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

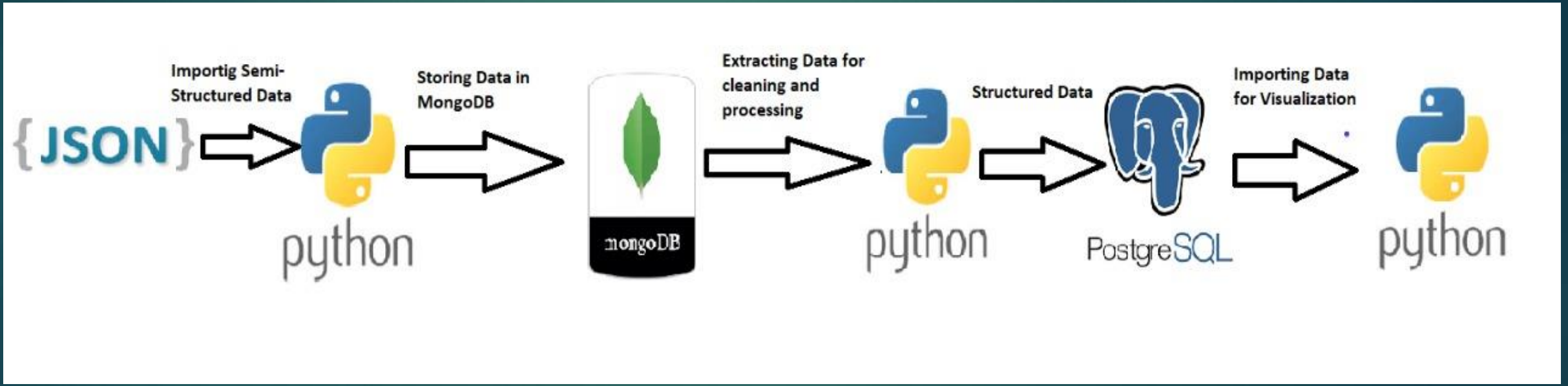
Statement of Objective

- 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

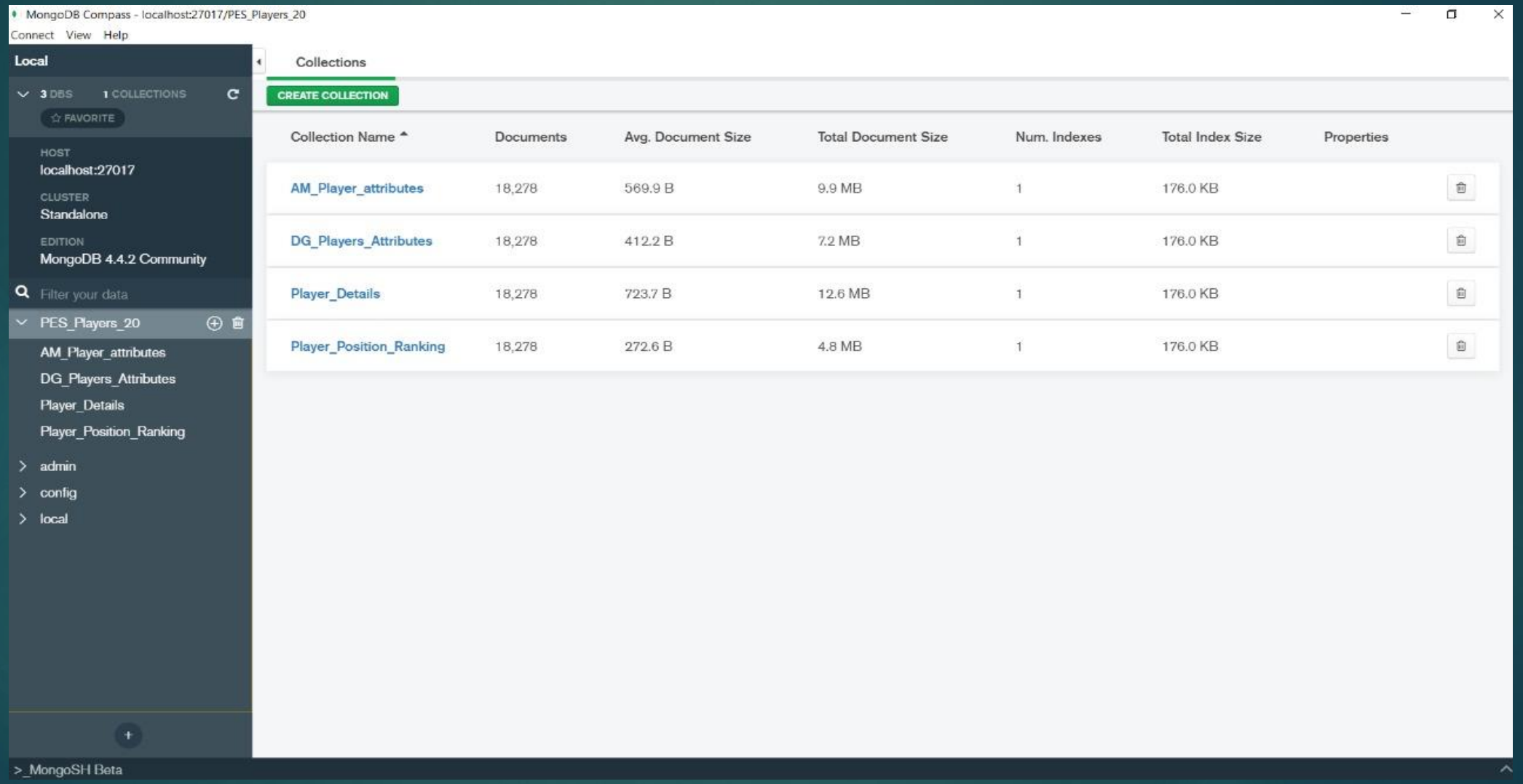
Methodology:



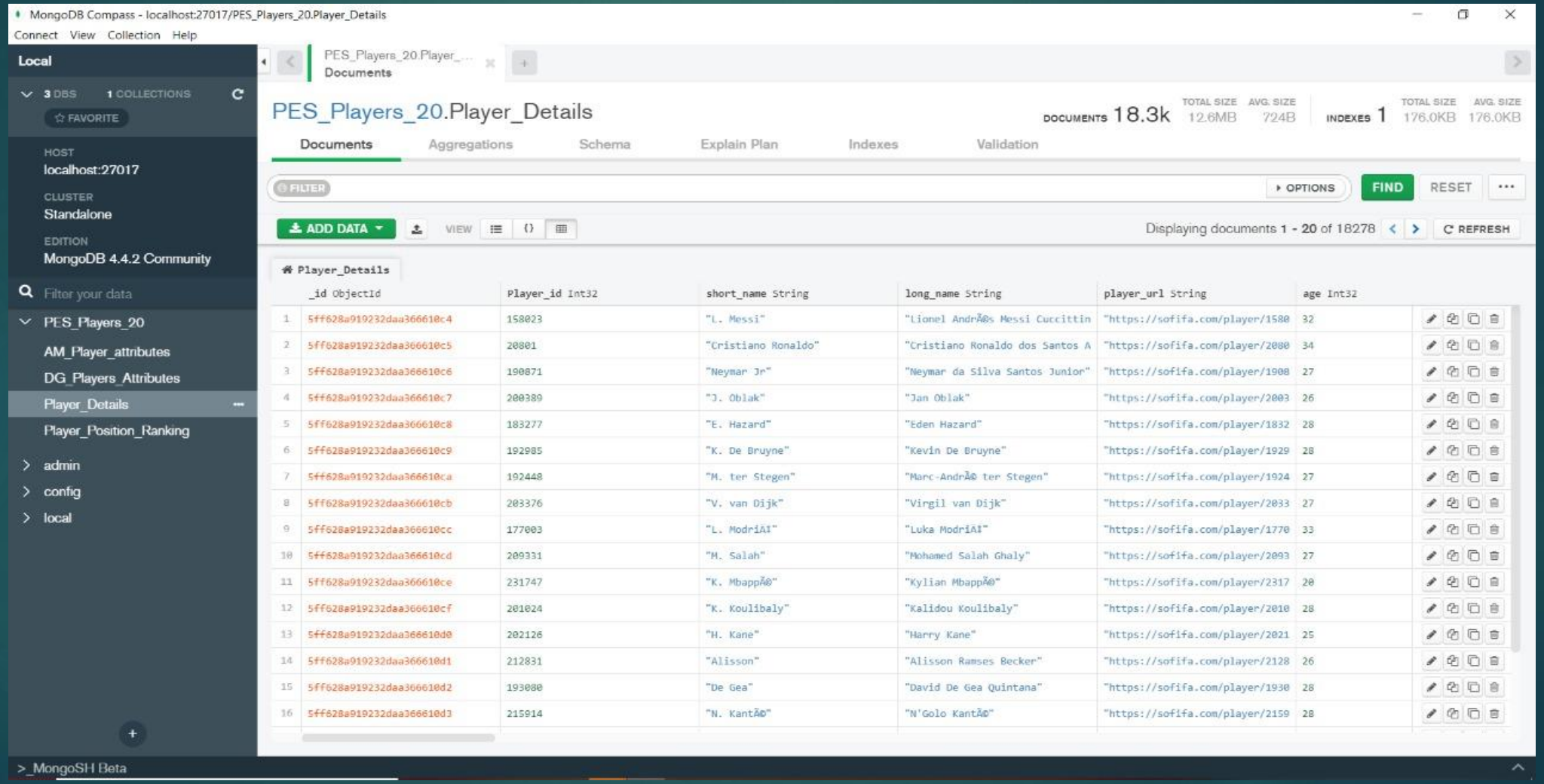
Semi Structured Dataset Programmatically Stored in Mongo DB

