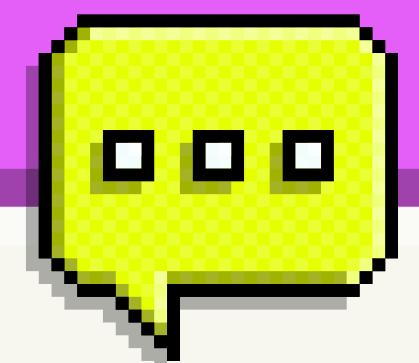
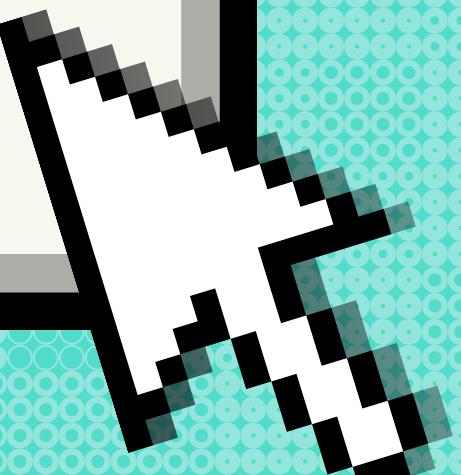


Equipo 30



VIDEOJUEGOS



Proyecto final

índice

01. Introducción
02. Objetivos
03. Planteamiento del problema
04. Resultados
05. Conclusiones

¡Comenzamos!

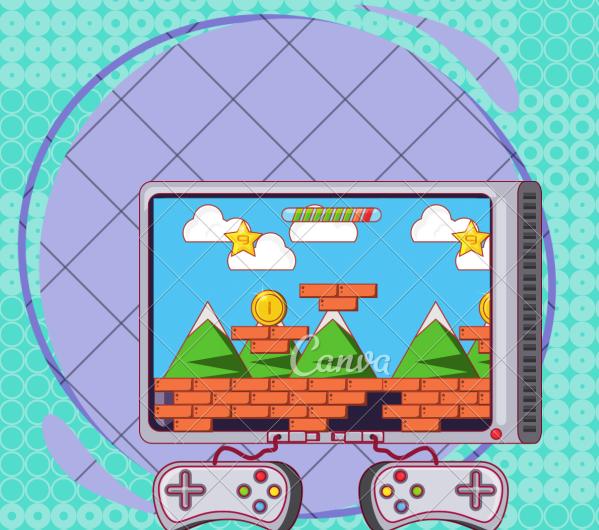
Introducción

Análisis de Videojuegos: Factores Clave del Éxito

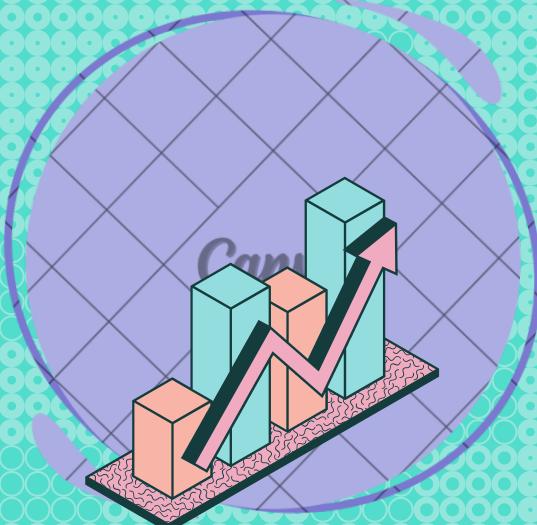
Popularidad mundial de los videojuegos como entretenimiento.



Impacto en las ventas globales y regionales.



