

NEOLAND

NBA 2023-24 Player Shooting data

Carlos Ramos
Enrique Lopez
Sara Fiorentini
Paloma González



NBA

Proyecto Fin de Bootcamp

- 01 OBJETIVO
- 02 ALCANCE DEL PROYECTO
- 03 HERRAMIENTAS
- 04 ANALISIS
- 05 CONCLUSION

Objetivos

GENERALES

- Analizar y modelar el comportamiento de los lanzamientos en la NBA.
- Predecir la probabilidad de acierto de un tiro.
- Entender qué factores influyen en el rendimiento.
- Proporcionar insights útiles para entrenadores y analistas deportivos.

ESPECIFICOS

- Limpieza y preparación de datos
- Exploración y visualización
- Modelado predictivo
- Análisis de rendimiento
- Aplicación práctica

Contexto

BASE DE DATOS

El proyecto se centra en una base de datos específica de lanzamientos de la NBA, que incluye información como la posición del jugador, la distancia al aro, el tipo de lanzamiento, el resultado del tiro, entre otros.

TÉCNICAS UTILIZADAS

Data Analytics: limpieza, exploración, visualización.
Machine Learning: clustering, regresión lineal, regresión logística, Random Forest, redes neuronales para el análisis y la predicción.

HERRAMIENTAS

Python para análisis y modelado, Power BI para visualización interactiva y bibliotecas de Machine Learning (scikit-learn, TensorFlow/Keras).

LIMITACIONES

No se incluyen datos en tiempo real ni se desarrolla una aplicación en producción.
No incluye factores externos como lesiones, fatiga del jugador o condiciones del partido.
Los modelos predictivos se evalúan en un contexto estático y no se integran en un sistema dinámico de toma de decisiones.

Contexto

IDENTIFICADORES

Esta BBDD contiene datos detallados de lanzamientos y eventos de juegos de la NBA. Incluye varias funciones, como tipo de tiro, ubicación del tiro, información del jugador y contexto del juego. El conjunto de datos está diseñado para ayudar a analizar el rendimiento de los tiros, la eficiencia de los jugadores y los resultados del juego.

- **GRID_TYPE**: Tipo de rejilla usada para rastrear los tiros (normalmente "Shot Chart Detail").
 - **GAME_ID**: ID único del partido.
 - **GAME_EVENT_ID**: ID único del evento dentro del partido (cada tiro, falta, etc.).
 - **PLAYER_ID**: ID único del jugador.
 - **PLAYER_NAME**: Nombre del jugador.
 - **TEAM_ID**: ID único del equipo.
 - **TEAM_NAME**: Nombre del equipo.

Contexto

DETALLES DEL TIRO

INFORMACION DEL PARTIDO

- **PERIOD**: Período del partido (1 = primer cuarto, 2 = segundo cuarto, etc.).
- **MINUTES_REMAINING**: Minutos restantes en el período cuando ocurrió el tiro.
- **SECONDS_REMAINING**: Segundos restantes en el período cuando ocurrió el tiro.
- **GAME_DATE**: Fecha del partido (formato AAAAMMDD).

- **EVENT_TYPE**: Tipo de evento (Ejemplo: "Shot" para un intento de tiro).
- **ACTION_TYPE**: Tipo de acción (Ejemplo: "Jump Shot", "Layup", "Dunk").
- **SHOT_TYPE**: Tipo de tiro (Ejemplo: "2PT Field Goal", "3PT Field Goal").
- **SHOT_ZONE_BASIC**: Zona básica del tiro (Ejemplo: "Mid-Range", "Restricted Area").
- **SHOT_ZONE_AREA**: Área específica del tiro (Ejemplo: "Left Corner", "Right Wing").
- **SHOT_ZONE_RANGE**: Distancia del tiro categorizada (Ejemplo: "<8 ft.", "16-24 ft.").
 - **SHOT_DISTANCE**: Distancia exacta del tiro en pies.
- **LOC_X**: Coordenada X de la ubicación del tiro en la cancha.
- **LOC_Y**: Coordenada Y de la ubicación del tiro en la cancha.
- **SHOT_ATTEMPTED_FLAG**: Indica si el tiro fue intentado (1 = Sí, 0 = No).
- **SHOT_MADE_FLAG**: Indica si el tiro fue encestado (1 = Sí, 0 = No).

Contexto

DATOS
ADICIONALES

- **`HOME_TEAM_NAME`**: Nombre del equipo local.
- **`AWAY_TEAM_NAME`**: Nombre del equipo visitante.
- **`HTM`**: Código del equipo local (Ejemplo: "TOR" para Toronto Raptors).
- **`VTM`**: Código del equipo visitante (Ejemplo: "BOS" para Boston Celtics).
- **`LOCATION`**: Si el tiro fue en casa ("HOME") o fuera ("AWAY").
- **`SEASON_TYPE`**: Tipo de temporada ("Regular Season" o "Playoffs"): Nombre del equipo.

Herramientas

Limpieza de datos

PHYTON

Análisis exploratorio

PHYTON

POWER BI

EXCEL

POWER POINT

Análisis predictivo

PHYTON

SCIKIT-LEARN

TENSORFLOW/KERAS



ANÁLISIS

ANÁLISIS

Limpieza de datos

Ingeniería de la base de datos: unión con otras tablas.
Integración de columnas clave y tratamiento de valores nulos.
Optimización y estructuración para análisis eficiente.

Análisis exploratorio

Mapas de calor de zonas de tiro.
Comparativa entre jugadores con gráficos de radar.
Tablero interactivo en Power BI para filtrar jugadores/equipos.

Análisis predictivo

Clusterización de jugadores según características.
Modelos para predecir si un tiro será encestado.
Predicción de puntos por partido por jugador.

Limpieza de datos

- Nuestra base de datos principal era **all_shot_data_final.csv**
Con las siguientes columnas: **NumColumns = 29**

```
Index(['GRID_TYPE', 'GAME_ID', 'GAME_EVENT_ID', 'PLAYER_ID', 'PLAYER_NAME',
       'TEAM_ID', 'TEAM_NAME', 'PERIOD', 'MINUTES_REMAINING',
       'SECONDS_REMAINING', 'EVENT_TYPE', 'ACTION_TYPE', 'SHOT_TYPE',
       'SHOT_ZONE_BASIC', 'SHOT_ZONE_AREA', 'SHOT_ZONE_RANGE', 'SHOT_DISTANCE',
       'LOC_X', 'LOC_Y', 'SHOT_ATTEMPTED_FLAG', 'SHOT_MADE_FLAG', 'GAME_DATE',
       'HTM', 'VTM', 'player_name', 'Season Type', 'HOME_TEAM_NAME',
       'AWAY_TEAM_NAME', 'LOCATION'],
      dtype='object')
```

Unir Tablas

- Cargamos las dos bases de Datos:
 - BBDD Principal : **all_shot_data_final.csv**
 - BBDD Secundaria: **NBA Stats 202324 All Stats NBA Player Props Tool.csv**
- Unión de columnas deseadas
 - "Claves de tablas ('PLAYER_NAME' y 'NAME')
 - “Columna agregada ('POS')
 - **LEFT JOIN**

Eliminar valores NULLS

- Cambiamos nombres de posiciones
- Reemplazamos los valores NULL

$("C": "P") = \text{Pivot}$
 $("G": "B") = \text{Base}$
 $("F-C": "AP") = \text{Ala-Pivot}$
 $("C-F": "AP") = \text{Ala-Pivot}$
 $("F-G": "ES") = \text{Escolta}$
 $("G-F": "ES") = \text{Escolta}$
 $("F": "A") = \text{Alero}$

{**Patrick Baldwin Jr. , D'Moi Hodge, GG Jackson II, AJ Lawson, Craig Porter Jr.**}

Arreglar columnas

- TEAM_NAME
- HOME_TEAM_NAME
- PLAYER_NAME
- AWAY_TEAM_NAME
- LOCATION

```
team_abbr = {  
    'Toronto Raptors': 'TOR', 'New York Knicks': 'NYK', 'Miami Heat': 'MIA', 'Utah Jazz': 'UTA',  
    'Memphis Grizzlies': 'MEM', 'Minnesota Timberwolves': 'MIN', 'Phoenix Suns': 'PHX',  
    'Cleveland Cavaliers': 'CLE', 'New Orleans Pelicans': 'NOP', 'Milwaukee Bucks': 'MIL',  
    'Orlando Magic': 'ORL', 'Washington Wizards': 'WAS', 'Portland Trail Blazers': 'POR',  
    'Detroit Pistons': 'DET', 'Charlotte Hornets': 'CHA', 'Philadelphia 76ers': 'PHI',  
    'Boston Celtics': 'BOS', 'San Antonio Spurs': 'SAS', 'Sacramento Kings': 'SAC',  
    'Brooklyn Nets': 'BKN', 'LA Clippers': 'LAC', 'Oklahoma City Thunder': 'OKC',  
    'Atlanta Hawks': 'ATL', 'Chicago Bulls': 'CHI', 'Denver Nuggets': 'DEN',  
    'Houston Rockets': 'HOU', 'Indiana Pacers': 'IND', 'Dallas Mavericks': 'DAL',  
    'Los Angeles Lakers': 'LAL', 'Golden State Warriors': 'GSW'  
}
```

Arreglar columnas

- Unidad de medida de pies a metros ($x * 0.3048$) col = "SHOT_DISTANCE"

"Less Than 8 ft." == "Less Than 2.44 m"

"24+"== "More Than 7.32 m"

"16-24" == "4.88-7.32 m"

"8-16" == "2.44-4.88 m"

- Cambiamos el formato de la fecha):

'[GAME_DATE]', format='%Y%m%d'

GAME_DATE
2023-11-17

Eliminar columnas y filas

- Eliminamos las siguientes columnas:

**`["HTM", "VTM", "SHOT_DISTANCE", "SHOT_ZONE_RANGE",
"player_name", "GRID_TYPE", "GAME_EVENT_ID"]`**

- Eliminamos si hubiera filas duplicadas:

*Número de filas antes de eliminar duplicados: **259017***

*Número de filas después de eliminar duplicados: **233999***

BBDD Final

- Damos un nuevo ID:
ID único del evento dentro del partido (cada tiro tiene uno).
- Comprobación de los nuevos ID:
df['NEW_GAME_EVENT_ID'].nunique() == 233999
- Nombre de la BBDD y columnas:
(NBA_BBDD_CLEAN.csv) // NumColumns = 26

```
[ 'GAME_ID', 'NEW_GAME_EVENT_ID', 'PLAYER_ID', 'PLAYER_NAME', 'TEAM_ID',
  'TEAM_NAME', 'PERIOD', 'MINUTES_REMAINING', 'SECONDS_REMAINING',
  'EVENT_TYPE', 'ACTION_TYPE', 'SHOT_TYPE', 'SHOT_ZONE_BASIC',
  'SHOT_ZONE_AREA', 'LOC_X', 'LOC_Y', 'SHOT_ATTEMPTED_FLAG',
  'SHOT_MADE_FLAG', 'GAME_DATE', 'Season Type', 'HOME_TEAM_NAME',
  'AWAY_TEAM_NAME', 'LOCATION', 'POS', 'SHOT_DISTANCE_METERS',
  'SHOT_ZONE_RANGE_METERS'],
```

