


|   |   |
|---|---|
|  <p><b>UNIVERSIDAD AUTÓNOMA DE BUCARAMANGA</b></p> | <p><b>FUNDAMENTOS DE CIENCIA DE DATOS</b><br/> <b>FACULTAD DE INGENIERÍA INGENIERÍA</b><br/> <b>MECATRÓNICA</b></p> |
|---|---|

|                    |   |
|--------------------|---|
| <b>TEMA</b>        | <b>Análisis EDA de un dataset de LoL</b>  |
| <b>Integrantes</b> | <b>JOHAN STEBAN PINEDA DIAZ</b><br><b>JOSE DAVID MORA JAIMES</b><br><b>LUIS GUILLERMO HIGUERA ROJAS</b> |

### Objetivo:

Explorar y comprender las características del dataset de partidas de LoL, identificar patrones de desempeño, distribución de variables, y relaciones que puedan explicar el resultado de las partidas.

### Preguntas de investigación:

- ¿Cuál es el factor más determinante para ganar una partida?
- ¿Cómo varían las métricas clave (kills, assists, gold.) entre roles o campeones?
- ¿Existe una duración de partida "óptima" para ciertos campeones o roles?

## Análisis Exploratorio de Datos (EDA)

### 1. Comprensión y limpieza del dataset

#### a. Estructura y Tipos de Datos.

El conjunto de datos, extraído de la API, detalla las estadísticas de 31 partidas de League of Legends. La tabla resume las 24 columnas de las variables, incluyendo el nombre del campeón, el rol, asesinatos, muertes y oro ganado, junto con sus tipos de datos.

En la siguiente imagen se muestran los tipos de datos y las cantidades según su clase.

```
dtypes: bool(1), float64(1), int64(19), object(3)
```

Data columns (total 24 columns):

| #  | Column                 | Non-Null Count | Dtype   |
|----|------------------------|----------------|---------|
| 0  | Unnamed: 0             | 31 non-null    | int64   |
| 1  | gameCreation           | 31 non-null    | int64   |
| 2  | gameDuration           | 31 non-null    | int64   |
| 3  | gameId                 | 31 non-null    | int64   |
| 4  | gameMode               | 31 non-null    | object  |
| 5  | tournamentCode         | 0 non-null     | float64 |
| 6  | win                    | 31 non-null    | bool    |
| 7  | championName           | 31 non-null    | object  |
| 8  | role                   | 31 non-null    | object  |
| 9  | assists                | 31 non-null    | int64   |
| 10 | champExperience        | 31 non-null    | int64   |
| 11 | champLevel             | 31 non-null    | int64   |
| 12 | deaths                 | 31 non-null    | int64   |
| 13 | kills                  | 31 non-null    | int64   |
| 14 | doubleKills            | 31 non-null    | int64   |
| 15 | tripleKills            | 31 non-null    | int64   |
| 16 | killingSpree           | 31 non-null    | int64   |
| 17 | pentaKills             | 31 non-null    | int64   |
| 18 | longestTimeSpentLiving | 31 non-null    | int64   |
| 19 | totalDamageTaken       | 31 non-null    | int64   |
| 20 | totalHeal              | 31 non-null    | int64   |
| 21 | goldEarned             | 31 non-null    | int64   |
| 22 | goldSpent              | 31 non-null    | int64   |
| 23 | itemsPurchased         | 31 non-null    | int64   |

## b. Manejo de Valores Faltantes (Nulos).

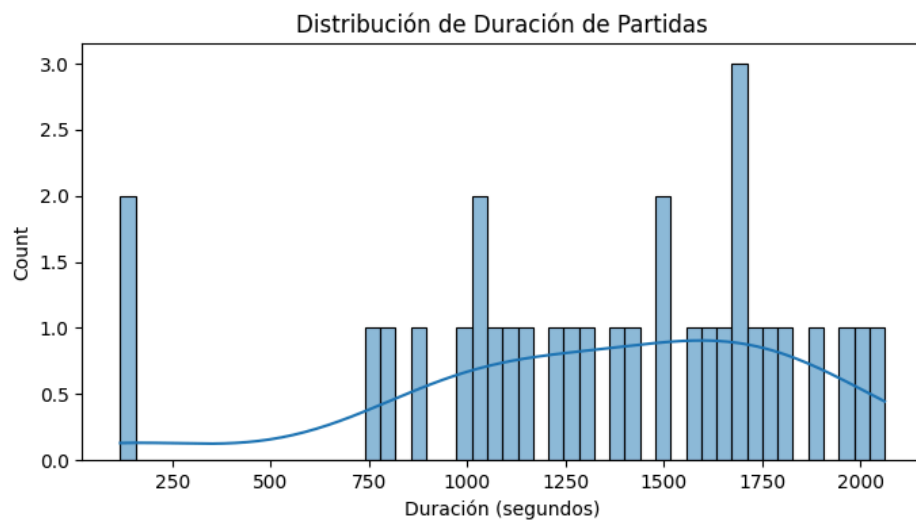
El conjunto de datos está completo y no presenta valores faltantes. La columna 'tournamentCode' está vacía, lo cual es un resultado esperado, ya que ninguna de las partidas analizadas fue de torneo.

```
--- Valores nulos ---
Unnamed: 0      0
gameCreation    0
gameDuration    0
gameId          0
gameMode        0
tournamentCode  31
win             0
championName    0
role            0
assists         0
champExperience  0
champLevel      0
deaths          0
kills           0
doubleKills     0
tripleKills     0
killingSpree    0
pentaKills      0
longestTimeSpentLiving  0
totalDamageTaken  0
totalHeal       0
goldEarned      0
goldSpent       0
itemsPurchased  0
dtype: int64
```

