

FIFA 26: Análise de Jogadores

Enzo Araújo

Dataset Usado:

EAFC26 Mens Player Data Analysis Modeling

<https://www.kaggle.com/code/devraai/eafc26-mens-player-data-analysis-modeling>

Etapa 1: Limpeza, Transformação, Criação de *Features* e Tratamento de Dados

Sobre os Datapoints

	ID	Rank	Name	GENDER	OVR	PAC	SHO	PAS	DRI	DEF	PHY
0	209331	1	Mohamed Salah	M	91	89	88	86	90	45	76
1	231747	3	Kylian Mbappé	M	91	97	90	81	92	37	76
2	231443	5	Ousmane Dembélé	M	90	91	88	83	93	50	69
3	231866	6	Rodri	M	90	65	80	86	84	86	85
4	203376	8	Virgil van Dijk	M	90	73	60	72	72	90	87
5	252371	9	Jude Bellingham	M	90	80	86	83	90	78	85
6	239085	10	Erling Haaland	M	90	86	91	70	80	45	88
7	233419	11	Raphinha	M	89	91	84	85	87	53	75
8	235212	12	Achraf Hakimi	M	89	92	79	82	83	82	79
9	277643	14	Lamine Yamal	M	89	85	81	86	90	23	53

Atributos

Pessoais: 'ID', 'Rank', 'Name', 'GENDER', 'Height', 'Weight', 'Age', 'Nation',
'League', 'Team', 'card', 'url', 'Weak foot', 'Skill moves', 'Preferred foot',
'Alternative positions', 'Position'

Principais: 'OVR', 'PAC', 'SHO', 'PAS', 'DRI', 'DEF', 'PHY'.

Ataque: 'Positioning', 'Finishing', 'Shot Power', 'Long Shots', 'Volleys', 'Penalties',
'Vision', 'Crossing', 'Free Kick Accuracy', 'Short Passing', 'Long Passing', 'Curve',
'Dribbling', 'Ball Control', 'Heading Accuracy'.

Físicos: 'Acceleration', 'Sprint Speed', 'Agility', 'Balance', 'Reactions',
'Composure', 'Jumping', 'Stamina', 'Strength', 'Aggression',

Defensivos: 'Interceptions', 'Def Awareness', 'Standing Tackle', 'Sliding Tackle'

Goleiro: 'GK Diving', 'GK Handling', 'GK Kicking', 'GK Positioning', 'GK Reflexes',

Remoção dos Atributos "ID", "card", "url",
"GENDER", "play style"

Completar com 0 valores faltantes
(Atributos de Goleiro em jogadores de linha)

Tratar dados de peso e altura para KG e CM,
respectivamente

Transformar em booleana a perna preferida
(Direita = 1)

Criação de Atributos:

1- Setor: Gol, Defesa, Ataque, Meio de Campo

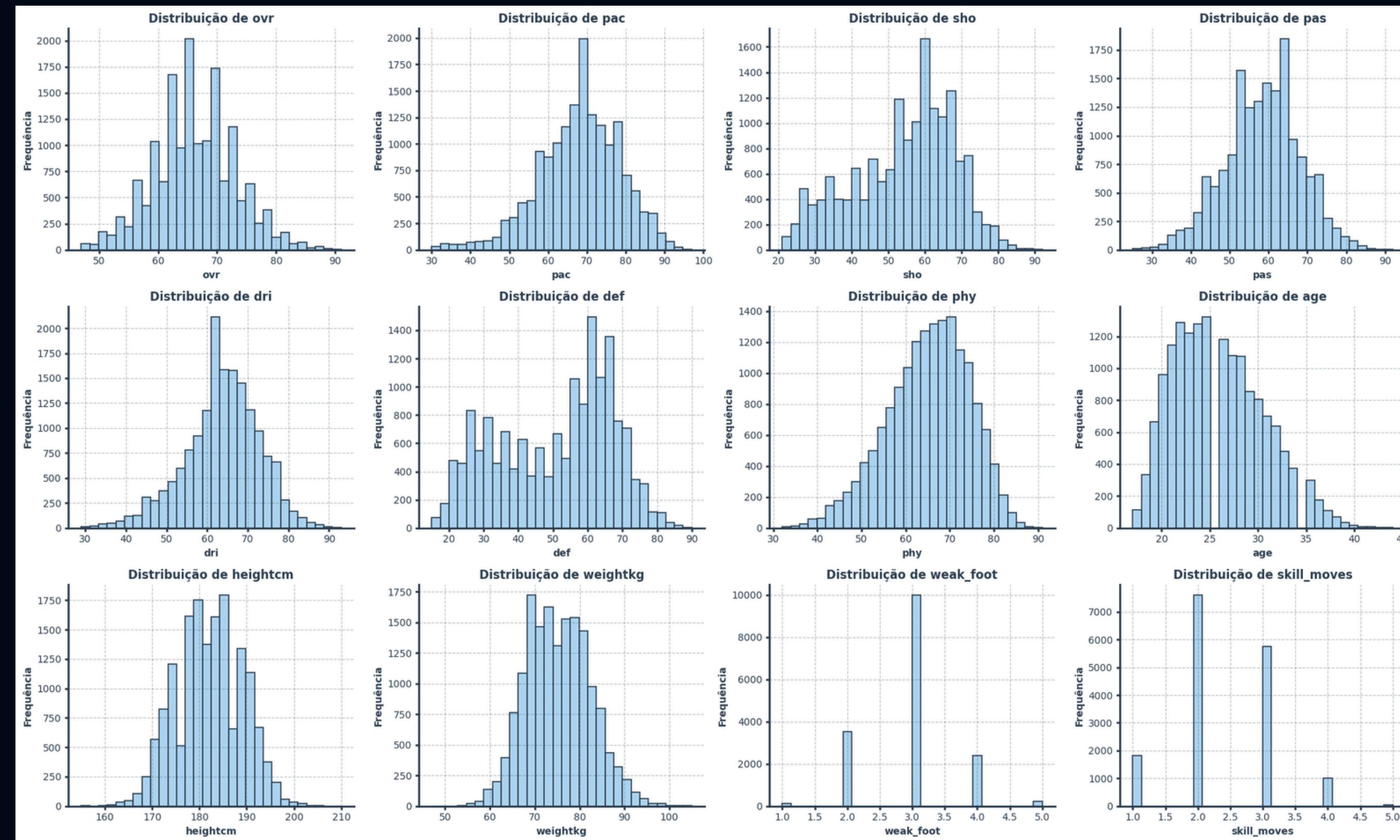
2- Versatilidade: Quantidade de Posições em que o jogador pode jogar

