

Bootcamp Data Science

EDA

**El rendimiento de los jugadores de la NBA**

**Autora:** Itziar Cid Garcia

**Fecha:** Diciembre 2023

**1.Introducción**

Este análisis se centra en el estudio de los registros de jugadores de la NBA de la temporada regular 2022-2023, a partir de características demográficas, por equipos y estadísticas en el juego.

Como principal objetivo se desea conocer que variables pueden estar relacionadas con el rendimiento de los jugadores y como afectan en el mismo.

**2.Descripción general de los datos**

El conjunto de datos contiene 679 entradas (jugadores) y diversas variables como edad, juegos jugados, minutos por juego, tiros de campo, porcentajes, rebotes, asistencias, robos, bloqueos, pérdidas, faltas personales y puntos por juego.

Nuestro DataFrame tiene 679 observaciones y 30 columnas. De las 30 variables que contiene, 3 son cualitativas, Player, Pos y Tm y el resto son cuantitativas. No tenemos valores duplicados.

Valores faltantes: No hay valores faltantes en ninguna de las columnas, lo que indica que el conjunto de datos está completo.

**3.Explicación de las variables**

-Rk: Rango. Posición relativa de un jugador en comparación con otros jugadores.

-Player: Nombre del jugador.

-Pos: Posición.

-Age: Edad del jugador.

-Tm: Equipo.

-G: Partidos disputados.

-GS: Partidos iniciados. Titularidades.

-MP: Minutos jugados por partido.

-FG: Canastas por partido.

-FGA: Intentos de canastas por partido.

-FG%: Porcentaje de canastas.

-3P: Triples por partido.

-3PA: Intentos de triples por partido.

-3P%: Porcentaje de triples.

-2P: Canastas de de 2 puntos por partido.

-2PA: Intentos de canastas de 2 puntos por partido.

-2P%: Porcentaje de canastas de 2 puntos

-eFG%: Porcentaje Efectivo de Tiros de Campo. Se calcula de la siguiente manera:

e**Fg% = (Canastas + 0.5 \* Triples Anotados) / Intentos de Canastas**

-FT: Tiros libres por partido.

-FTA: Intentos de tiros libres por partido.

-FT%: Porcentaje de tiros libres.

-ORB: Rebotes ofensivos por partido.

-DRB: Rebotes defensivos por partido.

-TRB: Rebotes totales por partido.

-AST: Asistencias por partido.

-STL: Robos por partido.

-BLK: Tapones por partido.

-TOV: Pérdidas de balón por partido.

-PF: Faltas personales por partido

-PTS: Puntos por partido.

**4.Metodología**

Este análisis propone introducir una metodología para evaluar el rendimiento de los jugadores, a través de la creación de una variable integral que considere múltiples aspectos del juego. Los registros de la NBA en este DataFrame nos muestra una serie de covariables que afectan directamente el rendimiento en la cancha.

**Es por ello que consideraremos al rendimiento (EFF) como una variable nueva creada a partir de la siguiente relación de covariables:**

**EFF=PTS+TRB+AST+STL+BLK−(FGA−FG)−(FTA−FT)−TOV**

Se considerará el rendimiento a través de la suma de puntos anotados, rebotes, asistencias, robos, tapones, y la resta de los intentos de tiro fallidos, los tiros libres fallidos y las pérdidas de balón. Es una medida integral que intenta capturar la contribución general de un jugador al juego.

Basándonos en esto, postulamos una serie de afirmaciones a analizar.

·¿La edad (Age) del jugador es un factor importante para el rendimiento?

·¿La posición (Pos) en la que se juega afecta el rendimiento?

·El rendimiento está relacionado con variables que miden la efectividad en los tiros

2p%

3P%

eFG%

FG%

FT%

(Alta eficiencia de tiro suele ser indicativa de un buen rendimiento tanto a nivel individual como de equipo)

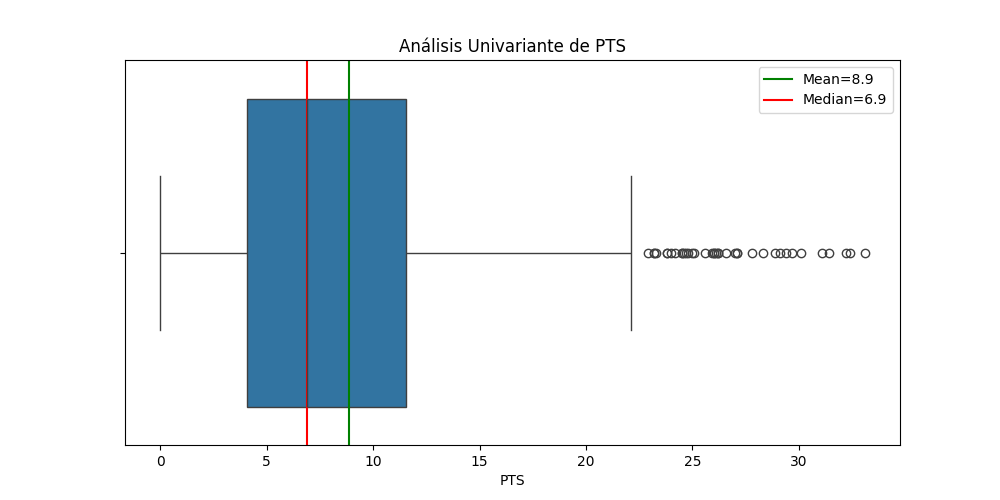
·Los minutos jugados (MP) afectan el rendimiento del jugador.

·Los jugadores que tienen mayor rendimiento son titulares en todos los partidos (GS).

·El número de partidos disputados (G) está relacionado con el rendimiento de los jugadores.

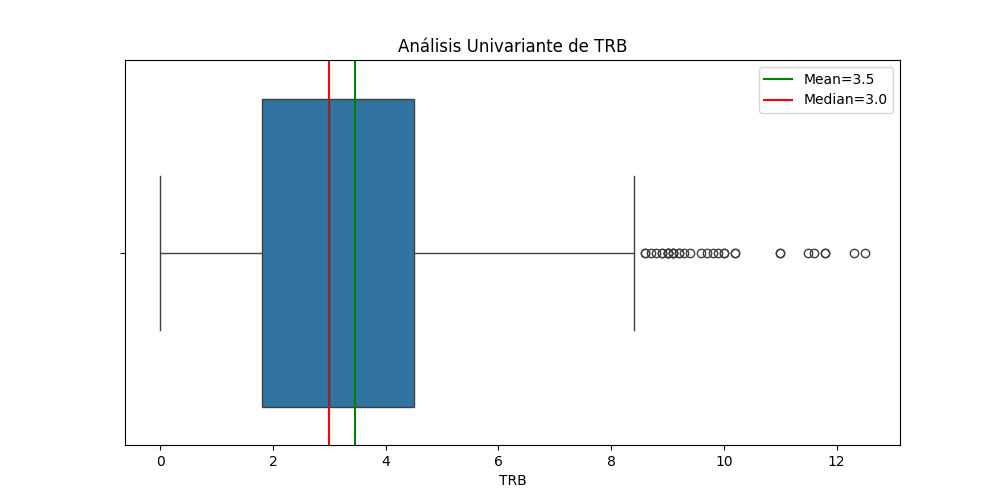
**5.Análisis**

**5.1.Análisis univariante**



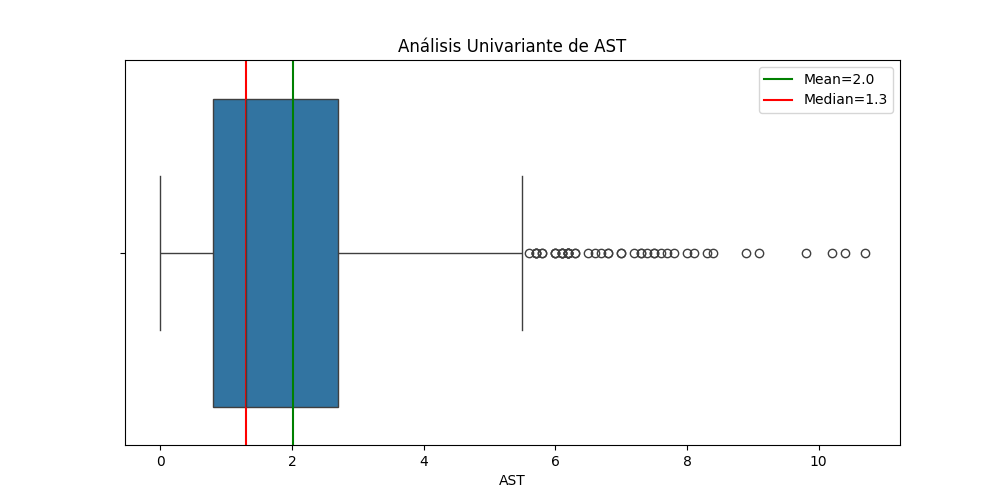
La gráfica sugiere una distribución sesgada hacia la derecha. Esto significa que hay jugadores que anotan muchos más puntos que la mayoría, posiblemente estrellas del equipo, la distribución de los puntos tiene una cantidad significativa de datos atípicos, indicados por los puntos individuales a la derecha del gráfico. Estos datos atípicos representan jugadores que anotan muchos más puntos por partido que sus compañeros.