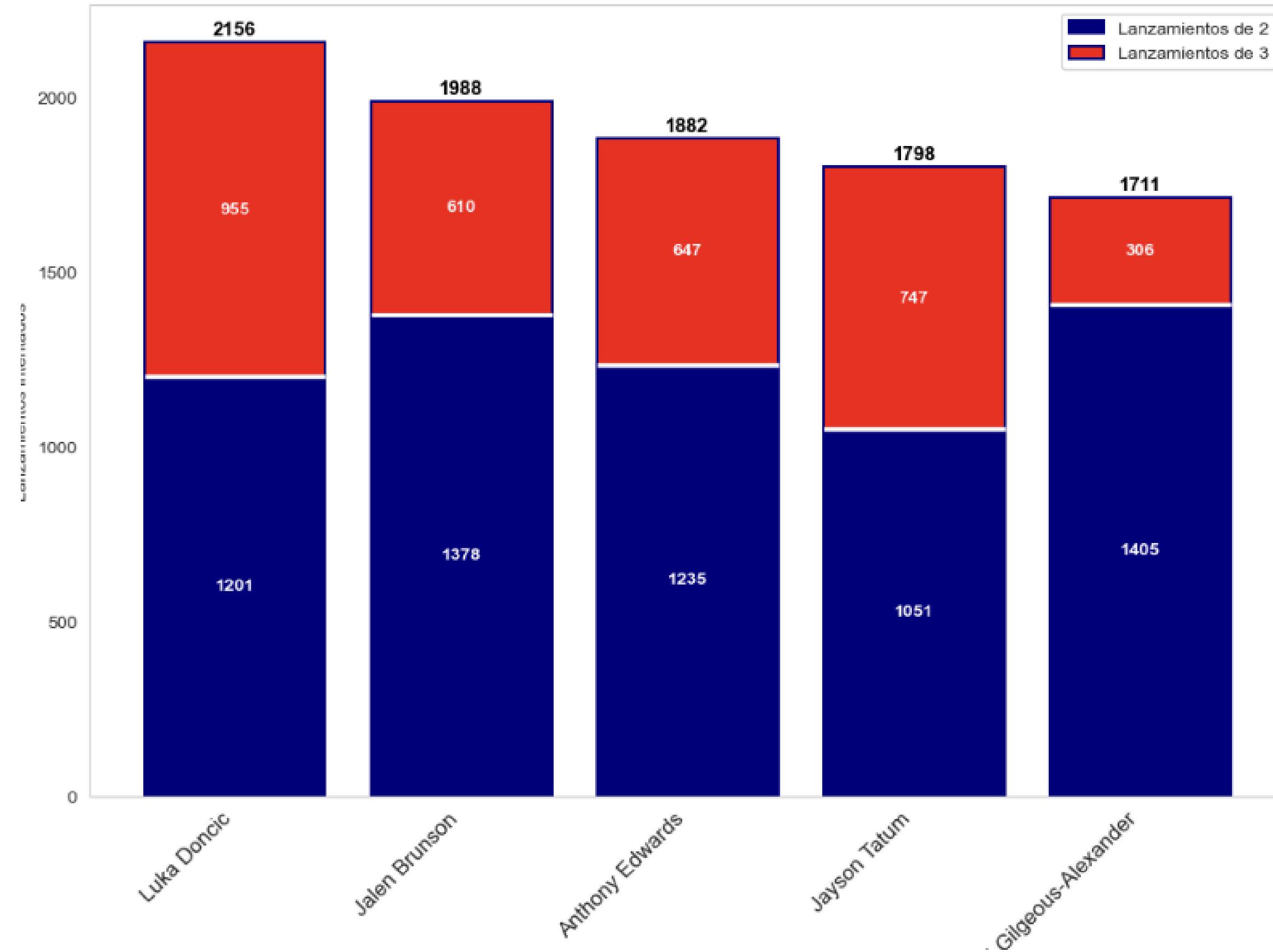
Análisis exploratorio

- PHYTON
- POWER BI



Análisis exploratorio

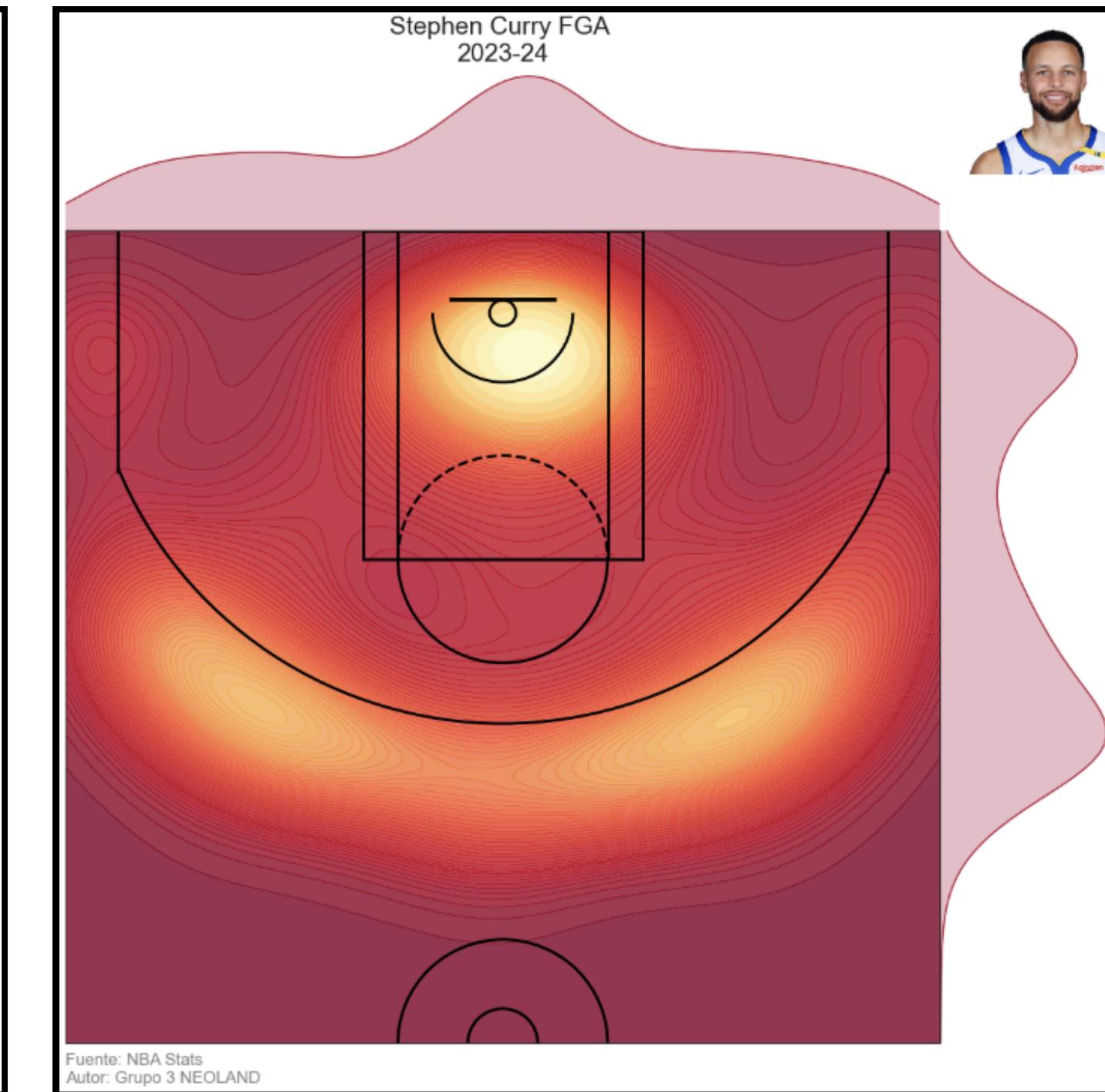
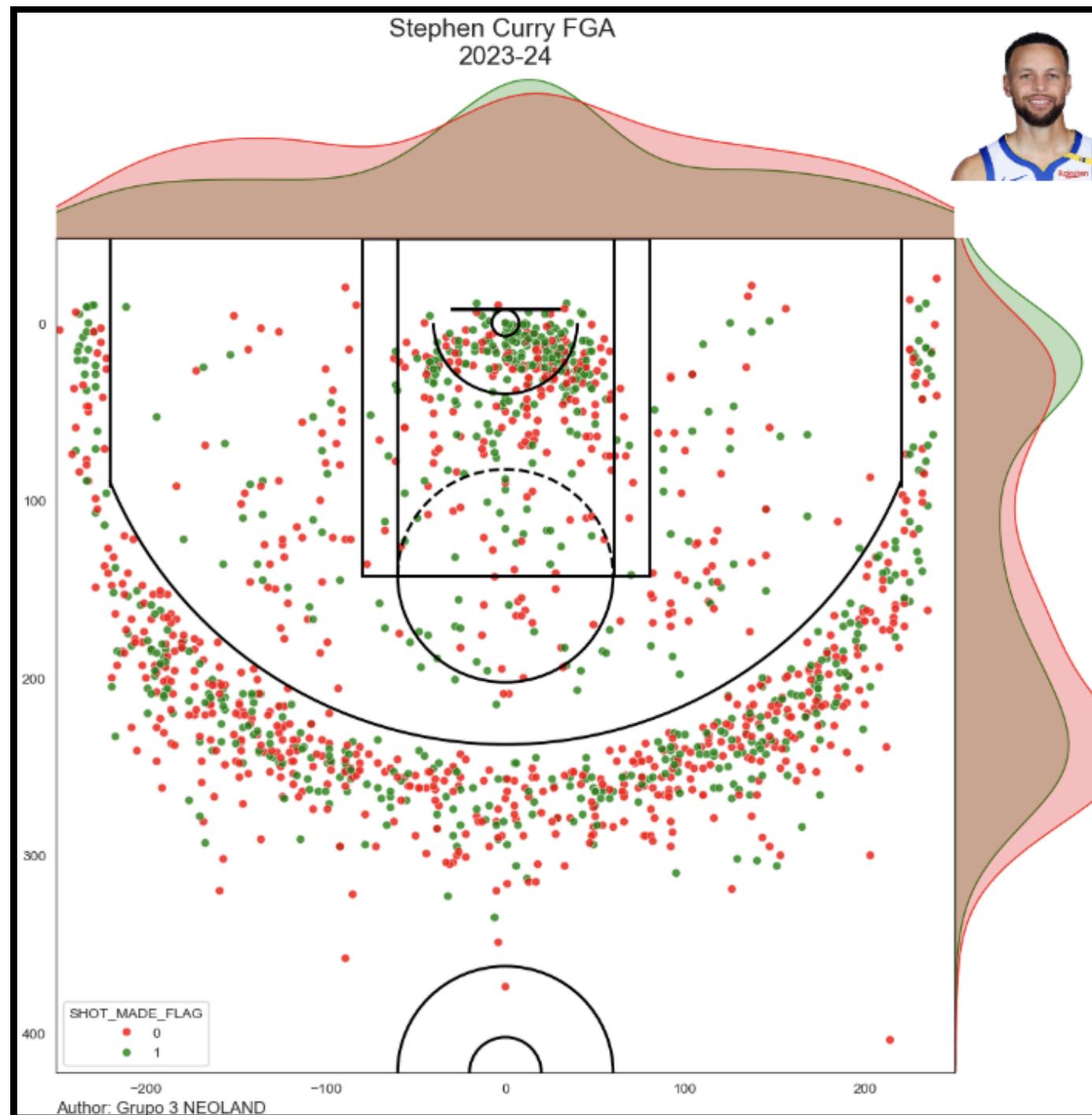
Distribución de lanzamientos intentados (TOP 5 Jugadores)



