

## Segunda Entrega Proyecto Final Probabilidad y Estadística 1

Daniel Morales, Laura Ortiz, Juan Reina

Adriana Pineda

### 1. Marco Metodológico

**Base de datos a Utilizar:** Valorant  
Champion Tour 2024 Data

La base de datos utilizada corresponde al Valorant Champions Tour 2024, que incluye estadísticas detalladas sobre los partidos, equipos y jugadores de las competiciones regionales y el torneo mundial Valorant Champions. Los datos son recolectados por vlr.gg a partir de transmisiones oficiales, registros de Riot Games y plataformas de estadísticas de e-sports.

#### Método de Recolección:

La entidad vlr.gg obtiene los datos directamente de las transmisiones oficiales de los partidos, los registros de Riot Games y otras plataformas de estadísticas de e-sports. Estos datos se almacenan en su base de datos y se actualizan regularmente para su consulta por parte de los usuarios.

#### Métodos de Análisis:

- **Estadística descriptiva:** Para resumir tendencias como daño promedio, tasa de victorias por mapa y frecuencias de agente.
- **Análisis multivariado:** Se utilizará el **análisis de agrupamiento** para

identificar estrategias similares entre equipos y jugadores.

- **Visualización de datos:** Se crearán **gráficos de barras, radar y mapas de calor** para representar estrategias y patrones de desempeño.
- **El objetivo**, es encontrar las estrategias óptimas (META) del torneo y analizarlas para mejorar el balance del juego y torneo **Valorant Champions 2024**.

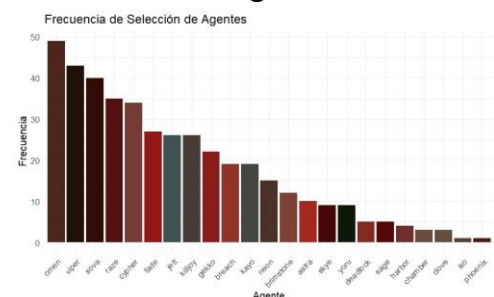
Datos:

- **Registros:** 46,152
- **Variables:** 21 (categóricas y numéricas)
  - Categóricas: Incluyen variables como **Tournament, Stage, Match Type, Player, Team, Map, Agents**.
  - Numéricas: Incluyen estadísticas como **Rating, Average Combat Score, Kills, Deaths, Assists**, entre otras.

### 2. Análisis Descriptivo Univariado

#### Categórico

##### 2.1 Selección de agentes

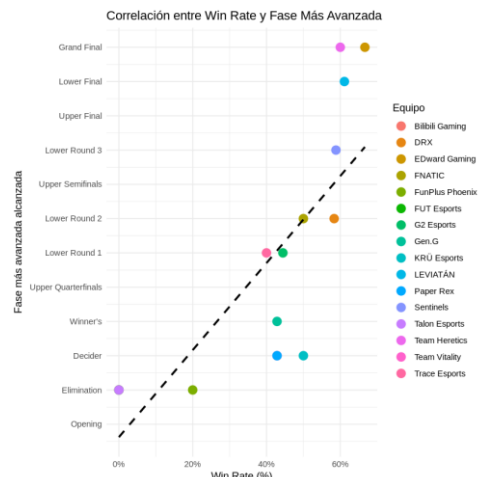


Para el torneo, se pudo ver que los agentes más escogidos fueron *Omen, Viper y Sova*. Demostrando

Para cada jugador, la página [vlr.gg](https://www.vlr.gg) tiene una métrica propia llamada Rating. Esta mide el rendimiento en general de los jugadores durante las partidas. Mediante esta gráfica podemos ver cómo los equipos que más avanzaron en la competencia están por encima de la media, y

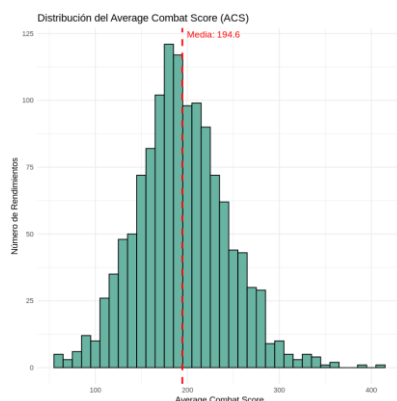
tienen una menor varianza de rating entre sus jugadores.

