

TRABAJO 2



-Ignacia Quezada
-Diego Santibañez
-Maximiliano Artiaga

19/06/2023

Ingeniería civil en ciencia de datos
Visualización de datos

ÍNDICE

ÍNDICE	1
RESUMEN EJECUTIVO	1
PROCEDIMIENTO	1
PRE PROCESAMIENTO	1
STORYTELLING	2
CONCLUSIONES	9
REFERENCIAS	9

RESUMEN EJECUTIVO

En este trabajo, nos enfocaremos en el análisis detallado del comportamiento de una partida del videojuego League of Legends. La información que se proporcionará será exhaustiva y cubrirá desde la selección de los personajes hasta las decisiones que se tomen a lo largo de la partida, incluyendo la toma de objetivos.

Aunque los talentos de los jugadores son importantes para el juego, no los consideraremos en este análisis debido a que no podemos medirlos de manera objetiva. En su lugar, nos centraremos en el entorno de la partida en sí, con el objetivo de identificar las principales variables que influyen en la obtención de una victoria o una derrota.

PROCEDIMIENTO

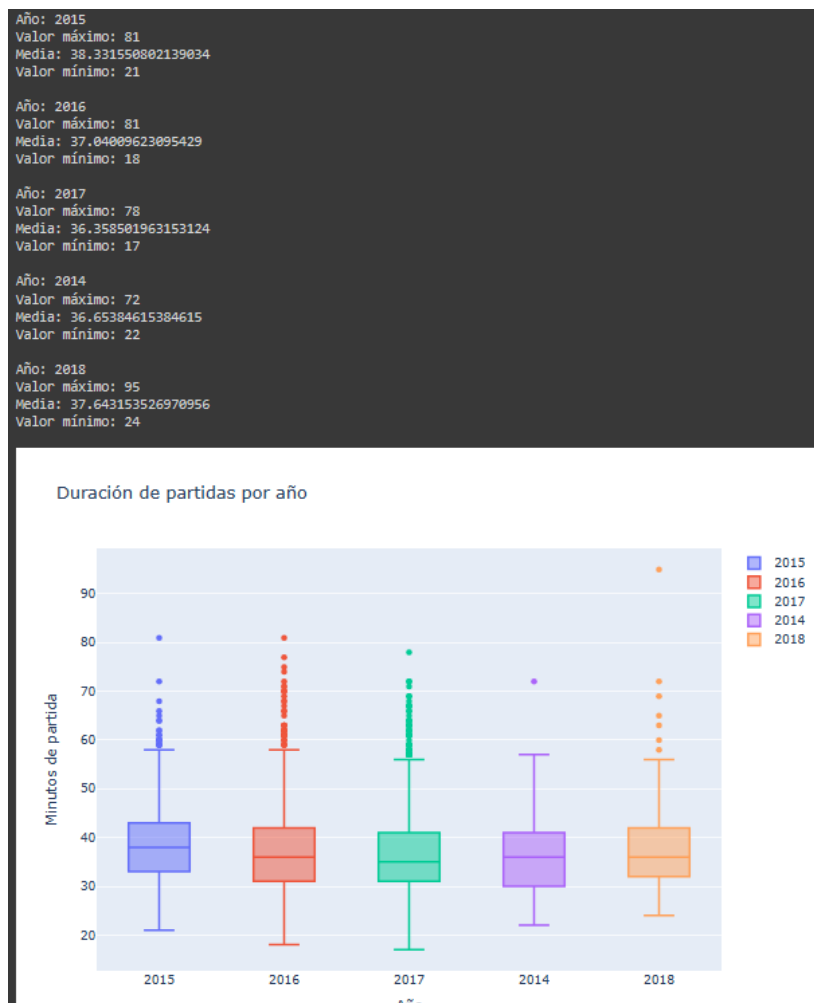
PRE PROCESAMIENTO

Como el estudio se enfocó en un set de datos donde se presentan 57 columnas, se tuvo que visualizar el contenido de cada una de estas, con las cuales se concluyó que para la limpieza de datos se eliminaron las columnas que no se utilizarán como estudio para este nuevo proyecto.

Ya evaluado esto se tomó la decisión de eliminar todas las columnas relacionadas a los equipos profesionales, ya que no serán de importancia para el estudio de datos realizados a continuación.

