


EDA Rendimiento de los Chicago Bulls durante la temporada 2023-2024

Ángela Franco

Bootcamp Data Science 2024

- 
1. - Introducción y contexto
 - a. - Hipótesis
 2. - Exploración de los datos
 3. -Análisis global de los datos:
univariante, bivalente y
multivalente
 - a. Análisis estadísticas ofensivas
 - b. Análisis estadísticas defensivas
 4. - Conclusiones

1-Introducción y contexto

La NBA (*National Basketball Association*) es la liga de baloncesto más importante del mundo. Una industria que mueve millones de dólares cada año. Cada temporada, 30 equipos, divididos en conferencia Este y Oeste, se miden para llegar a los play-off y hacerse con el anillo del campeonato.

La temporada 2023-2024 fue especialmente dura para los Chicago Bulls, ya que su pobre rendimiento hizo que sus resultados no fueran los esperados.

Este proyecto se centra en analizar y comparar las estadísticas de la plantilla de jugadores de los Chicago Bulls para comprobar si el bajo rendimiento de los jugadores comparado con los jugadores del resto de equipos fue la causa del mal resultado del equipo durante la pasada temporada.

El dataset utilizado para llevar a cabo este proyecto se puede encontrar aquí: <https://www.kaggle.com/datasets/bryanchungweather/nba-player-stats-dataset-for-the-2023-2024/data>

1.1 Hipótesis

Las hipótesis de este proyecto se centran en 2 ámbitos: el desempeño ofensivo y el desempeño defensivo de los jugadores:

Hipótesis sobre el rendimiento ofensivo:

1. Los jugadores de los Chicago Bulls tienen menos puntos por partido (PPG) de media que el resto de jugadores por una contribución ofensiva pobre.
2. Los jugadores de los Chicago Bulls tiene un volumen menor de tiros de 2 (2P) y tiros de 3 (3P) comparado con los jugadores de otros equipos. Esto puede deberse a dificultades en el ataque.

Hipótesis sobre el rendimiento defensivo:

1. Los jugadores de los Chicago Bulls tienen de media menos rebotes defensivos por partido (REB) que el resto de jugadores. Perder los rebotes defensivos contribuye a que el equipo contrario tenga más posibilidades de anotar.

2. Los jugadores de los Chicago Bulls tienen menos robos (STL) y bloqueos (BLK) que el resto de jugadores. De nuevo, esto contribuye a la anotación del equipo contrario.

2-Exploración de los datos

Primero se ha llevado a cabo una exploración general y limpieza de los datos. En este dataset, había nulos y repeticiones de jugadores. Esto se debe a que si un jugador ha sido traspasado durante la temporada, sus registros aparecen tantas veces como haya sido traspasado, pero bajo el nombre de su nuevo equipo y con sus estadísticas correspondientes.

Respecto a los nulos, aparecen en 4 columnas: 49 en 3P% (porcentaje de tiros de 3); 17 en 2P% (porcentaje de tiros de 2); 8 eFG% (porcentaje de tiros de campo efectivo) y 62 en FT% (porcentaje de tiros libres). Las columnas 3P%, 2P% y FT% se han eliminado del dataset ya que no son necesarias para el análisis. Los 8 nulos de eFG% se han mantenido ya que no afectaba a jugadores de los Chicago Bulls.

Respecto a la frecuencia de aparición de los valores únicos. Se observa que las frecuencias de aparición de son muy bajas en casi todo el dataset ya que se trata de estadísticas individuales, por lo que hay una alta cardinalidad.