- 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”

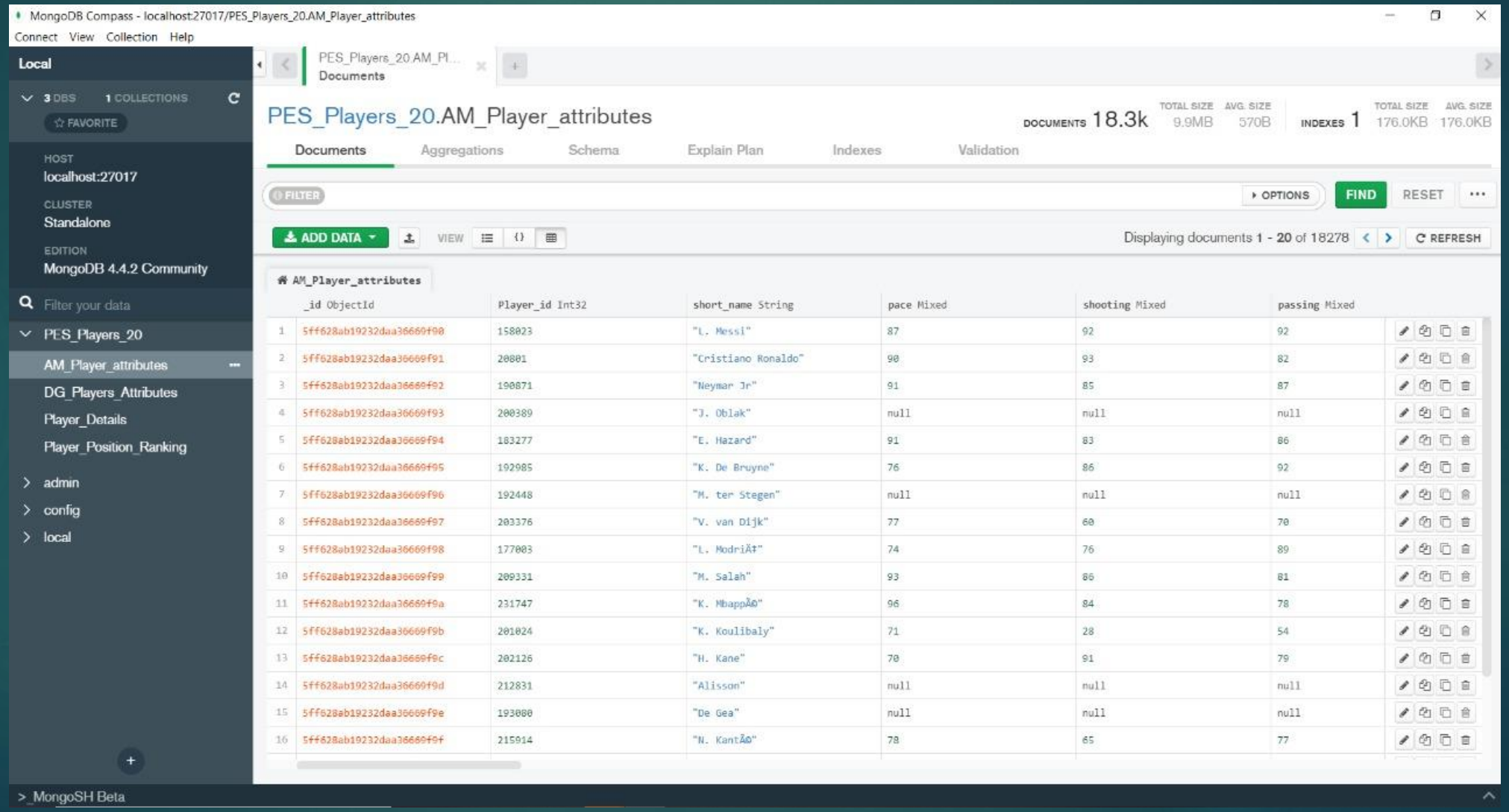
Database Structure



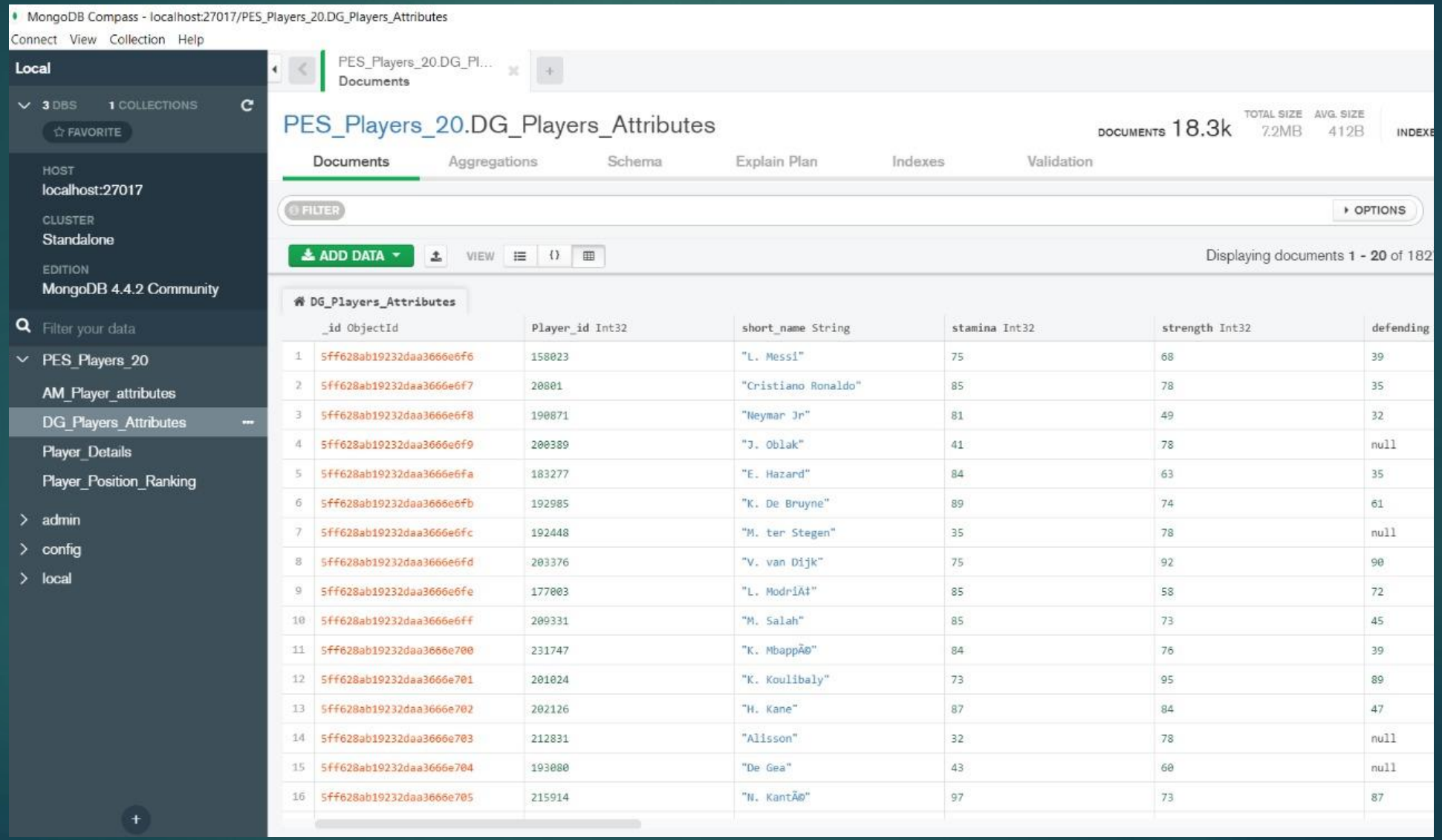
Player_Details



AM_Players_attributes

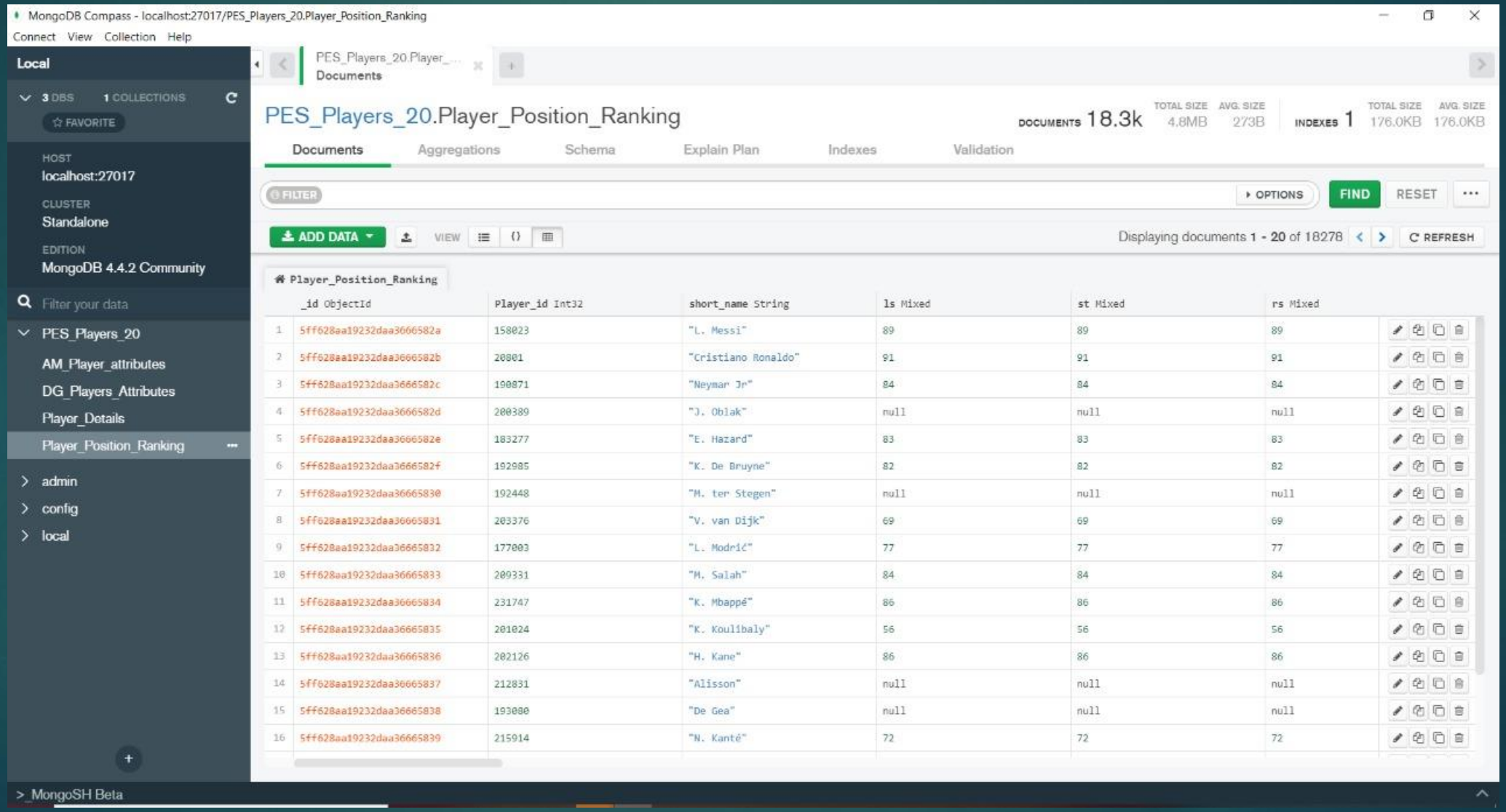


DG_Players_Attributes



	_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. ModriÄ‡"	85	58	72
10	5ff628ab19232daa3666e6ff	209331	"M. Salah"	85	73	45
11	5ff628ab19232daa3666e700	231747	"K. MbappÃ©"	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. KantÃ©"	97	73	87

Player_Position_Ranking