### 3.2 Correlación Taza de victoria (WR%) con fase más avanzada



La tasa de victoria, calculada como mapas ganados sobre el total jugado, muestra una correlación positiva con el avance en la competición: los equipos que llegaron más lejos (como EDward Gaming, LEVIATÁN y Team Heretics) presentan mayores porcentajes de victoria.

### 3.3 Distribución de ACS (Average Combat Score)

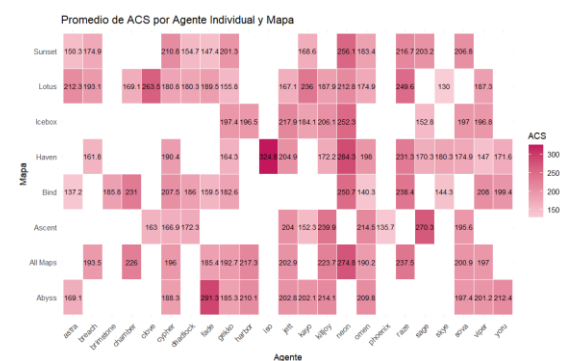


En Valorant, el ACS (Average Combat Score) es un indicador clave del rendimiento individual de

un jugador en una partida. Podemos ver que el ACS se distribuye de manera normal porque, al promediar el rendimiento de un jugador a lo largo de varias rondas, el Teorema Central del Límite asegura que estos promedios tienden a seguir una distribución normal.

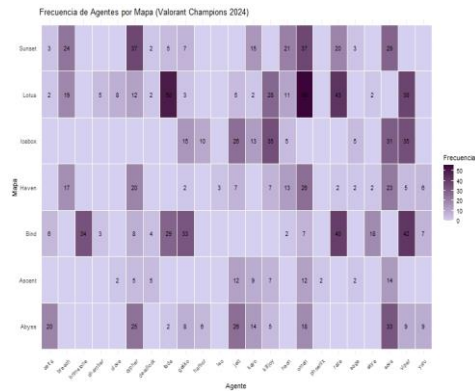
#### 4. Análisis Descriptivo Bivariado Categórico

#### 4.1 Promedio de ACS por Agente vs Mapa



Este gráfico de calor compara el ACS promedio por agente en cada mapa jugado durante el torneo Valorant Champions 2024. A partir de este análisis, identificamos los agentes que destacaron como MVPs (jugadores con el mejor ACS) en cada mapa

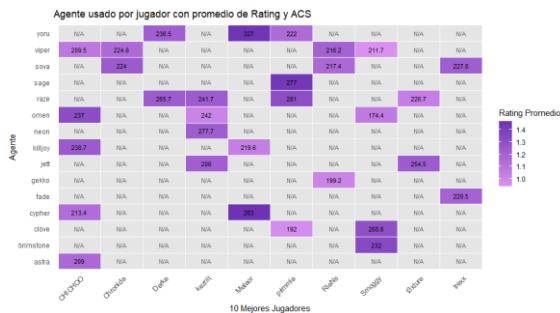
#### 4.2 Frecuencia de agente por Mapa vs Agente



Para cada mapa, ciertos personajes son más beneficiosos que otros estratégicamente. El siguiente mapa de calor muestra cuántas veces se escogió cierto agente para cada mapa. Las casillas sin número indican que el agente nunca se jugó en ese mapa durante el torneo.

