School of Computer Science and Engineering
(Computer Science & Engineering)
Faculty of Engineering & Technology
Jain Global Campus, Kanakapura Taluk - 562112
Ramanagara District, Karnataka, India

2023-2024
(IV Semester)

A Project Report on

"DATA ANALYSIS OF FIFA WORLD CUP"

Submitted in partial fulfilment for the award of the degree of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING

Submitted by

Khushal Vijay Donga, Chethan Y, Maridi Sai Sathvika Reddy
22BTRAD019, 22BTRAD011, 22BTRAD024

Under the guidance of

Mr. Arnab Roy
Project Practice Head and Mentor
Futurense Technologies

FACULTY OF ENGINEERING AND TECHNOLOGY

Department of Computer Science and Engineering

School of Computer Science & Engineering

Faculty of Engineering & Technology

Jain Global Campus, Kanakapura Taluk - 562112

Ramanagara District, Karnataka, India

CERTIFICATE

This is to certify that the project work titled **"DATA ANALYSIS OF FIFA"** is carried out by **Khushal Vijay Donga (22BTRAD019), Chethan Y (22BTRAD011), Maridi Sai Sathvika Reddy (22BTRAD024),** a bonafide student(s) of Bachelor / Master of Technology at the School of Engineering & Technology, Faculty of Engineering & Technology, JAIN (Deemed-to-be University), Bangalore in partial fulfillment for the award of degree in Bachelor / Master of Technology in Computer Science and Engineering, during the year **2023-2024**.

Mr. Arnab Roy
Project Practice Head and Mentor
Futurense Technologies

Date: 10-04-2024

Dr. Aditya Pai H ,
Program Head, Computer Science and Engineering,
School of Computer Science & Engineering(AI & DE)
Faculty of Engineering & Technology
JAIN (Deemed to-be University)

Date: 10-04-2024

Dr. Geetha G
Director,
School of Computer Science & Engineering
Faculty of Engineering Technology
JAIN (Deemed to-be University)

Date: 10-04-2024

Name of the Examiner

Signature of Examiner

DECLARATION

We, **Khushal Vijay Donga (22BTRAD019), Chethan Y (22BTRAD011), Maridi Sai Sathvika Reddy (22BTRAD024),** student of CSE (AI & DE) 4th semester B.Tech/ M.Tech in **Computer Science and Engineering**, at School of Engineering & Technology, Faculty of Engineering & Technology, **JAIN (Deemed to-be University)**, hereby declare that the internship work titled **"DATA ANALYSIS OF FIFA"** has been carried out by us and submitted in partial fulfilment for the award of degree in **Bachelor /Master of Technology in Computer Science and Engineering** during the academic year **2023-2024**. Further, the matter presented in the work has not been submitted previously by anybody for the award of any degree or any diploma to any other University, to the best of our knowledge and faith.

Khushal Vijay Donga: Signature

22BTRAD019:

Chethan Y: Signature

22BTRAD011:

Maridi Sai Sathvika Reddy: Signature

22BTRAD024:

Place : Bangalore

Date :

ACKNOWLEDGEMENT

It is a great pleasure for us to acknowledge the assistance and support of a large number of individuals who have been responsible for the successful completion of this project work.

First, I take this opportunity to express my sincere gratitude to Faculty of Engineering & Technology, JAIN (Deemed to-be University) for providing me with a great opportunity to pursue my Bachelors Degree in this institution.

I am deeply thankful to several individuals whose invaluable contributions have made this project a reality. I wish to extend my heartfelt gratitude to ***Dr. Chandraj Roy Chand, Chancellor****, for his tireless commitment to fostering excellence in teaching and research at Jain (Deemed-to-be-University). I am also profoundly grateful to the honorable* ***Vice Chancellor, Dr. Raj Singh, and Dr. Dinesh Nilkant, Pro Vice Chancellor****, for their unwavering support. Furthermore, I would like to express my sincere thanks to* ***Dr. Jitendra Kumar Mishra, Registrar****, whose guidance has imparted invaluable qualities and skills that will serve us well in our future endeavors.*

I extend my sincere gratitude to ***Dr. Hariprasad S A, Director*** *of the Faculty of Engineering & Technology,* ***and Dr. Geetha G, Director*** *of the School of Computer Science & Engineering within the Faculty of Engineering & Technology, for their constant encouragement and expert advice. Additionally, I would like to express my appreciation to* ***Dr. Krishnan Batri, Deputy Director (Course and Delivery), and Dr. V. Vivek, Deputy Director (Students & Industry Relations),*** *for their invaluable contributions and support throughout this project.*

It is a matter of immense pleasure to express my sincere thanks to ***Dr. Aditya Pai H, (Program Head), Computer Science and Engineering****, School of Computer Science & Engineering Faculty of Engineering & Technology for providing right academic guidance that made my task possible.*

I would like to thank our guide ***Mr. Akash Das****,* ***AVP and Project Manager at Futurense Technologies****, for sparing his valuable time to extend help in every step of my work, which paved the way for smooth progress and fruitful culmination of the project.*

*I would like to thank our Project Coordinator **Mr. Akash Das**, and all the staff members of Computer Science and Engineering for their support.*

I am also grateful to my family and friends who provided me with every requirement throughout the course.