El hecho de que haya jugadores con puntajes muy por encima de la media puede indicar una alta efectividad en los tiros. Si estos jugadores con altos PTS también tienen altos porcentajes en estas estadísticas de tiro, apoyaría la hipótesis de que la efectividad en los tiros está correlacionada con un mayor rendimiento.



En la gráfica se observa un sesgo leve hacia la derecha en la distribución de TRB. Esto indica que, mientras la mayoría de los jugadores tiene un número de rebotes cercano a la mediana, hay algunos jugadores que obtienen significativamente más rebotes por partido. Por lo tanto, un mayor número de TRB podría correlacionarse indirectamente con un mejor rendimiento del equipo, lo que afirma el hecho de incluirse para medir el rendimiento.

Se observan varios datos atípicos en el lado derecho del Boxplot. Estos representan jugadores que logran obtener un número de rebotes por partido mucho mayor que el promedio de sus compañeros.



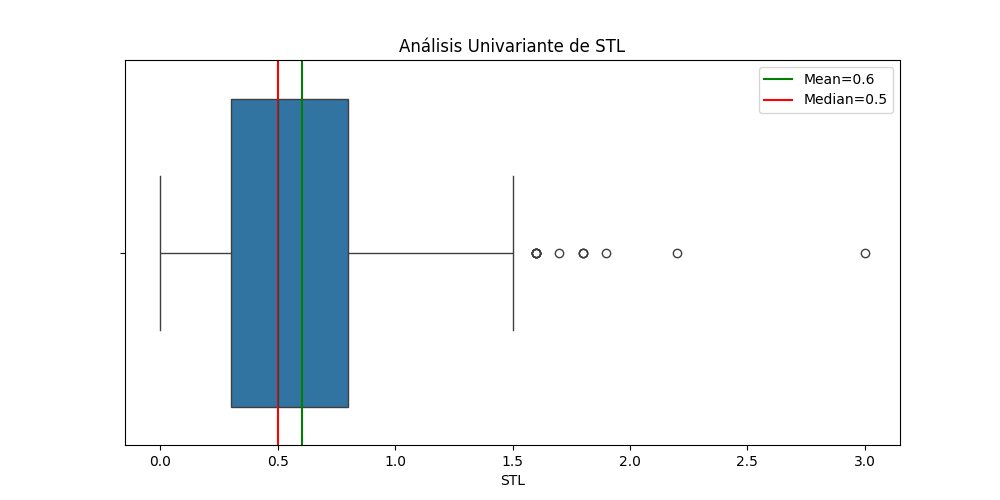
Se observa un sesgo positivo en la distribución de las asistencias por partido.

Este sesgo sugiere que hay unos pocos jugadores que están realizando muchas más asistencias que la mayoría, lo que eleva la media de asistencias.

Hay una cantidad notable de datos atípicos en el lado derecho del Boxplot. Estos representan jugadores que proporcionan muchas más asistencias que sus compañeros de equipo.

Este gráfico podría apoyar la hipótesis de que la posición influye en el rendimiento. Se espera que las posiciones que tradicionalmente manejan más el balón, como el base, tengan un número más alto de asistencias.

De los datos atípicos detectados, se resalta que en las asistencias la posición que más las realiza es la (PG) Base, quien dirige el ataque del equipo, manejando el balón y asegurándose de que llegue al jugador correcto en el momento adecuado, lo cual tiene coherencia con los resultados.

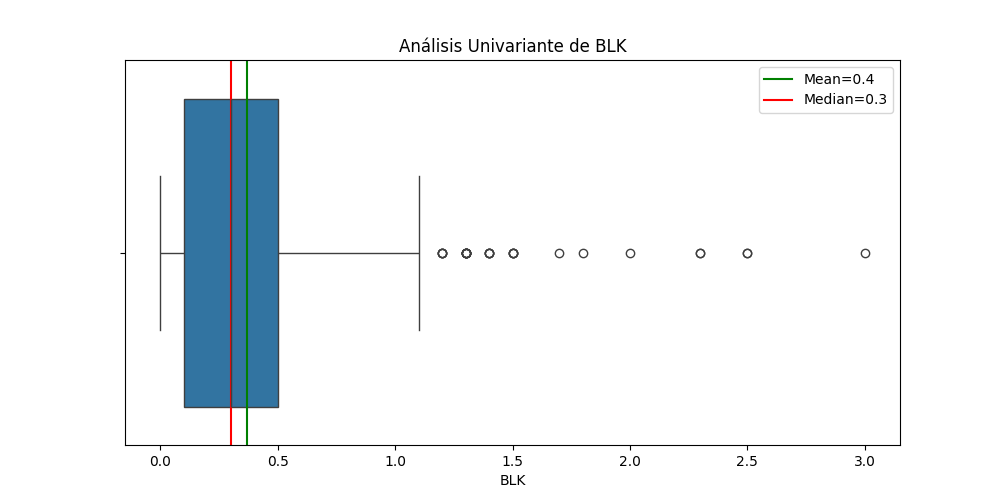


En la gráfica de robos por partidos la distribución no está fuertemente sesgada y la mayoría de los jugadores tienen un número de robos por partido cerca de la mediana.

Este gráfico puede respaldar la idea de que ciertas posiciones, como la del base, que suelen estar más involucrados en la defensa del perímetro, podrían tener más robos.

Se observan algunos datos atípicos a la derecha del Boxplot. Estos valores representan jugadores que consiguen más robos por partido de lo normal, lo que podría indicar jugadores con habilidades defensivas excepcionales, se procede a identificar casos atípicos.

En la identificación no se observa un número significativo de datos atípicos, lo que da respuesta a que pertenece a jugadores con habilidades excepcionales en la defensa.

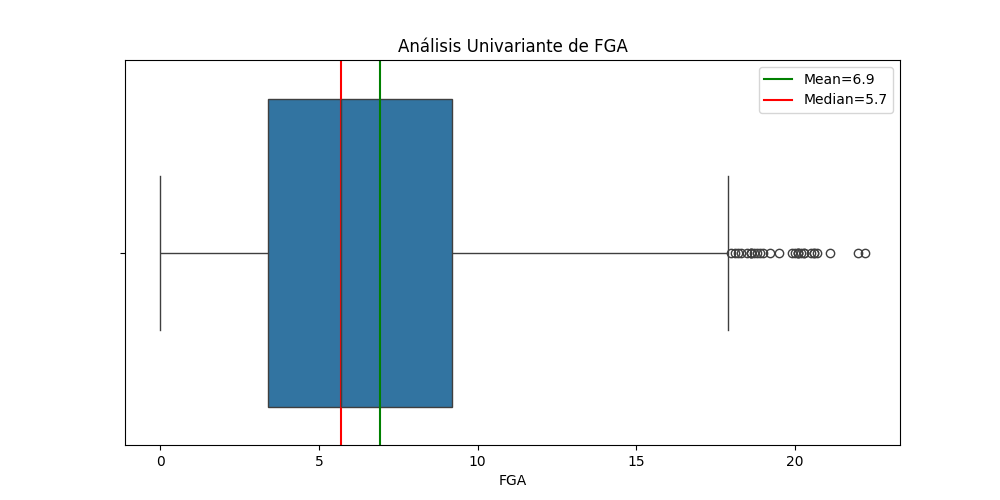


La gráfica muestra una proximidad de la media y la mediana sugiere que la distribución de los tapones por partido es relativamente simétrica con una ligera cola hacia la derecha debido a algunos valores más altos. Que indican que hay jugadores que consiguen significativamente más tapones que el promedio.

Este gráfico podría apoyar la hipótesis de que la posición afecta al rendimiento. Los jugadores que son responsables de proteger el aro (generalmente pívots y ala-pívots) son más propensos a tener un mayor número de BLK.

