



DATA ANALYSIS ON FIFA 19 COMPLETE PLAYER DATASET (KAGGLE)

Prepared by
Raj Tank

Notheastern University

Masters of Professional Studies in Analytics

Guided by
Shafiqul



Agenda

1. Data source, information and summary

.....

2. Visualizations and story telling

.....

3. Business questions

.....

4. Creating new attributes based on the data

.....

5. Insights of the new attributes

.....

6. Follow-up questions

.....

7. Conclusion

.....

8. References

Objective and aim

- The objective is to get all the information about footballers and how they performed in 2019.
- Many club owner acquire new talents now; thus, they can get one of these players, according to the analysis I did.
- The unending debate of who is the bet footballer player in the world statistically, mentally, and physically will be proven.

Introduction of Fifa 19 complete player dataset from Kaggle

The dataset I selected for analysis has all the information of Fifa 2019 and 18208 observations with 89 attributes in it.(Name, Club, Wage, and so on...)

Data set also needs to be clened before analyzing in R; thus, I needed to clean the data wherever necessary such as null values, column deletion, repetition, and correction.

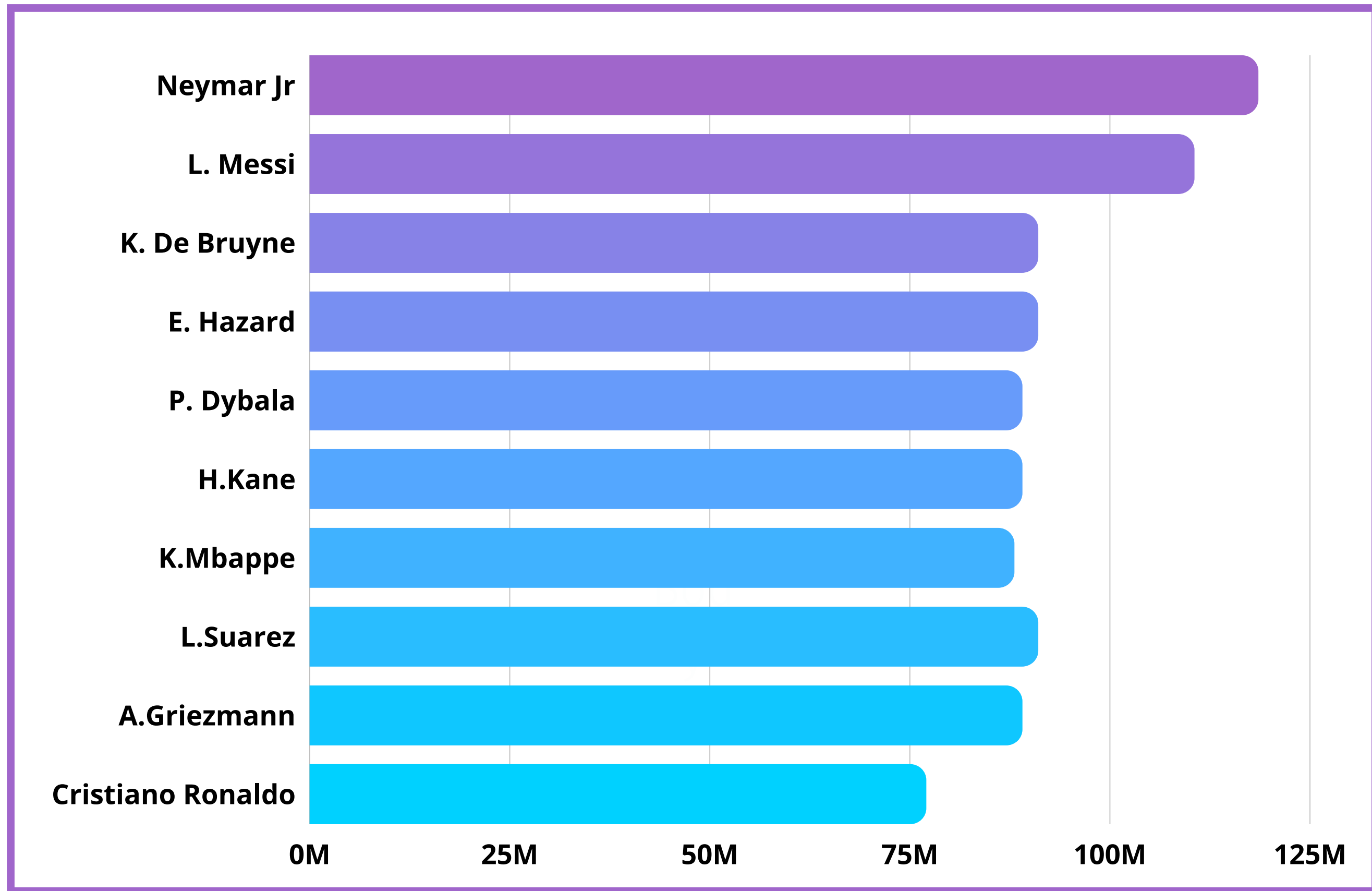
Name	Age	Nationality	Club	Value	Preferred.Foot
0	0	0	0	0	0
Weak.Foot	Work.Rate	Body.Type	Position	Crossing	Finishing
11	0	0	0	11	11
HeadingAccuracy	ShortPassing	Dribbling	Curve	BallControl	Acceleration
11	11	11	11	11	11
Jumping	LongShots				
11	11				