3- Ofensivo: Flag se o jogador é predominantemente ofensivo

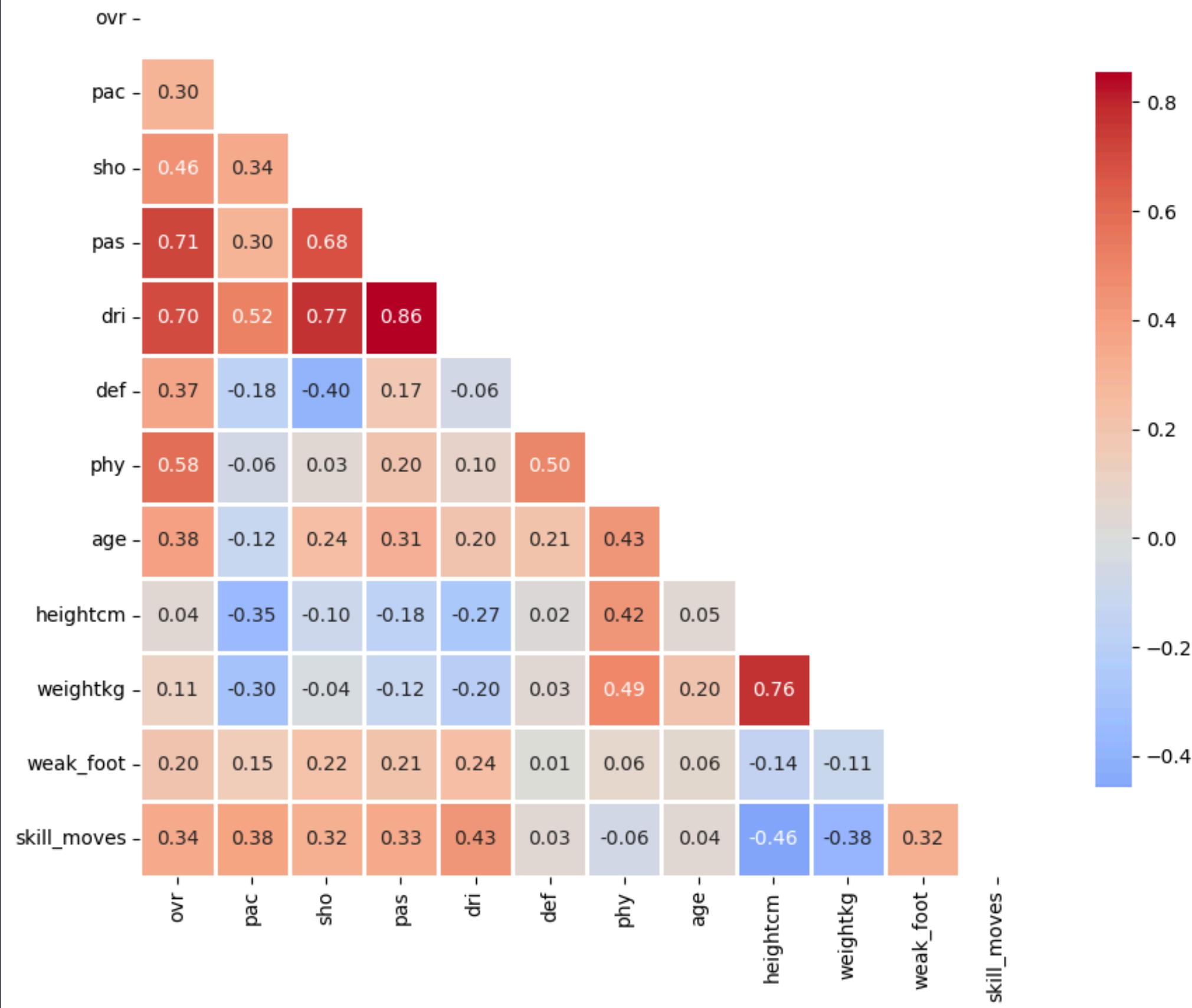
4- Defensive: Flag se o jogador é predominantemente defensivo

5- All around: Flag se o jogador domina os dois lados do campo

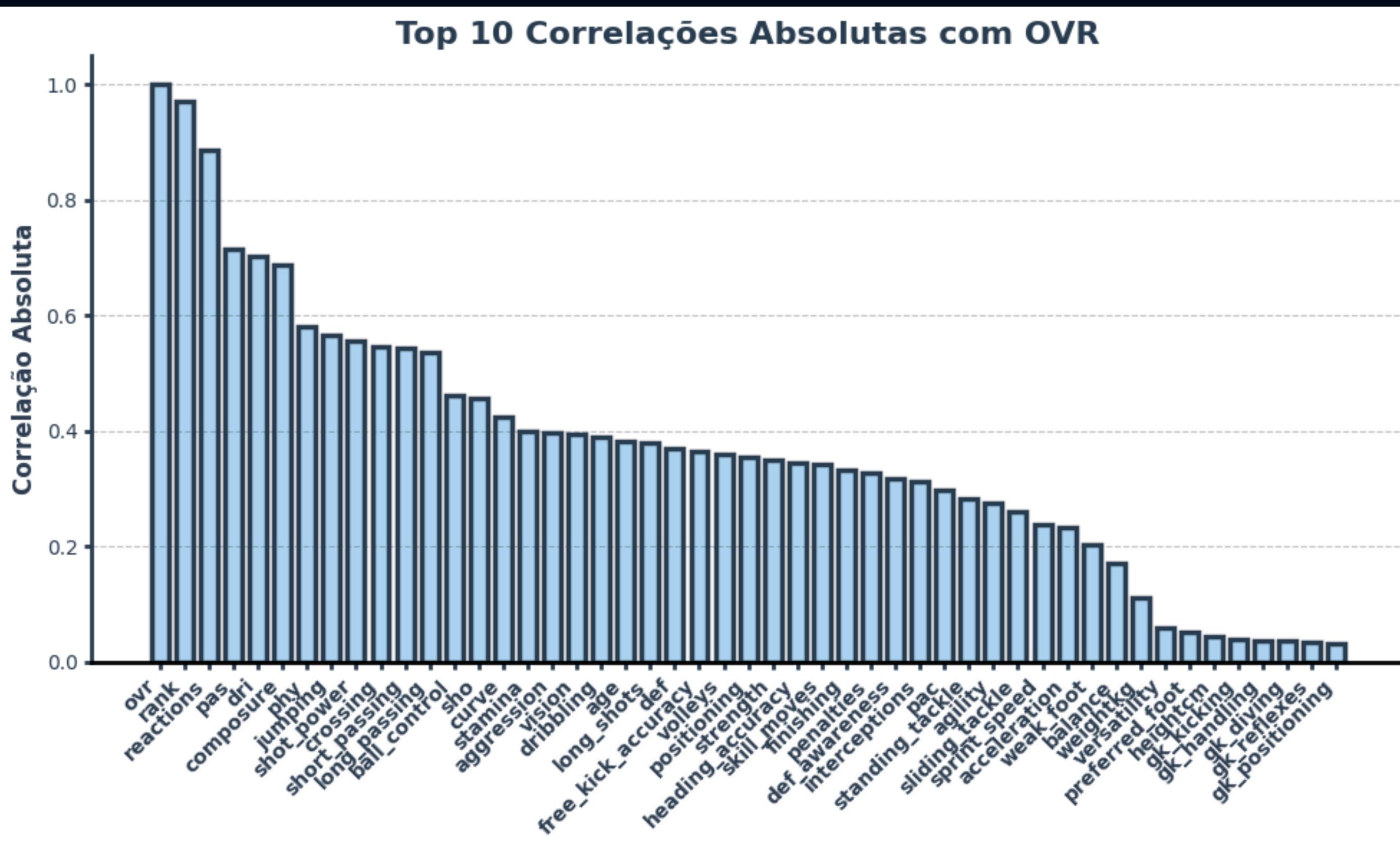
Etapa 2: Análise Exploratória de Dados (EDA)

