

FASE 1: Entendiendo el negocio

Contexto:

La empresa **SotoBurguer S.A.**, una compañía especializada en el análisis de datos para la industria de los videojuegos, ha sido contratada por Epic Games para ayudar a mejorar la experiencia de juego en Fortnite. El objetivo principal es utilizar los datos de rendimiento de los jugadores en modo solitario para desarrollar un modelo predictivo que pueda identificar a los jugadores con un alto potencial de rendimiento y aquellos que podrían estar en riesgo de abandonar el juego debido a un bajo desempeño.

Problemática:

Epic Games ha observado que la retención de jugadores en Fortnite es crucial para el éxito a largo plazo del juego. Sin embargo, algunos jugadores, especialmente aquellos con un bajo desempeño, tienden a abandonar el juego después de experimentar frustración por no lograr buenos resultados. La empresa quiere crear un sistema que pueda predecir el rendimiento futuro de los jugadores basándose en su historial de juegos en modo solitario.

Objetivo:

Desarrollar un modelo de predicción basado en las estadísticas actuales de un jugador (como el Solo kd, Solo winRatio, Solo matches y otras métricas), pueda predecir:

1. **El nivel de desempeño futuro** del jugador en términos de victorias y ratio de kills/deaths (KD).
2. **El riesgo de abandono** del jugador, definido como una disminución significativa en el número de partidas jugadas o un cambio abrupto en su rendimiento.
3. **¿Qué factores influyen en la puntuación (Solo score) de un jugador?**
4. **¿Existe una relación significativa entre el tiempo jugado (Solo minutesPlayed) y el ratio de victorias (Solo winRatio)?**

En este conjunto de datos hay 5 modos de juego (Solo, Duo, Trios. Squads y ltm) cada uno tiene:

Puntuación: Puntuación total. **top1:** Victorias totales. **kd:** Ratio de muertes. **WinRatio:** Ratio de victorias. **matches:** Número de partidos jugados. **kills:** Número de eliminaciones realizadas. **minutesPlayed:** Número total de minutos jugados.

```
# Se realiza las importaciones de las Librerías
import pandas as pd
import numpy as np
import seaborn as sb
import matplotlib.pyplot as plt
```

```

from sklearn.impute import KNNImputer
from sklearn.preprocessing import StandardScaler
from sklearn.preprocessing import MinMaxScaler
from sklearn.preprocessing import RobustScaler

#cargar dataset
df_completo = pd.read_csv('Fortnite_players_stats.csv', sep=",")

print(df_completo.columns)

Index(['Player', 'Solo score', 'Solo top1', 'Solo kd', 'Solo
winRatio',
      'Solo matches', 'Solo kills', 'Solo minutesPlayed', 'Duos
score',
      'Duos top1', 'Duos kd', 'Duos winRatio', 'Duos matches', 'Duos
kills',
      'Duos minutesPlayed', 'Trios score', 'Trios top1', 'Trios kd',
      'Trios winRatio', 'Trios matches', 'Trios kills', 'Trios
minutesPlayed',
      'Squads score', 'Squads top1', 'Squads kd', 'Squads winRatio',
      'Squads matches', 'Squads kills', 'Squads minutesPlayed', 'LTM
score',
      'LTM top1', 'LTM top3', 'LTM kd', 'LTM winRatio', 'LTM
matches',
      'LTM kills', 'LTM minutesPlayed'],
      dtype='object')

df = df_completo[['Player', 'Solo score', 'Solo top1', 'Solo kd',
                  'Solo winRatio', 'Solo matches', 'Solo
kills', 'Solo minutesPlayed']]

```

FASE 2: Data Understanding

Verificar columnas y tipos de datos

```

columnas = df.columns
for columna in columnas:
    if df[columna].dtype == int:
        print(f"Columna {columna} ----- Tipo de variable: Cuantitativa
discreta")
    elif df[columna].dtype == float:
        print(f"Columna {columna} ----- Tipo de variable: Cuantitativa
continua")
    elif df[columna].dtype == object:
        print(f"Columna {columna} ----- Tipo de variable: Cualitativa
nominal")

```

```

elif df[columna].dtype == bool:
    print(f"Columna {columna} ----- Tipo de variable: Cualitativa ordinal")

```

```

Columna Player ----- Tipo de variable: Cualitativa nominal
Columna Solo score ----- Tipo de variable: Cuantitativa discreta
Columna Solo top1 ----- Tipo de variable: Cuantitativa discreta
Columna Solo kd ----- Tipo de variable: Cuantitativa continua
Columna Solo winRatio ----- Tipo de variable: Cuantitativa continua
Columna Solo matches ----- Tipo de variable: Cuantitativa discreta
Columna Solo kills ----- Tipo de variable: Cuantitativa discreta
Columna Solo minutesPlayed ----- Tipo de variable: Cuantitativa discreta

```

```

# Se filtran las 10 primeras Filas
df.head(10)

```

```

{"column_count":8,"columns":
[{"dtype":"object","name":"Player","stats":{"categories":
[{"count":1,"name":"Prospering"}, {"count":1,"name":"BH nixxxay"},
{"count":8,"name":"8 others"}], "nan_count":0,"unique_count":10}},
{"dtype":"int64","name":"Solo score","stats":{"histogram":
[{"bin_end":691496.4,"bin_start":15872,"count":3},
{"bin_end":1367120.8,"bin_start":691496.4,"count":2},
{"bin_end":2042745.2000000002,"bin_start":1367120.8,"count":0},
{"bin_end":2718369.6,"bin_start":2042745.2000000002,"count":2},
{"bin_end":3393994,"bin_start":2718369.6,"count":1},
{"bin_end":4069618.4000000004,"bin_start":3393994,"count":0},
{"bin_end":4745242.8,"bin_start":4069618.4000000004,"count":1},
{"bin_end":5420867.2,"bin_start":4745242.8,"count":0},
{"bin_end":6096491.6000000001,"bin_start":5420867.2,"count":0},
{"bin_end":6772116,"bin_start":6096491.6000000001,"count":1}], "max":"67
72116", "min":"15872", "nan_count":0,"unique_count":10}},
{"dtype":"int64","name":"Solo top1","stats":{"histogram":
[{"bin_end":653.4,"bin_start":6,"count":3},
{"bin_end":1300.8,"bin_start":653.4,"count":0},
{"bin_end":1948.1999999999998,"bin_start":1300.8,"count":5},
{"bin_end":2595.6,"bin_start":1948.1999999999998,"count":0},
{"bin_end":3243,"bin_start":2595.6,"count":0},
{"bin_end":3890.3999999999996,"bin_start":3243,"count":0},
{"bin_end":4537.8,"bin_start":3890.3999999999996,"count":0},
{"bin_end":5185.2,"bin_start":4537.8,"count":1},
{"bin_end":5832.599999999999,"bin_start":5185.2,"count":0},
{"bin_end":6480,"bin_start":5832.599999999999,"count":1}], "max":"6480"
, "min":"6", "nan_count":0,"unique_count":10}},
{"dtype":"float64","name":"Solo kd","stats":{"histogram":
[{"bin_end":1.383,"bin_start":0.47,"count":3},
{"bin_end":2.296,"bin_start":1.383,"count":1},
{"bin_end":3.2089999999999996,"bin_start":2.296,"count":0},
{"bin_end":4.122,"bin_start":3.2089999999999996,"count":2},

```

```

{"bin_end":5.034999999999999,"bin_start":4.122,"count":1},
{"bin_end":5.947999999999995,"bin_start":5.034999999999999,"count":0}
,
{"bin_end":6.860999999999999,"bin_start":5.947999999999995,"count":1}
,
{"bin_end":7.773999999999999,"bin_start":6.860999999999999,"count":0},
{"bin_end":8.687,"bin_start":7.773999999999999,"count":0},
{"bin_end":9.6,"bin_start":8.687,"count":2}], "max": "9.6", "min": "0.47",
"nan_count": 0, "unique_count": 10}}, {"dtype": "float64", "name": "Solo
winRatio", "stats": {"histogram":
[{"bin_end":3.85,"bin_start":0.5,"count":3},
{"bin_end":7.2,"bin_start":3.85,"count":0},
{"bin_end":10.55,"bin_start":7.2,"count":2},
{"bin_end":13.9,"bin_start":10.55,"count":0},
{"bin_end":17.25,"bin_start":13.9,"count":0},
{"bin_end":20.6,"bin_start":17.25,"count":1},
{"bin_end":23.95,"bin_start":20.6,"count":0},
{"bin_end":27.3,"bin_start":23.95,"count":0},
{"bin_end":30.650000000000002,"bin_start":27.3,"count":2},
{"bin_end":34,"bin_start":30.650000000000002,"count":2}], "max": "34.0",
"min": "0.5", "nan_count": 0, "unique_count": 10}},
{"dtype": "int64", "name": "Solo matches", "stats": {"histogram":
[{"bin_end":2372.5,"bin_start":429,"count":2},
{"bin_end":4316,"bin_start":2372.5,"count":1},
{"bin_end":6259.5,"bin_start":4316,"count":2},
{"bin_end":8203,"bin_start":6259.5,"count":0},
{"bin_end":10146.5,"bin_start":8203,"count":0},
{"bin_end":12090,"bin_start":10146.5,"count":1},
{"bin_end":14033.5,"bin_start":12090,"count":2},
{"bin_end":15977,"bin_start":14033.5,"count":0},
{"bin_end":17920.5,"bin_start":15977,"count":0},
{"bin_end":19864,"bin_start":17920.5,"count":2}], "max": "19864", "min": "
429", "nan_count": 0, "unique_count": 10}}, {"dtype": "int64", "name": "Solo
kills", "stats": {"histogram":
[{"bin_end":8728.1,"bin_start":200,"count":3},
{"bin_end":17256.2,"bin_start":8728.1,"count":0},
{"bin_end":25784.3000000000003,"bin_start":17256.2,"count":2},
{"bin_end":34312.4,"bin_start":25784.3000000000003,"count":0},
{"bin_end":42840.5,"bin_start":34312.4,"count":3},
{"bin_end":51368.6000000000006,"bin_start":42840.5,"count":0},
{"bin_end":59896.7000000000004,"bin_start":51368.6000000000006,"count":0
}, {"bin_end":68424.8,"bin_start":59896.7000000000004,"count":1},
{"bin_end":76952.9000000000001,"bin_start":68424.8,"count":0},
{"bin_end":85481,"bin_start":76952.9000000000001,"count":1}], "max": "8548
1", "min": "200", "nan_count": 0, "unique_count": 10}},
{"dtype": "int64", "name": "Solo minutesPlayed", "stats": {"histogram":
[{"bin_end":29048.6,"bin_start":739,"count":3},
{"bin_end":57358.2,"bin_start":29048.6,"count":2},
{"bin_end":85667.79999999999,"bin_start":57358.2,"count":2},

```

```
{
  "bin_end": 113977.4, "bin_start": 85667.79999999999, "count": 1},
  {"bin_end": 142287, "bin_start": 113977.4, "count": 1},
  {"bin_end": 170596.59999999998, "bin_start": 142287, "count": 0},
  {"bin_end": 198906.19999999998, "bin_start": 170596.59999999998, "count": 0},
  {"bin_end": 227215.8, "bin_start": 198906.19999999998, "count": 0},
  {"bin_end": 255525.4, "bin_start": 227215.8, "count": 0},
  {"bin_end": 283835, "bin_start": 255525.4, "count": 1}], "max": "283835", "min": "739", "nan_count": 0, "unique_count": 10}},
{"dtype": "int64", "name": "_deepnote_index_column"}], "row_count": 10, "rows": [
  {"Player": "Prospering", "Solo kd": 4.37, "Solo kills": 36328, "Solo matches": 10150, "Solo minutesPlayed": 81389, "Solo score": 2476763, "Solo top1": 1828, "Solo winRatio": 18, "_deepnote_index_column": 0},
  {"Player": "BH nixxay", "Solo kd": 8.71, "Solo kills": 35895, "Solo matches": 5817, "Solo minutesPlayed": 12732, "Solo score": 439562, "Solo top1": 1694, "Solo winRatio": 29.1, "_deepnote_index_column": 1},
  {"Player": "Ranger", "Solo kd": 9.6, "Solo kills": 85481, "Solo matches": 13488, "Solo minutesPlayed": 122171, "Solo score": 4519465, "Solo top1": 4582, "Solo winRatio": 34, "_deepnote_index_column": 2},
  {"Player": "Idk_Pi", "Solo kd": 0.84, "Solo kills": 3005, "Solo matches": 3687, "Solo minutesPlayed": 32453, "Solo score": 752869, "Solo top1": 121, "Solo winRatio": 3.3, "_deepnote_index_column": 3},
  {"Player": "Twitch.GryphonRB", "Solo kd": 6.32, "Solo kills": 19591, "Solo matches": 4429, "Solo minutesPlayed": 36245, "Solo score": 1136282, "Solo top1": 1327, "Solo winRatio": 30, "_deepnote_index_column": 4},
  {"Player": "Twitch Kayotica", "Solo kd": 3.23, "Solo kills": 39131, "Solo matches": 13438, "Solo minutesPlayed": 96777, "Solo score": 2919037, "Solo top1": 1310, "Solo winRatio": 9.7, "_deepnote_index_column": 5},
  {"Player": "FaZe Replays", "Solo kd": 3.84, "Solo kills": 66161, "Solo matches": 18670, "Solo minutesPlayed": 76258, "Solo score": 2389537, "Solo top1": 1454, "Solo winRatio": 7.8, "_deepnote_index_column": 6},
  {"Player": "NiteGamerYT 190k", "Solo kd": 0.47, "Solo kills": 200, "Solo matches": 429, "Solo minutesPlayed": 739, "Solo score": 15872, "Solo top1": 6, "Solo winRatio": 1.4, "_deepnote_index_column": 7},
  {"Player": "Boss >", "Solo kd": 1.39, "Solo kills": 18610, "Solo matches": 19864, "Solo minutesPlayed": 283835, "Solo score": 6772116, "Solo top1": 6480, "Solo winRatio": 32.6, "_deepnote_index_column": 8},
  {"Player": "CIUPEA 144.HZ", "Solo kd": 0.61, "Solo kills": 1174, "Solo matches": 1938, "Solo minutesPlayed": 2441, "Solo score": 54479, "Solo top1": 9, "Solo winRatio": 0.5, "_deepnote_index_column": 9}]]
```

```
#Se Filtra el tipo de dato de la columnas
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1435 entries, 0 to 1434
Data columns (total 8 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Player                1435 non-null   object
1   Solo score            1435 non-null   int64
```

```

2 Solo top1      1435 non-null int64
3 Solo kd        1435 non-null float64
4 Solo winRatio  1435 non-null float64
5 Solo matches   1435 non-null int64
6 Solo kills     1435 non-null int64
7 Solo minutesPlayed 1435 non-null int64
dtypes: float64(2), int64(5), object(1)
memory usage: 89.8+ KB

```

Verificación de datos nulos y posibles errores en DataFrame.

```

# Verificación de datos nulos en el dataset
df.isnull().sum()

Player      0
Solo score  0
Solo top1   0
Solo kd     0
Solo winRatio 0
Solo matches 0
Solo kills  0
Solo minutesPlayed 0
dtype: int64

```

De momento los datos a utilizar no arrojan datos nulos, pero existe la posibilidad que hayan datos negativos que los interprete como datos no nulos por lo que se verificara con el siguiente código.

```

# Verificar si hay datos negativos
(df[['Solo score', 'Solo top1', 'Solo kd', 'Solo winRatio', 'Solo matches', 'Solo kills', 'Solo minutesPlayed']] < 0).sum()

Solo score      0
Solo top1       0
Solo kd         0
Solo winRatio   0
Solo matches    0
Solo kills      0
Solo minutesPlayed 0
dtype: int64

```

Al no existir datos negativos y nulos, estamos listo para comenzar las medidas de posición.