Análisis exploratorio

Mapas de Calor

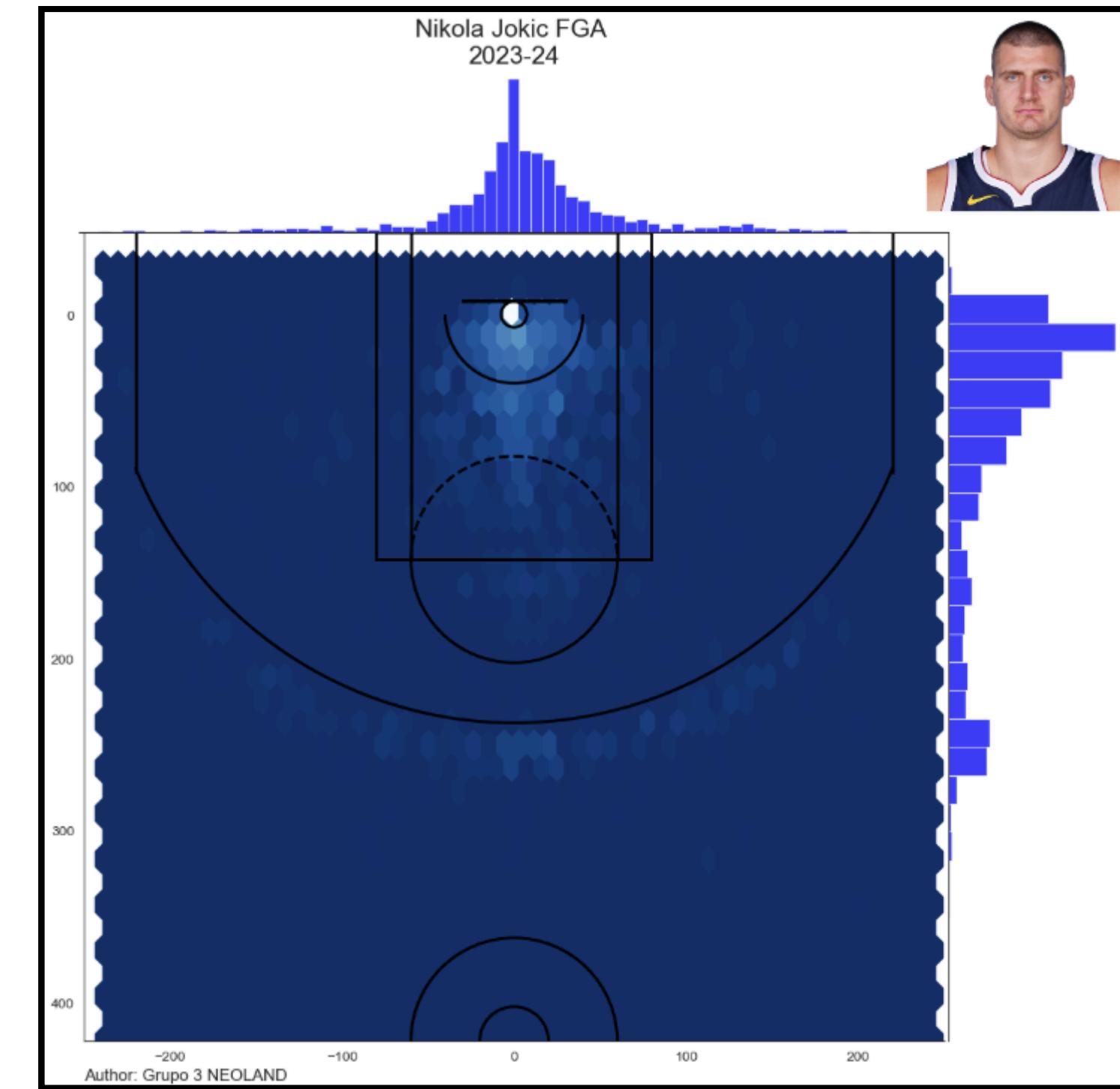
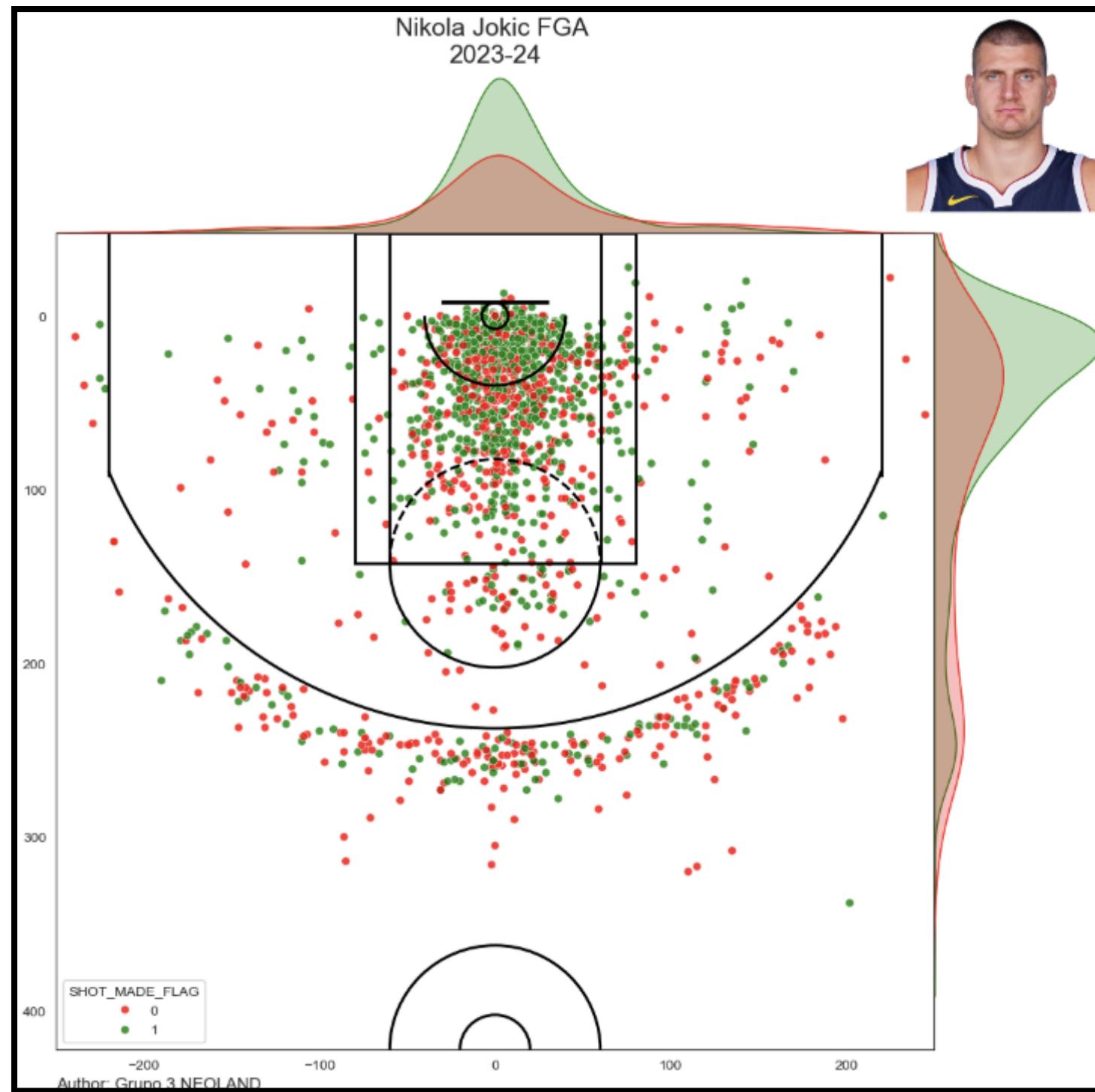
Jugador con más triples encestados: Stephen Curry con 357 triples



Análisis exploratorio

Mapas de Calor

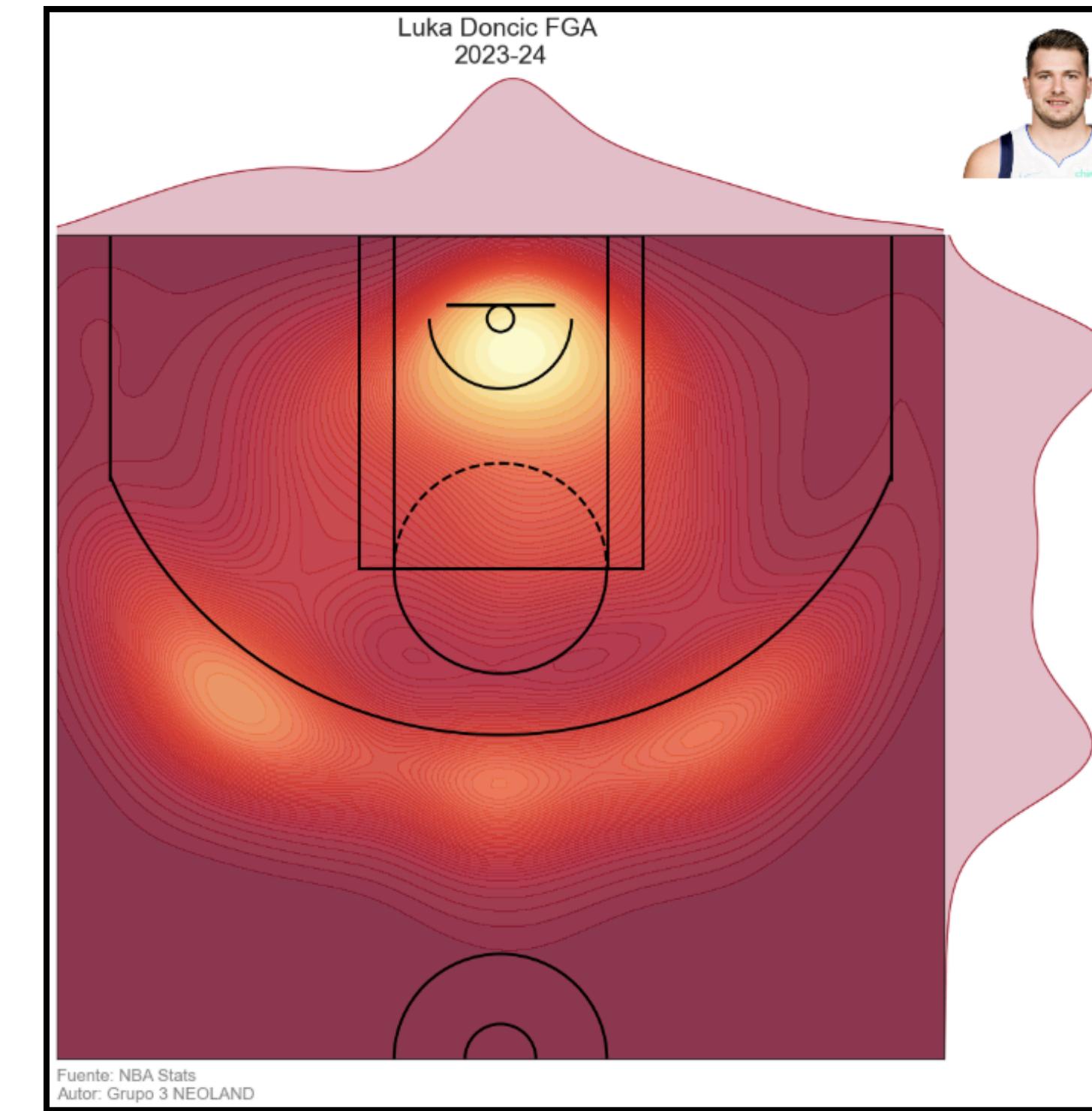
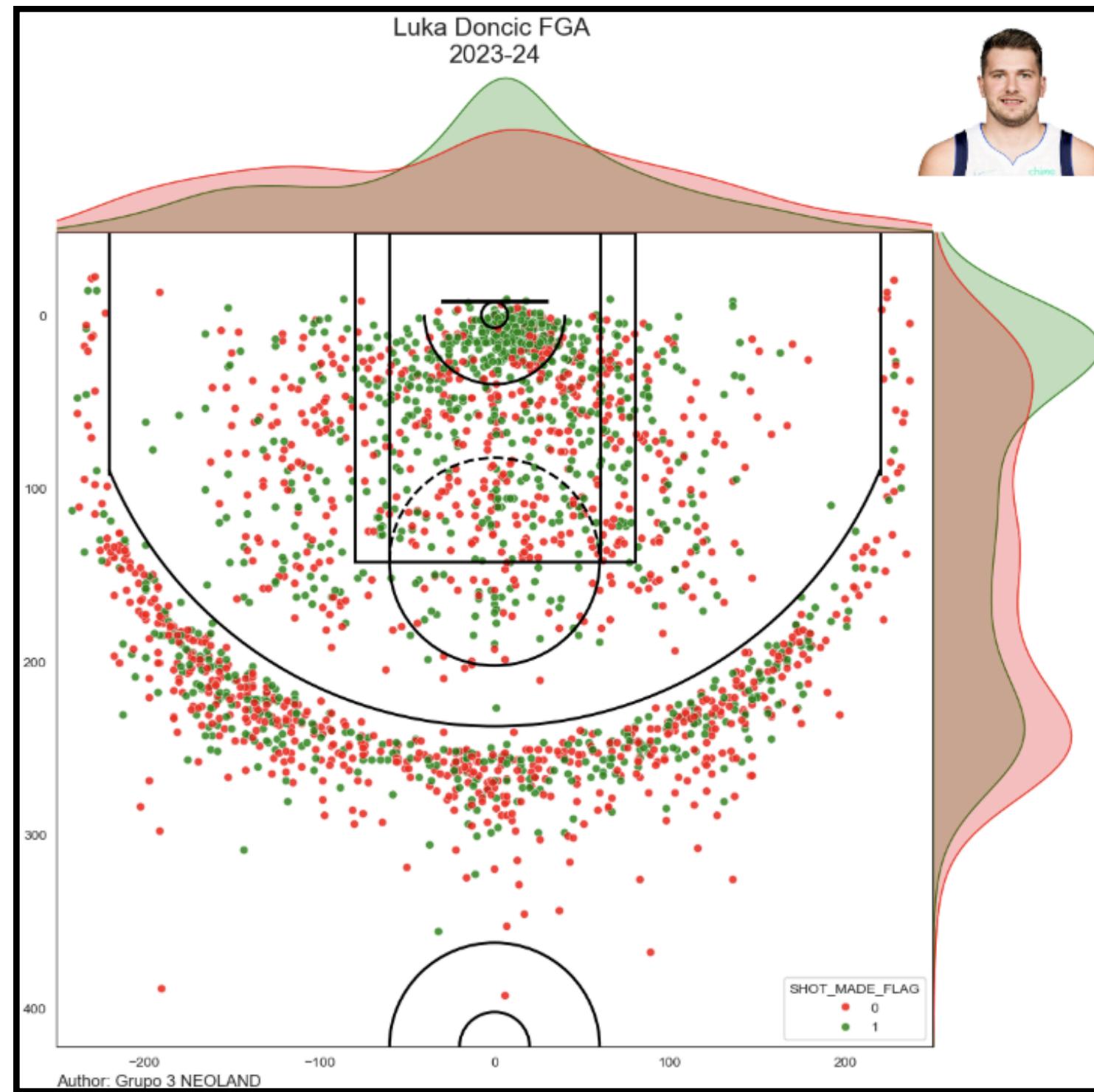
Jugador con más tiros de dos encestados: Nikola Jokic con 858 tiros de dos