Null values were replaced by its mean and also computed the column significance just to delete those.

Summary of Fifa 19 Complete player dataset

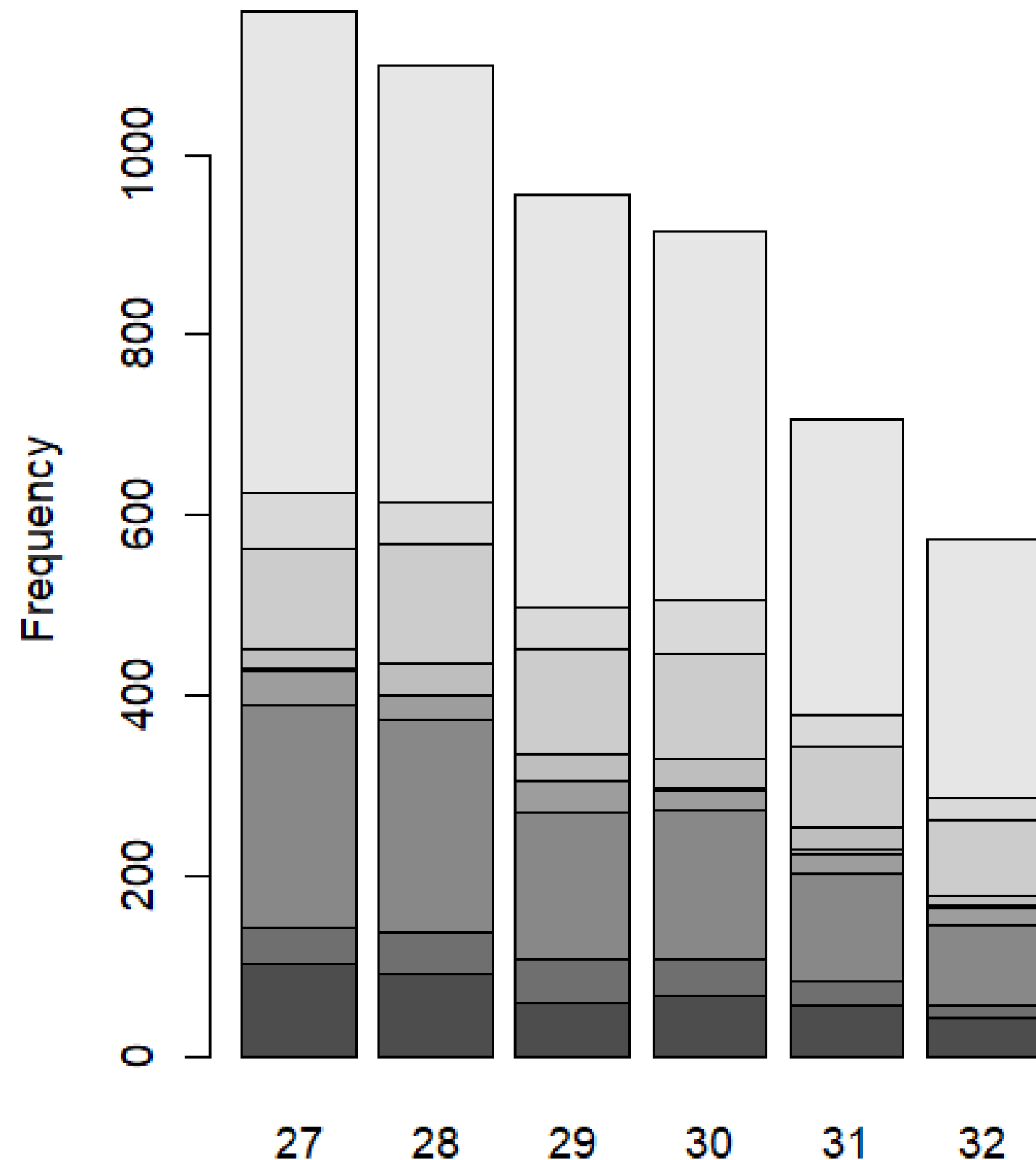
Name	Age	Nationality	Club	Value	
Length:5420	Min. :27.00	Length:5420	Length:5420	Min. : 0.0	
Class :character	1st Qu.:28.00	Class :character	Class :character	1st Qu.: 3.3	
Mode :character	Median :29.00	Mode :character	Mode :character	Median : 79.0	
	Mean :29.12			Mean :253.6	
	3rd Qu.:30.00			3rd Qu.:500.0	
	Max. :32.00			Max. :975.0	
Preferred.Foot	Weak.Foot	Work.Rate	Body.Type	Position	
Length:5420	Min. :1.000	Length:5420	Length:5420	Length:5420	
Class :character	1st Qu.:3.000	Class :character	Class :character	Class :character	
Mode :character	Median :3.000	Mode :character	Mode :character	Mode :character	
	Mean :3.001				
	3rd Qu.:3.000				
	Max. :5.000				
	NA's :11				
Crossing	Finishing	HeadingAccuracy	ShortPassing	Dribbling	Curve