Medidas de posición

```

#Resumen de Medidas de posición, para datos numéricos
print(df.describe())

```

	Solo score	Solo top1	Solo kd	Solo winRatio	Solo matches \
count	1.435000e+03	1435.000000	1435.000000	1435.000000	1435.000000
mean	1.088202e+06	328.909408	1.816300	4.853937	6911.060627
std	8.872639e+05	529.192568	1.156026	5.330249	5408.219523
min	4.453000e+03	0.000000	0.080000	0.000000	30.000000
25%	4.355170e+05	65.500000	1.050000	1.600000	2875.000000
50%	9.070010e+05	162.000000	1.480000	2.900000	5728.000000
75%	1.506555e+06	366.000000	2.200000	6.100000	9585.000000
max	6.772116e+06	6480.000000	9.600000	44.100000	36487.000000

	Solo kills	Solo minutesPlayed
count	1435.000000	1435.000000
mean	11781.843206	43517.135889
std	11417.340071	36314.877771
min	28.000000	210.000000
25%	3861.000000	16922.000000
50%	8452.000000	35762.000000
75%	15993.000000	59471.000000
max	89023.000000	283835.000000

Excluir columnas no numéricas

```
df_numeric = df.select_dtypes(include=[float, int])
```

Calcular la matriz de correlación solo con las columnas numéricas

```
correlation_matrix = df_numeric.corr()
```

Mostrar la matriz de correlación

```
print(correlation_matrix)
```

	Solo score	Solo top1	Solo kd	Solo winRatio	\
Solo score	1.000000	0.598385	0.203172	0.245719	
Solo top1	0.598385	1.000000	0.659677	0.768525	
Solo kd	0.203172	0.659677	1.000000	0.854501	
Solo winRatio	0.245719	0.768525	0.854501	1.000000	
Solo matches	0.860244	0.361533	0.036813	-0.022918	
Solo kills	0.715041	0.748277	0.613477	0.448650	
Solo minutesPlayed	0.979049	0.484908	0.067514	0.133830	

	Solo matches	Solo kills	Solo minutesPlayed
Solo score	0.860244	0.715041	0.979049
Solo top1	0.361533	0.748277	0.484908

Solo kd	0.036813	0.613477	0.067514
Solo winRatio	-0.022918	0.448650	0.133830
Solo matches	1.000000	0.705988	0.859965
Solo kills	0.705988	1.000000	0.601728
Solo minutesPlayed	0.859965	0.601728	1.000000

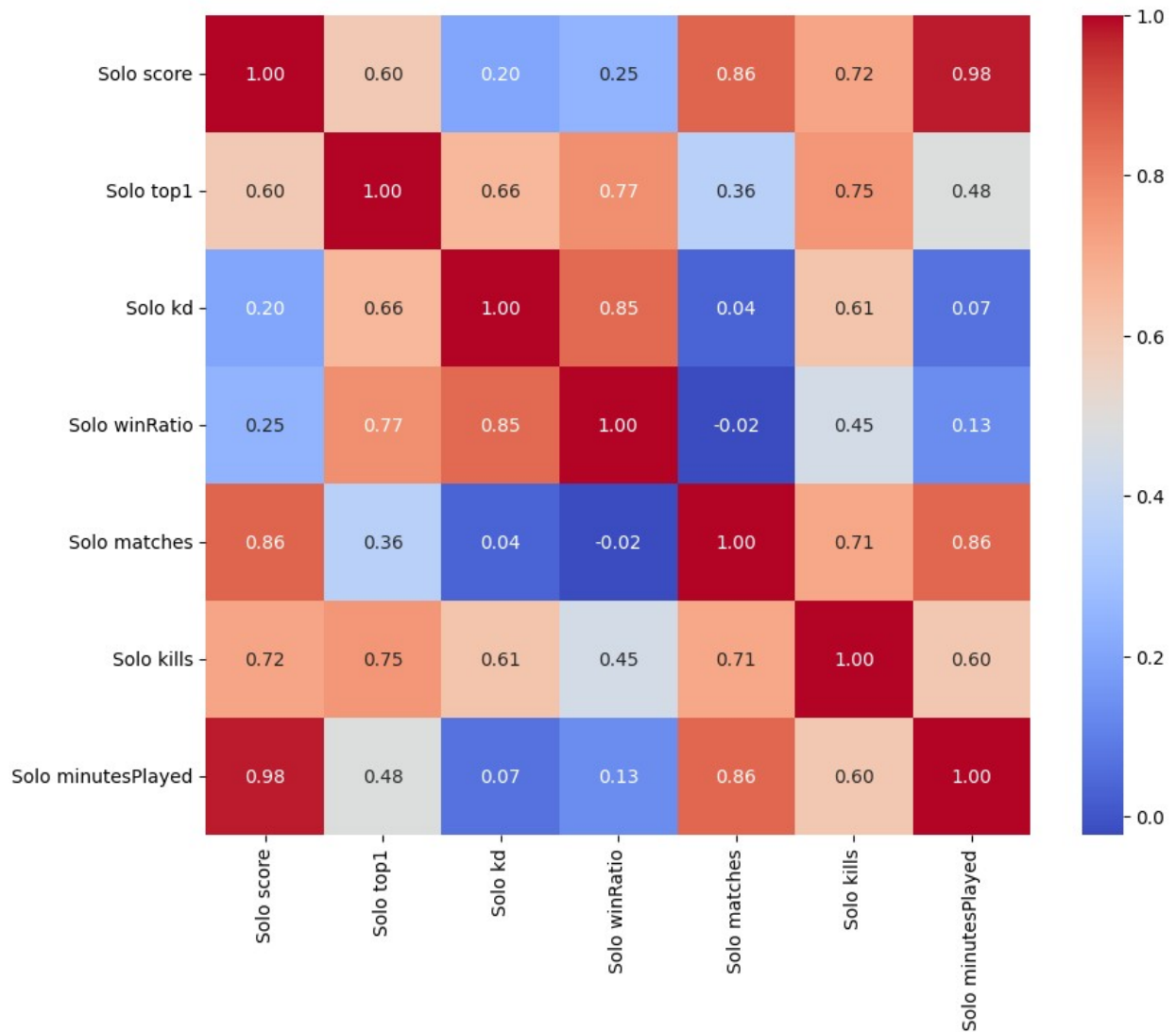
Matriz de correlación

Las variables Solo score, Solo matches y Solo minutesPlayed están fuertemente correlacionadas, podemos decir que el tiempo jugado y el número de partidas tienen un gran impacto en la puntuación total.

Solo winRatio está fuertemente correlacionado con la capacidad del jugador de estar en el Solo top1 y con su habilidad de combate (Solo kd).

No hay correlación significativa entre algunas variables, como Solo matches y Solo kd, lo que puede sugerir que jugar más partidas no mejora la relación de asesinatos/muertes de un jugador.

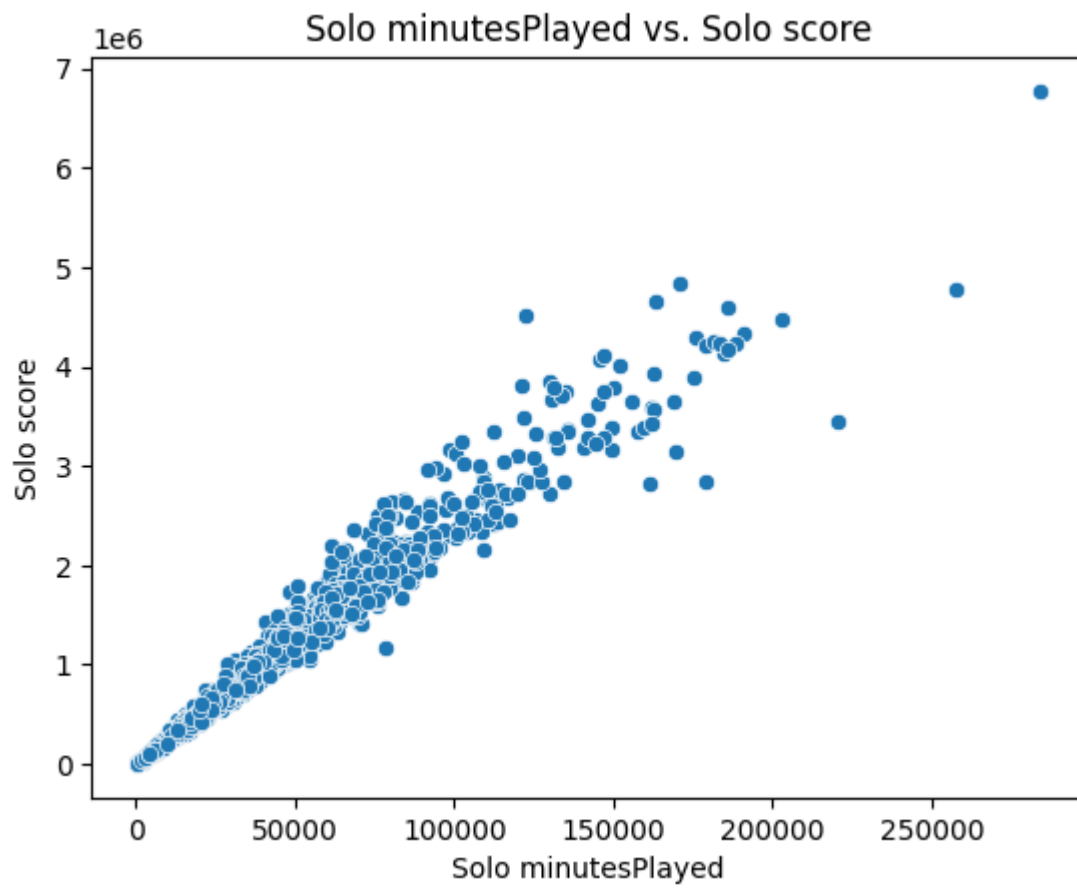
```
plt.figure(figsize=(10, 8))
sb.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt='.2f')
plt.show()
```

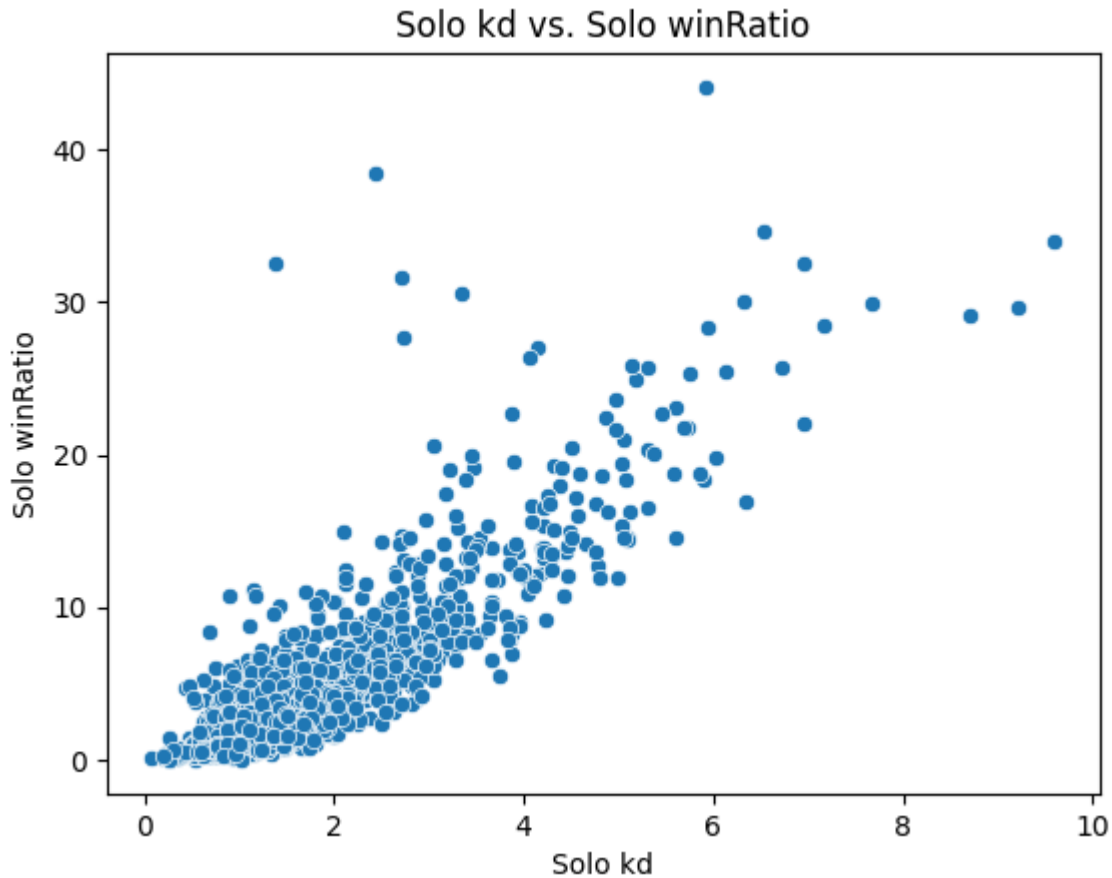



Scatter Plots (Diagramas de Dispersión):

```
sb.scatterplot(x='Solo minutesPlayed', y='Solo score', data=df)
plt.title('Solo minutesPlayed vs. Solo score')
plt.show()

sb.scatterplot(x='Solo kd', y='Solo winRatio', data=df)
plt.title('Solo kd vs. Solo winRatio')
plt.show()
```





