**Analyzing Stats for FIFA 23 Game and NBA Players**

Denzel Chike

Ijay Kaz-Onyekazi

Data Analysis and Knowledge Discovery, INFO 4670, Section number 002

Term Project

12/10/2023

1. **Introduction (Section I)**

This is a report for my Data Analysis class of me analyzing two datasets with Excel and rapid miner functions, methods, and analysis questions.

* 1. **Data (Section II)**

The domain of the two datasets I chose is sports. The dataset I chose are player's stats on FIFA 23 video game and NBA player's stats. Neither of my datasets included a data dictionary. My first data set FIFA 23 doesn’t include any author, but it includes three top contributors of the name Harbal Sidhu, Raphael Marconato, and Mira Ehab. For my second data set NBA players stats Eduardo Tocco created it. The people who benefit from analysis results will be sports fans and gamers. I removed multiple duplicates for my first data set, and for my second data set, no duplicates were found. I have many attributes for both data sets but the main ones I used to analyze for FIFA 23 were ID, Name, Age, Weight, Overall, and Nationality. For NBA stats I used Position, Points, and Scoring Tier.

1. **Data set 1 Excel Functions ( Section III)**

The data set I chose to use Excel to analyze is the FIFA 23 game. I chose this game because of my love for gaming and soccer. I chose 7 analysis goals questions I wanted to analyze and answer.

* First Question: Find the club of a player with a specific ID. For this question, I chose Messi whose ID number is 158023
* Answer: Paris Saint-Germain
* Explanation: The reason I chose this question was that I realized that most of the players’ names had special characters and accent marks that prevented me from using a Vlookup to search by name, so I decided to go with ID. This also shows the importance of why every player has an ID.
* Second Question: Find the overall rating of a player with a specific ID? For this question, I used ID 231747.
* Answer: 91 K.Mbappe
* Explanation: In this analysis goal we give you the overall rating of a player by entering the ID, As well this helps you find the name if a pivot table is created.
* Third Questions: Find the nationality of the player with the highest overall rating?
* Answer: Argentina and 91
* Explanation: This information can let any gamer know that Argentina has the best player in the world with an overall rating at 91.
* Fourth Question: The average player rating by club and position. For this analysis goal, I asked for the position CB and the team FC Barcelona.
* Answer: 83
* Explanation: Every club has an average player rating for each position. This can let anyone know how good each position is for every team.
* Fifth Question: Top 5 average positions? For this question, I chose FC Barcelona.
* Answer: CF: 85, CB:83, LB:82, LW:82, RB:79.5.
* Explanation: This is not the 11 positions, but this will let any gamer or analyst know what position FC Barcelona is lacking in or worse in. Any FC Barcelona analysis could use this and see that they might need to go to the market and buy a better player.
* Sixth Question: How many players for each age range from 21-25 are from a certain Nationality? For this question, I chose the age range 21 to 25 and England as the nationality that I will analyze.
* Answer: Ages:21(14), 22(6), 23(3), 24(1), 25(2),
* Explanation: This analysis question can help you discover where most young prospects and old mature players come from.
* Seventh Question: What is the average weight of athletes from a certain nationality? For this question I chose Spain.
* Answer: 73.79
* Explanation: I chose this analysis question because during international breaks players go back to their home country and play for their team. I think this can be used to help examine if their players are maintaining good weight compared to other teams.

1. **Data set 2 Rapidminer models (Section IV)**

I used Rapidminer on my second data set which was NBA stats. The association I used was the textual association. I wanted to associate the text for column “pos” with the “scoring\_tier”. The goal was to show which position scored the most points by associating every player's scoring\_tier to their position. This would then give the average scoring\_tier to each position. I removed all missing values, I then proceeded to use select attributes and only included columns “pos” and “scoring\_tier”, then added processed documents to tokenize all non-letters and their lengths. I then used fp-growth to reveal patterns and associations.

1. **Results Excel**

I would like to state that most of my pivot tables don’t represent my data goal because I used the Excel function on the pivot table.