Min. : 5.00	Min. : 5.00	Min. : 4.00	Min. : 7.00	Min. : 4.00	Min. : 7.00
1st Qu.:43.00	1st Qu.:32.00	1st Qu.:49.00	1st Qu.:58.00	1st Qu.:50.00	1st Qu.:37.00
Median :59.00	Median :52.00	Median :60.00	Median :64.00	Median :63.00	Median :54.00
Mean :53.17	Mean :47.92	Mean :56.06	Mean :61.39	Mean :56.83	Mean :50.75
3rd Qu.:67.00	3rd Qu.:64.00	3rd Qu.:68.00	3rd Qu.:70.00	3rd Qu.:69.00	3rd Qu.:66.00
Max. :93.00	Max. :95.00	Max. :94.00	Max. :93.00	Max. :97.00	Max. :93.00
NA's :11	NA's :11	NA's :11	NA's :11	NA's :11	NA's :11
BallControl	Acceleration	Jumping	LongShots		
Min. : 8.00	Min. :12.00	Min. :15.00	Min. : 4.00		
1st Qu.:57.00	1st Qu.:55.00	1st Qu.:61.00	1st Qu.:37.00		
Median :65.00	Median :67.00	Median :69.00	Median :57.00		
Mean :60.77	Mean :64.08	Mean :67.84	Mean :50.97		
3rd Qu.:71.00	3rd Qu.:75.00	3rd Qu.:76.00	3rd Qu.:66.00		
Max. :96.00	Max. :97.00	Max. :94.00	Max. :94.00		
NA's :11	NA's :11	NA's :11	NA's :11		