**c. Valores Atípicos (*Outliers*).**

i. Partidas extremadamente cortas

Al analizar el tiempo de duración de las partidas, encontramos que existen 2 partidas con una duración menor a 5 minutos, también se evidencian 4 partidas con una duración menor a 15 minutos,

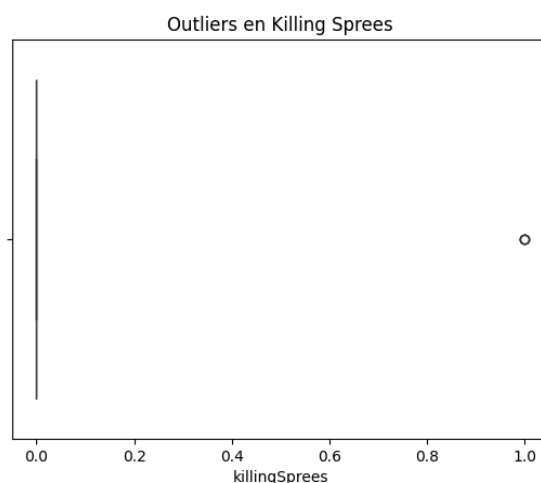


```
--- Partidas extremadamente cortas (< 1000 s) ---
   gameId  gameDuration   win championName
5   1662694108         118 False      Seraphine
25  1648427922         139  True       Nidalee
2   1663769486         744  True    Blitzcrank
3   1663765687         803  True       Aurora
9   1660681722         893  True      Seraphine
```

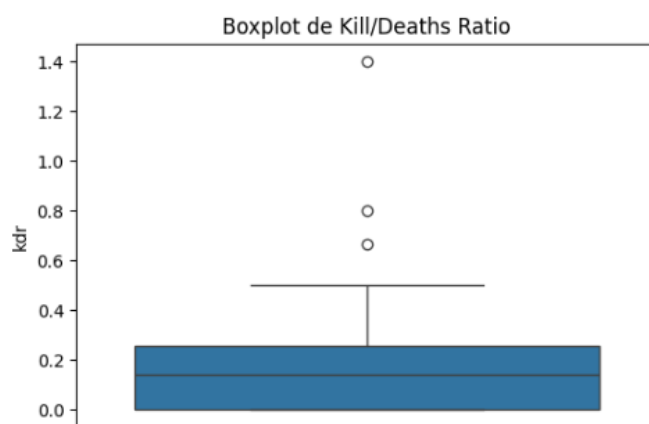
## ii. Performance excepcional ( killing Sprees)

El gráfico de caja para 'killing Sprees' (rachas de asesinatos) identifica un claro valor atípico (outlier) en 1.0. Esto indica que una sola partida logró una racha. La gran mayoría de los datos se concentran en 0.0, lo que significa que lo común fue no tener ninguna. Esta partida excepcional, jugada por Seraphine, es un evento estadísticamente raro en este dataset.

En esta partida tuvo 10 muertes, 5 kills, 1 double kills y Perdió.



Este gráfico de KDR (Kill/Death Ratio) muestra que la gran mayoría de las partidas tuvieron un rendimiento bajo, con la mediana y el rango principal (IQR) concentrados muy cerca de cero. Sin embargo, tres partidas destacan como valores atípicos con KDRs significativamente superiores: Aurora (1.4), Sona (0.8) y Yuumi (0.67). Estos *outliers* representan las mejores performances del conjunto de datos.



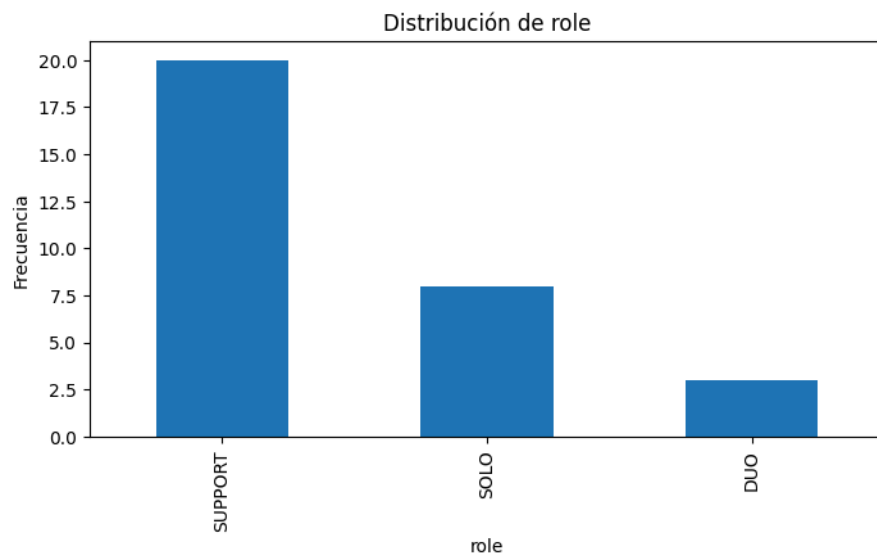
```
Campeones con Kill/Death Ratio fuera del rango normal:
 championName kills deaths kdr
0      Aurora    7      5  1.400000
08      Sona     4      5  0.800000
00      Yuumi    2      3  0.666667
```

