



TASK

Exploratory Data Analysis on the FIFA 17 video game Data Set

[Visit our website](#)

Introduction

I have selected the FIFA17 videogame data from Kaggle. It has data on 17000 players and sources attributes such as dribbling, tackling, skill level from the FIFA game. It also has information on what nationalities and clubs the players belong to.

In my EDA I'll be looking at various factors on player information.

DATA CLEANING

I started with loading the libraries I'll be using in my data analysis. Libraries such as pandas, matplotlib are important as they'll assist me in disseminating the data better.

I load the data using the `pd.read_csv` method as my data is within a csv file.

To clean the data I used the following:

- `df_drop`: I did this to drop 2 columns that I wouldn't be using in my data analysis. I dropped the 'National_Kit' & 'National_Position' columns
- `df.drop_duplicates`: I used this to drop duplicate rows in the dataset
- `df.isnull`: I used this to sum up any null columns in the data

After the data cleaning, I then used `temp_df` to create a temporary dataset from which I would be doing my data analysis from.

I also used the `df.describe` function so I could see a summary of my data such as the mean and standard deviation.

MISSING DATA

There was missing data in the 'National_Kit' & 'National_Position' columns. I removed these columns from the dataset and chose not to include them in my data analysis.

DATA STORIES AND VISUALIZATIONS

Exploratory data visualizations were captured in the **FIFA.ipynb** notebook.

