

NBA Games Analytics

Presentado por : Roberto Ragaglia
Data Scientist en proceso

para **B9TTT**



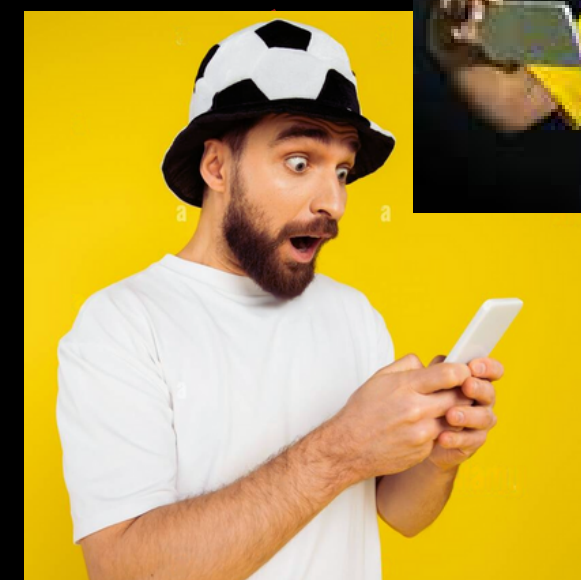
Contexto comercial

Los sistemas de apuestas existen desde hace muchos años y se han utilizado en distintas disciplinas, como carreras de caballos y deportes.

La competencia entre las plataformas de apuestas puede ser bastante intensa, ya que se trata de un mercado muy competitivo . Las empresas suelen ofrecer diferentes tipos de bonos y promociones para atraer a los clientes y fidelizarlos, además de tener una amplia variedad de apuestas disponibles para satisfacer las necesidades de los apostadores.

El cliente establece que las apuestas han disminuido significativamente durante los últimos años , debido a las distintas actividades que las personas comenzaron a desarrollar luego del paso de la pandemia.

Estas situaciones llevan a pensar nuevas estrategias a implementar intentando mejorar los beneficios/costos



Problema comercial

El equipo de análisis de datos de B9TTT busca mejorar su modelo de cuotas en el area de baloncesto y delegar otras tareas a su personal para aumentar sus ganancias y reducir las tareas repetitivas. El cliente supone que los resultados de los partidos pueden estar influenciados por factores por su participacion como local o visitante y el porcentaje de tiros de campo.

Ellos ofrecen un registro de datos en base a partidos anteriores para poder hacer un analisis estadistico de datos y poder armar una solucion automatizada

Contexto analitico

Este dataset brindado por el cliente es un registro de los resultados de los partidos de la NBA (National Basketball Association) durante las temporadas 2003-2021. Incluye información sobre la fecha , los equipos que jugaron, el resultado final ademas de estadísticas detalladas para cada equipo, incluyendo anotaciones, porcentaje de tiros de campo, porcentaje de tiros libres, etc.

Haciendo referencia a la variable "target" , debe ser una variable continua que represente el valor de la cuota de apuestas para un determinado partido. Es decir, la cuota debe ser un número real positivo que refleje la participacion en la ganancia de la apuesta recibiría si acierta su apuesta en ese partido.

Hipotesis y soluciones

El cliente propone la oportunidad que presenta los equipos participantes en cada partido cuando son locales o visitantes , y así predecir un cambio de probabilidad que determinara el valor de la cuota

¿Tiene alguna ventaja significativa el equipo local?

La hipótesis inicial es que la localía debería suponer una ventaja por diversas razones que a priori parecen lógicas: el apoyo del público, el descanso (ya que los jugadores juegan en su ciudad y no deben desplazarse), conocer la cancha y el efecto psicológico.

¿Los tiros de campo verifican lo mismo?

Otra forma de saber si es que el equipo local tiene una ventaja sobre el equipo visitante es revisando el rendimiento como porcentaje de tiros convertidos, es decir, de los tiros que intentó el equipo, cuantos anoto.

¿Los rebotes capturados tienen alguna relacion al estar de local?

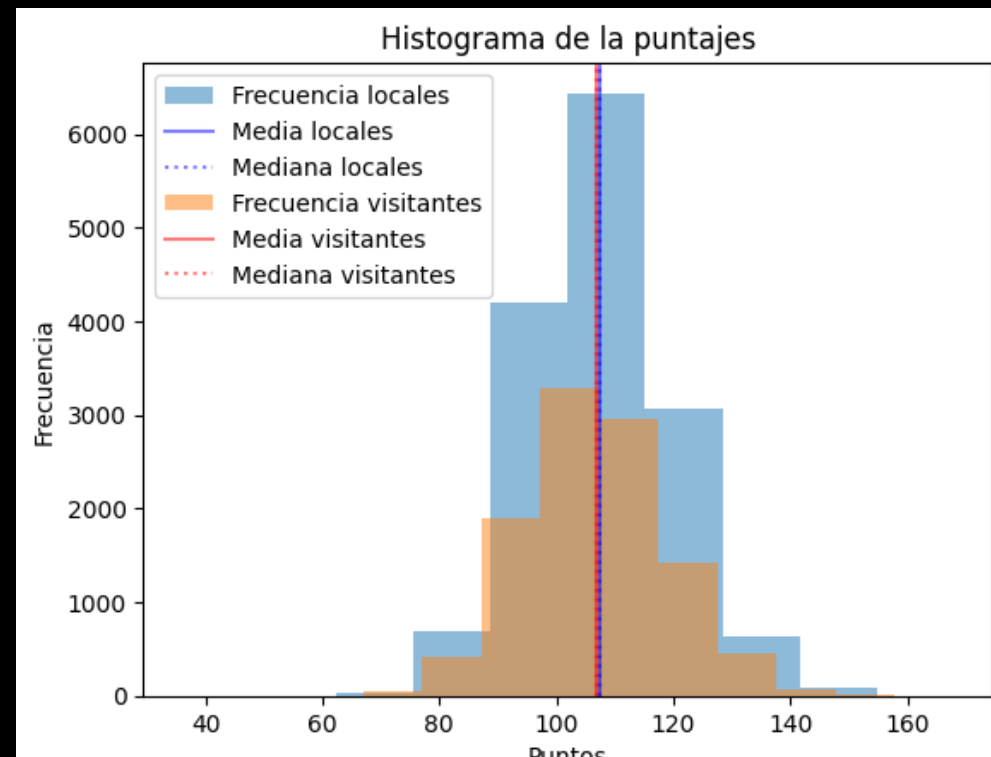
Al intentar buscar pruebas que demuestren que existen causas por las que ganen mas los locales que los visitantes , se puede plantear la hipotesis de que existe la posibilidad de que haya una mayor cantidad de rebotes capturados por el local cuando ganan.

