



Universidad Tecnológica de Morelia

Tecnologías de la información y comunicación Desarrollo
de software multiplataforma

Grupo: 9ºA

EXTRACCIÓN DE CONOCIMIENTOS USANDO EL MODELO CRISP-DM

ACTIVIDAD:
USO DEL MODELO CRISP-DM EN UN DATASET

Alumno: Joshua García Gamiño

Maestro: José Luis Cendejas Valdez

1 de diciembre del
2025

ETAPA 1 Comprensión del negocio (Business Understanding)

Objetivos de Negocio:

- Analizar el mercado global de videojuegos
- Identificar tendencias de ventas por región, género y plataforma
- Descubrir patrones de éxito de publishers y franquicias
- Analizar la evolución temporal del mercado

Objetivos de Minería de Datos:

- Identificar factores que influyen en el éxito comercial
- Segmentar juegos por desempeño comercial
- Predecir potencial de ventas basado en características
- Analizar distribución geográfica de preferencias

Criterios de Éxito:

- Identificar los 5 publishers más exitosos
- Determinar géneros con mayor tasa de éxito
- Analizar evolución de plataformas dominantes
- Identificar patrones temporales de lanzamientos exitosos

ETAPA 2 Comprensión de los datos (Data Understanding)

Estructura del Dataset:

- 16,328 registros de videojuegos
- Período: 1980-2020
- Variables principales:
 - Nombre del juego
 - Plataforma
 - Año de lanzamiento
 - Género
 - Publisher
 - Ventas por región (NA, EU, JP, Other)
 - Ventas globales Análisis

Exploratorio Inicial:

Principales observaciones:

1. **Juegos más vendidos:** Wii Sports (82.74M), Super Mario Bros. (40.24M), Mario Kart Wii (35.82M)
2. **Publisher dominante:** Nintendo (aparece en múltiples juegos top)
3. **Plataformas diversas:** Wii, NES, DS, PS2, X360
4. **Géneros principales:** Sports, Platform, Racing, Role-Playing

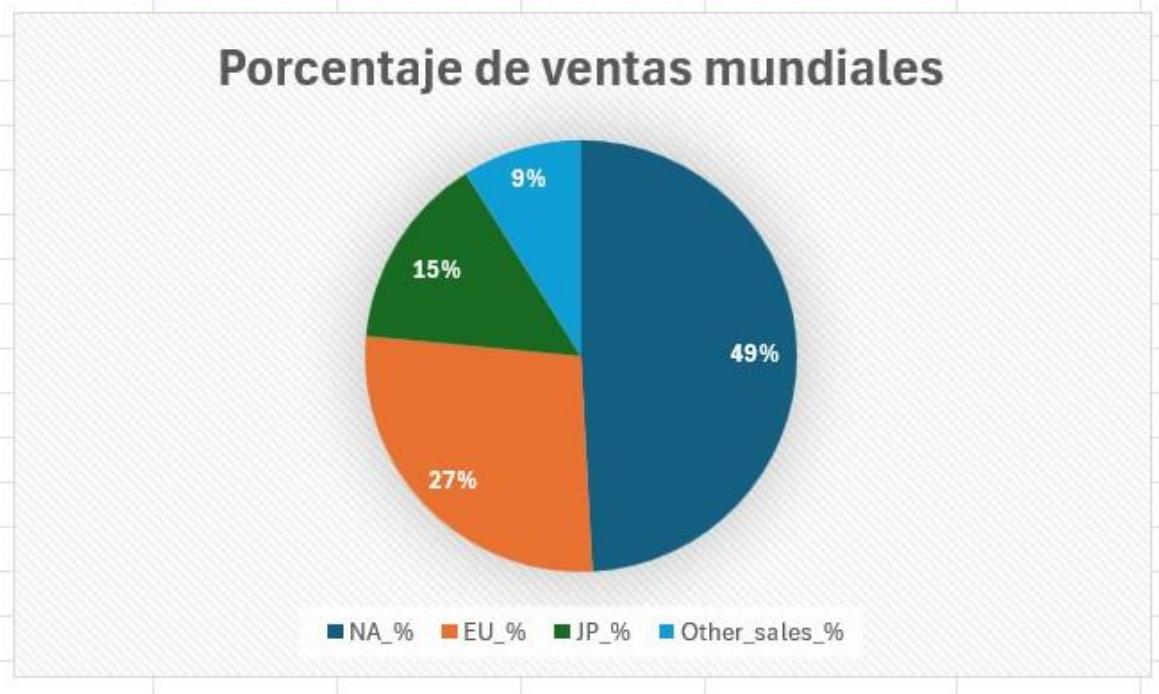
ETAPA 3 Preparación de datos (Data Preparation)

