

Consultas Avanzadas de Python

Python

Ventas Globales de Videojuegos

VENTAS GLOBALES DE VIDEO JUEGOS

```
# importación y lectura de datos
```

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
df = pd.read_csv('vg-sales.csv')
```

```
print("Head:\n", df.head())
print("\nTail:\n", df.tail())
print("\nSample:\n", df.sample(5, random_state=1))
print("\nDescripción:\n", df.describe(include='all'))
```

```
print("\nColumnas numéricas:", df.select_dtypes(include='number').columns.tolist())
```



VENTAS GLOBALES DE VIDEO JUEGOS

```
j]: # importación y lectura de datos
```

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

df = pd.read_csv('vgsales.csv')

print("Head:\n", df.head())
print("\nTail:\n", df.tail())
print("\nSample:\n", df.sample(5, random_state=1))
print("\nDescripción:\n", df.describe(include='all'))

print("\nColumnas numéricas:", df.select_dtypes(include='number').columns.tolist())
```

Head:

	Rank	Name	Platform	Year	Genre	Publisher
0	1	Wii Sports	Wii	2006.0	Sports	Nintendo
1	2	Super Mario Bros.	NES	1985.0	Platform	Nintendo
2	3	Mario Kart Wii	Wii	2008.0	Racing	Nintendo
3	4	Wii Sports Resort	Wii	2009.0	Sports	Nintendo
4	5	Pokemon Red/Pokemon Blue	GB	1996.0	Role-Playing	Nintendo

	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
0	41.49	29.02	3.77	8.46	82.74
1	29.08	3.58	6.81	0.77	40.24
2	15.85	12.88	3.79	3.31	35.82
3	15.75	11.01	3.28	2.96	33.00
4	11.27	8.89	10.22	1.00	31.37

Tail:

	Rank	Name	Platform
16593	16596	Woody Woodpecker in Crazy Castle 5	GBA
16594	16597	Men in Black II: Alien Escape	GC
16595	16598	SCORE International Baja 1000: The Official Game	PS2
16596	16599	Know How 2	DS
16597	16600	Spirits & Spells	GBA

	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales
16593	2002.0	Platform	Kemco	0.01	0.00	0.0
16594	2003.0	Shooter	Infogrames	0.01	0.00	0.0
16595	2008.0	Racing	Activision	0.00	0.00	0.0
16596	2010.0	Puzzle	7G//AMES	0.00	0.01	0.0
16597	2003.0	Platform	Wanadoo	0.01	0.00	0.0

	Other_Sales	Global_Sales
16593	0.0	0.01
16594	0.0	0.01
16595	0.0	0.01
16596	0.0	0.01
16597	0.0	0.01

Sample:

	Rank	Name	Platform	Year	Genre
4234	4236	Razor Freestyle Scooter	PS	1999.0	Sports
7119	7121	AC/DC LIVE: Rock Band Track Pack	X360	2008.0	Misc
106	107	Crash Bandicoot 3: Warped	PS	1998.0	Platform
5242	5244	Company of Heroes 2	PC	2013.0	Strategy
13547	13549	Advance Guardian Heroes	GBA	2004.0	Role-Playing

	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales
4234	Ubisoft	0.26	0.17	0.00	0.03
7119	MTV Games	0.21	0.00	0.00	0.02
106	Sony Computer Entertainment	3.68	1.75	1.42	0.28
5242	THQ	0.06	0.25	0.00	0.05
13547	Ubisoft	0.03	0.01	0.00	0.00

Global_Sales

4234	0.46
7119	0.23
106	7.13
5242	0.36
13547	0.04

Descripción:

	Rank	Name	Platform	Year
count	16598.000000	16598	16598	16327.000000
unique	NaN	11493	31	NaN
top	NaN	Need for Speed: Most Wanted	DS	NaN
freq	NaN	12	2163	NaN
mean	8300.605254	NaN	NaN	2006.406443
std	4791.853933	NaN	NaN	5.828981
min	1.000000	NaN	NaN	1980.000000
25%	4151.250000	NaN	NaN	2003.000000
50%	8300.500000	NaN	NaN	2007.000000
75%	12449.750000	NaN	NaN	2010.000000
max	16600.000000	NaN	NaN	2020.000000