Estas preguntas se pueden responder luego de un correcto analisis estadistico usando pruebas de hipotesis

Planteo rapido de datos estadistico para las preguntas propuestas

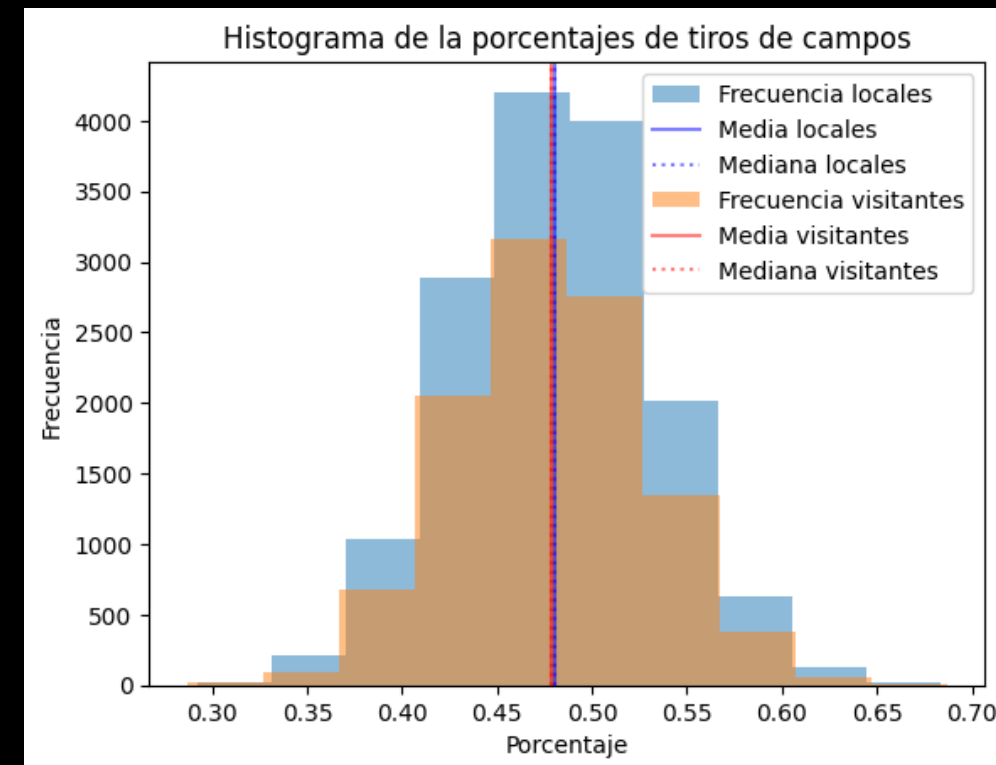
Hipotesis 1:

¿Tiene alguna ventaja significativa el equipo local?



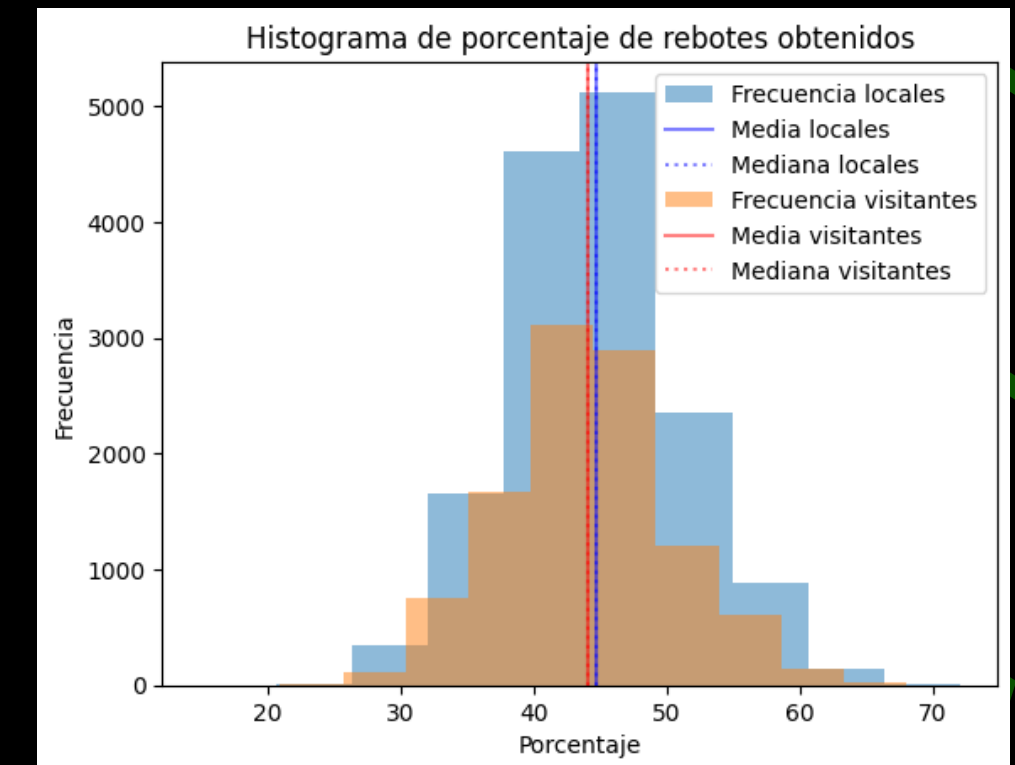
Hipotesis 2:

¿Los tiros de campo verifican lo mismo?



Hipotesis 3:

¿Los rebotes capturados tienen alguna relacion al estar de local?

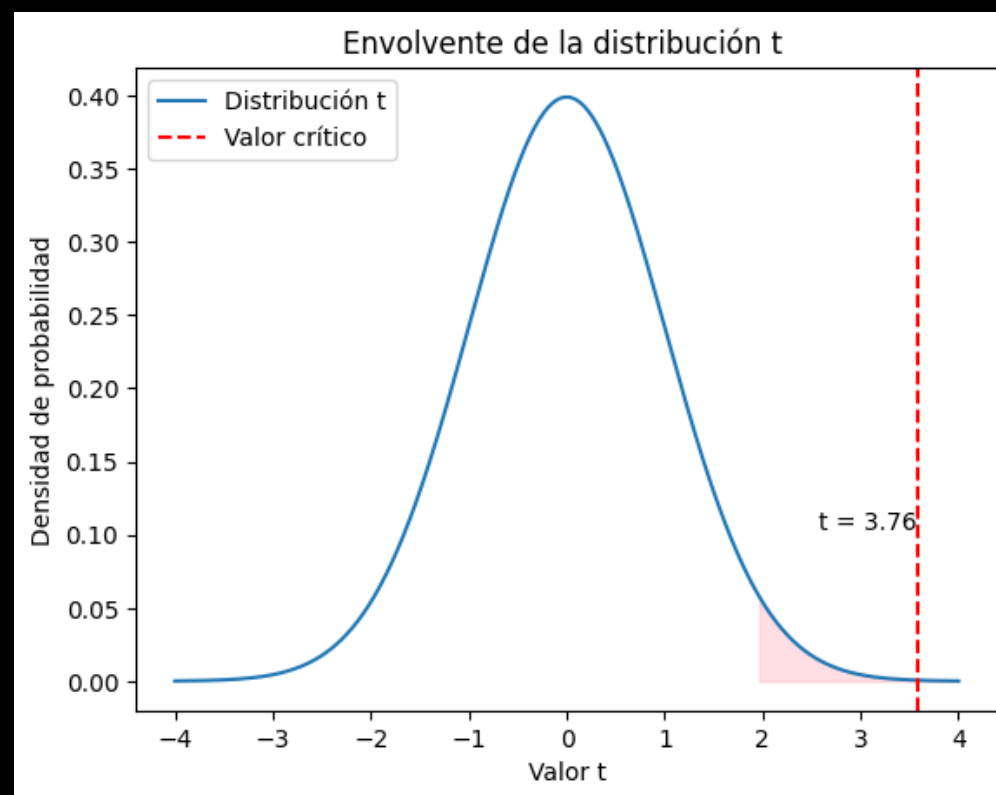


No hay conclusiones rapidas que verifiquen ventaja alguna!

