



Database and Analytics

Programming for Sports Analytics: Football

GROUP A TEAM – H

19241267 – POOJA MILIND RAKATE

19241607 - ANKIT RUNGTA

19241721 – PRIYANKA PRASHANT CHIMTHANKAR

20137265 – MAYURESH MOHAN LONDHE

Statement of Objective

- ▶ Implement an algorithm to select the best players for the user-defined formation and improve squad structure through a comprehensive study of the PES soccer data sets and improve the various decision-making stages of football.

Overview:

- ▶ Pro Evolution Soccer (PES)- Series of Football Simulation Video Game.
- ▶ PES Players 2020
 - Player s Demographics
 - Attacker and Midfielder Attributes
 - Defender and Goalkeeper Attributes
 - Player Ranking Position wise

Methodology:



Semi Structured Dataset Programmatically Stored in Mongo DB

- ▶ Database (PES_Players_20) created in MongoDB using Python Programming
- ▶ JSON Files loaded as collections in the created database using Python with below mentioned names.
 - Players Demographics - Player_Details
 - Attacker and Midfielder Attributes – AM_Players_attributes
 - Defender and Goalkeeper Attributes – DG_Players_Attributes
 - Player Ranking Position wise – Player_Position_Ranking
- ▶ All the collections are having a common attribute “Player_ID”

Database Structure

MongoDB Compass - localhost:27017/PES_Players_20

Connect View Help

Local

3 DBS 1 COLLECTIONS

☆ FAVORITE

HOST
localhost:27017

CLUSTER
Standalone

EDITION
MongoDB 4.4.2 Community

Filter your data

PES_Players_20

AM_Player_attributes
DG_Players_Attributes
Player_Details
Player_Position_Ranking

> admin
> config
> local

Collections

CREATE COLLECTION

Collection Name ^	Documents	Avg. Document Size	Total Document Size	Num. Indexes	Total Index Size	Properties
AM_Player_attributes	18,278	569.9 B	9.9 MB	1	176.0 KB	
DG_Players_Attributes	18,278	412.2 B	7.2 MB	1	176.0 KB	
Player_Details	18,278	723.7 B	12.6 MB	1	176.0 KB	
Player_Position_Ranking	18,278	272.6 B	4.8 MB	1	176.0 KB	

>_MongoSH Beta

Player_Details

MongoDB Compass - localhost:27017/PES_Players_20.Player_Details

Connect View Collection Help

Local

3 DBS 1 COLLECTIONS

FAVORITE

HOST
localhost:27017

CLUSTER
Standalone

EDITION
MongoDB 4.4.2 Community

Filter your data

PES_Players_20

AM_Player_attributes

DG_Players_Attributes

Player_Details

Player_Position_Ranking

admin

config

local

PES_Players_20.Player_Details

DOCUMENTS 18.3k TOTAL SIZE 12.6MB AVG. SIZE 724B INDEXES 1 TOTAL SIZE 176.0KB AVG. SIZE 176.0KB

Documents Aggregations Schema Explain Plan Indexes Validation

FILTER

ADD DATA

VIEW

Displaying documents 1 - 20 of 18278

REFRESH

#	Player_Details	_id ObjectId	Player_id Int32	short_name String	long_name String	player_url String	age Int32	
1		5ff628a919232daa366610c4	158023	"L. Messi"	"Lionel Andr��s Messi Cuccittin"	"https://sofifa.com/player/15800"	32	
2		5ff628a919232daa366610c5	20801	"Cristiano Ronaldo"	"Cristiano Ronaldo dos Santos A"	"https://sofifa.com/player/20800"	34	
3		5ff628a919232daa366610c6	190871	"Neymar Jr"	"Neymar da Silva Santos Junior"	"https://sofifa.com/player/19080"	27	
4		5ff628a919232daa366610c7	200389	"J. Oblak"	"Jan Oblak"	"https://sofifa.com/player/20030"	26	
5		5ff628a919232daa366610c8	183277	"E. Hazard"	"Eden Hazard"	"https://sofifa.com/player/18320"	28	
6		5ff628a919232daa366610c9	192985	"K. De Bruyne"	"Kevin De Bruyne"	"https://sofifa.com/player/19290"	28	
7		5ff628a919232daa366610ca	192448	"M. ter Stegen"	"Marc-Andr�� ter Stegen"	"https://sofifa.com/player/19240"	27	
8		5ff628a919232daa366610cb	203376	"V. van Dijk"	"Virgil van Dijk"	"https://sofifa.com/player/20330"	27	
9		5ff628a919232daa366610cc	177003	"L. Modri��"	"Luka Modri��"	"https://sofifa.com/player/17700"	33	
10		5ff628a919232daa366610cd	209331	"M. Salah"	"Mohamed Salah Ghaly"	"https://sofifa.com/player/20930"	27	
11		5ff628a919232daa366610ce	231747	"K. Mbapp��"	"Kylian Mbapp��"	"https://sofifa.com/player/23170"	20	
12		5ff628a919232daa366610cf	201024	"K. Koulibaly"	"Kalidou Koulibaly"	"https://sofifa.com/player/20100"	28	
13		5ff628a919232daa366610d0	202126	"H. Kane"	"Harry Kane"	"https://sofifa.com/player/20210"	25	
14		5ff628a919232daa366610d1	212831	"Alisson"	"Alisson Ramses Becker"	"https://sofifa.com/player/21280"	26	
15		5ff628a919232daa366610d2	193080	"De Gea"	"David De Gea Quintana"	"https://sofifa.com/player/19300"	28	
16		5ff628a919232daa366610d3	215914	"N. Kant��"	"N'Golo Kant��"	"https://sofifa.com/player/21590"	28	

> _MongoSH Beta

AM_Players_attributes

MongoDB Compass - localhost:27017/PES_Players_20.AM_Player_attributes

Connect View Collection Help

Local

3 DBS 1 COLLECTIONS

FAVORITE

HOST
localhost:27017

CLUSTER
Standalone

EDITION
MongoDB 4.4.2 Community

Filter your data

PES_Players_20

AM_Player_attributes

DG_Players_Attributes

Player_Details

Player_Position_Ranking

admin

config

local

PES_Players_20.AM_Player_attributes

Documents

Documents

Aggregations

Schema

Explain Plan

Indexes

Validation

DOCUMENTS 18.3k TOTAL SIZE 9.9MB AVG. SIZE 570B INDEXES 1 TOTAL SIZE 176.0KB AVG. SIZE 176.0KB

ADD DATA

VIEW

Displaying documents 1 - 20 of 18278

REFRESH

	_id ObjectId	Player_id Int32	short_name String	pace Mixed	shooting Mixed	passing Mixed	
1	5ff628ab19232daa36669f90	158023	"L. Messi"	87	92	92	
2	5ff628ab19232daa36669f91	20801	"Cristiano Ronaldo"	90	93	82	
3	5ff628ab19232daa36669f92	190871	"Neymar Jr"	91	85	87	
4	5ff628ab19232daa36669f93	200389	"J. Oblak"	null	null	null	
5	5ff628ab19232daa36669f94	183277	"E. Hazard"	91	83	86	
6	5ff628ab19232daa36669f95	192985	"K. De Bruyne"	76	86	92	
7	5ff628ab19232daa36669f96	192448	"M. ter Stegen"	null	null	null	
8	5ff628ab19232daa36669f97	203376	"V. van Dijk"	77	60	70	
9	5ff628ab19232daa36669f98	177003	"L. Modrić"	74	76	89	
10	5ff628ab19232daa36669f99	209331	"M. Salah"	93	86	81	
11	5ff628ab19232daa36669f9a	231747	"K. Mbappé"	96	84	78	
12	5ff628ab19232daa36669f9b	201024	"K. Koulibaly"	71	28	54	
13	5ff628ab19232daa36669f9c	202126	"H. Kane"	70	91	79	
14	5ff628ab19232daa36669f9d	212831	"Alisson"	null	null	null	
15	5ff628ab19232daa36669f9e	193080	"De Gea"	null	null	null	
16	5ff628ab19232daa36669f9f	215914	"N. Kanté"	78	65	77	