	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales
count	16598	16540	16598.000000	16598.000000	16598.000000
unique	12	578	NaN	NaN	NaN
top	Action	Electronic Arts	NaN	NaN	NaN
freq	3316	1351	NaN	NaN	NaN
mean	NaN	NaN	0.264667	0.146652	0.077782
std	NaN	NaN	0.816683	0.505351	0.309291
min	NaN	NaN	0.000000	0.000000	0.000000
25%	NaN	NaN	0.000000	0.000000	0.000000
50%	NaN	NaN	0.080000	0.020000	0.000000
75%	NaN	NaN	0.240000	0.110000	0.040000
max	NaN	NaN	41.490000	29.020000	10.220000

	Other_Sales	Global_Sales
count	16598.000000	16598.000000
unique	NaN	NaN
top	NaN	NaN
freq	NaN	NaN
mean	0.048063	0.537441
std	0.188588	1.555028
min	0.000000	0.010000
25%	0.000000	0.060000
50%	0.010000	0.170000
75%	0.040000	0.470000
max	10.570000	82.740000

Columnas numéricas: ['Rank', 'Year', 'NA_Sales', 'EU_Sales', 'JP_Sales', 'Other_Sales', 'Global_Sales']

```
#Totales generales
total_juegos = df['Name'].nunique()
total_generos = df['Genre'].nunique()
total_plataformas = df['Platform'].nunique()

print(f" Total de videojuegos: {total_juegos}")
print(f" Total de géneros: {total_generos}")
print(f" Total de plataformas: {total_plataformas}")
```

Total de videojuegos: 11493
Total de géneros: 12
Total de plataformas: 31

Ventas por año y por región

```
ventas_por_anio = df.groupby('Year')[['NA_Sales', 'EU_Sales', 'JP_Sales', 'Other_Sales', 'Global_Sales']].sum()
print("\nVentas por año:\n", ventas_por_anio)
```

Ventas por año:					
	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
Year					
1980.0	10.59	0.67	0.00	0.12	11.38
1981.0	33.40	1.96	0.00	0.32	35.77
1982.0	26.92	1.65	0.00	0.31	28.86
1983.0	7.76	0.80	8.10	0.14	16.79
1984.0	33.28	2.10	14.27	0.70	50.36
1985.0	33.73	4.74	14.56	0.92	53.94
1986.0	12.50	2.84	19.81	1.93	37.07
1987.0	8.46	1.41	11.63	0.20	21.74
1988.0	23.87	6.59	15.76	0.99	47.22
1989.0	45.15	8.44	18.36	1.50	73.45
1990.0	25.46	7.63	14.88	1.40	49.39
1991.0	12.76	3.95	14.78	0.74	32.23
1992.0	33.87	11.71	28.91	1.65	76.16
1993.0	15.12	4.65	25.33	0.89	45.98
1994.0	28.15	14.88	33.99	2.20	79.17
1995.0	24.82	14.90	45.75	2.64	88.11
1996.0	86.76	47.26	57.44	7.69	199.15
1997.0	94.75	48.32	48.87	9.13	200.98
1998.0	128.36	66.90	50.04	11.03	256.47
1999.0	126.06	62.67	52.34	10.05	251.27
2000.0	94.49	52.75	42.77	11.62	201.56
2001.0	173.98	94.89	39.86	22.76	331.47
2002.0	216.19	109.74	41.76	27.28	395.52
2003.0	193.59	103.81	34.20	26.01	357.85
2004.0	222.59	107.32	41.65	47.29	419.31
2005.0	242.61	121.94	54.28	40.58	459.94
2006.0	263.12	129.24	73.73	54.43	521.04
2007.0	312.05	160.50	60.29	77.60	611.13
2008.0	351.44	184.40	60.26	82.39	678.90
2009.0	338.85	191.59	61.89	74.77	667.30
2010.0	304.24	176.73	59.49	59.90	600.45
2011.0	241.06	167.44	53.04	54.39	515.99
2012.0	154.96	118.78	51.74	37.82	363.54
2013.0	154.77	125.80	47.59	39.82	368.11
2014.0	131.97	125.65	39.46	40.02	337.05
2015.0	102.82	97.71	33.72	30.01	264.44
2016.0	22.66	26.76	13.70	7.75	70.93
2017.0	0.00	0.00	0.05	0.00	0.05
2020.0	0.27	0.00	0.00	0.02	0.29