## 2. Análisis univariado

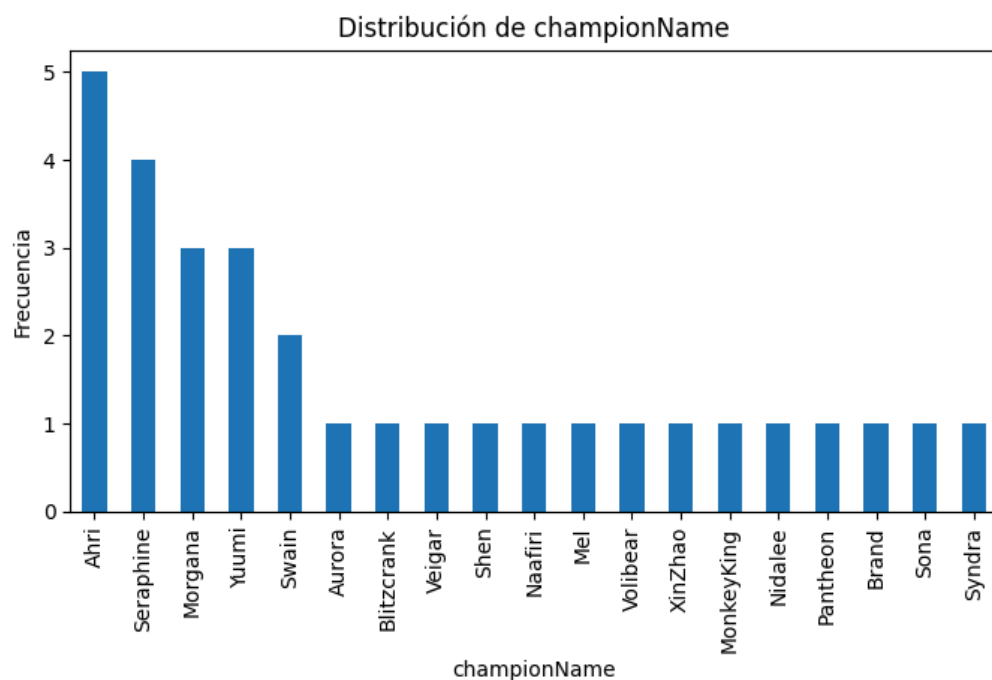
### a. Categóricas:

- role, championName, gameMode, win
- Gráficos: Barras, pie chart
- Pregunta: ¿qué roles o campeones se juegan más? ¿Qué porcentaje de victorias hay?

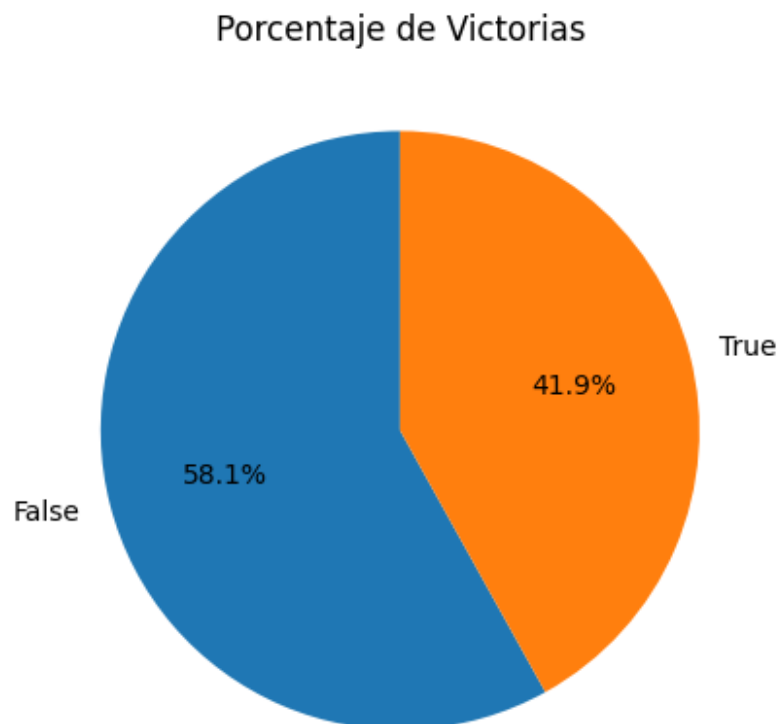
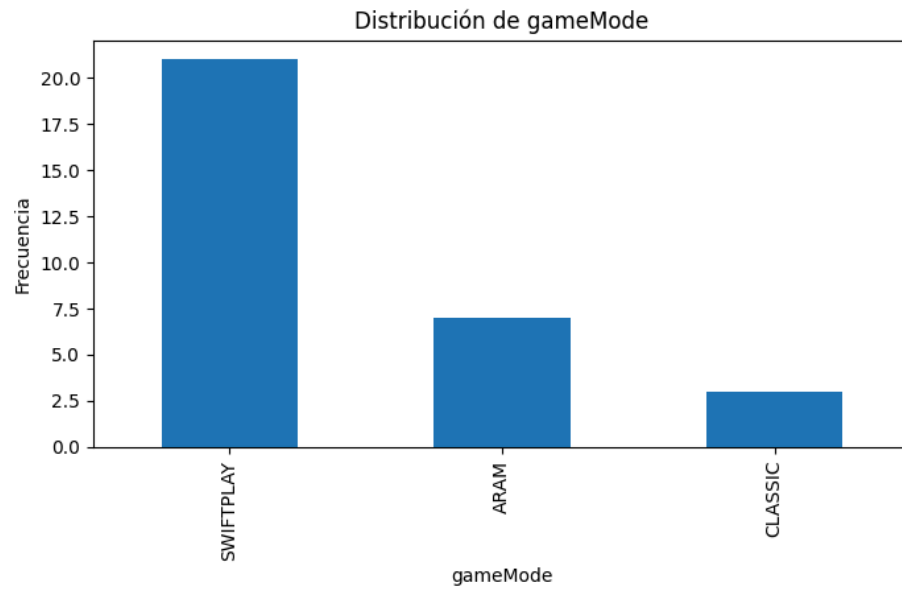
El gráfico de distribución de rol muestra que el rol preferido de la jugadora es Support con 20 partidas, seguido por solo con 8 partidas.