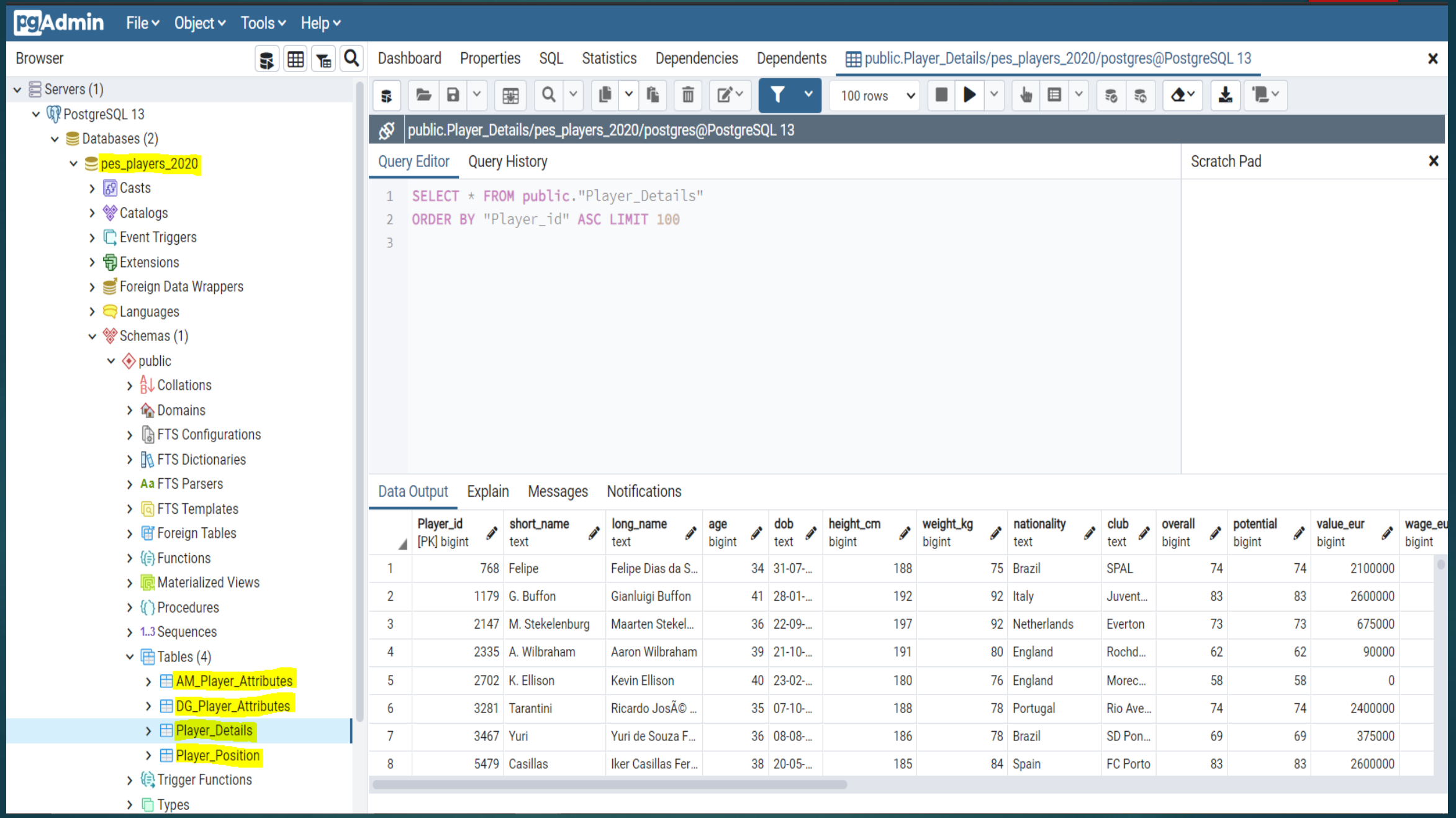


Data Pre-Processing:

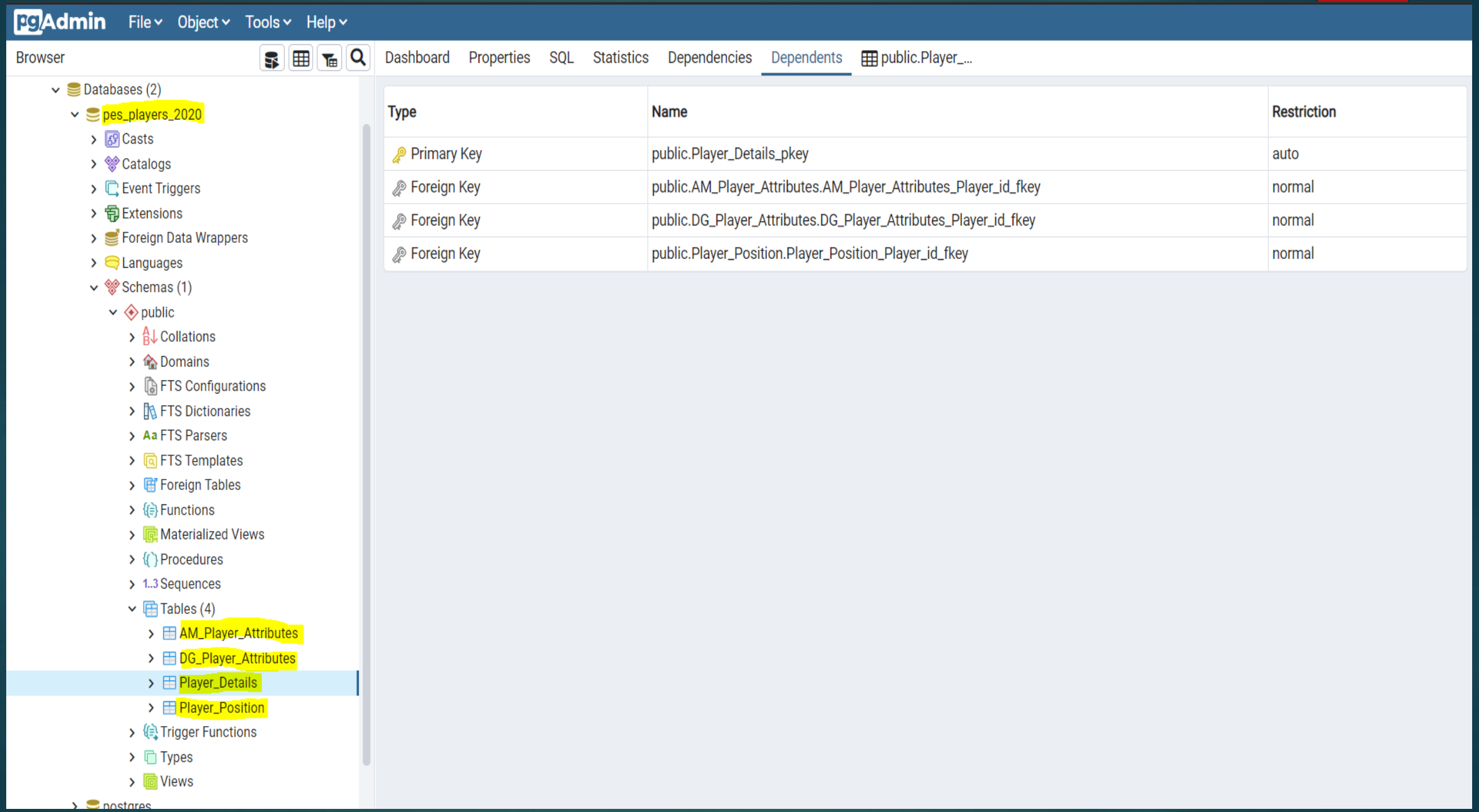
- Removed insignificant columns from all datasets like Id, Player Tags, Team Jersey Number, Player URL, etc.
- Handling Missing Data
  - Removed Nation Jersey Number column as it had 90% Null Values as many players just play for clubs and do not play for their nations.
  - Replaced null values with zero 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 postreSQL 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