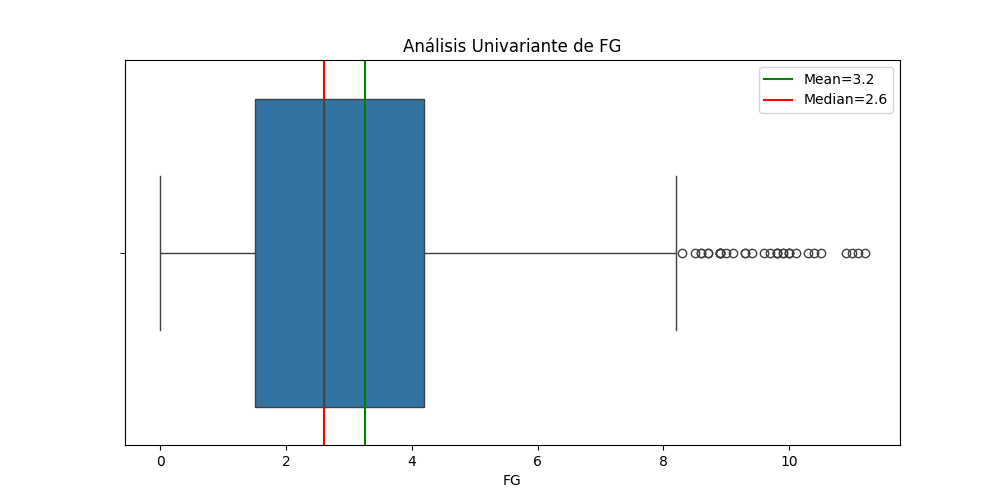
Los datos atípicos en esta gráfica son importantes para la evaluación del rendimiento defensivo, procedemos a la detención de los mismos.

En la identificación de los datos atípicos se observa una coherencia, en las posiciones que generan más tapones en este caso (C) el Pivot que se especializa en bloquear tiros, capturar rebotes y anotar cerca del aro. Lo que muestra que estos datos no son errores en los datos, sino rendimientos excepcionales en esa posición.



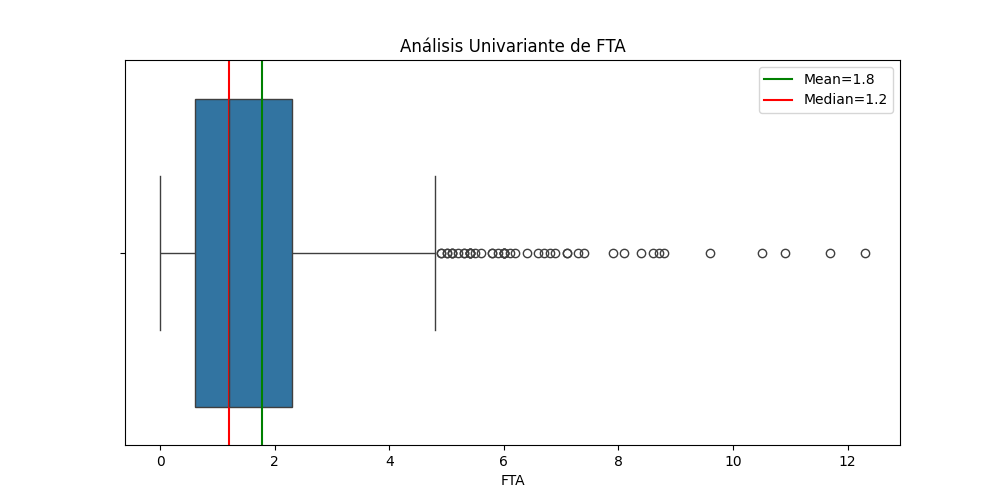
Se observa que la media (Mean = 6.9) es mayor que la mediana (Median = 5.7), lo cual indica un sesgo hacia la derecha en la distribución de FGA. Esto sugiere que hay jugadores que están tomando muchos más tiros que la mayoría y son datos atípicos.

Esta gráfica contribuye al análisis de la relación entre FGA y los porcentajes de tiro para ver si los jugadores que disparan más son también aquellos que anotan con mayor eficiencia.



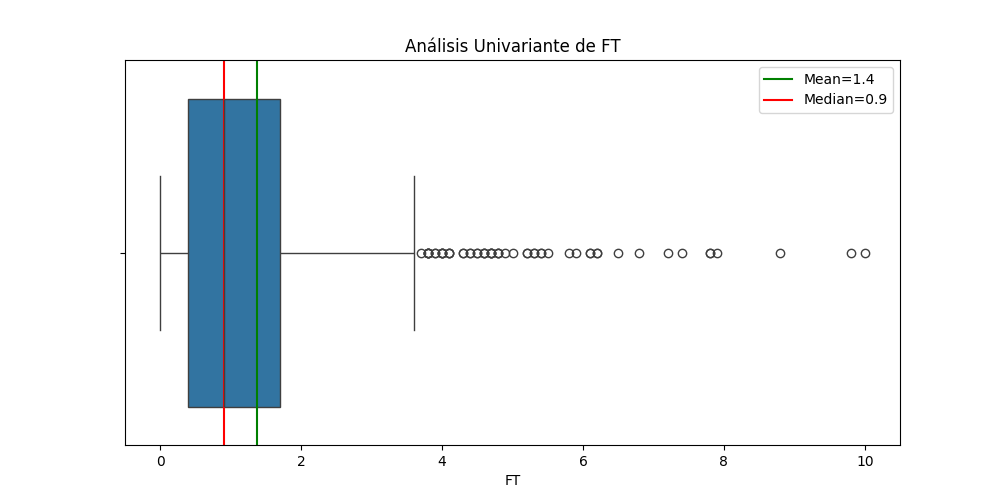
La gráfica se muestra un sesgo hacia la derecha en la distribución de FG. Esto sugiere que, aunque la mayoría de los jugadores anotan cerca de 2.6 canastas por partido, hay algunos que anotan significativamente más, lo que eleva la media.

La cantidad de canastas anotadas (FG) está directamente relacionada con la efectividad en los tiros. Los jugadores que tienen un alto FG y también mantienen buenos porcentajes de tiro probablemente son jugadores de alto rendimiento. Esto apoyaría la hipótesis de que una alta efectividad de tiro está asociada con un buen rendimiento.

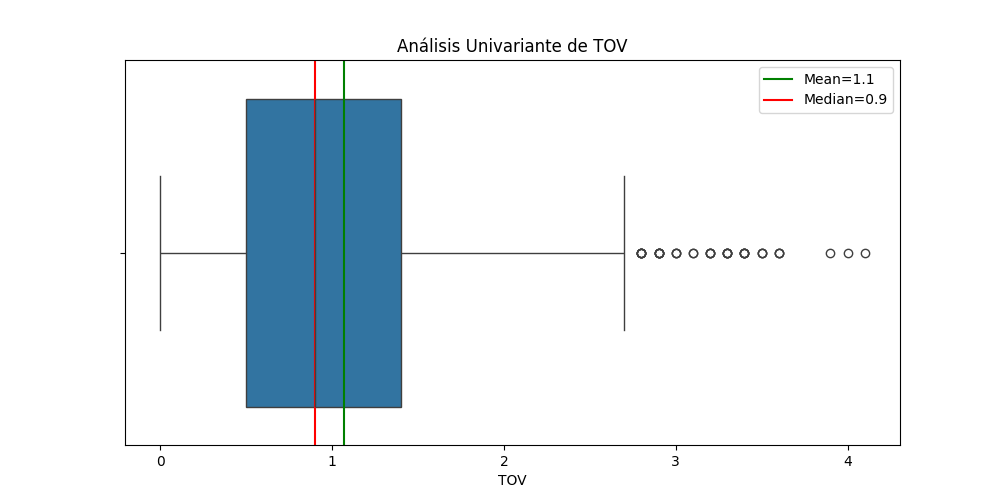


En el gráfico se observa que la media (Mean = 1.8) es mayor que la mediana (Median = 1.2), lo que sugiere un sesgo hacia la derecha en la distribución de FTA. Esto puede indicar que hay algunos jugadores que van a la línea de tiros libres con más frecuencia que la mayoría.

Anexo la cantidad de FTA está relacionada con la efectividad de los tiros libres (FT%). Los jugadores con un alto número de FTA que también tienen un alto porcentaje de FT% pueden ser vistos como jugadores efectivos, lo que contribuiría al rendimiento.