En el gráfico de distribución de campeones observamos que ha experimentado con 19 campeones, de los cuales con 14 solo los jugó 1 vez, y encontramos que sus campeones preferidos son Ahri y Seraphine



Al analizar la distribución de gameMode, encontramos que su modo preferido de juego es swift play el cual es un modo de juego acelerado del juego clásico que se juega en la grieta del invocador, Está diseñado para ofrecer una experiencia más rápida y casual que los modos de juego estándar, ideal para partidas cortas.



b. Numéricas:

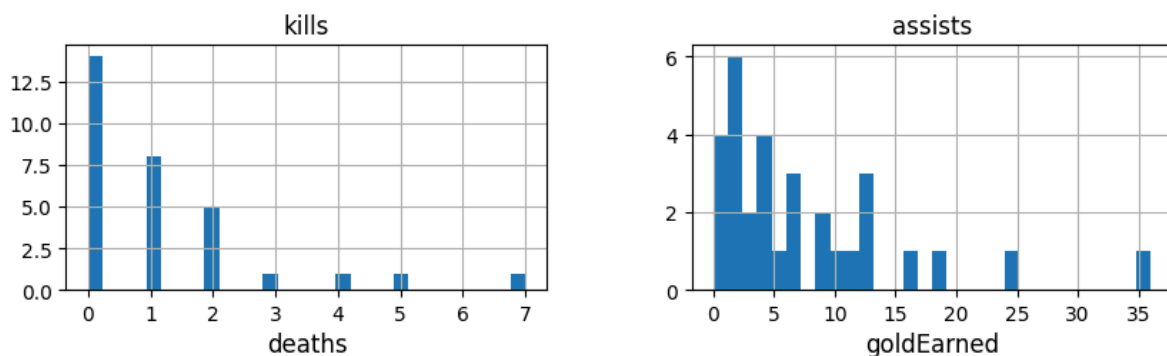
- kills, assists, deaths, goldEarned, champExperience,
- Gráficos: histogramas, boxplots
- Pregunta: ¿qué tan dispersas están las métricas de desempeño?

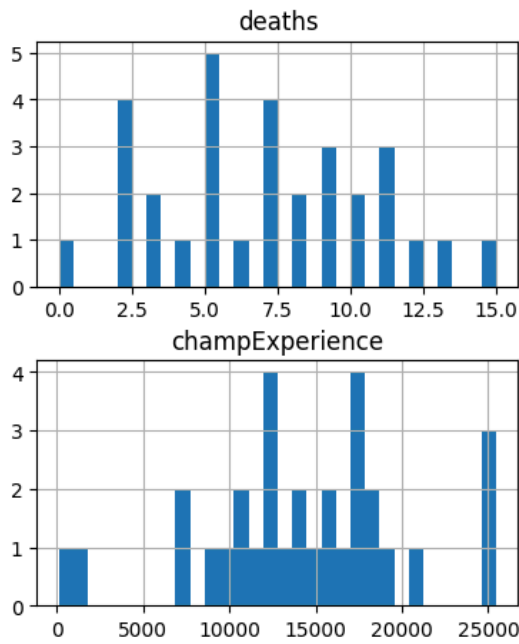
**Métricas como 'deaths' y 'champ Experience' están más extendidas** a lo largo de todo su rango, lo que indica una gran variabilidad y falta de consistencia en el desempeño entre una partida y otra.

Los histogramas muestran la distribución de cinco métricas clave de desempeño. Se observa un patrón claro en la mayoría de ellas:

- **Kills (Asesinatos):** La gran mayoría de los datos están fuertemente concentrados en 0 y 1. Es una distribución muy sesgada a la derecha, lo que indica que la mayoría de las partidas tuvieron muy pocos asesinatos, con muy pocas excepciones.
- **Assists (Asistencias):** Similar a 'kills', está sesgada a la derecha. La mayoría de las partidas tienen entre 0 y 10 asistencias, con valores atípicos que llegan hasta 35.
- **Deaths (Muertes):** Esta métrica está más distribuida que 'kills' o 'assists', con valores que se reparten de forma más uniforme entre 2.5 y 12.5, aunque el pico está alrededor de 5.
- **Gold Earned (Oro Ganado):** También muestra un sesgo a la derecha, con un pico notable de partidas alrededor de los 10,000 de oro, pero con muchas partidas por debajo y algunas por encima.
- **Champ Experience (Experiencia):** Esta es la variable más dispersa y uniformemente distribuida. No hay un pico dominante claro; en su lugar, los valores se reparten por todo el rango, desde 5,000 hasta 25,000.