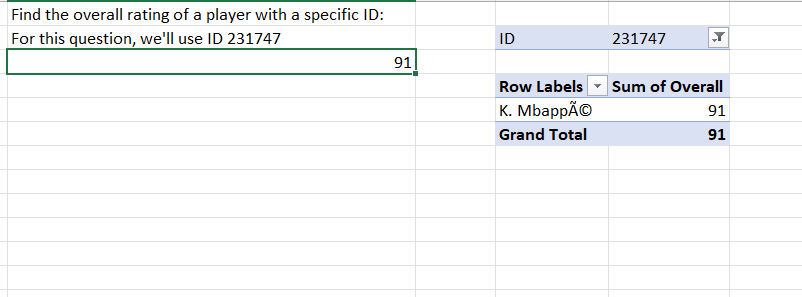
* **Question 1.**

**A screenshot of a spreadsheet

Description automatically generated**

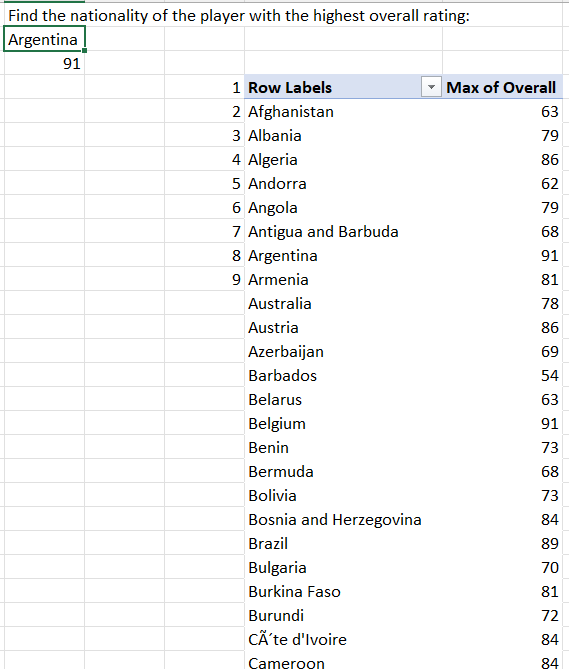
**1A.** Pivot Table shows the club the ID 158023 belongs to.

* **Question 2.**

****

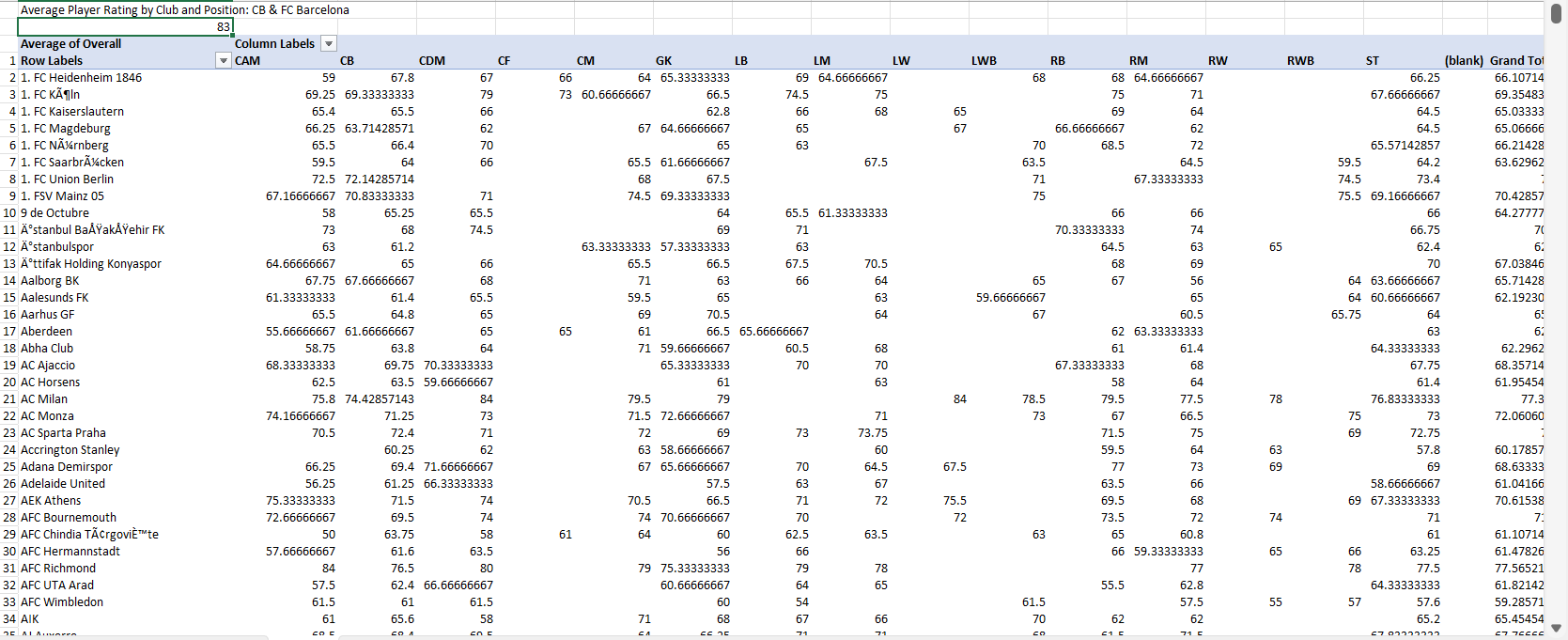
**1.B** Pivot Table shows the ID 231747 has an overall at 91.