> _MongoSH Beta

DG_Players_Attributes

MongoDB Compass - localhost:27017/PES_Players_20.DG_Players_Attributes

Connect View Collection Help

Local

3 DBS 1 COLLECTIONS

☆ FAVORITE

HOST
localhost:27017

CLUSTER
Standalone

EDITION
MongoDB 4.4.2 Community

Filter your data

PES_Players_20

AM_Player_attributes

DG_Players_Attributes

Player_Details

Player_Position_Ranking

> admin

> config

> local

PES_Players_20.DG_Players_Attributes

DOCUMENTS 18.3k TOTAL SIZE 7.2MB AVG. SIZE 412B INDEXES

Documents Aggregations Schema Explain Plan Indexes Validation

FILTER

ADD DATA VIEW

Displaying documents 1 - 20 of 182

	DG_Players_Attributes					
	_id ObjectId	Player_id Int32	short_name String	stamina Int32	strength Int32	defending
1	5ff628ab19232daa3666e6f6	158023	"L. Messi"	75	68	39
2	5ff628ab19232daa3666e6f7	20801	"Cristiano Ronaldo"	85	78	35
3	5ff628ab19232daa3666e6f8	190871	"Neymar Jr"	81	49	32
4	5ff628ab19232daa3666e6f9	200389	"J. Oblak"	41	78	null
5	5ff628ab19232daa3666e6fa	183277	"E. Hazard"	84	63	35
6	5ff628ab19232daa3666e6fb	192985	"K. De Bruyne"	89	74	61
7	5ff628ab19232daa3666e6fc	192448	"M. ter Stegen"	35	78	null
8	5ff628ab19232daa3666e6fd	203376	"V. van Dijk"	75	92	90
9	5ff628ab19232daa3666e6fe	177003	"L. Modrić"	85	58	72
10	5ff628ab19232daa3666e6ff	209331	"M. Salah"	85	73	45
11	5ff628ab19232daa3666e700	231747	"K. Mbappé"	84	76	39
12	5ff628ab19232daa3666e701	201024	"K. Koulibaly"	73	95	89
13	5ff628ab19232daa3666e702	202126	"H. Kane"	87	84	47
14	5ff628ab19232daa3666e703	212831	"Alisson"	32	78	null
15	5ff628ab19232daa3666e704	193080	"De Gea"	43	60	null
16	5ff628ab19232daa3666e705	215914	"N. Kanté"	97	73	87

Player_Position_Ranking

MongoDB Compass - localhost:27017/PES_Players_20.Player_Position_Ranking

Connect View Collection Help

Local

3 DBS 1 COLLECTIONS

☆ FAVORITE

HOST
localhost:27017

CLUSTER
Standalone

EDITION
MongoDB 4.4.2 Community

Filter your data

PES_Players_20

- AM_Player_attributes
- DG_Players_Attributes
- Player_Details
- Player_Position_Ranking

admin

config

local

PES_Players_20.Player_Position_Ranking

DOCUMENTS 18.3k TOTAL SIZE 4.8MB AVG. SIZE 273B INDEXES 1 TOTAL SIZE 176.0KB AVG. SIZE 176.0KB

Documents Aggregations Schema Explain Plan Indexes Validation

FILTER

ADD DATA

VIEW

Displaying documents 1 - 20 of 18278

REFRESH

	_id ObjectId	Player_id Int32	short_name String	ls Mixed	st Mixed	rs Mixed	
1	5ff628aa19232daa3666582a	158023	"L. Messi"	89	89	89	
2	5ff628aa19232daa3666582b	20801	"Cristiano Ronaldo"	91	91	91	
3	5ff628aa19232daa3666582c	190871	"Neymar Jr"	84	84	84	
4	5ff628aa19232daa3666582d	208389	"J. Oblak"	null	null	null	
5	5ff628aa19232daa3666582e	183277	"E. Hazard"	83	83	83	
6	5ff628aa19232daa3666582f	192985	"K. De Bruyne"	82	82	82	
7	5ff628aa19232daa36665830	192448	"M. ter Stegen"	null	null	null	
8	5ff628aa19232daa36665831	203376	"V. van Dijk"	69	69	69	
9	5ff628aa19232daa36665832	177003	"L. Modrić"	77	77	77	
10	5ff628aa19232daa36665833	209331	"M. Salah"	84	84	84	
11	5ff628aa19232daa36665834	231747	"K. Mbappé"	86	86	86	
12	5ff628aa19232daa36665835	201024	"K. Koulibaly"	56	56	56	
13	5ff628aa19232daa36665836	202126	"H. Kane"	86	86	86	
14	5ff628aa19232daa36665837	212831	"Alisson"	null	null	null	
15	5ff628aa19232daa36665838	193000	"De Gea"	null	null	null	
16	5ff628aa19232daa36665839	215914	"N. Kanté"	72	72	72	

> MongoSH Beta

Data Pre-Processing:

- ▶ Removed insignificant columns from all datasets like Id, Player Tags, Team Jersey Number, Player URL, etc.
- ▶ Handling Missing Data
 - Removed Nation Jersey Number column as it had 90% Null Values as many players just play for clubs and do not play for their nations.
 - Replaced null values with zero wherever player not eligible for that specific attribute. For example, Gk Speed Column attribute determines speed for goalkeepers and hence Players other than Goalkeepers will not have value assigned to this.
 - Imputed null values using mean value of the column Release Clause Euro.

Structured Data storage in PostgreSQL

- ▶ Pre processed structured data stored in the PostgreSQL using Python with below mentioned table names.
 - Players Demographics - Player_Details
 - Attacker and Midfielder Attributes – AM_Players_attributes
 - Defender and Goalkeeper Attributes – DG_Players_Attributes
 - Player Ranking Position wise – Player_Position_Ranking

Browser



Servers (1)

PostgreSQL 13

Databases (2)

pes_players_2020

Casts

Catalogs

Event Triggers

Extensions

Foreign Data Wrappers

Languages

Schemas (1)

public

Collations

Domains

FTS Configurations

FTS Dictionaries

FTS Parsers

FTS Templates

Foreign Tables

Functions

Materialized Views

Procedures

Sequences

Tables (4)

AM_Player_Attributes

DG_Player_Attributes

Player_Details

Player_Position

Trigger Functions

Types

Dashboard

Properties

SQL

Statistics

Dependencies

Dependents

public.Player_Details/pes_players_2020/postgres@PostgreSQL 13



public.Player_Details/pes_players_2020/postgres@PostgreSQL 13

Query Editor

Query History

Scratch Pad



```
1 SELECT * FROM public."Player_Details"  
2 ORDER BY "Player_id" ASC LIMIT 100  
3
```

Data Output

Explain

Messages

Notifications

	Player_id [PK] bigint	short_name text	long_name text	age bigint	dob text	height_cm bigint	weight_kg bigint	nationality text	club text	overall bigint	potential bigint	value_eur bigint	wage_eu bigint
1	768	Felipe	Felipe Dias da S...	34	31-07-...	188	75	Brazil	SPAL	74	74	2100000	
2	1179	G. Buffon	Gianluigi Buffon	41	28-01-...	192	92	Italy	Juvent...	83	83	2600000	
3	2147	M. Stekelenburg	Maarten Stekel...	36	22-09-...	197	92	Netherlands	Everton	73	73	675000	
4	2335	A. Wilbraham	Aaron Wilbraham	39	21-10-...	191	80	England	Rochd...	62	62	90000	
5	2702	K. Ellison	Kevin Ellison	40	23-02-...	180	76	England	Morec...	58	58	0	
6	3281	Tarantini	Ricardo JosÃ© ...	35	07-10-...	188	78	Portugal	Rio Ave...	74	74	2400000	
7	3467	Yuri	Yuri de Souza F...	36	08-08-...	186	78	Brazil	SD Pon...	69	69	375000	
8	5479	Casillas	Iker Casillas Fer...	38	20-05-...	185	84	Spain	FC Porto	83	83	2600000	

- ▼ Databases (2)
 - ▼ pes_players_2020
 - > Casts
 - > Catalogs
 - > Event Triggers
 - > Extensions
 - > Foreign Data Wrappers
 - > Languages
 - ▼ Schemas (1)
 - ▼ public
 - > Collations
 - > Domains
 - > FTS Configurations
 - > FTS Dictionaries
 - > FTS Parsers
 - > FTS Templates
 - > Foreign Tables
 - > Functions
 - > Materialized Views
 - > Procedures
 - > Sequences
 - ▼ Tables (4)
 - > AM_Player_Attributes
 - > DG_Player_Attributes
 - > Player_Details
 - > Player_Position
 - > Trigger Functions
 - > Types
 - > Views

Type	Name	Restriction
Primary Key	public.Player_Details_pkey	auto
Foreign Key	public.AM_Player_Attributes.AM_Player_Attributes_Player_id_fkey	normal
Foreign Key	public.DG_Player_Attributes.DG_Player_Attributes_Player_id_fkey	normal
Foreign Key	public.Player_Position.Player_Position_Player_id_fkey	normal