Análisis exploratorio

Mapas de Calor

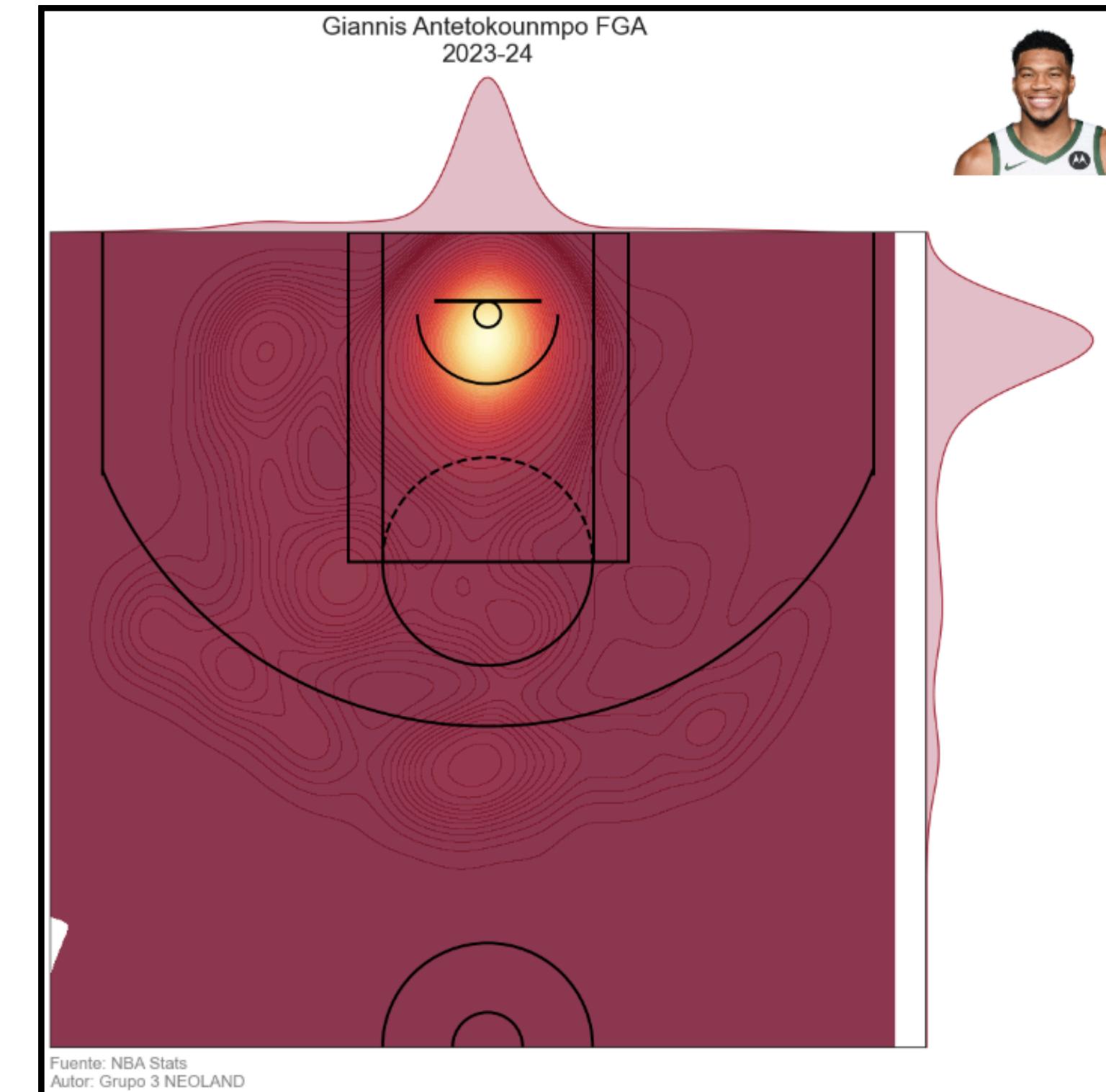
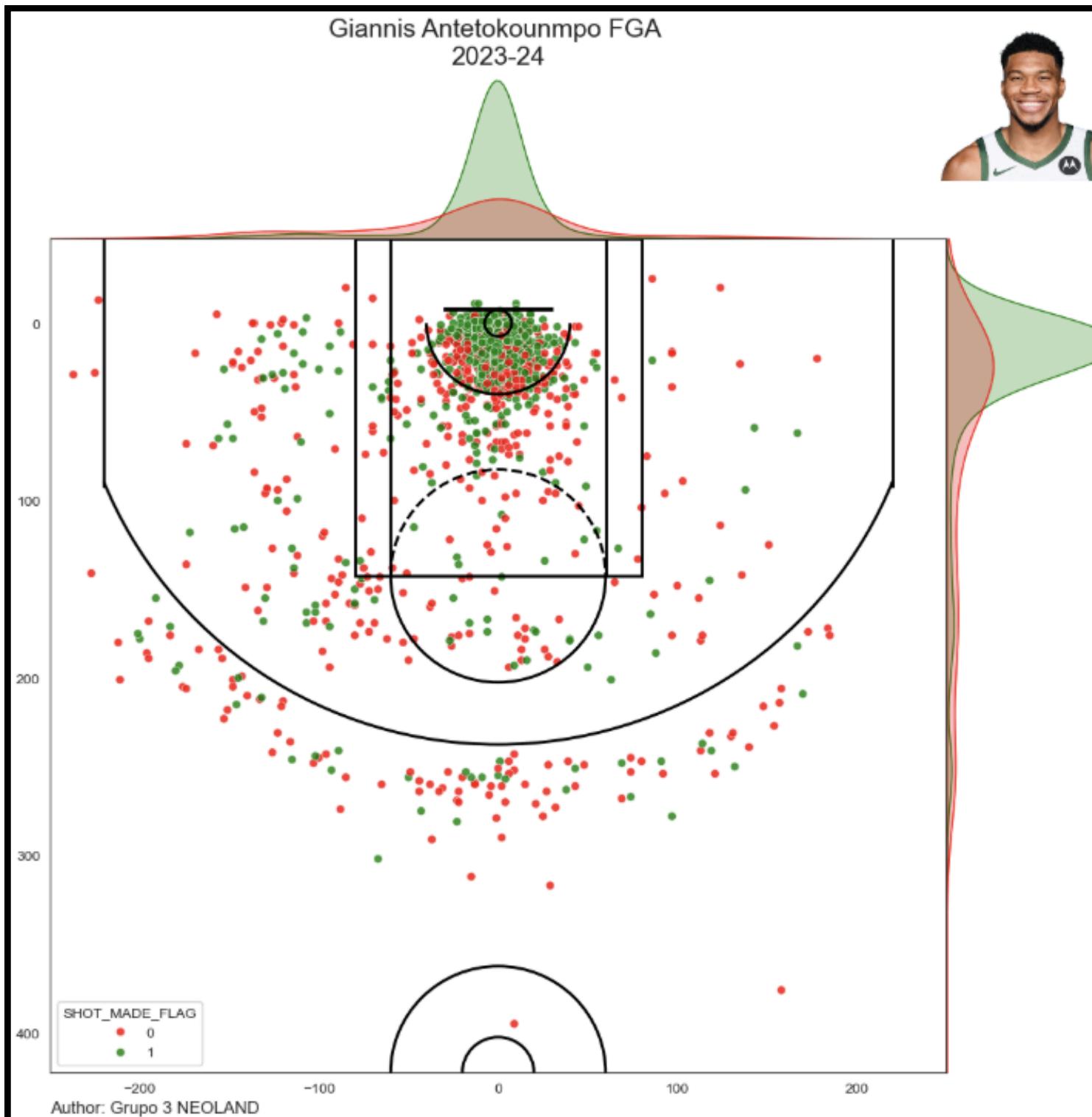
Jugador con más lanzamientos intentados: Luka Doncic con 2156 intentos