* **Question 3.**

****

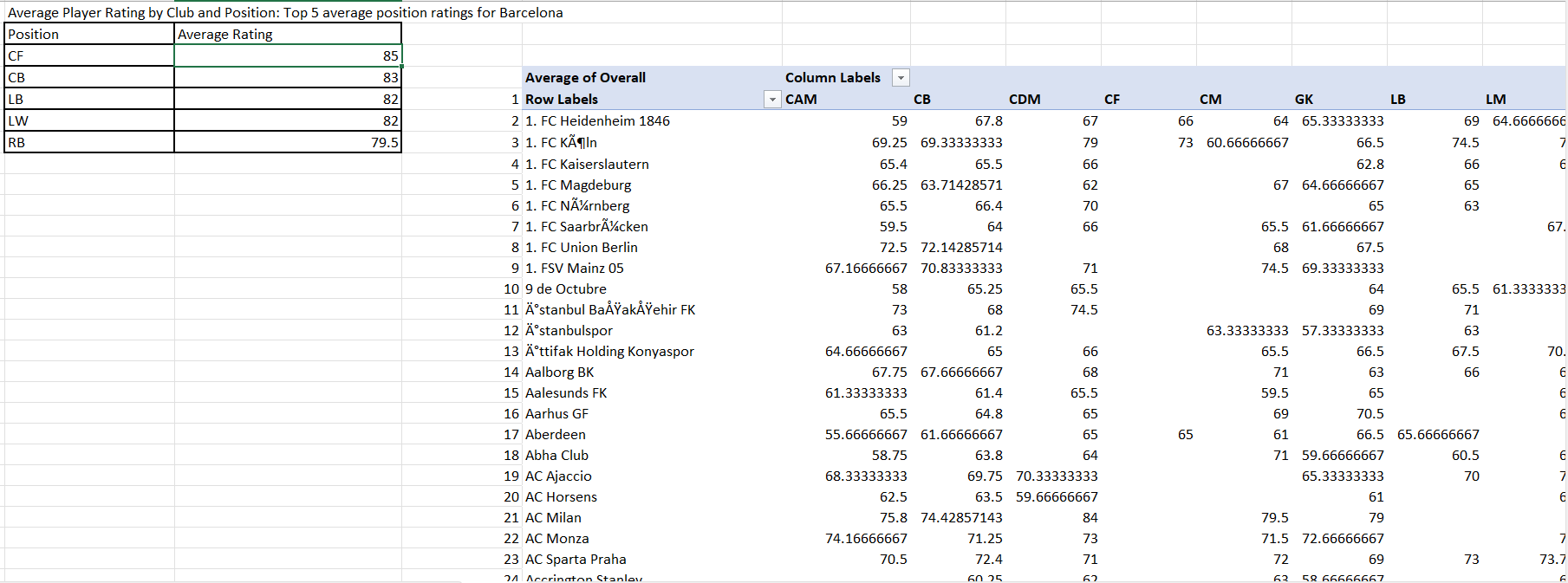
**1.C** Pivot Table shows the list of the country's best player ratings.