Result of Implemented Algorithm

-Best Team Selection

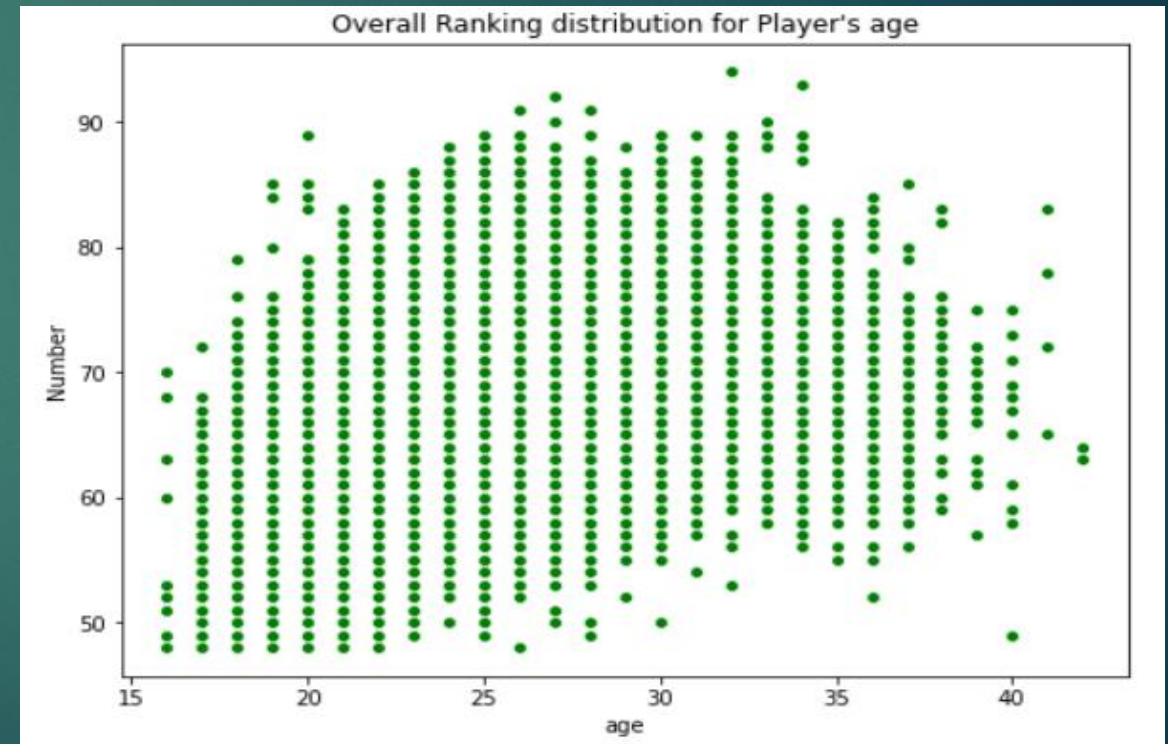
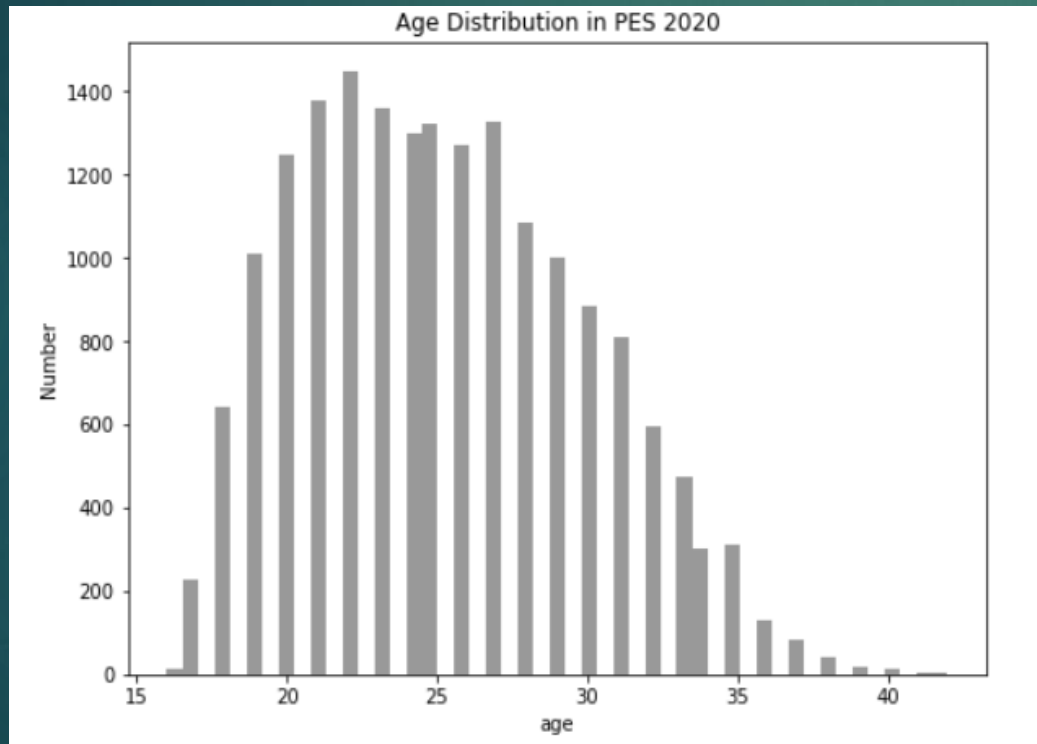
4-3-3			
	Position	Player	Overall
0	GK	J. Oblak	91
1	LB	Jordi Alba	87
2	CB	S. de Vrij	84
3	CB	F. Acerbi	83
4	RB	J. Kimmich	86
5	LM	P. Aubameyang	88
6	CDM	Sergio Busquets	89
7	RM	C. Eriksen	88
8	LW	Cristiano Ronaldo	93
9	ST	H. Kane	89
10	RW	L. Messi	94

4-3-3			
Team Selection Based on Club : Real Madrid			
	Position	Player	Overall
0	GK	T. Courtois	88
1	LB	Marcelo	85
2	RB	Carvajal	85
3	LCB	Sergio Ramos	89
4	RCB	R. Varane	85
5	LCM	T. Kroos	88
6	CDM	Casemiro	87
7	RCM	L. Modrić	90
8	LW	E. Hazard	91
9	CF	K. Benzema	87
10	RW	Vinícius Jr.	79