Análisis exploratorio

Mapas de Calor

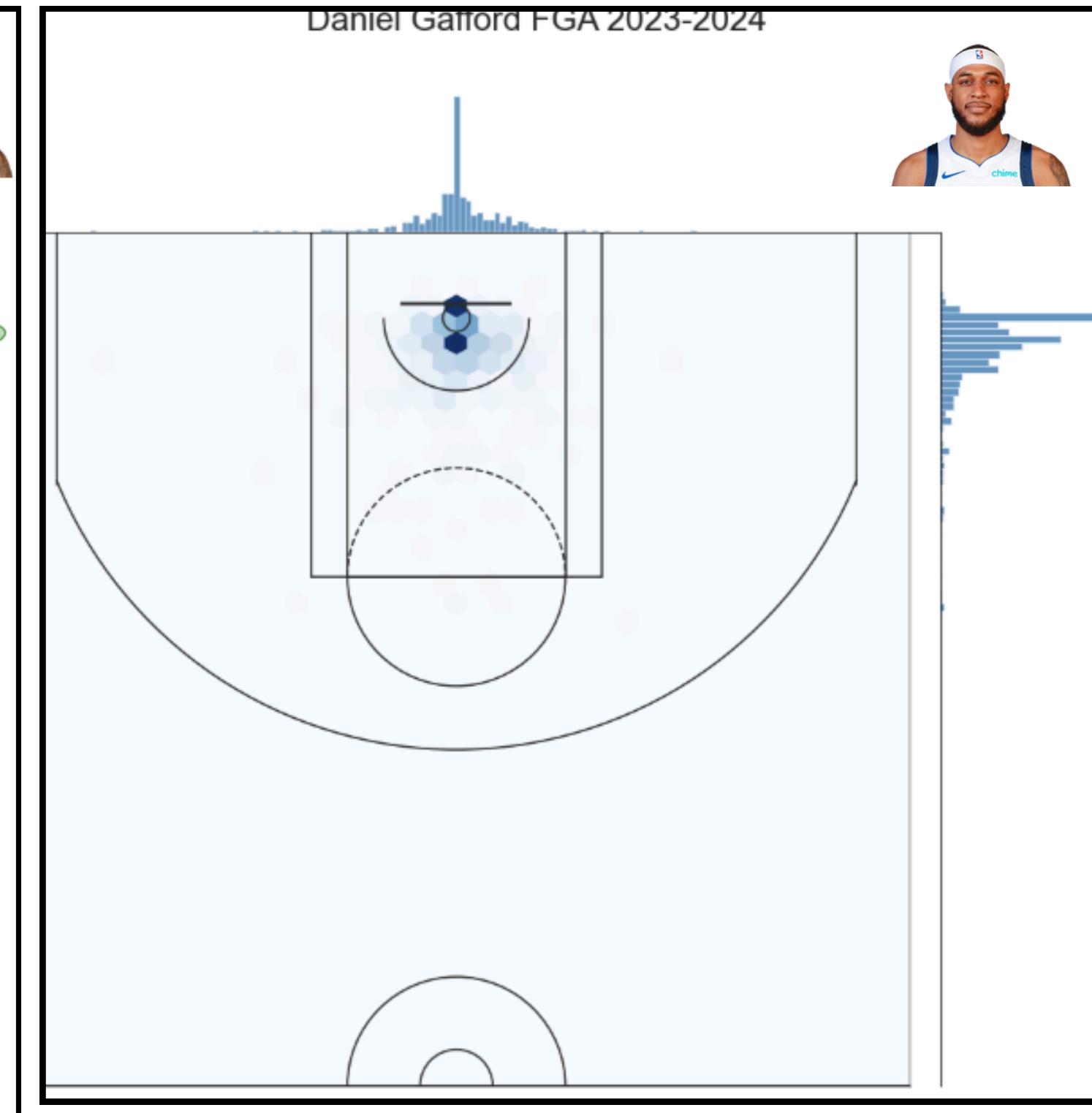
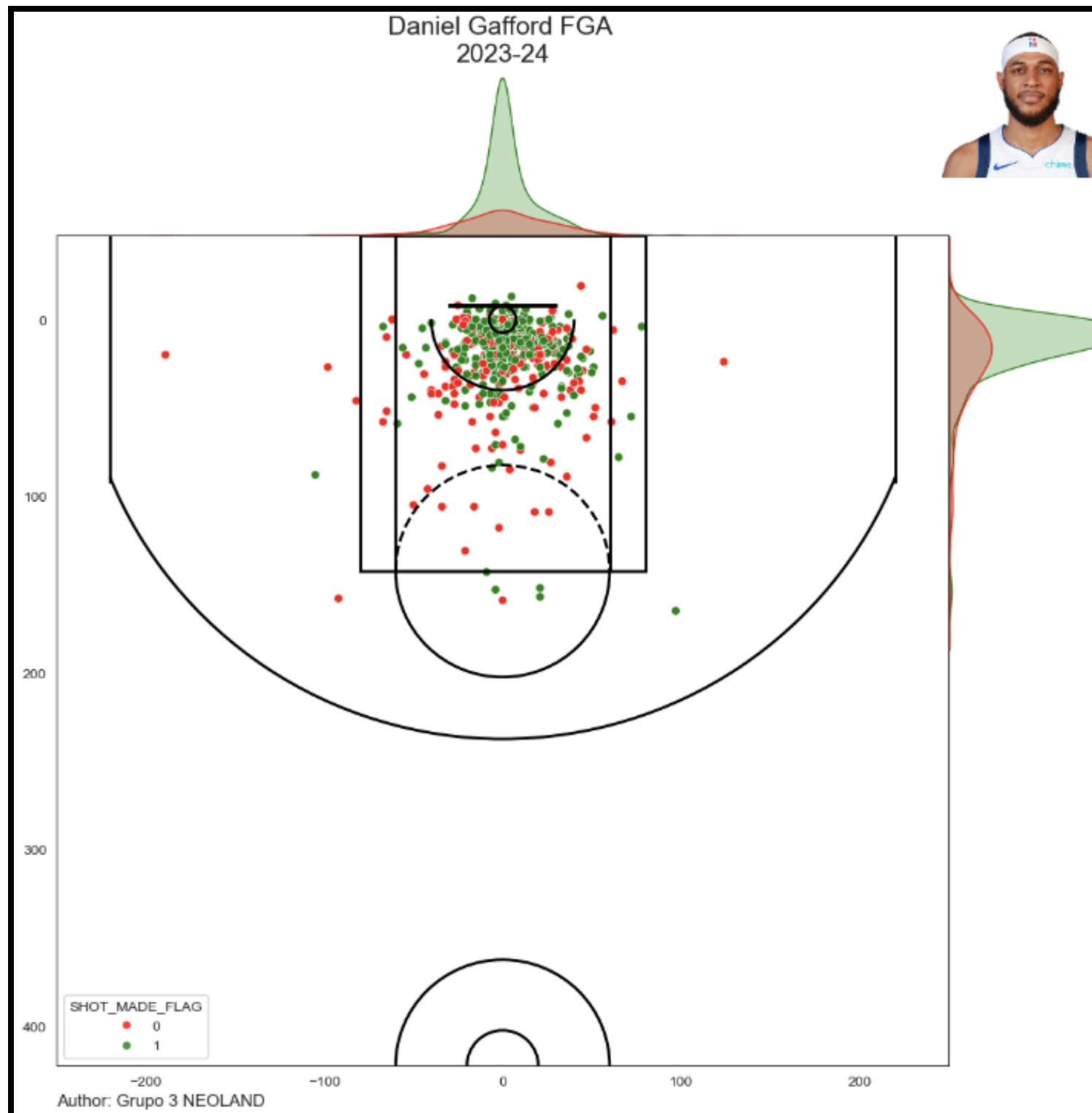
Jugador con el mejor promedio de aciertos partido : Giannis Antetokounmpo con 11.47 aciertos partido



Análisis exploratorio

Mapas de Calor

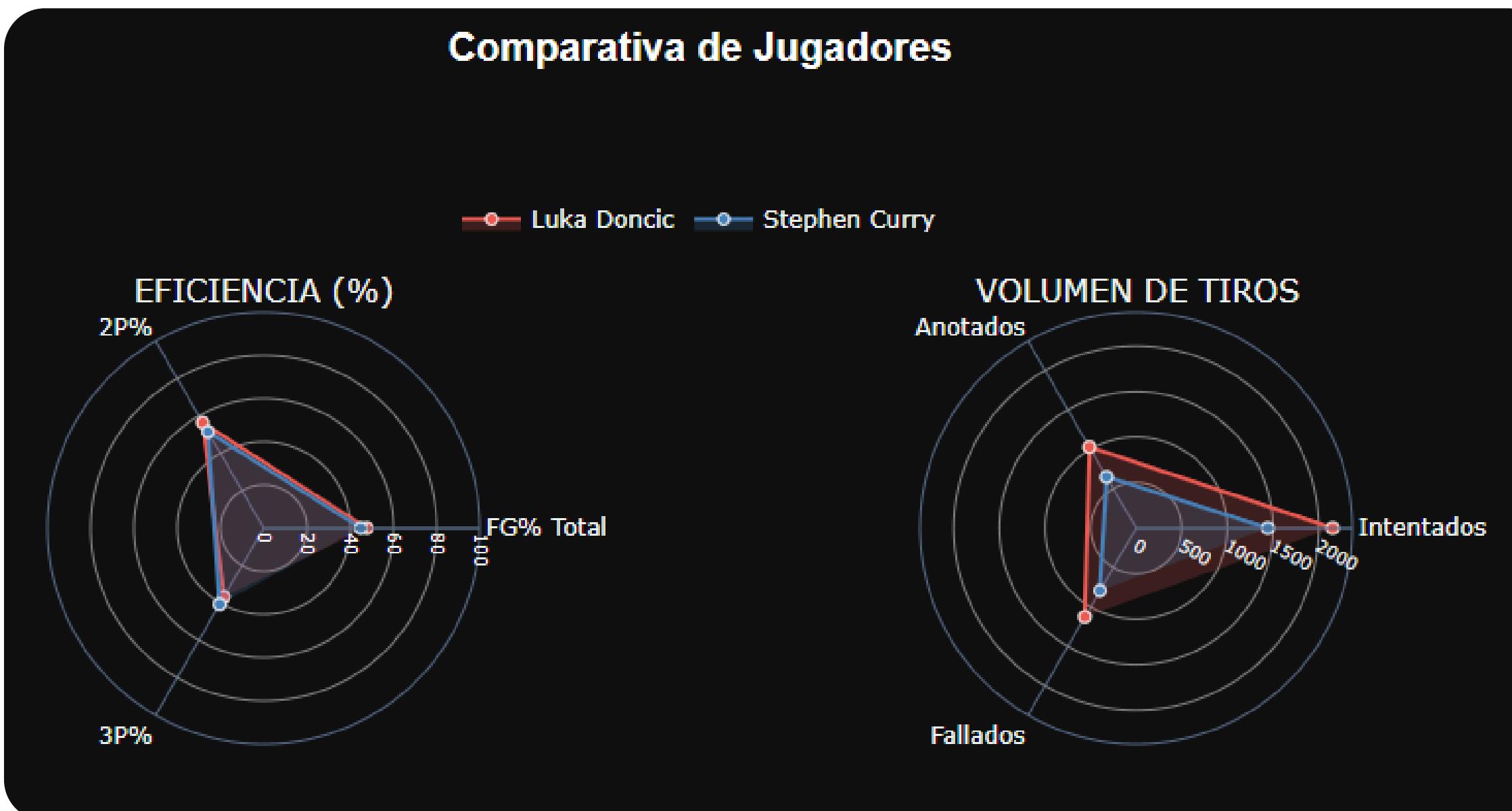
Jugador con mejor porcentaje acierto (mínimo 400 intentos): Daniel Gafford con un 70.65% de aciertos.



Análisis exploratorio

Comparativa Jugadores

Luka Doncic vs Stephen Curry (Jugador con mas triples intentados vs Jugador con mas triples encestados)



Volumen vs eficiencia en triples: Curry convierte 5 triples más que Doncic 357 vs 352 con 79 intentos menos.

