

# Proyecto ciencia de datos.

Arquieta Cortez Mauricio Marcial  
Ramírez Peral Jesús Eduardo.  
*Universidad Nacional Autónoma de México.*

4 de junio de 2023.

## Resumen.

Se analizó un conjunto de datos con información de la venta de videojuegos de varias plataformas entre los años de 1980 y 2016. Haciendo énfasis en los videojuegos más vendidos, así como el genero y la plataforma en la que fueron lanzados, además se analizó que empresas dominaron el mercado con sus lanzamientos en los años con mayores lanzamientos, de igual manera cuál fue el juego que se lanzo en más plataformas. Por otra parte, se analizaron la venta por género, descubriendo que el genero mas vendido a nivel global fue el de Acción, para las ventas en Japón el genero más vendido fue el de rol. De igual manera se encontró que la plataforma que más ventas tuvo fue la PS2 y la empresa que mas ventas tuvo fue Nintendo.

## Objetivos.

- Realizar un análisis en base a la venta de videojuegos.
- Este data set tiene ventas de 3 regiones y ventas del mundo, el nombre del juego, el año de lanzamiento, la industria de la consola, el Rank, el año de lanzamiento y el género del juego.
- Ver cuál es la plataforma que vende más juegos.

## Hipótesis.

A nivel consumismo esperamos que las ventas globales se parezcan a las ventas de las 3 regiones (Norteamérica, Europa y Japón).

También esperamos que los videojuegos en tengan una tendencia a aumentar a medida que avanzan los años.

Por del genero esperamos que los distintos países jueguen cosas similares en este aspecto así como el uso de consolas.

Además suponemos que sea cual sea el año con mayor lanzamiento de juegos, este año tiene que tener juegos que se encuentren en al menos en el top 100 del Rank.

## Introduccion.

El data set a utilizar es uno de venta de videojuegos.

Las librerías a usar son pandas, numpy, matplotlib y seaborn para el análisis.

Primeramente realizamos un head para ver que el contenido del dataset. Después ordenamos de acuerdo al Rank solo para ver los primeros juegos que encabezan este dataset.

Con esto utilizamos el comando ***data.info()***, esto para ver con que tipos de variables estamos tratando, numero de columnas y filas, además de ver si todas las filas de las distintas columnas tenían un valor asignado.

Seguido de eso utilizamos ***data.describe(include='all').T***, para que nos den un análisis estadístico y ver con que variables podemos trabajar en base a esto. Al parecer podemos trabajar con el año de lanzamiento, tiene sentido, pues son las únicas variables con números que tienen estadística por así decirlo. Asi podemos usar los parámetros que nos dan y hacer un

histograma de la cantidad de juegos lanzados por año.

Siguiendo con esto nos preguntamos **¿Qué año ha lanzado mayores juegos hasta el 2016?** Usando `data['Year'].value_counts()` podemos ver fácilmente de que año se trata, en nuestro caso fue el **2009**.

Como habíamos dicho en la hipótesis esperamos que algunos juegos que estén en el top 100 se encuentren en este año. Programamos para que nos arrojen los juegos que se encuentran en el top 100 de este año con un total de 7 juegos, y 2 de ellos se encuentran en el top 10.

Ya que estamos en esto también nos dimos a la tarea de buscar el juego con menor Rank, ósea el parámetro máximo dentro de la columna Rank.

Pasado esto procedemos a realizar un análisis de los géneros. No sin antes usar **`unique()`** para ver los géneros del dataset y después realizamos un histograma. En nuestro análisis el género más lanzado fue el de **acción**.

Esto nos dice el total de juegos lanzados de acuerdo al género, pero no sus ventas. Entonces realizamos un análisis del género y sus ventas de las distintas regiones para comprobar nuestra hipótesis acerca del comportamiento del género.

Respecto a la plataforma o consola, podemos ver para que plataforma se venden más juegos en el mundo, también para cada país. Esto lo restringimos a las primeras 10 plataformas.

Respecto también a las plataformas, no porque las consolas saquen más juegos implica que van a tener un mayor número de ventas, entonces nos preguntamos **¿Qué consola tuvo la mayor cantidad de juegos y sus ventas en el mundo?** Para dar respuesta a ello hicimos un análisis de los juegos por

consola y el número de ventas globales encontrando cosas interesantes.

En cuanto a los Publishers o empresas. Podemos las ventas de juegos por empresa.