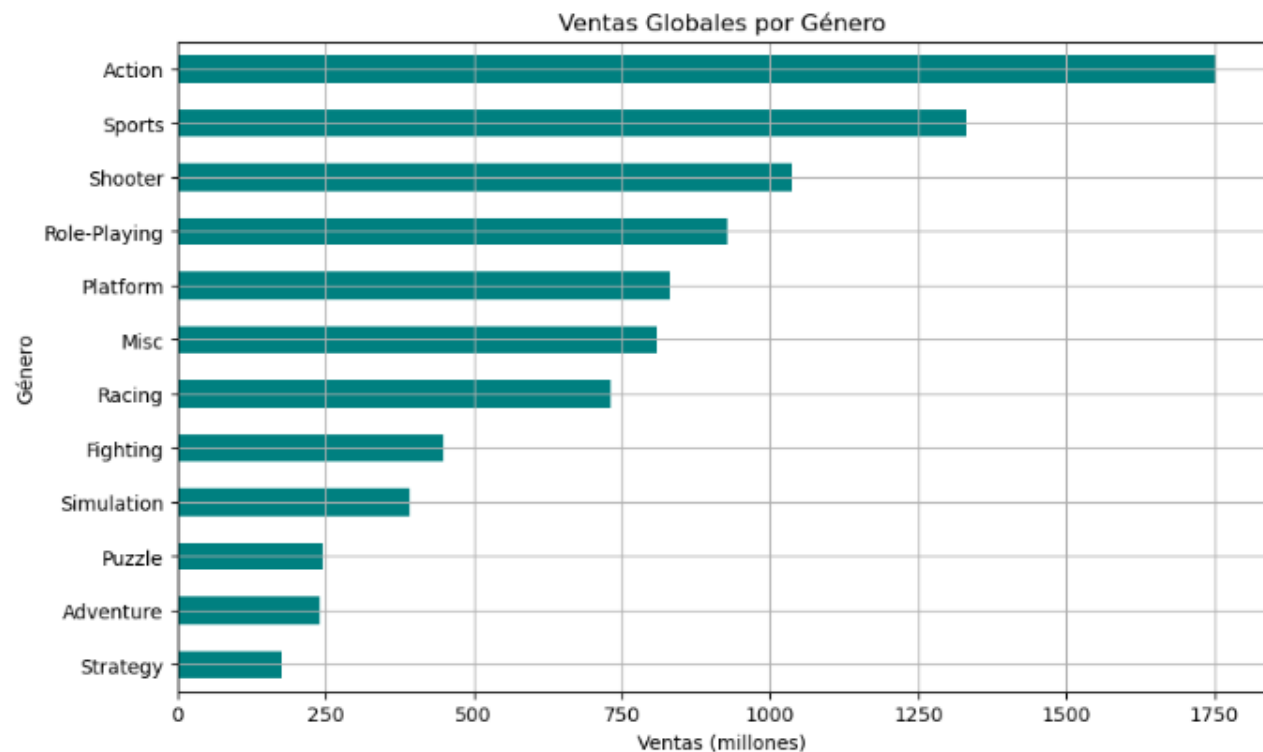
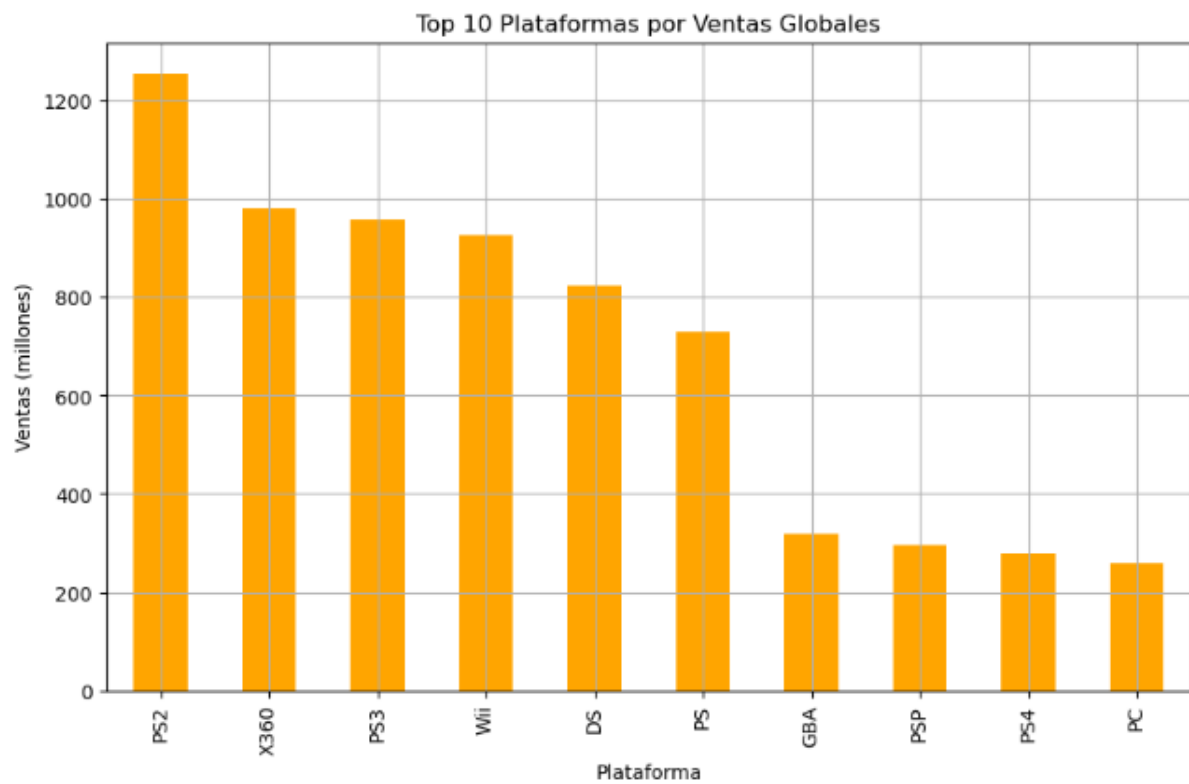

```
# Insights adicionales graficos
```

```
# Top 10 plataformas con mayores ventas
```

```
top_plataformas = df.groupby('Platform')['Global_Sales'].sum().sort_values(ascending=False).head(10)
top_plataformas.plot(kind='bar', figsize=(10,6), color='orange')
plt.title('Top 10 Plataformas por Ventas Globales')
plt.xlabel('Plataforma')
plt.ylabel('Ventas (millones)')
plt.grid(True)
plt.show()
```

```
# Distribución de géneros por total de venta
```

```
genero_ventas = df.groupby('Genre')['Global_Sales'].sum().sort_values(ascending=True)
genero_ventas.plot(kind='barh', figsize=(10,6), color='teal')
plt.title('Ventas Globales por Género')
plt.xlabel('Ventas (millones)')
plt.ylabel('Género')
plt.grid(True)
plt.show()
```



```
# crecimiento en últimos 4 años (2017-2020)
ultimos_4_anios = df[df['Year'] >= 2017]

# Género con mayor y menor crecimiento
crecimiento_genero = ultimos_4_anios.groupby('Genre')['Global_Sales'].sum().sort_values(ascending=False)
print(" Crecimiento por Género:\n", crecimiento_genero)

# Plataforma con mayor y menor crecimiento
crecimiento_plataforma = ultimos_4_anios.groupby('Platform')['Global_Sales'].sum().sort_values(ascending=False)
print("\n Crecimiento por Plataforma:\n", crecimiento_plataforma)

# Juego con mayor y menor crecimiento
crecimiento_juego = ultimos_4_anios.groupby('Name')['Global_Sales'].sum().sort_values(ascending=False)
print("\n Videojuegos más vendidos en los últimos 4 años:\n", crecimiento_juego.head(5))
print("\n Videojuegos con menos ventas en los últimos 4 años:\n", crecimiento_juego.tail(5))
```