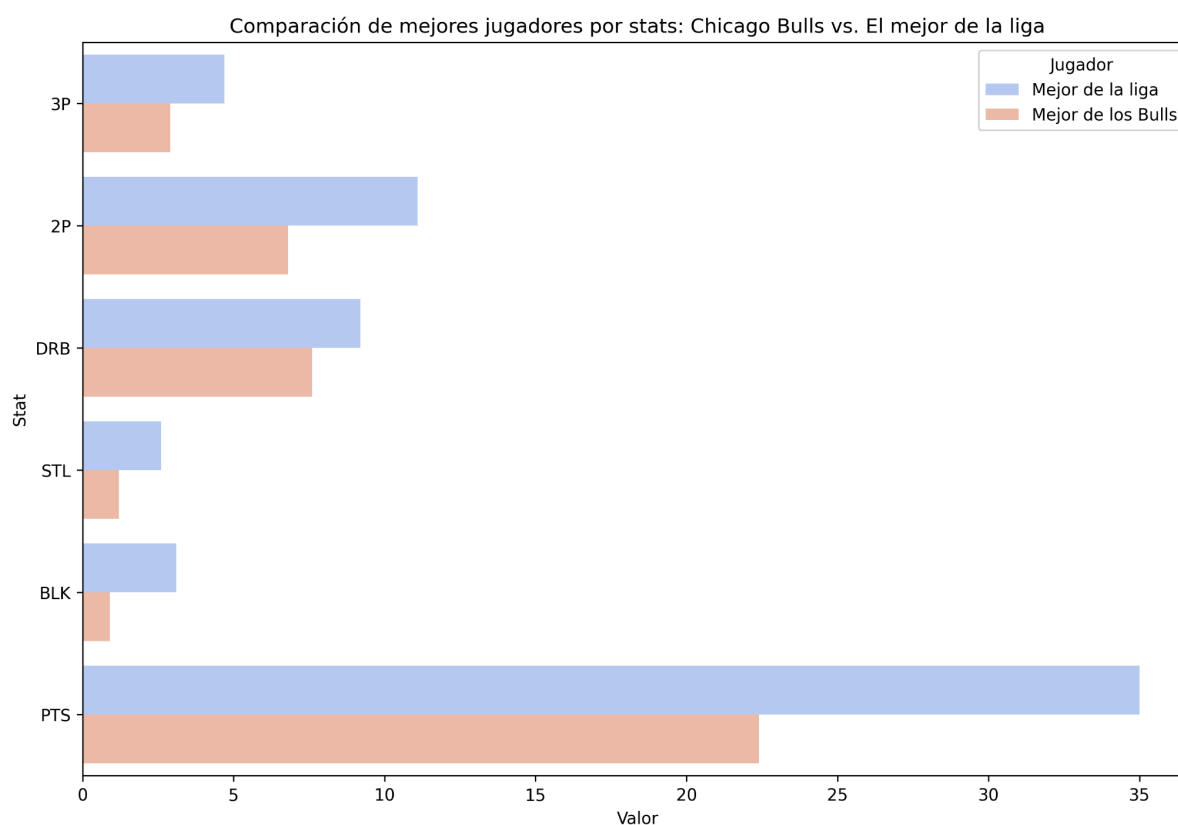
Respecto a la tipificación de los datos y clasificación de las variables, dado el número elevado de columnas, esto se ha realizado en la hoja excel: Resumen_columnas. En esa hoja se han categorizado y explicado las columnas, además se ha identificado las columnas target para realizar el análisis de las hipótesis.

Respecto a las medidas de tendencia central, el dataset ya está formado por estadísticas medias por jugador de la pasada temporada. Pero sí se han extraído los jugadores con los valores más altos para las columnas de interés para poder compararlo con las estadísticas de los jugadores de los Chicago Bulls a modo de comparativa general.

3-Análisis de los datos: univariante, bivariante y multivariante

Primero se realiza una comparación para entender el salto de nivel que hay entre los mejores jugadores de la liga y los jugadores de los Chicago Bulls para visualizar la diferencia de capacidades, lo cual también está ligado al mal resultado del equipo.

Para ello se seleccionan las columnas directoras (marcadas con un 1 en el excel Resumen_columnas) y se obtiene el mejor jugador de la liga para esas columnas y el mejor jugador de los Bulls, luego con un gráfico de barras se muestra visualmente la diferencia.



Como se observa, hay una gran diferencia entre las estadísticas de los mejores jugadores y las estadísticas de los mejores jugadores de los Bulls.