Realizando las pruebas de hipotesis..

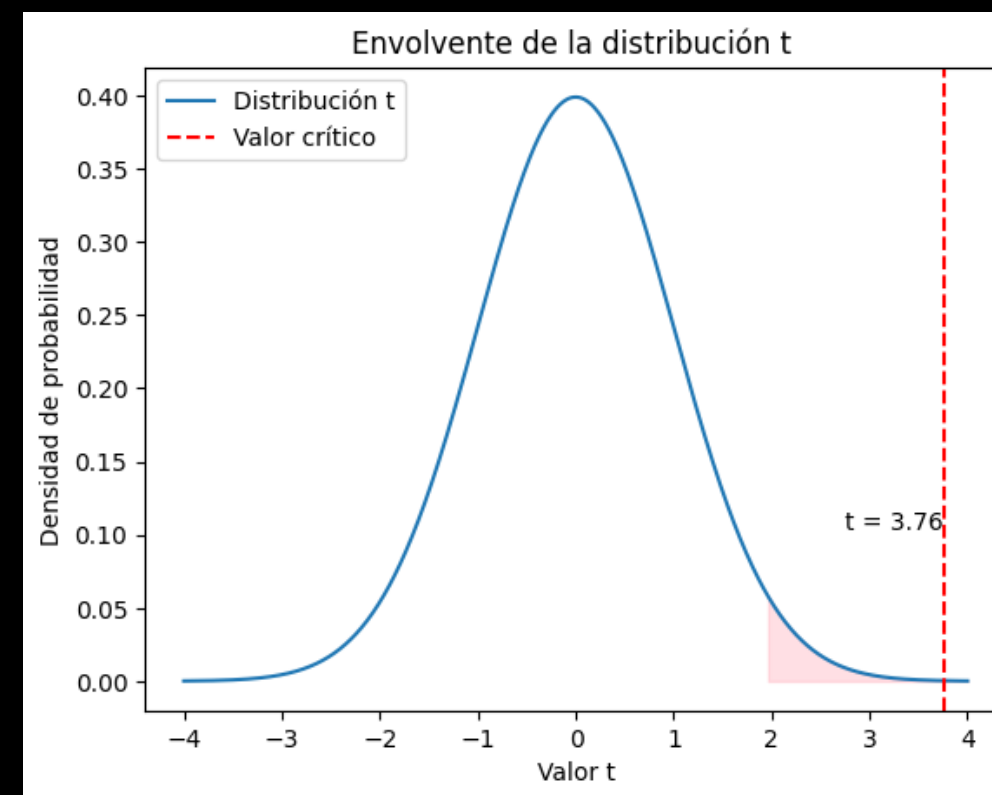
Prueba de hipotesis 1:

¿Tiene alguna ventaja significativa el equipo local?



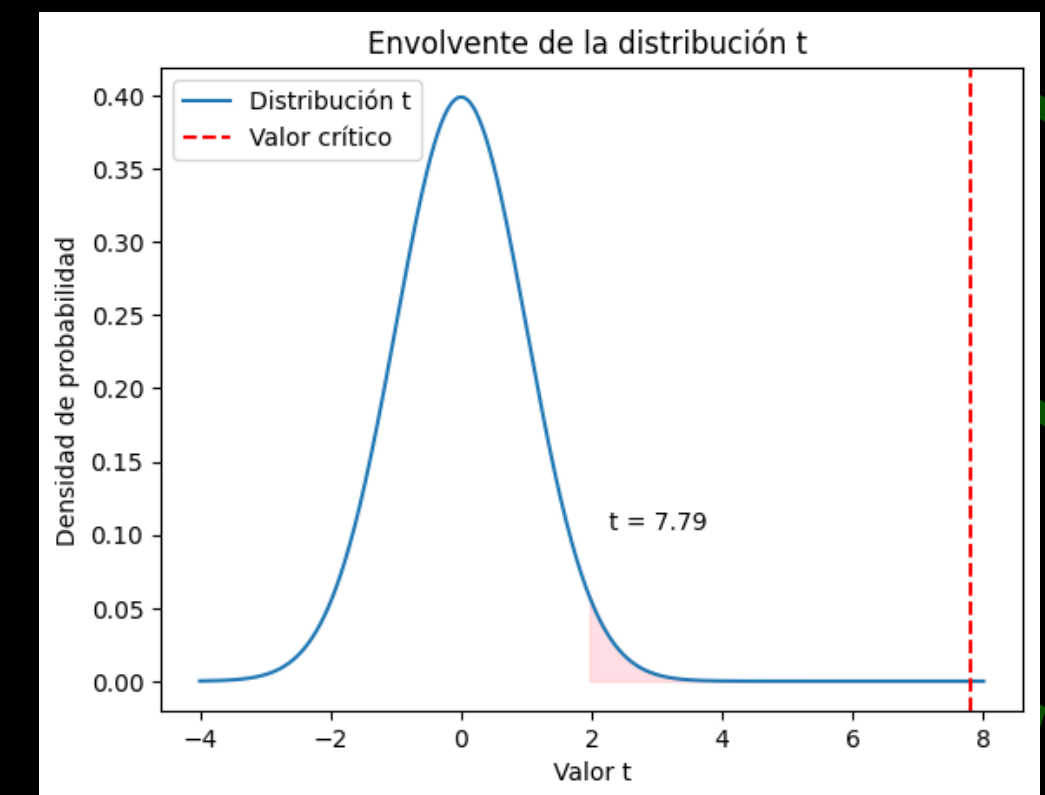
Prueba de hipotesis 2:

¿Los tiros de campo verifican lo mismo?



Prueba de hipotesis 3:

¿Los rebotes capturados tienen alguna relación al estar de local?