I would like to thank one and all who directly or indirectly helped me in completing the work successfully.

Signature of Students

ABSTRACT

The initiative offers a thorough examination of FIFA player data with the goal of revealing patterns that will guide strategic choices in the football world. The analysis explores the finer points of player performance, team tactics, and market trends by utilizing a dataset that includes important player traits, positional dynamics, and performance indicators. The research uses a methodical technique to investigate several aspects of FIFA data, such as positional duties, player evaluations, potential, and favorite foot. Correlations between features are investigated using exploratory data analysis approaches, providing insight into the relationship between player characteristics and performance outcomes. Important discoveries emphasize the importance of specific characteristics, like passing accuracy, defensive prowess, and pace, in determining a player's effectiveness. Furthermore, strategic factors for player placement and team composition are revealed by positional analysis, which has an impact on tactical planning and game preparation. The insights derived from the research are applicable to a wide range of stakeholders, such as game aficionados, football clubs. Stakeholders may maximize performance in the competitive FIFA arena, improve team plans, and recruit players more efficiently by utilizing data analytics. The initiative concludes by highlighting the importance of data-driven decision-making for improving football success and providing practical insights to direct strategic excellence in FIFA games.

Chapter 1

1.1 Background and Motivation

This CSV file appears to be a comprehensive dataset containing information about football (soccer) players. It likely originates from a database used for managing player information, scouting purposes, or fantasy football applications.

Here's a breakdown of the motivation behind capturing this data:

- Player Management: Clubs and leagues might use this data to track player performance, contracts, and development. It helps analyze strengths, weaknesses, and potential of players within a team or across the league.

- Scouting: Talent scouts can utilize this data to identify promising players based on various attributes like age, nationality, skills, and physical characteristics.

- Fantasy Football: Fantasy football games often rely on player statistics for drafting and managing virtual teams. This data provides the foundation for assigning points and simulating real-world performance.

1.2 Overall Objective: Report on Football Player Potential

To identify and analyze the potential of football players for transfer, development, or scouting purposes.To create a comprehensive player database that allows for easy comparison and evaluation of talent based on various attributes.

Here's how the data in your CSV file can help you achieve this objective:

- Player Attributes: Data like Age, Nationality, Potential, Club, Value, Wage, etc., provides a basic overview of the player's experience, current standing, and market worth.

- Technical Skills: Attributes like Crossing, Finishing, Heading Accuracy, Short Passing, etc., indicate the player's technical proficiency in various aspects of the game.
- Physical Attributes: Data on Height, Weight, Acceleration, Sprint Speed, etc., provides insights into the player's physical capabilities.
- Mental Attributes: Attributes like Work Rate, Aggression, Interceptions, Vision, etc., indicate the player's mentality, decision-making skills, and playing style.

1.2 Delimitation of research

Delimitations in research define the boundaries and limitations within which the study will be conducted. Here are some delimitations specific to analyzing the FIFA dataset:

- Focus: This dataset likely focuses on professional male football (soccer) players. You can delimit your research to this specific sport and gender.
- League/Competition: The data might encompass players from various leagues or competitions. You can choose to analyze players from a specific league (e.g., Premier League, La Liga) or competition (e.g., FIFA World Cup).
- Player Experience: The data might include players of all experience levels. You could delimit your research to established professional players (excluding youth academy players or those with minimal playing time).
- Data Points: The dataset contains a vast range of player attributes. You can choose to focus on a specific subset of these attributes relevant to your research question. For example, if analyzing transfer values, you might delimit to attributes like age, potential, and playing position.
- Timeframe: The data likely reflects player information at a specific point in time. You can delimit your analysis to that timeframe (e.g., data collected during a specific transfer window) and acknowledge that player attributes can change over time.

1.3 Benefits of research

Research on the analysis of FIFA datasets can yield several benefits, including:

Improved Player Evaluation and Scouting:

- Identify hidden gems: By analyzing statistical attributes alongside potential and other metrics, you can unearth undervalued players who might be good signings for a club.
- Predict future performance: Through statistical modeling, you can estimate a player's potential for improvement and future value.
- Compare players objectively: Analyze combinations of stats to compare players competing for the same position or role.