Distribución de Variables Numéricas



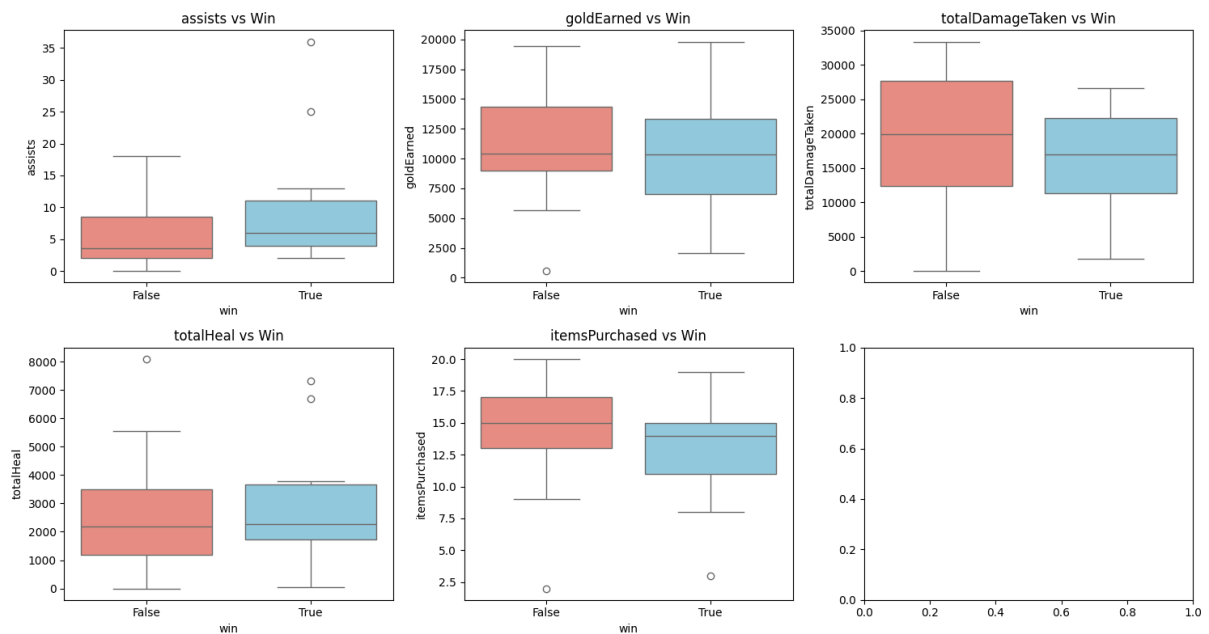


### 3. Análisis bivariado

#### a. Rendimiento vs resultado:

- Comparar kills, assists, deaths, goldEarned, etc. contra win = True/False

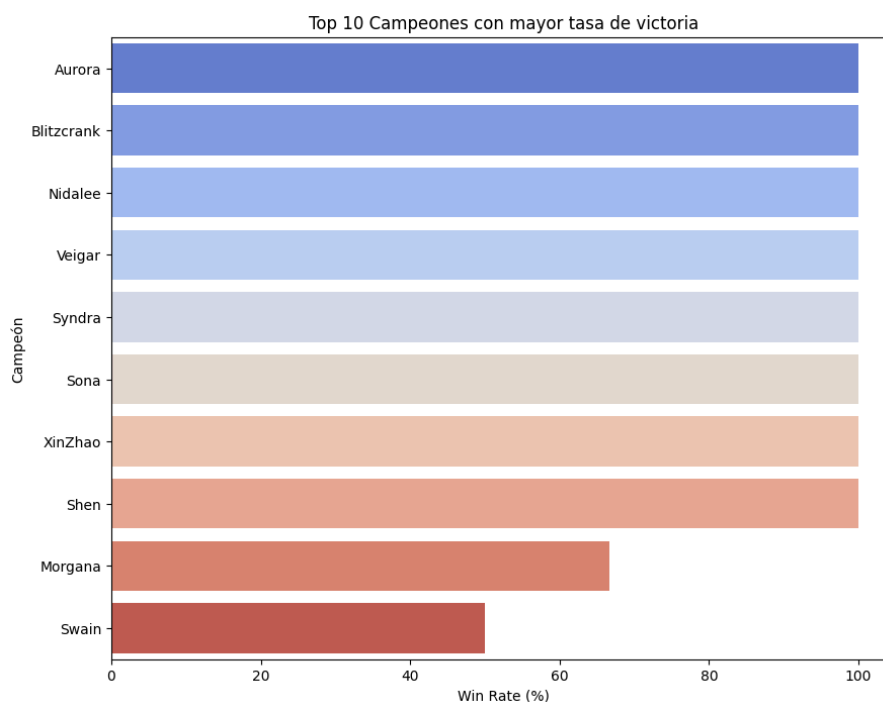
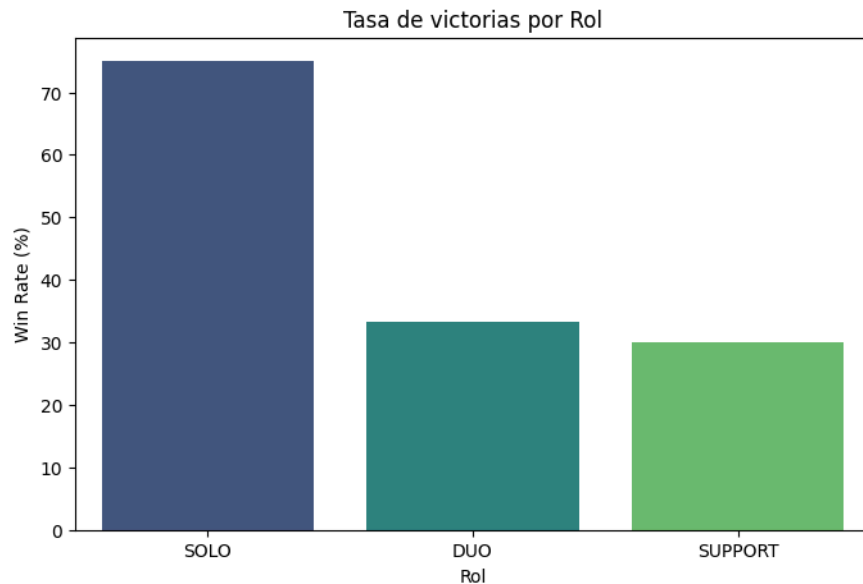
En esta sección se analizaron las relaciones entre las variables de desempeño del jugador y el resultado final de la partida (win = True/False), utilizando diagramas de caja (boxplots) para comparar la distribución de los valores.