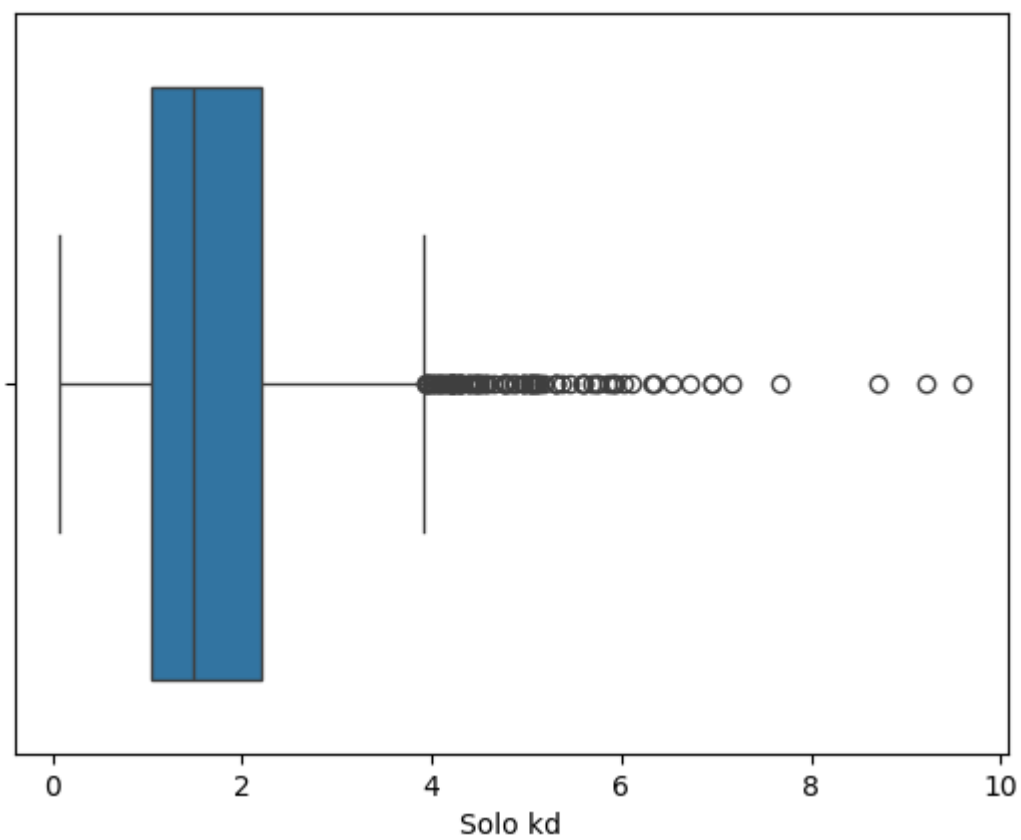
Solo minutesPlayed vs. Solo score: Dado que existe una fuerte correlación entre el tiempo jugado y el puntaje, un scatter plot puede mostrar cómo el puntaje de los jugadores aumenta con el tiempo jugado. Solo kd vs. Solo winRatio: Con una correlación alta entre el ratio KD y el winRatio, un scatter plot puede ayudar a visualizar cómo se relacionan estas dos métricas.

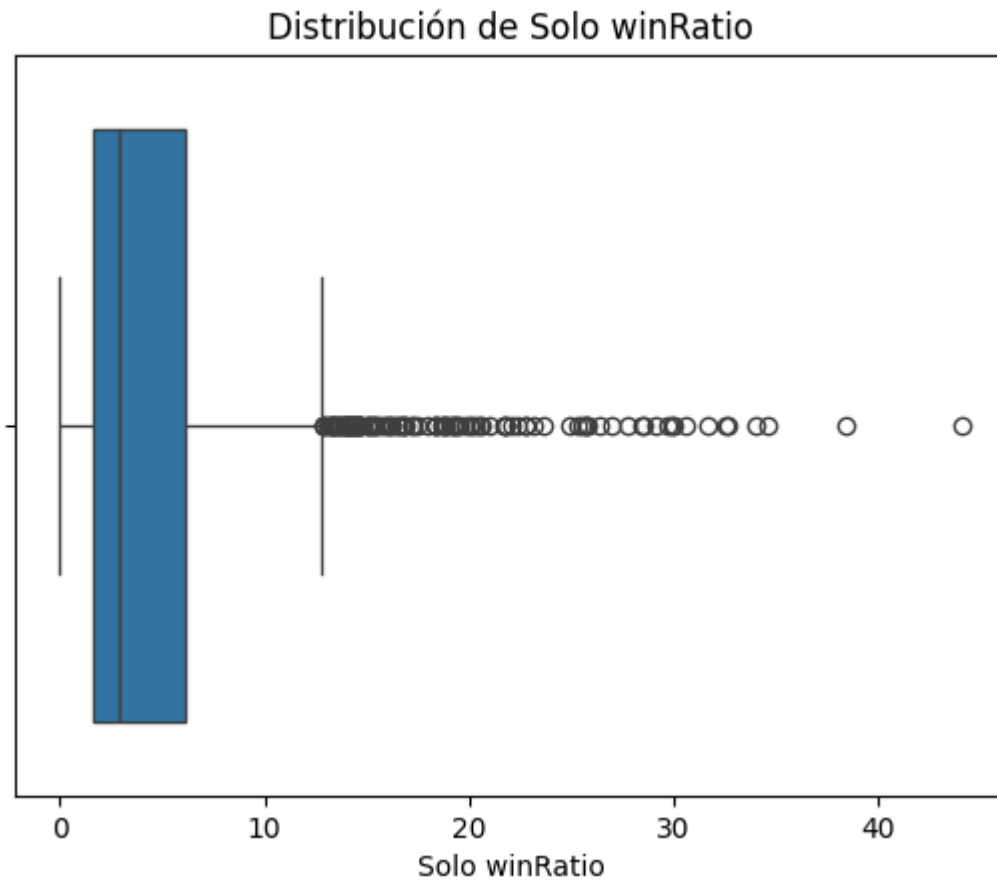
Box Plot

```
sb.boxplot(x='Solo kd', data=df)
plt.title('Distribución de Solo kd')
plt.show()

sb.boxplot(x='Solo winRatio', data=df)
plt.title('Distribución de Solo winRatio')
plt.show()
```

Distribución de Solo kd





Solo KD: La línea dentro de la caja representa la mediana del Solo kd, que parece estar alrededor de 1. Esto significa que el 50% de los jugadores tienen un ratio de eliminación-muerte menor o igual a 1.

Se observa un grupo considerable de outliers a la derecha del gráfico, es decir, jugadores que tienen un Solo kd significativamente mayor que el promedio. Algunos jugadores tienen valores de Solo kd que llegan hasta 10, pero son casos poco comunes.

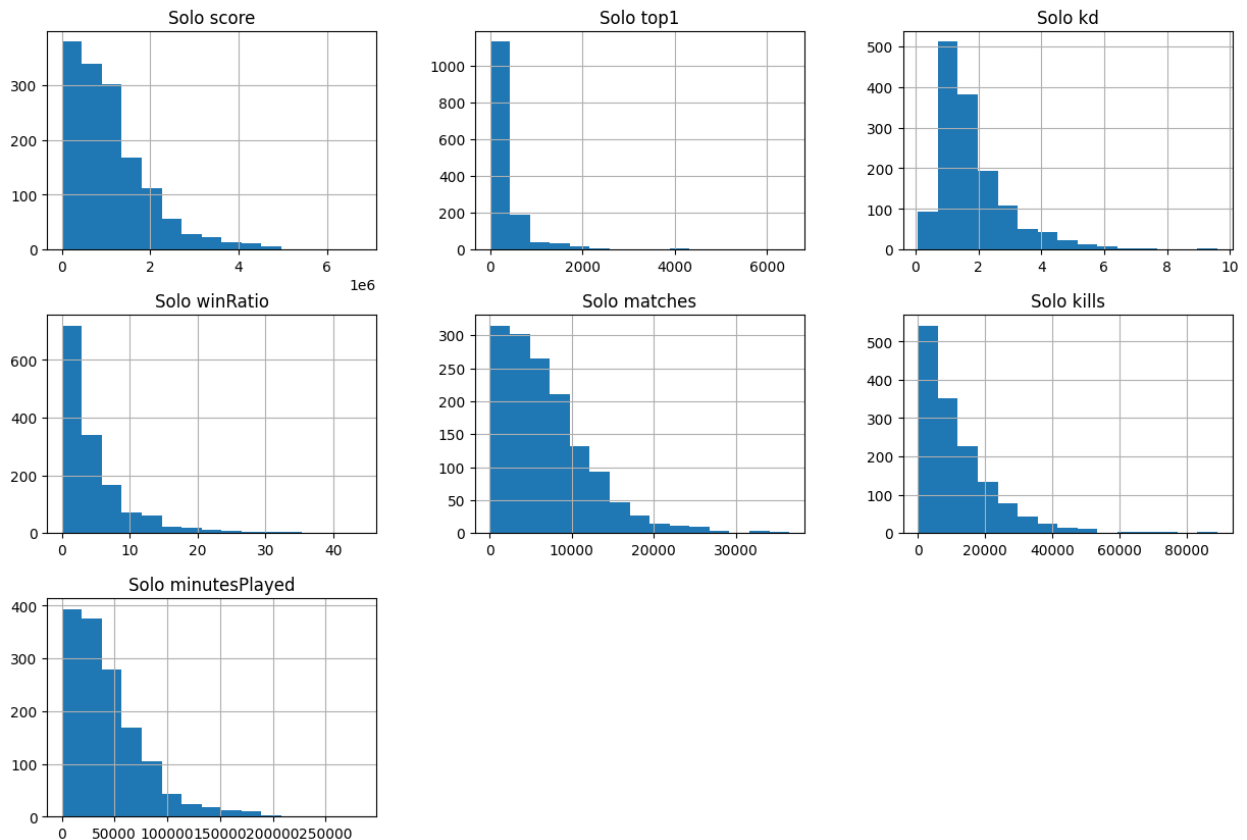
Solo winRatio: La línea en la caja sugiere que la mediana del win ratio está entre 5% y 10%, lo que indica que el 50% de los jugadores ganan menos del 10% de las partidas que juegan.

Nuevamente, se observan muchos outliers hacia la derecha, con algunos jugadores logrando ratios de victoria que superan el 40%. Estos jugadores son excepcionales en comparación con la mayoría.

Histogramas

```
df_numeric.hist(bins=15, figsize=(15, 10), layout=(3, 3))
plt.suptitle('Distribuciones de Variables Numéricas')
plt.show()
```

Distribuciones de Variables Numéricas



Solo score: -Hay una concentración de valores bajos y una larga cola hacia la derecha, lo que nos indicaría que la mayoría de los jugadores tienen una puntuación baja, mientras que muy pocos tienen puntuaciones extremadamente altas. Solo top1: -La mayoría de los jugadores tienen pocas o ninguna victoria, mientras que solo unos pocos tienen un gran número de victorias. Solo kd: -La mayoría de los jugadores tienen un kd bajo, con pocos jugadores teniendo un kd muy alto. -Tiene una distribución sesgada a la derecha, indicando que la mayoría de los jugadores no tienen una relación de asesinatos/muertes alta. Solo winRatio: -La mayoría de los jugadores tienen una proporción de victorias baja (menor a 10%), con solo unos pocos jugadores alcanzando porcentajes mayores. Solo matches: -Se tiene una distribución sesgada a la derecha, nos indica que es común tener pocos jugadores que juegan muchas partidas y que la mayoría han jugado un número relativamente bajo de partidas, mientras que unos pocos han jugado muchas más. Solo kills: -La mayoría de los jugadores tienen un número bajo de asesinatos, y pocos jugadores tienen un número muy alto. -Esta variable también muestra una distribución sesgada a la derecha. Solo minutesPlayed: -La mayoría de los jugadores tienen menos tiempo de juego, y hay algunos jugadores con una cantidad de tiempo de juego muy alta.

Identificar los jugadores con el mejor rendimiento en Modo Solitario

```
# Ordenar por rendimiento en el modo Solo
mejores_jugadores_solo = df.sort_values(by=['Solo score', 'Solo
```

```
winRatio', 'Solo kd'], ascending=False)

# Mostrar los jugadores con el mejor rendimiento en cada modo
print("Mejores jugadores en Solo:\n",
mejores_jugadores_solo[['Player', 'Solo score', 'Solo winRatio', 'Solo
kd']].head(10))
```

Mejores jugadores en Solo:

	Player	Solo score	Solo winRatio	Solo kd
8	Bossy	6772116	32.6	1.39
11	DarkAssaSSin._	4840719	30.6	3.34
110	TEBELKING	4784847	5.5	0.95
14	MachiTv	4659548	19.5	3.90
53	Ninjadad5	4595164	0.8	1.01
2	Ranger	4519465	34.0	9.60
144	RailCart	4479084	1.8	0.73
37	Ci2i	4323464	1.6	0.84
17	pr0Mancos	4300077	1.8	1.27
186	facu giorgetta29	4253032	7.0	1.36