* **Question 4.**

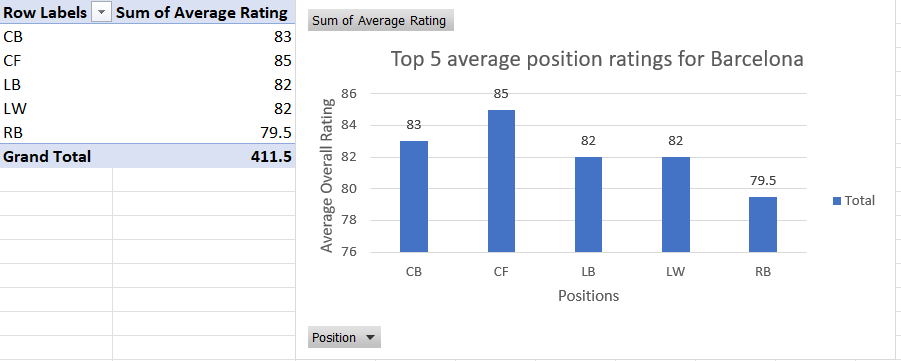
****

**1.D** Pivot Table shows the players’ ratings for each position for their club.

* **Question 5.**

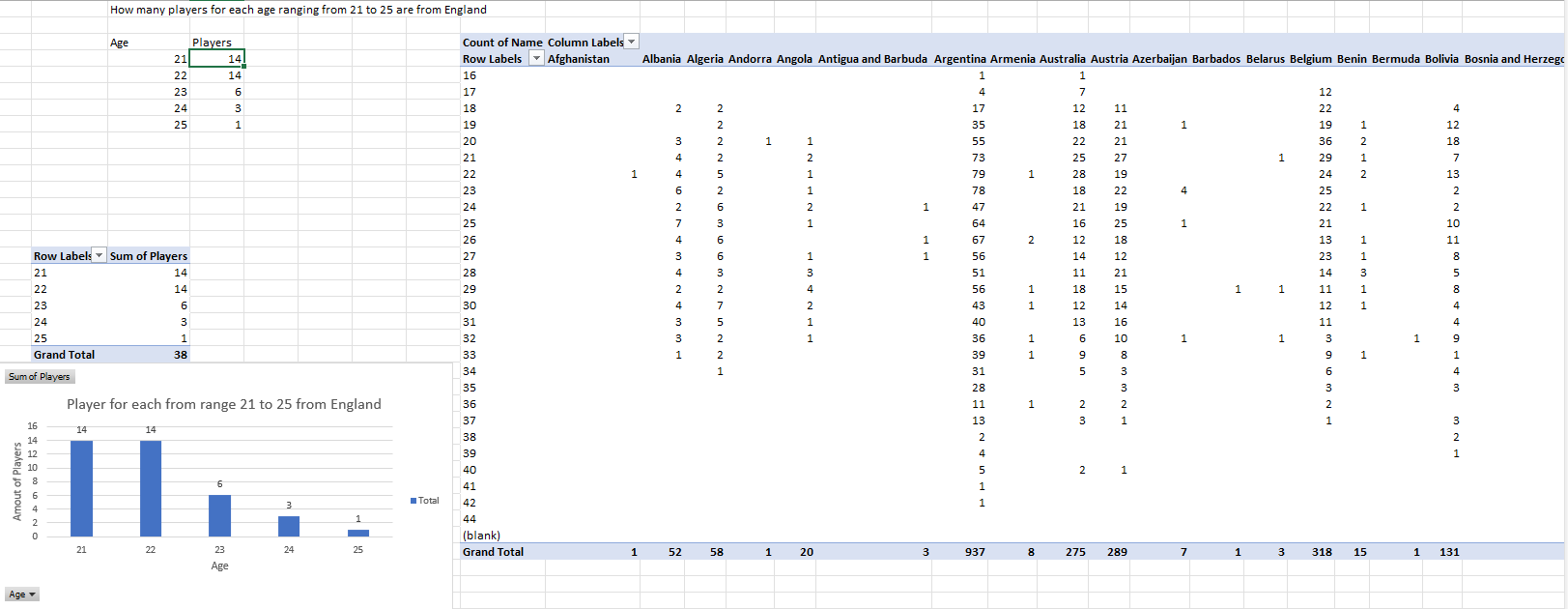
****

**1.E** Average Player Rating by Club and Position.

****

**2.E** Top 5 average position ratings for Barcelona.

* **Question 6.**

****

**1.F** Pivot Chart shows how many players for each age ranging from 21 to 25 are from England.

* **Question 7.**

****

**1.G** Pivot chart shows the average weight of athletes from Spain.

1. **Results Rapidminer**

**A diagram of a diagram

Description automatically generated**

**1.A** This shows all the operators used on Rapidminer.

**A screenshot of a computer

Description automatically generated**

**2.A** Results

1. **Discussion**

* The information I got from my results from Excel was about the best players in the world, how good each position is in a team, and whether the average weight of each national team is good or bad. The information I was hoping to get from Rapidminer was which positions score the highest based on their scoring tier levels. For example, if the position Center was continuously associated with High then a user would understand that most centers in the dataset score a lot of points.

1. **Evaluation and conclusion**

* What I found out on analyzing the FIFA 23 data set was insights on popular players, player performance, and trends. For the second data set that was analyzed with Rapidminer, I tried my best to operate Rapidminer but I understand what textual association is and the two text columns I wanted to associate with each other.

**References**

*NBA Player Stats - dataset by etocco. (2023, December 9). data.world.* [*https://data.world/etocco/nba-player-stats*](https://data.world/etocco/nba-player-stats)

*Basketball Statistics & History of Every Team & NBA and WNBA Players | Basketball-Reference.com. (n.d.). Basketball-Reference.com.* [*https://www.basketball-reference.com/*](https://www.basketball-reference.com/)

*FIFA 23 Complete Player Dataset [UPD:29/09/22]. (2022, September 29). Kaggle.* [*https://www.kaggle.com/datasets/cashncarry/fifa-23-complete-player-dataset?select=players\_fifa23.csv*](https://www.kaggle.com/datasets/cashncarry/fifa-23-complete-player-dataset?select=players_fifa23.csv)