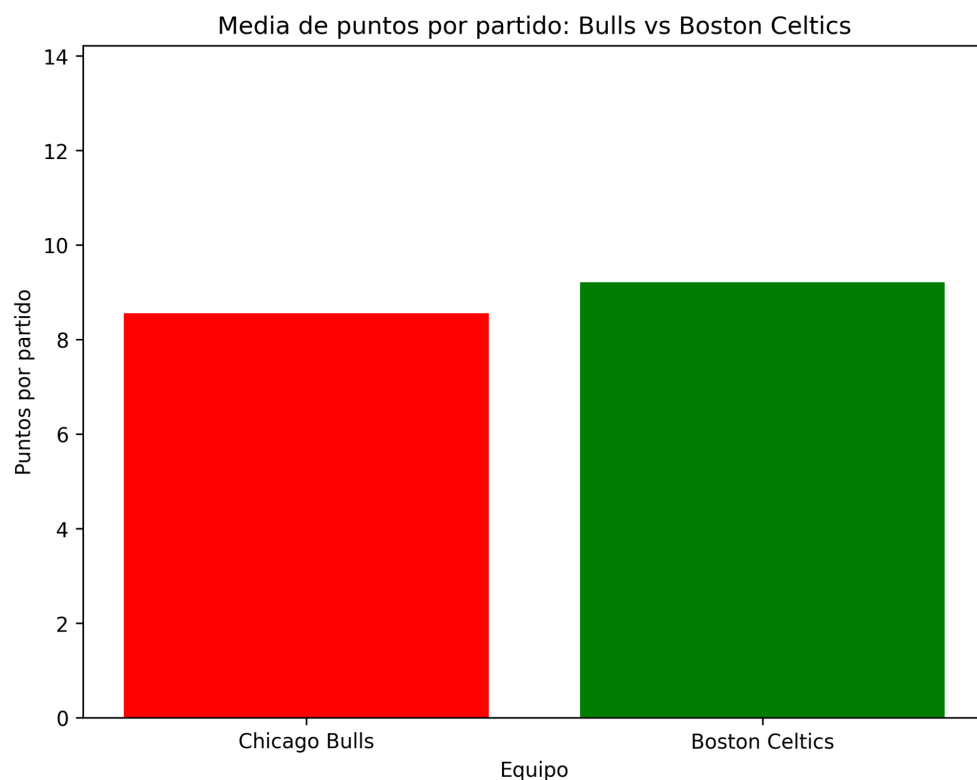
Una vez realizado este primer análisis visual y general, pasamos a realizar análisis específicos de las columnas de interés para verificar nuestras hipótesis.

a. Análisis estadísticas ofensivas

En primer lugar, se analiza la eficiencia de anotación entre los jugadores de los Bulls y el resto de equipos de la liga. Para esto se lleva a cabo un análisis univariante sobre la columna PTS y se genera un boxplot. Primero se realiza el análisis sobre el resto de equipos de la liga y luego sobre el ganador de la conferencia Este los Boston Celtics.