The best top 10 highly paid footballers



Work rate of footballer by their age

Work rate distribution by age of the footballer



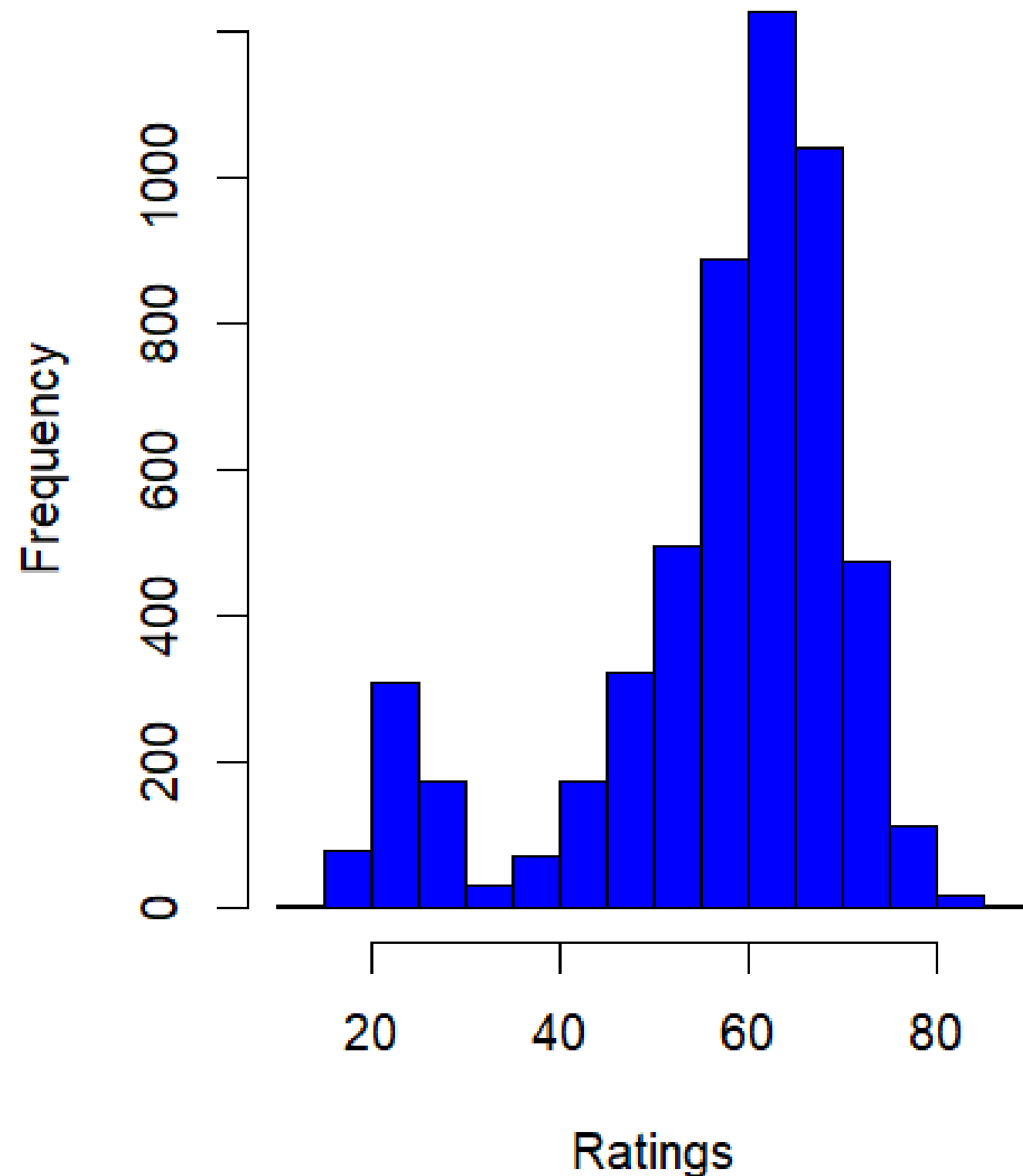
Analysis and Visualization

The footballers in the analysis has shown negative co-relation. In other words, as they grow older, their work ration went down.