## 5. Análisis Descriptivo Bivariado Numérico

### 5.1 Rating de Jugador vs Agente



Los jugadores suelen utilizar distintos agentes, lo que puede influir en su rendimiento. La gráfica presenta a los 10 mejores jugadores del torneo, mostrando su rating promedio según el agente usado (representado con un gradiente) y el ACS en cada casilla. Las casillas con "N/A" indican agentes no utilizados. Esta

visualización permite analizar si un jugador mantiene un alto desempeño siendo versátil o si destaca más al especializarse en uno o pocos agentes.

### 5.2 Equipo vs Tipo de Victoria en Ronda



Existen diversas formas de ganar una ronda en el juego, como eliminar al equipo rival, detonar o desactivar la spike, o dejar que el tiempo se agote sin que se plante. La gráfica muestra un mapa de calor con la cantidad de victorias por tipo para cada equipo. Los colores indican la frecuencia relativa y los números el conteo exacto. Las casillas sin número representan cero victorias en ese tipo. Esta visualización permite identificar las estrategias predominantes de cada equipo.

## 6. Pruebas de Hipótesis

### 6.1 ¿La proporción de picks de Omen es más que el de Viper?

Hipótesis Alternativa: La proporción de escoger a Omen es significativamente Mayor o menor a la de Viper.

Hipótesis Nula: La proporción de escoger a Omen es significativamente igual a la de Viper.

Test usado: Normal Estándar.

Resultados:

```
2-sample test for equality of proportions without continuity correction
data:  c(x1, x2) out of c(n1, n2)
X-squared = 1.12, df = 1, p-value = 0.2899
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.07238163  0.24382020
sample estimates:
 prop 1    prop 2 
0.6857143 0.6000000
```

Conclusión:

El valor  $p = 0.2899 > 0.05$ , por lo que no hay evidencia estadísticamente significativa para decir que Omen se selecciona más (o menos) que Viper.

## 6.2 Media de HS% de EDward Gaming es menor que la de Team Heretics

Hipótesis nula: La media de porcentaje de tiro a la cabeza de Team Heretics es menor o igual que la de EDward Gaming.

Hipótesis alternativa: La media de porcentaje de tiro a la cabeza de Team Heretics es mayor que la de EDward Gaming.

Resultados:

```
Welch Two Sample t-test
data:  eg_hs and th_hs
t = -4.9925, df = 224.74, p-value = 5.974e-07
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
 -inf -0.03775923
sample estimates:
mean of x mean of y 
0.2913333 0.3477600
```

Conclusión:

Dado que el valor  $p$  de la prueba  $t$  (Welch Two Sample  $t$ -test) fue muy

cercano a 0, rechazamos la hipótesis nula. Por lo tanto, hay evidencia estadística suficiente para afirmar que el porcentaje de tiros a la cabeza (HS%) de Team Heretics es significativamente mayor que el de EDward Gaming en el torneo Valorant Champions 2024.

## 7. Análisis de Normalidad

A las siguientes variables cuantitativas continuas les hicimos la prueba de Shapiro-Wilk para comprobar si estas se distribuyen de manera normal. Estos son los resultados

### 7.1 Variable: Rating

- $p$ -valor original: 0 No normal
- $p$ -valor sin outliers: 0.0197 No normal
- $p$ -valor  $\log(x+1)$ :  $1e-04$  No normal

### 7.2 Variable: Average.Combat.Score

- $p$ -valor original: 0.0029 No normal
- $p$ -valor sin outliers: 0.2033 Normal

### 7.3 Variable: Kill..Assist..Trade..Survive..

- $p$ -valor original: 0.0319 No normal
- $p$ -valor sin outliers: 0.0043 No normal
- $p$ -valor  $\log(x+1)$ : 0 No normal

### 7.4 Variable: Headshot..

- p-valor original: 0.0028 No normal
- p-valor sin outliers: 0.1423 Normal

#### 7.5 Variable: Average.Damage.Per.Round

- p-valor original: 0 No normal
- p-valor sin outliers: 0.0018 No normal
- p-valor log(x+1): 0 No normal