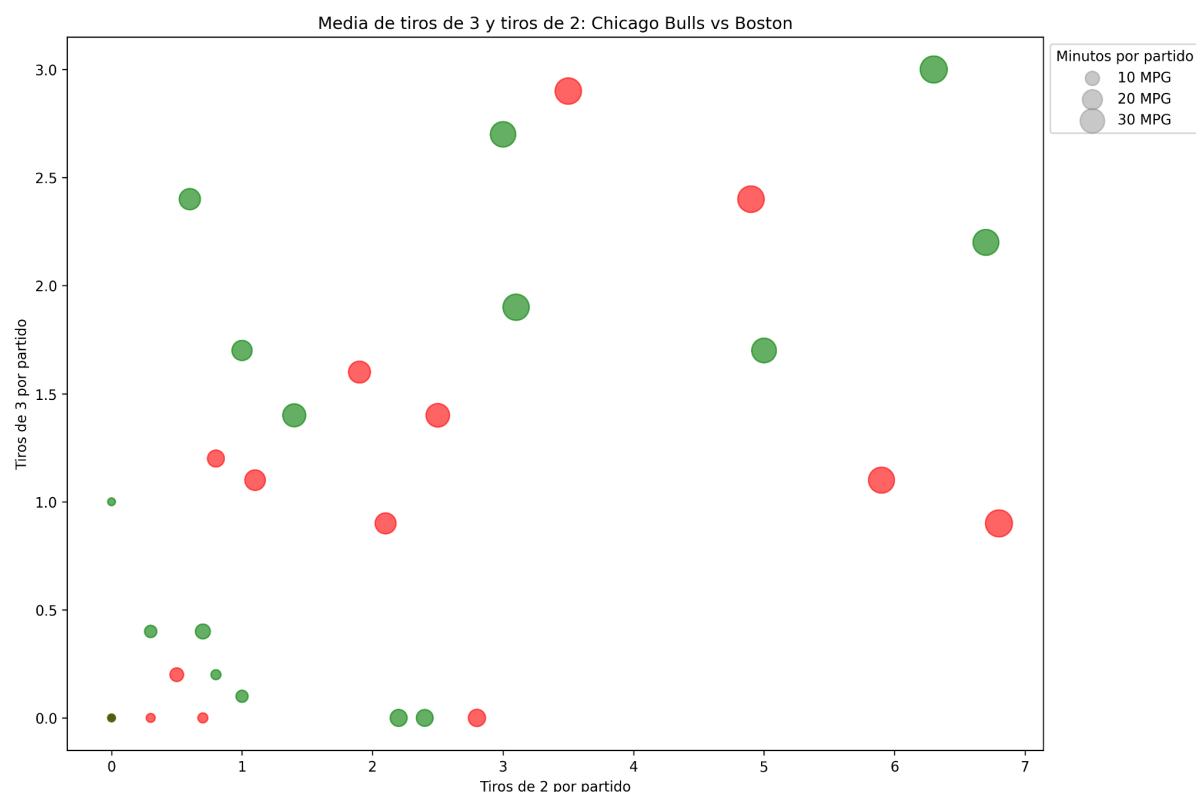
El resultado del boxplot muestra que, comparado con el resto de la liga, la mediana de los puntos por partido de los Bulls es ligeramente superior a la del resto de equipos. Además, es más consistente, teniendo un rango más compacto y menos outliers, pero es cierto que en este caso se ha realizado la comparación con los 29 equipos restantes.

Cuando la comparación se realiza solo vs Boston Celtics, es notable la diferencia en el rango de los bigotes, siendo la de los jugadores de Boston muy superior. Aunque la media es similar, se observa que el equipo de Boston anota más puntos por partido. El resultado se muestra con un gráfico de barras para que sea más sencilla su interpretación.