## Resultados, análisis y Discusion.

Primeramente iremos poniendo en orden los comandos usados en base a la introducción.

Primeramente el primer vistazo fue el top de los juegos.

Tabla 1. Top 5 juegos.

	Rank	Name	Platform	Genre	Publisher
6112	1	Wii Sports	Wii	Sports	Nintendo
122	2	Super Mario Bros.	NES	Platform	Nintendo
8322	3	Mario Kart Wii	Wii	Racing	Nintendo
9750	4	Wii Sports Resort	Wii	Sports	Nintendo
705	5	Pokemon Red/Pokemon Blue	GB	Role-Playing	Nintendo

Como podemos ver el juego que encabeza esta lista es Wii sports y seguido Super Mario Bros.

Usando `data.info()` obtuvimos

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 16324 entries, 0 to 16323
Data columns (total 11 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   Rank            16324 non-null  int64
1   Name            16324 non-null  object
2   Platform        16324 non-null  object
3   Year            16324 non-null  int64
4   Genre           16324 non-null  object
5   Publisher       16288 non-null  object
6   NA_Sales        16324 non-null  float64
7   EU_Sales        16324 non-null  float64
8   JP_Sales        16324 non-null  float64
9   Other_Sales     16324 non-null  float64
10  Global_Sales    16324 non-null  float64
dtypes: float64(5), int64(2), object(4)
memory usage: 1.4+ MB
```

Figura 1. Información del dataset con 16324 entradas y 11 columnas.

Como podemos ver en la figura 1 tenemos 11 columnas x 16324 filas, nos muestra el tipo de variables que tenemos. Además al parecer las columnas no tienen valores

nulos. En cuanto a análisis estadístico solo hay números en la columna Rank y Year, Como Rank no se repite la única variable a la que podemos sacar información estadística es a “Year”, aun así hacemos un análisis estadístico de todo por si aparece algo interesante que analizar.

Para los datos también obtuvimos el mapa de datos faltantes



Figura 2: Mapa de datos faltantes

Como podemos observar en la columna de Platform hay datos faltantes, no afecta a nuestro análisis.

Usando la matriz de correlación observamos que datos se relacionan de mejor manera con los demás.

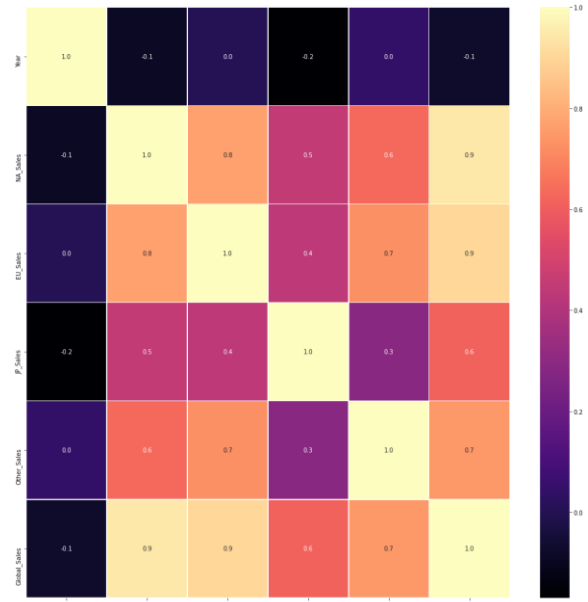


Figura 3: Matriz de correlación

Usando describe() obtuvimos el siguiente análisis estadístico.

Tabla 2. Estadística obtenida del dataset de videojuegos usando describe().

	count	unique		top	freq	mean	std	min	25%	50%	75%	max
Rank	16324.0	NaN		NaN	NaN	8291.50827	4792.043734	1.0	4135.75	8293.5	12439.25	16600.0
Name	16324	11358	Need for Speed: Most Wanted	12	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Platform	16324	31	DS	2133	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Year	16324.0	NaN		NaN	NaN	2006.404251	5.826744	1980.0	2003.0	2007.0	2010.0	2016.0
Genre	16324	12	Action	3252	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Publisher	16288	576	Electronic Arts	1339	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
NA_Sales	16324.0	NaN		NaN	NaN	0.265464	0.821658	0.0	0.0	0.08	0.24	41.49
EU_Sales	16324.0	NaN		NaN	NaN	0.147581	0.508809	0.0	0.0	0.02	0.11	29.02
JP_Sales	16324.0	NaN		NaN	NaN	0.078673	0.311584	0.0	0.0	0.0	0.04	10.22
Other_Sales	16324.0	NaN		NaN	NaN	0.048334	0.189902	0.0	0.0	0.01	0.04	10.57
Global_Sales	16324.0	NaN		NaN	NaN	0.540328	1.56586	0.01	0.06	0.17	0.48	82.74

Podemos ver que en efecto “Year” es la única columna en la cual el comando realizó un análisis estadístico bien.

Como podemos ver la información que podemos explotar es la de Year o año de lanzamiento con una media= 2006.4 años. Primer juego que salió de este dataset es del año 1980 y el último que fue del 2016. Por lo que tomare esta columna y hare un respectivo análisis mediante los parámetros que nos arrojan.

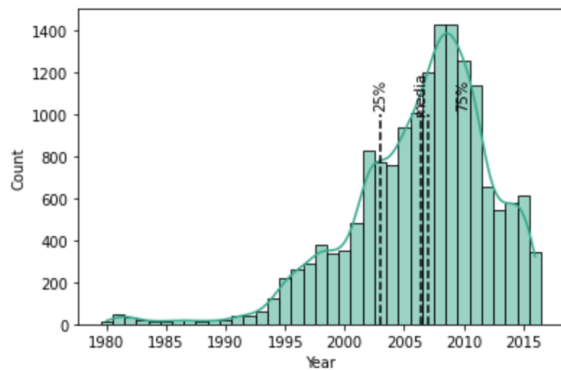


Figura 4. Grafica de los juegos lanzados por año realizado con los parámetros del análisis que nos dios la tabla 2.

De la figura 4 podemos obtener que de 1980 hasta el 2003 se encuentra el 25% de los juegos. y hasta un 75 % de los juegos estan hasta el año 2010. Notamos que del 2003 al 2010 se encuentran el 50% de la cantidad de juegos que salieron hasta el 2016.

Parece que apartir de 2009 hay un descenso en los juegos lanzados. Al parecer nuestra hipótesis de que el lanzamiento de juegos por año iba a aumentar con año no es verdadero.

Siguiendo con la pregunta ¿Qué año ha lanzado mayores juegos hasta el 2016? El siguiente comando nos arroja que el año 2009 fue el año con mayor cantidad de juegos lanzados con un valor de 1433.

Tabla 3. Año con mayor lanzamiento de juegos.