Este tipo de distribución sugiere que hay algunos jugadores que anotan más tiros libres que la mayoría, lo que eleva la media. Los FT son una medida directa de efectividad en una faceta del juego, y un alto número de FT anotados sugiere una buena eficacia en esta área. Los jugadores que tienen un alto número de FT y también un alto FT% probablemente contribuyen positivamente al rendimiento de su equipo.



La media (Mean = 1.1) es ligeramente superior a la mediana (Median = 0.9), lo que sugiere un sesgo leve hacia la derecha en la distribución de TOV. Esto puede indicar que, aunque la mayoría de los jugadores tienen un número de pérdidas cercano a la mediana, hay algunos jugadores que cometen más pérdidas que otros.

Aunque los TOV no están directamente relacionados con la efectividad de los tiros, sí afectan la eficiencia ofensiva de un jugador. Los jugadores que mantienen un bajo número de TOV mientras tienen una alta efectividad en los tiros podrían ser considerados más eficientes y, por tanto, tener un mejor rendimiento.

De lo realizado anteriormente, hemos identificado la distribución de cada variable que compone la métrica del rendimiento (EFF).

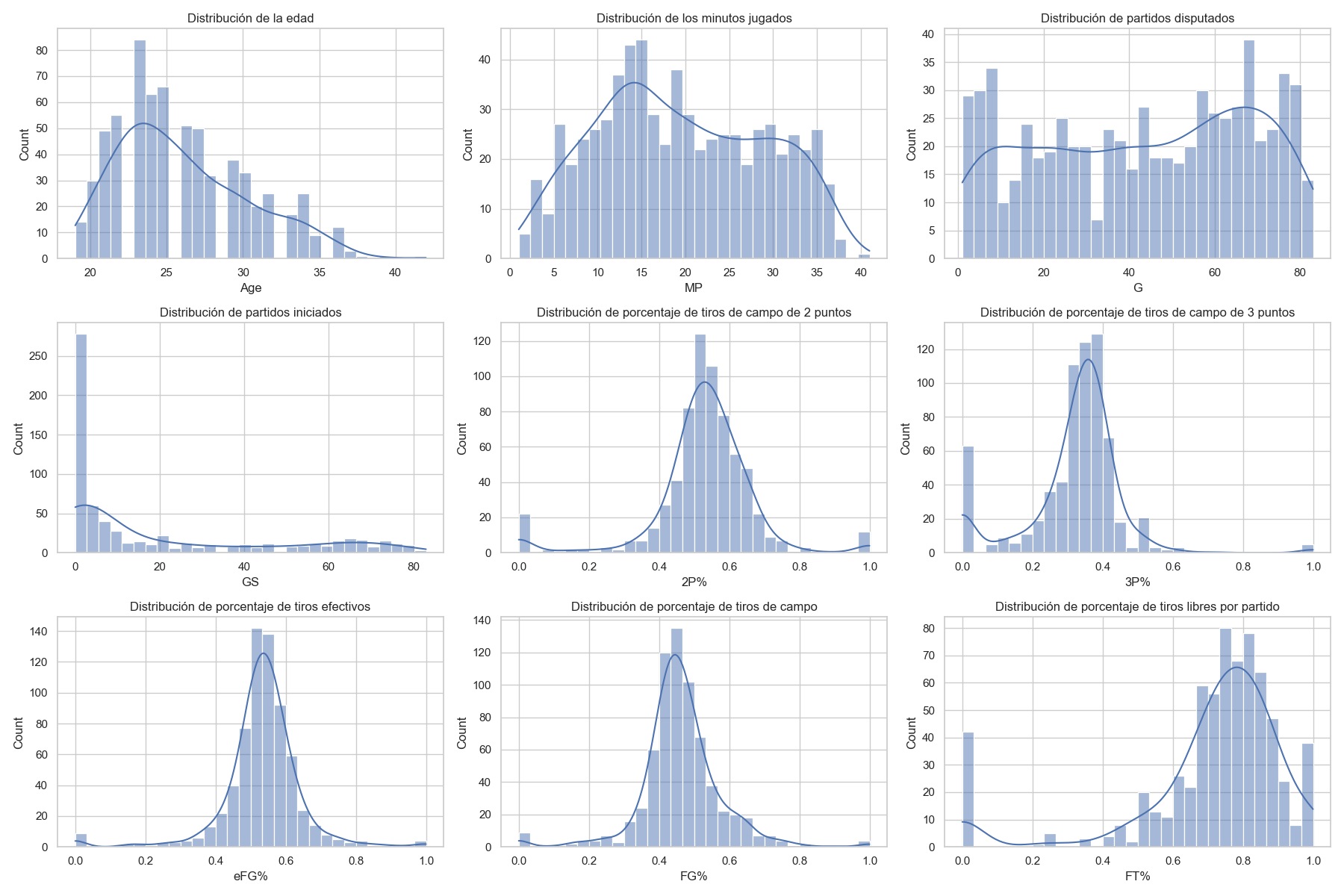
Y se observaron los siguientes resultados generales:

-Hay sesgo a la derecha en todas las variables que podrían afectar la valoración general del rendimiento. Una variable que tiene muchos valores extremos (outliers), podría influir desproporcionadamente en la puntuación total de EFF.

-A partir de la forma de las variables resulta importante más adelante revisar la distribución normal de cada variable puede observar si se cumplen estos supuestos o si se necesitan transformaciones o métodos estadísticos alternativos.

-Nos ha permitido revisar si es necesario la corrección de Datos Atípicos, los cuales podrían representar errores de medición o situaciones excepcionales que no deberían influir indebidamente en la métrica de eficiencia. Se han considerado como jugadores excepcionales que contribuyen al juego a partir del rol y la posición que tienen.

A continuación, se proporciona una serie de histogramas que representan la distribución del resto de variables estadísticas claves en las hipótesis planteadas se observara Age (Edad), MP (Minutos jugados), G (Partidos jugados), GS (Partidos iniciados), 2p%, 3P%, eFG% (Alta eficiencia de tiro suele ser indicativa de un buen rendimiento tanto a nivel individual como de equipo), FG%, Ft%.

-

Distribución de la edad (Age):

La distribución parece ser ligeramente sesgada hacia la derecha, con una concentración de jugadores en el rango de 20 a 30 años, lo cual es típico en atletas profesionales.

Contribución a la hipótesis: Esto puede indicar que los jugadores más jóvenes son más prevalentes en la liga, pero no necesariamente que la edad es un factor importante en el rendimiento sin más contexto.

-Distribución de los minutos jugados (MP):

La distribución de MP parece ser aproximadamente normal, lo que indica una distribución equitativa de los minutos jugados entre los jugadores.

Contribución a la hipótesis: Un análisis más profundo podría revelar si los jugadores que juegan más minutos tienen un mayor rendimiento.

-Distribución de partidos jugados (G):

Los jugadores parecen participar en una cantidad variada de partidos, con algunas concentraciones alrededor de valores específicos, posiblemente relacionados con la duración de la temporada o lesiones.

Contribución a la hipótesis: Esto podría apoyar la idea de que la experiencia de juego (número de partidos disputados) influye en el rendimiento.

-Distribución de partidos iniciados (GS):

Hay una gran cantidad de jugadores con pocos o ningún partido iniciado, con un descenso progresivo en el número de jugadores a medida que aumenta el número de partidos iniciados.

Contribución a la hipótesis: Esto puede sugerir que solo algunos jugadores son constantemente titulares, lo cual podría estar relacionado con un mayor rendimiento.

-Distribución de porcentaje de tiros de campo de 2 puntos (2P%):

La distribución del 2P% parece tener una forma de campana, lo que indica una distribución normal. La mayoría de los jugadores tienen un porcentaje alrededor de un valor medio, con pocos extremos en ambos lados.

Contribución a la hipótesis: Esto apoya la idea de que la efectividad en los tiros es un componente crítico del rendimiento.

-Distribución de porcentaje de triples (3P%):

Similar al 2P%, la distribución de 3P% parece aproximadamente normal, aunque con una cola más larga hacia los valores más bajos, lo que sugiere que algunos jugadores tienen mucha menor efectividad en triples.

Contribución a la hipótesis: Indica que no todos los jugadores tienen la misma capacidad para anotar de triples, lo cual es un factor importante en el rendimiento.

-Distribución de porcentaje de tiros efectivos (eFG%):

La distribución de eFG% también sigue una distribución normal, con la mayoría de los jugadores concentrados alrededor del promedio.

Contribución a la hipótesis: Un eFG% alto es una buena indicación de rendimiento eficiente, apoyando la hipótesis de que la eficacia en la anotación es clave para el rendimiento general.