Cálculo de los jugadores de menor rendimiento, riesgo de abandono en el Modo Solitario y disminución significativa en el número de partidas jugadas

```
# Filtrar jugadores con Solo kd y Solo winRatio bajos (por debajo de
la mediana)
jugadorBajoRendimiento = df[(df['Solo kd'] < df['Solo kd'].median()) &
(df['Solo winRatio'] < df['Solo
winRatio'].median())]

# Filtrar jugadores que han jugado pocas partidas en comparación con
el promedio
jugadoresPocasPartidas = df[df['Solo matches'] < df['Solo
matches'].quantile(0.25)]

# Filtrar jugadores con bajo rendimiento y baja actividad
jugadorBajoRendimientoYBajaAct = df[(df['Solo kd'] < df['Solo
kd'].median()) &
(df['Solo winRatio'] < df['Solo
winRatio'].median()) &
(df['Solo matches'] < df['Solo
matches'].quantile(0.25))]

jugadorBajoRendimiento.head(10)

{"column_count":8,"columns":
[{"dtype":"object","name":"Player","stats":{"categories":
[{"count":1,"name":"NiteGamerYT 190k"}, {"count":1,"name":"CIUPEA
```

```

144.HZ"}}, {"count":8, "name": "8
others"}], "nan_count":0, "unique_count":10}},
{"dtype": "int64", "name": "Solo score", "stats": {"histogram":
[{"bin_end":444292.5, "bin_start":15872, "count":5},
{"bin_end":872713, "bin_start":444292.5, "count":1},
{"bin_end":1301133.5, "bin_start":872713, "count":1},
{"bin_end":1729554, "bin_start":1301133.5, "count":1},
{"bin_end":2157974.5, "bin_start":1729554, "count":0},
{"bin_end":2586395, "bin_start":2157974.5, "count":1},
{"bin_end":3014815.5, "bin_start":2586395, "count":0},
{"bin_end":3443236, "bin_start":3014815.5, "count":0},
{"bin_end":3871656.5, "bin_start":3443236, "count":0},
{"bin_end":4300077, "bin_start":3871656.5, "count":1}], "max": "4300077", "
min": "15872", "nan_count":0, "unique_count":10}},
{"dtype": "int64", "name": "Solo top1", "stats": {"histogram":
[{"bin_end":63, "bin_start":6, "count":6},
{"bin_end":120, "bin_start":63, "count":2},
{"bin_end":177, "bin_start":120, "count":0},
{"bin_end":234, "bin_start":177, "count":0},
{"bin_end":291, "bin_start":234, "count":0},
{"bin_end":348, "bin_start":291, "count":1},
{"bin_end":405, "bin_start":348, "count":0},
{"bin_end":462, "bin_start":405, "count":0},
{"bin_end":519, "bin_start":462, "count":0},
{"bin_end":576, "bin_start":519, "count":1}], "max": "576", "min": "6", "nan_
count":0, "unique_count":9}}, {"dtype": "float64", "name": "Solo
kd", "stats": {"histogram":
[{"bin_end":0.5499999999999999, "bin_start":0.47, "count":1},
{"bin_end":0.63, "bin_start":0.5499999999999999, "count":1},
{"bin_end":0.71, "bin_start":0.63, "count":2},
{"bin_end":0.79, "bin_start":0.71, "count":0},
{"bin_end":0.87, "bin_start":0.79, "count":1},
{"bin_end":0.95, "bin_start":0.87, "count":0},
{"bin_end":1.03, "bin_start":0.95, "count":1},
{"bin_end":1.1099999999999999, "bin_start":1.03, "count":3},
{"bin_end":1.19, "bin_start":1.1099999999999999, "count":0},
{"bin_end":1.27, "bin_start":1.19, "count":1}], "max": "1.27", "min": "0.47"
, "nan_count":0, "unique_count":9}}, {"dtype": "float64", "name": "Solo
winRatio", "stats": {"histogram":
[{"bin_end":0.73, "bin_start":0.5, "count":3},
{"bin_end":0.96, "bin_start":0.73, "count":0},
{"bin_end":1.19, "bin_start":0.96, "count":2},
{"bin_end":1.42, "bin_start":1.19, "count":1},
{"bin_end":1.65, "bin_start":1.42, "count":0},
{"bin_end":1.88, "bin_start":1.65, "count":1},
{"bin_end":2.11, "bin_start":1.88, "count":0},
{"bin_end":2.34, "bin_start":2.11, "count":1},
{"bin_end":2.57, "bin_start":2.34, "count":1},
{"bin_end":2.8, "bin_start":2.57, "count":1}], "max": "2.8", "min": "0.5", "n

```



```

an_count":0,"unique_count":10}},{"dtype":"int64","name":"Solo
matches","stats":{"histogram":
[{"bin_end":3675.9,"bin_start":429,"count":6},
{"bin_end":6922.8,"bin_start":3675.9,"count":0},
{"bin_end":10169.7,"bin_start":6922.8,"count":1},
{"bin_end":13416.6,"bin_start":10169.7,"count":2},
{"bin_end":16663.5,"bin_start":13416.6,"count":0},
{"bin_end":19910.4,"bin_start":16663.5,"count":0},
{"bin_end":23157.3,"bin_start":19910.4,"count":0},
{"bin_end":26404.2,"bin_start":23157.3,"count":0},
{"bin_end":29651.100000000002,"bin_start":26404.2,"count":0},
{"bin_end":32898,"bin_start":29651.100000000002,"count":1}], "max":"328
98","min":"429","nan_count":0,"unique_count":10}},
{"dtype":"int64","name":"Solo kills","stats":{"histogram":
[{"bin_end":4274.2,"bin_start":200,"count":6},
{"bin_end":8348.4,"bin_start":4274.2,"count":1},
{"bin_end":12422.599999999999,"bin_start":8348.4,"count":2},
{"bin_end":16496.8,"bin_start":12422.599999999999,"count":0},
{"bin_end":20571,"bin_start":16496.8,"count":0},
{"bin_end":24645.199999999997,"bin_start":20571,"count":0},
{"bin_end":28719.399999999998,"bin_start":24645.199999999997,"count":0},
{"bin_end":32793.6,"bin_start":28719.399999999998,"count":0},
{"bin_end":36867.799999999996,"bin_start":32793.6,"count":0},
{"bin_end":40942,"bin_start":36867.799999999996,"count":1}], "max":"409
42","min":"200","nan_count":0,"unique_count":10}},
{"dtype":"int64","name":"Solo minutesPlayed","stats":{"histogram":
[{"bin_end":18229.4,"bin_start":739,"count":5},
{"bin_end":35719.8,"bin_start":18229.4,"count":1},
{"bin_end":53210.200000000004,"bin_start":35719.8,"count":0},
{"bin_end":70700.6,"bin_start":53210.200000000004,"count":2},
{"bin_end":88191,"bin_start":70700.6,"count":0},
{"bin_end":105681.400000000001,"bin_start":88191,"count":1},
{"bin_end":123171.800000000002,"bin_start":105681.400000000001,"count":0},
{"bin_end":140662.2,"bin_start":123171.800000000002,"count":0},
{"bin_end":158152.6,"bin_start":140662.2,"count":0},
{"bin_end":175643,"bin_start":158152.6,"count":1}], "max":"175643","min
":"739","nan_count":0,"unique_count":10}},
{"dtype":"int64","name":"_deepnote_index_column"}], "row_count":10,"row
s":[{"Player":"NiteGamerYT 190k","Solo kd":0.47,"Solo kills":200,"Solo
matches":429,"Solo minutesPlayed":739,"Solo score":15872,"Solo
top1":6,"Solo winRatio":1.4,"_deepnote_index_column":7},
{"Player":"CIUPEA 144.HZ","Solo kd":0.61,"Solo kills":1174,"Solo
matches":1938,"Solo minutesPlayed":2441,"Solo score":54479,"Solo
top1":9,"Solo winRatio":0.5,"_deepnote_index_column":9},
{"Player":"UçaN_HamSi","Solo kd":0.68,"Solo kills":1471,"Solo
matches":2182,"Solo minutesPlayed":3199,"Solo score":67758,"Solo
top1":25,"Solo winRatio":1.1,"_deepnote_index_column":28},
{"Player":"Trips","Solo kd":0.7,"Solo kills":1785,"Solo
matches":2581,"Solo minutesPlayed":21679,"Solo score":478280,"Solo

```

```

top1":25,"Solo winRatio":1,"_deepnote_index_column":13},
{"Player":"KING nax nax","Solo kd":0.8,"Solo kills":8356,"Solo
matches":10460,"Solo minutesPlayed":63105,"Solo score":1407541,"Solo
top1":70,"Solo winRatio":0.7,"_deepnote_index_column":24},
{"Player":"Deathfury420","Solo kd":0.95,"Solo kills":8101,"Solo
matches":8568,"Solo minutesPlayed":55671,"Solo score":1257973,"Solo
top1":51,"Solo winRatio":0.6,"_deepnote_index_column":23},
{"Player":"Scribell","Solo kd":1.03,"Solo kills":1505,"Solo
matches":1491,"Solo minutesPlayed":8945,"Solo score":212388,"Solo
top1":34,"Solo winRatio":2.3,"_deepnote_index_column":18},
{"Player":"Qx_9u انسنا","Solo kd":1.03,"Solo kills":3281,"Solo
matches":3269,"Solo minutesPlayed":8903,"Solo score":228620,"Solo
top1":79,"Solo winRatio":2.4,"_deepnote_index_column":19},
{"Player":"Kid_Fighter","Solo kd":1.06,"Solo kills":12348,"Solo
matches":11967,"Solo minutesPlayed":103551,"Solo score":2428145,"Solo
top1":331,"Solo winRatio":2.8,"_deepnote_index_column":29},
{"Player":"pr0Mancos","Solo kd":1.27,"Solo kills":40942,"Solo
matches":32898,"Solo minutesPlayed":175643,"Solo score":4300077,"Solo
top1":576,"Solo winRatio":1.8,"_deepnote_index_column":17}}]

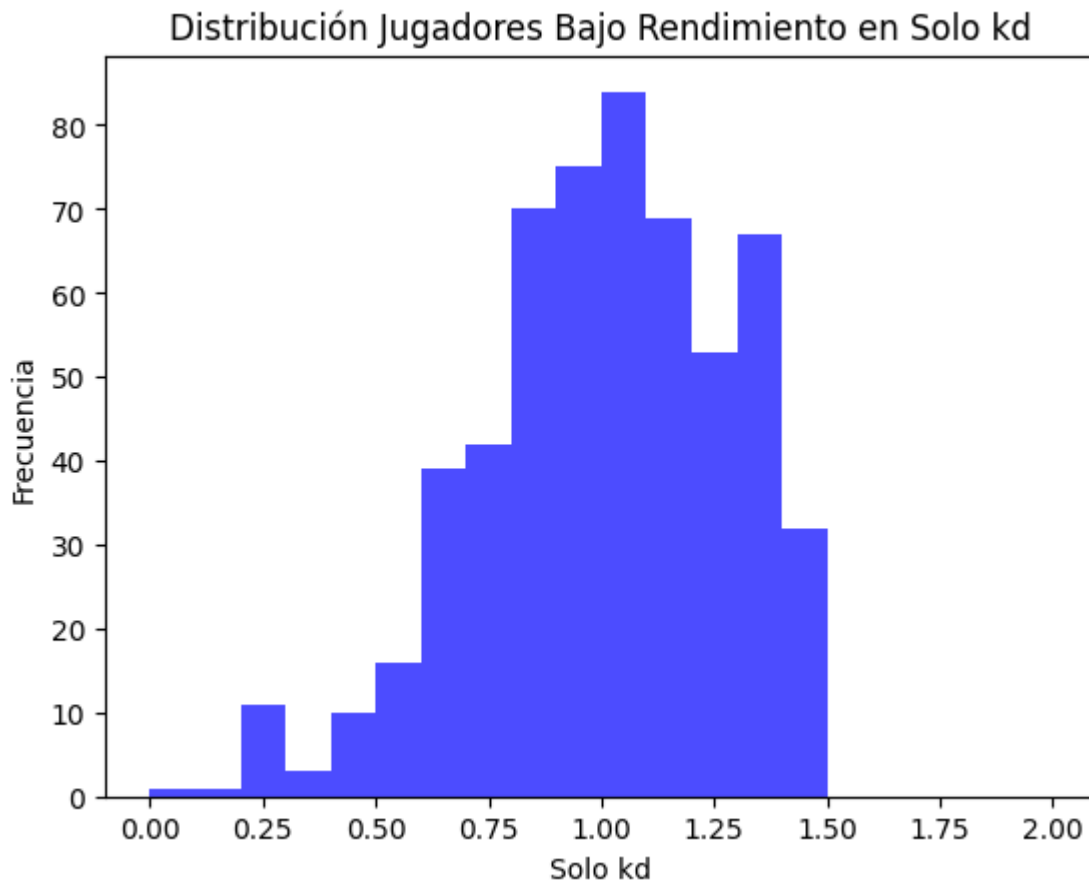
```

```
print(jugadorBajoRendimiento.describe())
```

	Solo score	Solo top1	Solo kd	Solo winRatio	Solo matches \
count	5.730000e+02	573.000000	573.000000	573.000000	573.000000
mean	9.795136e+05	99.668412	1.001850	1.451134	7292.431065
std	9.044167e+05	92.986643	0.277303	0.693415	6444.610264
min	4.453000e+03	0.000000	0.080000	0.000000	30.000000
25%	2.862850e+05	28.000000	0.830000	0.900000	2256.000000
50%	7.330330e+05	74.000000	1.020000	1.400000	5752.000000
75%	1.359115e+06	143.000000	1.230000	2.000000	10247.000000
max	4.595164e+06	591.000000	1.470000	2.800000	36487.000000

	Solo kills	Solo minutesPlayed
count	573.000000	573.000000
mean	7384.303665	42627.394415
std	6923.464940	40263.510377
min	28.000000	210.000000
25%	2147.000000	12418.000000
50%	5460.000000	31656.000000
75%	10606.000000	59160.000000
max	45588.000000	220536.000000

```
# Histograma de Solo kd para jugadores de bajo rendimiento
plt.hist(jugadorBajoRendimiento['Solo kd'], bins=20, color='blue',
alpha=0.7, range=(0,2))
plt.title('Distribución Jugadores Bajo Rendimiento en Solo kd')
plt.xlabel('Solo kd')
plt.ylabel('Frecuencia')
plt.show()
```



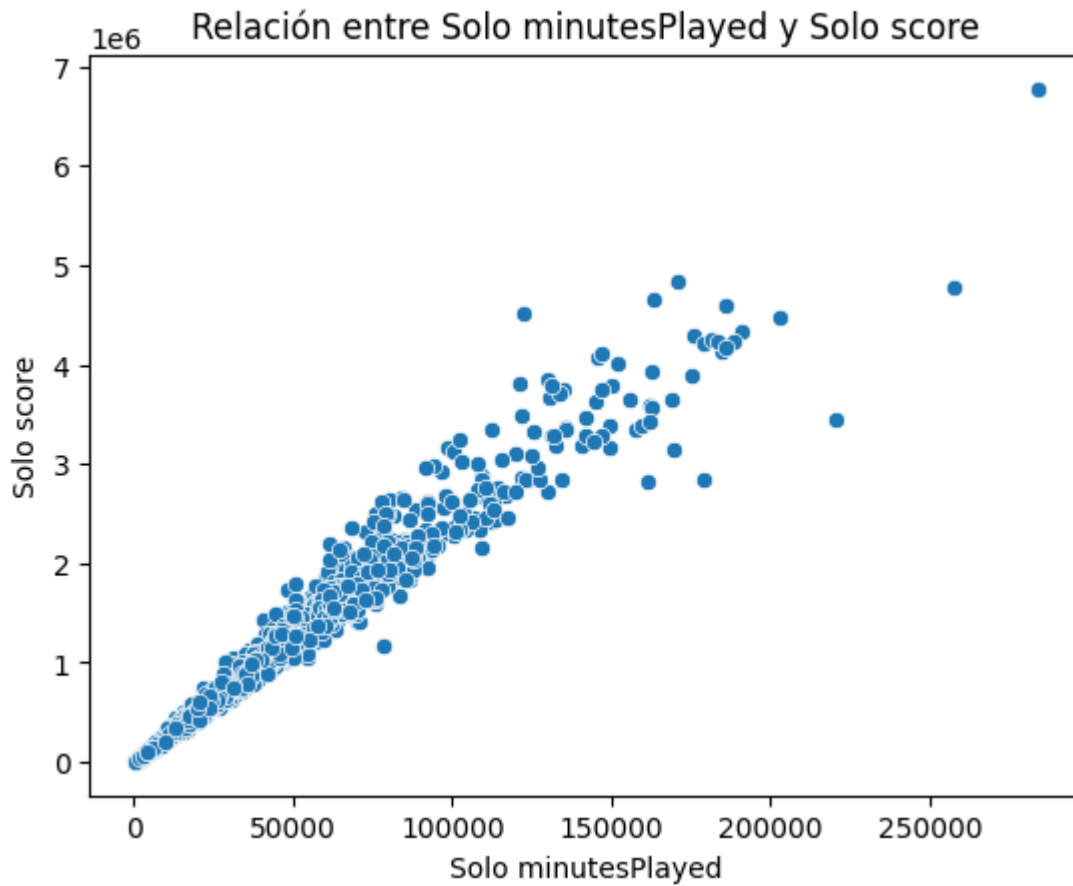
La mayor concentración de jugadores de menor rendimiento está alrededor de los valores de 1.0 y 1.25 en Solo kd. Esto significa que la mayoría de los jugadores considerados de bajo rendimiento tienen una proporción de 1 o un poco más de 1 muerte por cada vez que mueren.