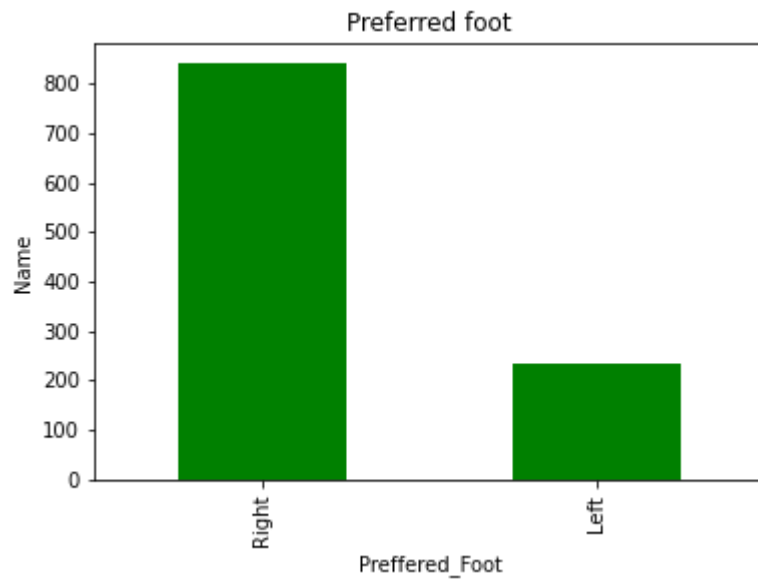


Figure 1: Split between Right & left footed players

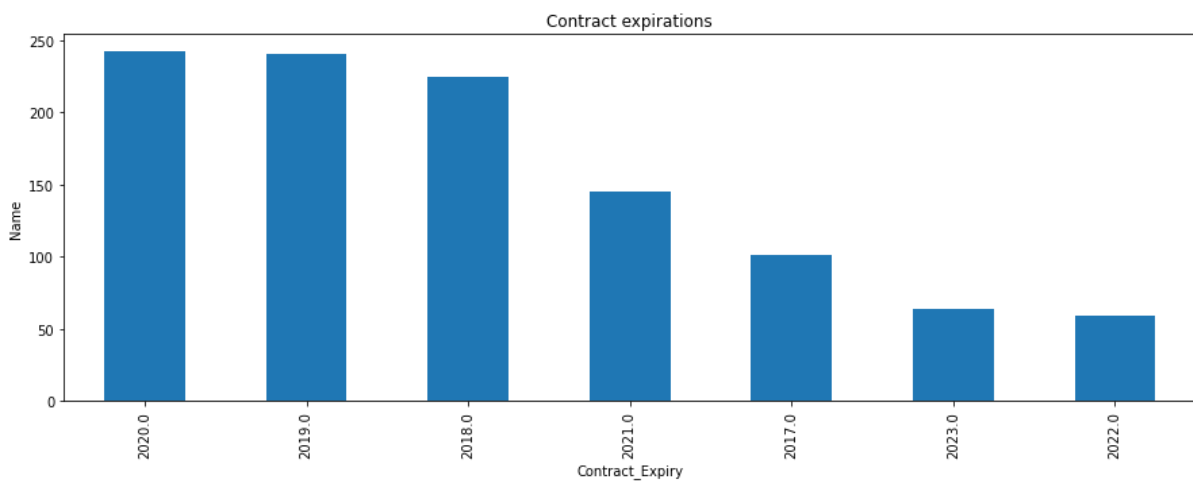


Figure 2: The year in which player contracts expire

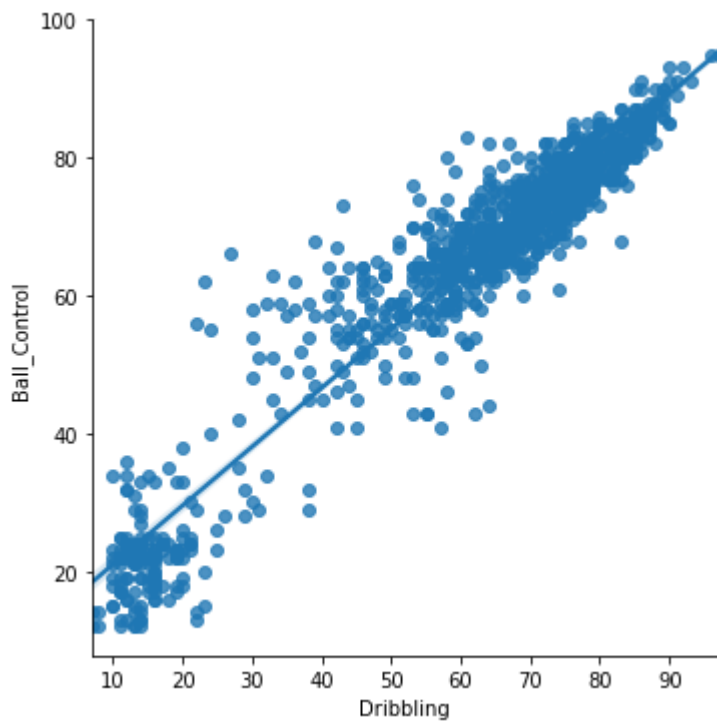


Figure 3: Correlation between player Ball control & Dribbling

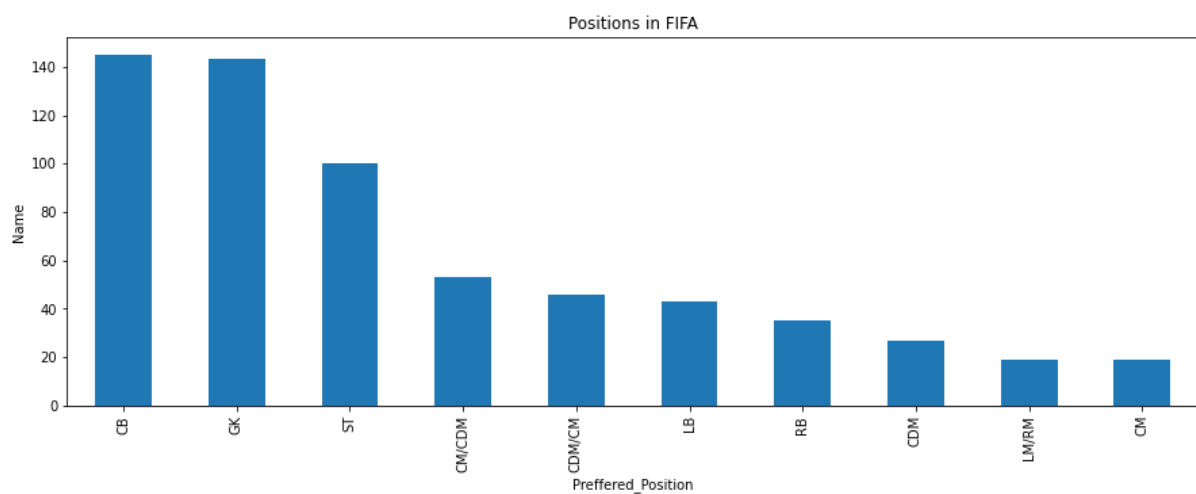


Figure 4: Popular positions in FIFA17

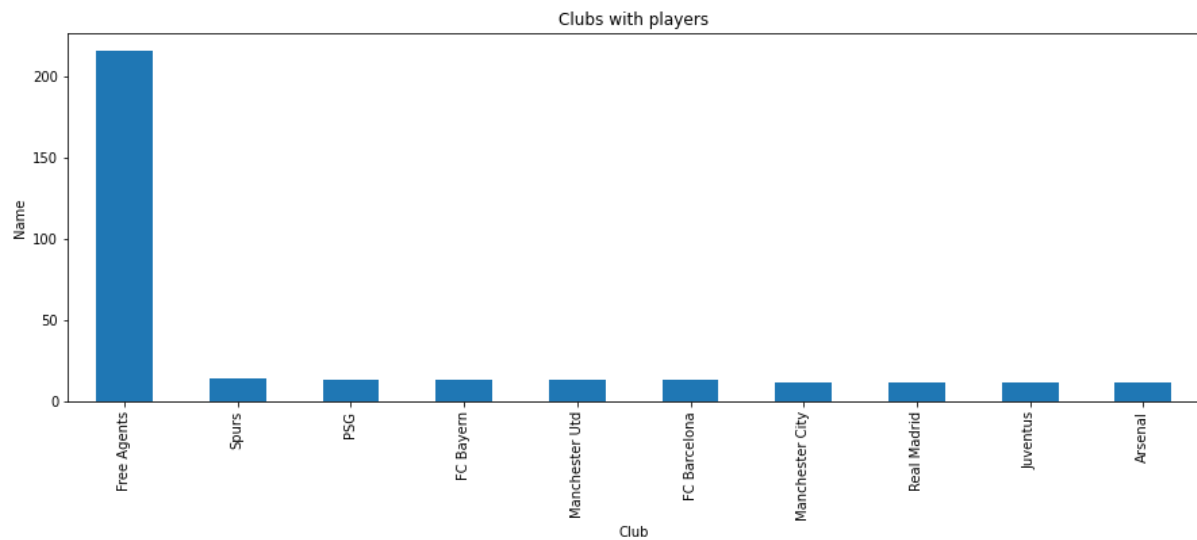


Figure 5: Clubs with the most players

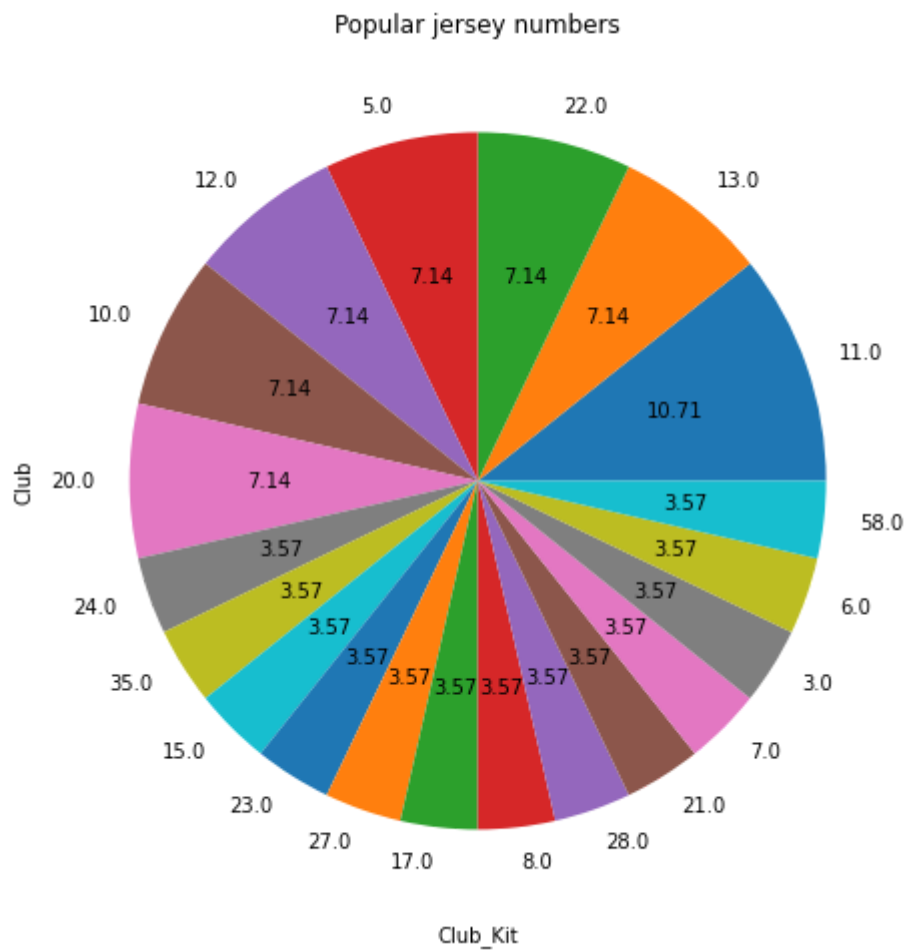


Figure 6: Most popular jersey numbers

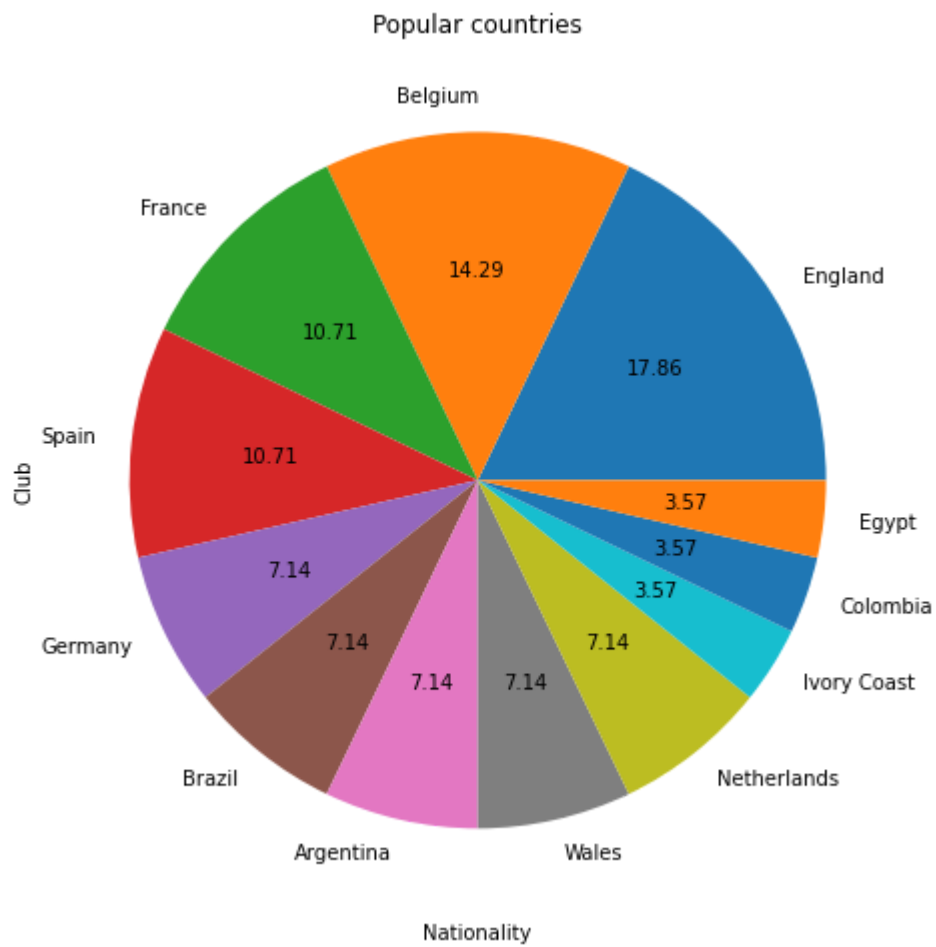