-Distribución de porcentaje de tiros de campo (FG%):

El FG% muestra una distribución normal, similar al eFG%, con la mayoría de los jugadores agrupados cerca del promedio.

Contribución a la hipótesis: Al igual que con eFG%, un alto FG% es indicativo de un rendimiento eficiente en tiros de campo.

-Distribución de porcentaje de tiros libres por partido (FT%):

La distribución de FT% parece ser aproximadamente normal, con una ligera inclinación hacia porcentajes más altos.

Contribución a la hipótesis: Esto sugiere que la mayoría de los jugadores son eficientes en tiros libres, lo cual contribuye al rendimiento general, especialmente en situaciones de juego cerrado.

**Es importante considerar:**

- Distribución de Puntos por Juego (PTS):

Contribuye a la hipótesis de que los puntos efectivos (puntos por juego) son un indicador de rendimiento.

- Distribución de Asistencias (AST):

Apoya la hipótesis de que las asistencias por partido son un indicador de rendimiento, ya que implican una participación activa en la creación de oportunidades de anotación para el equipo.

- Distribución de Rebotes Totales (TRB):

Relaciona con la idea de que los rebotes totales por partido son un indicador de rendimiento, particularmente en la contribución de un jugador a la posesión del balón y la defensa.

-Distribución de Porcentaje de Tiro Efectivo (eFG%):

Directamente relevante para la hipótesis que relaciona un alto porcentaje de tiro efectivo con buen rendimiento, ya que considera el valor adicional de los triples.

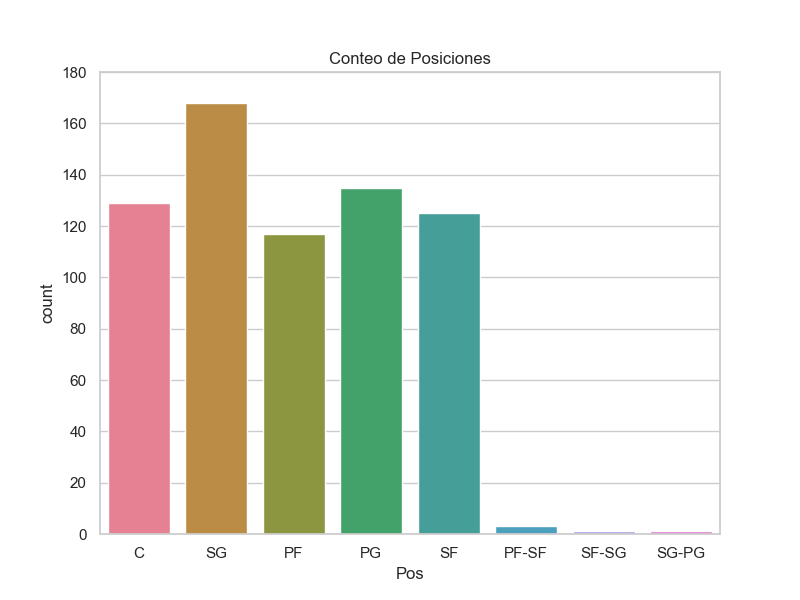
- Distribución de Minutos por Juego (MP):

Se alinea con la hipótesis de que los minutos jugados afectan el rendimiento de los jugadores, posiblemente reflejando la confianza del entrenador y la resistencia del jugador.

- Distribución de Partidos Disputados (G):

Contribuye a la hipótesis de que el número de juegos jugados está relacionado con el rendimiento, potencialmente indicando la consistencia y la disponibilidad del jugador.

**A** **continuación nos centramos en conocer el comportamiento de la variable *posición* (Pos)**



**Conteo de valores:**

Posición

SG 168

PG 135

C 129

SF 125

PF 117

PF-SF 3

SF-SG 1

SG-PG 1

**Porcentaje de valores:**

Posición

SG 24.74%

PG 19.88%

C 19.00%

SF 18.41%

PF 17.23%

PF-SF 0.44%

SF-SG 0.15%

SG-PG 0.15%

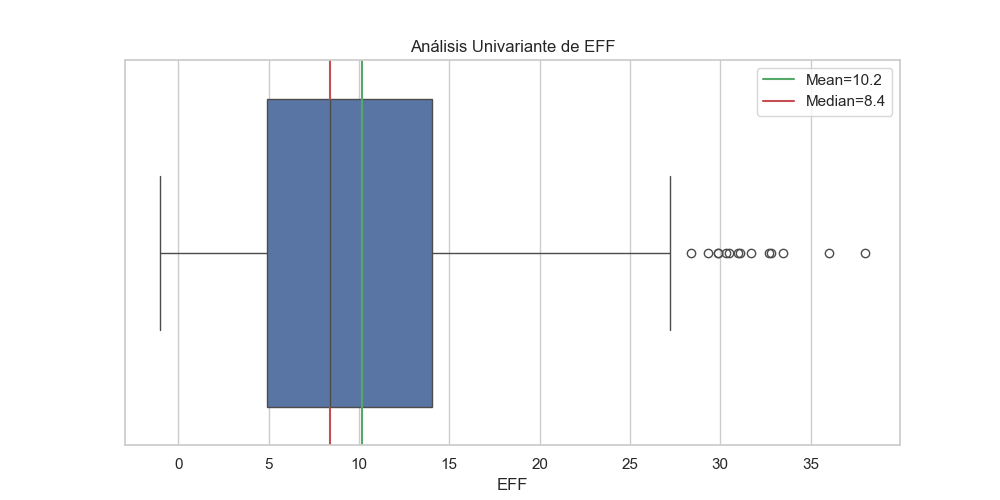
Las posiciones tradicionales C, SG, PF, PG, SF, (Pívot, Escolta, Ala-Pívot, Base, Alero) respectivamente, se puede observar que todas las posiciones tradicionales tienen representaciones similares en términos de conteo, aunque con ligeras variaciones en la cantidad.

Las barras etiquetadas como PF-SF, SF-SG, y SG-PG representan posiciones híbridas o combinadas. Esto refleja una tendencia moderna en el baloncesto donde los jugadores son más versátiles y pueden ocupar más de una posición en la cancha.

La distribución de las posiciones puede proporcionar una base para analizar cómo la posición específica o las posiciones combinadas afectan el rendimiento de los jugadores.

Este gráfico es una herramienta útil para comenzar a explorar el rendimiento y cómo las diferentes posiciones están representadas.

**Finalmente, se analiza el comportamiento de la variable creada de rendimiento global (EFF)**



De la gráfica se resalta que dado que EFF es una combinación de estadísticas positivas (puntos, rebotes, asistencias, etc.) y negativas (tiros y tiros libres fallidos, pérdidas de balón), una distribución con sesgo hacia la derecha.