```
data['Year'].value_counts()
```

Year	
2009	1433
2008	1428
2010	1259
2007	1202
2011	1139
2006	1008

Indagamos un poco más en que empresa de videojuegos lanzo más videojuegos en esos años.

Activision	121
Electronic Arts	112
Ubisoft	104
THQ	72
Namco Bandai Games	62

Tabla 4. Empresas con mayores lanzamientos en el año 2009

Observamos que Activision fue la empresa que más juegos lanzo, seguido de EA y Ubisoft.

Para el segundo año con mayores lanzamientos.

Electronic Arts	120
Ubisoft	112
Activision	88
Sega	67
Konami Digital Entertainment	64

Tabla 5. Empresas con mayores lanzamientos en el año 2008.

Para este año, 2008, EA estuvo en primer lugar seguido de Ubisoft y Activision.

Para el 2010.

Activision	89
Electronic Arts	88
Ubisoft	72
Namco Bandai Games	70
Konami Digital Entertainment	68

Tabla 6. Empresas con mayores lanzamientos en 2010.

Para el año 2010 se repitió el orden de las empresas, siendo Activision la de mayores lanzamientos, seguido de EA y Activision.

Observamos que en estos 3 años de mayores lanzamientos la presencia de empresas orientales no falto. Para el año de 1980 apenas se lanzaron 9 juegos, siendo Activision la empresa que más juegos lanzo. Para 2016 se lanzaron apenas 344 de los cuales Namco Bandai Games lanzo más.

Además, con base en esto quisimos ver si alguno de estos juegos se encuentra en el top 100 del Rank, esto debido a que hay mayor probabilidad de ello.

Tabla 7. Juegos que se encuentran dentro del top 100 del Rank del año 2009.

Rank		Name	Platform	Year	Genre	Publisher
9750	4	Wii Sports Resort	Wii	2009	Sports	Nintendo
9751	9	New Super Mario Bros. Wii	Wii	2009	Platform	Nintendo
9752	15	Wii Fit Plus	Wii	2009	Sports	Nintendo
9753	37	Call of Duty: Modern Warfare 2	X360	2009	Shooter	Activision
9754	46	Pokemon HeartGold/Pokemon SoulSilver	DS	2009	Action	Nintendo
9755	56	Call of Duty: Modern Warfare 2	PS3	2009	Shooter	Activision
9756	84	The Sims 3	PC	2009	Simulation	Electronic Arts

Vemos en la tabla 7 que hay 6 juegos que están en el top 100, de hecho, 2 de ellos se encuentran dentro del top 10. Comprobamos así nuestra hipótesis planteada unas líneas atrás.

Por mera curiosidad también encontramos cual es el peor calificado.

Tabla 8. Peor juego del Rank.