Tactical Analysis and Team Building:

- Identify team weaknesses and strengths: Analyze team statistics based on positions to see areas needing improvement or areas to capitalize on.
- Develop effective playing styles: Explore correlations between player attributes and playing styles to build a team suited to a specific strategy.
- Optimize player placement: Analyze individual player stats to determine the best position for their strengths within the team's formation.

Market Trend Analysis:

- Identify transfer market trends: Analyze trends in player value, wages, and positions to inform transfer strategies.
- Predict market fluctuations: Analyze historical data to potentially forecast future trends in player valuation.
- Assess the impact of player attributes on value: Determine which attributes have the most significant influence on a player's market value.

Chapter 2

2.1 Details of the dataset:

Number of Records: 18,208
Number of Columns: 54

Column Descriptions:

- **Player Information:**
  - Name
  - Age
  - Nationality
  - Potential (future growth potential)
  - Club (current club)
  - Value (estimated transfer market value)
  - Wage (weekly wage)
  - Preferred Foot (dominant foot used for playing)
  - International Reputation (level of recognition in international football)
- **Physical Attributes:**
  - Weak Foot (strength of non-dominant foot)
  - Skill Moves (technical ability with the ball)
  - Work Rate (effort and stamina on the field)
  - Body Type (physical build of the player)
  - Position (playing position on the field)
  - Jersey Number
- **Contract Details:**
  - Joined (date the player joined the club)
  - Loaned From (club the player is on loan from, if applicable)
  - Contract Valid Until (date the player's contract expires)
- **Physical Measurements:**
  - Height
  - Weight
- **Technical Skills (all likely rated on a scale):**
  - Crossing (ability to deliver accurate crosses)
  - Finishing (ability to score goals)
  - Heading Accuracy (skill in heading the ball)
  - Short Passing (accuracy of short passes)
  - Volleys (ability to shoot accurately on the first touch)
  - Dribbling (skill in controlling the ball while moving)
  - Curve (ability to bend the ball with shots and passes)
  - FK Accuracy (accuracy of free kicks)

    - Long Passing (accuracy of long passes)
    - Ball Control (ability to control the ball)
- **Physical Attributes (all likely rated on a scale):**
    - Acceleration (speed of acceleration)
    - Sprint Speed (maximum running speed)
    - Agility (ability to change direction quickly)
    - Reactions (reflexes and decision-making speed)
    - Balance (ability to maintain balance while moving)
    - Shot Power (strength of shots)
    - Jumping (jumping ability)
    - Stamina (physical endurance)
    - Strength (physical power)
- **Mental Attributes (all likely rated on a scale):**
    - Long Shots (ability to score goals from long distances)
    - Aggression (level of aggressiveness)
    - Interceptions (ability to anticipate and intercept passes)
    - Positioning (ability to be in the right place at the right time)
    - Vision (ability to see and anticipate plays)
    - Penalties (ability to score penalties)
    - Composure (ability to stay calm under pressure)
- **Goalkeeper Specific Attributes (likely only for players with Goalkeeper position):**
    - Marking (ability to mark and challenge opposing attackers)
    - Standing Tackle (ability to win the ball in a standing tackle)
    - Sliding Tackle (ability to win the ball in a sliding tackle)
    - GKDiving (skill in diving to save shots)
    - GKHandling (ability to control the ball in the hands)
    - GKKicking (ability to distribute the ball with kicks)
    - GKPositioning (ability to position oneself effectively in goal)
    - GKReflexes (reaction speed and agility in goal)

2.2 Data Preprocessing:

1. Handling Missing Values:

- Identify Missing Values: Check for missing values (empty cells) in each column. You can use functions like isnull() or countna() in libraries like Pandas to identify missing entries.
- Decide on Strategy: Choose a strategy to handle missing values. Options include:
    - Deletion: Remove rows or columns with a high percentage of missing values (be cautious, this can lose data).
    - Imputation: Fill missing values with estimated values based on other data (e.g., mean/median for numerical data, mode for categorical data). You might consider advanced techniques like K-Nearest Neighbors (KNN) imputation.

2. Data Type Verification:

- Check Data Types: Ensure each column has the appropriate data type (e.g., integer for Age, string for Nationality). Tools like Pandas dtypes attribute can help verify this.
- Correct Inconsistencies: Fix any inconsistencies in data types. For example, convert inconsistent date formats to a single format.

3. Additional Considerations:

- Handling Dates: If the "Joined" or "Contract Valid Until" columns contain dates, convert them to a consistent format (e.g., YYYY-MM-DD) or consider extracting features like years as a player (Joined - Birth Year).