Crecimiento por Género:

Genre	
Simulation	0.29
Role-Playing	0.04
Action	0.01

Name: Global_Sales, dtype: float64

Crecimiento por Plataforma:

Platform	
DS	0.29
PS4	0.03
PSV	0.02

Name: Global_Sales, dtype: float64

Videojuegos más vendidos en los últimos 4 años:

Name	
Imagine: Makeup Artist	0.29
Phantasy Star Online 2 Episode 4: Deluxe Package	0.04
Brothers Conflict: Precious Baby	0.01

Name: Global_Sales, dtype: float64

Videojuegos con menos ventas en los últimos 4 años:

Name	
Imagine: Makeup Artist	0.29
Phantasy Star Online 2 Episode 4: Deluxe Package	0.04
Brothers Conflict: Precious Baby	0.01

Name: Global_Sales, dtype: float64

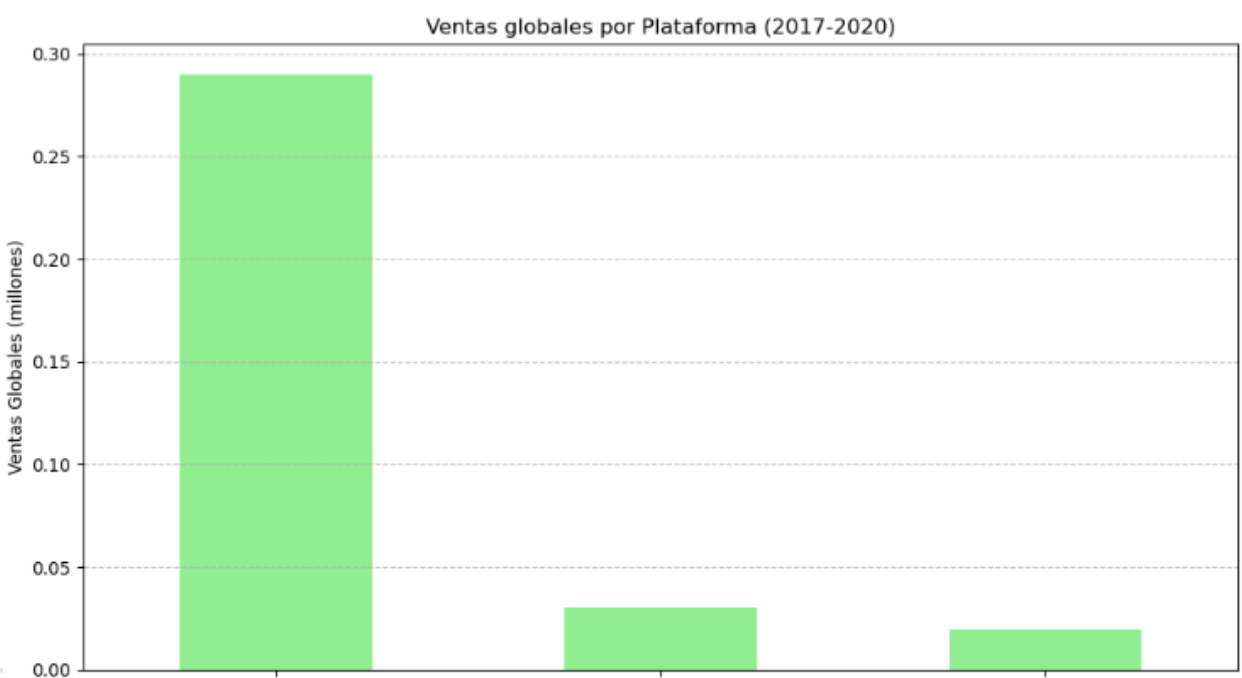
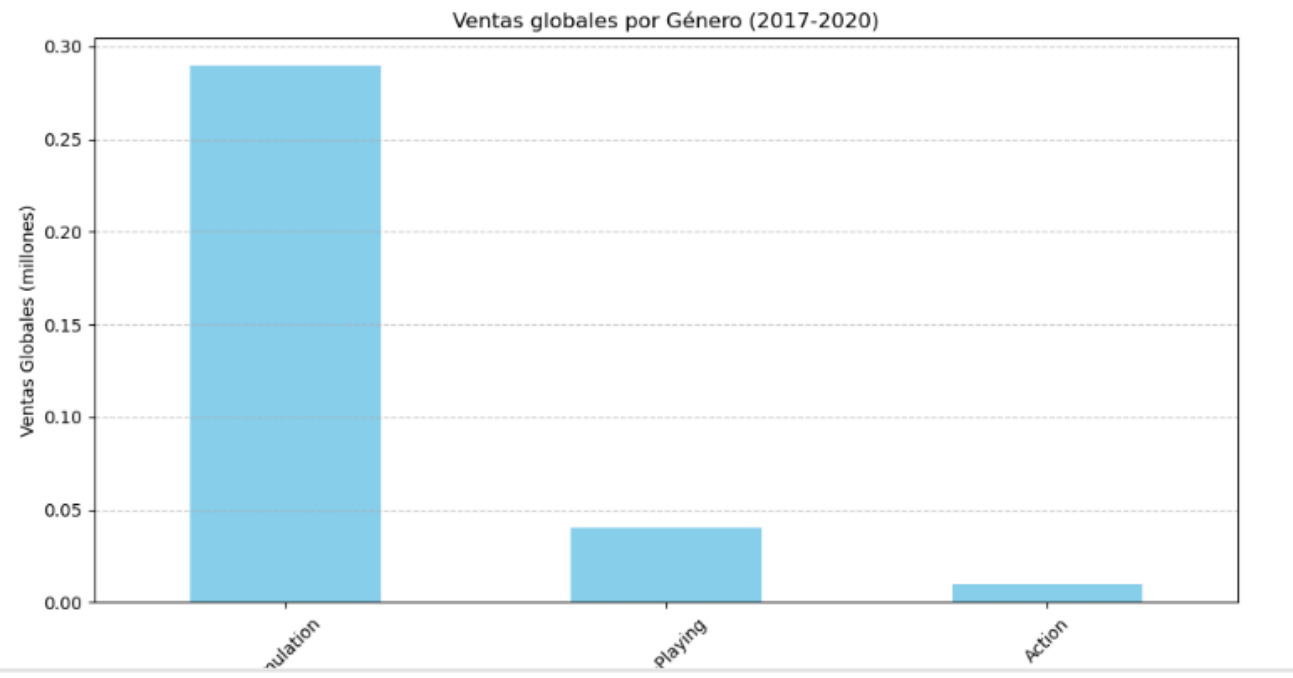
```
ultimos_4_anios = df[df['Year'] >= 2017]

# Agrupar y sumar ventas globales por género
crecimiento_genero = ultimos_4_anios.groupby('Genre')['Global_Sales'].sum().sort_values(ascending=False)

# Agrupar y sumar ventas globales por plataforma
crecimiento_plataforma = ultimos_4_anios.groupby('Platform')['Global_Sales'].sum().sort_values(ascending=False)

# Gráfico de barras para crecimiento por género
plt.figure(figsize=(10,6))
crecimiento_genero.plot(kind='bar', color='skyblue')
plt.title('Ventas globales por Género (2017-2020)')
plt.ylabel('Ventas Globales (millones)')
plt.xlabel('Género')
plt.xticks(rotation=45)