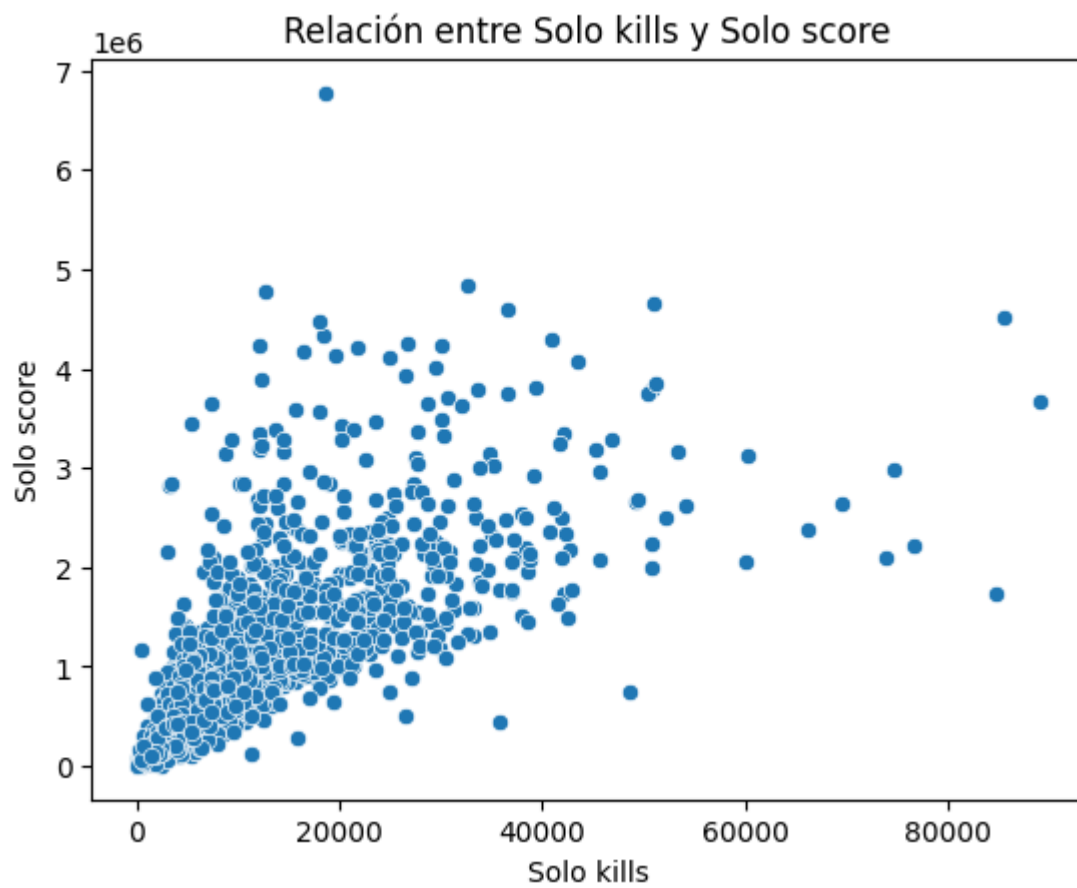
Casi no hay jugadores con un Solo kd superior a 1.5, lo que refuerza la idea de que estos jugadores están en la categoría de bajo rendimiento y con altas posibilidades de abandono de juego.

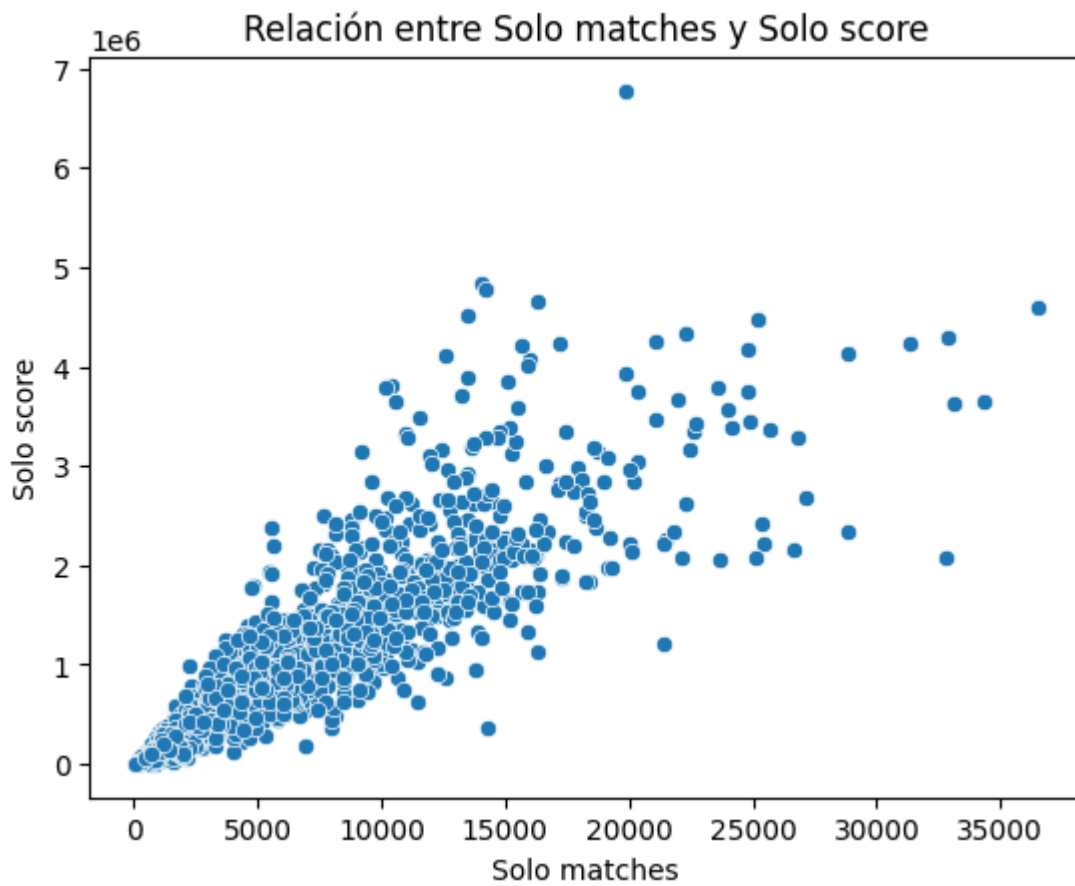
¿Qué factores influyen en la puntuación (Solo score) de un jugador?

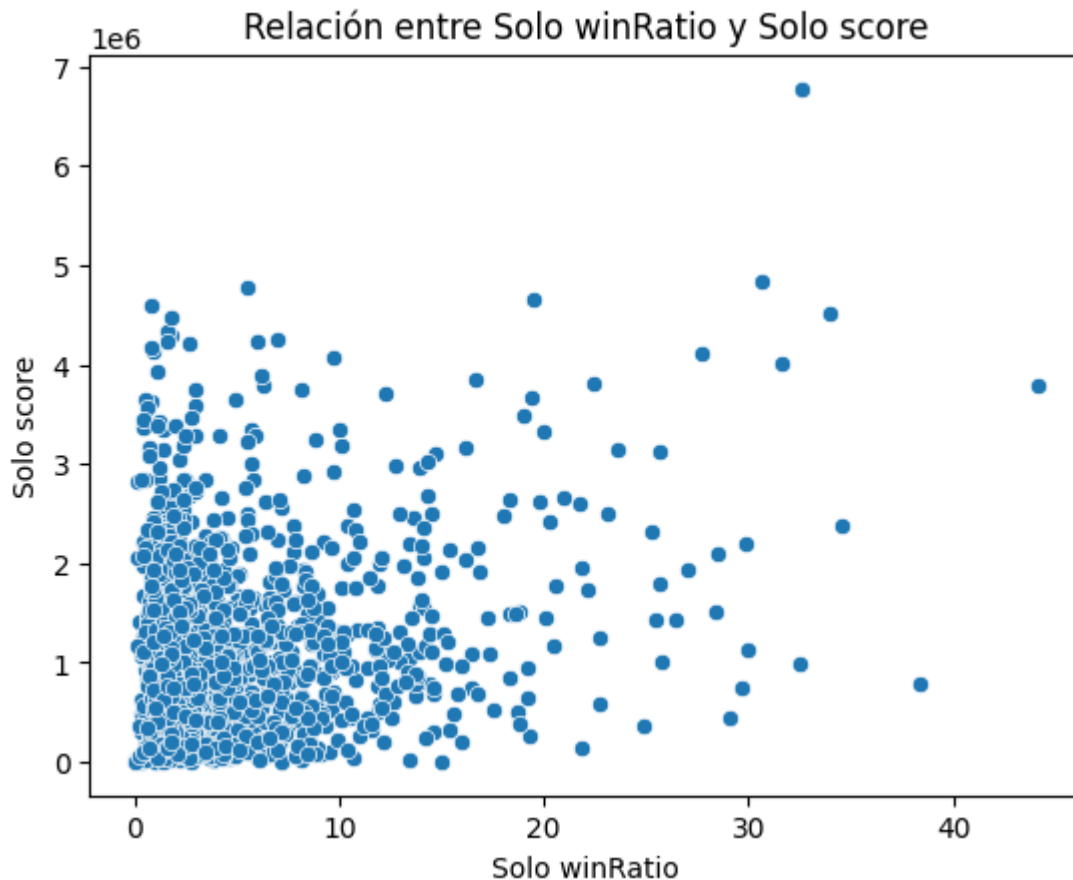
```
variables = ['Solo minutesPlayed', 'Solo kills', 'Solo matches', 'Solo winRatio']
```

```
# Gráficos de dispersión
for var in variables:
    plt.figure()
    sb.scatterplot(x=df[var], y=df['Solo score'])
    plt.title(f'Relación entre {var} y Solo score')
    plt.xlabel(var)
    plt.ylabel('Solo score')
    plt.show()
```









Existe una fuerte correlación entre el tiempo jugado (Solo minutesPlayed), el número de partidas (Solo matches), y el puntaje (Solo score). Esto indica que el puntaje es influenciado significativamente por la cantidad de tiempo y partidas jugadas. Aunque Solo kills también muestra una correlación positiva con el Solo Score, la dispersión muestra que la presencia de otros factores influyentes. Solo winRatio muestra la menor correlación, lo que indica que ganar una mayor proporción de partidas no siempre asegura un puntaje más alto.

Existe una relación significativa entre el tiempo jugado (Solo minutesPlayed) y el ratio de victorias (Solo winRatio)?

```
# Crear gráficos de dispersión
plt.figure(figsize=(14, 6))

# Scatter plot entre Solo minutesPlayed y Solo kills
plt.subplot(1, 2, 1)
sb.scatterplot(x=df['Solo minutesPlayed'], y=df['Solo kills'])
plt.title('Relación entre Solo minutesPlayed y Solo kills')
plt.xlabel('Solo minutesPlayed')
plt.ylabel('Solo kills')

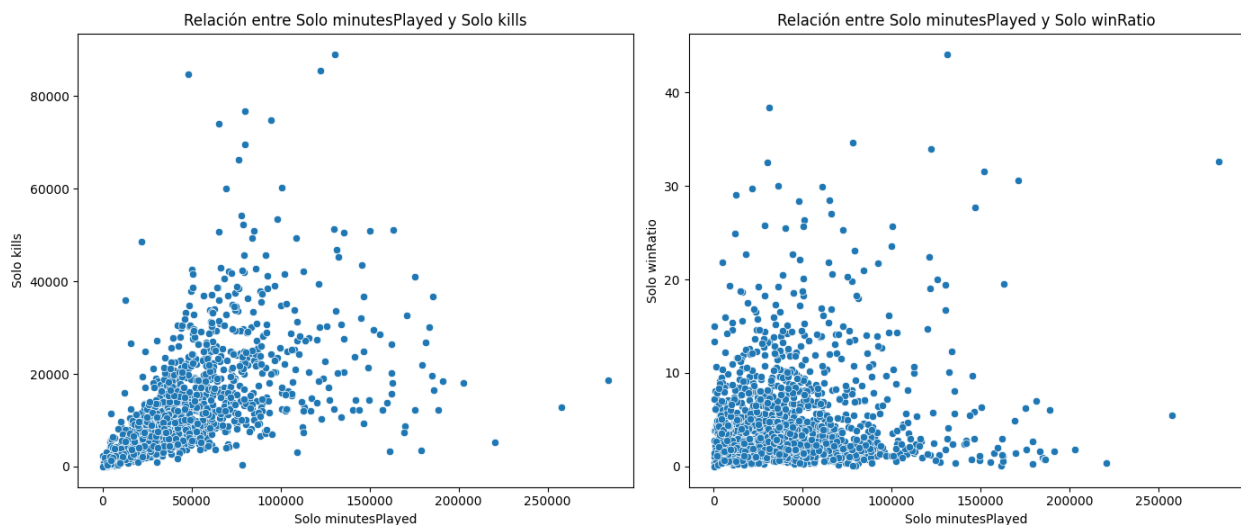
# Scatter plot entre Solo minutesPlayed y Solo winRatio
plt.subplot(1, 2, 2)
```

```

sb.scatterplot(x=df['Solo minutesPlayed'], y=df['Solo winRatio'])
plt.title('Relación entre Solo minutesPlayed y Solo winRatio')
plt.xlabel('Solo minutesPlayed')
plt.ylabel('Solo winRatio')

plt.tight_layout()
plt.show()

```



En el primer gráfico se observa que hay una tendencia positiva entre SoloMinutesPlayed y Solo Kills, según la matriz de correlación es de un 0.60, lo que indica que a medida que el tiempo de juego aumenta, también aumenta el número de eliminaciones.

Podemos decir que los jugadores que juegan más tiempo tienden a conseguir más eliminaciones. Pero, la separación de puntos nos demuestra que son muy dispersas, lo que nos señala que no todos los jugadores con mucho tiempo de juego tienen un alto número de eliminaciones.