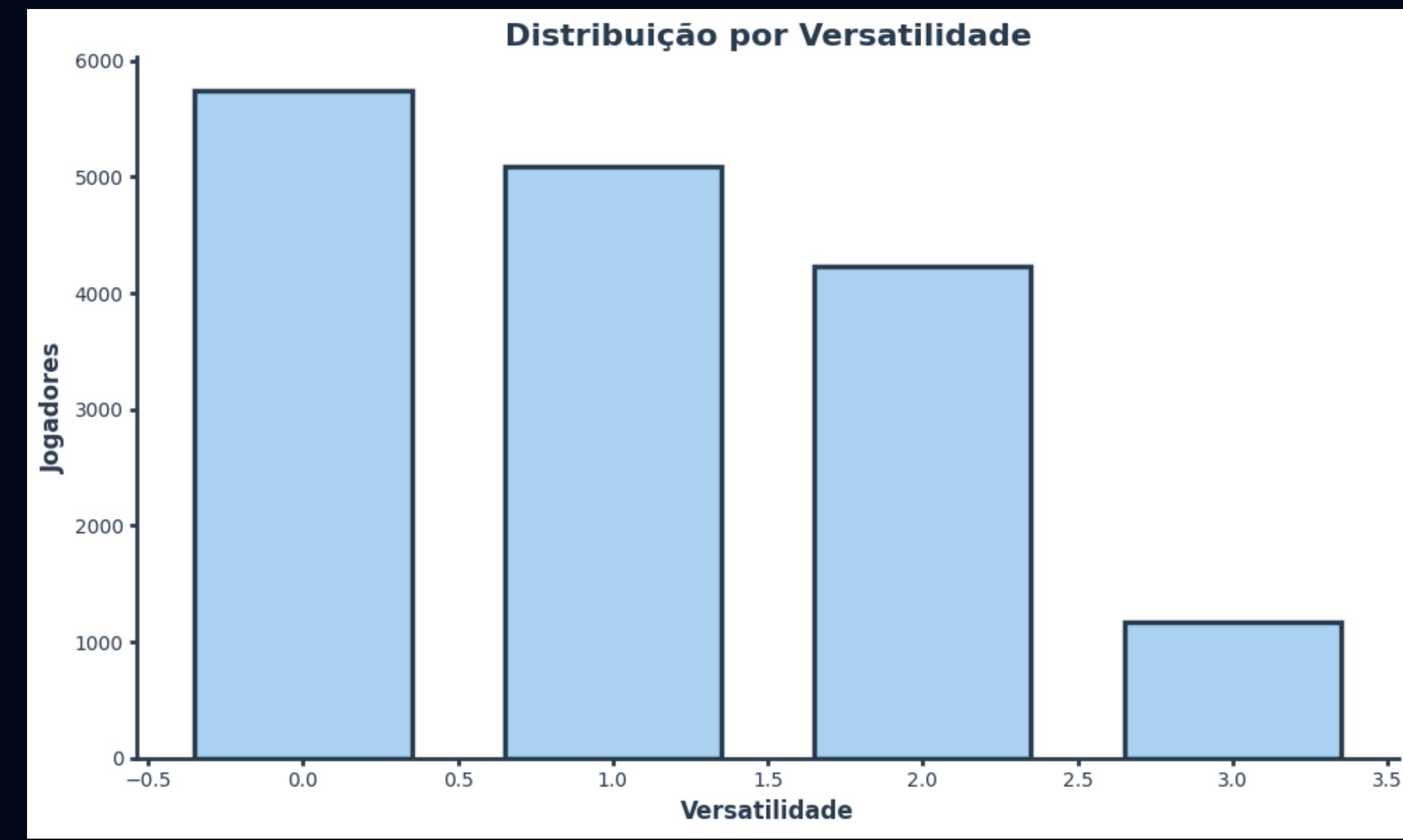


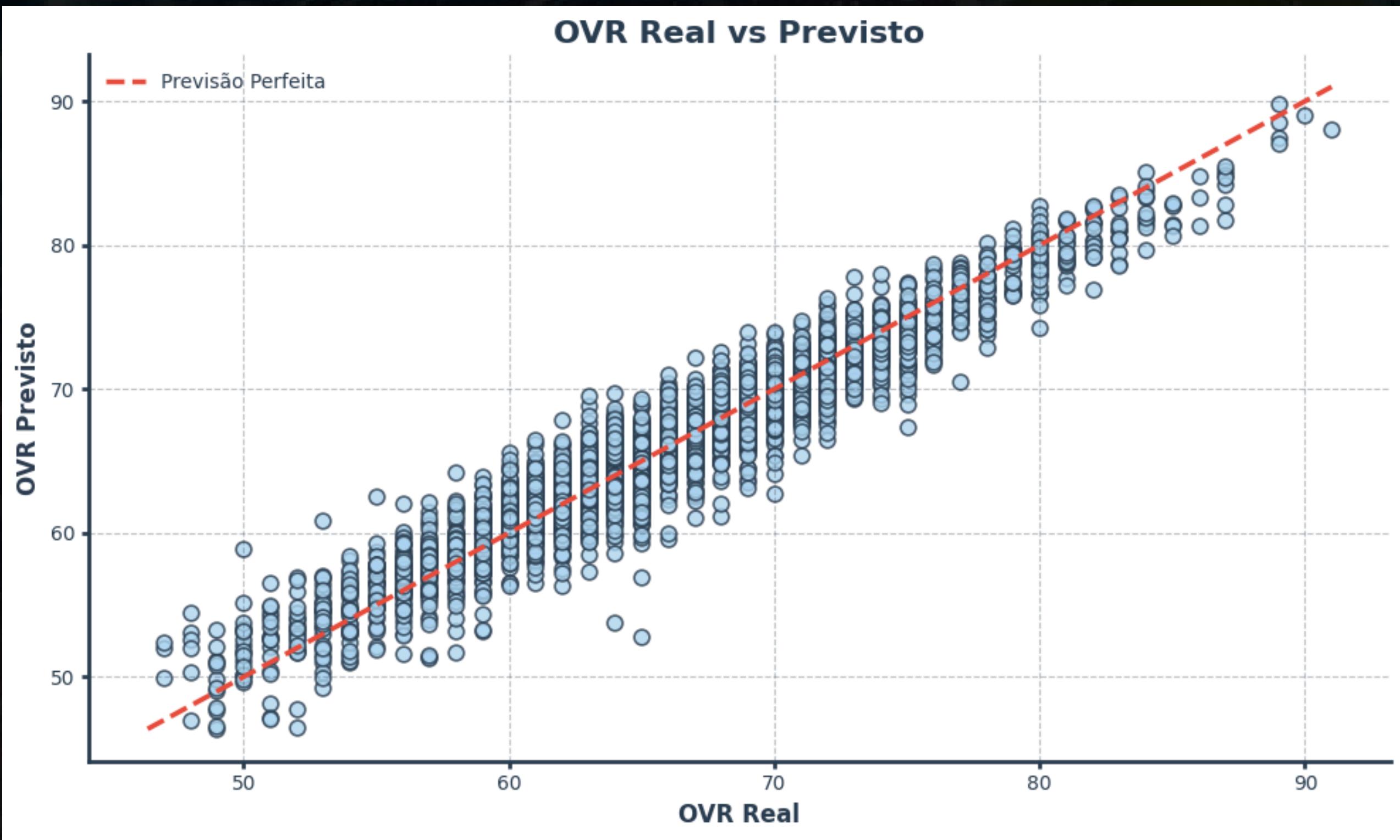
Matriz de Correlação - Atributos Principais