STORYTELLING

Se estudió la variable de duración de partidas como primer lugar, lo cual da como resultado que en todas las gráficas se obtuvieron unos valores fuera de la tendencia y del rango normal de por sí, por lo que se requirió una investigación de qué hacer con este valor o si era erróneo, con lo cual se verificó que es un dato real con lo cual no se eliminará por conservar todo dato real para nuestro estudio, luego de esto visualizamos los valores máximos, mínimos y sus medias que esto transformado a lenguaje más entendible se refiere a cuánto se demoró la partida mas larga, mas corta y cuanto dura una partida promedio y todo esto para cada año.



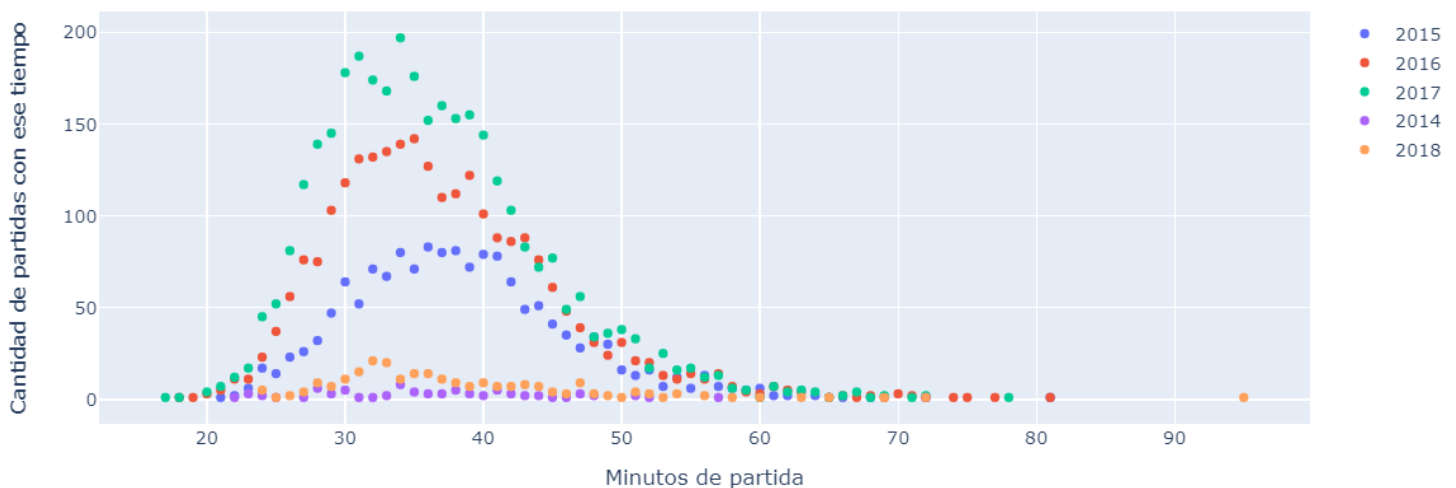
Por ejemplo acá se observa como hay varios datos fuera de la respectiva tendencia pero

todos estos valores no son atípicos, sino también valores reales que nos dan a entender que no todas las partidas están dentro del rango sino que muchas veces estas se pueden extender mucho más de lo que normalmente dura una partida, con esa justificación, se tomó la decisión de no eliminar estos valores ya que son igual de importantes que los que están en el rango, y la razón es que se estudiaron las condiciones de victoria tanto en el trabajo pasado como en este, la cual en el anterior se llegó a la conclusión de que la duración de partidas está relacionada directamente con la victoria con lo cual es importante no tocar estos datos, siempre y cuando no estén erróneos.

En el código de trabajo se dejó un link de la partida que más tiempo tomó, que sería la de 95 minutos, dándonos a entender que hay partidas de esa duración también. Estas medidas de tendencia se quisieron mostrar por año para un mayor identificación de los datos y que no se nos sobrecargue un solo Box.