Chapter 3

Unveiling the Trends

- NUMBER OF PLAYER IN EACH TEAM AND NATIONALITY

```
Number of players in each club:
FC Barcelona               33
Southampton                33
Cardiff City               33
TSG 1899 Hoffenheim        33
Wolverhampton Wanderers    33
                           ..
Vitória                    20
Bahia                      20
Sligo Rovers               19
Limerick FC                19
Derry City                 18

Number of players of each nationality:
England      1655
Germany      1195
Spain        1071
Argentina     935
France        910
             ...
Guam            1
Fiji            1
Liberia         1
Mauritius       1
Botswana        1
```

- TOP 10 PLAYERS BY OVERALL RATING

	ID	Name	Age	Nationality	Overall	Potential
0	158023	L. Messi	31	Argentina	94	94
1	20801	Cristiano Ronaldo	33	Portugal	94	94
2	190871	Neymar Jr	26	Brazil	92	93
3	193080	De Gea	27	Spain	91	93
4	192985	K. De Bruyne	27	Belgium	91	92
5	183277	E. Hazard	27	Belgium	91	91
6	177003	L. Modrić	32	Croatia	91	91
7	176580	L. Suárez	31	Uruguay	91	91
8	155862	Sergio Ramos	32	Spain	91	91
12	182493	D. Godín	32	Uruguay	90	9

- GOALKEEPERS WITH BEST REFLEXES

	Name	GK_Avg
3	De Gea	81.166667
22	M. Neuer	80.666667
9	J. Oblak	79.000000
18	M. ter Stegen	78.333333
57	Ederson	77.666667
41	G. Buffon	76.833333
40	S. Handanovič	76.833333
180	J. Pickford	76.833333
19	T. Courtois	76.666667
46	K. Navas	76.166667

- TOP 10 PENALTY TAKERS

	Name	Penalties
206	M. Balotelli	92.0
118	Fabinho	91.0
297	M. Kruse	90.0
16	H. Kane	90.0
507	R. Boudebouz	90.0
823	R. Jiménez	90.0
384	D. Perotti	90.0
945	L. Baines	90.0
68	M. Reus	89.0
109	Z. Ibrahimović	89.0

- TOP 10 PLAYERS BY VALUE

Top 10 Players by Value:

	Name	Nationality	Club	Value
2	Neymar Jr	Brazil	Paris Saint-Germain	11851000000
0	L. Messi	Argentina	FC Barcelona	11051000000
16	H. Kane	England	Tottenham Hotspur	8351000000
11	T. Kroos	Germany	Real Madrid	7651000000
31	C. Eriksen	Denmark	Tottenham Hotspur	7351000000
30	Isco	Spain	Real Madrid	7351000000
32	Coutinho	Brazil	FC Barcelona	6951000000
28	J. Rodríguez	Colombia	FC Bayern München	6951000000
26	M. Salah	Egypt	Liverpool	6951000000
23	S. Agüero	Argentina	Manchester City	6451000000

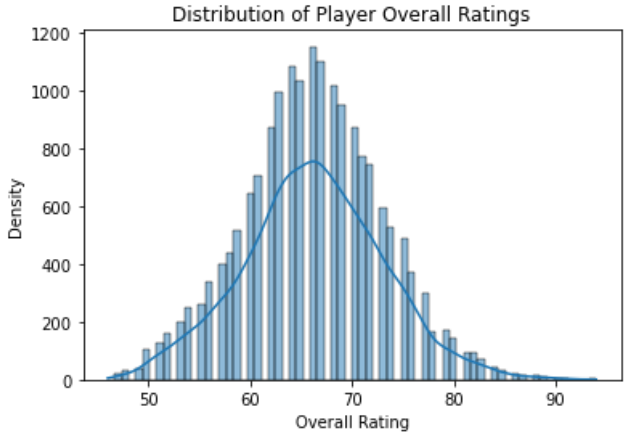
- Top 10 Players by Potential

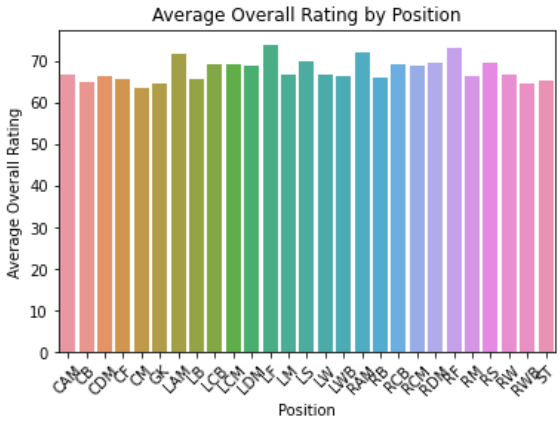
	Name	Potential	Overall
25	K. Mbappé	95	88
0	L. Messi	94	94
1	Cristiano Ronaldo	94	94
15	P. Dybala	94	89
2	Neymar Jr	93	92
3	De Gea	93	91
9	J. Oblak	93	90
229	G. Donnarumma	93	82
4	K. De Bruyne	92	91
18	M. ter Stegen	92	89