b. Rol o campeón vs desempeño:

- Analizar la tasa de victorias (Win Rate) por championName y role.  
¿Hay roles o campeones que tienen un mayor impacto en la victoria?



En cuanto al análisis de Win Rate por rol, se observó que el rol SOLO presenta la tasa de victorias más alta, con más del 60% de win rate, seguido del rol DUO con aproximadamente 32%, y finalmente el rol SUPPORT con una tasa cercana al 18%. Este patrón sugiere que los jugadores en posiciones individuales o con mayor control sobre el combate directo tienen más impacto en el resultado.

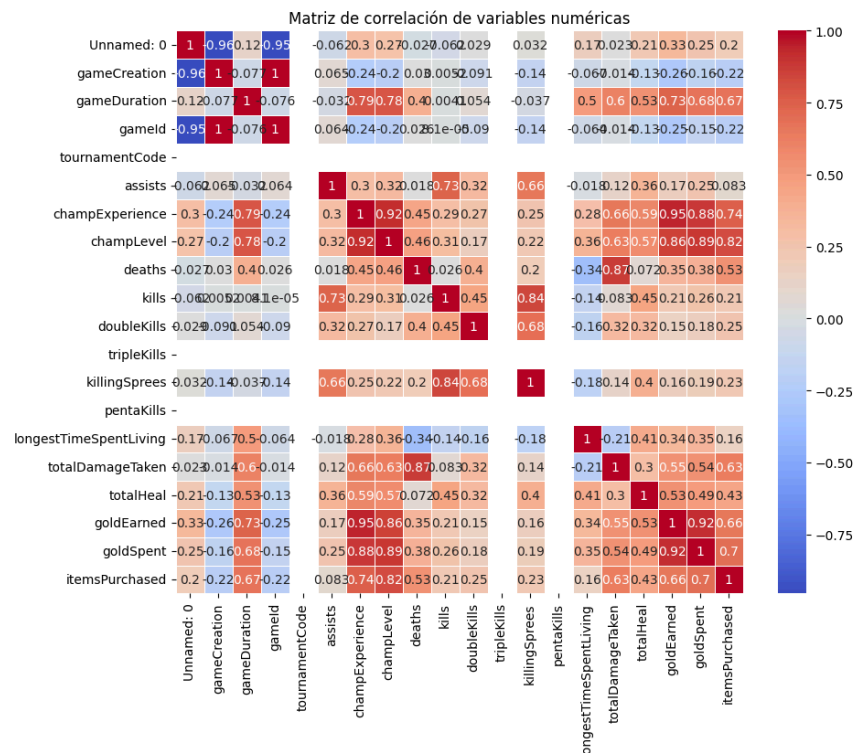
Por último, en la comparación de Win Rate por campeón, los campeones Aurora, Blitzcrank, Nidalee, Veigar, Syndra, Sona, Xin Zhao y Shen muestran un 100% de tasa de victoria, mientras que otros campeones se mantienen en valores superiores al 65%. Este resultado

refleja que ciertos personajes presentan ventajas tácticas o se adaptan mejor a las estrategias utilizadas en las partidas analizadas.

## 4. Análisis Multivariado

### a. Correlaciones numéricas:

- Matriz de correlación (heatmap) para ver qué variables están más relacionadas.