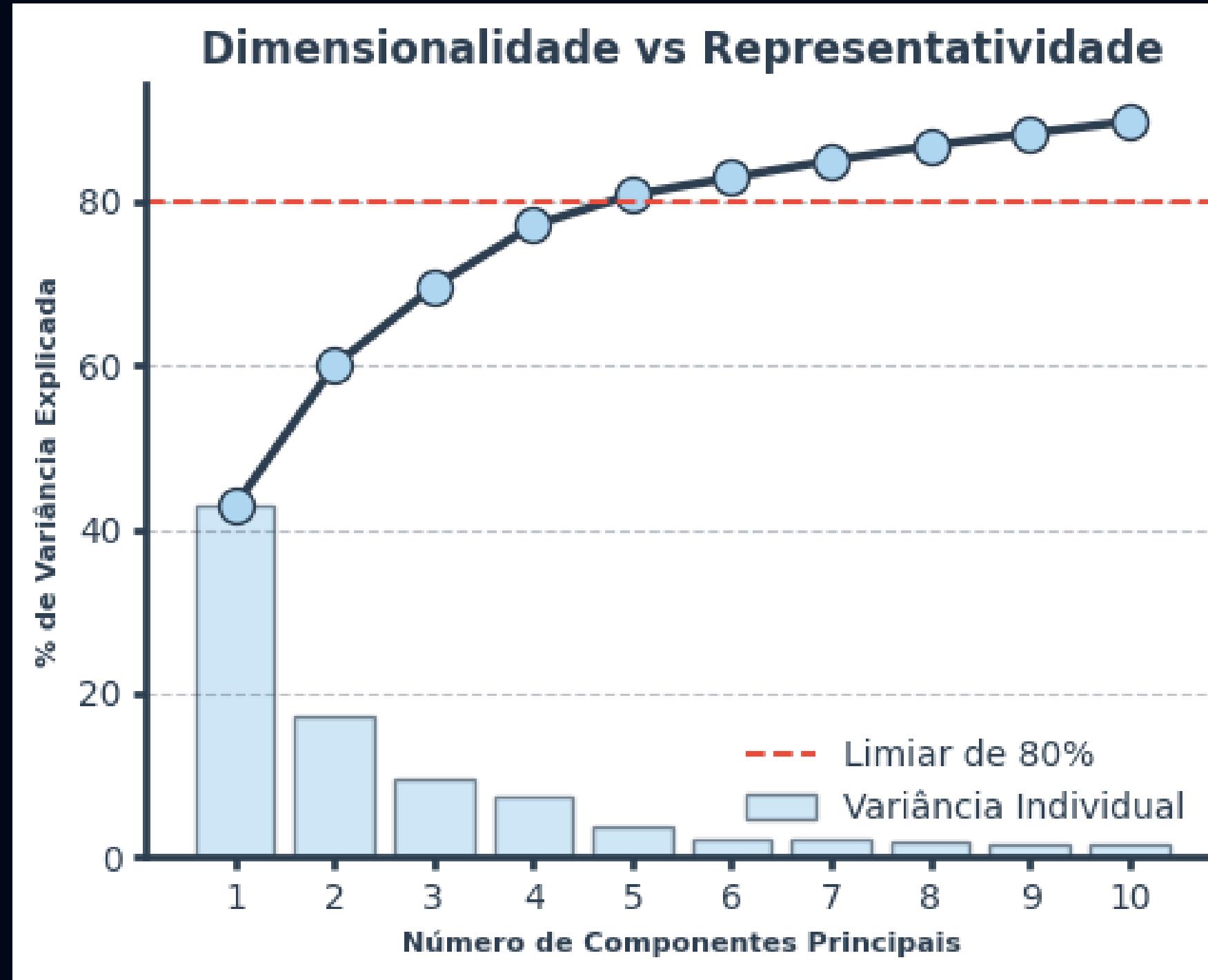
Top 10 Correlações Absolutas com OVR







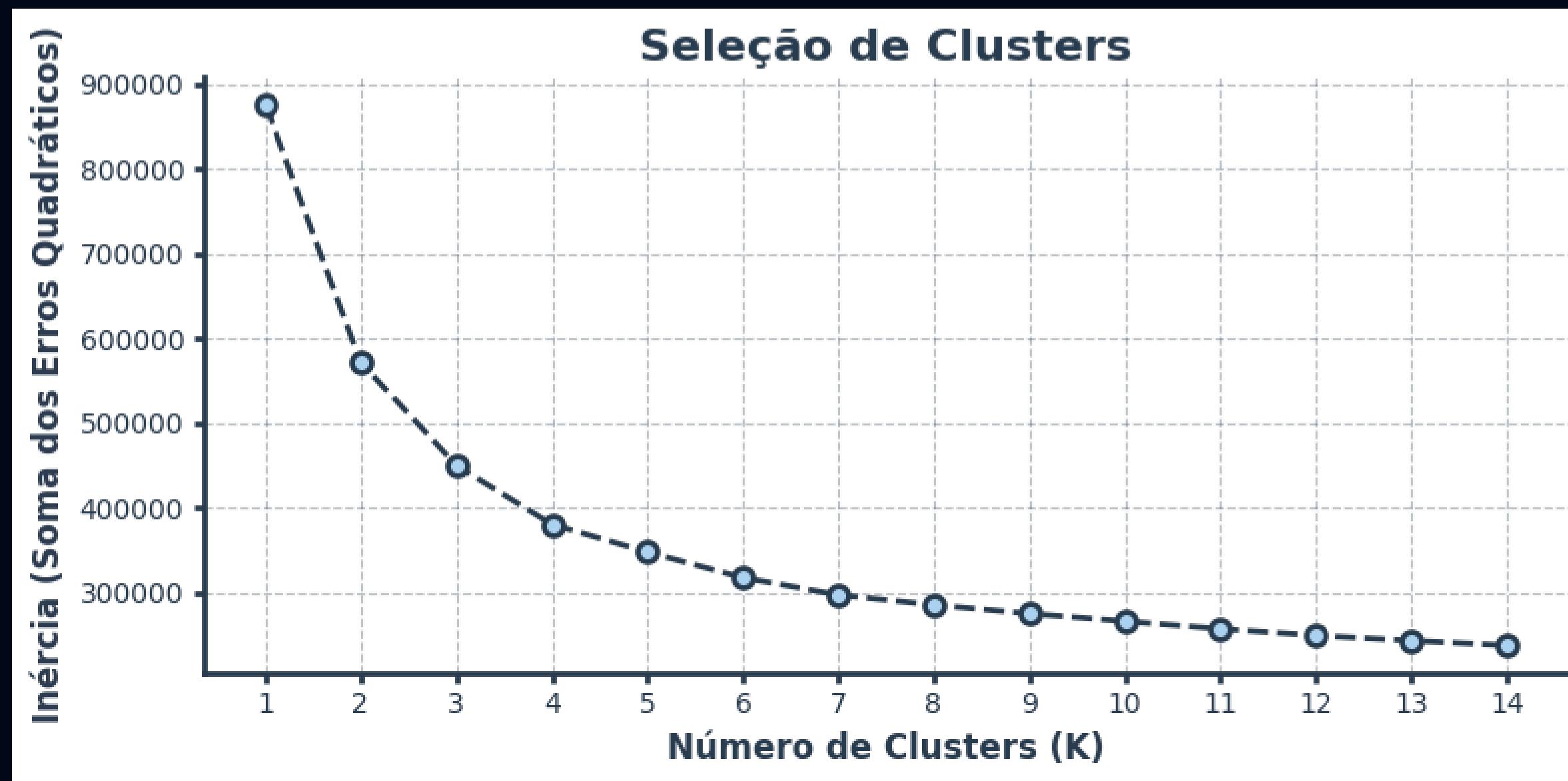
Etapa 3: Treinamento do Modelo Linear: PCA + K-Means



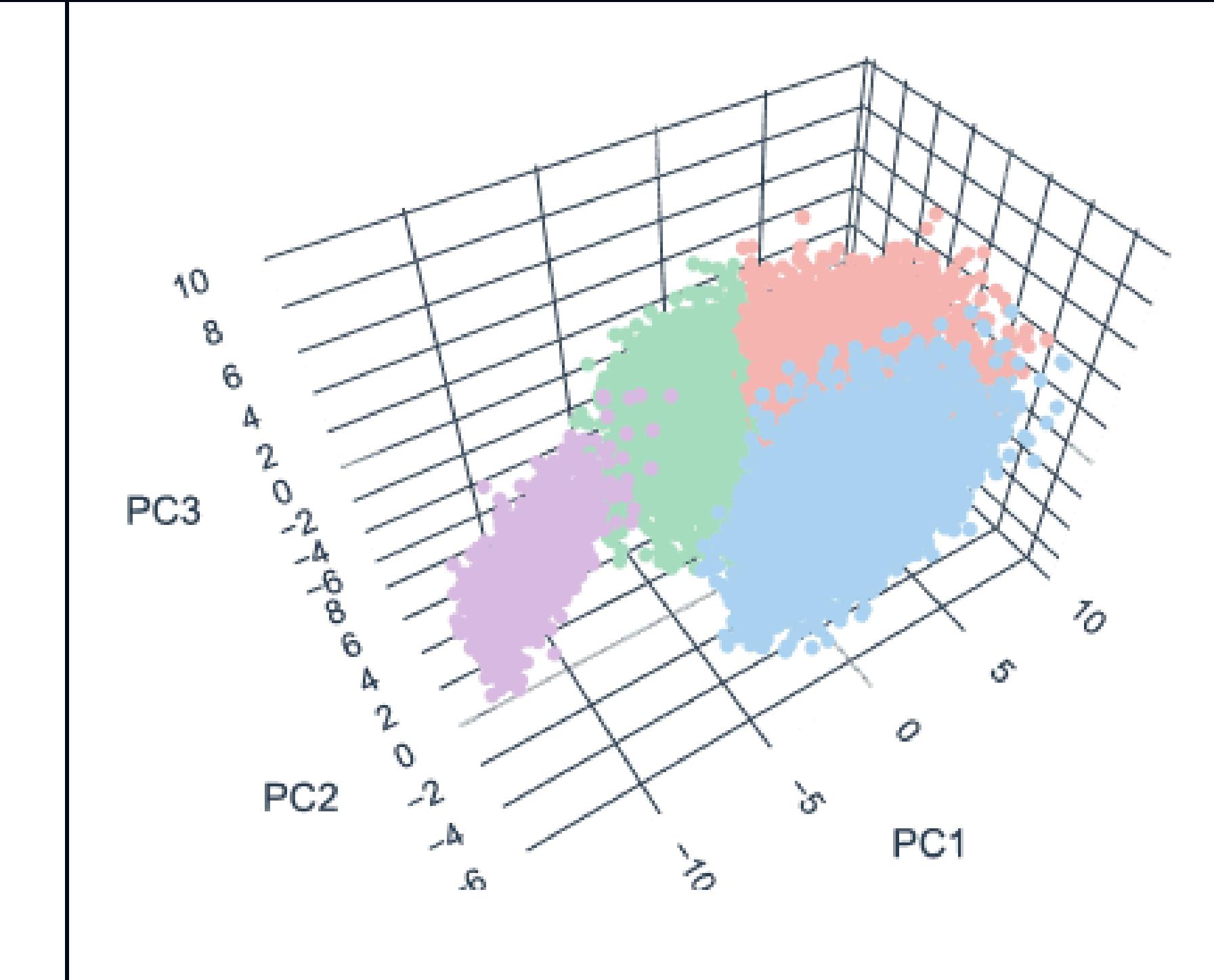
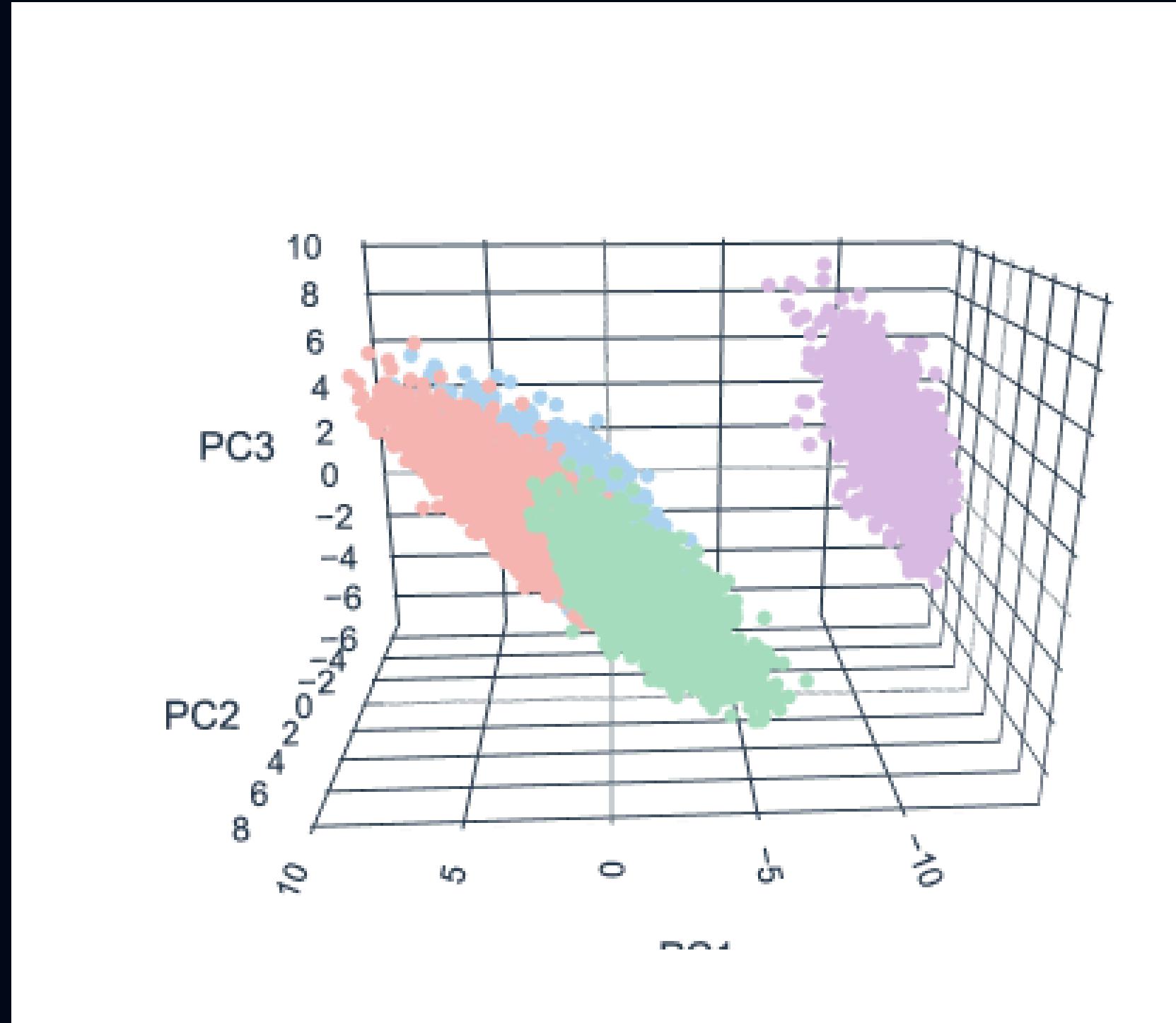
Passo 1:
Padronização dos
Dados;