```
data.Rank.max()
```

```
16600
```

```
data[data["Rank"]==16600]
```

Rank	Name	Platform	Year	Genre	Publisher
4407	Spirits & Spells	GBA	2003	Platform	Wanadoo

Spirits & spells al parecer es el peor calificado.

También observamos cual fue el juego con mayor lanzamiento en distintas plataformas. Need for Speed: Most Wanted, el cuál fue lanzado en 10 plataformas ["PS2" 'X360' 'XB' 'GC' 'PC' 'DS' 'GBA' 'PS3' 'PSV' 'WiiU']".

Ahora ya empezamos con las ventas, pero primero, respecto a los géneros obtuvimos la siguiente gráfica.

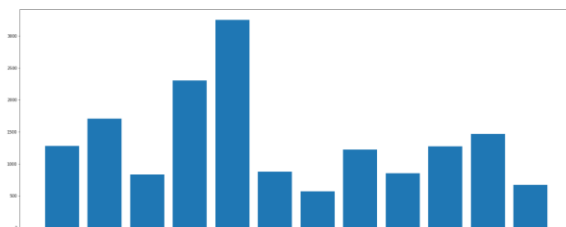


Figura 5. Juegos lanzados por género.

Donde el género acción es el que más ha lanzado juegos, pero eso no nos dice nada acerca de las ventas. Así que hicimos varias graficas de las ventas por género globales y de distintas regiones.

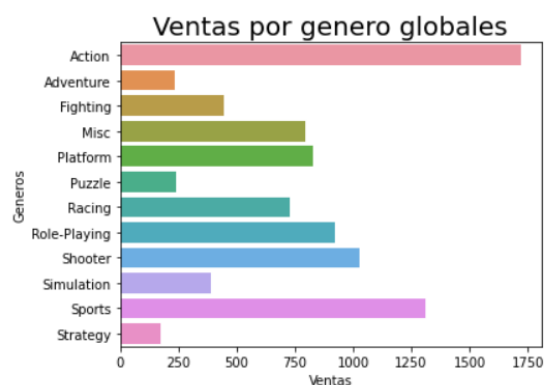


Figura 6. Ventas por género globales.

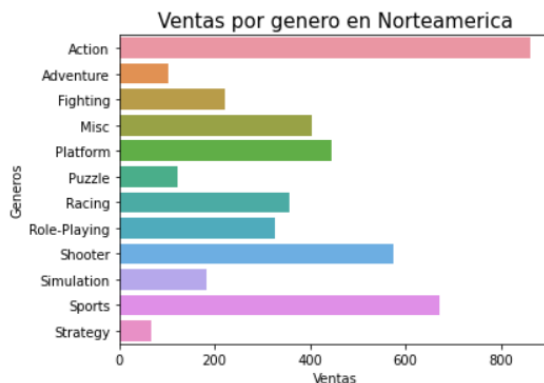


Figura 7. Ventas por género en Norteamérica.

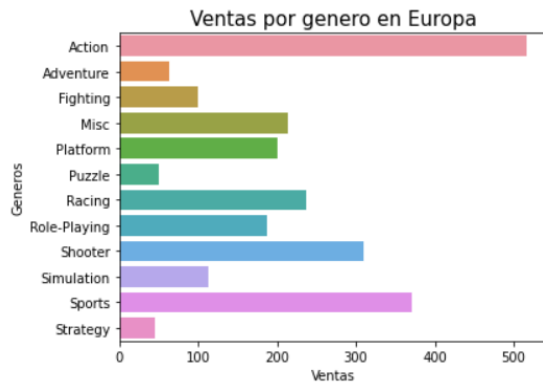


Figura 8. Ventas por género en Europa.

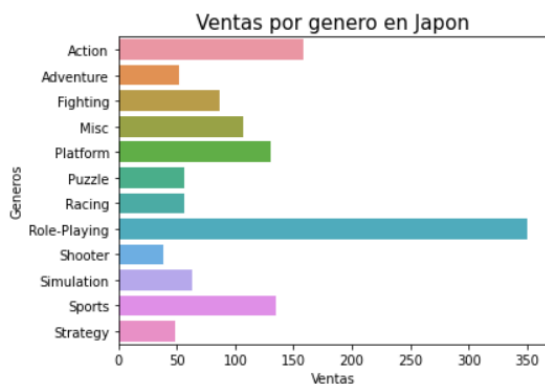


Figura 9. Ventas por género en Japón.

Como podemos ver globalmente se venden más los juegos de acción como se muestra en la figura 6, en general esperaríamos esto de los demás países como lo es Norteamérica y Europa, pero la figura 9 nos dice lo contrario, pues en Japón lo que más se venden son los juegos de rol. Esto es importante pues, si deseas vender juegos en Japón, apostaría a lanzar más juegos de rol en ese país.

Siguiendo con lo propuesto en la introducción, tenemos que las consolas que más se venden ordenadas globalmente son

Tabla 9. Juegos más vendidos globalmente por consola.

	NA_Sales	EU_Sales	JP_Sales	Global_Sales
<b>Platform</b>				
<b>PS2</b>	572.92	332.63	137.54	1233.46
<b>X360</b>	594.33	278.01	12.30	969.61