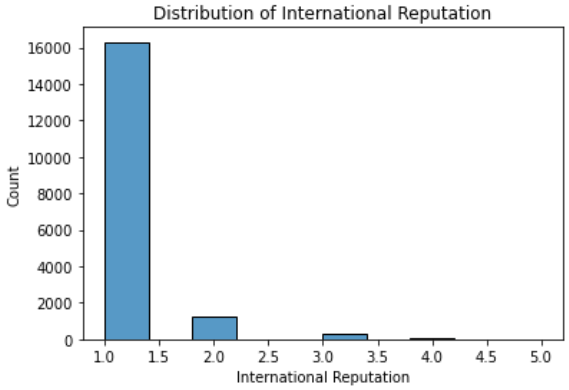
- TOP 10 PLAYERS BY BEST AVERAGE OVERALL

	Name	Avg_Attributes
35	Marcelo	69.147059
45	P. Pogba	68.705882
4	K. De Bruyne	68.676471
6	L. Modrić	68.588235
36	G. Bale	68.558824
101	R. Nainggolan	68.294118
96	A. Vidal	68.029412
85	D. Alaba	67.882353
428	M. Acuña	67.794118
258	A. Florenzi	67.676471

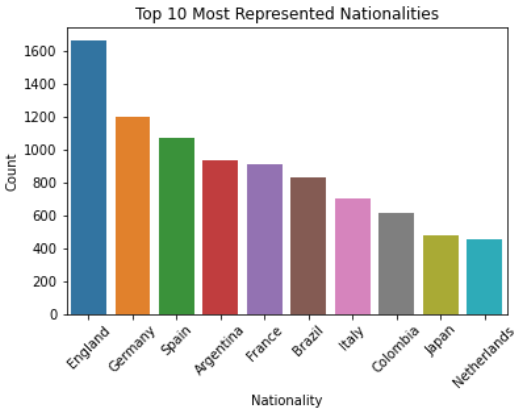
- DISTRIBUTION OF OVERALL PLAYER RATINGS

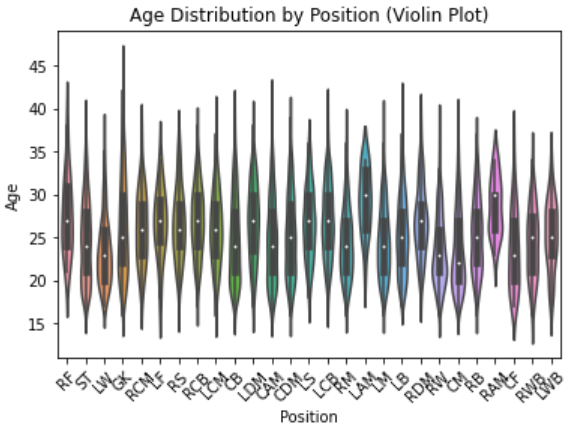






- TOP 10 MOST REPRESENTED NATIONALITIES



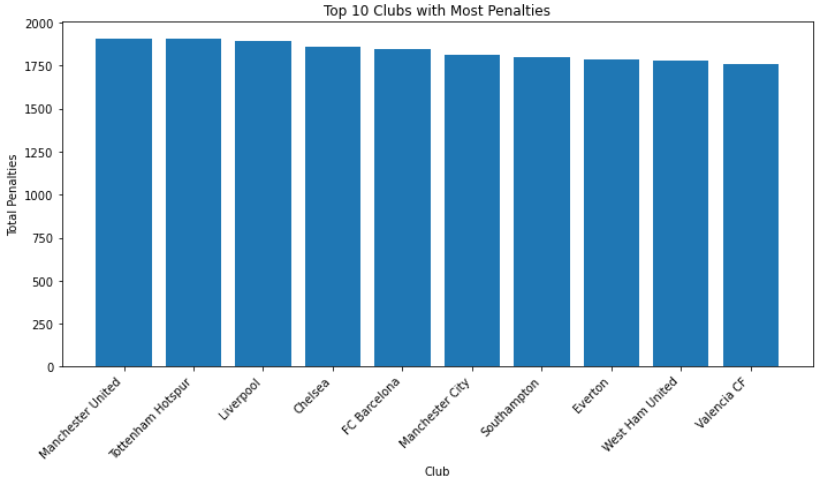


- TOP PLAYERS BY POSITION (BASED ON AVERAGE SKILL SCORE)