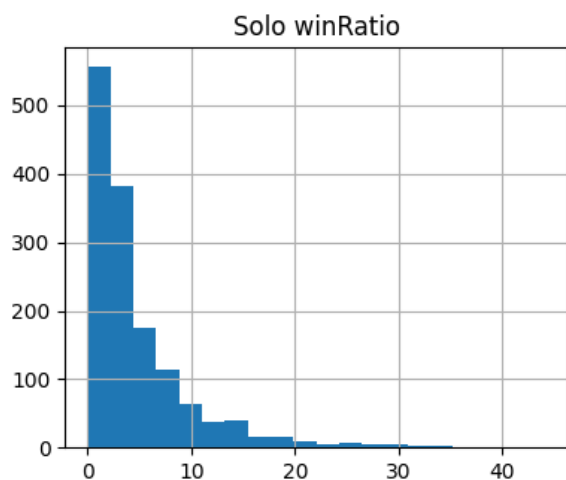
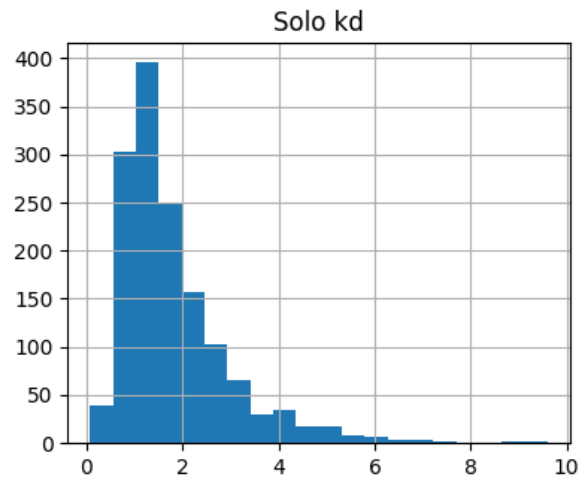
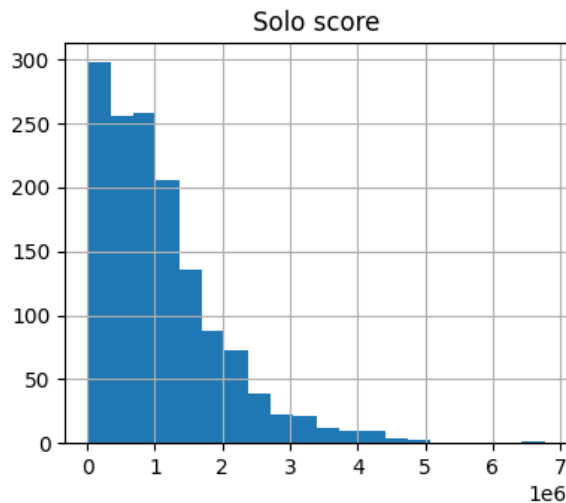
En el gráfico de Relación entre SoloMinutesPlayed y SoloWinRatio no hay una relación clara entre estas dos variables, ya que los puntos están distribuidos de manera más dispersa y no existe una tendencia clara.

Fase 3 Data Preparation

```

# Histograma para visualizar la distribución
df[['Solo score', 'Solo kd', 'Solo winRatio']].hist(bins=20,
figsize=(10, 8))
plt.show()

```

Solo score (Puntuación en Solitario): Este tipo de distribución sugiere que la mayor parte de los jugadores están concentrados en un rango bajo de puntuación, con un número reducido de jugadores que logran puntuaciones muy altas. Esto puede indicar una disparidad en el rendimiento, donde solo unos pocos jugadores son significativamente más exitosos que el promedio.

Solo kd (Relación K/D en Solitario): Esto indica que muchos jugadores tienden a mantener un rendimiento equilibrado (un número similar de bajas y muertes), mientras que solo algunos jugadores tienen un rendimiento mucho más alto (con un kd por encima de 2 o 3). Los valores extremos pueden ser jugadores con mucha habilidad o que se enfrentan a oponentes de menor nivel.

Solo winRatio (Porcentaje de Victorias en Solitario): La mayoría de los jugadores probablemente tengan dificultades para ganar en partidas en solitario. Solo un pequeño porcentaje logra un winRatio alto, lo cual es esperable, ya que las victorias son un resultado más exclusivo en un entorno competitivo como Fortnite.

Todas las variables muestran una distribución sesgada a la derecha, lo que sugiere que la mayoría de los jugadores tienen un rendimiento moderado, mientras que solo unos pocos logran resultados significativamente mejores.

Para mejorar el rendimiento del modelo vamos a normalizar los datos así obtener un mejor nivel predictivo de los datos.

Normalización de los datos

Primero seleccionamos las columnas que vamos a normalizar y utilizamos MinMaxScaler para escalar a valores entre 0 y 1.

```
columnasNormalizar = ['Solo score', 'Solo top1', 'Solo kd', 'Solo winRatio', 'Solo matches', 'Solo kills', 'Solo minutesPlayed']
```

Columnas antes de la normalización

```
df.head(10)
```

```
{
  "column_count": 8,
  "columns": [
    {
      "dtype": "object",
      "name": "Player",
      "stats": {
        "categories": [
          {
            "count": 1,
            "name": "Boss"
          },
          {
            "count": 1,
            "name": "Ranger"
          },
          {
            "count": 8,
            "name": "8 others"
          }
        ],
        "nan_count": 0,
        "unique_count": 10
      }
    },
    {
      "dtype": "int64",
      "name": "Solo score",
      "stats": {
        "histogram": [
          {
            "bin_end": 691496.4,
            "bin_start": 15872,
            "count": 3
          },
          {
            "bin_end": 1367120.8,
            "bin_start": 691496.4,
            "count": 2
          },
          {
            "bin_end": 2042745.2,
            "bin_start": 1367120.8,
            "count": 0
          },
          {
            "bin_end": 2718369.6,
            "bin_start": 2042745.2,
            "count": 2
          },
          {
            "bin_end": 3393994,
            "bin_start": 2718369.6,
            "count": 1
          },
          {
            "bin_end": 4069618.4,
            "bin_start": 3393994,
            "count": 0
          },
          {
            "bin_end": 4745242.8,
            "bin_start": 4069618.4,
            "count": 1
          },
          {
            "bin_end": 5420867.2,
            "bin_start": 4745242.8,
            "count": 0
          },
          {
            "bin_end": 6096491.6,
            "bin_start": 5420867.2,
            "count": 0
          },
          {
            "bin_end": 6772116,
            "bin_start": 6096491.6,
            "count": 1
          }
        ],
        "max": 6772116,
        "min": 15872,
        "nan_count": 0,
        "unique_count": 10
      }
    },
    {
      "dtype": "int64",
      "name": "Solo top1",
      "stats": {
        "histogram": [
          {
            "bin_end": 653.4,
            "bin_start": 6,
            "count": 3
          },
          {
            "bin_end": 1300.8,
            "bin_start": 653.4,
            "count": 0
          },
          {
            "bin_end": 1948.1999999999998,
            "bin_start": 1300.8,
            "count": 5
          },
          {
            "bin_end": 2595.6,
            "bin_start": 1948.1999999999998,
            "count": 0
          },
          {
            "bin_end": 3243,
            "bin_start": 2595.6,
            "count": 0
          },
          {
            "bin_end": 3890.3999999999996,
            "bin_start": 3243,
            "count": 0
          },
          {
            "bin_end": 4537.8,
            "bin_start": 3890.3999999999996,
            "count": 0
          },
          {
            "bin_end": 5185.2,
            "bin_start": 4537.8,
            "count": 1
          },
          {
            "bin_end": 5832.599999999999,
            "bin_start": 5185.2,
            "count": 0
          },
          {
            "bin_end": 6480,
            "bin_start": 5832.599999999999,
            "count": 1
          }
        ],
        "max": 6480,
        "min": 6,
        "nan_count": 0,
        "unique_count": 10
      }
    },
    {
      "dtype": "float64",
      "name": "Solo kd",
      "stats": {
        "histogram": [
          {
            "bin_end": 1.383,
            "bin_start": 0.47,
            "count": 3
          },
          {
            "bin_end": 2.296,
            "bin_start": 1.383,
            "count": 1
          },
          {
            "bin_end": 3.2089999999999996,
            "bin_start": 2.296,
            "count": 0
          },
          {
            "bin_end": 4.122,
            "bin_start": 3.2089999999999996,
            "count": 2
          },
          {
            "bin_end": 5.034999999999999,
            "bin_start": 4.122,
            "count": 1
          }
        ]
      }
    }
  ]
}
```

```

{"bin_end":5.9479999999999995,"bin_start":5.034999999999999,"count":0}
,
{"bin_end":6.860999999999999,"bin_start":5.9479999999999995,"count":1}
,
{"bin_end":7.773999999999999,"bin_start":6.860999999999999,"count":0},
{"bin_end":8.687,"bin_start":7.773999999999999,"count":0},
{"bin_end":9.6,"bin_start":8.687,"count":2}], "max":"9.6", "min":"0.47",
"nan_count":0,"unique_count":10}}, {"dtype":"float64", "name":"Solo
winRatio", "stats":{"histogram":
[{"bin_end":3.85,"bin_start":0.5,"count":3},
{"bin_end":7.2,"bin_start":3.85,"count":0},
{"bin_end":10.55,"bin_start":7.2,"count":2},
{"bin_end":13.9,"bin_start":10.55,"count":0},
{"bin_end":17.25,"bin_start":13.9,"count":0},
{"bin_end":20.6,"bin_start":17.25,"count":1},
{"bin_end":23.95,"bin_start":20.6,"count":0},
{"bin_end":27.3,"bin_start":23.95,"count":0},
{"bin_end":30.650000000000002,"bin_start":27.3,"count":2},
{"bin_end":34,"bin_start":30.650000000000002,"count":2}], "max":"34.0",
"min":"0.5", "nan_count":0,"unique_count":10}},
{"dtype":"int64", "name":"Solo matches", "stats":{"histogram":
[{"bin_end":2372.5,"bin_start":429,"count":2},
{"bin_end":4316,"bin_start":2372.5,"count":1},
{"bin_end":6259.5,"bin_start":4316,"count":2},
{"bin_end":8203,"bin_start":6259.5,"count":0},
{"bin_end":10146.5,"bin_start":8203,"count":0},
{"bin_end":12090,"bin_start":10146.5,"count":1},
{"bin_end":14033.5,"bin_start":12090,"count":2},
{"bin_end":15977,"bin_start":14033.5,"count":0},
{"bin_end":17920.5,"bin_start":15977,"count":0},
{"bin_end":19864,"bin_start":17920.5,"count":2}], "max":"19864", "min":
"429", "nan_count":0,"unique_count":10}}, {"dtype":"int64", "name":"Solo
kills", "stats":{"histogram":
[{"bin_end":8728.1,"bin_start":200,"count":3},
{"bin_end":17256.2,"bin_start":8728.1,"count":0},
{"bin_end":25784.3000000000003,"bin_start":17256.2,"count":2},
{"bin_end":34312.4,"bin_start":25784.3000000000003,"count":0},
{"bin_end":42840.5,"bin_start":34312.4,"count":3},
{"bin_end":51368.6000000000006,"bin_start":42840.5,"count":0},
{"bin_end":59896.7000000000004,"bin_start":51368.6000000000006,"count":0
}, {"bin_end":68424.8,"bin_start":59896.7000000000004,"count":1},
{"bin_end":76952.9000000000001,"bin_start":68424.8,"count":0},
{"bin_end":85481,"bin_start":76952.9000000000001,"count":1}], "max":"8548
1", "min":"200", "nan_count":0,"unique_count":10}},
{"dtype":"int64", "name":"Solo minutesPlayed", "stats":{"histogram":
[{"bin_end":29048.6,"bin_start":739,"count":3},
{"bin_end":57358.2,"bin_start":29048.6,"count":2},
{"bin_end":85667.79999999999,"bin_start":57358.2,"count":2},
{"bin_end":113977.4,"bin_start":85667.79999999999,"count":1}],

```

```
{
  "bin_end": 142287, "bin_start": 113977.4, "count": 1},
  {"bin_end": 170596.59999999998, "bin_start": 142287, "count": 0},
  {"bin_end": 198906.19999999998, "bin_start": 170596.59999999998, "count": 0},
  {"bin_end": 227215.8, "bin_start": 198906.19999999998, "count": 0},
  {"bin_end": 255525.4, "bin_start": 227215.8, "count": 0},
  {"bin_end": 283835, "bin_start": 255525.4, "count": 1}], "max": "283835", "min": "739", "nan_count": 0, "unique_count": 10}},
{"dtype": "int64", "name": "_deepnote_index_column"}], "row_count": 10, "rows": [
  {"Player": "Boss", "Solo kd": 1.39, "Solo kills": 18610, "Solo matches": 19864, "Solo minutesPlayed": 283835, "Solo score": 6772116, "Solo top1": 6480, "Solo winRatio": 32.6, "_deepnote_index_column": 8},
  {"Player": "Ranger", "Solo kd": 9.6, "Solo kills": 85481, "Solo matches": 13488, "Solo minutesPlayed": 122171, "Solo score": 4519465, "Solo top1": 4582, "Solo winRatio": 34, "_deepnote_index_column": 2},
  {"Player": "Twitch Kayotica", "Solo kd": 3.23, "Solo kills": 39131, "Solo matches": 13438, "Solo minutesPlayed": 96777, "Solo score": 2919037, "Solo top1": 1310, "Solo winRatio": 9.7, "_deepnote_index_column": 5},
  {"Player": "Prospering", "Solo kd": 4.37, "Solo kills": 36328, "Solo matches": 10150, "Solo minutesPlayed": 81389, "Solo score": 2476763, "Solo top1": 1828, "Solo winRatio": 18, "_deepnote_index_column": 0},
  {"Player": "FaZe Replays", "Solo kd": 3.84, "Solo kills": 66161, "Solo matches": 18670, "Solo minutesPlayed": 76258, "Solo score": 2389537, "Solo top1": 1454, "Solo winRatio": 7.8, "_deepnote_index_column": 6},
  {"Player": "Twitch.GryphonRB", "Solo kd": 6.32, "Solo kills": 19591, "Solo matches": 4429, "Solo minutesPlayed": 36245, "Solo score": 1136282, "Solo top1": 1327, "Solo winRatio": 30, "_deepnote_index_column": 4},
  {"Player": "Idk_Pi", "Solo kd": 0.84, "Solo kills": 3005, "Solo matches": 3687, "Solo minutesPlayed": 32453, "Solo score": 752869, "Solo top1": 121, "Solo winRatio": 3.3, "_deepnote_index_column": 3},
  {"Player": "BH nixxay", "Solo kd": 8.71, "Solo kills": 35895, "Solo matches": 5817, "Solo minutesPlayed": 12732, "Solo score": 439562, "Solo top1": 1694, "Solo winRatio": 29.1, "_deepnote_index_column": 1},
  {"Player": "CIUPEA 144.HZ", "Solo kd": 0.61, "Solo kills": 1174, "Solo matches": 1938, "Solo minutesPlayed": 2441, "Solo score": 54479, "Solo top1": 9, "Solo winRatio": 0.5, "_deepnote_index_column": 9},
  {"Player": "NiteGamerYT 190k", "Solo kd": 0.47, "Solo kills": 200, "Solo matches": 429, "Solo minutesPlayed": 739, "Solo score": 15872, "Solo top1": 6, "Solo winRatio": 1.4, "_deepnote_index_column": 7}]]
```