Limpieza de Datos Requerida:

1. **Manejo de valores nulos/missing:**
 - Verificar y tratar años faltantes
2. **Estandarización:**
 - Uniformizar nombres de columnas
 - Estandarizar formatos de género
3. **Eliminación de columnas:**
 - Eliminar columnas vacías o duplicadas
4. **Integración:**
 - Unificar datos de ventas en estructura consistente
 - Crear variables derivadas (ej: % de ventas por región)

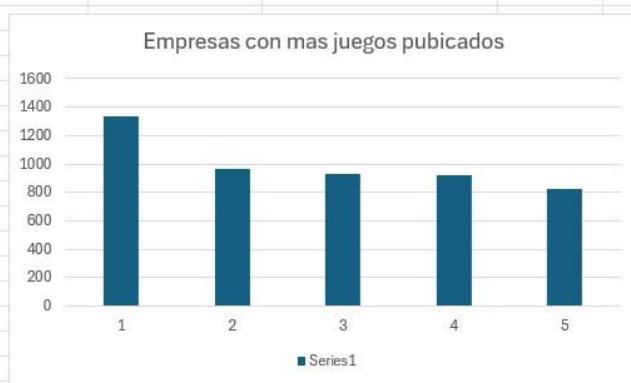
Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
Wii Sports	Wii	2006	### Sports	Nintendo	414.9	29.02	3.77	8.46	82.74
Super Mario Bros.	NES	1985	### Platform	Nintendo	29.08	3.58	6.81	0.77	40.24
Mario Kart Wii	Wii	2008	### Racing	Nintendo	15.85	12.88	3.79	3.31	35.82
Wii Sports Resort	Wii	2010	### Sports	Nintendo	15.75	11.01	3.28	2.96	33
Pokemon Red/Pokemon Blue	GB	1996	### Role-Play	Nintendo	11.27	8.89	10.22	1	31.37
Tetris	GB	1989	### Puzzle	Nintendo	23.2	2.26	4.22	0.58	30.26
New Super Mario Bros.	DS	2006	### Platform	Nintendo	11.38	9.23	6.5	2.9	30.01
Wii Play	Wii	2006	### Misc	Nintendo	14.03	9.2	2.93	2.85	29.02
New Super Mario Bros. Wii	Wii	2009	### Platform	Nintendo	14.59	7.06	4.7	2.26	28.62
Duck Hunt	NES	1984	### Shooter	Nintendo	26.93	0.63	0.28	0.47	28.31
Nintendogs	DS	2005	### Simulation	Nintendo	9.07	11	1.93	2.75	24.76
Mario Kart DS	DS	2004	### Racing	Nintendo	9.81	7.57	4.13	1.92	23.42
Pokemon Gold/Pokemon Silver	GB	1999	### Role-Play	Nintendo	9	6.18	7.2	0.71	23.1
Wii Fit	Wii	2005	### Sports	Nintendo	8.94	8.03	3.6	2.15	22.72
Wii Fit Plus	Wii	2009	### Sports	Nintendo	9.09	8.59	2.53	1.79	22
Kinect Adventures!	X360	2010	### Misc	Microsoft Game Studios	14.37	4.94	0.24	1.67	21.82
Grand Theft Auto V	PS3	2013	### Action	Take-Two Interactive	7.01	9.27	0.97	4.14	21.4
Grand Theft Auto: San Andreas	PS2	2004	### Action	Take-Two Interactive	9.43	0.4	0.41	10.57	20.81
Super Mario World	SNES	1990	### Platform	Nintendo	12.78	3.75	3.54	0.55	20.61
Brain Age: Train Your Brain in Minutes a Day	DS	2006	### Misc	Nintendo	4.75	9.26	4.16	2.05	20.22
Pokemon Diamond/Pokemon Pearl	DS	2007	### Role-Play	Nintendo	6.42	4.52	6.04	1.37	18.36
Super Mario Land	GB	1989	### Platform	Nintendo	10.83	2.71	4.18	0.42	18.14
Super Mario Bros. 3	NES	1988	### Platform	Nintendo	9.54	3.44	3.84	0.46	17.28
Grand Theft Auto V	X360	2013	### Action	Take-Two Interactive	9.63	5.31	0.06	1.38	16.38
Grand Theft Auto: Vice City	PS2	2002	### Action	Take-Two Interactive	8.41	5.49	0.47	1.78	16.15