	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	


pgAdmin
File
Object
Tools
Help
Browser
Dashboard
Properties
SQL
Statistics
Dependencies
Dependents
public.Player_...
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
postgres
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

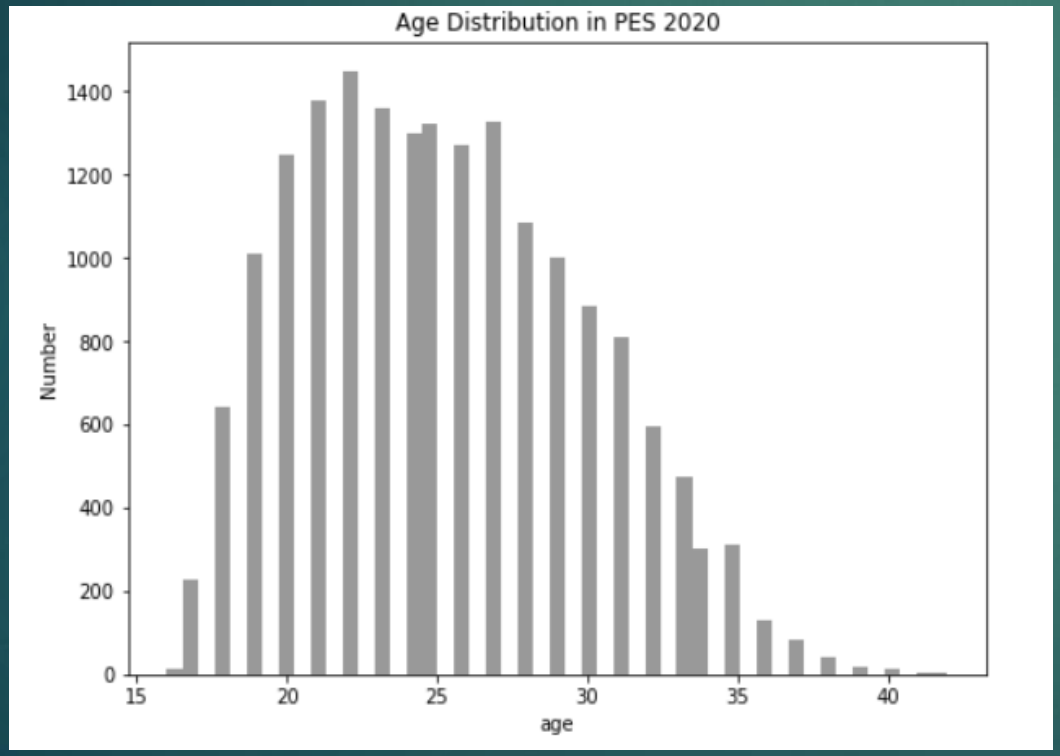
	Position	Player	Overall
0	GK	T. Courtois	88
1	LB	Marcelo	85
2	RB	Carvajal	85
3	LCB	Sergio Ramos	89
4	RCB	R. Varane	85
5	LCM	T. Kroos	88
6	CDM	Casemiro	87
7	RCM	L. ModriÄ‡	90
8	LW	E. Hazard	91
9	CF	K. Benzema	87
10	RW	VinÃcius Jr.	79

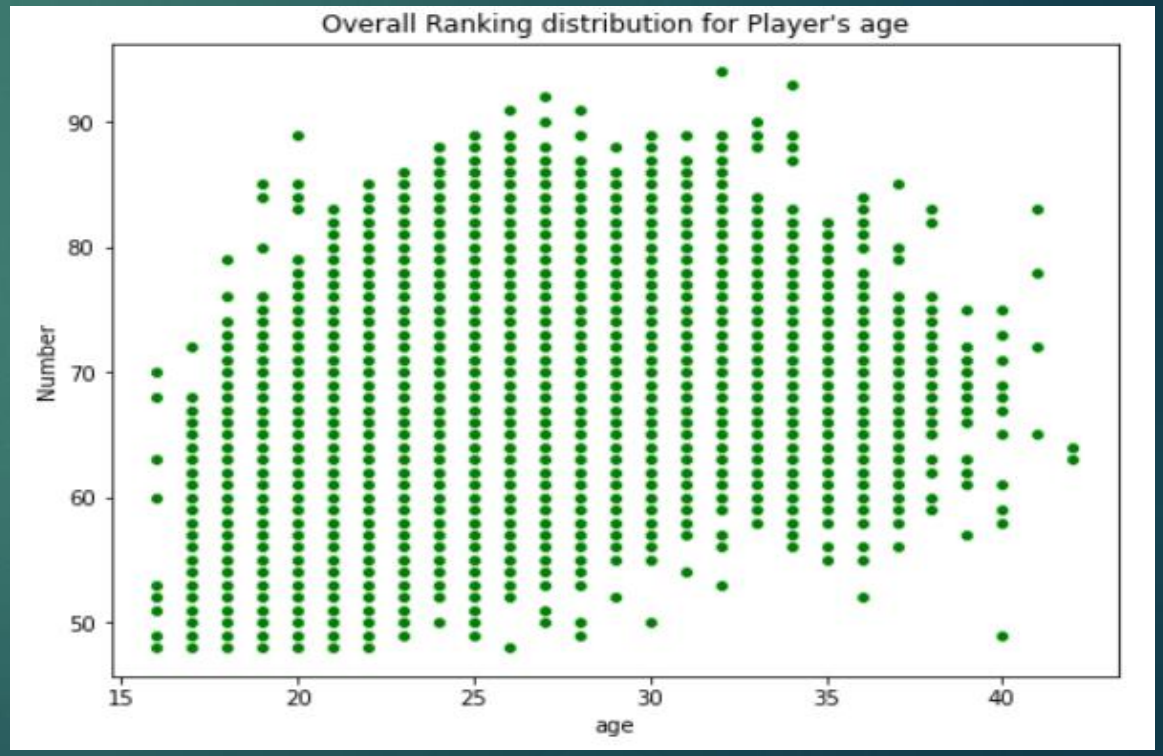
Team Selection Based on Nation : Spain

	Position	Player	Overall
0	GK	De Gea	89
1	LB	Jordi Alba	87
2	RB	Carvajal	85
3	LCB	Sergio Ramos	89
4	RCB	PiquÃ©	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	IÃ±aki Williams	82