<b>PS3</b>	388.90	340.47	79.21	949.35
<b>Wii</b>	497.37	264.35	68.28	909.81
<b>DS</b>	388.56	194.07	175.02	818.96
<b>PS</b>	334.71	212.38	139.78	727.39
<b>GBA</b>	184.12	74.59	46.56	313.56
<b>PSP</b>	107.09	67.16	75.89	291.71
<b>PS4</b>	96.80	123.70	14.27	278.07
<b>PC</b>	92.10	137.64	0.17	255.05

De igual manera tenemos su gráfica y también se esperaría un comportamiento similar de los demás países, pero eso lo comprobaremos, para ello usaremos las siguientes graficas de juegos vendidos por consola globalmente y por país.

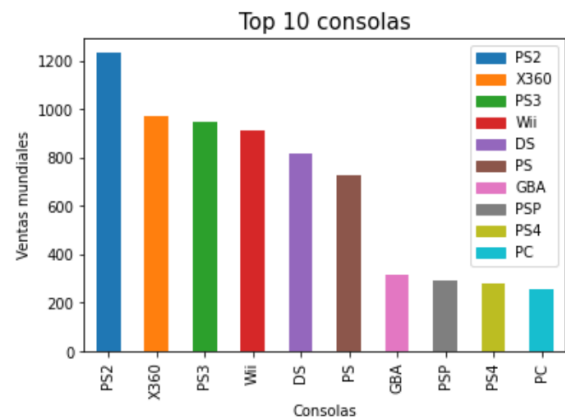


Figura 10. Ventas de juegos por consola en el mundo.

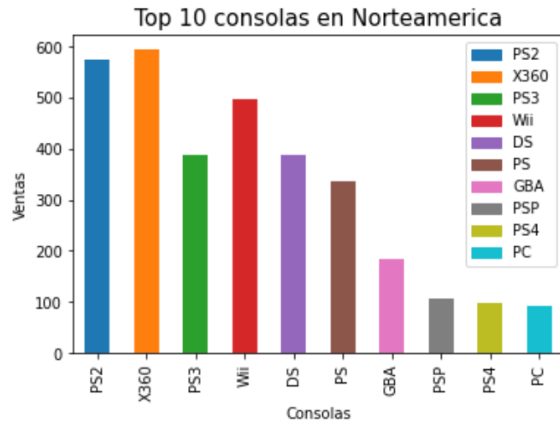


Figura 11. Ventas de juegos por consola Norteamérica.

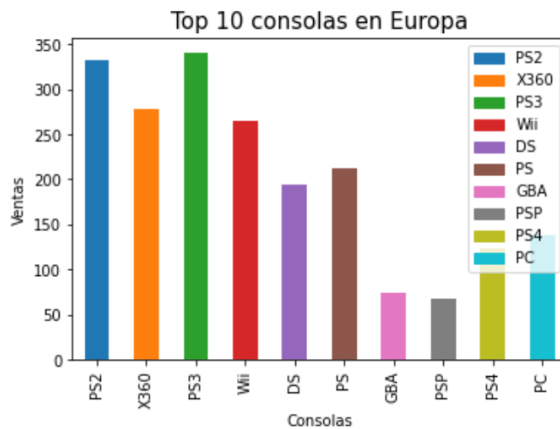


Figura 12. Ventas de juegos por consola Europa.

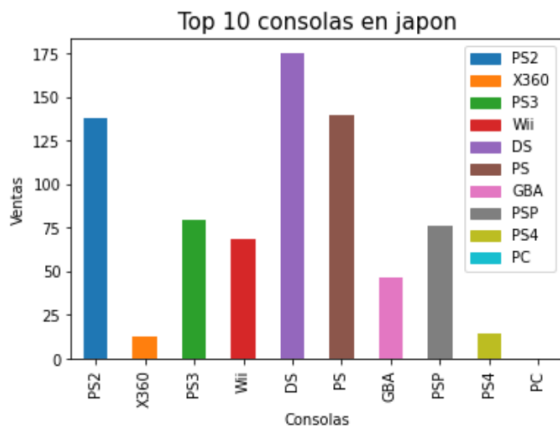


Figura 13. Ventas de juegos por consola en Japón.

Aquí la cosa si cambia bastante, pues globalmente vende más juegos PS2, en Norteamérica vende más Xbox360 por muy poco, en Europa vende más PS3 por muy poco, finalmente de la figura 13 tenemos a

Japón, nuevamente son muy diferentes al resto. Pues ahí el Nintendo DS lidera la venta de juegos por consola seguida de PS, y no tan notorio pero PSP está en cuarto puesto seguido de PS2. Interesante, pues una consola portátil lidera el top y en cuarto puesto se encuentra el PSP, a simple vista al parecer a los japoneses les gustan las consolas portátiles.

Lo mismo que hicimos aquí aplica para los Publishers. Las empresas que crean los juegos y sus ventas, en la siguiente tabla presentamos datos ordenados en base a sus ventas globales

Tabla 10. Ventas por Publisher, top 10.

	NA_Sales	EU_Sales	JP_Sales	Global_Sales
<b>Publisher</b>				
Nintendo	815.75	418.30	454.99	1784.43
Electronic Arts	584.22	367.38	13.98	1093.39