This analysis helps manager and owners to purchase the player who has young age and higher work ratio except they are exceptional players like Messi and Ronaldo

Histogram of players' rating

Histogram of Players ratings



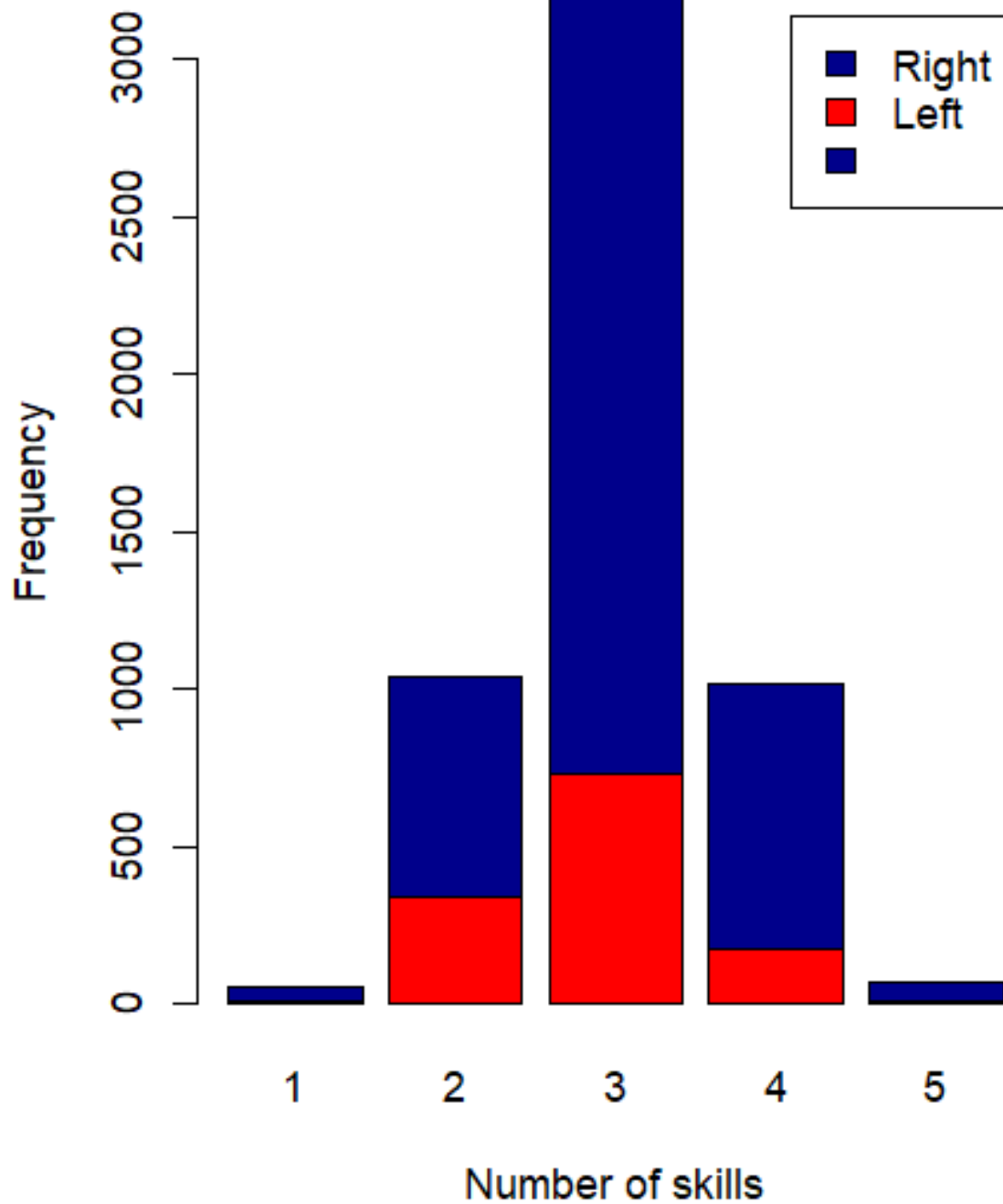
Analysis and Visualization

The season 2019-2020 was the worst season for many footballers as there are merely 2-3 player who has 94,92, and 89 rating.