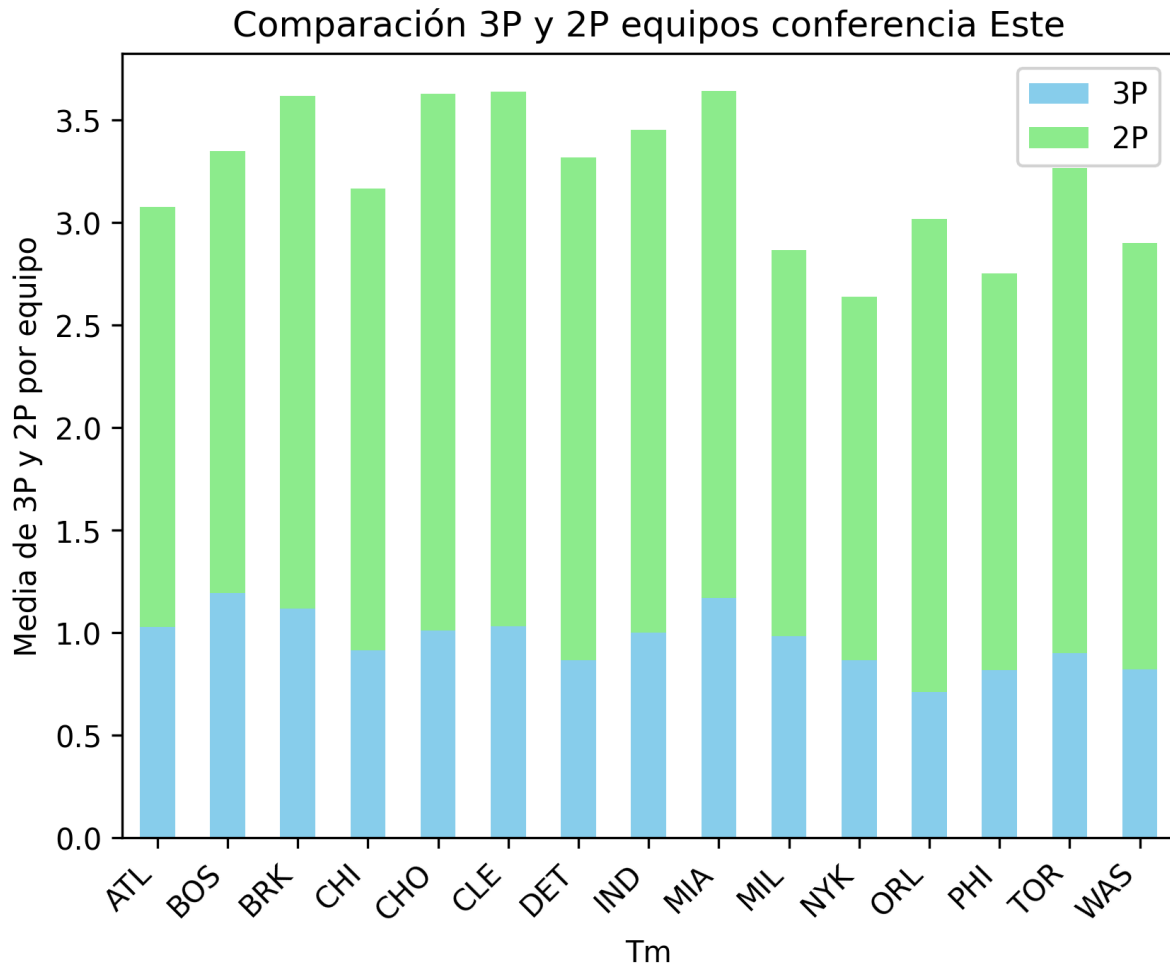


Con esto, la primera hipótesis se confirma parcialmente. Es cierto que comparado con el resto de equipos de la liga, el desempeño en cuanto a puntos por partido es superior en el caso de los Bulls, pero, cuando se compara con el ganador de la conferencia Este, la media de puntos por partido de los Bulls es menor y esto puede haber contribuido al mal resultado de la temporada anterior.

A continuación se realiza el análisis bivalente del número de tiros de 2 (2P) y de tiros de 3 (3P) entre los Bulls comparado con Boston para comprobar la segunda hipótesis. En este caso se ha añadido una tercera columna en el análisis, el tamaño de las burbujas está definido por los minutos jugados de media por partido de cada jugador.



El gráfico no muestra una diferencia notable respecto al desempeño entre los jugadores de ambos equipos. Se realiza otra comparación, esta vez incluyendo solo los 15 equipos de la conferencia Este y se muestra el resultado en un gráfico de barras.