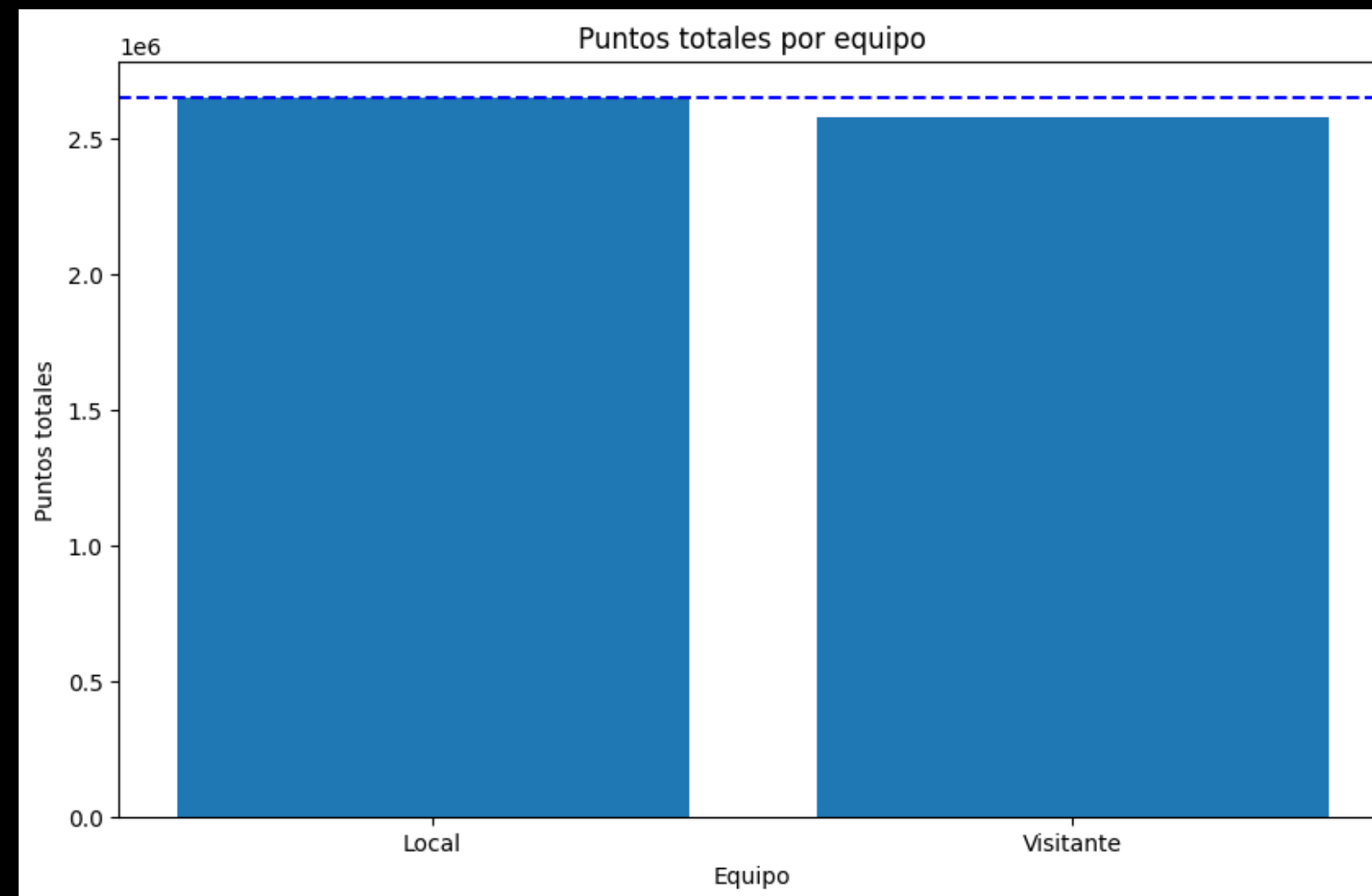


Los valores del examen de medias en las distribuciones de t de Student demuestran una ventaja en las medias de los puntajes tiros de campo y rebotes capturados cuando el local gana!!!

Analisis Exploratorio

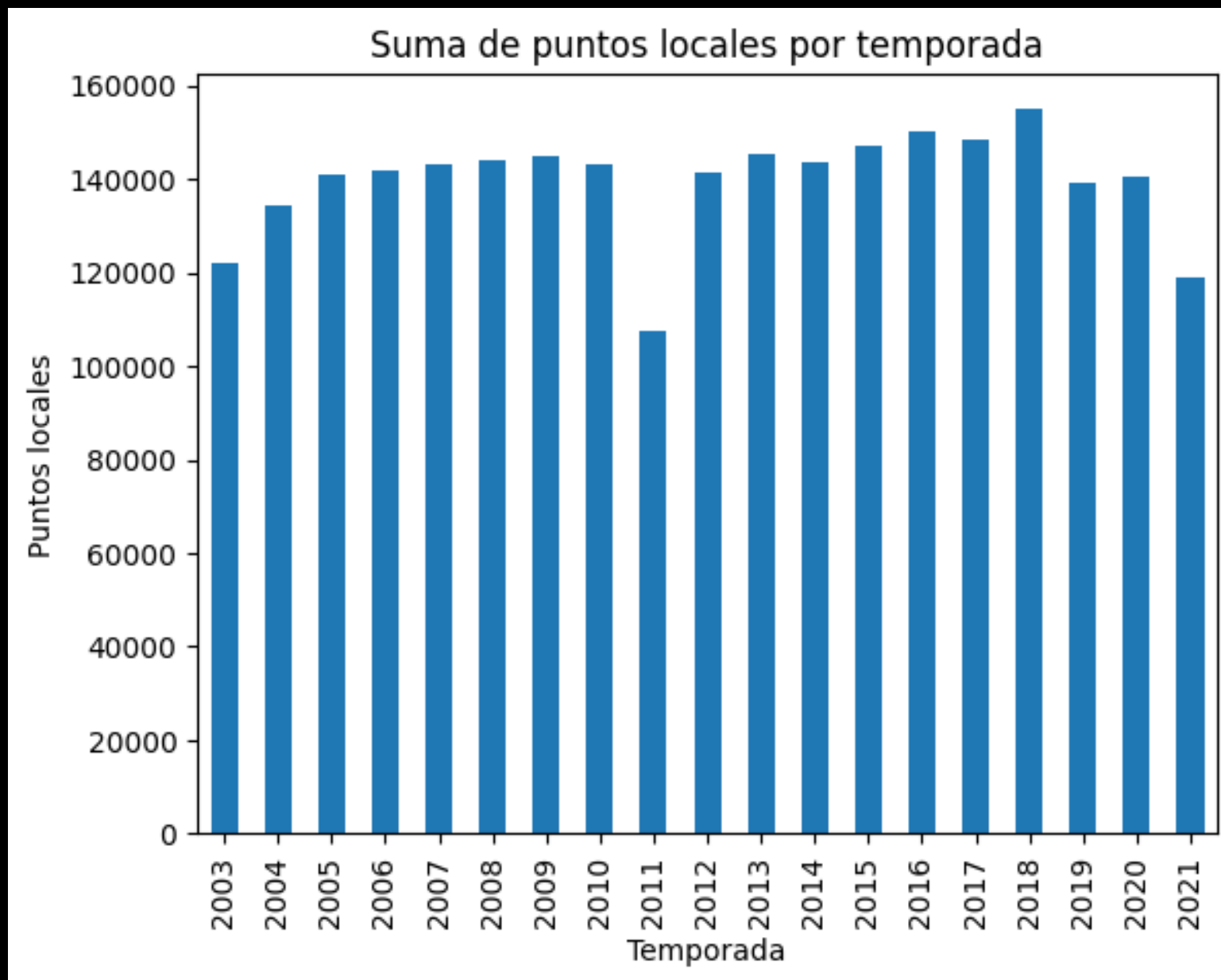
Al tener una verificacion de la certeza , se realizara una exploracion para encontrar algunos patrones visuales entre los datos