Factores del éxito: plataforma, género, críticas y ventas

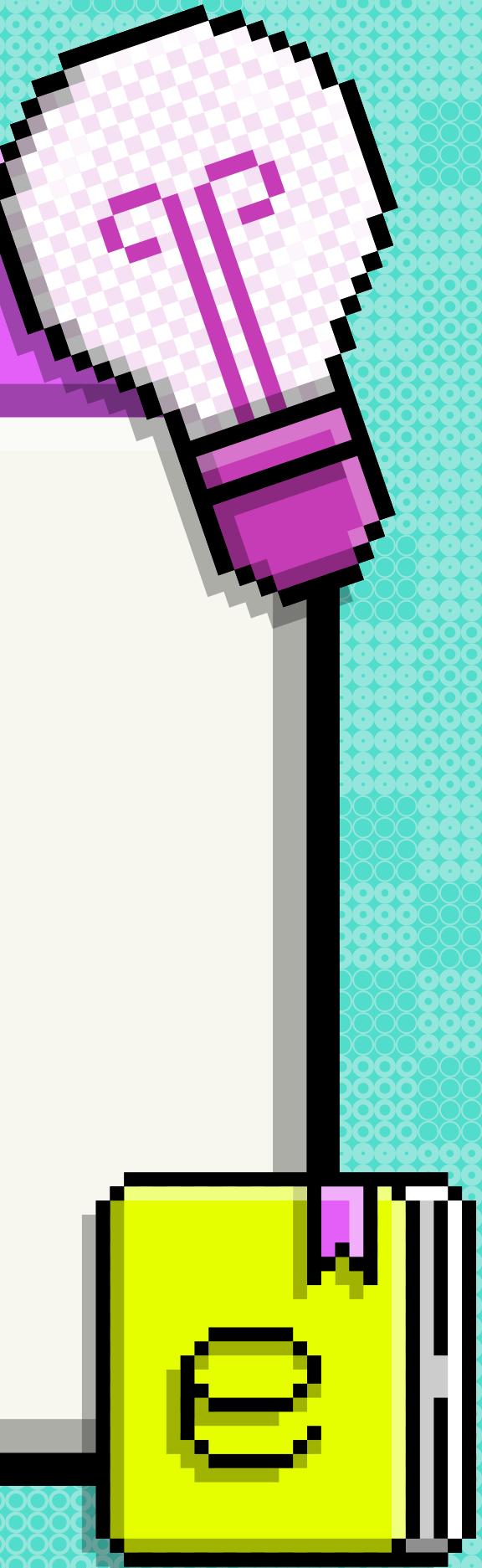
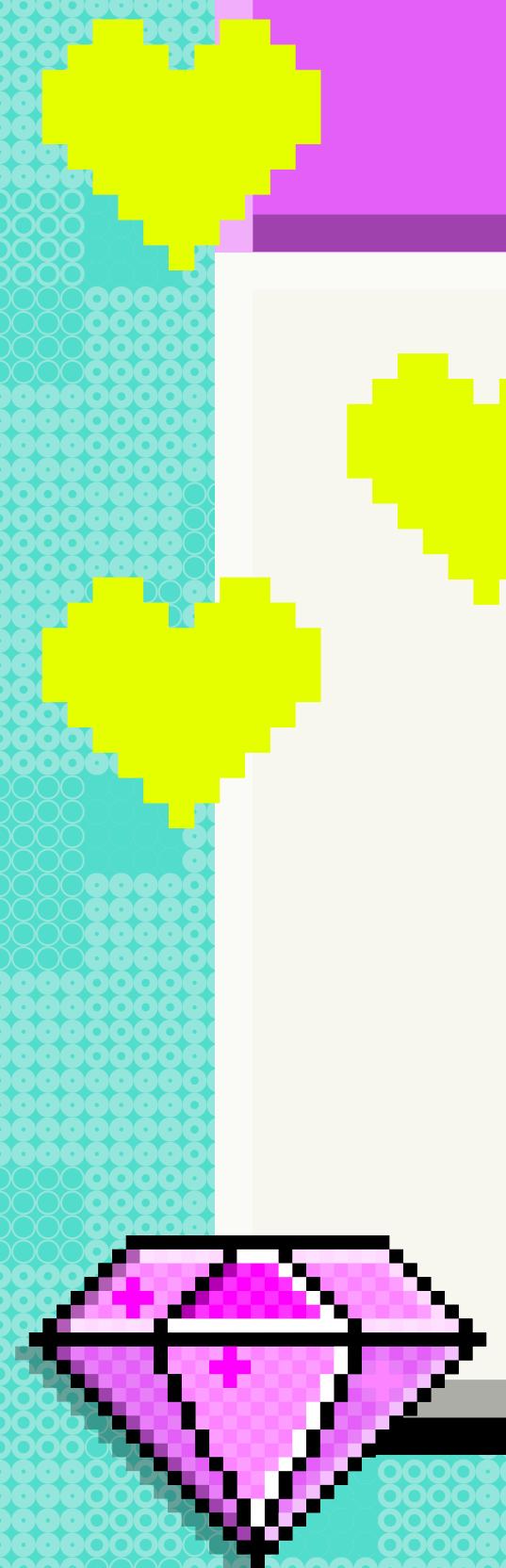