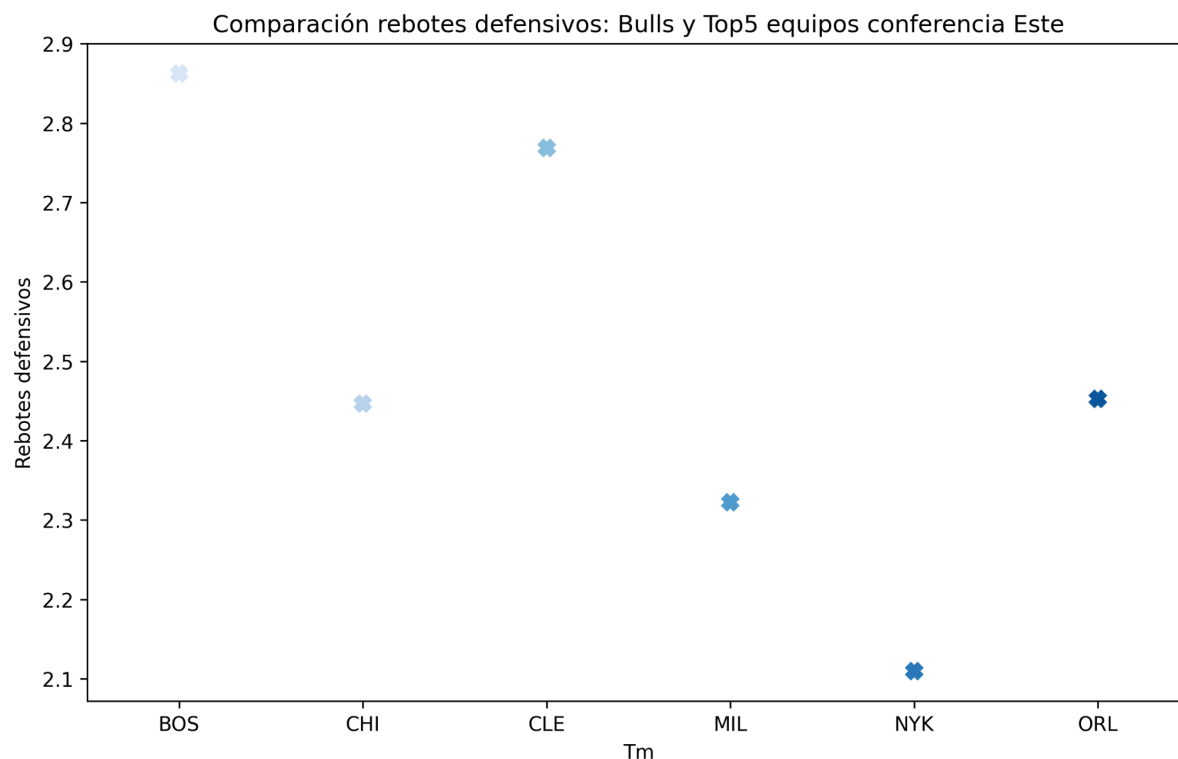
Se observa mucha variabilidad entre los equipos. Respecto a la media de tiros de 3, tanto los New York Knicks como Philadelphia por ejemplo muestran valores menores a los Bulls y lo mismo sucede con los valores para tiros de 2 puntos, en este caso incluyendo también a Orlando Magic.

Por lo que la segunda hipótesis se confirma parcialmente, los jugadores de los Bulls han anotado menos tiros de 3 y tiros de 2 que muchos de los equipos de la conferencia pero no son el equipo con los peores resultados, de hecho, equipos cuya clasificación fue superior en la conferencia, tienen peores resultados en cuanto a tiros de 2 y de 3.

b. Análisis estadísticas defensivas

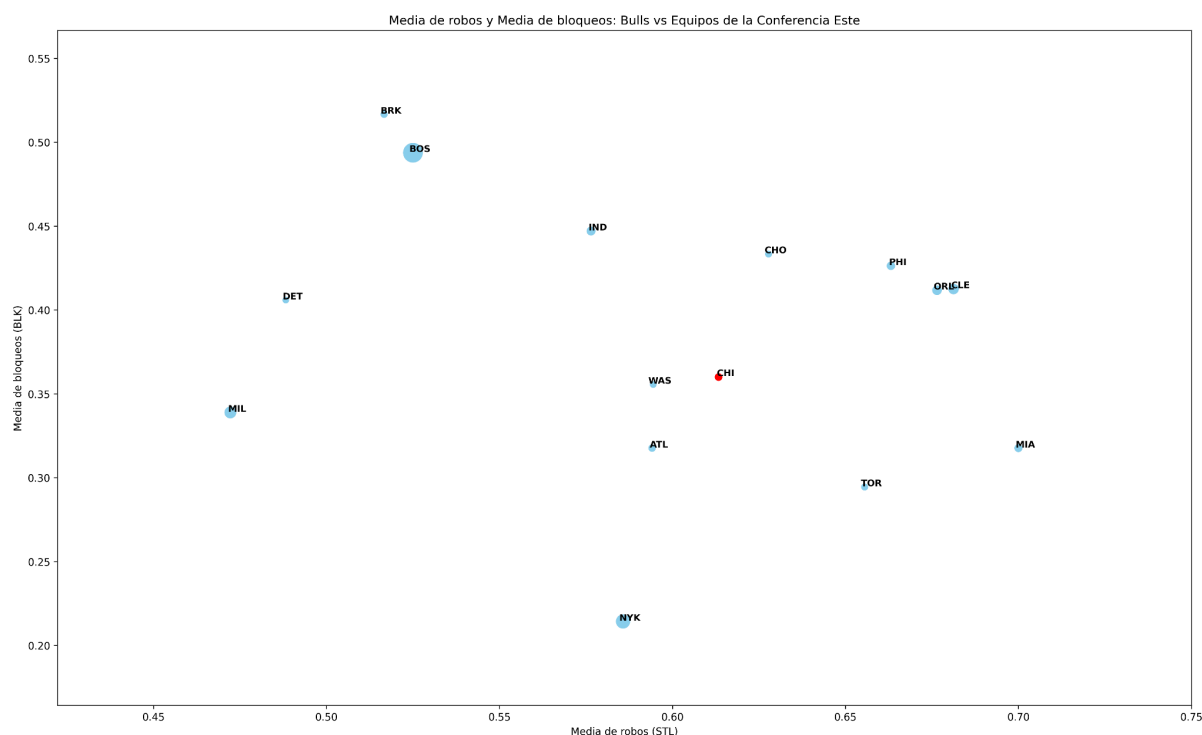
Respecto a los rebotes defensivos por partido, este valor influye también en el rendimiento del equipo ya que perder los rebotes contribuye a que el equipo