Puntos totales

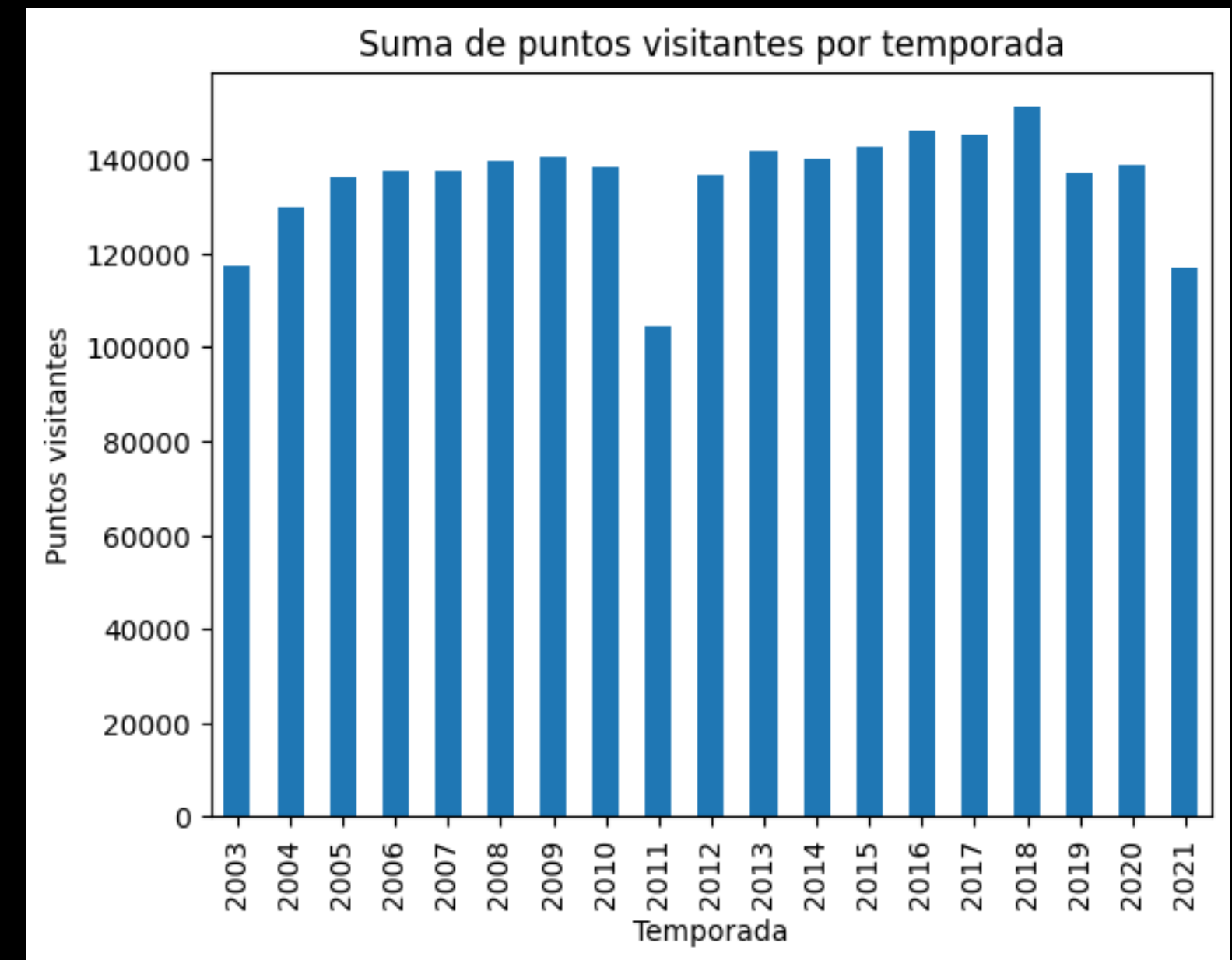


Analisis Exploratorio

Sumatoria de locales

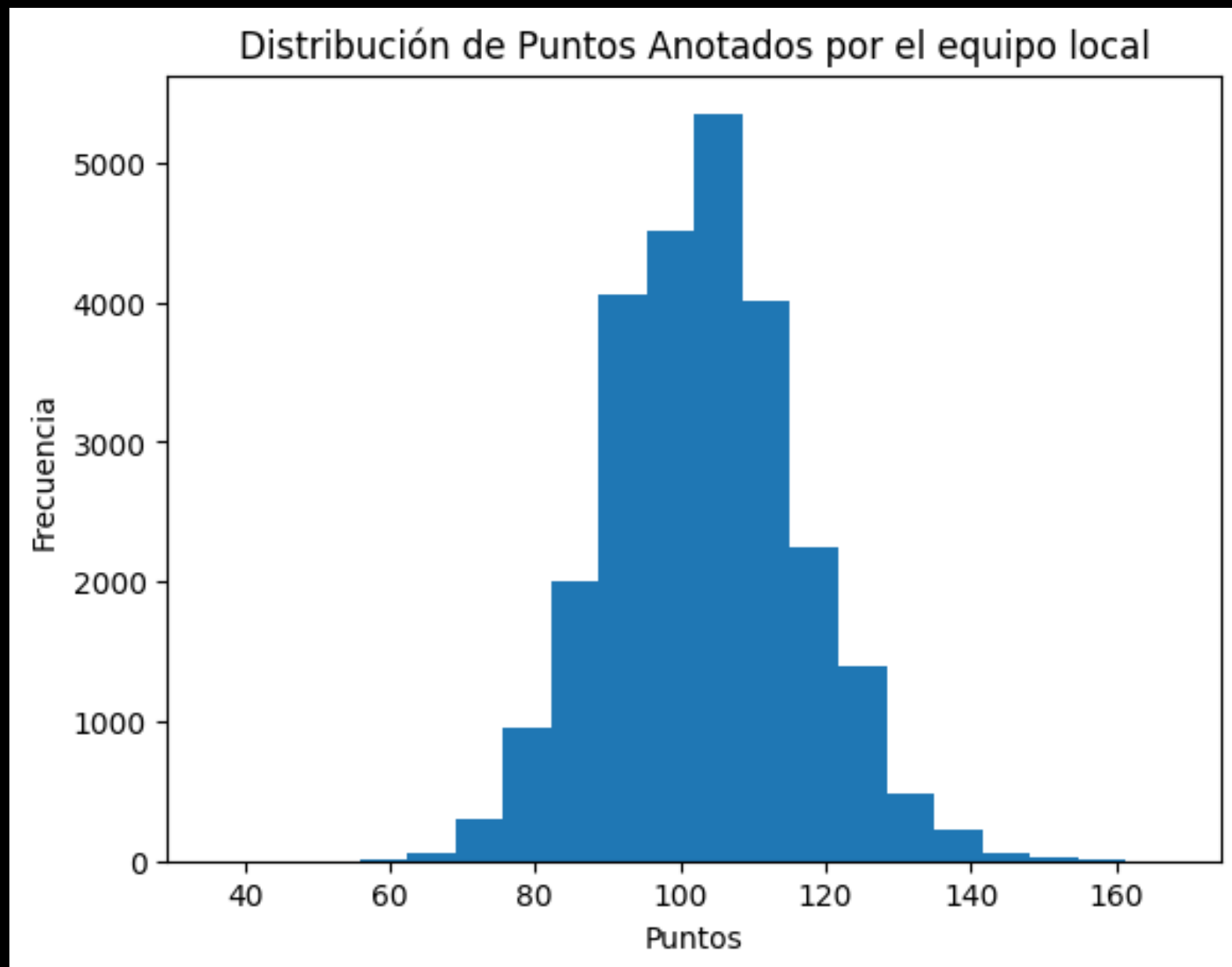


Sumatoria de visitantes

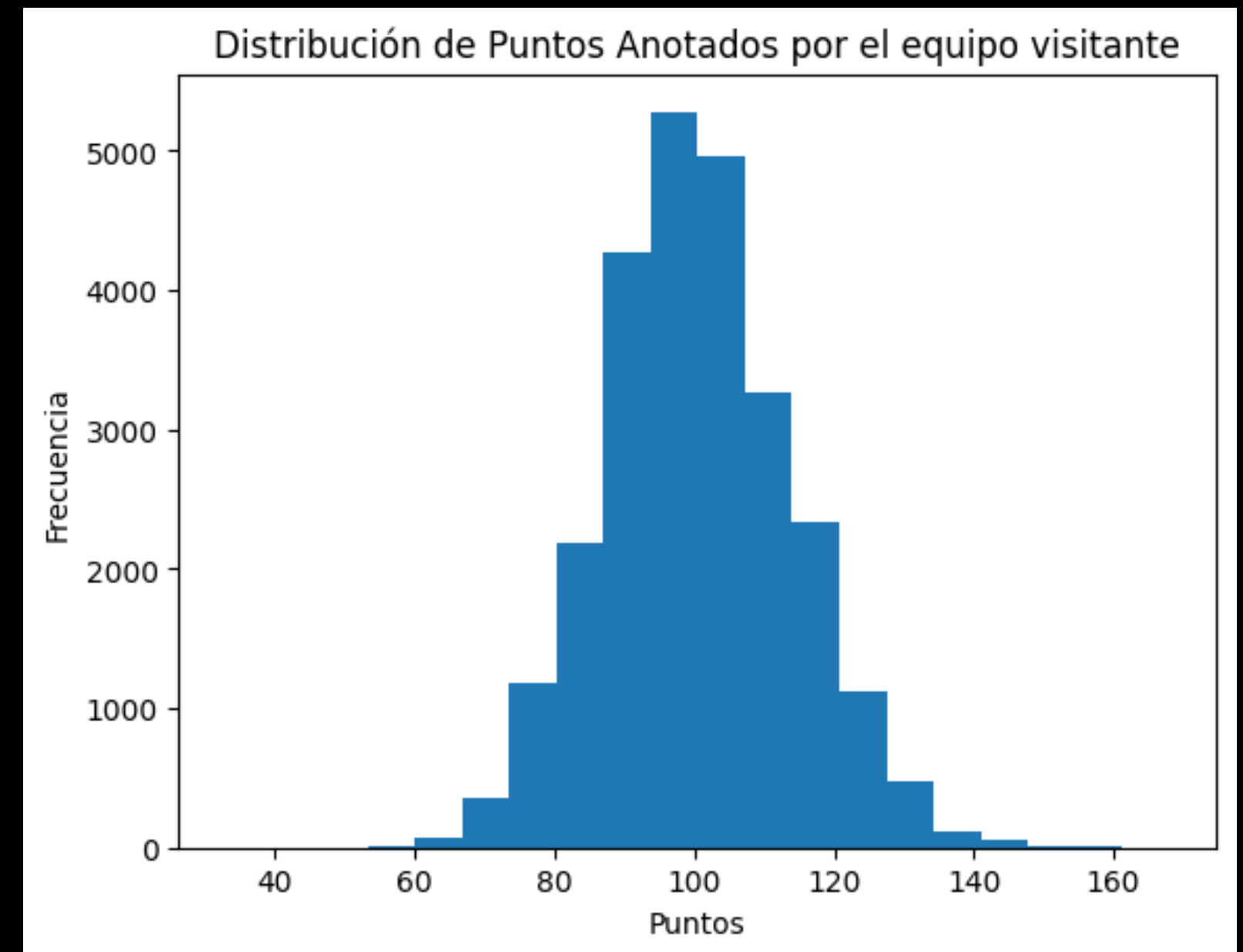