Estudio de datos globales: plataforma, género y calificaciones.



Patrones que explican el éxito de ciertos videojuegos.

Objetivos



Generales

- Analizar la relación entre ventas, plataformas, géneros y puntuaciones para identificar los factores que influyen en el éxito comercial de los videojuegos.

Específicos

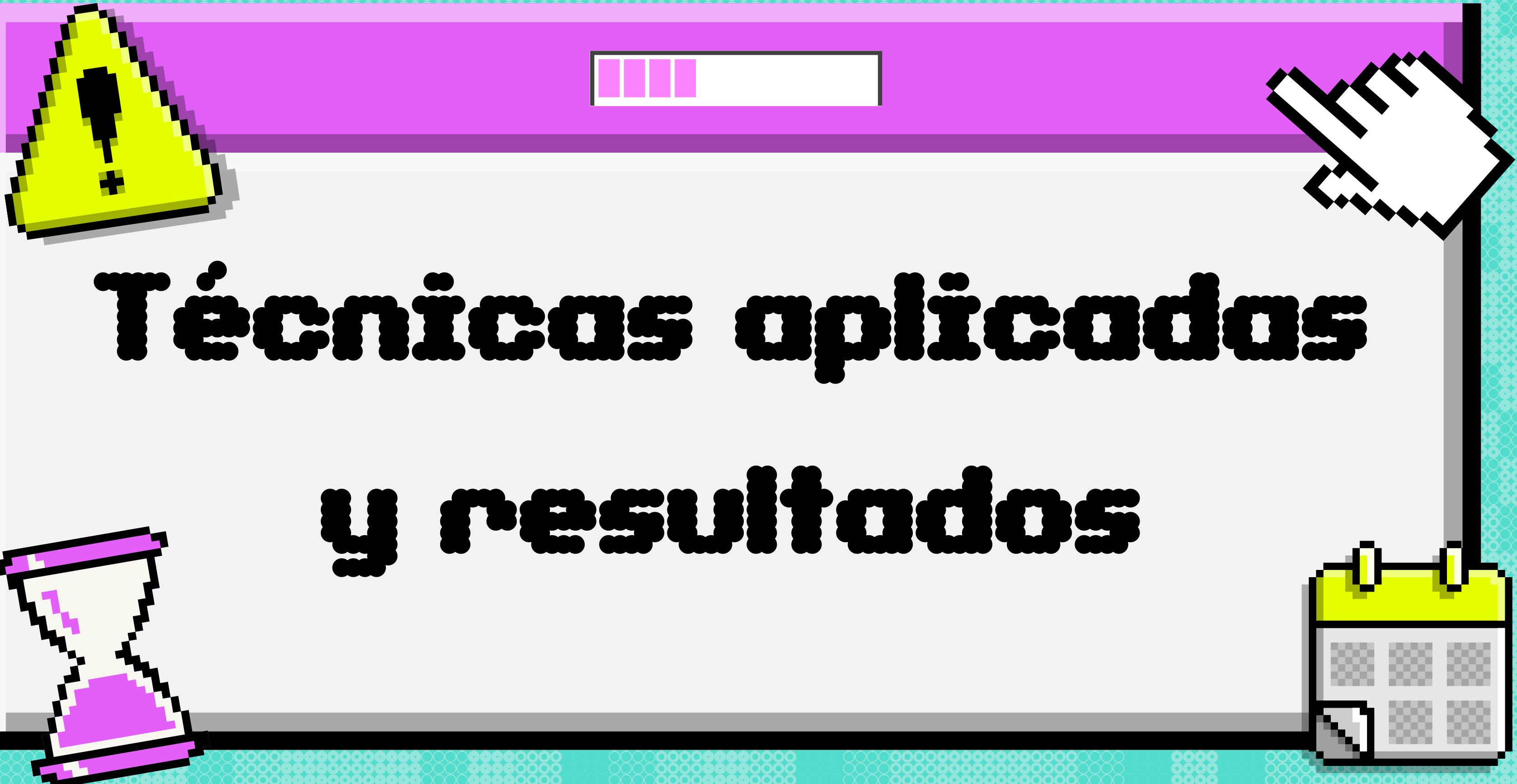
- Preparar los datos de ventas, plataformas, géneros y puntuaciones.
- Estudiar la relación entre ventas globales y regionales.
- Comparar el éxito de diferentes plataformas en ventas.
- Evaluar el impacto de las puntuaciones en las ventas.
- Identificar tendencias de ventas por género y año de lanzamiento.