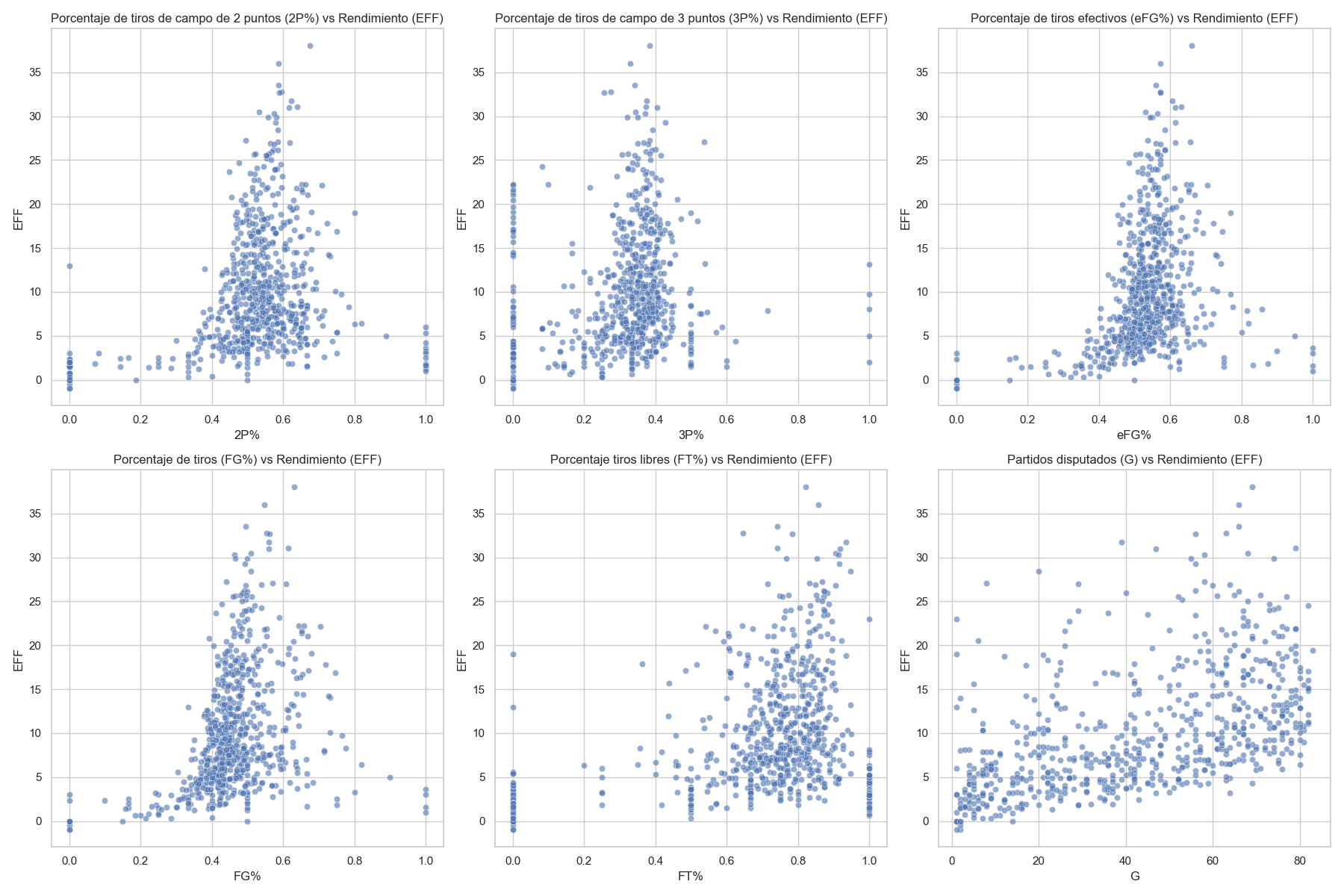
La presencia de datos atípicos puede sugerir que aquellos jugadores que tienen más tiempo en la cancha o que son titulares frecuentes tienen la oportunidad de acumular estadísticas que contribuyen positivamente a su EFF.

Esta gráfica es un instrumento valioso para evaluar el rendimiento global de los jugadores y para identificar tanto el rendimiento típico como las excepciones destacables dentro de un conjunto de datos.

**5.****2.Análisis bivariante**

A continuación se observa la relación entre dos variables, permitiendo descubrir correlaciones entre variables que afecten el rendimiento y tendencia.

Se inicia observando la correlación entre algunas variables que miden la efectividad y el rendimiento.



Podemos interpretar los siguientes resultados:

-Porcentaje de tiros de campo de 2 puntos (2P%) vs Rendimiento (EFF):

Parece haber una tendencia positiva donde un mayor porcentaje de tiros de 2 puntos se asocia con un mayor EFF.

Esto sugiere que la eficiencia en tiros cercanos podría ser un indicador significativo del rendimiento general de un jugador.

-Porcentaje de tiros de campo de 3 puntos (3P%) vs Rendimiento (EFF):

La dispersión muestra una concentración de datos que indica una relación positiva, aunque menos pronunciada que para los tiros de 2 puntos, entre el 3P% y el EFF.

Esto podría indicar que mientras los tiros de 3 puntos son valiosos, no todos los jugadores con alto 3P% necesariamente tienen un alto rendimiento, posiblemente debido a la naturaleza de riesgo de estos tiros.

-Porcentaje de tiros efectivos (eFG%) vs Rendimiento (EFF):

La relación parece ser fuertemente positiva, lo que indica que el eFG%, que ajusta el porcentaje de tiros de campo para reflejar el valor extra de los tiros de 3 puntos, es un buen predictor del rendimiento.

Esto refuerza la idea de que la eficiencia general en tiros de campo es un componente clave del rendimiento de un jugador.

-Porcentaje de tiros de campo (FG%) vs Rendimiento (EFF):

El FG% también muestra una asociación positiva con el EFF, lo que respalda la importancia de la eficiencia de tiro en la valoración general del rendimiento de un jugador.

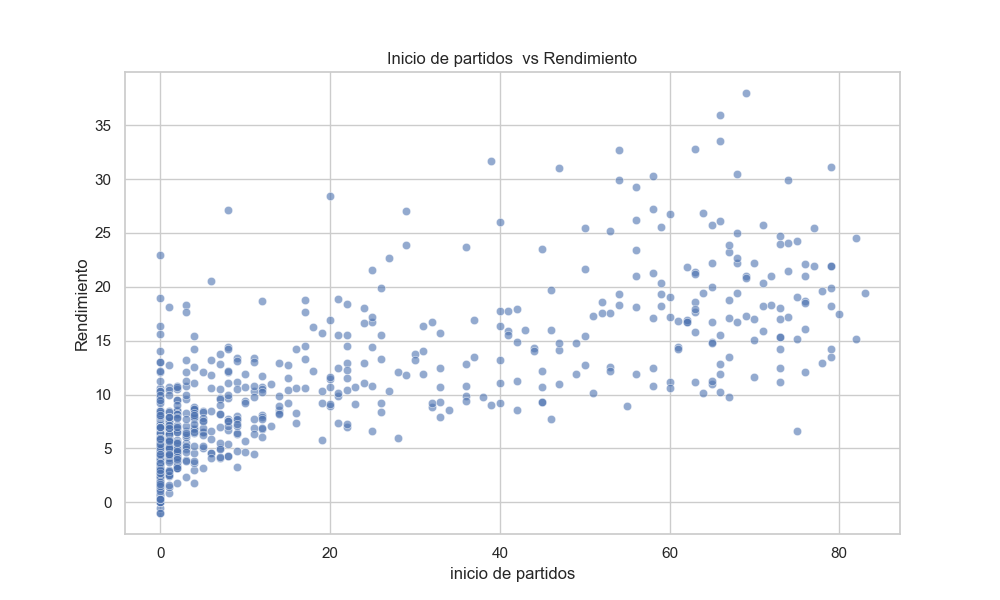
-Porcentaje de tiros libres (FT%) vs Rendimiento (EFF):

Los datos sugieren una tendencia positiva, aunque con bastante dispersión, lo que podría indicar que aunque los tiros libres son importantes, hay otros factores que también influyen significativamente en el rendimiento.

-Partidos jugados (G) vs Rendimiento (EFF):

La relación entre G y EFF no muestra una tendencia clara, lo que sugiere que simplemente jugar más partidos no garantiza un mayor rendimiento. Esto podría deberse a que el rendimiento depende de la calidad de las contribuciones en el juego, no solo de la cantidad.

Anexo a lo anterior revisamos en un gráfico a parte la relación entre el rendimiento y el inicio de partidos.



De la gráfica resaltamos:

-A primera vista, parece haber una correlación positiva leve entre el número de partidos iniciados y el rendimiento. Los jugadores que han iniciado más partidos tienden a tener un rendimiento ligeramente más alto, pero la tendencia no es fuertemente pronunciada.

-Los jugadores que han iniciado muchos partidos (a la derecha del gráfico) muestran una amplia gama de puntuaciones de rendimiento, lo que sugiere que ser titular frecuente no garantiza automáticamente un alto rendimiento.

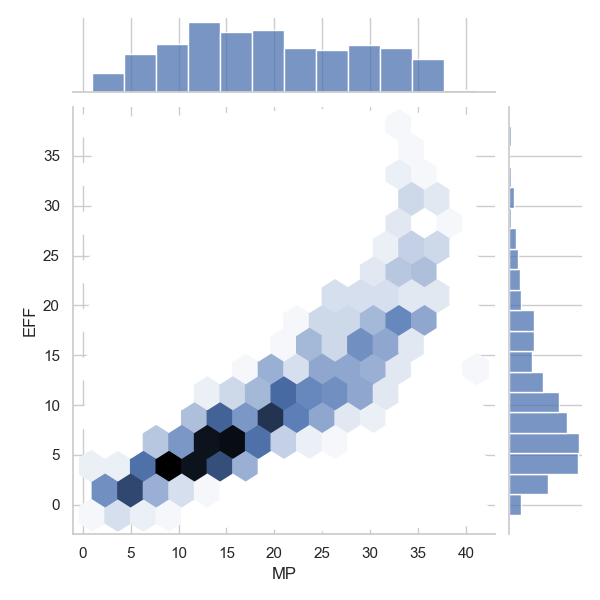
-Hay una concentración de datos en el extremo inferior del eje X, lo que indica que muchos jugadores han iniciado pocos o ningún partido. Estos jugadores tienden a tener un rendimiento más bajo, pero aún hay una variabilidad considerable.