Analisis Exploratorio (Graficos univariados)

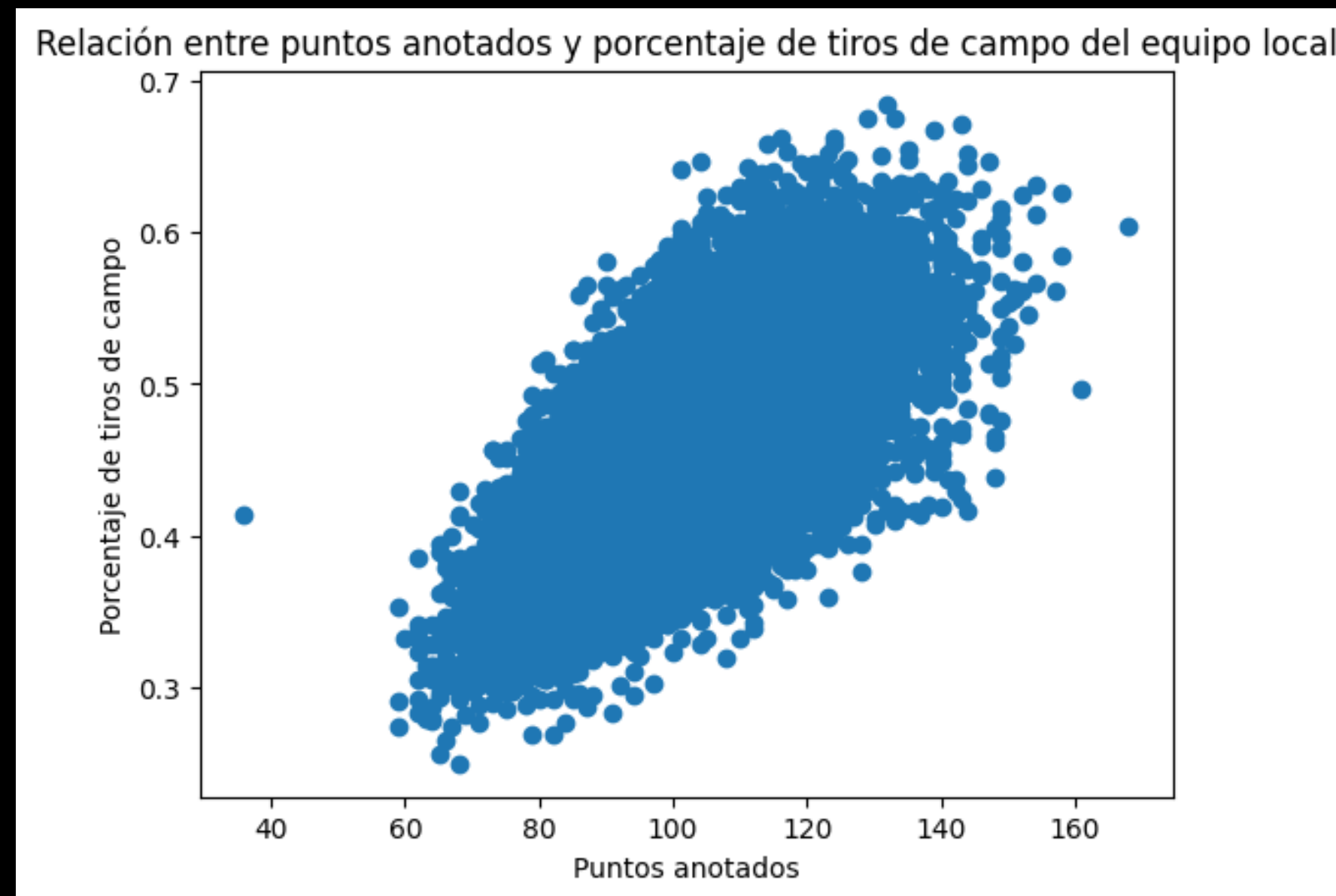
Histograma de locales



Histograma de visitantes



Puntos de local vs. Porcentaje de tiros de campo



Analisis Exploratorio (Graficos bivariados)

