Introduction

* + By performing this analysis, we will get top players whose standard rating is greater than rapid and blitz rating from the dataset

### Description

* + The analysis is based on standard rating, blitz rating and rapid rating of the dataset.

### Specific requirements, functions and formulas

* + Nested if else formula is used to get the output
  + Formula: - IF(B7>C7, IF(B7>D7, Cleaned\_Chess\_Dataset!B6117,0),0)
  + Filter is used to get the top 10 values and arranging

### Analysis results

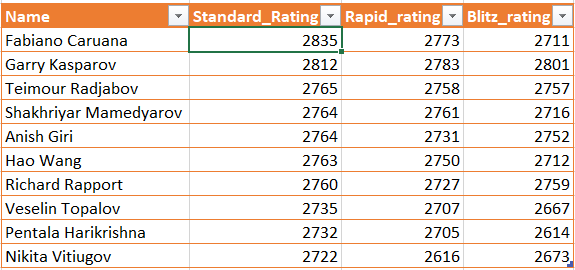
****

Table 1.5 Top Player having Greater Standard Rating

**Visualization**

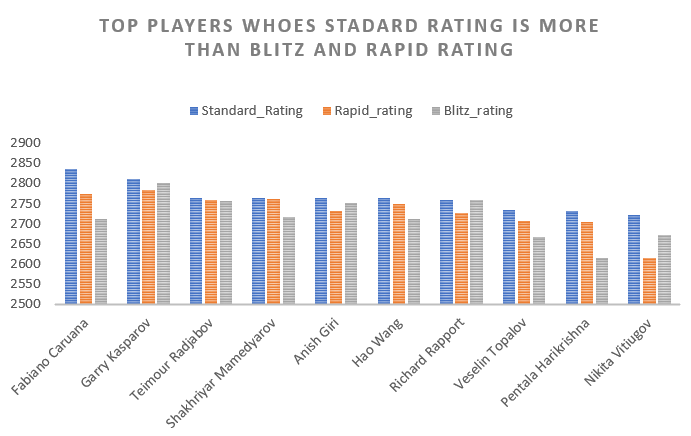
****

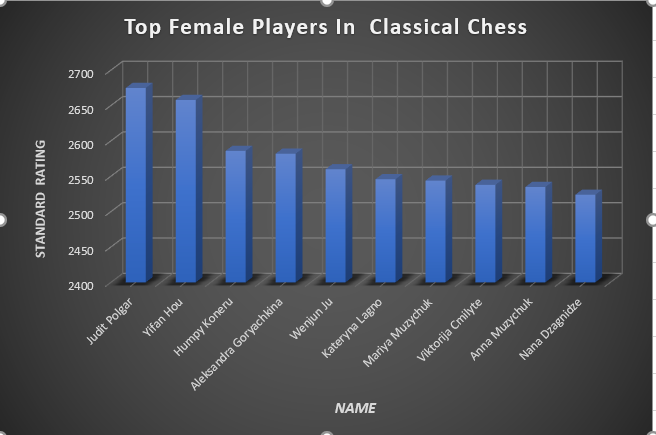
Fig 4.2 Top Players In standard rating

# List of Analysis with results

### Top Female Players according to standard rating

### Top 10 Female Players:

1. Judit Polgar
2. Yifan Hou
3. Humpy Koneru
4. Aleksandra Goryachkina
5. Wenjun Ju
6. Kateryna Lagno
7. Mariya Muzychuk
8. Viktorija Cmilyte
9. Anna Muzychuk
10. Nana Dzagnidze



### Top 10 male Players:

1. Magnus Carlsen
2. Fabiano Caruana
3. Garry Kasparov
4. Liren Ding
5. Ian Nepomniachtchi
6. Maxime Vachier-Lagrave
7. Alexander Grischuk
8. Levon Aronian
9. Wesley So
10. Teimour Radjabov
11. **Top Male Players who are born after 1990 in Blitz game**

### Top 10 male Players:

1. Wesley So
2. Yangyi Yu
3. Jan-Krzysztof Duda
4. Liren Ding
5. Vladislav Artemiev
6. Alireza Firouzja
7. Maxim Matlakov
8. Richard Rapport
9. Vladimir Fedoseev

10) Anish Giri

## Top Grand Masters from Russia Federation

### Top 10 Players from Russia:

1. Garry Kasparov
2. Ian Nepomniachtchi
3. Alexander Grischuk
4. Vladimir Kramnik
5. Sergey Karjakin
6. Dmitry Andreikin
7. Peter Svidler
8. Nikita Vitiugov
9. Vladislav Artemiev
10. Evgeny Tomashevsky

## Top International Masters federations in Classical Game

### Top 10 IM federations :

1. TPE
2. CPV
3. AND
4. FIN
5. YEM
6. CHN
7. WLS
8. NED
9. ARU
10. ALB

### Top Players whose standard rating is more than blitz and rapid rating

* + The top 10 players:
  1. Fabiano Caruana
  2. Garry Kasparov
  3. Teimour Radjabov
  4. Shakhriyar Mamedyarov
  5. Anish Giri
  6. Hao Wang
  7. Richard Rapport
  8. Veselin Topalov
  9. Pentala Harikrishna

10) Nikita Vitiugov

# FINAL DASHBOARD

****

**BIBLIOGRAPHY**

* Dataset source:

https://www.kaggle.com/vikasojha98/world-top-chess-players-august-2020

* Dashboard Background Image:

https://wallpaperaccess.com/black-king-chess-piece

* Information about Data Management:

https://[www.blue-pencil.ca/what-is-data-management-and-why-it-is-](http://www.blue-pencil.ca/what-is-data-management-and-why-it-is-) important/