Columnas después de la normalización

```
# Aplicar Min-Max Scaler
scaler = MinMaxScaler()
dfNormalizado = df.copy() # Crear una copia del DataFrame
dfNormalizado[columnasNormalizar] =
scaler.fit_transform(df[columnasNormalizar])
dfNormalizado.head(10)
```

```

{"column_count":8,"columns":
[{"dtype":"object","name":"Player","stats":{"categories":
[{"count":1,"name":"Boss ン"}, {"count":1,"name":"Ranger"},
{"count":8,"name":"8 others"}]}, {"nan_count":0,"unique_count":10}},
{"dtype":"float64","name":"Solo score","stats":{"histogram":
[{"bin_end":0.10151855965641314,"bin_start":1.6872885071257243e-
3,"count":3},
{"bin_end":0.20134983080570054,"bin_start":0.10151855965641314,"count":
2},
{"bin_end":0.301181101954988,"bin_start":0.20134983080570054,"count":0
},
{"bin_end":0.4010123731042754,"bin_start":0.301181101954988,"count":2}
,
{"bin_end":0.5008436442535628,"bin_start":0.4010123731042754,"count":1
},
{"bin_end":0.6006749154028502,"bin_start":0.5008436442535628,"count":0
},
{"bin_end":0.7005061865521376,"bin_start":0.6006749154028502,"count":1
},
{"bin_end":0.8003374577014251,"bin_start":0.7005061865521376,"count":0
},
{"bin_end":0.9001687288507125,"bin_start":0.8003374577014251,"count":0
},
{"bin_end":0.9999999999999999,"bin_start":0.9001687288507125,"count":1
}], "max":"0.9999999999999999", "min":"0.0016872885071257243", "nan_count":
0, "unique_count":10}}, {"dtype":"float64","name":"Solo top1","stats":
{"histogram":
[{"bin_end":0.10083333333333334,"bin_start":9.25925925925926e-
4,"count":3},
{"bin_end":0.20074074074074075,"bin_start":0.10083333333333334,"count":
0},
{"bin_end":0.30064814814814816,"bin_start":0.20074074074074075,"count":
5},
{"bin_end":0.40055555555555556,"bin_start":0.30064814814814816,"count":
0},
{"bin_end":0.500462962962963,"bin_start":0.40055555555555556,"count":0}
,
{"bin_end":0.6003703703703703,"bin_start":0.500462962962963,"count":0}
,
{"bin_end":0.7002777777777778,"bin_start":0.6003703703703703,"count":0
},
{"bin_end":0.8001851851851852,"bin_start":0.7002777777777778,"count":1
},
{"bin_end":0.9000925925925927,"bin_start":0.8001851851851852,"count":0
},
{"bin_end":1,"bin_start":0.9000925925925927,"count":1}], "max":"1.0", "m
in":"0.000925925925925926", "nan_count":0, "unique_count":10}},
{"dtype":"float64","name":"Solo kd","stats":{"histogram":
[{"bin_end":0.13686974789915968,"bin_start":4.096638655462185e-
2,"count":3},

```

```
{"bin_end":0.23277310924369748,"bin_start":0.13686974789915968,"count":1},
{"bin_end":0.32867647058823535,"bin_start":0.23277310924369748,"count":0},
{"bin_end":0.42457983193277316,"bin_start":0.32867647058823535,"count":2},
{"bin_end":0.520483193277311,"bin_start":0.42457983193277316,"count":1},
{"bin_end":0.6163865546218488,"bin_start":0.520483193277311,"count":0},
{"bin_end":0.7122899159663866,"bin_start":0.6163865546218488,"count":1},
{"bin_end":0.8081932773109244,"bin_start":0.7122899159663866,"count":0},
{"bin_end":0.9040966386554622,"bin_start":0.8081932773109244,"count":0},
{"bin_end":1,"bin_start":0.9040966386554622,"count":2}], "max": "1.0", "min": "0.04096638655462185", "nan_count": 0, "unique_count": 10}},
{"dtype": "float64", "name": "Solo winRatio", "stats": {"histogram": [{"bin_end": 8.730158730158731e-2, "bin_start": 1.1337868480725623e-2, "count": 3}, {"bin_end": 0.163265306122449, "bin_start": 8.730158730158731e-2, "count": 0}, {"bin_end": 0.2392290249433107, "bin_start": 0.163265306122449, "count": 2}, {"bin_end": 0.31519274376417233, "bin_start": 0.2392290249433107, "count": 0}, {"bin_end": 0.391156462585034, "bin_start": 0.31519274376417233, "count": 0}, {"bin_end": 0.46712018140589573, "bin_start": 0.391156462585034, "count": 1}, {"bin_end": 0.5430839002267573, "bin_start": 0.46712018140589573, "count": 0}, {"bin_end": 0.6190476190476191, "bin_start": 0.5430839002267573, "count": 0}, {"bin_end": 0.6950113378684808, "bin_start": 0.6190476190476191, "count": 2}, {"bin_end": 0.7709750566893424, "bin_start": 0.6950113378684808, "count": 2}], "max": "0.7709750566893424", "min": "0.011337868480725623", "nan_count": 0, "unique_count": 10}}, {"dtype": "float64", "name": "Solo matches", "stats": {"histogram": [{"bin_end": 6.425377842389664e-2, "bin_start": 1.0944400252352085e-2, "count": 2}, {"bin_end": 0.1175631565954412, "bin_start": 6.425377842389664e-2, "count": 1}, {"bin_end": 0.17087253476698575, "bin_start": 0.1175631565954412, "count": 2}, {"bin_end": 0.2241819129385303, "bin_start": 0.17087253476698575, "count": 0}, {"bin_end": 0.27749129111007487, "bin_start": 0.2241819129385303, "count": 0},
```

```
{"bin_end":0.33080066928161944,"bin_start":0.27749129111007487,"count":1},
{"bin_end":0.38411004745316396,"bin_start":0.33080066928161944,"count":2},
{"bin_end":0.43741942562470854,"bin_start":0.38411004745316396,"count":0},
{"bin_end":0.4907288037962531,"bin_start":0.43741942562470854,"count":0},
{"bin_end":0.5440381819677976,"bin_start":0.4907288037962531,"count":2}],
{"max":"0.5440381819677976","min":"0.010944400252352085","nan_count":0,
"unique_count":10}},{"dtype":"float64","name":"Solo kills","stats":{"histogram":
[{"bin_end":9.775942468678016e-2,"bin_start":1.9326928479128042e-3,"count":3},
{"bin_end":0.1935861565256475,"bin_start":9.775942468678016e-2,"count":0},
{"bin_end":0.28941288836451484,"bin_start":0.1935861565256475,"count":2},
{"bin_end":0.3852396202033822,"bin_start":0.28941288836451484,"count":0},
{"bin_end":0.48106635204224957,"bin_start":0.3852396202033822,"count":3},
{"bin_end":0.5768930838811168,"bin_start":0.48106635204224957,"count":0},
{"bin_end":0.6727198157199842,"bin_start":0.5768930838811168,"count":0},
{"bin_end":0.7685465475588515,"bin_start":0.6727198157199842,"count":1},
{"bin_end":0.8643732793977189,"bin_start":0.7685465475588515,"count":0},
{"bin_end":0.9602000112365863,"bin_start":0.8643732793977189,"count":1}],
{"max":"0.9602000112365863","min":"0.0019326928479128042","nan_count":0,
"unique_count":10}},{"dtype":"float64","name":"Solo minutesPlayed","stats":{"histogram":
[{"bin_end":0.10167862494490965,"bin_start":1.8651388276773907e-3,"count":3},
{"bin_end":0.2014921110621419,"bin_start":0.10167862494490965,"count":2},
{"bin_end":0.30130559717937416,"bin_start":0.2014921110621419,"count":2},
{"bin_end":0.40111908329660645,"bin_start":0.30130559717937416,"count":1},
{"bin_end":0.5009325694138387,"bin_start":0.40111908329660645,"count":1},
{"bin_end":0.6007460555310709,"bin_start":0.5009325694138387,"count":0},
{"bin_end":0.7005595416483031,"bin_start":0.6007460555310709,"count":0},
{"bin_end":0.8003730277655354,"bin_start":0.7005595416483031,"count":0},
{"bin_end":0.9001865138827677,"bin_start":0.8003730277655354,"count":0}
```



```

},
{"bin_end":1,"bin_start":0.9001865138827677,"count":1}], "max":"1.0", "min":
"0.0018651388276773907", "nan_count":0, "unique_count":10}},
{"dtype":"int64", "name": "_deepnote_index_column"}], "row_count":10, "rows":
[{"Player": "Boss ン", "Solo kd":0.1376050420168067, "Solo kills":0.20879824709253328, "Solo matches":0.5440381819677976, "Solo minutesPlayed":1, "Solo score":0.9999999999999999, "Solo top1":1, "Solo winRatio":0.7392290249433107, "_deepnote_index_column":8}, {"Player": "Ranger", "Solo kd":1, "Solo kills":0.9602000112365863, "Solo matches":0.3691472145267027, "Solo minutesPlayed":0.4300079330101366, "Solo score":0.6671449213709371, "Solo top1":0.7070987654320987, "Solo winRatio":0.7709750566893424, "_deepnote_index_column":2}, {"Player": "Twitch Kayotica", "Solo kd":0.33088235294117646, "Solo kills":0.4393842350693859, "Solo matches":0.3677757357983378, "Solo minutesPlayed":0.340474217717056, "Solo score":0.43066328805083826, "Solo top1":0.2021604938271605, "Solo winRatio":0.21995464852607707, "_deepnote_index_column":5}, {"Player": "Prospering", "Solo kd":0.45063025210084034, "Solo kills":0.4078880836002023, "Solo matches":0.2775872946210604, "Solo minutesPlayed":0.2862194799471133, "Solo score":0.36531222077695064, "Solo top1":0.2820987654320988, "Solo winRatio":0.40816326530612246, "_deepnote_index_column":0}, {"Player": "FaZe Replays", "Solo kd":0.3949579831932773, "Solo kills":0.7431091634361481, "Solo matches":0.5112872699344433, "Solo minutesPlayed":0.2681286910533275, "Solo score":0.3524235766467686, "Solo top1":0.2243827160493827, "Solo winRatio":0.17687074829931973, "_deepnote_index_column":6}, {"Player": "Twitch.GryphonRB", "Solo kd":0.6554621848739497, "Solo kills":0.2198213382774313, "Solo matches":0.12066269852154593, "Solo minutesPlayed":0.12705156456588806, "Solo score":0.16724074470020153, "Solo top1":0.20478395061728394, "Solo winRatio":0.6802721088435374, "_deepnote_index_column":4}, {"Player": "Idk_Pi", "Solo kd":7.983193277310924e-2, "Solo kills":3.345131748974662e-2, "Solo matches":0.10030995419261048, "Solo minutesPlayed":0.1136817981489643, "Solo score":0.11058706676145073, "Solo top1":1.867283950617284e-2, "Solo winRatio":7.482993197278912e-2, "_deepnote_index_column":3}, {"Player": "BH nixxay", "Solo kd":0.9065126050420169, "Solo kills":0.40302264172144503, "Solo matches":0.1587349480209562, "Solo minutesPlayed":4.4149845747025115e-2, "Solo score":6.429235616489769e-2, "Solo top1":0.26141975308641974, "Solo winRatio":0.6598639455782314, "_deepnote_index_column":1}, {"Player": "CIUPEA 144.HZ", "Solo kd":5.567226890756302e-2, "Solo kills":1.287712792853531e-2, "Solo matches":5.2335628274405466e-2, "Solo minutesPlayed":7.866020273248127e-3, "Solo score":7.391916530122732e-3, "Solo top1":1.388888888888889e-3, "Solo winRatio":1.1337868480725623e-2, "_deepnote_index_column":9}, {"Player": "NiteGamerYT 190k", "Solo kd":4.096638655462185e-2, "Solo kills":1.9326928479128042e-3, "Solo matches":1.0944400252352085e-

```



```
2, "Solo minutesPlayed":1.8651388276773907e-3, "Solo score":1.6872885071257243e-3, "Solo top1":9.25925925925926e-4, "Solo winRatio":3.1746031746031744e-2, "_deepnote_index_column":7}}]
```

aValor cercano a 0: El jugador tiene un bajo rendimiento en comparación con otros jugadores para esa métrica específica. Valor cercano a 1: El jugador tiene un alto rendimiento en comparación con otros jugadores para esa métrica específica.

"Boss" es el mejor jugador en términos de Solo score, Solo top1, Solo kd y Solo minutesPlayed, porque sus valores están en 1, lo que indica que está al tope de las métricas de rendimiento en comparación con los demás jugadores.

La normalización nos ayuda a estandarizar las escalas de las diferentes variables para que los algoritmos de machine learning puedan interpretar correctamente los datos, evitando que variables con mayor rango afecten de manera desproporcionada los resultados.