Puntos de local y visitante agrupados por mes y por año

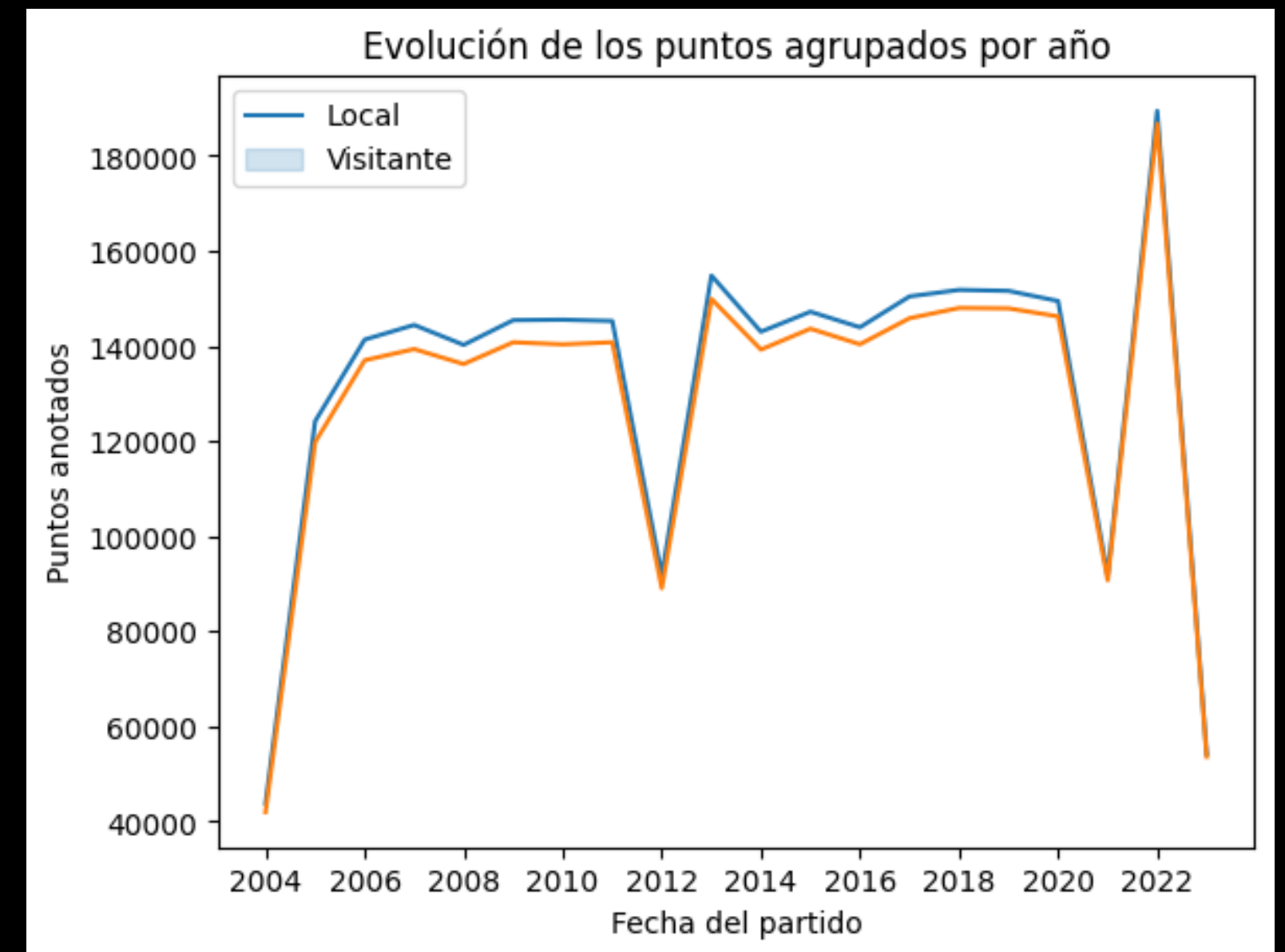
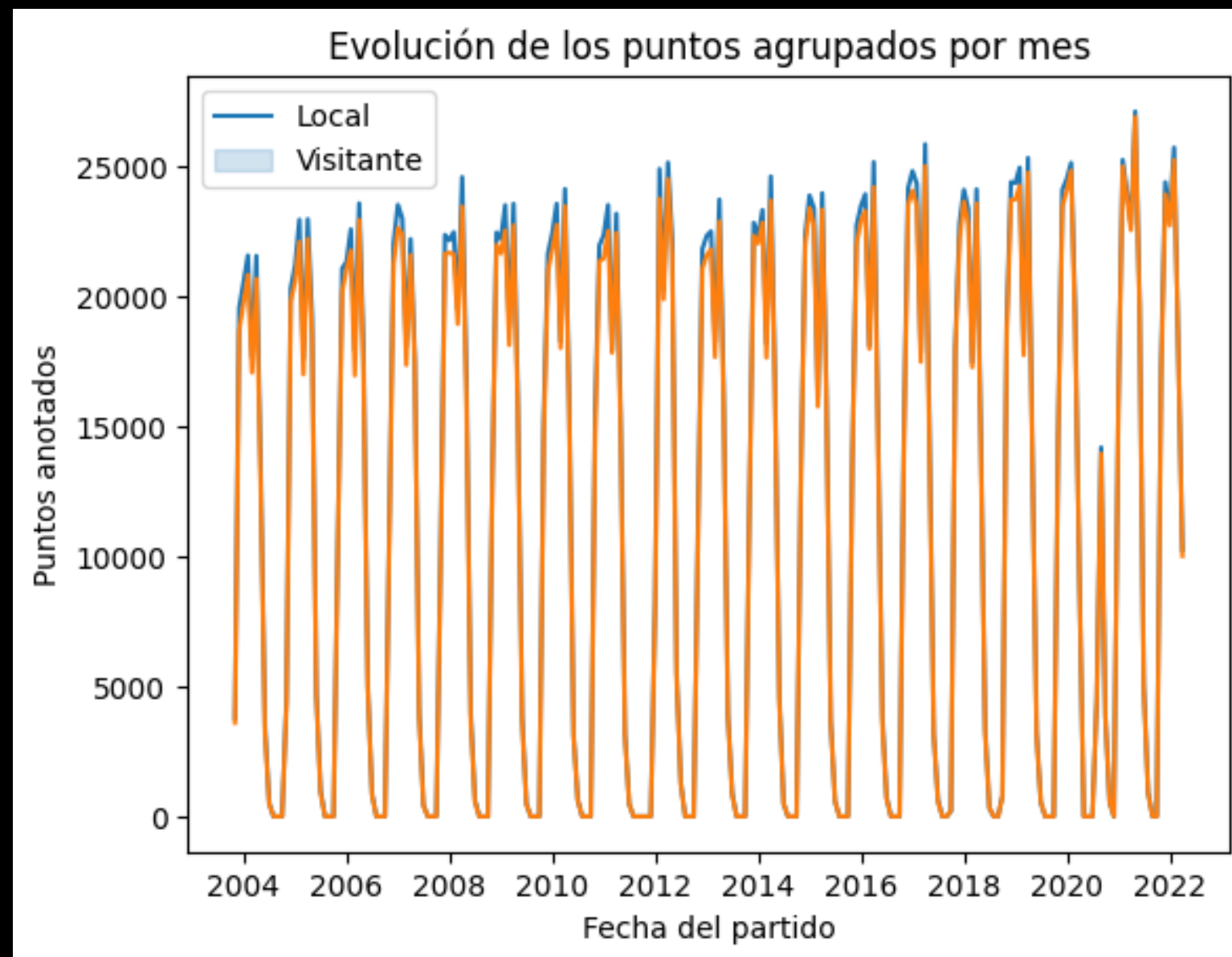
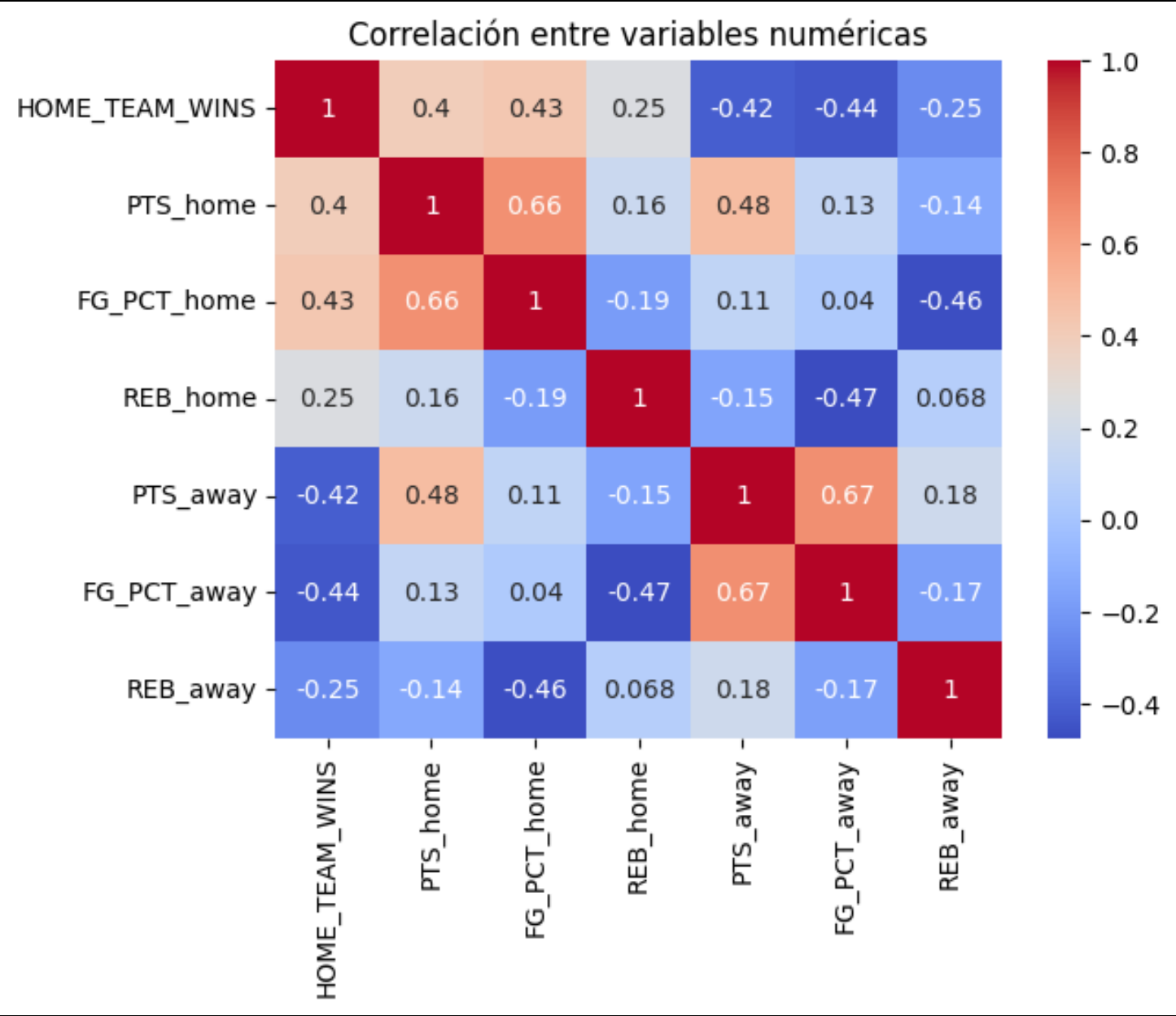