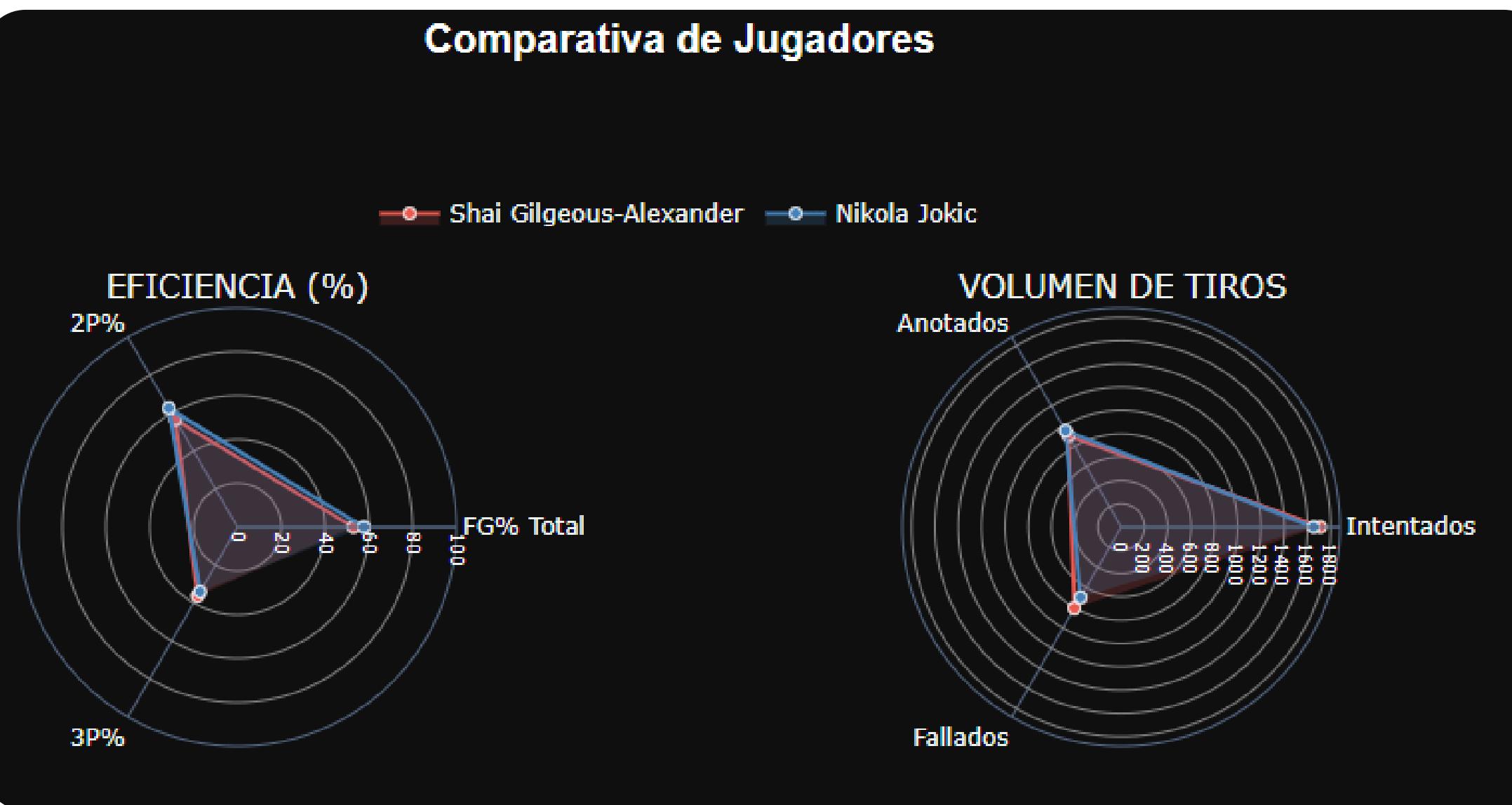
Impacto en el juego interior: Doncic domina en 2PT 56.37% de efectividad (677 canastas de dos puntos vs 293 de Curry)

Eficiencia Global: Doncic tiene mejor % total de tiros (47.73% vs 44.98%), gracias a su superioridad en tiros de 2PT.

Análisis exploratorio

Comparativa Jugadores

Shai Gilgeous-Alexander vs Nikola Jokic (Jugador de nuestro top con mas canastas de dos encestadas vs Lider de Liga)



Volumen vs eficiencia en triples: SGA: 111 triples (36.27%) en 306 intentos. Jokic: 97 triples (34.15%) en 284 intentos (menos volumen, similar eficiencia).

Impacto en el juego interior: Jokic: 858 canastas de 2PT (62.63%) en 1,370 intentos. GA: 796 canastas de 2PT (56.65%) en 1,405 intentos (Jokic domina con menos tiros).

Eficiencia Global: Jokić lidera en FG% total (57.74% vs 53.01%), gracias a su dominio absoluto cerca del aro.

Análisis exploratorio

Comparativa Jugadores

Jalen Brunson vs Grayson Allen (Jugador de nuestro top con mayor porcentaje de tiros de 3 vs Lider Liga)

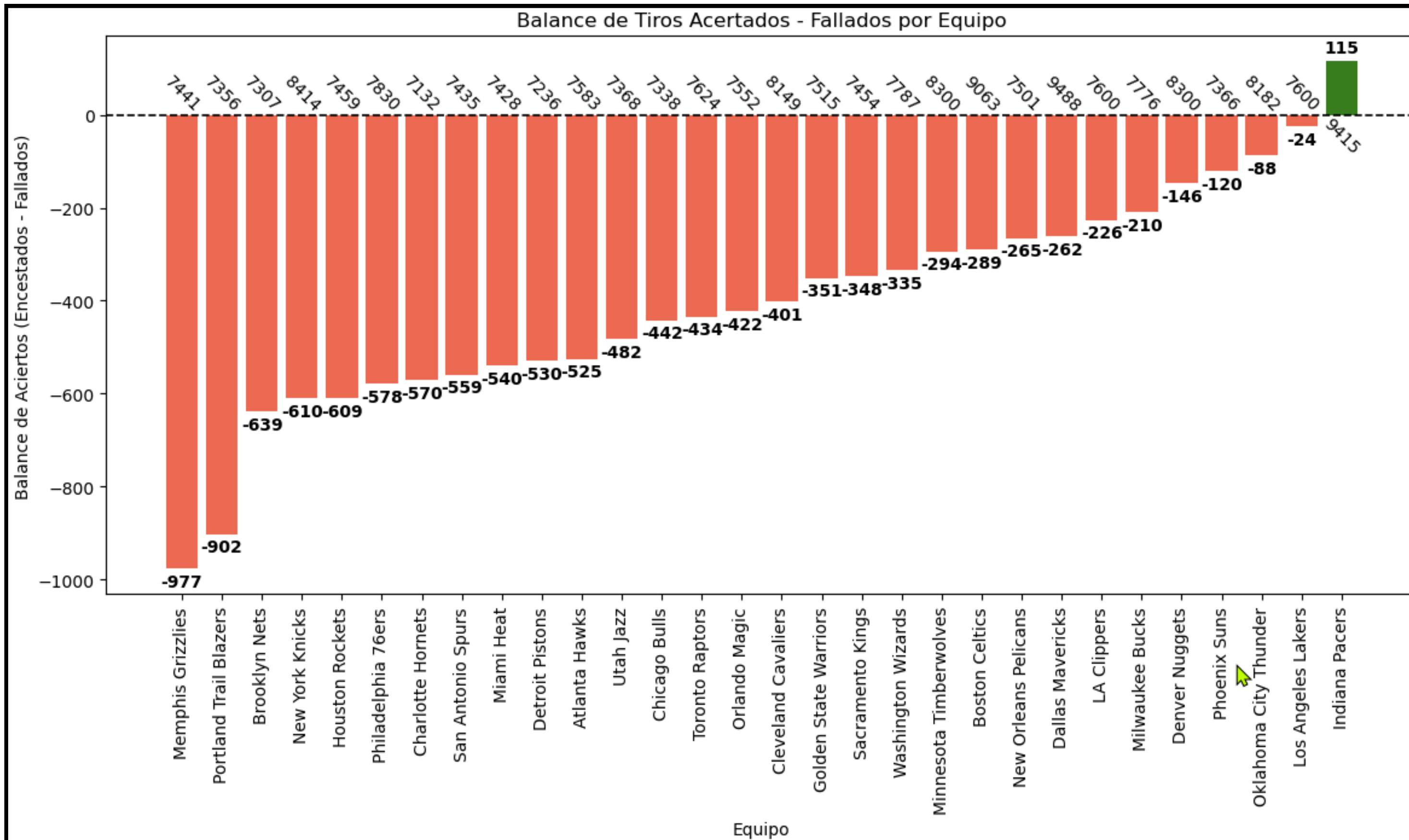


Volumen vs eficiencia en triples: Brunson: 237 triples (38.85%) en 610 intentos. Allen: 206 triples (45.78%) en 450 intentos (+7% eficiencia).

Impacto en el juego interior: Brunson: 704 canastas de 2PT (51.09%) en 1,378 intentos. Allen: 135 canastas de 2PT (56.96%) en 237 intentos (eficiencia táctica).