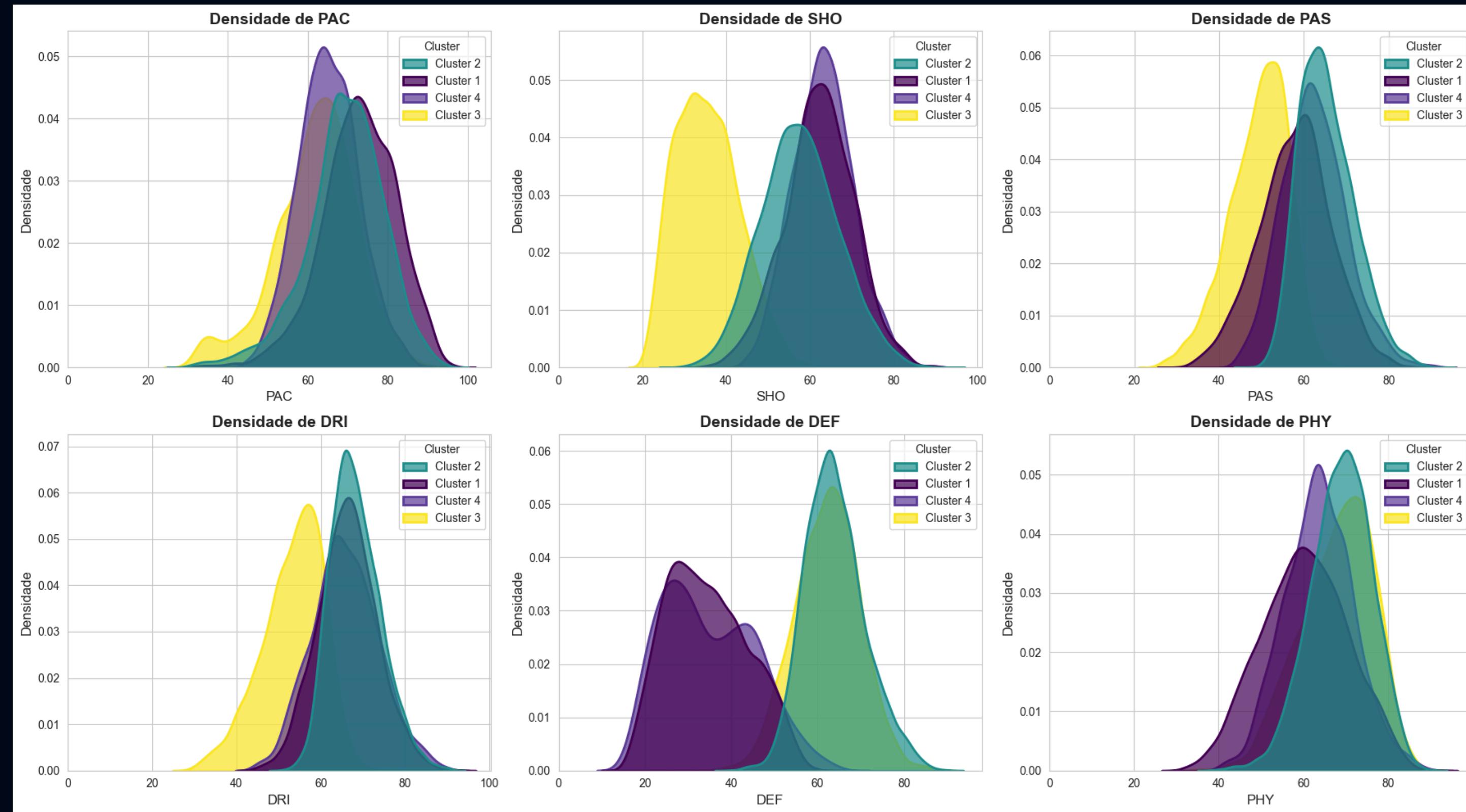
Passo 2: Escolher
PCA para 3 Features
(Para representar
em 3d)

Treinamento do **K-Means** de 1 até 9 com seleção pelo **método do Cotovelo**



Plot Tridimensional da Clusterização





A partir da análise
dos *clusters*, esses
foram os perfis
traçados:

Cluster 1 e 3: Jogadores “de Nicho”