Planteamiento del problema

La industria de los videojuegos busca identificar qué factores determinan el éxito comercial.

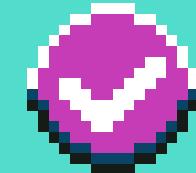
Aunque la plataforma y el género son claves, otros elementos como las calificaciones, preferencias regionales y el año de lanzamiento también influyen.

Existe una desconexión entre críticas y ventas, ya que los juegos más vendidos no siempre son los mejor valorados. Este proyecto analizará datos para descubrir qué características impactan más en las ventas globales y regionales.

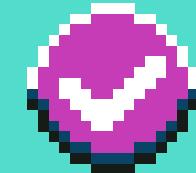


Técnicos aplicados y resultados

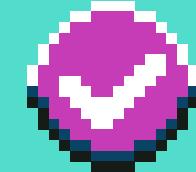
Limpieza de la BD



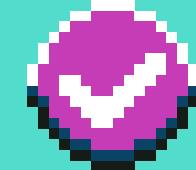
01. Importar bibliotecas



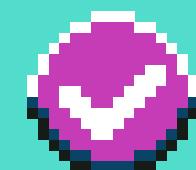
02. Cargar datos (csv, api)



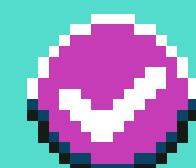
03. Hacer la exploración



04. Limpiar los datos



05. Transformar los datos



06. Dar formato a la BD



#1. Bibliotecas

Import requests / pandas / numpy / pprint / display

#2. Datos

df = pd.read_csv('ruta')

response = requests.get(url)

#3. Exploración

df.head / df.shape / df.isnull().sum() / df.nunique() / df.dtypes /
df.duplicated().sum() / df.columns

#4. Limpiar Datos

df.dropna(axis=1, how='all').reset_index(drop=True)
df.drop_duplicates()

df.drop(columns=['columna'])

#5. Transformar

.astype('Int64')

.apply(pd.to_numeric, errors='coerce')

.fillna() #fechas aleatoria en rango, promedios, 0

.str.title/upper/replace/strip()

#6. Formato

.rename #columnas a través de diccionario

df['platform'].map(mapping) #aplicar diccionario con + descripción

.reset_index(drop=True)

¿Cuántos videojuegos tienen información incompleta o nula en términos de ventas o puntuaciones?

```
print('Cantidad de datos nulos por columna: ')
display(df.isnull().sum().to_frame(name='Nulos'))
print("-" * 50)
```

Información general del DataFrame	
Cantidad de filas y columnas: (16928, 17)	
Cantidad de datos nulos por columna:	
Nulos	
index	0
Name	2
Platform	0
Year_of_Release	273
Genre	2
Publisher	55
NA_Sales	0
EU_Sales	0
JP_Sales	0
Other_Sales	0
Global_Sales	0
Critic_Score	8668
Critic_Count	8668
User_Score	6769
User_Count	9210
Developer	6688
Rating	6836

La función isnull() busca los valores nulos. La aplicamos en un principio para la limpieza de datos, nos muestra la cantidad de datos nulos en cada columna.