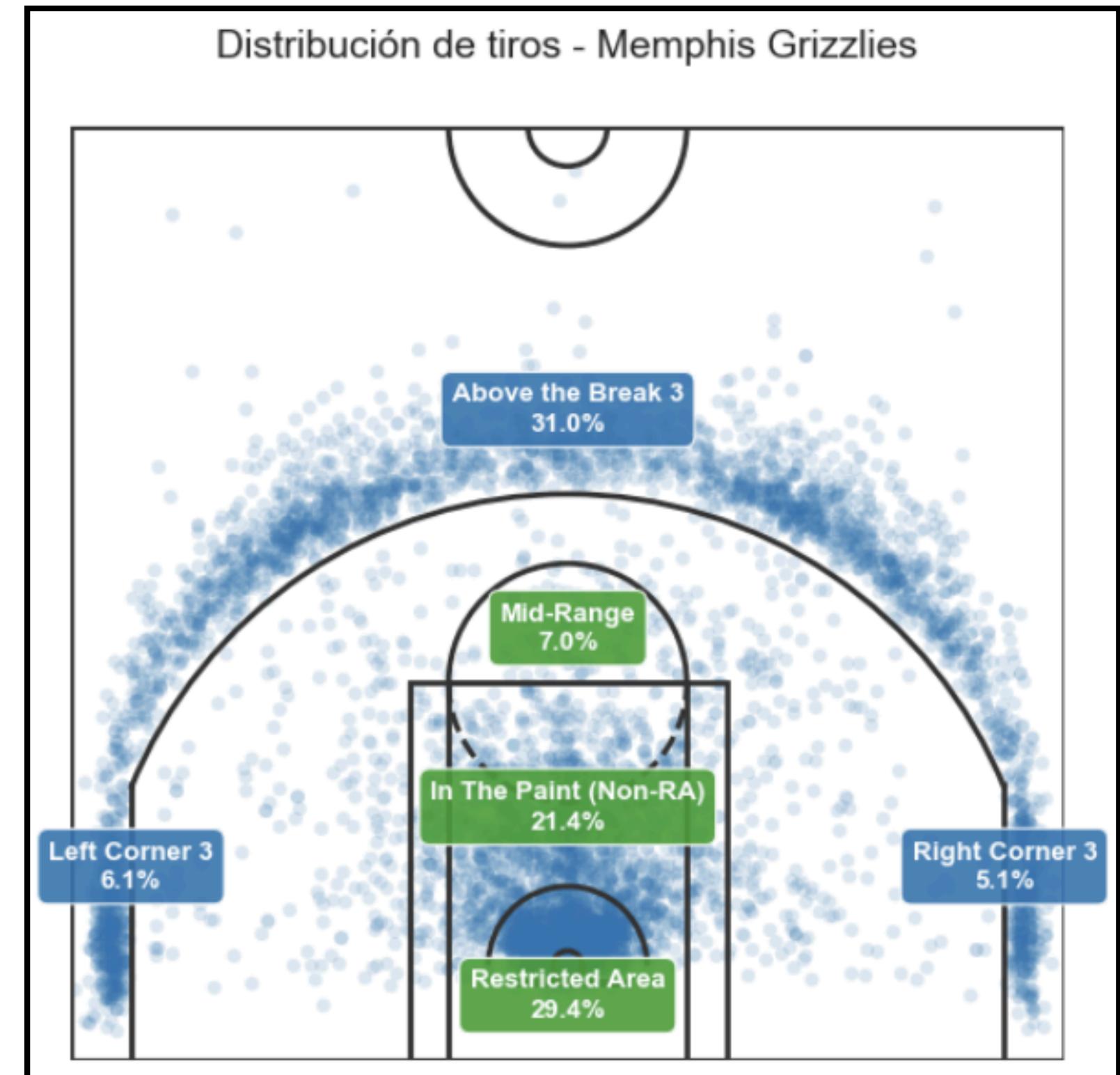
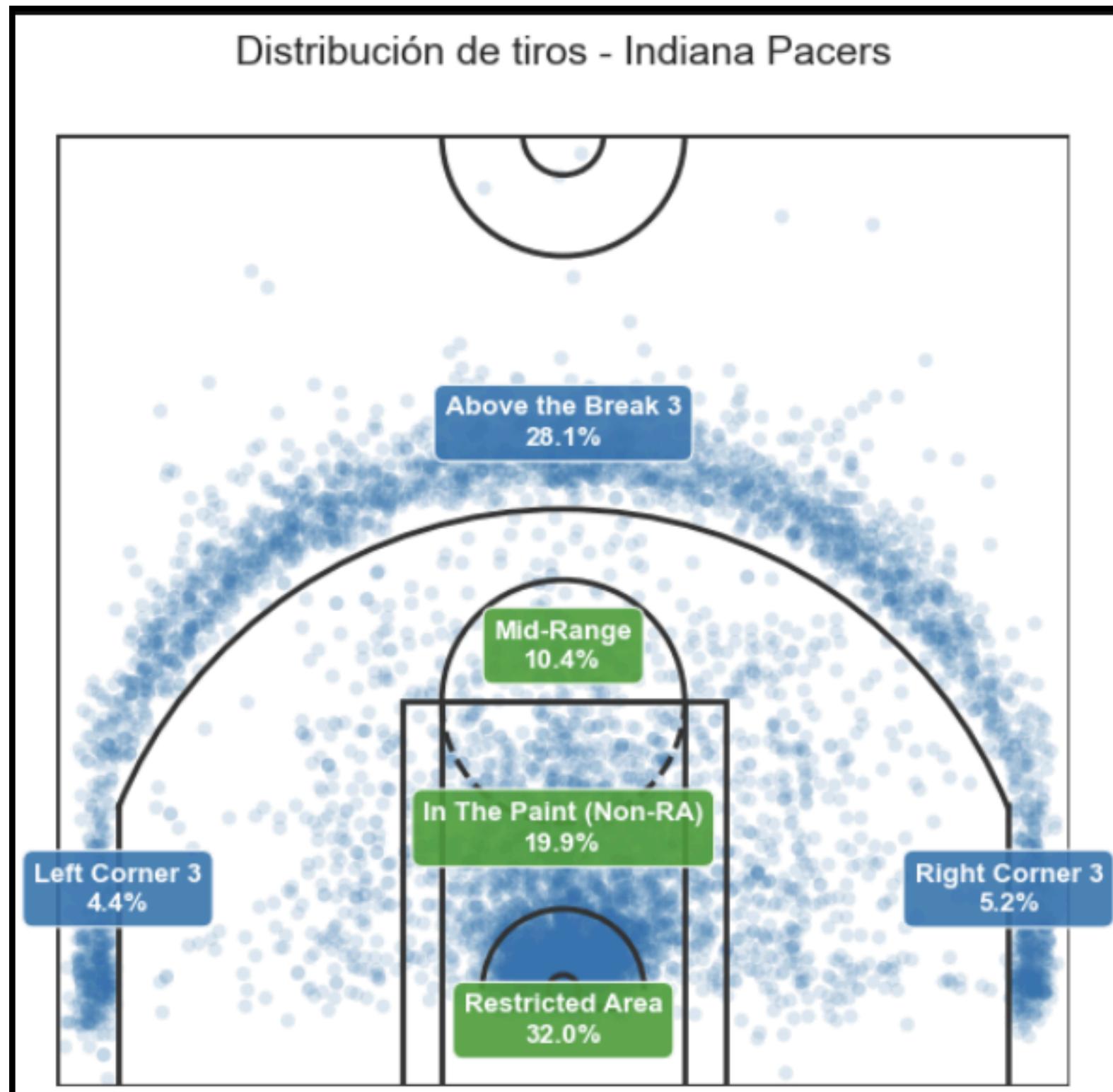
Eficiencia Global: Allen lidera (49.64% FG total vs 47.33% de Brunson), gracias a su precisión en 2PT y 3PT.

Análisis exploratorio



Análisis exploratorio

Distribución de tiros del mejor y peor equipo en base al balance de tiros encestados - fallados



Análisis exploratorio



Análisis exploratorio



Análisis exploratorio

JUGADORES PRINCIPALES

Luka Doncic

2156
Total_Intentos_Tiro

1029
Total_Tiros_Encestados

47,73 %
Porcentaje_Tiros_Encestados

TOTAL DE TIROS POR ZONA

Zona	Cantidad	Porcentaje
Backcourt	677	65,79%
Otro	352	34,21%

SHOT_TYPE
● 2PT Field...
● 3PT Field...

TIROS TOTALES POR ZONA

Zona	Cantidad
Above the Rim	344
In The Paint	252
Restricted Area	249
Mid-Range	176
Left Corner	5
Right Corner	3
Backcourt	0

TIROS TOTALES POR PERIODO

Periodo	Cantidad	Porcentaje
1	30,52%	
2	26,34%	
3	24,68%	
4	18,37%	
5	0,10%	

PERIOD ● 1 ● 2 ● 3 ● 4 ● 5



Análisis exploratorio



Análisis exploratorio

JUGADORES PRINCIPALES

Jalen Brunson

1988 Total_Intentos_Tiro

941 Total_Tiros_Encestados

47,33 % Porcentaje_Tiros_Encestados

TOTAL DE TIROS POR ZONA

SHOT_TYPE

- 2PT Field...
- 3PT Field...

237 (25,19%)
704 (74,81%)

TIROS TOTALES POR ZONA

Zona	Cantidad
In The Paint	328
Restricted Area	211
Above the Rim	201
Mid-Range	165
Right Corner	26
Left Corner	10
Backcourt	0

TIROS TOTALES POR PERIODO

PERIOD ● 3 ● 1 ● 2 ● 4 ● 5

0,43%
21,04%
22,53%
26,25%
29,76%

Análisis exploratorio

JUGADORES PRINCIPALES

Shai Gilgeous-Alexander

1711
Total_Intentos_Tiro

907
Total_Tiros_Encestados

53,01 %
Porcentaje_Tiros_Encestados

TOTAL DE TIROS POR ZONA

SHOT_TYPE

- 2PT Field...
- 3PT Field...

Zona	Cantidad
2PT Field...	796
3PT Field...	111

TIROS TOTALES POR ZONA

Zona	Cantidad
Restricted...	342
In The Pai...	251
Mid-Range	203
Above th...	100
Left Corn...	9
Right Cor...	2

TIROS TOTALES POR PERIODO

Periodo	Porcentaje
3	34,18%
1	31,09%
2	18,...
4	15,44%
5	0,99%
6	...

PERIOD ● 3 ● 1 ● 2 ● 4 ● 5 ● 6

Análisis exploratorio

JUGADORES PRINCIPALES

Anthony Edwards ▼

1882
Total_Intentos_Tiro

874
Total_Tiros_Encestados

46,44 %
Porcentaje_Tiros_Encestados

TOTAL DE TIROS POR ZONA

SHOT_TYPE
● 2PT Field...
● 3PT Field...

ZONA	CANTIDAD
Restricted...	355
Above th...	212
In The Pai...	148
Mid-Range	135
Left Corn...	16
Right Cor...	8
Backcourt	0

TIROS TOTALES POR ZONA

ZONA	CANTIDAD
Restricted...	355
Above th...	212
In The Pai...	148
Mid-Range	135
Left Corn...	16
Right Cor...	8
Backcourt	0

TIROS TOTALES POR PERIODO

PERIOD ● 3 ● 1 ● 4 ● 2 ● 5

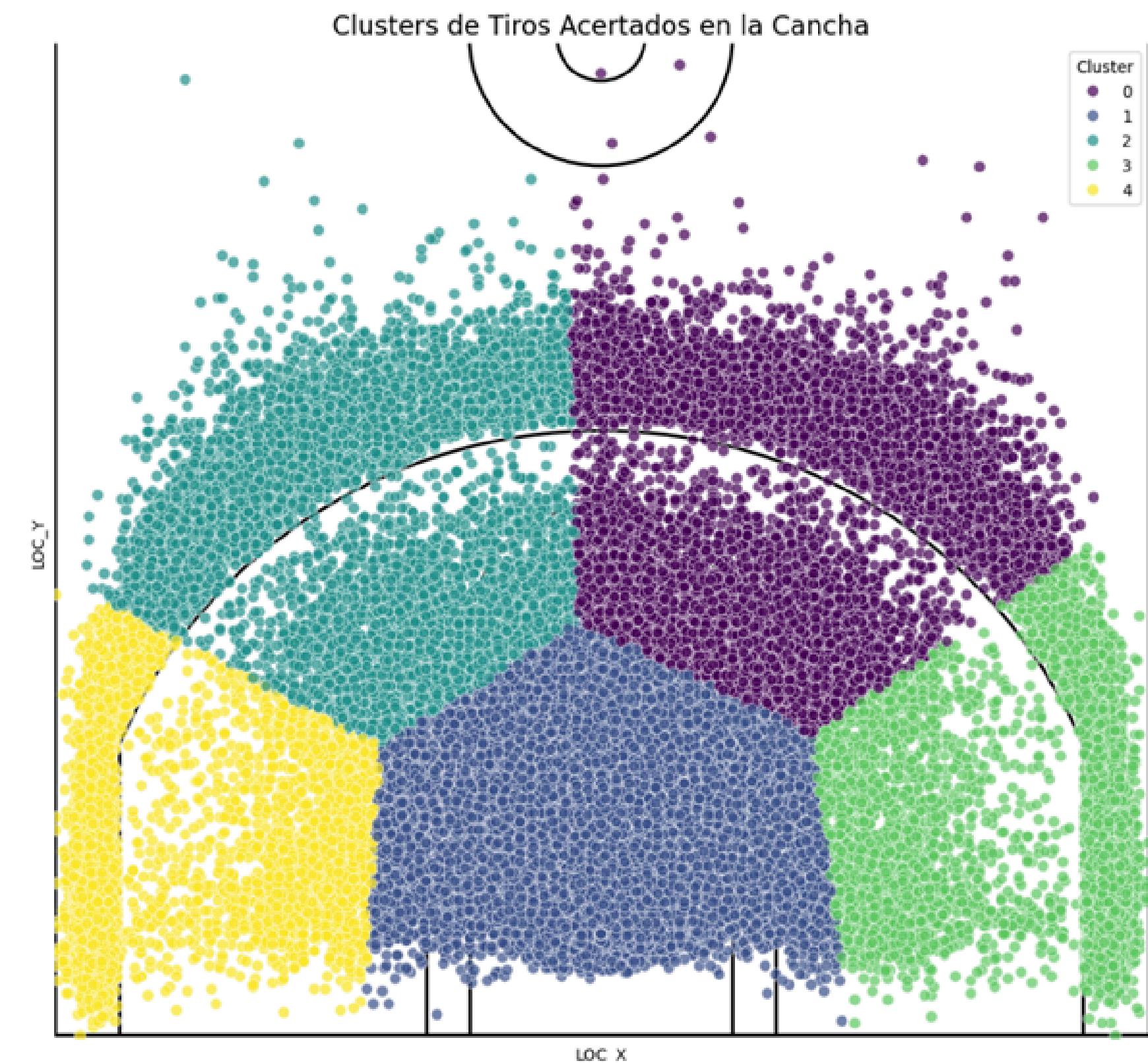
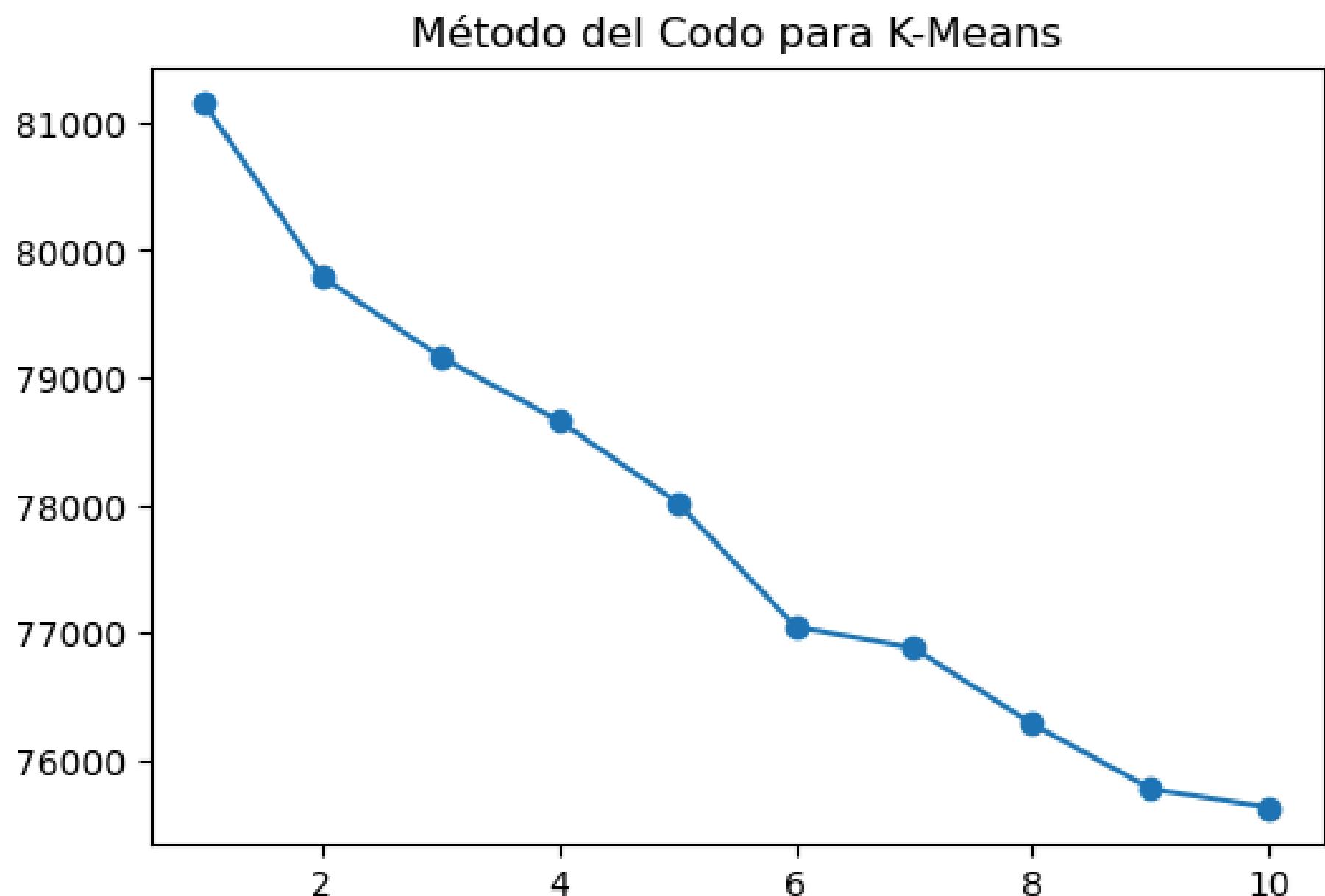