Position			
CAM	17	A. Griezmann	CAM
	11829	J. Calle	CAM
	11548	M. Laurent	CAM
CDM	20	Sergio Busquets	CDM
	11685	D. Potter	CDM
	11444	Adrián Rocheira	CDM
GK	3	De Gea	GK
	22	M. Neuer	GK
	18	M. ter Stegen	GK
LB	35	Marcelo	LB
	12255	I. Vujica	LB
	12247	N. Pantaleone	LB
LCB	24	G. Chiellini	LCB
	9068	D. Balanta	LCB
	8715	F. Carvalho	LCB
LCM	11	T. Kroos	LCM
	9060	D. Blanco	LCM
	8967	L. Angulo	LCM
LF	5	E. Hazard	LF
	15	P. Dybala	LF
	76	Iniesta	LF
LW	2	Neymar Jr	LW
	11513	I. Boonen	LW
	11794	K. Hazard	LW
RB	69	Azpilicueta	RB
	12422	C. Özkan	RB
	12123	M. Gomes	RB
RCB	8	Sergio Ramos	RCB
	8505	S. Hefti	RCB
	8521	R. Fennell	RCB
RCM	4	K. De Bruyne	RCM
	8667	J. Spearing	RCM
	9605	G. Bijl	RCM
RF	0	L. Messi	RF
	50	D. Mertens	RF
	528	D. Valeri	RF
RW	56	Bernardo Silva	RW
	11413	K. Anderson	RW
	11783	M. Ortíz	RW
ST	1	Cristiano Ronaldo	ST
	12694	L. John-Lewis	ST
	12796	F. Schubert	ST

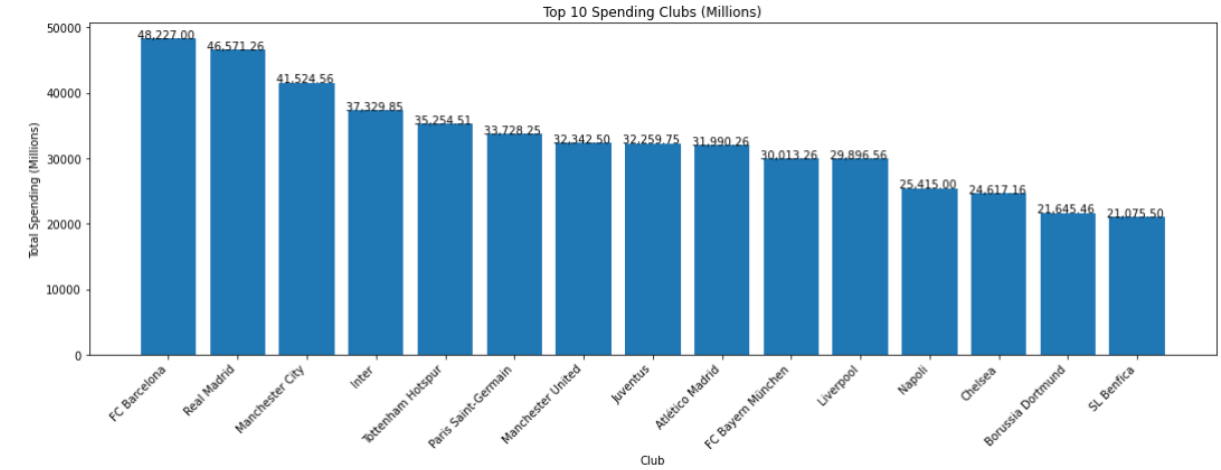
- TOP 10 CLUBS WITH MOST PENALTIES TAKEN

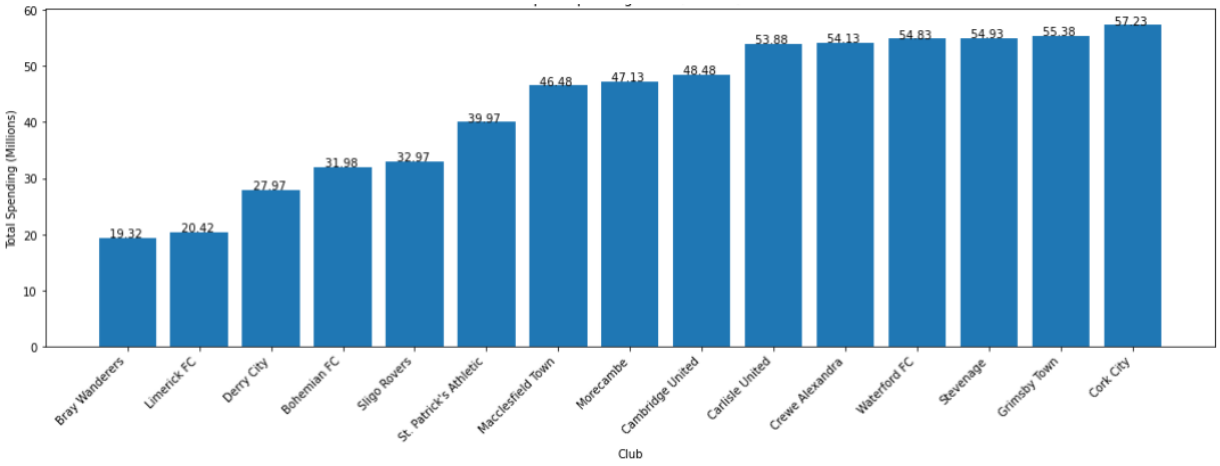
```
Club
Manchester United    1909.0
Tottenham Hotspur    1907.0
Liverpool            1894.0
Chelsea              1862.0
FC Barcelona         1849.0
Manchester City      1816.0
Southampton          1801.0
Everton              1786.0
West Ham United      1777.0
Valencia CF          1757.0
Name: Penalties,
```