#### 7.6 Variable: Loadout.Value

- p-valor original: 0 No normal
- p-valor sin outliers: 0 No normal
- p-valor log(x+1): 0 No normal

#### 7.7 Variable: Remaining.Credits

- p-valor original: 0 No normal
- p-valor sin outliers: 0 No normal
- p-valor log(x+1): 0 No normal

#### 7.8 Variable: DurationSeconds

- p-valor original: 3e-04 No normal
- p-valor sin outliers: 0.017 No normal
- p-valor log(x+1): 0.0294 No normal

Dado que nuestros p\_valores para todas nuestras variables continuas en nuestra base de datos son menores que nuestro nivel de significancia ( $\alpha = 0.05$ ), podemos concluir que el Rating se distribuye normal.

### 8. Test de hipótesis diferencia de medias por categoría binaria

Para esta prueba de hipótesis, vamos a revisar si el porcentaje de tiros a la cabeza (HS%) de EDward Gaming (EG), el equipo ganador del torneo, es mayor que el de todos los otros equipos de la competencia.

#### a) Prueba de Cola Mayor

Para probar si el HS% de EG hicimos un t Test de cola superior, del cual obtuvimos los siguientes resultados:

```
> result_greater
      Welch Two Sample t-test

data:  df_headshot_EG$HS and df_headshot_others$HS
t = 0.55098, df = 145.29, p-value = 0.2912
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 -0.009761213      Inf
sample estimates:
mean of x mean of y
0.2913333 0.2864636
```

Donde la columna X son las entradas de HS% para EG y la columna Y la de todos los demás equipos. De esta prueba obtuvimos el p\_valor = 0.2912. Lo cual no rechaza la hipótesis nula de que el EG% es menor o igual que el de todos los equipos. De acuerdo con esto decidimos hacer la misma prueba verificando otra cola:

#### b) Prueba de dos colas

Dado que de la anterior prueba no rechazó que  $\overline{HS\%}_{EG} \leq \overline{HS\%}_{Otros}$ , decidimos hacer la prueba a dos colas de hipótesis nula simple tal que las medias son iguales. Para eso obtuvimos los siguientes resultados.

```
> result_two_sided

Welch Two Sample t-test

data: df_headshot_EG$HS and df_headshot_others$HS
t = 0.55098, df = 145.29, p-value = 0.5825
alternative hypothesis: true difference in means is not equal
95 percent confidence interval:
 -0.01259846  0.02233797
sample estimates:
mean of x mean of y
0.2913333 0.2864636
```

Para la cual obtuvimos un p\_valor = 0.5825, la cual es aproximadamente el doble de prueba de una cola. Lo anterior es posible porque podemos concluir que no hay evidencia estadística suficiente para rechazar que el porcentaje de tiro a la cabeza de EG y los otros equipos sean iguales.

### c) Prueba de homogeneidad de varianzas.

```
F test to compare two variances

data: df_headshot_EG$HS and df_headshot_others$HS
F = 0.75286, num df = 104, denom df = 754, p-value = 0.06994
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.5718561 1.0238550
sample estimates:
ratio of variances
0.7528633
```

Dado que el p-valor = 0.06994 > 0.05, no se rechaza la hipótesis nula al 5% de significancia. Esto significa que:

No hay evidencia estadística suficiente para afirmar que las varianzas del porcentaje de tiros a la cabeza (HS%) entre EDward Gaming y los demás equipos sean

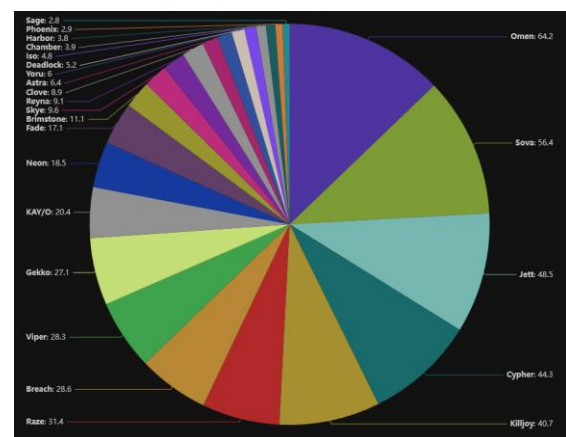
diferentes. Es decir, se puede asumir homogeneidad de varianzas entre los grupos.