En ventas no tenemos valores nulos , en el caso de puntuaciones obtenemos bastantes valores NaNs. Solución:
Usarfillna() reemplazando los valores de las puntuaciones con valores generados aleatoriamente en base al promedio de los datos de la columna correspondiente.

¿Cuál es el rango de ventas en las diferentes regiones (Norteamérica, Europa, Japón)?

El código calcula el rango de ventas de videojuegos en varias regiones [Norteamérica, Europa, Japón y otras] tomando la diferencia entre las ventas máximas y mínimas de cada área. Utiliza un diccionario para asociar las columnas de ventas con sus nombres y luego imprime los resultados por región.

Hay variaciones significativas en las ventas de videojuegos entre las diferentes regiones. La cobertura de ventas en Norteamérica y Europa es particularmente alta en comparación con otras regiones.

```
1 #Definimos un diccionario que contenga las columnas por region y su el nombre de la region
2 regions = {
3     'na_sales': 'Norteamérica',
4     'eu_sales': 'Europa',
5     'jp_sales': 'Japón',
6     'other_sales': 'otras regiones'
7 }
8 #Creamos un for que itere cada region dentro del diccionario
9 #Para sacar el rango de ventas debemos buscar la diferencia entre el valor maximo y minimo de las ventas. Con cada iteracion se busca el valor maximo con la funcion ".max()" y hacemos una resta con el valor minimo con la funcion ".min()", este proceso se repite por cada region que exista.
10 for region, name in regions.items():
11     sale_range = df_clean[region].max() - df_clean[region].min()
12     print("—" * 20)
13     print(f'Rango de ventas en {name}: {sale_range} millones')
14     print("—" * 20)
15
```

```
Rango de ventas en Norteamérica: 41.36 millones
-----
Rango de ventas en Europa: 28.96 millones
-----
Rango de ventas en Japón: 10.22 millones
-----
Rango de ventas en otras regiones: 10.57 millones
-----
```

```
[ ] 1 #Para sacar el rango de ventas a nivel mundial podemos repetir el mismo proceso.
2 global_sales_range = df_clean['global_sales'].max() - df_clean['global_sales'].min()
3 print("—" * 20)
4 print(f'Rango de ventas globales: {global_sales_range} millones')
5 print("—" * 20)
```

```
Rango de ventas globales: 82.52 millones
-----
```

¿Qué plataformas generan mayores ventas globales?

```
[44] #Primero usamos el metodo "groupby()" para ordenar las ventas globales por la columna de las plataformas  
sales_by_platform = df_clean.groupby('platform')['global_sales'].sum()  
  
#Usamos la nueva variable y con "sort_values()" ordenamos los valores de mayor a menor.  
sales_by_platform_sorted = sales_by_platform.sort_values(ascending=False)  
  
#Mostramos con "head()" el top 10 de las plataformas con mayores ventas  
top_platforms = sales_by_platform_sorted.head(10)  
print("—" * 20)  
print(top_platforms)  
print("—" * 20)
```

```
-----  
platform  
PlayStation 2      1283.08  
Xbox 360          988.36  
PlayStation 3      962.71  
Nintendo Wii       937.27  
Nintendo DS        813.86  
PlayStation         759.88  
Game Boy Advance   327.11  
PlayStation 4       324.48  
PlayStation Portable 298.02  
Nintendo 3DS       271.21  
Name: global_sales, dtype: float64  
-----
```

Para agrupar los datos de ventas por plataforma se utilizó `groupby()`, además de `sum()` para obtener el total de ventas globales, posteriormente `sort_values(ascending=False)` para ordenar las plataformas de mayor a menor en función a lo anterior.

PlayStation 2 es la consola con mayores ingresos globales en la historia de los videojuegos, seguida de Xbox 360 y PlayStation 3.

¿Qué géneros de videojuegos son los que mayor se venden?

Se realizó un análisis de las ventas globales totales de cada género de videojuego. Se agruparon los datos por género y luego se sumaron las ventas globales para cada uno de ellos.