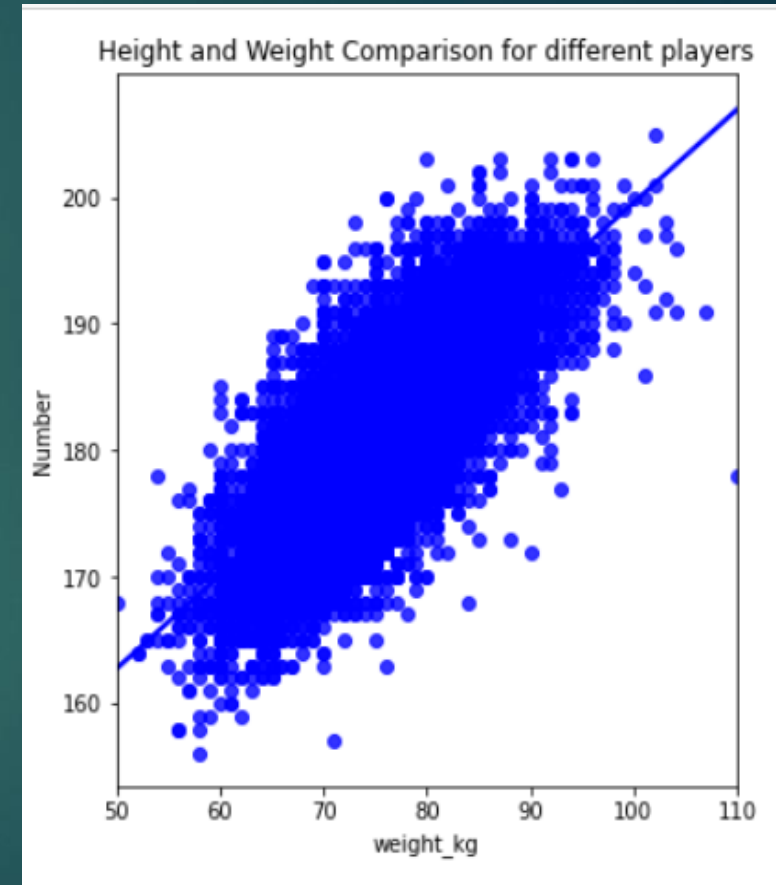
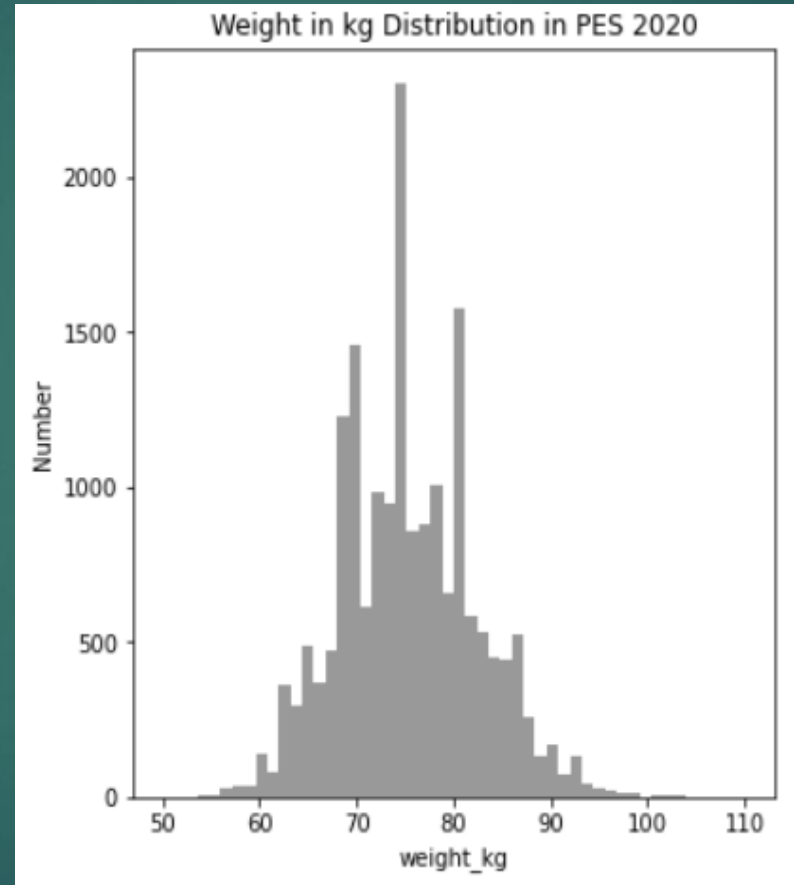
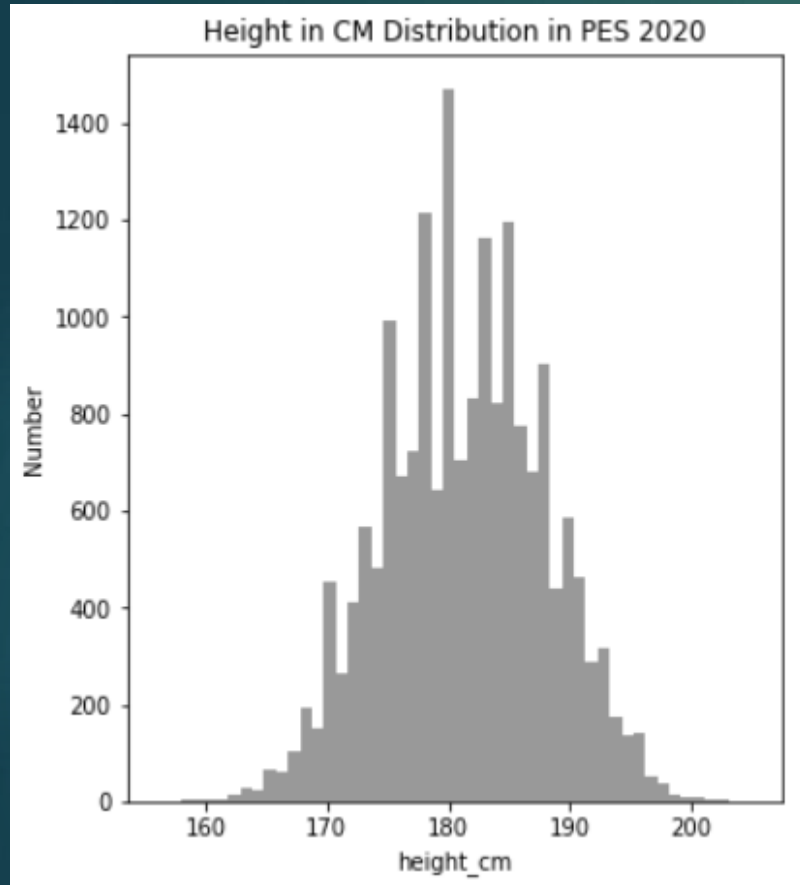
Team Selection Based on Nation : Spain			
	Position	Player	Overall
0	GK	De Gea	89
1	LB	Jordi Alba	87
2	RB	Carvajal	85
3	LCB	Sergio Ramos	89
4	RCB	Piqué	88
5	LCM	David Silva	88
6	CDM	Sergio Busquets	89
7	RCM	Parejo	86
8	LW	Oyarzabal	82
9	ST	Borja Iglesias	83
10	RW	Iñaki Williams	82

Analysis and Visualization of Data

- Age Distribution and overall rating of all the PES players:

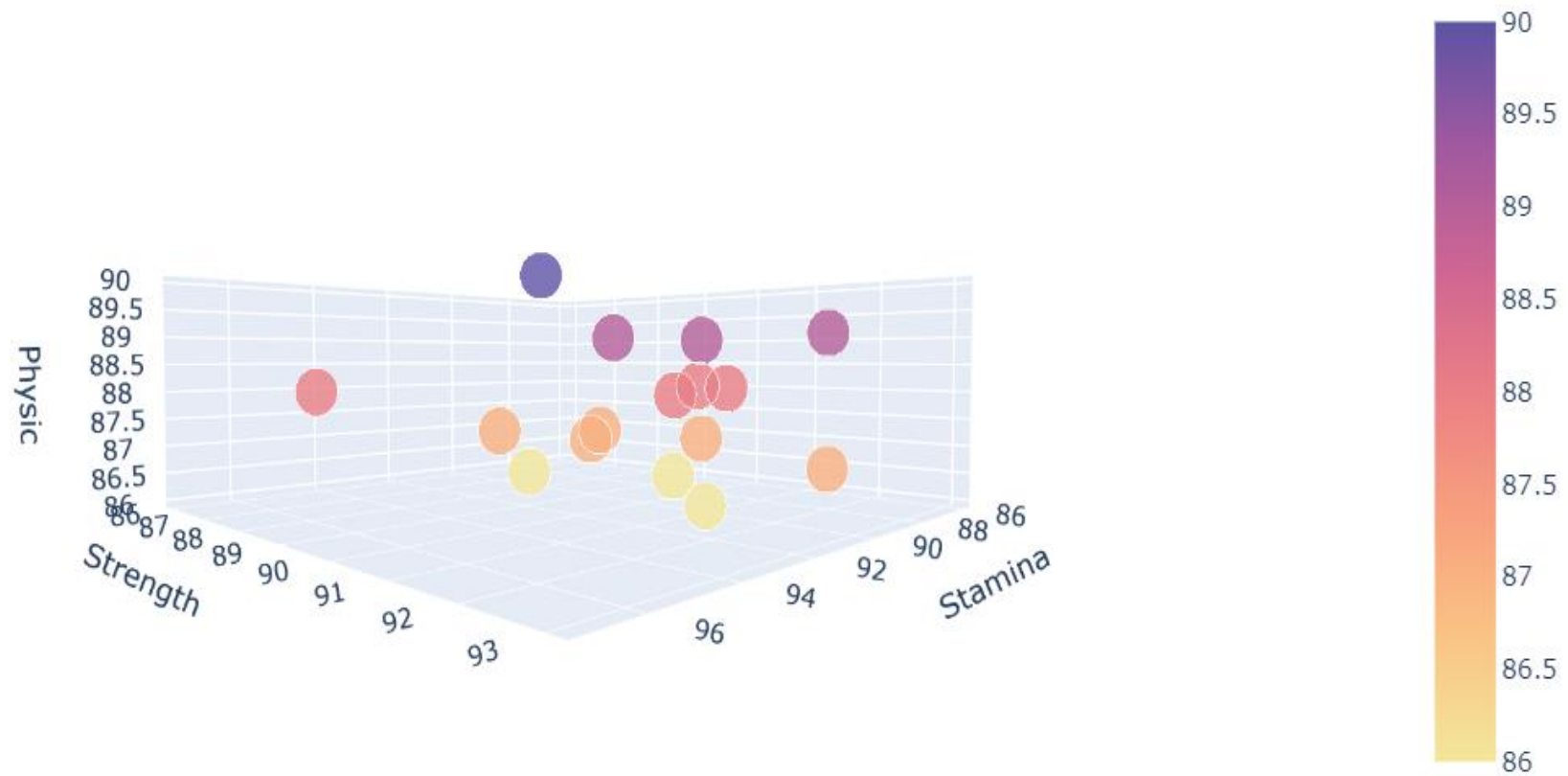


Fitness of the PES Players:

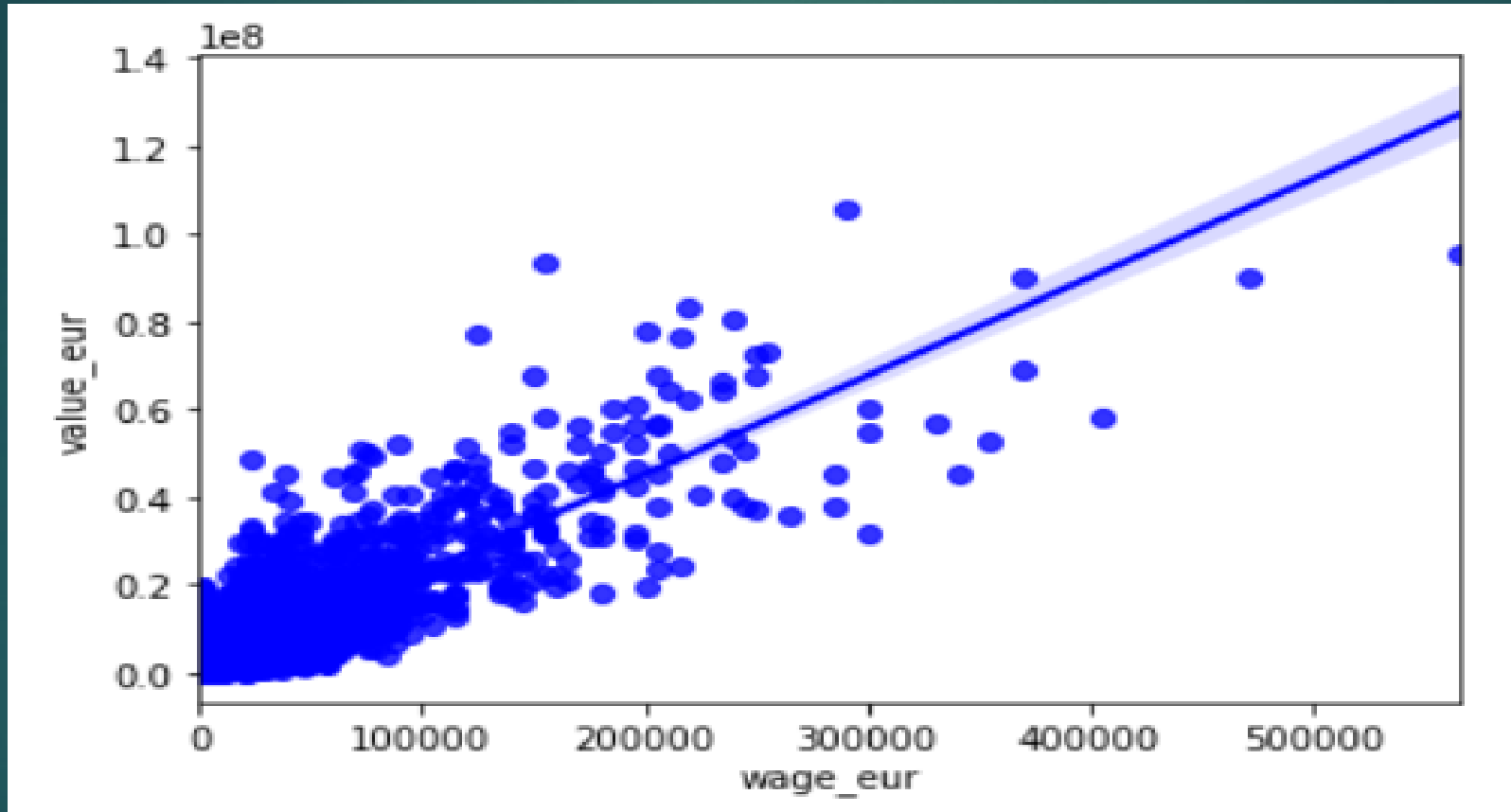


Physics, Strength, Stamina:

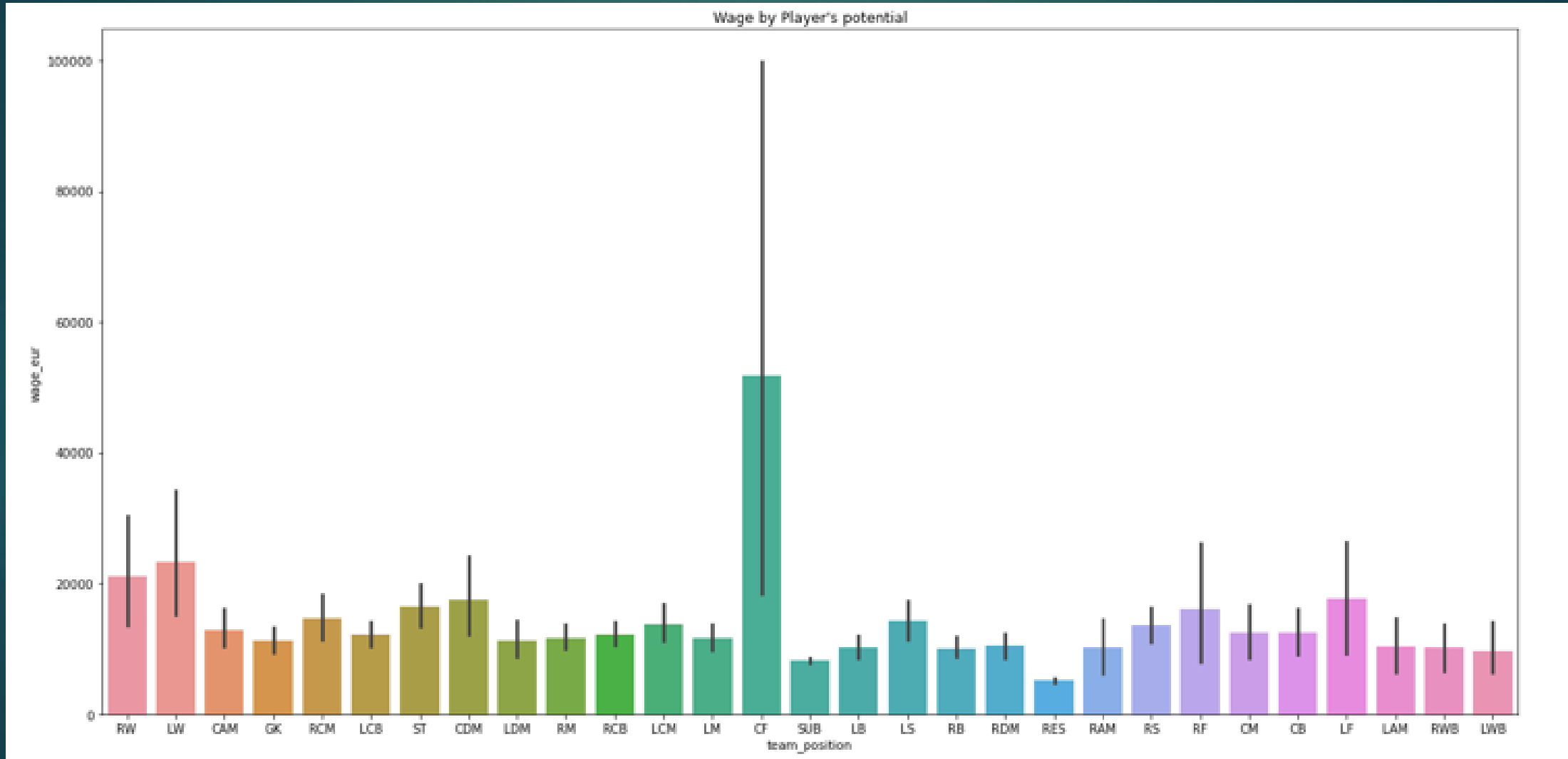
Players with Stamina, Strength and Physic greater than 85