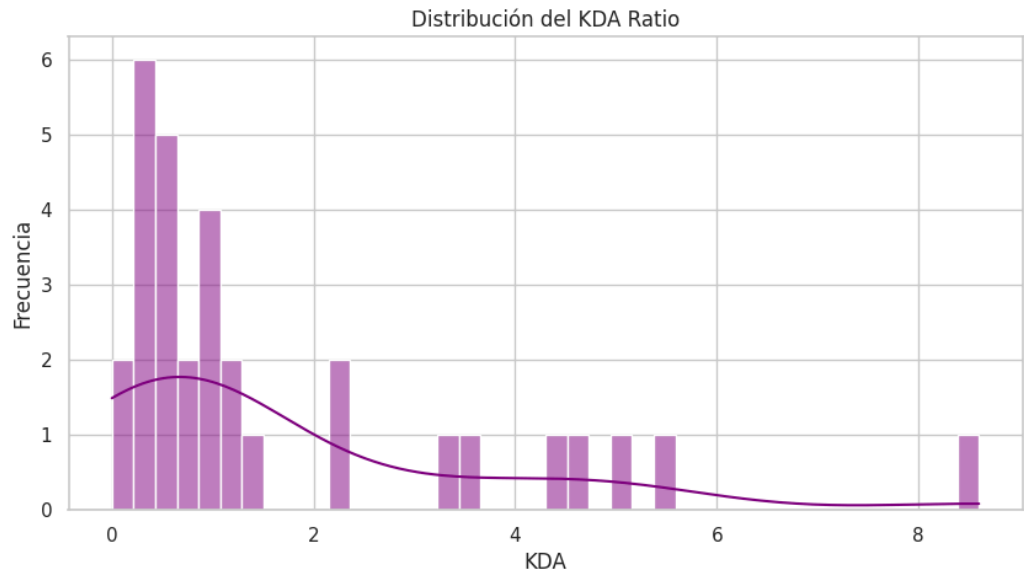


champ Experience, champLevel, goldEarned y items Purchased presentan correlaciones altas (superiores a 0.85), indicando que a mayor experiencia y nivel del campeón, mayor es el oro obtenido y la cantidad de objetos comprados.

En general, el heatmap evidencia que las variables relacionadas con el desempeño del jugador están fuertemente interconectadas, mientras que las variables de identificación o tiempo de creación de partida (Unnamed: 0, game Creation, gameId) no presentan correlaciones significativas con las variables de rendimiento.

## 5. Métricas Derivadas

- KDA ratio:  $(\text{kills} + \text{assists}) / \max(\text{deaths}, 1)$



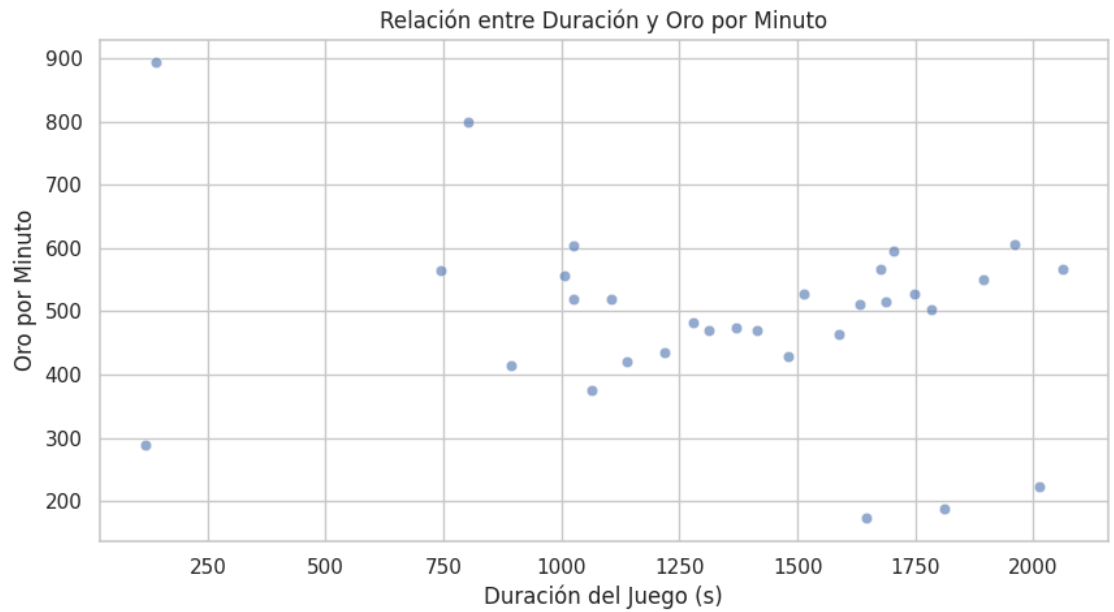
### Análisis:

- Promedio general de Majia: **1.71**
- Un KDA superior a 3 indica un excelente desempeño, mientras que valores bajos (<1) reflejan más muertes que participación en eliminaciones.

### Conclusión:

Majia presenta un **rendimiento medio**, con un balance entre asesinatos, asistencias y muertes. En algunas partidas logró un impacto notable con altos KDA, pero en otras su desempeño fue más discreto.