Pokemon Ruby/Pokemon Sapphire	GBA	2002	### Role-Play	Nintendo	6.06	3.9	5.38	0.5	15.85
Pokemon Black/Pokemon White	DS	2010	### Role-Play	Nintendo	5.57	3.28	5.65	0.62	15.32
Brain Age 2: More Training in Minutes a Day	DS	2007	### Puzzle	Nintendo	3.44	5.36	5.32	1.18	15.3
Gran Turismo 3: A-Spec	PS2	2003	### Racing	Sony Computer Entertainment	6.85	5.09	1.87	1.16	14.98
Call of Duty: Modern Warfare 3	X360	2011	### Shooter	Activision	9.03	4.28	0.13	1.32	14.76
Pokémon Yellow: Special Pikachu Edition	GB	1999	### Role-Play	Nintendo	5.89	5.04	3.12	0.53	14.64
Call of Duty: Black Ops	X360	2010	### Shooter	Activision	9.67	3.73	0.11	1.13	14.64
Pokemon X/Pokemon Y	3DS	2014	### Role-Play	Nintendo	5.17	4.05	4.34	0.79	14.35
Call of Duty: Black Ops 3	PS4	2015	### Shooter	Activision	5.77	5.81	0.35	2.31	14.24
Call of Duty: Black Ops II	PS3	2012	### Shooter	Activision	4.99	5.88	0.65	2.52	14.03
Call of Duty: Black Ops II	X360	2012	### Shooter	Activision	8.25	4.3	0.07	1.12	13.73
Call of Duty: Modern Warfare 2	X360	2009	### Shooter	Activision	8.52	3.63	0.08	1.29	13.51
Call of Duty: Modern Warfare 3	PS3	2012	### Shooter	Activision	5.54	5.82	0.49	1.62	13.46
Grand Theft Auto III	PS2	2001	### Action	Take-Two Interactive	6.99	4.51	0.3	1.3	13.1
Super Smash Bros. Brawl	Wii	2008	### Fighting	Nintendo	6.75	2.61	2.66	1.02	13.04
Call of Duty: Black Ops	PS3	2011	### Shooter	Activision	5.98	4.44	0.48	1.83	12.73
Animal Crossing: Wild World	DS	2005	### Simulation	Nintendo	2.55	3.52	5.33	0.68	12.27
Mario Kart 7	3DS	2012	### Racing	Nintendo	4.74	3.91	2.67	0.89	12.21
Halo 3	X360	2007	### Shooter	Microsoft Game Studios	7.97	2.83	0.13	1.21	12.14
Grand Theft Auto V	PS4	2013	### Action	Take-Two Interactive	3.8	5.81	0.36	2.02	11.98
Pokemon HeartGold/Pokemon SoulSilver	DS	2009	### Action	Nintendo	4.4	2.77	3.96	0.77	11.9
Super Mario 64	N64	1996	### Platform	Nintendo	6.91	2.85	1.91	0.23	11.89
Gran Turismo 4	PS2	2005	### Racing	Sony Computer Entertainment	3.01	0.01	1.1	7.53	11.66
Super Mario Galaxy	Wii	2007	### Platform	Nintendo	6.16	3.4	1.2	0.76	11.52
Pokemon Omega Ruby/Pokemon Alpha Sapi	3DS	2014	### Role-Play	Nintendo	4.23	3.37	3.08	0.65	11.33
Super Mario Land 2: 6 Golden Coins	GB	1992	### Adventure	Nintendo	6.16	2.04	2.69	0.29	11.18
Grand Theft Auto IV	X360	2008	### Action	Take-Two Interactive	6.76	3.1	0.14	1.03	11.02
Gran Turismo	PS	2000	### Racing	Sony Computer Entertainment	4.02	3.87	2.54	0.52	10.95
Super Mario 3D Land	3DS	2011	### Platform	Nintendo	4.83	2.99	2.13	0.78	10.79