contrario tenga otra posesión y pueda anotar de nuevo. En este apartado se analizan solo los rebotes defensivos y se compara la media de rebotes defensivos de los Bulls vs el top 5 de equipos de la conferencia Este en la temporada 23-24.



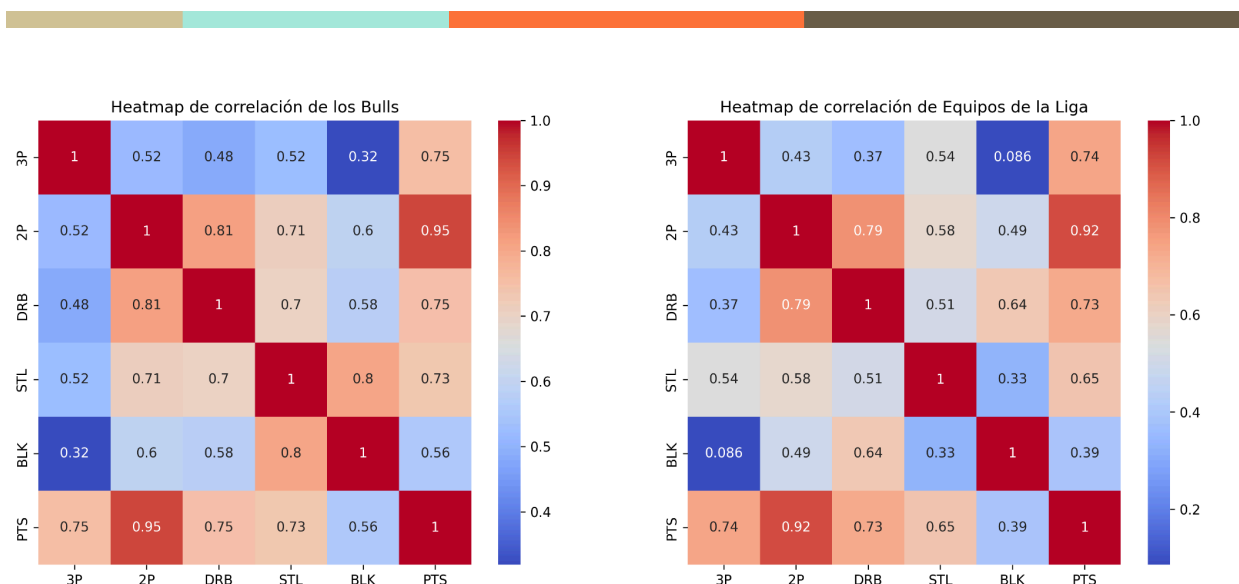
De nuevo, se aprecia que los Chicago Bulls no son el peor equipo respecto a rebotes defensivos, con una media de casi 2,5 por partido. Tanto Milwaukee como New York Knicks y Orlando tienen una media inferior por lo que no parece que haya relación directa entre la media de rebotes defensivos por partido y el rendimiento del equipo. Aunque es cierto que Boston es el equipo con mejor media y fueron los campeones de conferencia, el resto de equipos, excepto los Bulls, estuvieron en el top 5 de la conferencia.