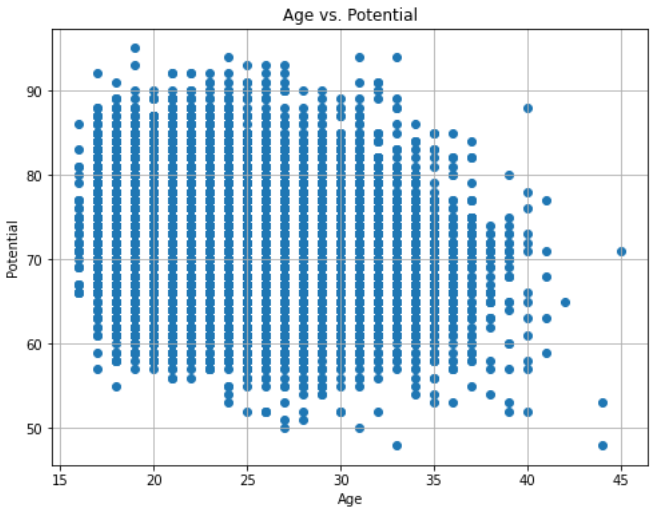
- TOP 10 CLUBS AND THEIR TOTAL SPENDING

```
Top 10 clubs and their total spending (in millions):
Club
FC Barcelona           48,227.00 Million
Real Madrid            46,571.26 Million
Manchester City        41,524.56 Million
Inter                  37,329.85 Million
Tottenham Hotspur      35,254.51 Million
Paris Saint-Germain    33,728.25 Million
Manchester United      32,342.50 Million
Juventus               32,259.75 Million
Atlético Madrid        31,990.26 Million
FC Bayern München      30,013.26 Million
```