Tratamiento de outliers

```
# Calcular el IQR
Q1 = df['Solo kd'].quantile(0.25)
Q3 = df['Solo kd'].quantile(0.75)
IQR = Q3 - Q1

# Definir los límites inferior y superior
lim_inferior = Q1 - 1.5 * IQR
lim_superior = Q3 + 1.5 * IQR

# Filtrar datos dentro de los límites
df_solo_sin_outliers = df[(df['Solo kd'] >= lim_inferior) & (df['Solo kd'] <= lim_superior)]

print("Número de filas antes de eliminar outliers:", len(df))
print("Número de filas después de eliminar outliers:", len(df_solo_sin_outliers))

Número de filas antes de eliminar outliers: 1435
Número de filas después de eliminar outliers: 1343

# Verificar límites del IQR
print(f"Límite inferior: {lim_inferior}, Límite superior: {lim_superior}")

Límite inferior: -0.675, Límite superior: 3.9250000000000003

# Cálculo del IQR para Solo minutesPlayed
Q1_mp = df['Solo minutesPlayed'].quantile(0.25)
Q3_mp = df['Solo minutesPlayed'].quantile(0.75)
IQR_mp = Q3_mp - Q1_mp
```

```

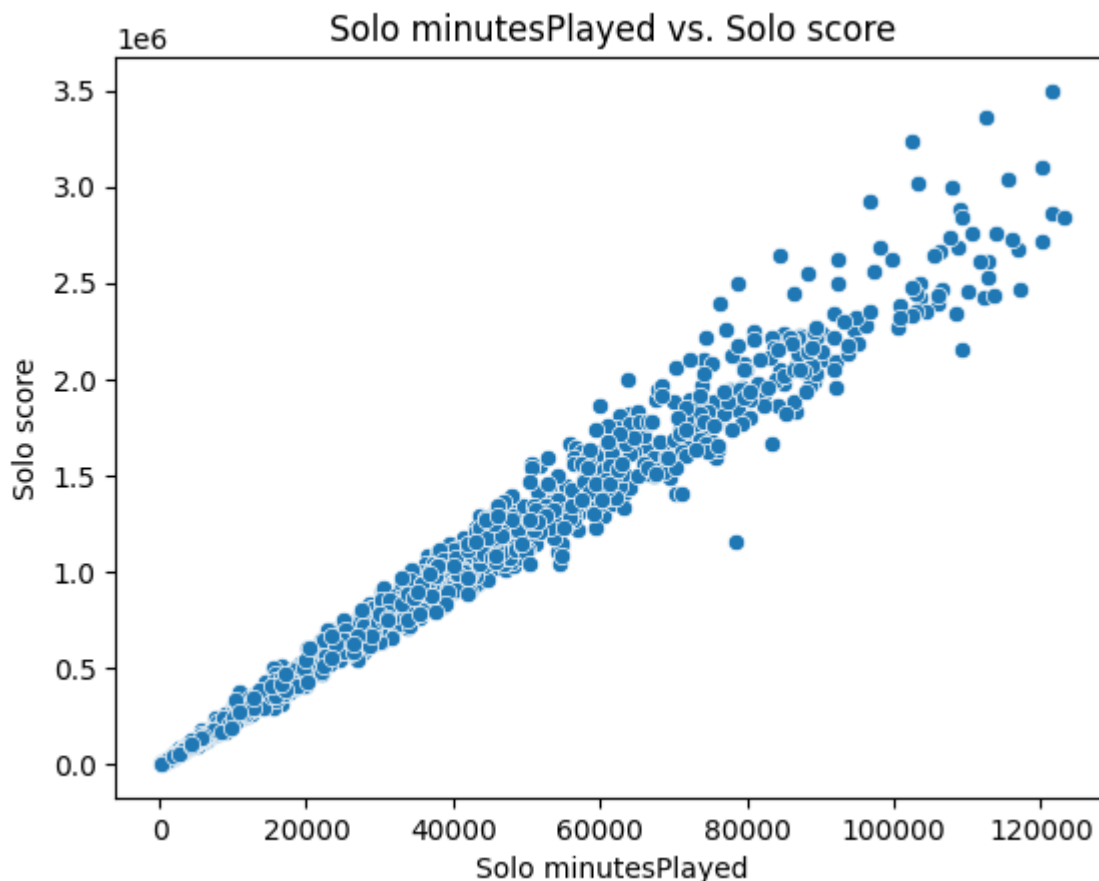
# Límite inferior y superior para Solo minutesPlayed
lim_inferior_mp = Q1_mp - 1.5 * IQR_mp
lim_superior_mp = Q3_mp + 1.5 * IQR_mp

# Filtrar datos sin outliers en Solo kd y Solo minutesPlayed
df_solo_sin_outliers =
df_solo_sin_outliers[(df_solo_sin_outliers['Solo minutesPlayed'] >=
lim_inferior_mp) &

(df_solo_sin_outliers['Solo minutesPlayed'] <= lim_superior_mp)]

# Visualización final del scatterplot sin outliers
sb.scatterplot(x='Solo minutesPlayed', y='Solo score',
data=df_solo_sin_outliers)
plt.title('Solo minutesPlayed vs. Solo score')
plt.show()

```



El gráfico resultante muestra una relación positiva entre el tiempo jugado en modo solo (Solo minutesPlayed) y la puntuación obtenida en ese modo (Solo score). Es decir, a medida que los jugadores pasan más tiempo jugando, tienden a obtener puntuaciones más altas.

Esta relación es consistente y sigue una tendencia lineal, aunque se observan algunas dispersiones (pequeñas variaciones) a medida que aumenta el número de minutos jugados. Al

haber filtrado los outliers, se han eliminado valores atípicos que podrían haber distorsionado la interpretación de la tendencia general. El gráfico se concentra en los datos que representan el comportamiento típico de los jugadores.

Tratamiento de outliers para conocer la existencia de relación significativa entre el tiempo jugado (Solo minutesPlayed) y el ratio de victorias (Solo winRatio)?

```
# Cálculo del IQR para Solo minutesPlayed
Q1_minutes = df['Solo minutesPlayed'].quantile(0.25)
Q3_minutes = df['Solo minutesPlayed'].quantile(0.75)
IQR_minutes = Q3_minutes - Q1_minutes

# Cálculo del IQR para Solo kills
Q1_kills = df['Solo kills'].quantile(0.25)
Q3_kills = df['Solo kills'].quantile(0.75)
IQR_kills = Q3_kills - Q1_kills

# Definir límites para Solo minutesPlayed
lower_bound_minutes = Q1_minutes - 1.5 * IQR_minutes
upper_bound_minutes = Q3_minutes + 1.5 * IQR_minutes

# Definir límites para Solo kills
lower_bound_kills = Q1_kills - 1.5 * IQR_kills
upper_bound_kills = Q3_kills + 1.5 * IQR_kills

# Filtrar datos para eliminar outliers
df_sin_outliers = df[(df['Solo minutesPlayed'] >= lower_bound_minutes)
& (df['Solo minutesPlayed'] <= upper_bound_minutes) &
(df['Solo kills'] >= lower_bound_kills) &
(df['Solo kills'] <= upper_bound_kills)]

# Mostrar los datos sin outliers
print(df_sin_outliers)
```

	Player	Solo score	Solo top1	Solo kd	Solo winRatio
3	Idk_Pi	752869	121	0.84	3.3
4	Twitch.GryphonRB	1136282	1327	6.32	30.0
7	NiteGamerYT 190k	15872	6	0.47	1.4
9	CIUPEA 144.HZ	54479	9	0.61	0.5
10	曼巴精神の R6	208411	103	4.17	
12.2					
...

1430	im bloom	106294	26	2.03	3.5
1431	Twitch ka0zs_	803918	396	3.42	13.2
1432	slxyher	603939	310	1.69	5.1
1433	RD-Antony	747295	71	1.10	1.9
1434	O-H0-H0-H0!	1378689	161	1.68	2.3

	Solo matches	Solo kills	Solo minutesPlayed
3	3687	3005	32453
4	4429	19591	36245
7	429	200	739
9	1938	1174	2441
10	844	3093	6661
...
1430	736	1440	4470
1431	3001	8911	27375
1432	6035	9679	20545
1433	3763	4048	31147
1434	7105	11661	57572

[1322 rows x 8 columns]

Usamos RobustScaler de scikit-learn, para reducir impacto de los outliers escalando los datos de acuerdo al IQR

```
scaler = RobustScaler()
df[['Solo minutesPlayed', 'Solo kills']] =
scaler.fit_transform(df[['Solo minutesPlayed', 'Solo kills']])

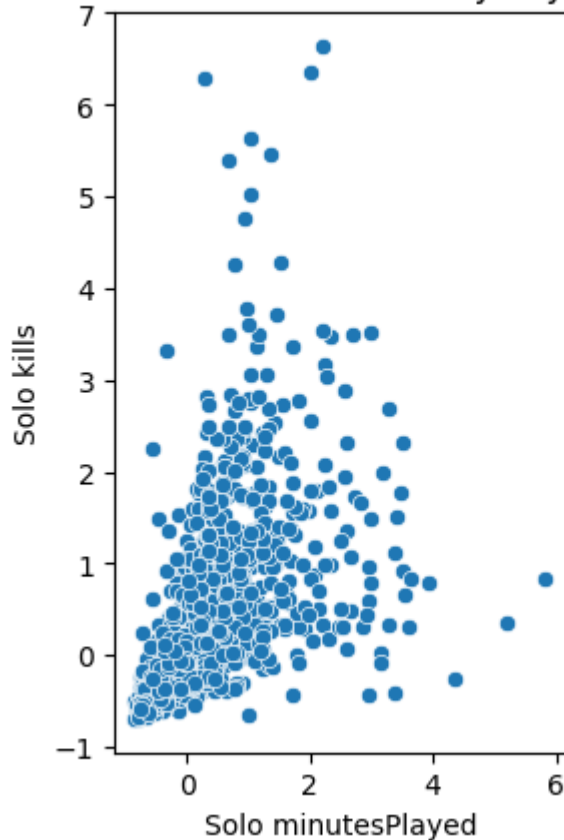
/tmp/ipykernel_65/3731571679.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy
df[['Solo minutesPlayed', 'Solo kills']] =
scaler.fit_transform(df[['Solo minutesPlayed', 'Solo kills']])

# Visualizar outliers
plt.subplot(1, 2, 1)
sb.scatterplot(x=df['Solo minutesPlayed'], y=df['Solo kills'])
plt.title('Relación entre Solo minutesPlayed y Solo kills')
plt.xlabel('Solo minutesPlayed')
plt.ylabel('Solo kills')
```

```
Text(0, 0.5, 'Solo kills')
```

Relación entre Solo minutesPlayed y Solo kills



Tratamiento de outliers para conocer qué factores influyen en la puntuación (Solo score) de un jugador

```
Q1 = df['Solo score'].quantile(0.25)
Q3 = df['Solo score'].quantile(0.75)
IQR = Q3 - Q1

# Filtra los datos que no son outliers
dfSinOutliers = df[~((df['Solo score'] < (Q1 - 1.5 * IQR)) | (df['Solo score'] > (Q3 + 1.5 * IQR)))]

# Mostrar los datos sin outliers
print(dfSinOutliers)
```

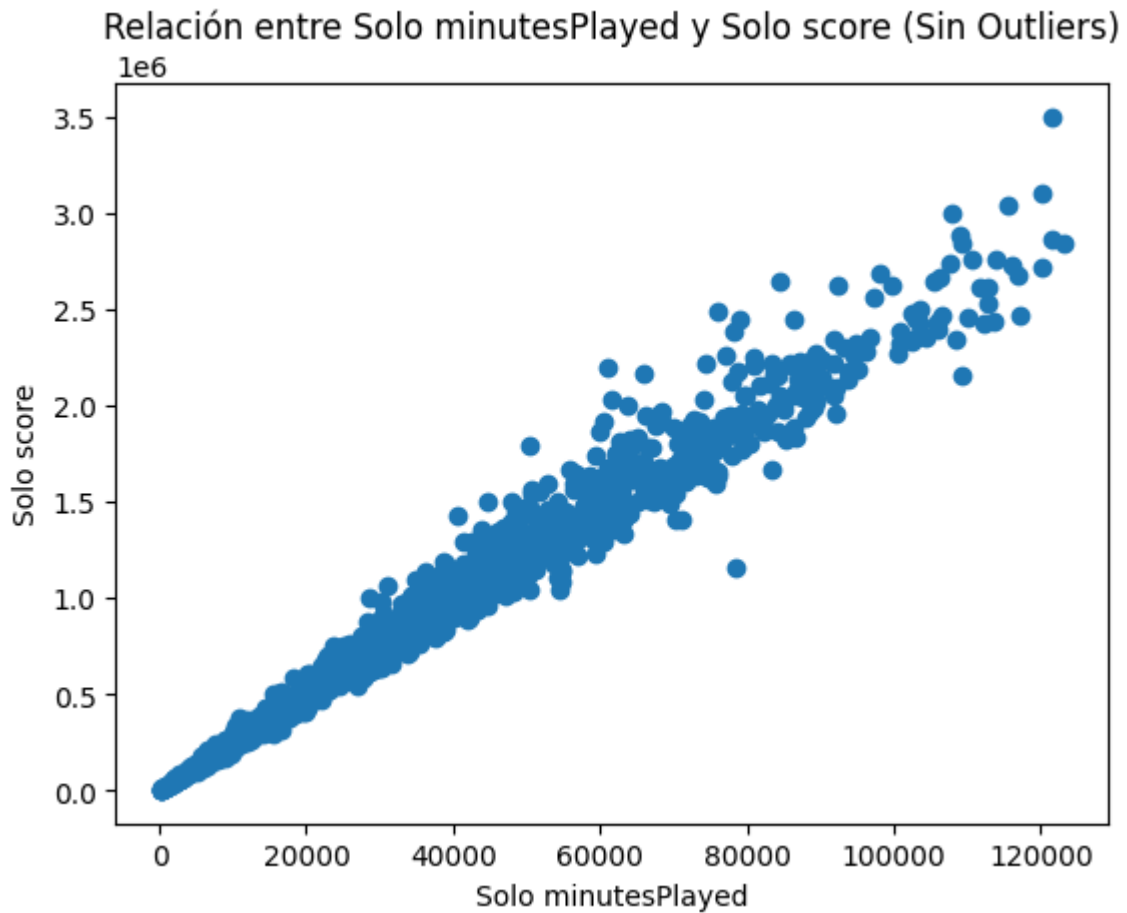
	Player	Solo score	Solo top1	Solo kd	Solo winRatio
0	Prospering	2476763	1828	4.37	18.0
1	BH nixxxay	439562	1694	8.71	29.1

3	Idk_Pi	752869	121	0.84	3.3
4	Twitch.GryphonRB	1136282	1327	6.32	30.0
5	Twitch Kayotica	2919037	1310	3.23	9.7
...
1430	im bloom	106294	26	2.03	3.5
1431	Twitch ka0zs_	803918	396	3.42	13.2
1432	slxyher	603939	310	1.69	5.1
1433	RD-Antony	747295	71	1.10	1.9
1434	O-HO-HO-HO!	1378689	161	1.68	2.3

	Solo matches	Solo kills	Solo minutesPlayed
0	10150	2.297725	1.072340
1	5817	2.262034	-0.541258
3	3687	-0.448978	-0.077769
4	4429	0.918150	0.011352
5	13438	2.528767	1.433994
...
1430	736	-0.577976	-0.735434
1431	3001	0.037834	-0.197114
1432	6035	0.101137	-0.357635
1433	3763	-0.363007	-0.108463
1434	7105	0.264507	0.512585

[1378 rows x 8 columns]

```
plt.scatter(df_sin_outliers['Solo minutesPlayed'],
df_sin_outliers['Solo score'])
plt.xlabel('Solo minutesPlayed')
plt.ylabel('Solo score')
plt.title('Relación entre Solo minutesPlayed y Solo score (Sin
Outliers)')
plt.show()
```



Sin los outliers, podemos observar una tendencia lineal clara y más concentrada entre los minutos jugados en solitario (SoloMinutesPlayed)) y la puntuación (SoloScore) obtenida. A medida que el número de minutos jugados en solitario aumenta, también lo hace la puntuación en solitario.

Created in Deepnote