Ya con todo eso en cuenta sólo se hizo un muestreo por año y con un conteo de partidas dependiendo de su duración, con lo cual llegamos a lo siguiente gráfico:

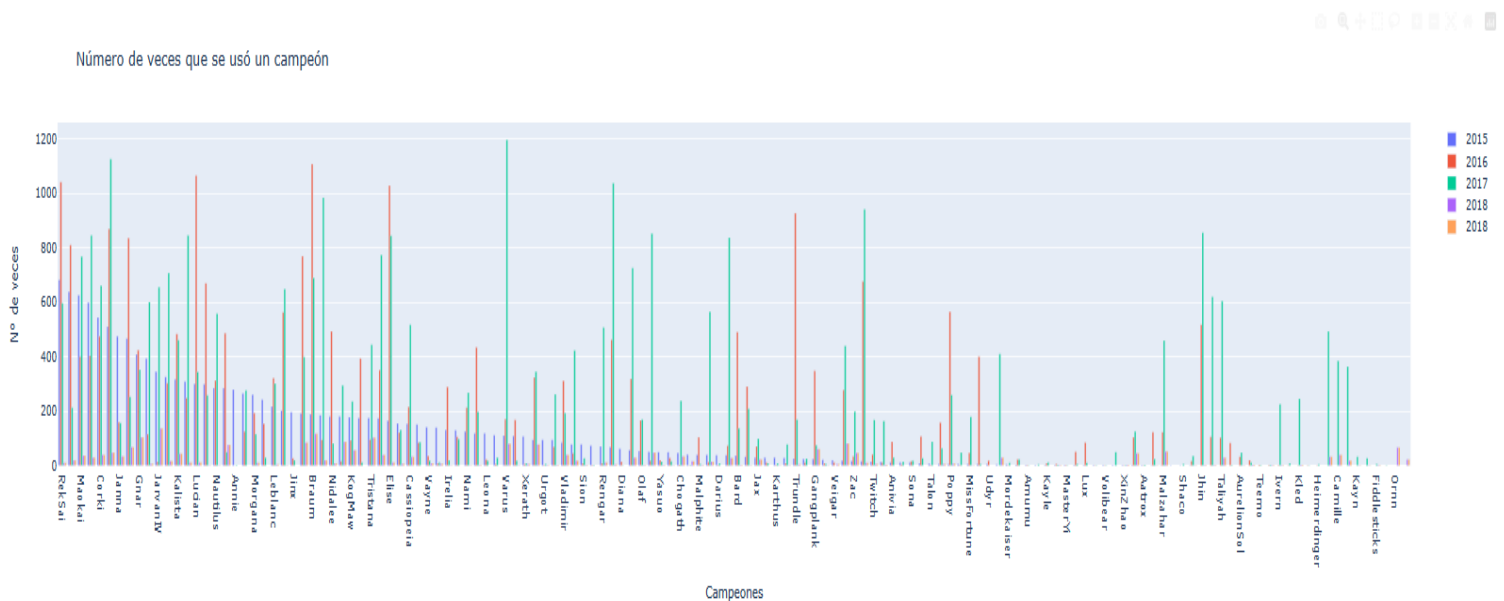
Duración de partidas por año



Dándonos como *eje 'x'* la duración por partida y por otro lado *eje 'y'* la respectiva sumatoria. Dependiendo de la duración, el gráfico se realizó para que este se visualice de manera interactiva para que al momento de querer ver un año en particular, solo se seleccione y así sea más amigable a la vista. Este gráfico nos ayuda a ver con más detalle la tendencia de los datos, ya que mientras más en la 'punta' de la pirámide vista, más datos hay con esos valores por año.

Desde el punto de vista como un cientista de datos, se puede se puede sugerir una condición como filtrado para evitar que estos datos extremadamente fuera del rango. Como un ejemplo de esto sería, si la partida dura más de 30 minutos que se visualice los datos en los cuales hay victorias y en cuáles no para luego evaluar si está relacionada de igual manera. Todo esto nos daría gráficas más “parejas” y con una tendencia más ordenada por la condición propuesta.

Otra variable que es importante en el set de datos y está relacionada con la victoria de una partida, son los tipo de campeones que se utilizan en las partidas. Para estos datos no se requirió una limpieza debido a que no tenían valores Null, por ende están correctos y por lo tanto solo se requirió una visualización por año, quedando de la siguiente manera:

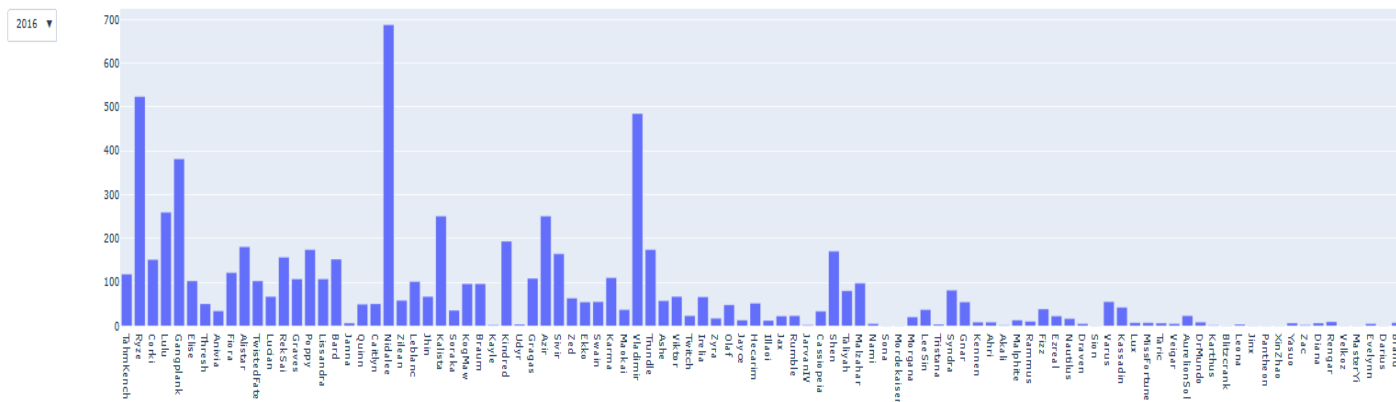


Esto nos muestra como va cambiando el uso de los campeones dependiendo del año en el que se encuentre, esta elección por parte de los jugadores no es azar, ya que a medida de que pasa el tiempo a los campeones se les va realizando ajuste constantemente en “parches” que se les van adhiriendo al juego para que así la victoria no depende de ellos, ya que al tener una gran variedad de campeones no es correcto que se ocupen solo algunos por tener más “Posibilidades” de victoria, sino hacer que estos dependan cada vez menos del tipo de campeón que se utiliza y que cada vez más dependa de la habilidad de los jugadores y el que hacen en la partida para poder obtener la victoria, una sugerencia para estos datos es evaluarlos por roles y a su vez clasificarlos por porcentaje de victoria. concluyendo se observa que a medida que se usó cierto campeón,

ayudará a establecer una relación más directa a la vez que se ajuste un campeón al “meta” respectivamente de ese año.

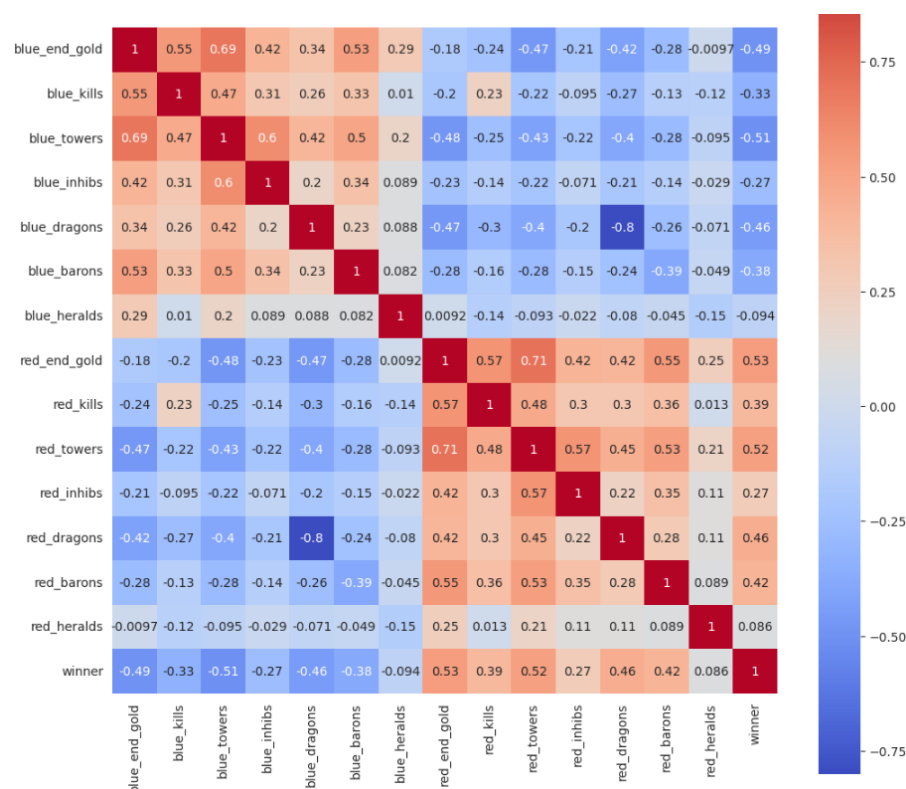
Junto con el estudio del gráfico anterior se tomó en cuenta la columna de “baneos”, la cual tiene una deficiencia, que es el cómo se registraron los datos, ya que con una investigación sobre el tema y en la visualización de los datos, nos dimos cuenta que desde cierto año, se empezaron a registrar más baneos, por ejemplo un valor del 2015 tenía en la columna de baneos 3 datos, en cambio en un valor de 2017 la columna no tenía 3 datos sino 5, (Todo esto fue debido a un parche en el juego que cambiaría a que todos los jugadores puedan banear un campeón y así sea más justo, ya que los baneos del 2015 sólo los hacía la cabecilla del team que se elegía aleatoriamente al momento de entrar a la selección de campeones) con lo cual para separarlos y que no tenga relevancia al estudio de datos separar los datos con un filtrado por año, dividiendo estos datos desde el año 2015 hasta el 2016. El código en esta parte fue más extensa debido a que se dificultó el cómo separar todos los datos de la columna para poder enumerar cuántas veces se ha baneado un cierto campeón, el cual se tiene entendido por la comunidad de que si se banea es porque tiene más “posibilidades de ganar” por sus “buffs” que se le hayan asignado en algún parche, así que con ese filtrado de datos para mantener la regla de los 3 baneos y no una mezcla de ambos años al banear, junto con lo últimamente visto en la materia de botones se aplicó un nuevo método actualizando la visualización dependiendo del año:

Cantidad de veces que aparece cada campeón en 2016



han baneado en el respectivo año, mostrando la cantidad de veces que se baneo por cada campeón, una sugerencia para la utilización de esto también es filtrarlos por roles y ver que tanto se logró una victoria baneando cierto campeón, esto ayudaría a identificar si hay alguna relación con la victoria directamente solo con algunos campeones en particular.

El uso de estos datos para establecer una victoria es importante junto con la toma de decisiones y también toma de objetivos, esta vez nos centraremos en el estudio de la prepartida que es la fase de selección, que muchas veces es conocida por la comunidad el “ganar antes del juego”, ya que una composición realmente “fuerte” por los ajustes realizados por las mismas personas de la empresa tiene como una “ventaja” sobre el otro, por eso como modelo a sugerir es establecer unas nuevas columnas donde almacene los campeones jugados en cierto partida por equipos y de eso establecer el si ganaron o no dependiendo de la composición, dándonos a ver si realmente es cierto eso que es conocidos por la comunidad que se puede ganar un juego con solo seleccionar una composición con una buena sinergia.



Como primera instancia del modelo, se visualizará la matriz de correlación para ver básicamente el comportamiento de las variables y su relación con las otras, en esta ocasión se añadió la columna “winner” que reemplaza a las columnas “blue_result” y

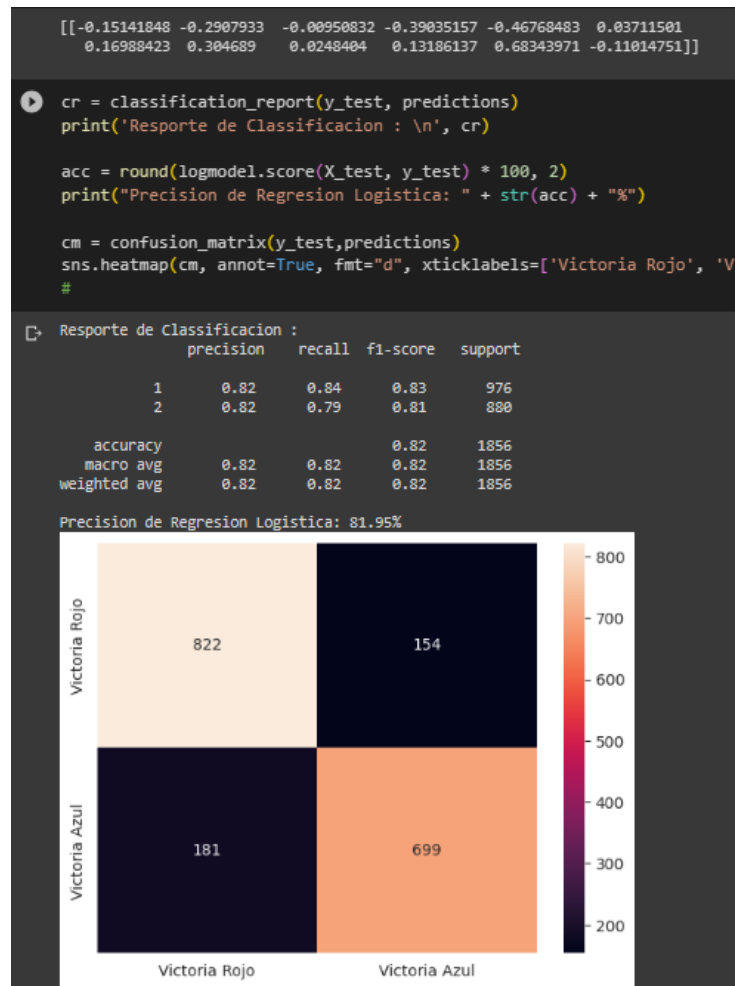
“*red_result*”, transformandolas de int64 a binarias haciendo más manipulable la columna y más fácil de visualizar.