Also, the rating generally stayed at the range from 60-70, 80 has very few players on it.

Histogram of players' ratings

Distribution of goal scored by both the foot

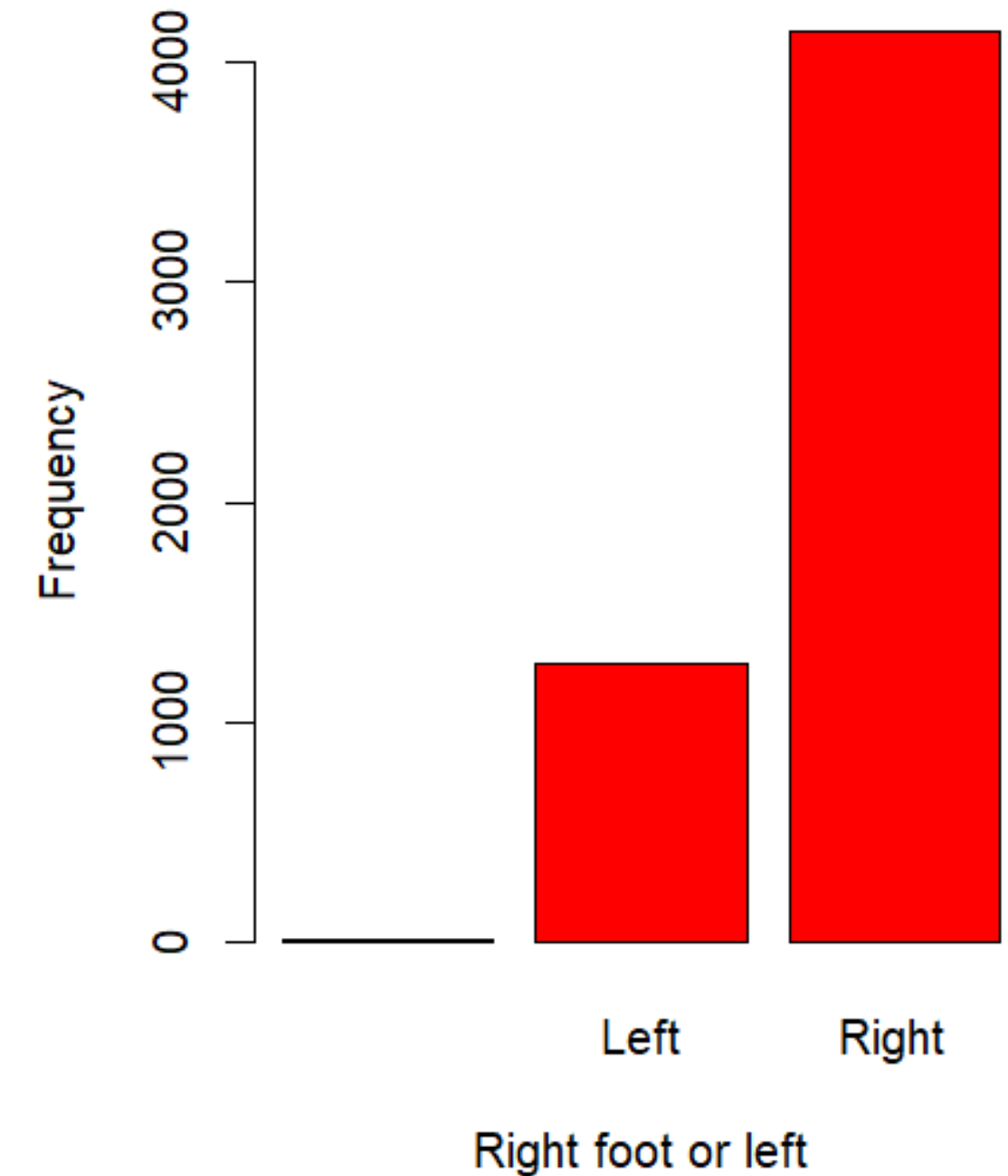


Analysis and Visualization

Right footers had more number of skills and players that scored more goals than left footers.