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.tight_layout()
plt.show()

# Gráfico de barras para crecimiento por plataforma
plt.figure(figsize=(10,6))
crecimiento_plataforma.plot(kind='bar', color='lightgreen')
plt.title('Ventas globales por Plataforma (2017-2020)')
plt.ylabel('Ventas Globales (millones)')
plt.xlabel('Plataforma')
plt.xticks(rotation=45)
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.tight_layout()
plt.show()
```




```
#Oportunidad de mercado

# Oportunida 1 Géneros poco explotados pero con potencial
media_ventas = df.groupby('Genre')['Global_Sales'].mean().sort_values(ascending=False)
```

```
# Gráfico horizontal
plt.figure(figsize=(10,6))
media_ventas.plot(kind='barh', color='salmon')
plt.title('Promedio de ventas por juego según Género')
plt.xlabel('Promedio de Ventas Globales (millones)')
plt.ylabel('Género')
plt.grid(axis='x', linestyle='--', alpha=0.7)
plt.tight_layout()
plt.show()
```

```
# Oportunidad 2 Plataformas activas en años recientes
```

```
recientes = df[df['Year'] >= 2015]
plataformas_recientes = recientes.groupby('Platform')['Global_Sales'].sum().sort_values(ascending=False)
```

```
# Gráfico de pastel
plt.figure(figsize=(8,8))
plataformas_recientes.head(6).plot(kind='bar', color='orange')
plt.title('Participación de plataformas con ventas (2015-2020)')
plt.ylabel('')
plt.tight_layout()
plt.show()
```