Vinicius Júnior:
Passe, Finalização, Drible,
Veloz, Explosivo



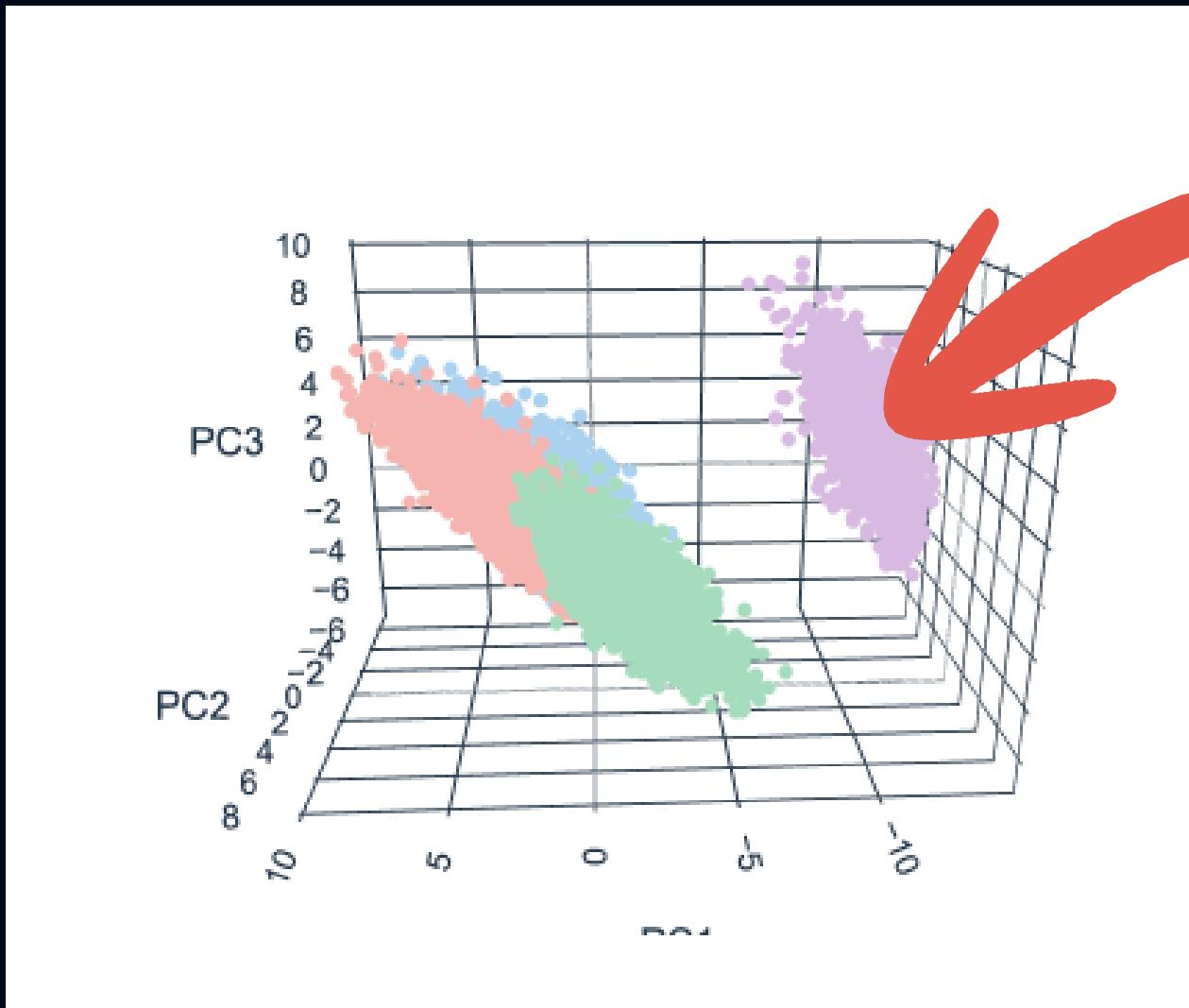
Gabriel Magalhães:
Defesa, Altura, Força,
Lento

Cluster 2: Jogadores Completos



Composto de jogadores com **todos os atributos em alto nível**, especialmente atributos **físicos**, como força, velocidade e altura. Por exemplo, Bellingham possui 78+ em todas as notas principais.

Cluster 4: Goleiros



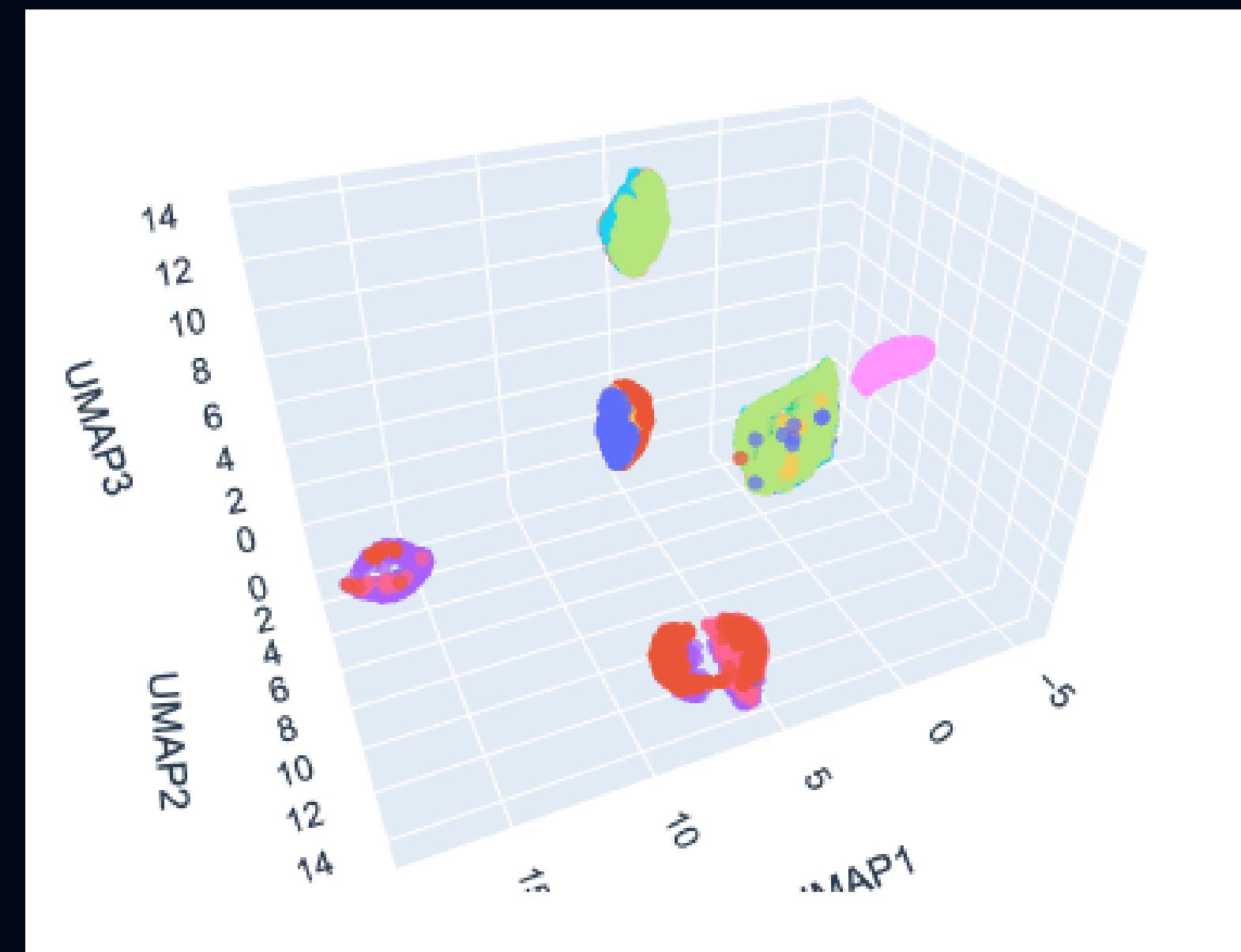
Manuel Neuer

Etapa 4:

Treinamento do

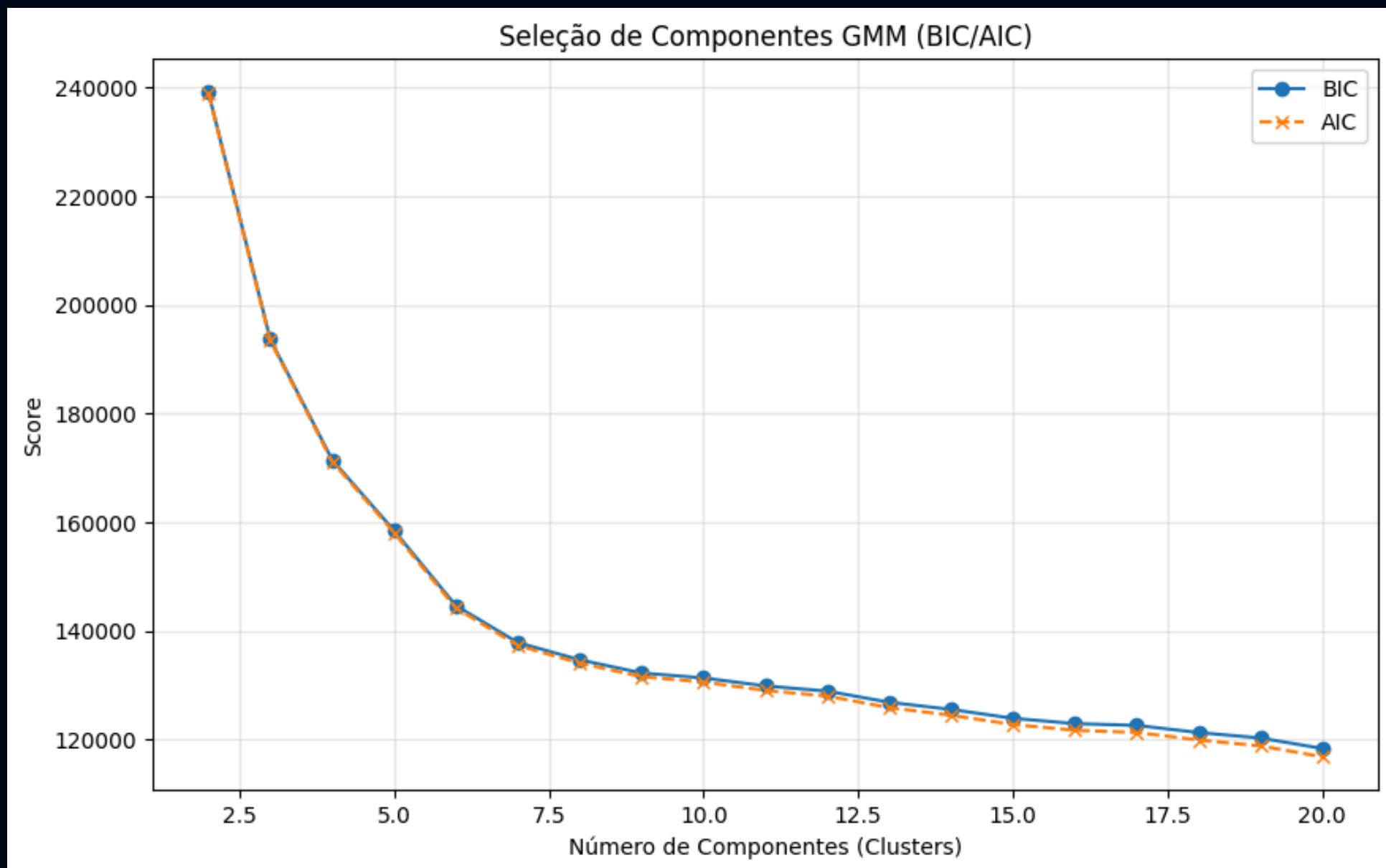
Modelo GMM + UMAP

O ajuste do UMAP foi feito por tentativa e erro, em que chegou ao parâmetro final de:
n_neighbors=50,
min_dist=0.15



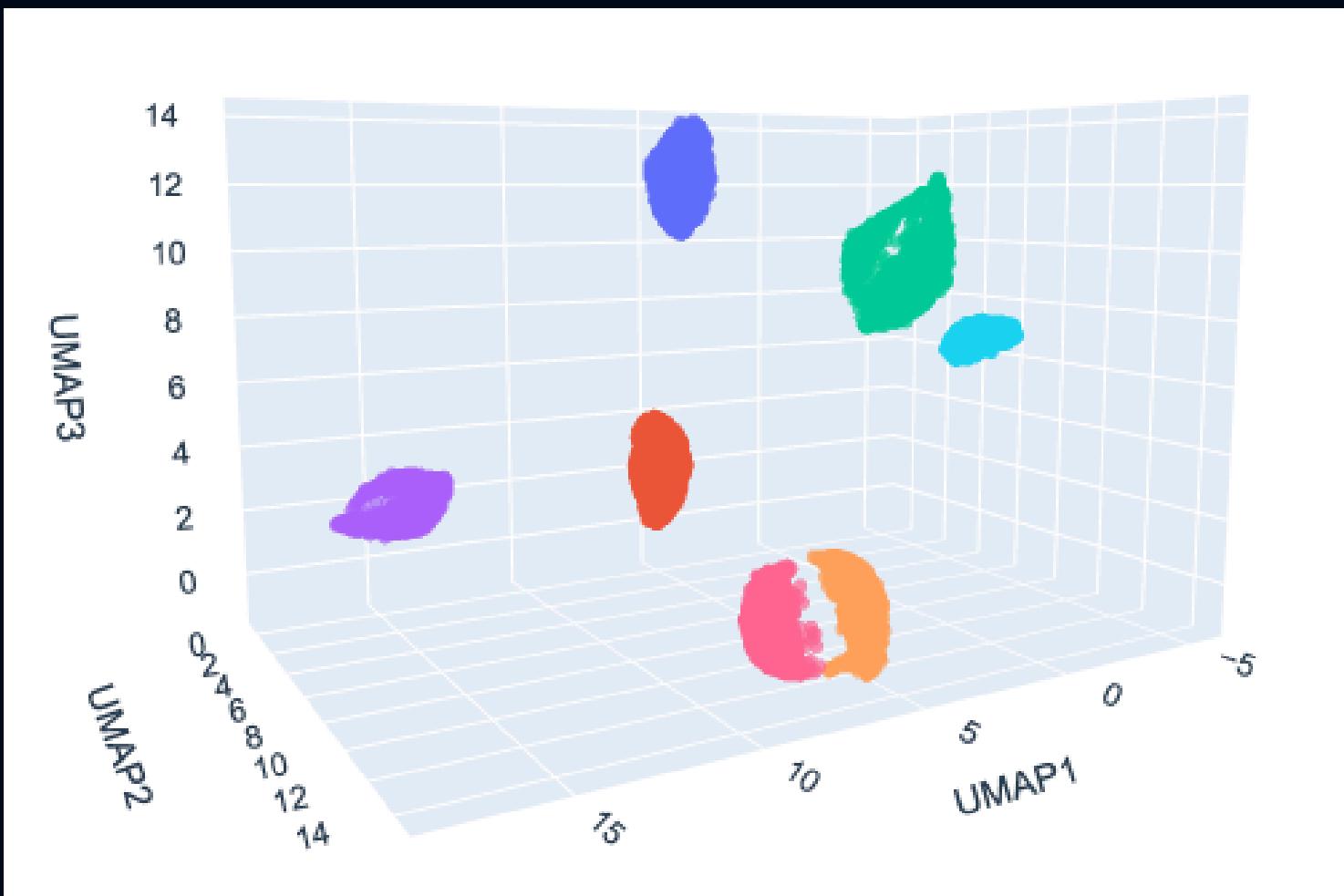
O Gaussian Mixture Models(**GMM**) é um modelo probabilístico que identifica *clusters* de formatos flexíveis calculando a **probabilidade** de um dado ponto pertencer a um cluster.

O número de clusters foi definido pelo **Método do Cotovelo**, usando o BIC/AIC Scores.