- Gold per minute:  $\text{goldEarned} / (\text{gameDuration} / 60)$



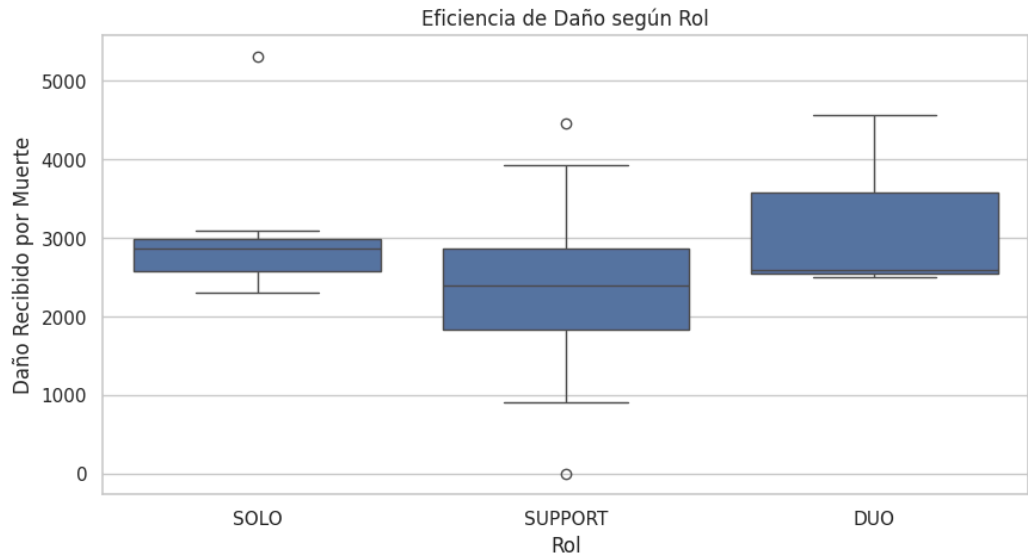
### Análisis:

- Promedio de Majia: **491.31 de oro por minuto**
- Un valor por encima de 600 suele indicar una excelente eficiencia en farreo y control de recursos.

### Conclusión:

Majia mantiene un **ritmo de oro estable**, suficiente para adquirir ítems clave, aunque todavía hay margen para optimizar el farreo y la toma de objetivos para aumentar su impacto económico en las partidas.

- Damage efficiency:  $\text{totalDamageTaken} / \text{deaths}$



### Análisis:

Promedio: **2647.15**

Un valor alto puede reflejar un estilo más resistente (tanques o soportes), mientras que uno bajo suele indicar poca supervivencia o posicionamiento riesgoso.

### Conclusión:

Majia muestra una **eficiencia de daño equilibrada**, resistiendo una cantidad considerable de daño antes de morir. Esto sugiere una buena gestión del posicionamiento y un rol posiblemente más defensivo o de apoyo en combates.

## 6. Conclusiones

- a. Identificar principales factores de victoria.

El análisis revela que los factores que más influyen en las victorias de Majia son el oro por minuto (GPM) y el KDA ratio. Las partidas donde Majia mantuvo una buena eficiencia económica (superior a 500 de oro por minuto) y una proporción alta entre eliminaciones y muertes coincidieron con la mayoría de sus victorias.

Esto sugiere que el control de recursos, la obtención de experiencia y la supervivencia durante las peleas son determinantes para el éxito. Por el contrario, las partidas con muchas muertes o bajo farmeo suelen terminar en derrota, evidenciando la importancia de mantener estabilidad y consistencia a lo largo del juego.

- b. Determinar qué roles/champions dominan ciertas métricas.

Al segmentar las partidas por rol, se encontró que Majia obtuvo su mayor tasa de victorias jugando en rol SOLO, con más del 60 % de win rate, mientras que DUO y SUPPORT presentaron tasas inferiores.

Esto indica que Majia rinde mejor en posiciones donde puede influir directamente en el resultado del enfrentamiento sin depender tanto de otros jugadores.

En cuanto a los campeones, los resultados muestran que Aurora, Blitzcrank, Nidalee, Veigar, Syndra, Sona, Xin Zhao y Shen fueron campeones con 100 % de efectividad en las partidas jugadas.

Esto sugiere que Majia obtiene mejores resultados con personajes que combinan control, daño sostenido o utilidad táctica, aprovechando sus fortalezas en coordinación y posicionamiento.

- c. Detectar tendencias o patrones de comportamiento (por ejemplo, supports con más asistencias pero menos oro).

A lo largo de las partidas analizadas, se identificaron patrones claros en el estilo de juego de Majia:

- Los supports tienden a registrar más asistencias, pero menos oro y kills, lo que refleja un rol enfocado en apoyo y visión del mapa.
- En roles solitarios, Majia presenta mayor impacto directo en daño y oro, correlacionándose con las victorias.
- El modo de juego preferido, *Swift Play*, refuerza un estilo de juego rápido y adaptable, enfocado en partidas cortas donde las decisiones tempranas son decisivas.
- Las métricas más correlacionadas con el éxito (como experiencia, nivel, oro y objetos comprados) confirman que un crecimiento sostenido dentro de la partida es clave para la victoria.

## Síntesis final

En conjunto, los datos reflejan que Majia tiene un estilo de juego consistente pero dependiente del rol: destaca más cuando asume control individual, mientras que en roles de soporte su desempeño es correcto, aunque con menor impacto directo.

La eficiencia económica, el posicionamiento y la supervivencia son los elementos más determinantes en sus partidas, y optimizarlos podría traducirse en una mejora significativa de su tasa de victorias.