**Conclusión:** Si bien EG fue el equipo ganador del torneo, estadísticamente podemos inferir que, aunque en los datos EDward Gaming tiene un HS% promedio ligeramente mayor a los otros equipos, este no fue un factor que diferenciara en gran medida a EDward Gaming con los demás equipos.

## 9. Bondad de ajuste

### 9.1 Cualitativa

Para nuestra variable cualitativa, decidimos comparar la frecuencia de selección de agentes del torneo con la selección de agentes del juego normal en el mismo periodo de tiempo (Parche 9.02, agosto 2024). Es decir, queremos ver si los profesionales seleccionan con la misma frecuencia los mismos agentes que la gente común y corriente. La distribución esperada se tomó de [https://liquipedia.net/valorant/Patch\\_9.02/Statistics](https://liquipedia.net/valorant/Patch_9.02/Statistics) y se ve de la siguiente forma:





Luego, contamos por cada jugador cuántas veces se han seleccionado los mismos agentes y utilizando la prueba de Xi cuadrado obtuvimos los siguientes resultados:

```
> chisq.test(x = observed_counts, p = expected_prob, rescale.p = TRUE)

Chi-squared test for given probabilities

data:  observed_counts
X-squared = 109.06, df = 22, p-value = 1.62e-13
```

Dado que existían en ese tiempo 23 agentes, los grados de libertad fueron 22. Para así obtener un p\_valor = 1.62e-13, lo cual es muchísimo más pequeño que nuestro nivel de significancia. De este modo concluimos que hay evidencia suficiente para decir que los profesionales escogen con distinta frecuencia sus personajes que la gente del común.

II. Cuantitativa Discreta:

Para nuestra variable escalar discreta, decidimos analizar si el número de kills por equipo en cada mapa del torneo Valorant Champions 2024 seguía una distribución Poisson, considerando como exposición el número de rondas jugadas por cada equipo.

Se estimó el parámetro  $\lambda$  como la tasa promedio de kills por ronda, y se calcularon las frecuencias esperadas para cada equipo, agrupando luego en clases de kills para cumplir con los requisitos de la prueba.

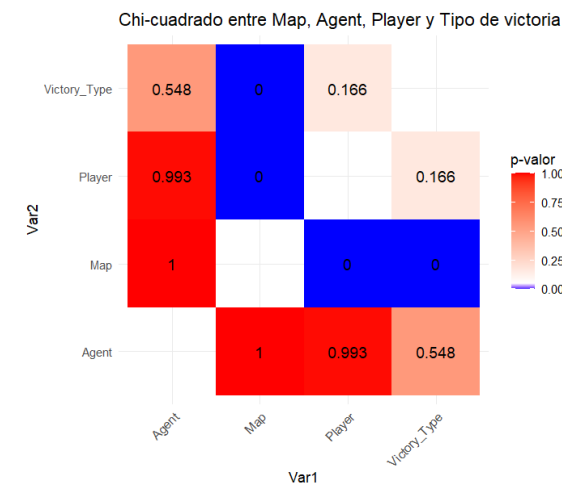
Aplicamos la prueba de  $\chi^2$  (chi-cuadrado) para comparar las frecuencias observadas y esperadas bajo dicho modelo. Se obtuvieron los siguientes resultados:

```
Chi-squared test for given probabilities

data:  observed
X-squared = 59.025, df = 3, p-value = 9.495e-13
```

Dado que el p-valor es muchísimo más pequeño que nuestro nivel de significancia  $\alpha=0.05$ , concluimos que existe evidencia suficiente para rechazar la hipótesis de que las kills por equipo siguen una distribución de Poisson ajustada por rondas jugadas. Esto sugiere que otros factores (como estrategia, sinergia de equipo, diferencias de nivel, etc.) influyen significativamente en la generación de eliminaciones.