Grafico de correlacion entre variables del registro



Comentarios ante el comportamiento de los datos

Anteriormente vimos la simpleza que trae la visualización de los datos para poder encontrar un camino mas adecuado del proyecto hacia el final del mismo.

Las medidas de dispersion y los graficos variados facilitaron la obtencion de insights rapidamente por el comportamiento de los datos :

- Al analizar la sumatoria de los cohortes por temporada tanto para visita como para local los resultados son similares por lo que **no se pueden concluir grandes diferencias**.
- La distribución de puntos entre el equipo de casa y el visitante es **relativamente similar**.
- En el gráfico de caja podemos observar en general **cuando el equipo de casa ganó, convirtió mas triples** durante el partido.
- Se puede observar un **relación positiva** entre la cantidad de tiros y el porcentaje de tiros de campo como es esperable por lo tanto la data es consistente.
- Cuando el equipo juega de local, gana en un 60% de los casos lo que arroja la idea de que podría ser correcta nuestra hipótesis inicial.
- Se puede observar un **relación positiva** entre la cantidad de tiros y el porcentaje de tiros de campo como es esperable por lo tanto la data es consistente.
- El equipo local obtiene **mayor cantidad de puntos** con el paso del tiempo que el equipo de visitante.
- Hay correlaciones altas entre los puntos anotados tanto de visita como local y los porcentajes de tiros de campo.

Insights y conclusiones

Hemos conseguido comportamientos mejores a los esperados , anticipando un gran exito en la desicion de usar modelos para solucionar estos problemas. Los datos permitiran predecir ciertas tendencias que seran:

- La distribucion de los puntajes de locales agrupados y visitantes agrupados es similar
- Cuando el local ganá, posee un porcentaje mayor de triples durante el partido
- El equipo local tiene mayor probabilidad de anotar que el visitante
- El equipo local tiene mayor probabilidad de ganar (del 60% aproximadamente)

Invertir en este proyecto es una oportunidad prometedora gracias a su gran potencial de éxito.