Figure 6: Player representation shown by country

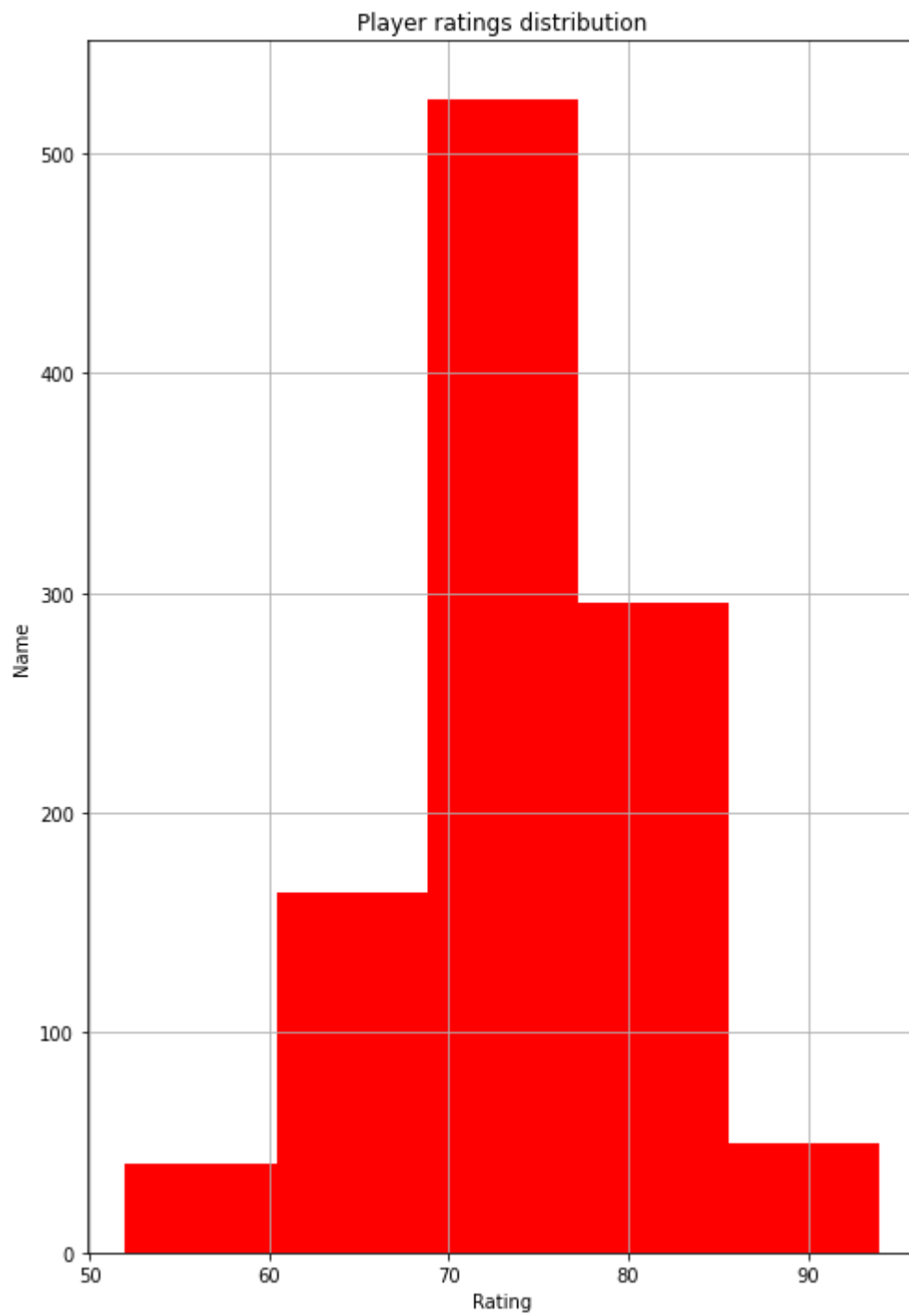


Figure7: Player rating distribution

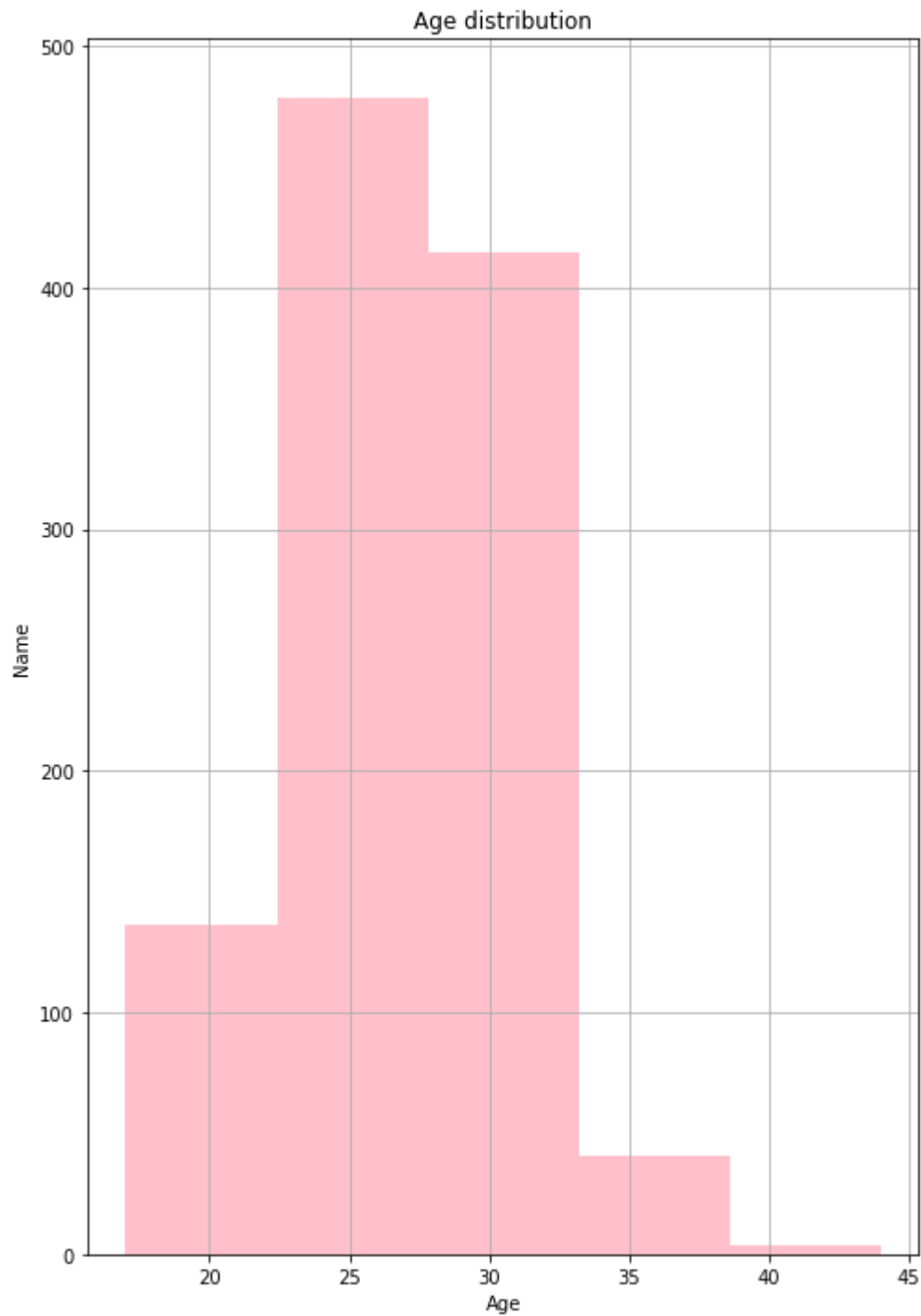


Figure7: Player age distribution

FINDINGS & CONCLUSIONS

- There are more right footed than left footed players in the video game
- At the time in 2017, the majority of players were going to be out of contract between 2018 & 2020. There were very few players on long term contracts expiring between 2022 & 2023

- When we look at the relationship between Dribbling & Ball control, we see that players with dribbling