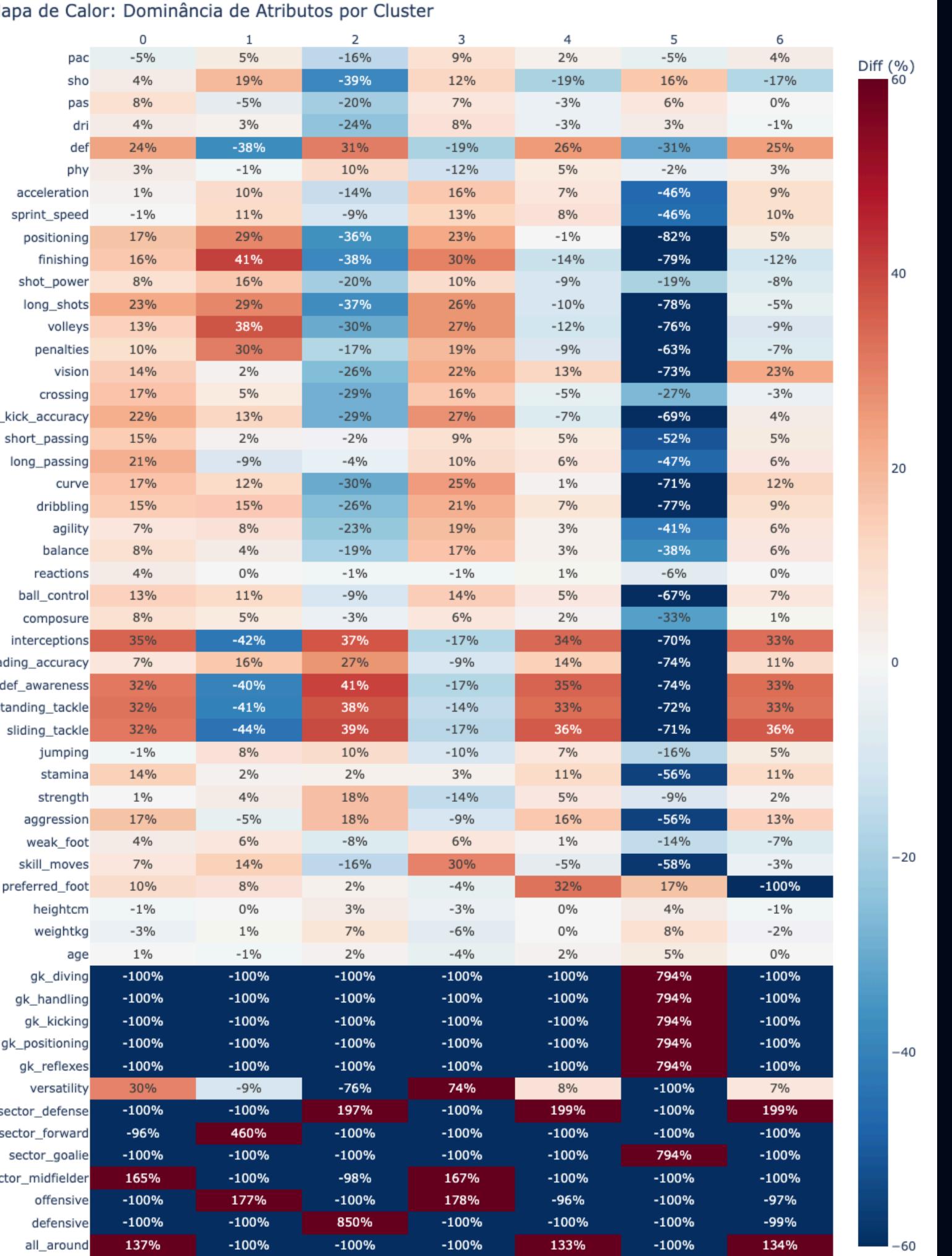
Escolha final: **7 Clusters**

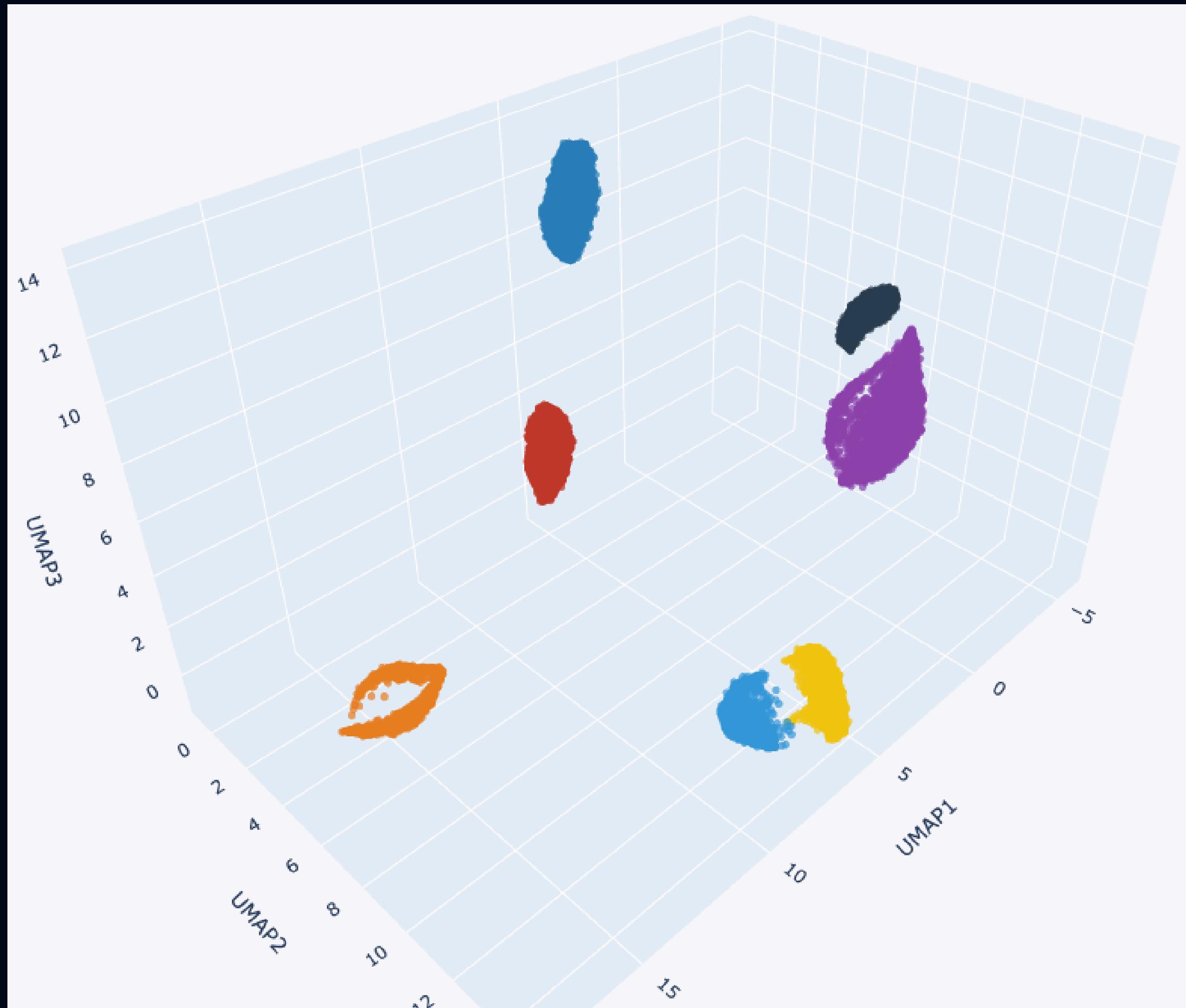
Clusters no Espaço Vetorial 3D



A seguir, a **Matriz de Dominância**.

Representa a variação de cada atributo em cada classe em relação à média global.





Função Tática

- Defensor Canhoto (LB/LCB)
- Atacante (ST/FW)
- Volante (CDM/CM)
- Meia Criativo (CAM/Winger)
- Goleiro (GK)
- Zagueiro Xerife (CB)
- Defensor Destro (RB/RCB)

Identidade dos 7 Clusters do GMM:

Cluster 0 – Volante / Meia Central

- O motor do time.
- Jogadores de meio-campo (midfielder +165%) com perfil all_around (+137%).
- Diferem dos criativos pela defesa forte: interceptions (+35%) e desarmes (+32%).



Rodri, Meia do Man. City (ING)

Cluster 1 - Atacante Finalizador

- A linha de frente
- Grupo massivo de atacantes (sector_forward +460%) focados em terminar a jogada
- Dominam finishing (+41%) e voleios (+38%)



Vini JR, Atacante do Real Madrid (ESP)

Cluster 2 - Zagueiro "Raiz"

- A força bruta.
- Zagueiros puros (97% CB) com a flag `defensive` explodindo (+850%).
- Focam apenas em destruir, sem a característica `all_around` dos outros defensores.



Virgil Van Dijk, zagueiro do Liverpool (ING)

Cluster 3 – Defensor Canhoto

- A raridade tática.
- O algoritmo isolou zagueiros e laterais (`sector_defense` +199%) que são exclusivamente canhotos(`preferred_foot` -100%).
- Essenciais para o equilíbrio da saída de bola.



Nuno Mendes, Lateral do PSG (FRA)

Cluster 4 - Goleiro



Alisson, goleiro do Liverpool (ING)

Cluster 5 – Meia Criativo

- A classe e o drible.
- Jogadores de meio-campo, mas ofensivos (`offensive` +178%).
- Destacam-se pela `versatility` (+74%) e habilidade técnica (`skill_moves` +29%).



Lamine Yamal, Atacante do Barcelona (ESP)

Cluster 6 - Defensor Destro

- A defesa técnica destra.
- Laterais direitos e zagueiros que sabem jogar (`all_around` +134%).
- Diferenciam-se do Cluster 3 por serem destros (`preferred_foot` positivo) e do Cluster 2 por terem técnica.



Carvajal, Lateral do Real Madrid(ESP)