After the season, a manager can make decision to sign right footer over left one.

preferred foot



New attributes need to be generated to get more Information

Questions generated while doing analysis

- What is the weekly and monthly salary of the footballers?
- How can I get 110.5M into the numbers?
- What is the overall rating of the footballers; so can managers, club- owners, and people know the potential of them?
- How can I calculate various criteria from different statistic parameters?

New attributes that can answer the questions

Salary in millions:

- $\text{football\$salary_million} = \text{football\$Value} * 1000000$

Weekly and Monthly salary:

- $\text{football\$weekly_salary} = \text{football\$salary_million} / 52$
- $\text{football\$monthly_salary} = \text{football\$salary_million} / 12$

Overall rating:

- $\text{football\$Overall_Ratings} = \text{Crossing} + \text{Finishing} + \text{HeadingAccuracy} + \text{ShortPassing} + \text{Dribbling} + \text{Curve} + \text{BallControl} + \text{Acceleration} + \text{Jumping} + \text{LongShots} / 10$

Computing the mean and median for the attributes

Summary of the new attributes

```
summary(football$Sal_mill)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0	3300000	70000000	253591533	500000000	975000000

```
summary(football$weekly_sal)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0	63462	1346154	4876760	9615385	18750000

```
summary(football$overall_rating)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
12.30	52.60	60.80	56.98	66.10	87.80

What does data have in it?

Data has all the information regarding footballers' statistics and depicts that which players are the greatest. Also, managers have a chance to know after the completion of the season whether they want to acquire the player or not.

People who are fans of football can judge their favourite player, club, and country statistically, and mentally.

Follow- up questions after analysis of Fifa 19 Complete player dataset

- Can Machine Learning model help a manager or owner to predict footballers performance precisely?
- Will Data Analytics change football industry entirely such as making the right decision everytime?
- which are the other most important criteria to judge footballers?
- Will Data Analytics, and Machine Learning replace Video Assistance Referee system in the future?
- Are there any easy way to do text mining and cleaning in the data?
- Can footballers rely on the analysis in order to improve their performance?

Conclusion

After visualization and analysis, it can be reiterated that:

- Neymar, and Leo Messi was the best player in terms of money and statistically, followed by Cristiano Ronaldo who had great season.
- Right footers contributed more than left foot in terms of preferred foot, scoring, and skillwise.
- Ratings of footballers accounted for 66-78(mean), which means that most of the footballers had the worst season.
- As the footballers grow older, their work rate went down; thus, manager knows that younger player have higher ration of work than older one.

References

DataCamp (2022, March 17), Quick-R: Sorting, Datacamp. Retrieved from: <https://www.statmethods.net/management/sorting.html#:~:text=To%20sort%20a%20data%20frame,sign%20to%20indicate%20DESCENDING%20order>, Last accessed: February 18, 2022.

DeWitt, P(2014, July 14), Formatted Summary Statistics and Data Summary Tables with qwraps2. Rcran. Retrieved from: <https://cran.r-project.org/web/packages/qwraps2/vignettes/summary-statistics.html>, Last accessed: February 18, 2022.

Gadiya, K.(2018, December 21).,FIFA 19 complete player dataset. Kaggle. Retrieved from: <https://www.kaggle.com/karangadiya/fifa19>, Last accessed: February 18, 2022.

Jeng, K (2018, August 28), Awesome Fifa 19 Wallpapers!, WallpaperAccess. Retrieved from : https://wallpaperaccess.com/fifa-19#google_vignette, Last accessed: February 18, 2022.

Marsja, E.(2021, November 3),How to Add a Column to a Dataframe in R with tibble & dplyr,Erik Marsja. Retrieved from: <https://www.marsja.se/how-to-add-a-column-to-dataframe-in-r-with-tibble-dplyr/#:~:text=How%20do%20I%20add%20a,new%20variable%20to%20your%20dataset>, Last accessed: February 18, 2022.



Thank You