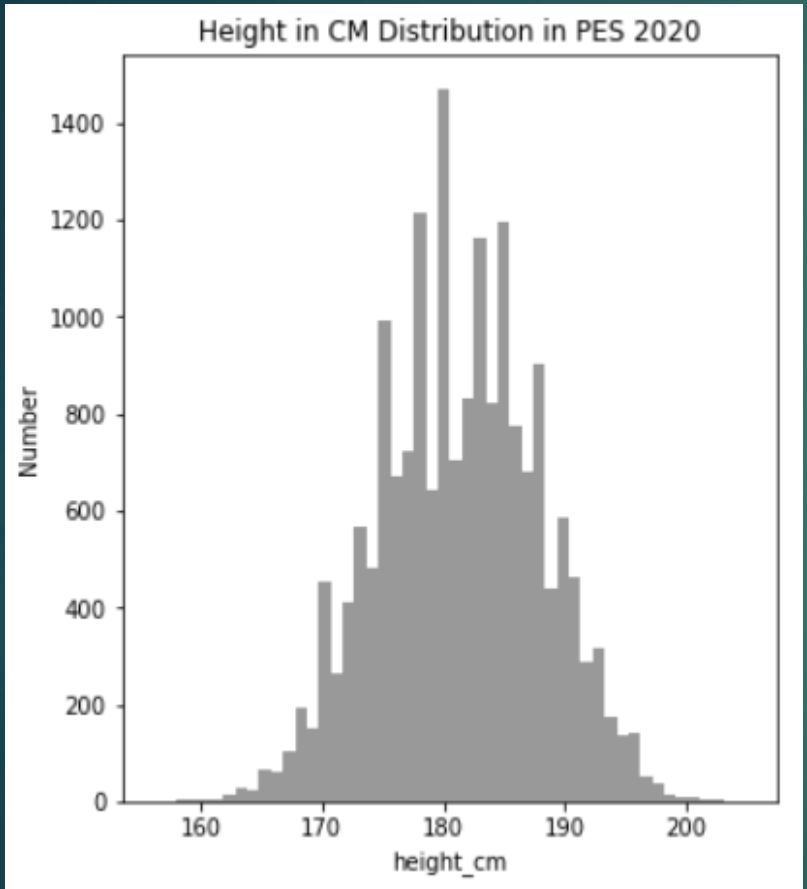
Analysis and Visualization of Data

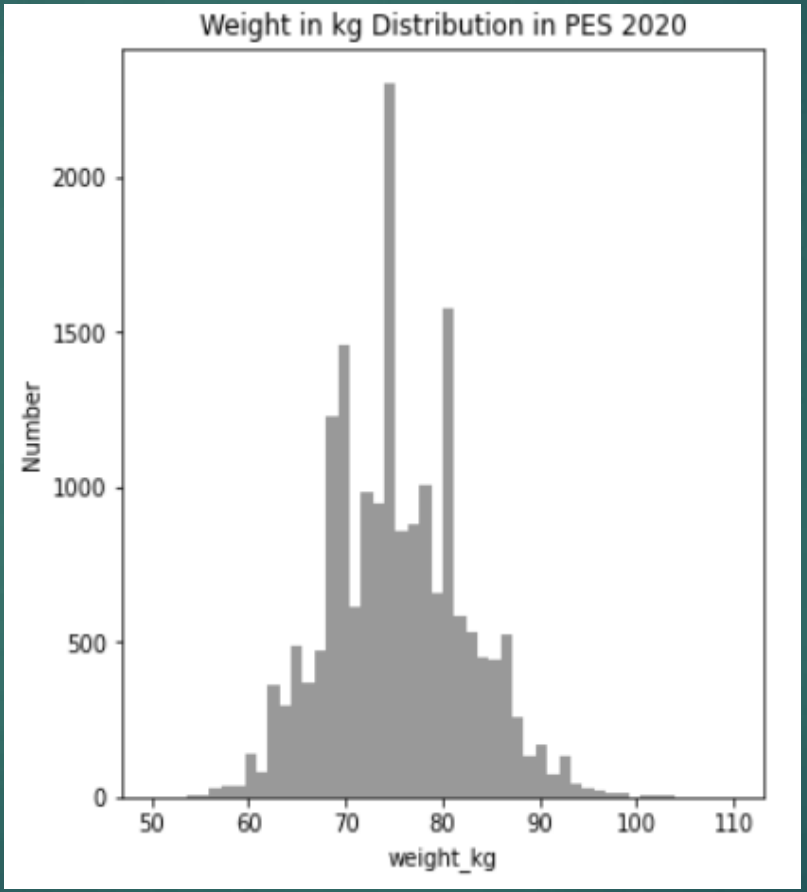
- Age Distribution and overall rating of all the PES players:

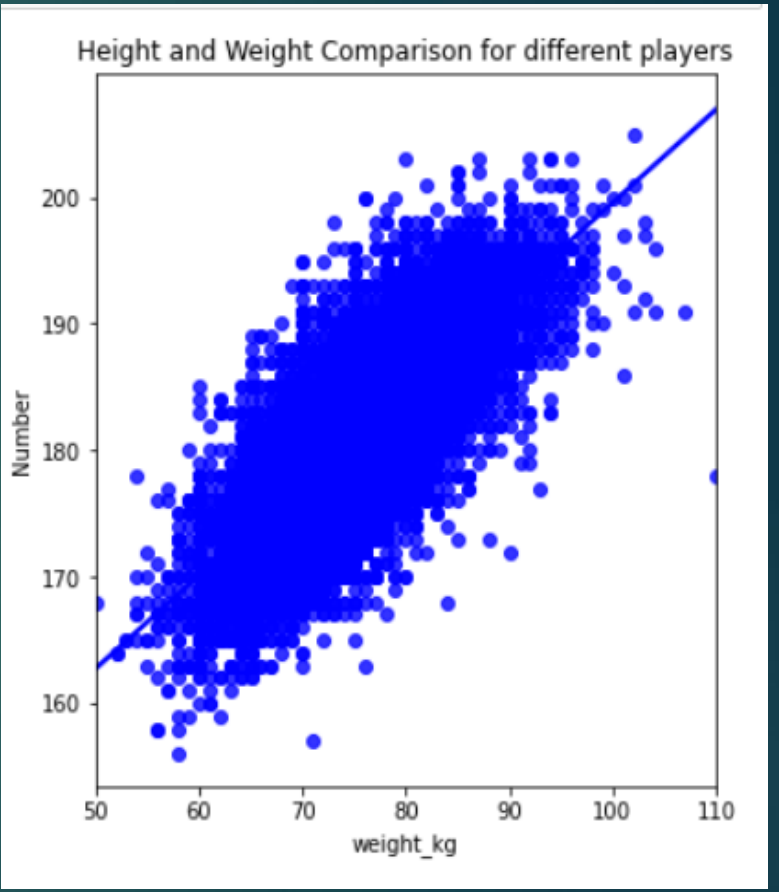




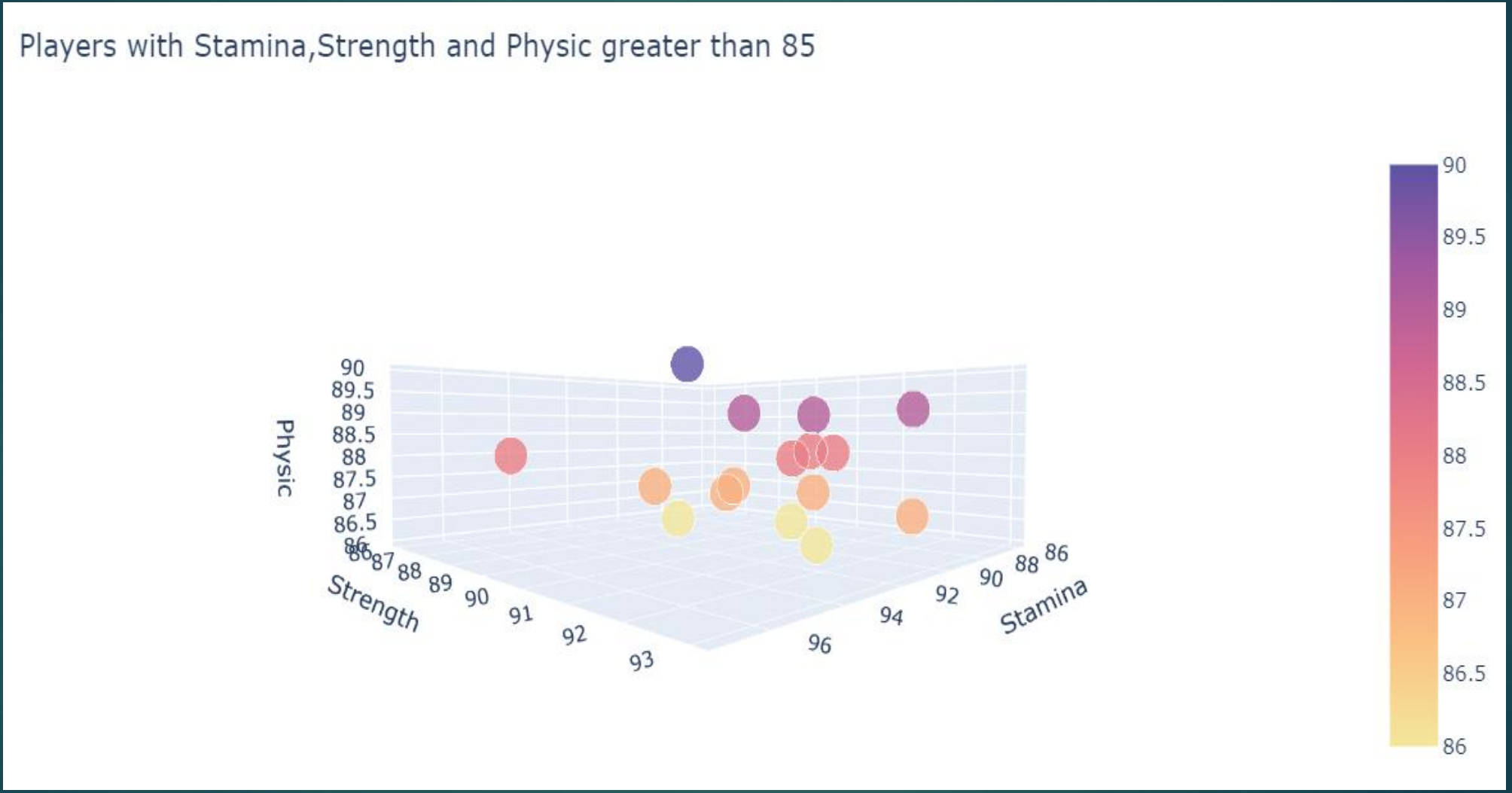
Fitness of the PES Players:



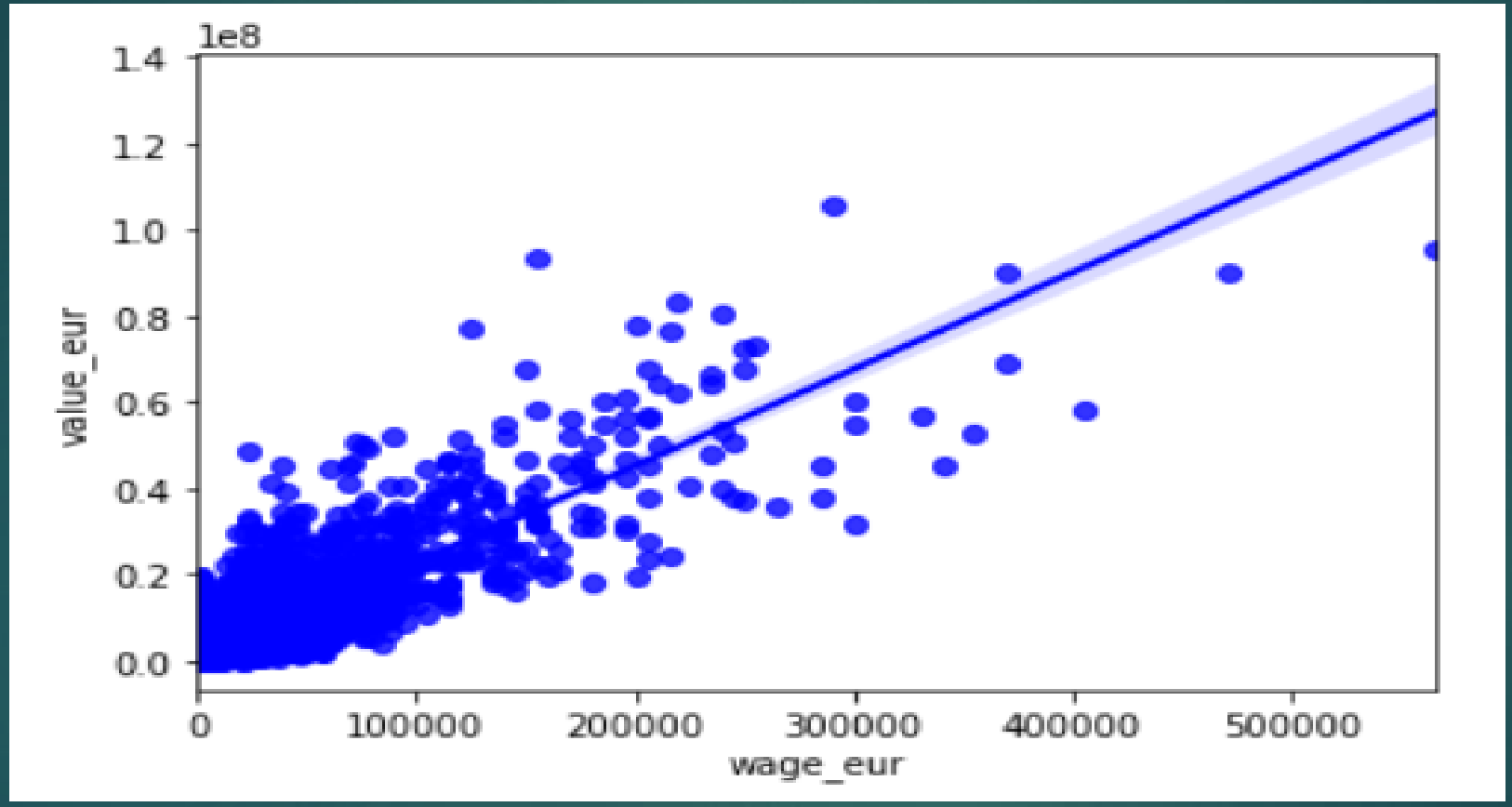




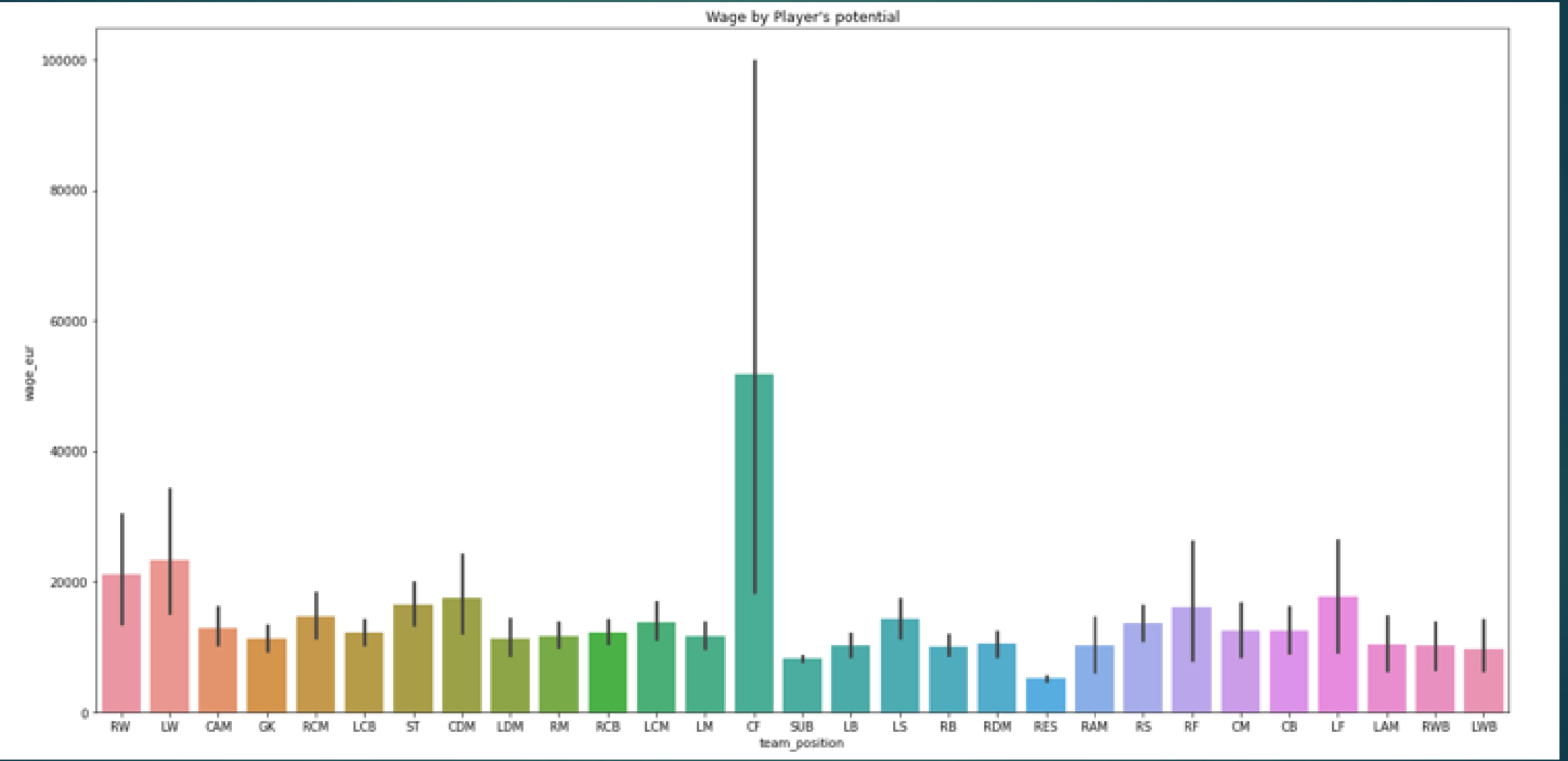
Physics, Strength, Stamina:



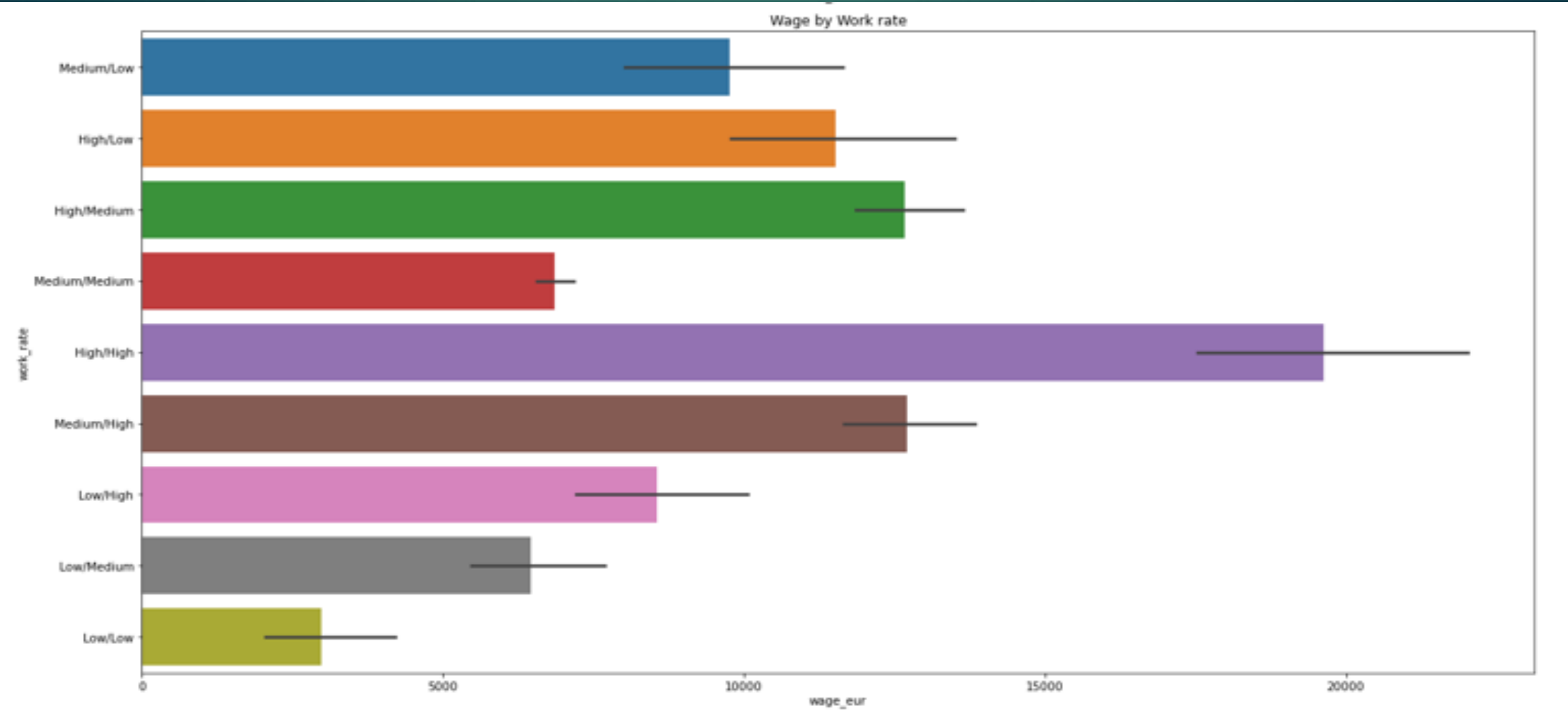
Wage and Value of Player:



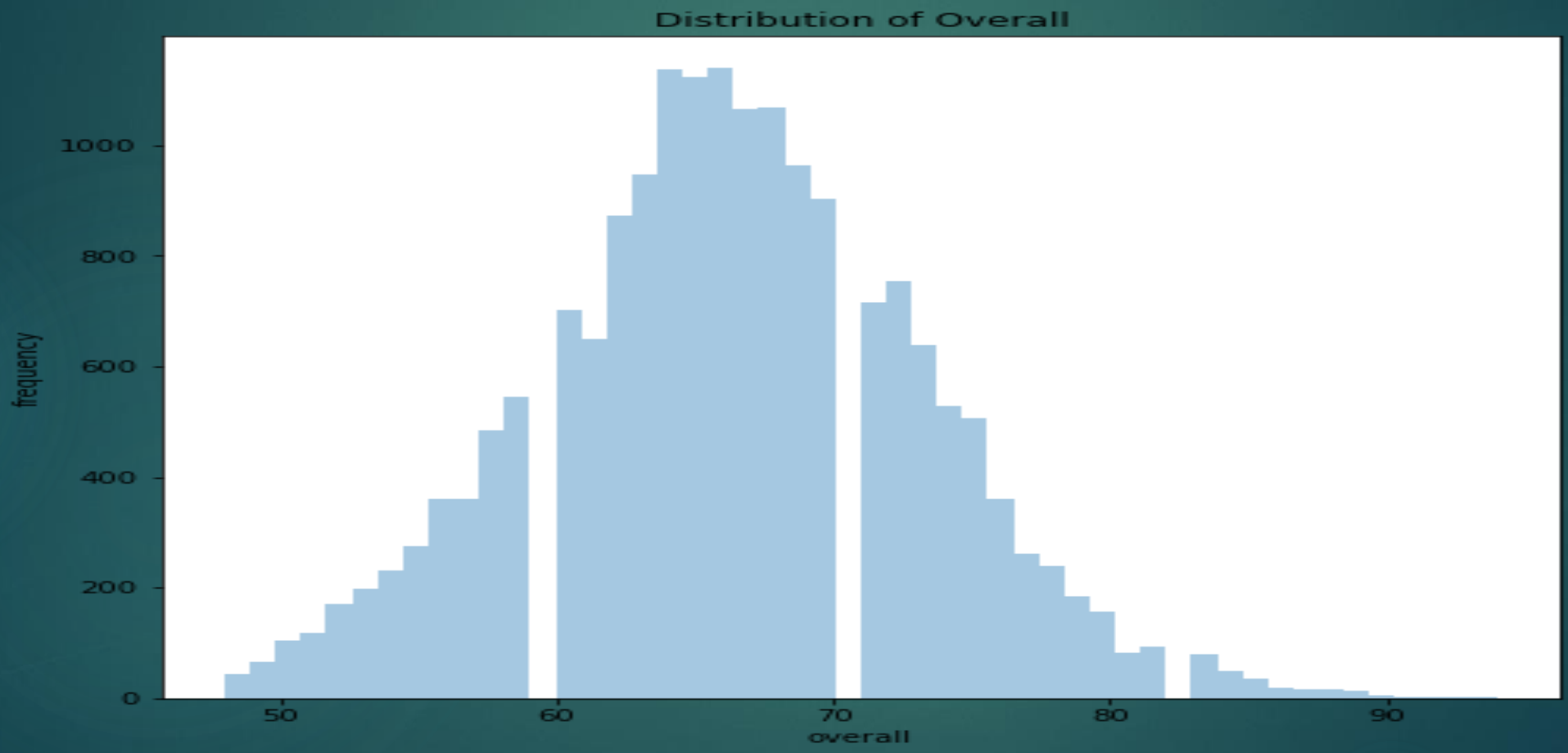
Average Wage by Player Position:



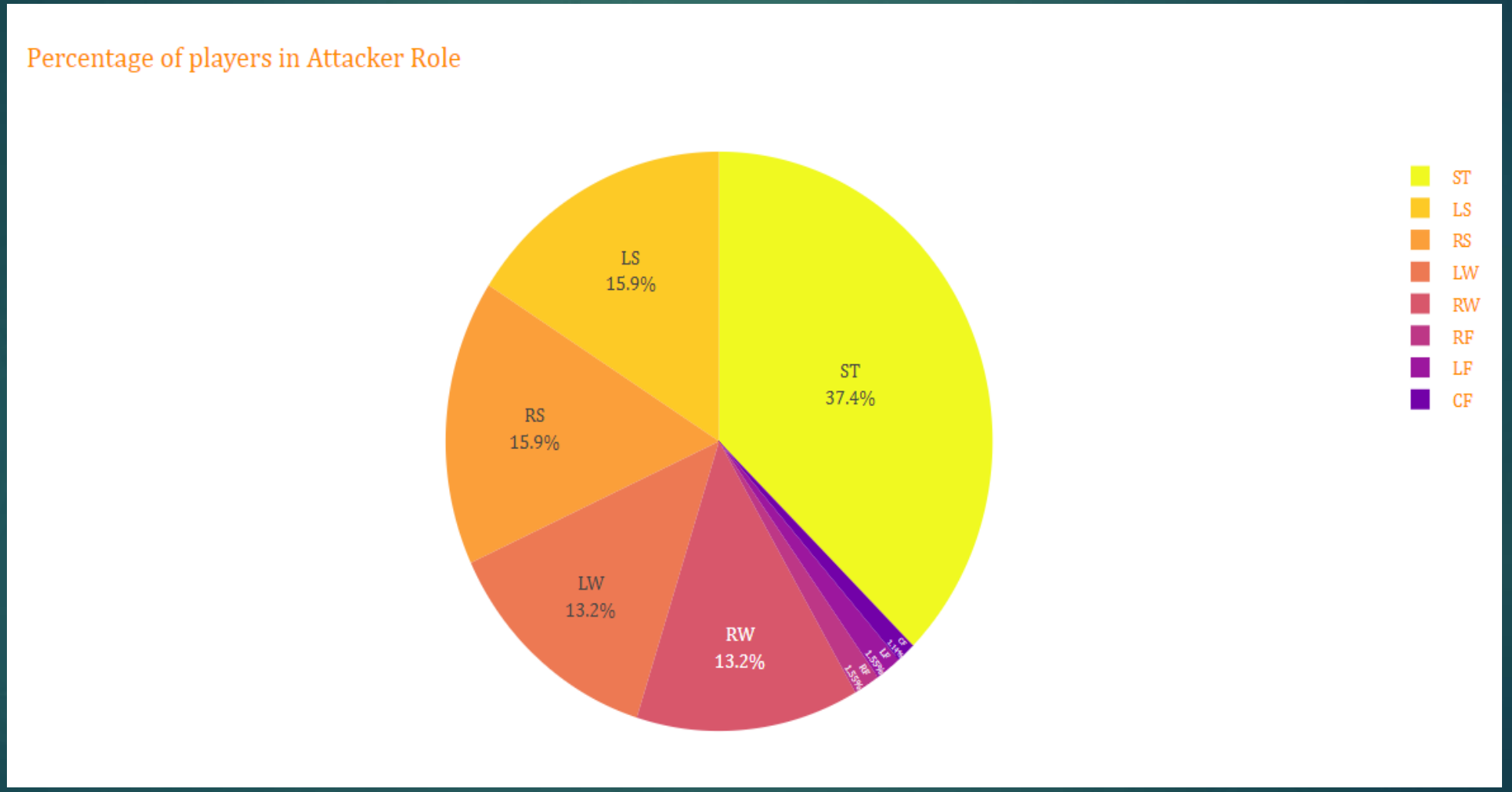
Average Wage by Work Load:



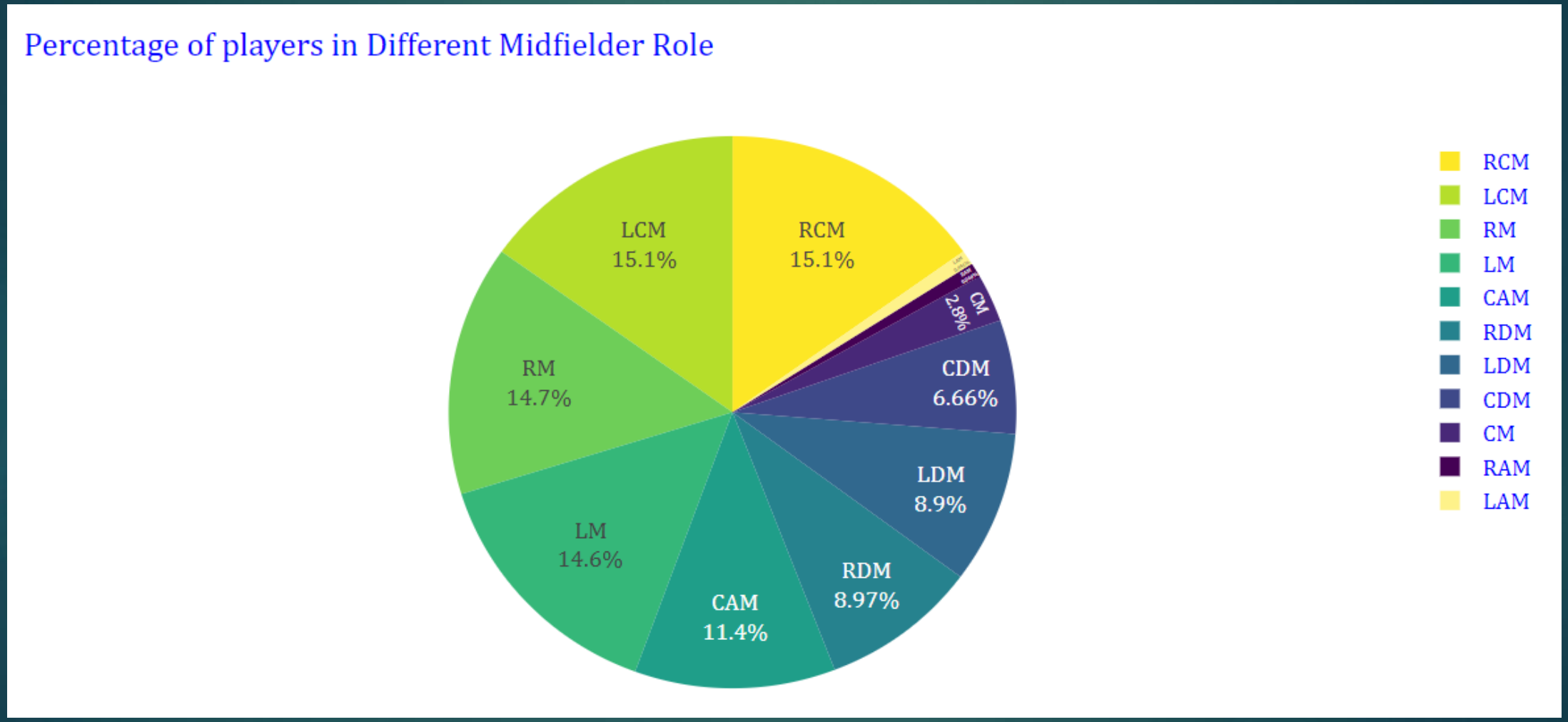
Player distribution according to overall rating:



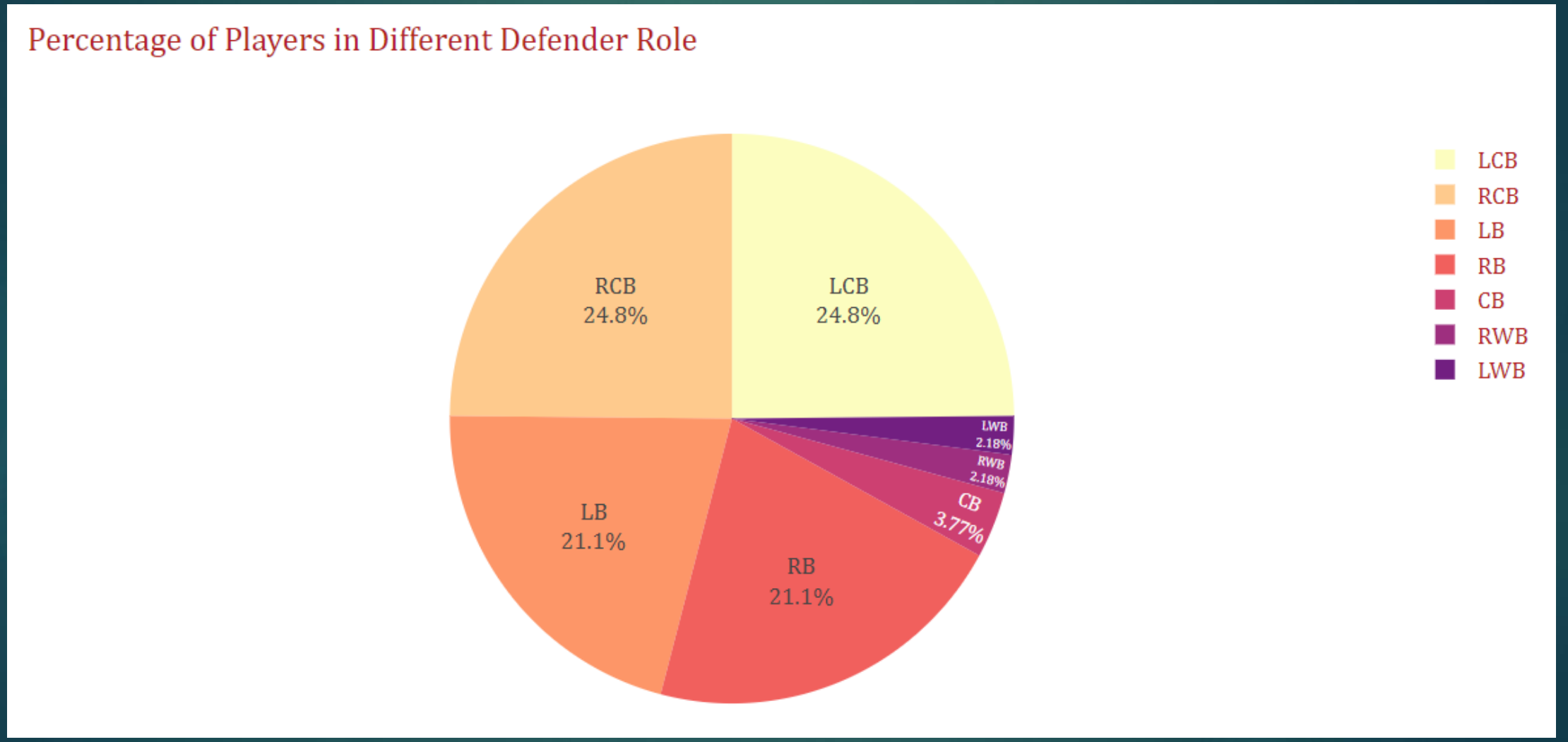
Attackers Distribution:



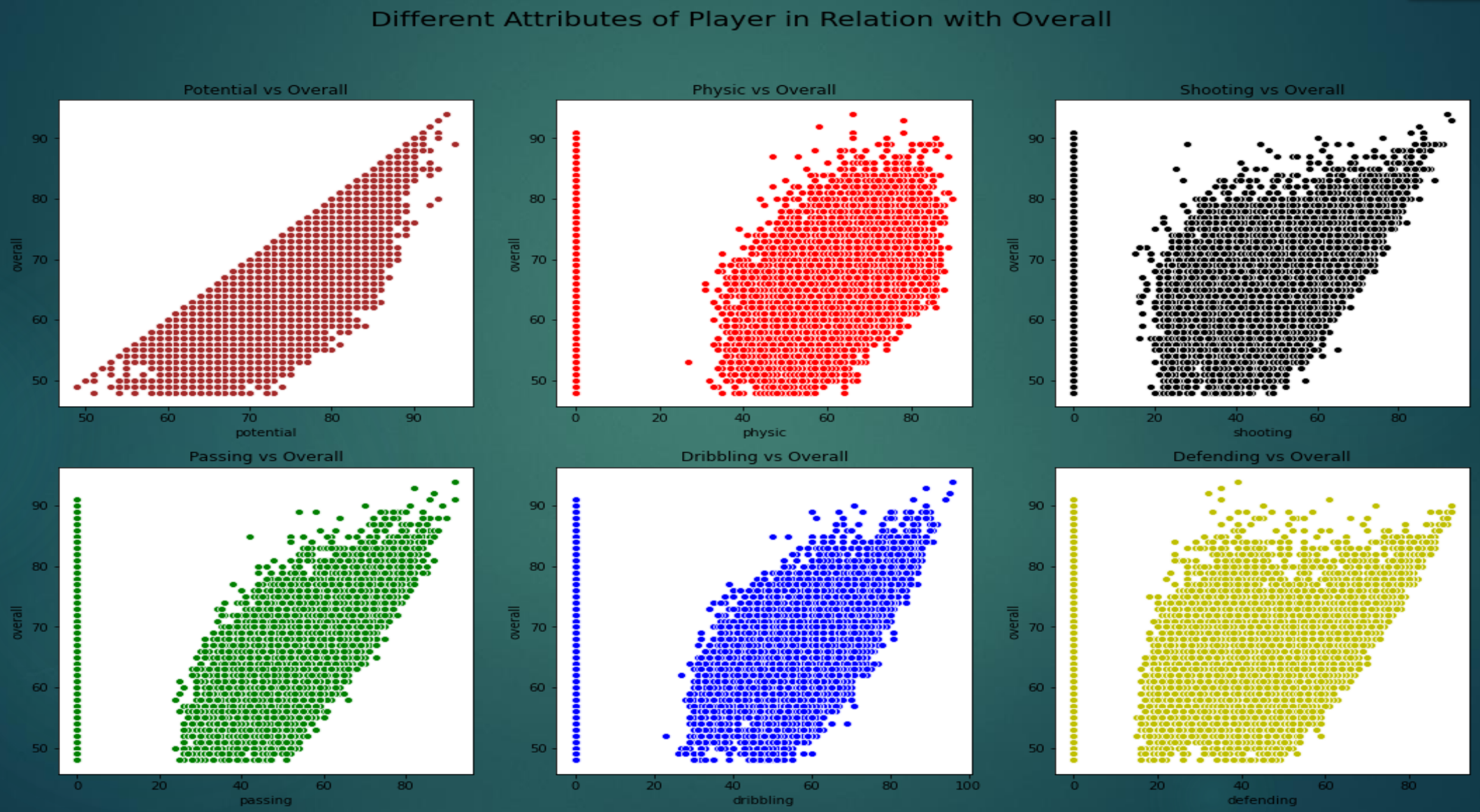
Midfielder Distribution:



Defender Distibution:



Attributes of a player against Overall Rating:



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. SuÃ¡rez	32	FC Barcelona	Uruguay	89
31	M. Neuer	33	FC Bayern MÃ¼nchen	Germany	88
96	Z. IbrahimoviÄ‡	37	LA Galaxy	Sweden	85

Thank You !!!