Análisis predictivo

- Clustering
- Regresión Lineal
- Regresión Logistica
- Random Forest
- Red Neuronal



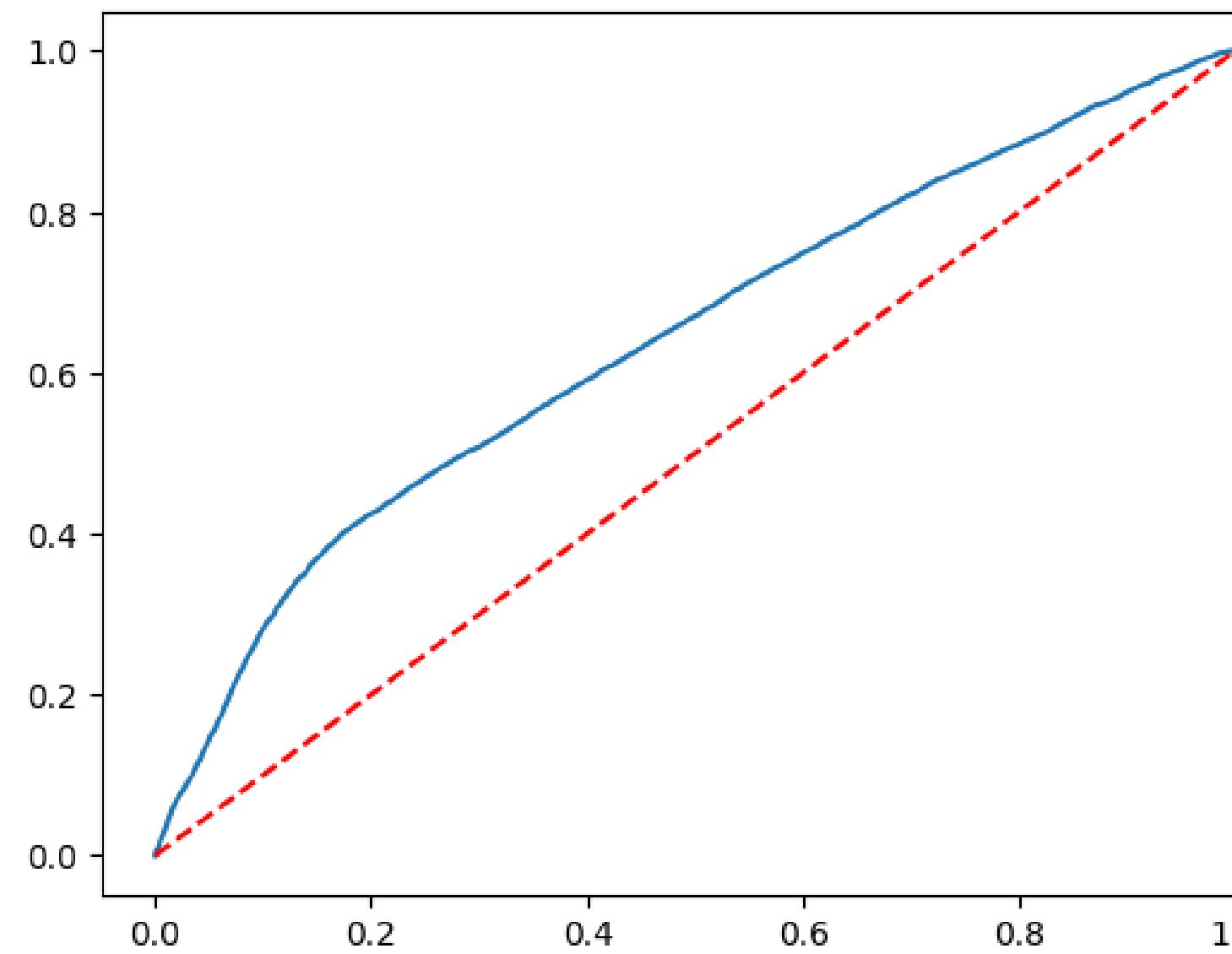
Clustering no supervisado



Modelos Predictivos

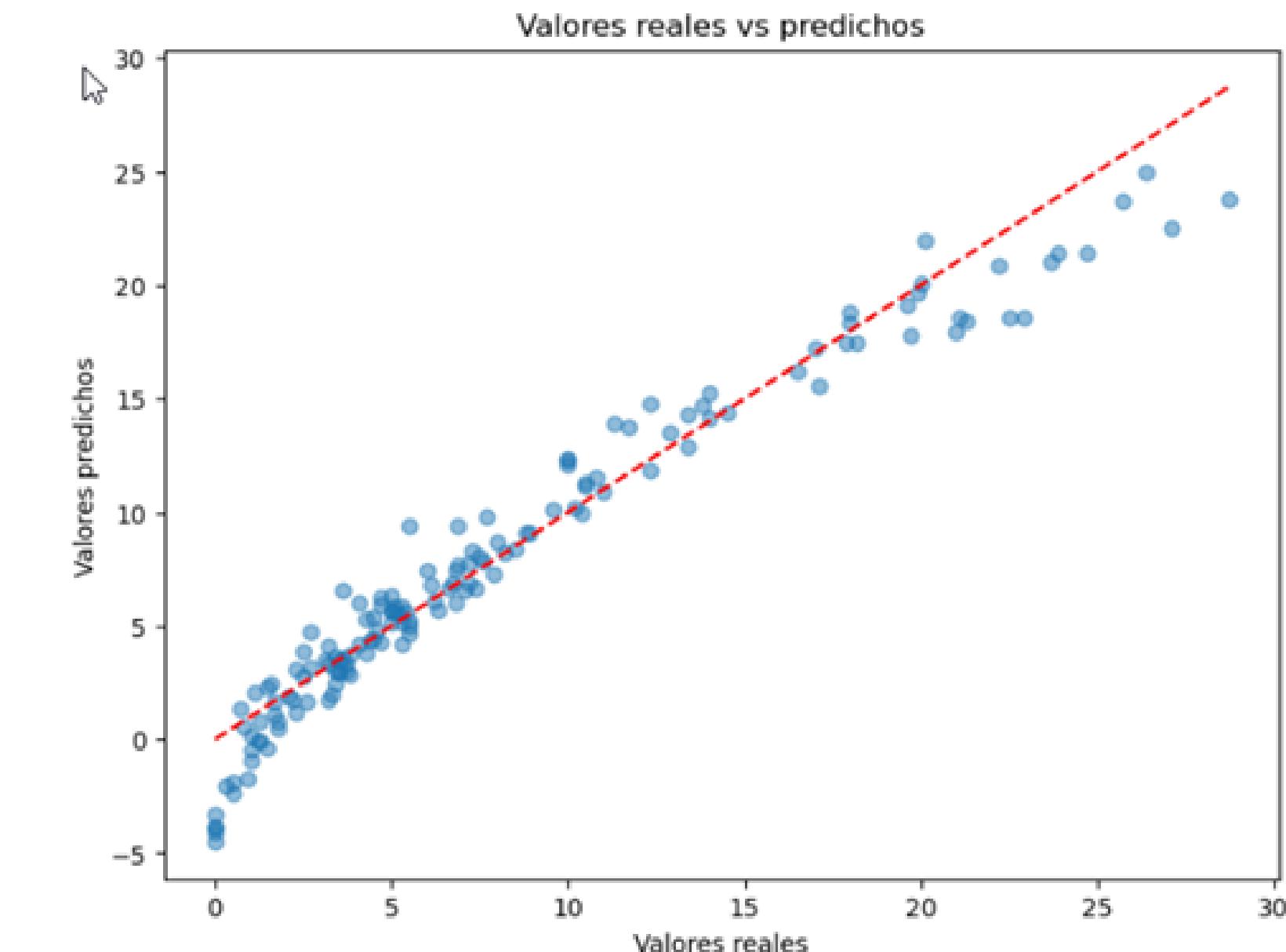
Predecir si un tiro será encestado

- Regresión logística
- Random forest
- Redes neuronales



Predecir puntuación de un partido

- Regresión lineal multiple



Conclusiones

Porcentaje 52.56% tiros fallados

Tendencia muy alta hacia tiros de 3 puntos

Menos tiros de media distancia

Evolución de los aleros

Puntos balanceados en cada cuarto



Next steps

- Herramienta para entrenadores
- Recomendaciones personalizadas
- Estudio del impacto de los jugadores
- Estrategias de equipos





Thank you!