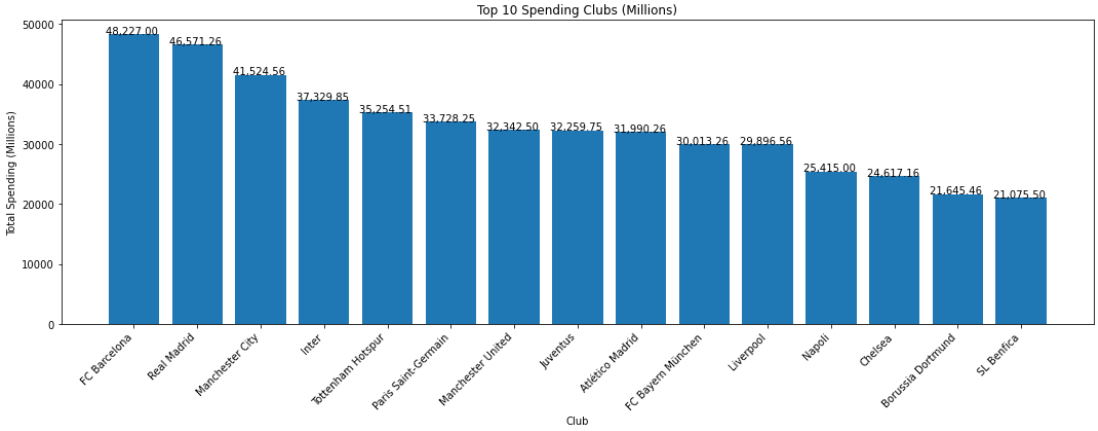




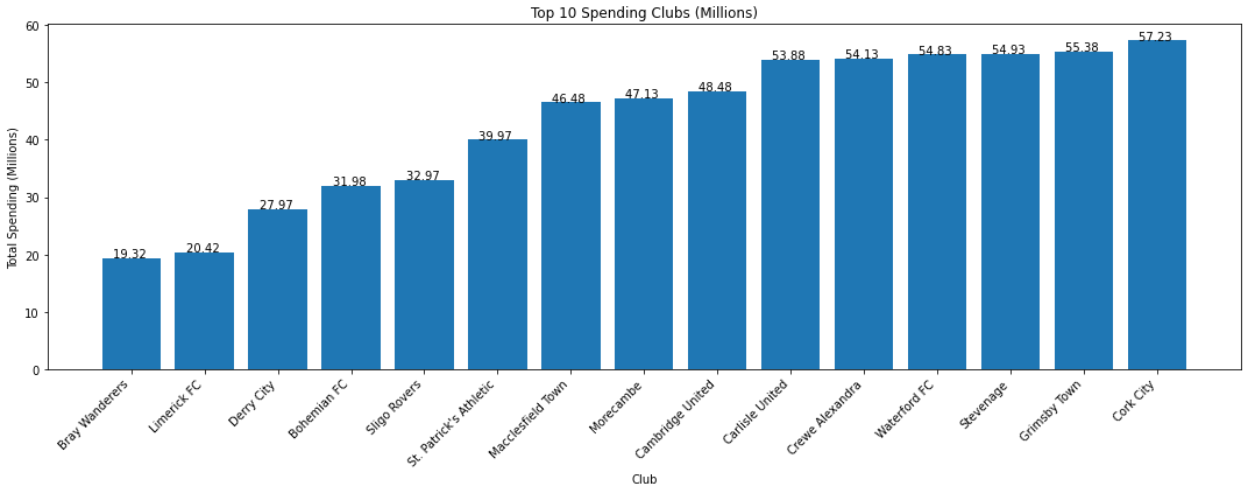
- AGE VS. POTENTIAL



- TOP 10 MOST SPENDING CLUBS



- TOP 10 LESS SPENDING CLUBS



- Players Contract Valid Until

```
Number of players with contract valid until each year:
2019.0    4819
2021.0    4360
2020.0    4027
2022.0    1477
2023.0    1053
2018.0     886
2024.0      23
2025.0       7
2026.0       2
```

- PLAYERS WITH LOW VALUE BUT HIGH AVERAGE

```
Players with low value but high average (within 1 standard deviation):
                     Name      Value
41              G. Buffon   41000000
102                 Naldo   91000000
108                  Pepe   91000000
201            D. Subašić  131000000
218           S. Mandanda  131000000
```

- BEST PLAYERS IN SOME BEST CATEGORY

```
Best players in each category:
Crossing: K. De Bruyne
Finishing: L. Messi
HeadingAccuracy: Naldo
ShortPassing: L. Modrić
Dribbling: L. Messi
LongPassing: T. Kroos
Jumping: Cristiano Ronaldo
Positioning: Cristiano Ronaldo
Marking: A. Barzagli
StandingTackle: G. Chiellini
SlidingTackle: Sergio Ramos
GKDiving: De Gea
GKHandling: J. Oblak
GKKicking: M. Neuer
GKPositioning: G. Buffon
```

- BEST TEAM FORM UNDER THE AGE 30

Total Budget: € 598100000.0

	ID	Name	Age	Club	Value	Overall
1	193080	De Gea	27	Manchester United	72000000.0	91
2	183277	E. Hazard	27	Chelsea	93000000.0	91
3	202126	H. Kane	24	Tottenham Hotspur	83500000.0	89
4	190460	C. Eriksen	26	Tottenham Hotspur	73500000.0	88
5	190483	Douglas Costa	27	Juventus	46500000.0	86
6	189509	Thiago	27	FC Bayern München	45500000.0	86
7	180206	M. Pjanić	28	Juventus	44000000.0	86
8	184267	Y. Brahimi	28	FC Porto	39000000.0	85
9	205498	Jorginho	26	Chelsea	380000.00	84
10	201956	S. Sané	27	FC Schalke 04	18500000.0	82
11	186345	K. Trippier	27	Tottenham Hotspur	18500000.0	82
12	212187	P. Max	24	FC Augsburg	11000000.0	78
13	213956	Adama	22	Wolverhampton Wanderers	10500000.0	75
14	215228	C. Lema	28	SL Benfica	4600000.0	74

- PLAYER WITH HIGH POTENTIAL AND LOWER AGE

	Name	Age	Potential
25	K. Mbappé	19	95
229	G. Donnarumma	19	93
155	O. Dembélé	21	92
1143	Vinícius Júnior	17	92
77	M. Škriniar	23	92
79	Marco Asensio	22	92
55	L. Sané	22	92
156	Gabriel Jesus	21	92
226	M. de Ligt	18	91
177	Kepa	23	91
56	Bernardo Silva	23	91
115	N. Süle	22	90
116	A. Martial	22	90
117	D. Alli	22	90
415	H. Aouar	20	90
83	Saúl	23	90
78	S. Milinković-Savić	23	90
228	Arthur	21	90
734	A. Lafont	19	90
6102	Riqui Puig	18	89
1149	R. Bentancur	21	89

- Literature Survey
- References