Los géneros más vendidos son Action, Sports y Shooter. Destacando la popularidad de los juegos dinámicos y competitivos a nivel mundial, además ofrecen experiencias intensas y variadas destacan entre su popularidad.

```
[ ] #Primero usamos el metodo "groupby()" para ordenar las ventas globales por la columna de sales_by_gender = df_clean.groupby('genre')['global_sales'].sum()
```

```
usamos la nueva variable y con "sort_values()" ordenamos los valores de mayor a menor.  
sales_by_gender_sorted = sales_by_gender.sort_values(ascending=False)
```

```
#Mostramos con "head()" el top 10 de los géneros con mayores ventas
```

```
top_genders = sales_by_gender_sorted.head(10)  
print("... " * 20)  
print(top_genders)  
print("... " * 20)
```

```
...  
genre  
Action      1771.73  
Sports       1350.61  
Shooter      1086.67  
Role-Playing  958.62  
Platform     850.73  
Misc         830.19  
Racing        757.92  
Fighting      467.91  
Simulation    394.12  
Puzzle        243.65  
Name: global_sales, dtype: float64  
...
```

¿Cuál es la distribución de los videojuegos por género y plataforma?

```
5 distribution_genre_platform = df_clean.groupby(['genre', 'platform'])['name'].count().reset_index(name='videogames')
6
7
8 # Mostramos la tabla de distribución combinada
9 print(distribution_genre_platform)
10
```

```
genre      platform  videogames
0  Action      Atari 2600       63
1  Action        Game Boy        6
2  Action  Game Boy Advance     171
3  Action      GameCube      102
4  Action     Nintendo 3DS     193
...
288 Strategy    Sega Saturn       18
289 Strategy   WonderSwan        2
290 Strategy        Xbox        21
291 Strategy     Xbox 360        29
292 Strategy      Xbox One        3
[293 rows x 3 columns]
```

```
1 #Ahora ordenamos de forma descendente con "sort_values()" para poder visualizar
  por genero y plataforma los 10 mayores usamos "head(10)".
2 top_genre_platform = distribution_genre_platform.sort_values(by='videogames',
  ascending=False).head(10)
3
4 # Mandamos a imprimir el top 10
5 print(top_genre_platform)
```

```
genre      platform  videogames
260    Sports  PlayStation 2      402
80      Misc   Nintendo DS      392
15    Action  PlayStation 3      383
14    Action  PlayStation 2      350
6     Action   Nintendo DS      341
22    Action     Xbox 360      329
83      Misc   Nintendo Wii      288
229   Simulation   Nintendo DS      285
256    Sports   Nintendo Wii      262
31 Adventure   Nintendo DS      240
```

Se realizó un análisis de cómo se agrupan los videojuegos según su género y plataforma. Iniciamos cuantificando los videojuegos según su tipo, plataforma, ambos y finalmente mostramos un top 10.

Los resultados muestran que Action, Sports, y Misc, son los géneros con más videojuegos. En términos de plataformas, las más populares son PlayStation 2, Nintendo DS, y PlayStation 3 [1,359].

```
#Primero hacemos un filtro para obtener las columnas que vamos a usar
df_scores_sales = df_clean[['name', 'critic_score', 'user_score', 'global_sales']]
#Con el metodo "sort_values()" hacemos un ordenamiento de forma descendente
df_descending = df_scores_sales.sort_values(by=['critic_score', 'user_score'], ascending=False)
#Mostramos los primeros 10 juegos
df_descending.head(10)
```

		name	critic_score	user_score	global_sales
5349		Soulcalibur	98	8.8	0.34
51		Grand Theft Auto Iv	98	7.9	11.01
227		Tony Hawk'S Pro Skater 2	98	7.7	4.68
57		Grand Theft Auto Iv	98	7.5	10.50
517		Metroid Prime	97	9.3	2.82
97		Super Mario Galaxy 2	97	9.1	7.51
49		Super Mario Galaxy	97	8.9	11.35
129		Halo: Combat Evolved	97	8.6	6.43
38		Grand Theft Auto Iii	97	8.5	13.10
42		Grand Theft Auto V	97	8.3	12.61

```
[ ] #Mostramos con "tail()" los ultimos datos
df_descending.tail(10)
```