-Datos Atípicos: Hay algunos datos atípicos, en particular jugadores con rendimientos muy altos que no han iniciado muchos partidos. Estos casos podrían ser jugadores especialistas o que juegan minutos de alta calidad pero no necesariamente comienzan el juego.

- El gráfico podría apoyar la hipótesis de que los jugadores que son titulares más frecuentemente tienen un mayor rendimiento, pero también sugiere que hay más factores en juego, ya que no hay una relación directa y fuerte.

Aunque iniciar partidos puede estar relacionado con la experiencia, el gráfico indica que la experiencia por sí sola (en términos de partidos iniciados) no es un predictor infalible del rendimiento.

**Ahora nos interesamos en conocer si existe relación del tiempo jugado y el rendimiento, como clave para el análisis**



La gráfica nos muestra una correlación positiva entre los minutos jugados y la eficiencia, donde los jugadores con más minutos tienden a tener una mayor eficiencia. Esto se muestra por la concentración de hexágonos más oscuros que se extienden desde la parte inferior izquierda hacia la parte superior derecha del gráfico.

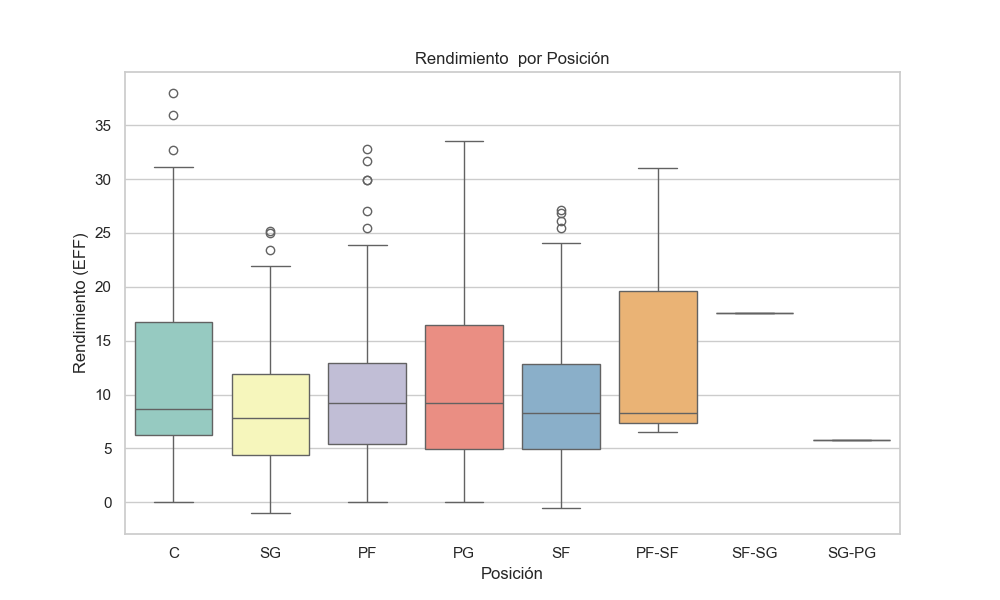
Los histogramas marginales sugieren que la mayoría de los jugadores juegan entre 10 y 20 minutos por partido y que la mayoría de los valores de eficiencia están entre 5 y 15.

Hay una notable concentración de jugadores en un rango medio de minutos jugados con eficiencias moderadas.

El gráfico apoya la hipótesis de que los jugadores que tienen más minutos en el juego tienden a tener un mayor rendimiento, medido por el EFF. Esto podría sugerir que los jugadores más talentosos o fundamentales para su equipo reciben más tiempo de juego y contribuyen más significativamente a las métricas de rendimiento.

El patrón observado en el gráfico puede indicar que no solo jugar más tiempo está asociado con un mayor rendimiento, sino que también puede ser un reflejo de una mayor oportunidad para contribuir en varias facetas del juego que componen la métrica EFF.

**Teniendo en cuenta la importancia que nos mostró la variable posición en el análisis univariante, observamos como se distribuye en rendimiento en las diferentes posiciones**



El grafico nos muestra:

-Las posiciones tradicionales (C, SG, PF, PG, SF) tienen rangos de rendimiento variados, con algunas posiciones mostrando una mayor dispersión que otras.

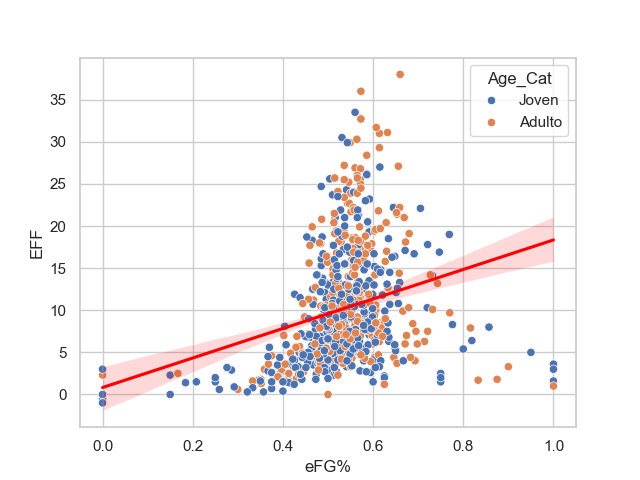
-Las posiciones híbridas (PF-SF, SF-SG, SG-PG) tienen una variabilidad de rendimiento que podría reflejar la versatilidad de las responsabilidades en la cancha.

-Las medianas varían entre las posiciones, lo que podría sugerir que ciertas posiciones tienen un mayor impacto en el rendimiento que otras, o que hay roles específicos dentro de un equipo que facilitan un mayor rendimiento estadístico.

Por lo tanto, este gráfico podría respaldar la hipótesis de que la posición afecta al rendimiento (EFF). Podría sugerir que ciertas posiciones están asociadas con un mayor rendimiento debido a sus roles en el campo, como más oportunidades para anotar, asistir o defender.

**Finalmente nos interesamos en conocer si la edad afecta al rendimiento de los jugadores, para ello se realiza una categorización de la variable edad, teniendo en cuenta la mediana de la variable**

Y a través de un diagrama de dispersión observamos la relación de los tiros efectivos y el rendimiento, teniendo en cuenta la clasificación de la edad.



Del gráfico se observa:

-Una correlación positiva entre eFG% y EFF, indicada por la línea de regresión ascendente. Esto significa que a medida que el eFG% aumenta, generalmente también lo hace el EFF.

-Diferencias por edad, los puntos están coloreados según la categoría de edad. A primera vista, no parece haber una clara división entre las categorías de 'Joven' y 'Adulto' en términos de su relación entre eFG% y EFF. Ambas categorías están dispersas a lo largo de la línea de regresión, lo que sugiere que la eficiencia en tiros de campo es un buen indicador de rendimiento independientemente de la edad.

Banda de Confianza:

- La banda de color rojo claro alrededor de la línea de regresión representa la banda de confianza, indicando el grado de incertidumbre alrededor de la estimación de la línea de regresión. La banda parece ser bastante estrecha, lo que sugiere que hay una fuerte confianza en la relación estimada entre eFG% y

EFF.

-Concentración de Datos: La mayoría de los datos se concentran alrededor del rango medio de eFG%, con menos datos hacia los extremos más altos y bajos.

-Efectividad de Tiros y Rendimiento:

El gráfico refuerza la hipótesis de que una mayor efectividad en tiros de campo está asociada con un mayor rendimiento, como se refleja en la métrica EFF.