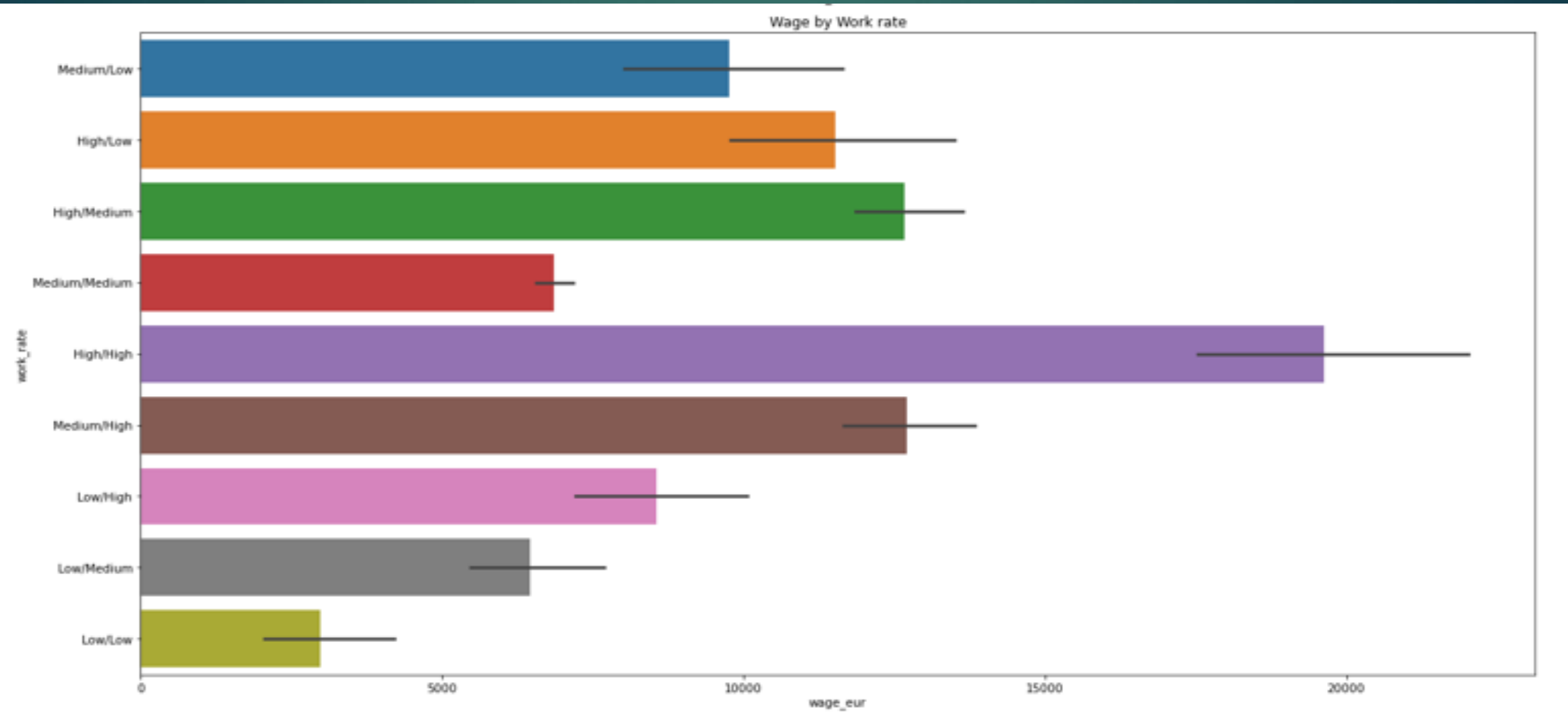
Wage and Value of Player:



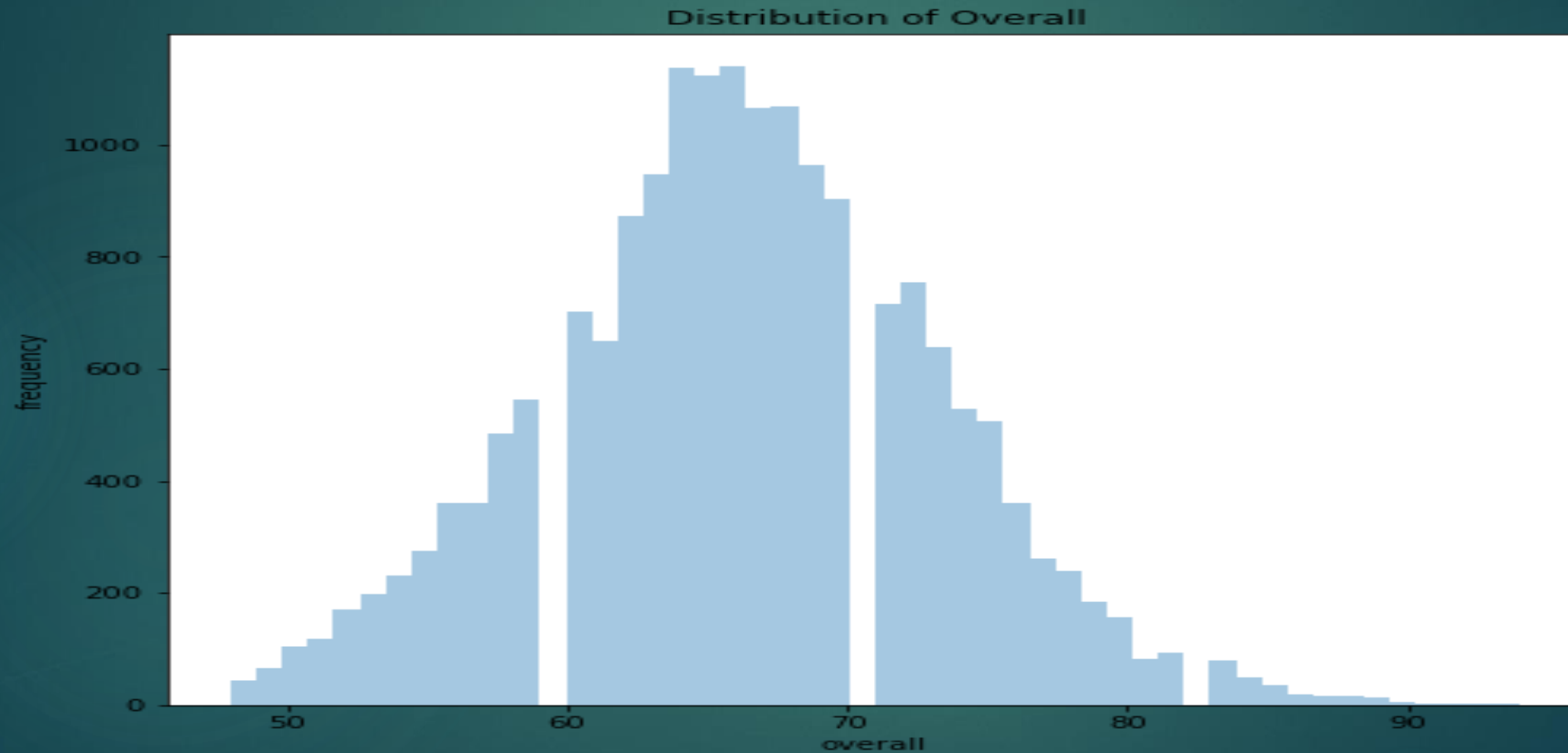
Average Wage by Player Position:



Average Wage by Work Load:

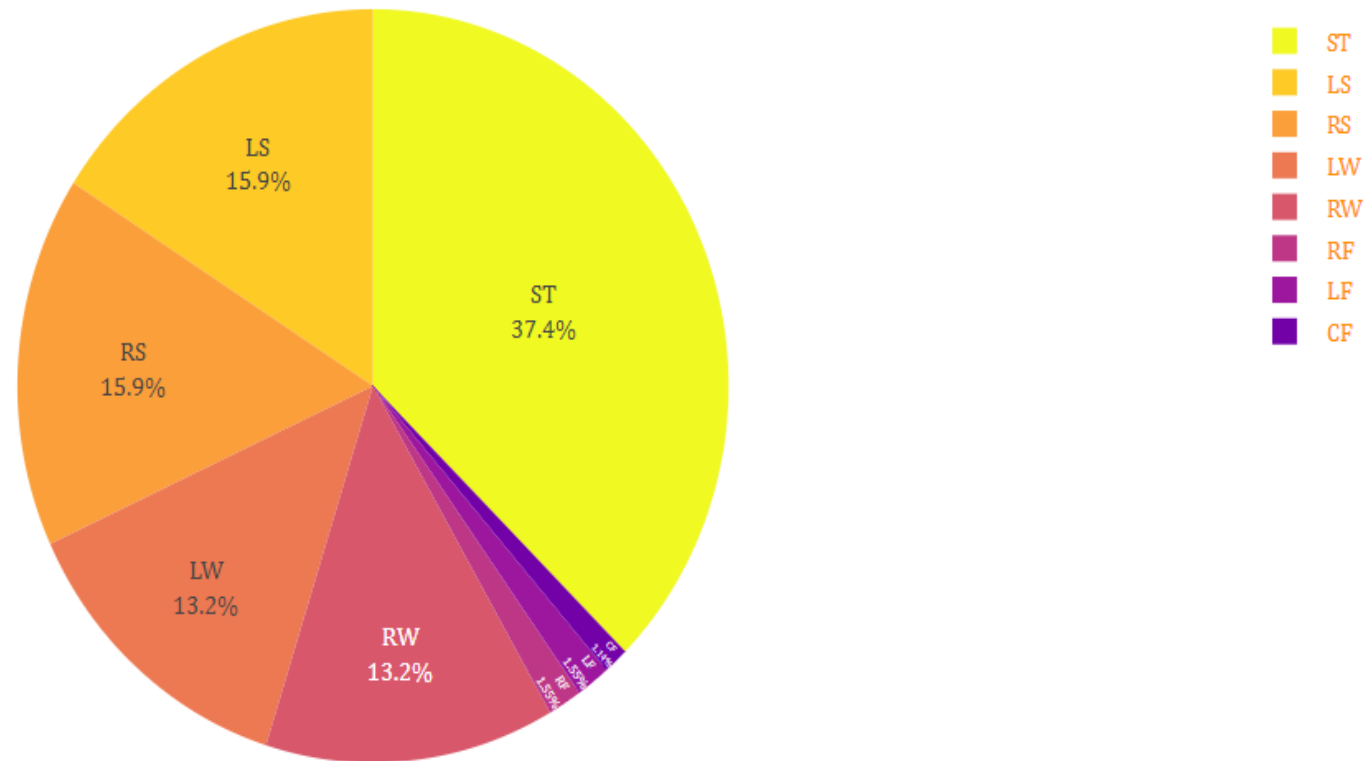


Player distribution according to overall rating:



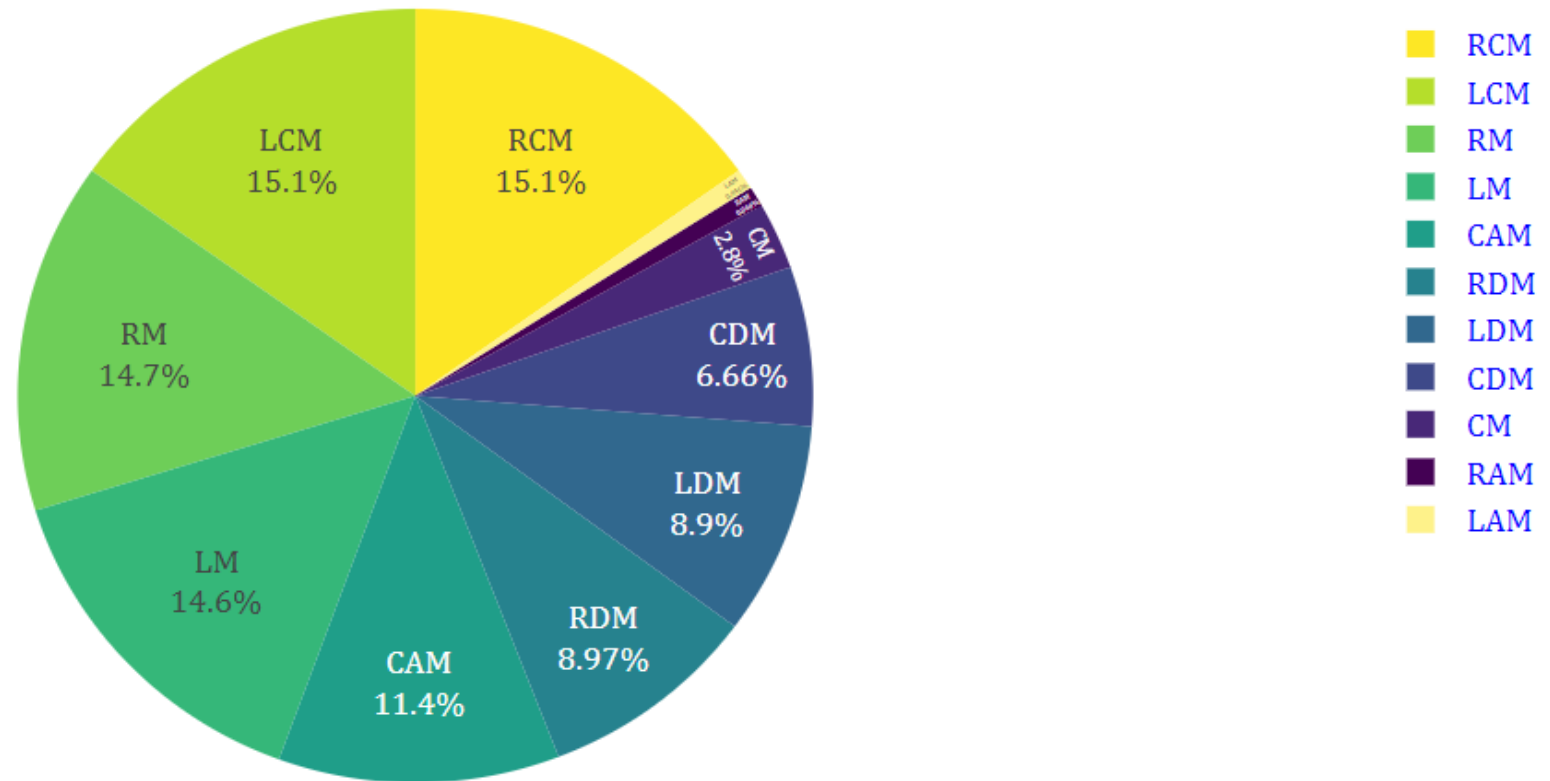
Attackers Distribution:

Percentage of players in Attacker Role



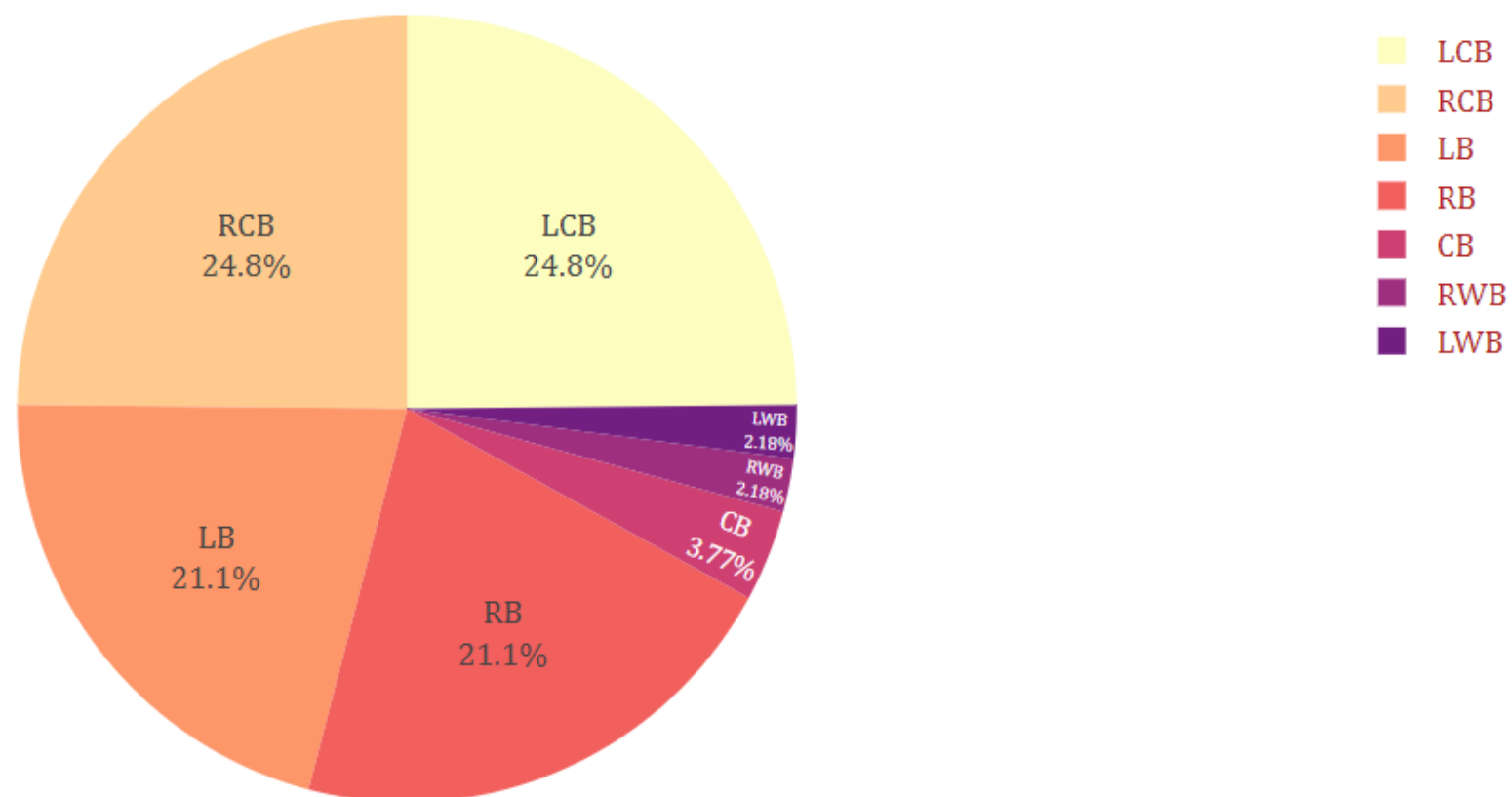
Midfielder Distribution:

Percentage of players in Different Midfielder Role



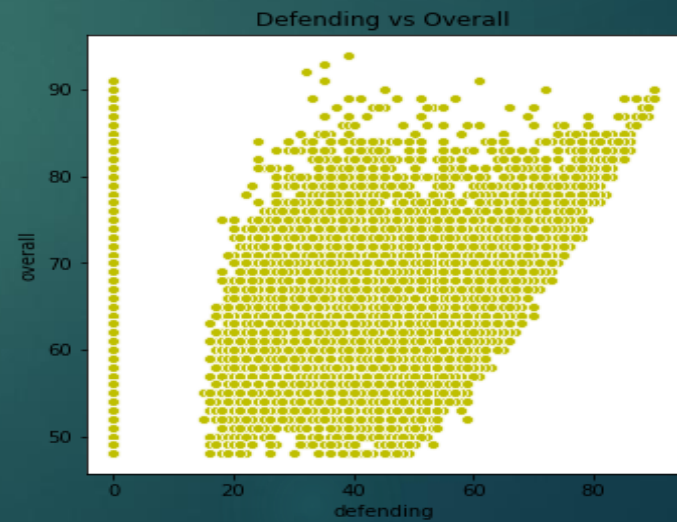
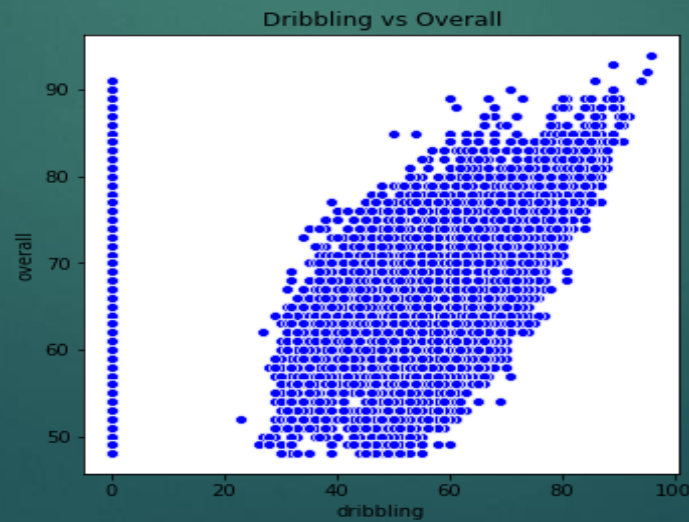
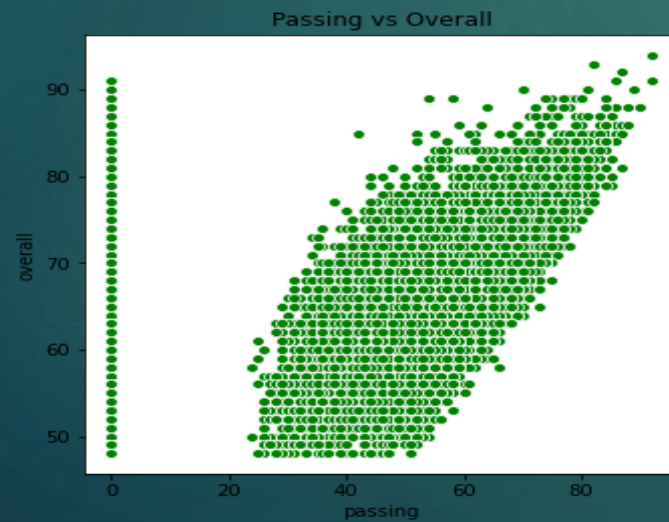
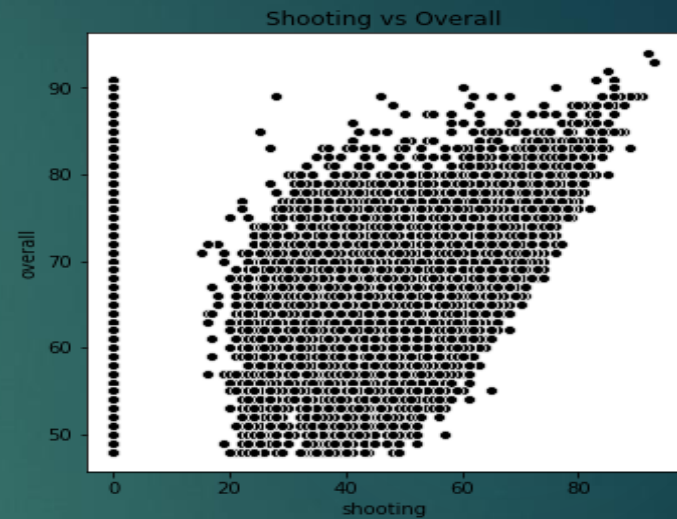
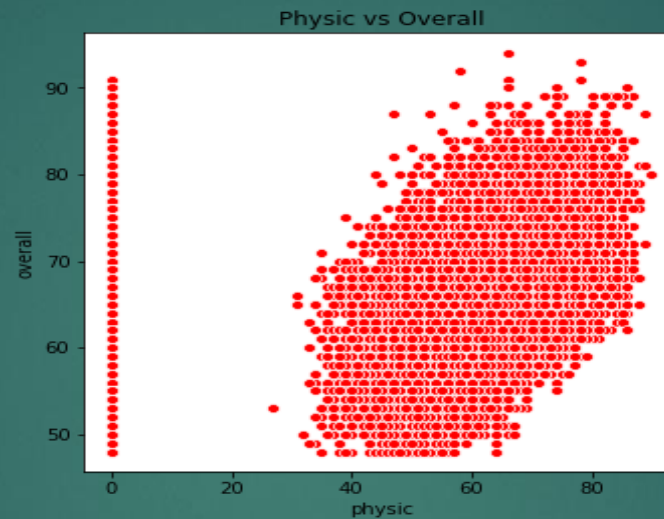
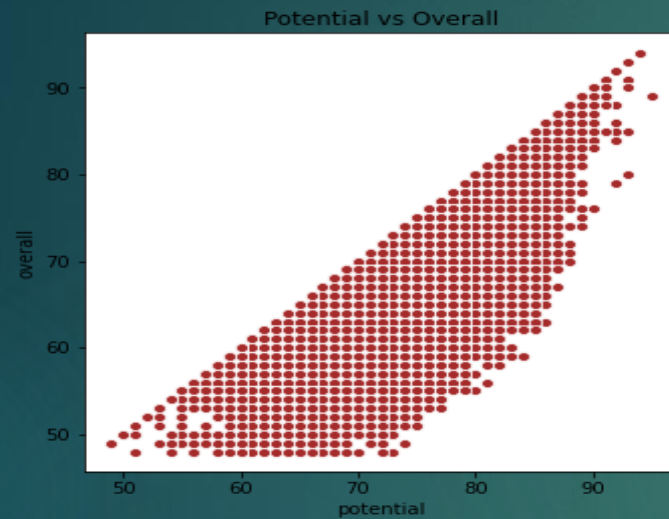
Defender Distribution:

Percentage of Players in Different Defender Role



Attributes of a player against Overall Rating:

Different Attributes of Player in Relation with Overall



International Reputation:

	short_name	age	club	nationality	overall
0	L. Messi	32	FC Barcelona	Argentina	94
1	Cristiano Ronaldo	34	Juventus	Portugal	93
2	Neymar Jr	27	Paris Saint-Germain	Brazil	92
19	L. Suárez	32	FC Barcelona	Uruguay	89
31	M. Neuer	33	FC Bayern München	Germany	88
96	Z. Ibrahimović	37	LA Galaxy	Sweden	85



Thank You !!!