		name	critic_score	user_score	global_sales
1049		Deal Or No Deal	20	1.9	1.70
11442		Ninjabread Man	20	1.7	0.08
7218		Nickelodeon Party Blast	19	4.3	0.22
10015		Rugby 15	19	3.4	0.11
9105		Chicken Shoot	19	3.1	0.14
15343		Balls Of Fury	19	2.8	0.02
12131		Anubis Ii	19	2.7	0.07
12822		Ride To Hell	19	1.3	0.06
10662		Leisure Suit Larry: Box Office Bust	17	1.7	0.10
13689		Ride To Hell	13	1.0	0.04

¿Qué relación existe entre las puntuaciones de críticos y usuarios y las ventas globales?

Se filtraron las columnas clave [nombre del juego, puntuaciones y ventas globales] y se ordenaron los juegos por puntuaciones más altas. Luego, se imprimieron los 10 mejores para observar la relación entre calificaciones y ventas.

Usando tail(), se mostraron los 10 juegos peor calificados, que a pesar de sus bajas puntuaciones, tienen algunas ventas, aunque mucho menores.

Juegos como Grand Theft Auto V y Super Mario Galaxy tienen calificaciones excelentes

Juegos como Ride to Hell y Ninjabread Man tienen puntuaciones extremadamente bajas

¿Existen diferencias significativas en las ventas según el año de lanzamiento de los videojuegos?

Se filtran las columnas relevantes para el análisis, seleccionando el año de lanzamiento y las ventas globales. Luego, se ordenan los datos de forma descendente utilizando `sort_values()`. Finalmente, se muestran los 10 años con mayores ventas.

Los datos muestran que algunos años, como 2006 y 1985, tuvieron ventas significativamente más altas que otros.

```
▶ #Primero hacemos un filtro para obtener las columnas que vamos a usar  
sales_by_year = df_clean[['year_of_release', 'global_sales']]  
#Con el metodo "sort_values()" hacemos un ordenamiento de forma descendente  
sales_by_year_desc = sales_by_year.sort_values('global_sales', ascending=False)  
#Mostramos los años que mayores ventas tuvieron  
sales_by_year_desc.head(10)
```

	year_of_release	global_sales
0	2006	82.53
1	1985	40.24
2	2008	35.52
3	2009	32.77
4	1996	31.37
5	1989	30.26
6	2006	29.80
7	2006	28.92
8	2009	28.32
9	1984	28.31



1. Rendimiento de ventas por plataforma

Las plataformas de videojuegos varían en su éxito global.

PlayStation 2
Xbox 360
PlayStation 3

Refleja la preferencia por estas consolas a nivel mundial.



2. Popularidad de géneros de videojuegos

Tipo de contenido que atrae más a los jugadores.

Los géneros de videojuegos que más venden son Acción, Deportes y Shooters.

RESULTADOS

3. Distribución por género y plataforma

Concentración significativa de juegos en estas consolas.

La mayoría de los videojuegos se agrupan en géneros como Acción y Deportes, distribuyéndose principalmente en plataformas como PlayStation 2 y Nintendo DS.



4. Relación entre puntuaciones y ventas

La relación entre las puntuaciones de críticos y usuarios y las ventas es débil, lo que indica que las ventas no dependen principalmente de la crítica, sino de otros factores como el marketing o la popularidad de la plataforma.



Conclusiones

- Plataformas más vendidas: PlayStation 2, Xbox 360 y PlayStation 3 lideran en ventas globales.
- Géneros más exitosos: Acción, deportes y shooters son los géneros con mayores ventas.
- Distribución de videojuegos: PlayStation 2 y Nintendo DS tienen la mayor cantidad de títulos en varios géneros clave.
- Puntuaciones y ventas: Juegos con buenas críticas suelen vender más, pero hay excepciones donde títulos mal calificados logran ventas significativas.

MUCHAS GRACIAS