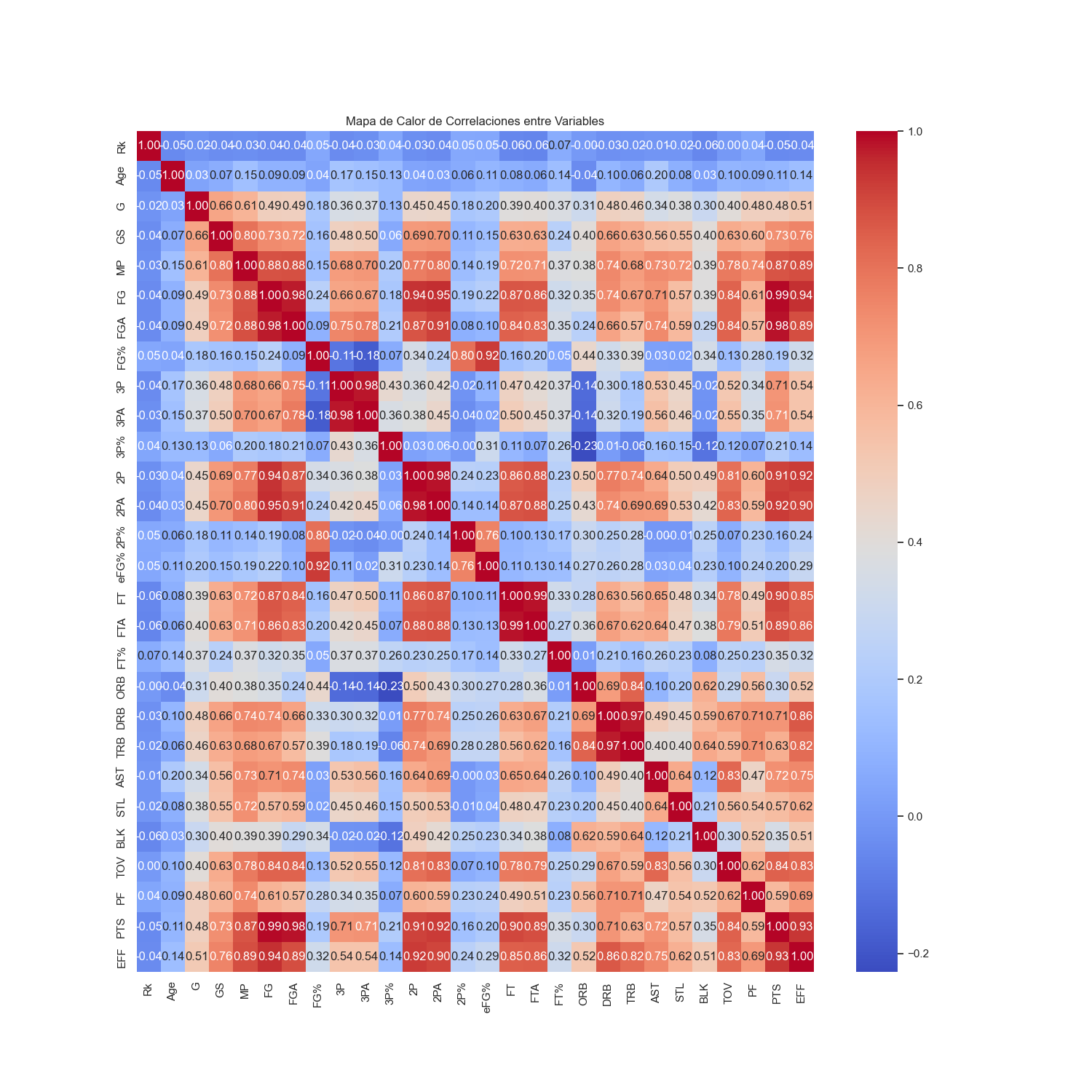
-Impacto de la Edad en el Rendimiento:

Si bien el gráfico no muestra una diferencia distintiva en el rendimiento basado en la edad, la superposición de las categorías en todo el rango de eFG% sugiere que la edad por sí sola no es un determinante claro del rendimiento en términos de la eficiencia de tiro.

**5.3.Análisis multivariante**

Para finalizar proponemos la realización de un análisis multivariante a partir de una matriz de correlación que muestre que variables numéricas presentan mejor asociación con el rendimiento (EFF)

**·Análisis de correlación**



De la matriz de correlación confirmamos de manera objetiva la relación de ciertas variables con el rendimiento, resaltamos aquellas que no han sido consideradas con su relación.

- GS y rendimiento = 0.76

- MP y rendimiento= 0.89

- 2P y rendimiento = 0.92

- 2PA y rendimiento= 0.90

- DRB y rendimiento= 0.86

En resumen, estos coeficientes altos señalan que tanto la participación (en términos de minutos jugados y titularidades) como la contribución efectiva en el juego (a través de anotaciones y rebotes defensivos) están fuertemente vinculadas al rendimiento de un jugador.

Sin embargo, solo tomaremos los minutos jugados y los tiros de dos puntos para ser evaluados en un contraste de hipótesis en el rendimiento.

**6.Contraste de hipótesis**

Teniendo en cuenta todo el análisis realizado hasta ahora sobre el rendimiento de los jugadores se propone el desarrollo de las siguientes hipótesis.

**1.Rendimiento Vs Edad categórica ----> T de Student**

**H0:** No hay evidencia significativa en rendimiento entre los diferentes grupos etarios.

**H1:** Existe una diferencia significativa entre rendimiento entre diferentes grupos etarios.

**Resultado:**

No rechazamos la hipótesis nula, es decir "No hay evidencia significativa entre el rendimiento y entre los diferentes grupos etarios".

**2.Rendimiento Vs Posición ----> ANOVA**

**H0:** No hay diferencias significativas en rendimiento entre las diferentes posiciones de los jugadores

**H1**: Existe una diferencia significativa entre rendimiento entre las diferentes posiciones de los jugadores

**Resultado:**

Rechazamos la hipótesis nula, es decir, "Existe una diferencia significativa entre rendimiento y entre las diferentes posiciones de los jugadores"

**3.Rendimiento Vs Minutos Jugados ----> Correlación de Pearson**

**H0**: No hay correlación entre los minutos jugados y el rendimiento

**H1:** Existe una diferencia significativa entre los minutos jugados y el rendimiento

**Resultado:**

Rechazamos la hipótesis nula, es decir "Existe una diferencia significativa entre los minutos jugados y el rendimiento"

**4.Rendimiento Vs tiros de 2 puntos ----> Correlación de Pearson**

**H0**: No hay correlación entre los tiros de dos puntos y el rendimiento

**H1**: Existe una diferencia significativa entre los tiros de dos puntos y el rendimiento

**Resultado:**

Rechazamos la hipótesis nula, es decir "Existe una diferencia significativa entre los tiros de dos puntos y el rendimiento"

**5.Normalidad ----> Shapiro-Wilk o D'Agostino's K-squared**

**H0:** La distribución de las columnas numéricas sigue una distribución normal.

**H1:** La distribución de las columnas numéricas NO sigue una distribución normal.

**Resultado:**

Rechazamos la hipótesis nula "No Hay evidencia significativa en la normalidad de las variables

**7.Conclusiones**

Con base en la serie de análisis realizados y los resultados de los contrastes de hipótesis, podemos concluir lo siguiente sobre el rendimiento de los jugadores en la NBA:

Primero, en cuanto a la influencia de la edad en el rendimiento, los datos no proporcionan evidencia suficiente para afirmar que existen diferencias significativas entre diferentes grupos etarios. Esto indica que la edad, como factor aislado, no es un predictor definitivo del rendimiento en la NBA.

En segundo lugar, la posición de los jugadores parece tener un impacto notable en su rendimiento. Los resultados sugieren que hay diferencias significativas en el rendimiento entre las distintas posiciones que los jugadores ocupan en la cancha.

La cantidad de minutos jugados y el número de partidos en los que un jugador comienza como titular tienen una fuerte correlación positiva con su rendimiento, lo que implica que aquellos que pasan más tiempo en la cancha y que comienzan los partidos en la alineación inicial tienden a tener un mayor impacto en el juego y, por lo tanto, un rendimiento superior.

Además, existe una relación estrecha entre la efectividad en tiros de campo, tanto de dos como de tres puntos, y el rendimiento. Los jugadores que tienen un alto porcentaje de éxito en sus tiros contribuyen significativamente a su rendimiento general, lo que resalta la importancia de la eficiencia ofensiva.

Por último, los resultados de las pruebas de normalidad indican que ninguna de las variables estadísticas sigue una distribución normal. Este hallazgo es importante porque afecta la elección de métodos estadísticos apropiados para futuros análisis y la interpretación de los resultados en el contexto del rendimiento de los jugadores.

En resumen, el rendimiento en la NBA es multifacético y está influenciado por una combinación de factores, incluyendo las posiciones en la cancha, la experiencia de juego y la eficiencia ofensiva. La edad, por otro lado, no parece desempeñar un papel significativo por sí sola. Estos hallazgos pueden proporcionar a los equipos y entrenadores información valiosa para la toma de decisiones estratégicas, la evaluación de jugadores y el desarrollo de tácticas de juego.

**8.Bibliografia**

Este conjunto de datos contiene estadísticas de jugadores de la NBA de la temporada regular 2022-2023 por partido.

Origen: <https://www.kaggle.com/datasets/vivovinco/20222023-nba-player-stats-regular/data>