skills between 60 -80 also have ball control between 60 -80. This shows us dribbling skills & ball control go hand in hand
- We can also see that players with low dribbling skills would also have low ball control scores
- Ball control & dribbling go hand in hand
- The most popular positions in the game are that of the Centre Back (CB) & Goalkeeper(GK). This could tell us in the larger scheme of the global game, there is an oversupply of defensive players
- There aren't very many Centre Mids(CM) & Left/Right Mids(LM/RM) in the global football game. This also tells us that there could be a dearth in two footed players playing in midfield as can be seen by the little LM/RM players in the dataset
- At the time, there were more free agents represented than any other club
- The data also tells us that top European clubs choose to keep their squad sizes at more or less the same sizes
- The player distribution histogram shows that as much as there's poor players based on ability there's also as many excellent players based on player ability. I could say poor players balance out the excellent players
- We can also see that the majority of players are rated between 70 -80. This we could call the average range of player abilities in the game
- The age distribution histogram shows us most players are between 25 - 30 years of age
- There's also quite a few youngsters entering the professional ranks and also little players beyond the age of 40 still playing
- The Top 7 popular club jersey numbers are: 11; 13; 22; 5; 12; 10; 20
- The country that is best represented in the game is England which has 17% of players
- African countries with the best representation of players are Egypt & Ivory Coast

THIS REPORT WAS WRITTEN BY : MATETE NCHABELENG