Para esta matriz de correlaciones tenemos las siguientes columnas que serán definidas en palabras bastante simple:

- *blue_end_gold*: Significa la cantidad de oro al final de una partida.
- *blue_kills*: Bajas del equipo azul.
- *blue_towers*: Cantidad de torres derrumbadas por el equipo azul.
- *blue_inhibs*: Cantidad de inhibidores derrumbados por el equipo azul.
- *blue_dragons*: Cantidad de dragones por el equipo azul.
- *blue_barons*: Cantidad de barones por el equipo azul.
- *blue_heralds*: Cantidad de heraldos por el equipo azul.

Estas variables son tanto para el equipo blue (azul) como para el equipo red (rojo)

Un análisis hacia la matriz nos deja ver que las Torres destruidas, el oro ganado, los dragones por equipo y los barones tienen una correlación mediana, en cambio las bajas, destrucción de inhibidores y heraldos tienen una correlación baja, que nos quiere decir todo esto, que los que tienen una correlación media poseen un cierto impacto y no poco, pero no lo suficiente para determinar radicalmente que depende de ellos el ganar o perder una partida, en cambio los que tienen una correlación baja nos indica que no tiene tanta relevancia para determinar un resultado de victoria.



Ya como última instancia y como se ha mencionado antes del modelo un supuesto es determinar que tan es posible ganar una partida si dura más de 30 minutos, esto es lograble con un modelo de regresión logística, previamente se hizo el filtrado de los datos para obtener los valores que no tengan la hora corte y quedarnos con ese set de datos ya que será mucho más útil que el que ya teníamos, aparte de proporcionar más información.

Esta regresión logística es viable para nuestro dataset ya que este nos ayudará a ver el peso de cada variable con respecto a la variable “winner” con lo cual después solo bastaría ver la tasa de porcentaje de cada equipo ver tanto como cuantos ganan y cuantos pierden con respecto a la condición propuesta, y por el peso de cada variable y el preprocesamiento fue bastante notorio en esta parte del código, ya que se tuvo que manipular el dataset para nuestro filtrado dándonos mayor comodidad para después solo realizar el muestreo de los datos y su respectivo análisis.

CONCLUSIONES

En conclusión, hemos logrado obtener un modelo preciso utilizando los datos disponibles, a pesar de haber excluido información acerca de los jugadores y la duración de las partidas. Nos hemos enfocado en el análisis de los objetivos tomados por los equipos, los cuales han demostrado ser un factor clave en el éxito de cada partida.

En particular, hemos observado la importancia de los barones, uno de los objetivos más relevantes en el juego. La comunidad de jugadores reconoce que este objetivo puede tener un gran impacto en la partida, ya que no solo otorga beneficios en términos de oro, sino que también proporciona un buff o ventaja a cada campeón, lo que facilita la toma de otros objetivos. En resumen, cuantos más barones tenga un equipo, mayores serán las posibilidades de obtener otros objetivos y aumentar su ventaja en el juego.

REFERENCIAS

Murray, S. (2017). Interactive Data Visualization for the Web: An Introduction to Designing with D3. Taiwán: O'Reilly Media.

[Interactive Data Visualization with Python - Second Edition | Packt](#)