10. Análisis de independencia:



Del heatmap con p-valores podemos concluir:

Par de variables	p-valor	¿Rechaza $H_0$ (independencia)?
Map vs Agent	0.000	Sí
Map vs Player	0.000	Sí



<b>Map vs Victory_Type</b>	0.000	Sí
<b>Agent vs Player</b>	0.993	No
<b>Agent vs Victory_Type</b>	0.548	No
<b>Player vs Victory_Type</b>	0.166	No

Este análisis tiene como propósito evaluar si existen asociaciones significativas entre pares de variables, es decir, si la distribución conjunta difiere de lo esperado bajo la suposición de independencia estadística.

Se construyeron tablas de contingencia para cada par de variables y se aplicó la prueba de independencia  $\chi^2$  (chi-cuadrado), utilizando un nivel de significancia de  $\alpha = 0.05$ . Para cada combinación de variables, se calcularon los p-valores correspondientes. Los resultados se representaron mediante un heatmap, que permitió identificar visualmente los pares con mayor evidencia de dependencia.

Posteriormente, para aquellos pares de variables en los que se rechazó la hipótesis nula de independencia, se aplicó un análisis post-hoc basado en los residuos estandarizados, con el fin de identificar qué combinaciones específicas contribuyen más a la asociación.

Para los pares significativos, se calcularon los residuos estandarizados:

Map vs Player:

- Se detectaron combinaciones específicas de jugadores que aparecen con más (o menos) frecuencia de lo esperado en ciertos mapas.
- Esto sugiere que algunos jugadores se desempeñan de forma sistemáticamente distinta dependiendo del mapa, lo cual puede relacionarse con su especialización o estrategias de equipo.

Map vs Victory\_Type:

- Se identificaron tipos de victoria más comunes en ciertos mapas, lo que podría relacionarse con las características estructurales del mapa o la economía de juego (ej. mapas que favorecen ataques rápidos o defensas más cerradas).

El análisis muestra evidencia sólida para rechazar la hipótesis de independencia entre el mapa y las variables agente, jugador y tipo de victoria, lo que indica que:

- La selección de agentes,
- El rendimiento individual,
- Y la forma en la que se define la victoria

están significativamente condicionadas por el mapa en el que se disputa la partida.

Este hallazgo refuerza la idea de que los mapas en Valorant no son escenarios neutrales, sino que influyen activamente en las dinámicas del juego. Se recomienda considerar estos factores en análisis de

rendimiento, selección de estrategias, y  
entrenamiento personalizado.