En relación a la última hipótesis: los jugadores de los Chicago Bulls realizan menos robos y tapones que el resto de jugadores de otros equipos, se realiza un análisis bivalente de estas dos columnas para comparar la media de los Bulls para estos dos valores vs la del resto de equipos de la conferencia Este. En esta ocasión, se ha añadido el nombre de los equipos y el tamaño de los puntos determina su posición en la clasificación del año pasado.



Como se observa, el desempeño de los jugadores de los Bulls para estas dos estadísticas está lejos de ser el mejor pero tampoco es el peor, por lo que la cuarta y última hipótesis tampoco se cumple por completo. Aunque la media de robos y bloqueos es importante para contribuir al buen desempeño del equipo, hay equipos como Boston cuya media de robos fue peor pero sin embargo su clasificación fue mucho mejor o New York Knicks cuya clasificación también fue mucho mejor pese a tener una media de bloqueos menor que los Bulls.

Finalmente, y como parte extra del análisis, se ha realizado un análisis multivariante mediante un mapa de calor para analizar la correlación entre las estadísticas ofensivas y defensivas de los Bulls vs el resto de equipos de la liga.




Se observan algunas diferencias:

1. Los Bulls muestran una correlación mayor entre los robos (STL) y los bloqueos (BLK), comparado con el resto de la liga (0.8 vs 0.33), esto muestra la sinergia defensiva del equipo.
2. La relación entre los puntos (PTS) y los tiros de 2 (2P) es ligeramente mayor (0.95) comparada con el resto de la liga (0.92), indicando que los Bulls se centran más en los tiros de 2.
3. Comparado con el resto de la liga, la relación entre los puntos y los tiros de 3 (3P) es similar a los Bulls (0.74), pero en conjunto los Bulls se basan menos en los tiros de 3.
4. La relación entre los rebotes defensivos (DRB) y los tiros de 2 (0.81) es ligeramente superior comparada con el resto de la liga (0.79) lo cual implica que las jugadas que parten de un rebote defensivo suelen acabar con un tiro de 2 efectivo.

5- Conclusiones

Tras haber hecho el EDA sobre el dataset se ha confirmado que ninguna de las 4 hipótesis planteadas inicialmente se cumple por completo en base a los resultados obtenidos.

Por un lado, es cierto que en general los jugadores de los Chicago Bulls tienen menos puntos por partido de media, pero esto solo se confirma al compararlos con el equipo campeón de la liga. Por otro lado, los jugadores de los Bulls tampoco tienen un volumen menor de tiros de 2 y de 3 significativamente bajo respecto al



de los equipos de la conferencia Este. Esto indica que ninguna de las dos hipótesis se ha cumplido totalmente y no se puede establecer una relación entre estas dos hipótesis y el mal rendimiento en la pasada temporada.

Respecto a las hipótesis defensivas, el volumen de rebotes ofensivos del equipo es superior al de otros equipos cuya clasificación fue más alta, por lo que tampoco se cumple esta hipótesis. Y lo mismo sucede con la relación entre los robos y bloqueos ya que en algunos casos los resultados de estas dos variables son superiores a las de equipos mejor clasificados.

En base a lo anterior se puede concluir que es necesario seguir investigando el porqué del mal rendimiento de los Chicago Bulls la pasada temporada. El equipo acabó la temporada en novena posición, muy alejado de los puestos esperados, pero a raíz de este estudio se concluye que influyeron otros factores más allá del mal o buen desempeño de los jugadores.

Por otro lado, se han excluido algunas columnas en este análisis, por lo que sería interesante retomar esas variables e incluirlas también en un análisis futuro.