Activision	426.01	213.72	6.54	721.41
Sony Computer Entertainment	265.22	187.55	74.10	607.28
Ubisoft	252.81	163.03	7.33	473.54
Take-Two Interactive	220.47	117.95	5.83	399.30
THQ	208.60	94.60	5.01	340.44
Konami Digital Entertainment	88.91	68.62	90.93	278.56
Sega	108.78	81.41	56.15	270.66
Namco Bandai Games	69.38	42.61	126.84	253.65

La tabla 10 nos muestra el top 10 de las ventas por empresa globales y de distintos países ordenados por ventas globales. Nintendo lidera las ventas globales. Similarmente se pudo hacer el análisis anterior (géneros) con ventas, pero por falta de tiempo lo dejaremos así.

## Conclusiones.

Basándonos en nuestras hipótesis y lo que encontramos y desarrollamos mediante avanzábamos el análisis del dataset, en principio esperábamos un aumento en el lanzamiento de los juegos, pero en 2009 tuvo un descenso, intuyo que es debido a la complejidad con la se hacen los juegos. Pues cada vez queremos juegos más realistas gráficamente y se tiene que innovar.

En 2009 obtuvimos el mayor número de juegos lanzados con 1433 juegos lanzados, de los cuales 7 se encuentran en el top 100 y 2 en el top 10.

En cuanto a los juegos lanzados por género obtuvimos que el género “**acción**” es el más lanzado, en ventas globales también tiene a la misma situación, y tiene un comportamiento similar en las ventas de regiones como Norteamérica y Europa. Japón fue la excepción, pues en ese país se venden más los **juegos de roles**.

De las ventas de juegos por consola vemos que la que globalmente lidera es PS2 y difiere un poco de las regiones analizadas, nuevamente la mayor diferencia se la llevan los japoneses, pues ellos compran más juegos de Nintendo DS y seguido de PS.

El análisis de la columna Publisher nos dice que la mayor venta de juegos por Publisher es Nintendo y seguido Electronic Arts.