**Anexo:** [Repositorio Github](#)

## DICCIONARIO:

- **Tournament ID** – ID del torneo
- **Stage ID** – ID de la etapa o fase del torneo
- **Match Type ID** – ID del tipo de enfrentamiento (ej. BO1, BO3, BO5)
- **Match ID** – ID del enfrentamiento
- **Game ID** – ID del mapa jugado en ese enfrentamiento
- **Pick Rate** – Cuántas veces fue seleccionado un agente
- **Total Maps Played** – Cuántas veces se ha jugado ese mapa
- **Attacker Side Win Percentage** – Porcentaje de rondas ganadas por el lado atacante en ese mapa
- **Defender Side Win Percentage** – Porcentaje de rondas ganadas por el lado defensor en ese mapa
- **Total Wins By Map** – Cuántas veces ganó el equipo usando ese agente en ese mapa
- **Total Loss By Map** – Cuántas veces perdió el equipo usando ese agente en ese mapa
- **Total Maps Played** – Número total de veces que el equipo jugó en ese mapa
- **Loadout Value** – Valor total del equipo comprado (armas, habilidades, escudos) por ronda
- **Remaining Credits** – Créditos restantes del equipo después de comprar
- **Type** – Tipo de ronda económica (eco, semi-eco, semi-buy, full buy)
- **Pistol Round** – Si el equipo ganó la ronda de pistolas
- **Eco (won)** – Veces que el equipo ganó gastando entre 0-5k créditos
- **\$ (Semi-eco)** – Veces que el equipo ganó gastando entre 5-10k créditos
- **\$\$ (Semi-buy)** – Veces que el equipo ganó gastando entre 10-20k créditos
- **\$\$\$ (Full buy)** – Veces que el equipo ganó gastando más de 20k créditos
- **Initiated** – Veces que el equipo inició ese tipo de ronda económica
- **Won** – Veces que el equipo ganó ese tipo de ronda económica
- **Eliminator Team** – Equipo del jugador que eliminó
- **Eliminator** – Jugador que realizó la eliminación
- **Eliminator Agent** – Agente utilizado por el jugador que eliminó
- **Eliminated Team** – Equipo del jugador eliminado
- **Eliminated** – Jugador que fue eliminado
- **Eliminated Agent** – Agente utilizado por el jugador eliminado
- **Kills** – Número de veces que un jugador eliminó a otro con su agente
- **2k / 3k / 4k / 5k** – Número de veces que el jugador consiguió 2, 3, 4 o 5 eliminaciones en una sola ronda
- **1v1 / 1v2 / 1v3 / 1v4** – Número de veces que el jugador ganó un clutch estando solo contra 1, 2, 3 o 4 oponentes
- **Econ** – Eficiencia económica del jugador (daño por cada 1000 créditos gastados)

- **Spike Plants** – Número de veces que el jugador plantó la spike
- **Spike Defuse** – Número de veces que el jugador desactivó la spike
- **Team Score** – Puntuación total del equipo en la partida
- **Team Attacker Score / Defender Score / Overtime Score** – Puntuación del equipo como atacante, defensor o en overtime
- **Duration** – Duración del enfrentamiento (horas, minutos, segundos)
- **Rating (R)** – Calificación del jugador considerando múltiples estadísticas
- **Average Combat Score (ACS)** – Puntuación de combate promedio (daño, kills, multikills, asistencias útiles)
- **Kills / Deaths / Assists** – Eliminaciones, muertes y asistencias
- **Kills - Deaths (FD)** – Diferencia entre kills y muertes
- **KAST** – % de rondas en las que el jugador consiguió un kill, asistencia, trade o sobrevivió
- **ADR** – Daño promedio por ronda  
**KPR / APR / FKPR / FDPR** – Kills, asistencias, primeras kills y primeras muertes promedio por ronda
- **Headshot %** – Porcentaje de disparos que fueron headshots
- **Clutch Success %** – % de clutches ganados sobre los jugados
- **Clutches (won/played)** – Ratio de clutches ganados respecto a los jugados
- **Maximum Kills in a Single Map** – Máximo número de kills en un solo mapa
- **First Kills / First Deaths / FKD** – Primeras eliminaciones, primeras muertes y su diferencia
- **Elimination** – Veces que el equipo ganó eliminando al equipo rival
- **Detonated** – Veces que el equipo ganó por detonación exitosa de la spike
- **Defuse** – Veces que el equipo ganó desactivando la spike
- **Time Expiry (No Plant)** – Veces que el equipo ganó porque el enemigo no plantó a tiempo
- **Eliminated** – Veces que el equipo perdió al ser eliminado
- **Defuse Failed** – Veces que el equipo perdió por no desactivar la spike
- **Detonation Denied** – Veces que el equipo perdió porque la spike fue desactivada
- **Time Expiry (Failed to Plant)** – Veces que el equipo perdió por no plantar a tiempo