Gran Turismo 5	PS3	2010	### Racing	Sony Computer Entertainment	2.96	4.88	0.81	2.12	10.77
Call of Duty: Modern Warfare 2	PS3	2009	### Shooter	Activision	4.93	3.69	0.38	1.63	10.69
Grand Theft Auto IV	PS3	2008	### Action	Take-Two Interactive	4.76	3.76	0.44	1.62	10.57
Super Mario All-Stars	SNES	1996	### Platform	Nintendo	5.93	2.15	2.12	0.29	10.55
Pokemon FireRed/Pokemon LeafGreen	GBA	2002	### Role-Play	Nintendo	4.34	2.65	3.15	0.35	10.49
Super Mario 64	DS	2002	### Platform	Nintendo	5.08	3.11	1.25	0.98	10.42
Just Dance 3	Wii	2010	### Misc	Ubisoft	6.05	3.15	0	1.07	10.26
Call of Duty: Ghosts	X360	2013	### Shooter	Activision	6.72	2.63	0.04	0.82	10.21
Halo: Reach	X360	2010	### Shooter	Microsoft Game Studios	7.03	1.98	0.08	0.78	9.88
Mario Kart 64	N64	1996	### Racing	Nintendo	5.55	1.94	2.23	0.15	9.87
New Super Mario Bros. 2	3DS	2012	### Platform	Nintendo	3.66	3.07	2.47	0.63	9.82
Halo 4	X360	2012	### Shooter	Microsoft Game Studios	6.63	2.36	0.04	0.73	9.76
Final Fantasy VII	PS	1997	### Role-Play	Sony Computer Entertainment	3.01	2.47	3.28	0.96	9.72
Call of Duty: Ghosts	PS3	2013	### Shooter	Activision	4.03	3.73	0.38	1.38	9.59
Just Dance 2	Wii	2009	### Misc	Ubisoft	5.84	2.89	0.01	0.78	9.52
Gran Turismo 2	PS	2000	### Racing	Sony Computer Entertainment	3.88	3.42	1.69	0.5	9.49
Call of Duty 4: Modern Warfare	X360	2007	### Shooter	Activision	5.91	2.38	0.13	0.9	9.32
Donkey Kong Country	SNES	1994	### Platform	Nintendo	4.36	1.71	3	0.23	9.3
Minecraft	X360	2012	### Misc	Microsoft Game Studios	5.58	2.83	0.02	0.77	9.2
Animal Crossing: New Leaf	3DS	2013	### Simulation	Nintendo	2.01	2.32	4.36	0.41	9.09
Mario Party DS	DS	2005	### Misc	Nintendo	4.46	1.88	1.98	0.7	9.02
The Elder Scrolls V: Skyrim	X360	2011	### Role-Play	Bethesda Softworks	5.03	2.86	0.1	0.85	8.84
Super Mario Kart	SNES	1992	### Racing	Nintendo	3.54	1.24	3.81	0.18	8.76
FIFA 16	PS4	2015	### Sports	Electronic Arts	1.11	6.06	0.06	1.26	8.49
Wii Party	Wii	2009	### Misc	Nintendo	1.79	3.53	2.49	0.68	8.49
Halo 2	XB	2004	### Shooter	Microsoft Game Studios	6.82	1.53	0.05	0.08	8.49
Mario Party 8	Wii	2007	### Misc	Nintendo	3.81	2.3	1.58	0.73	8.42
Pokemon Black 2/Pokemon White 2	DS	2011	### Role-Play	Nintendo	2.91	1.86	3.14	0.43	8.33
FIFA Soccer 13	PS3	2012	### Action	Electronic Arts	1.06	5.05	0.13	2.01	8.24
The Sims 3	PC	2009	### Simulation	Electronic Arts	n/a	8.42	n/a	0.71	8.11

NA_%	EU_%	JP_%	Other_sales_%
49.13%	27.31%	14.56%	8.95%



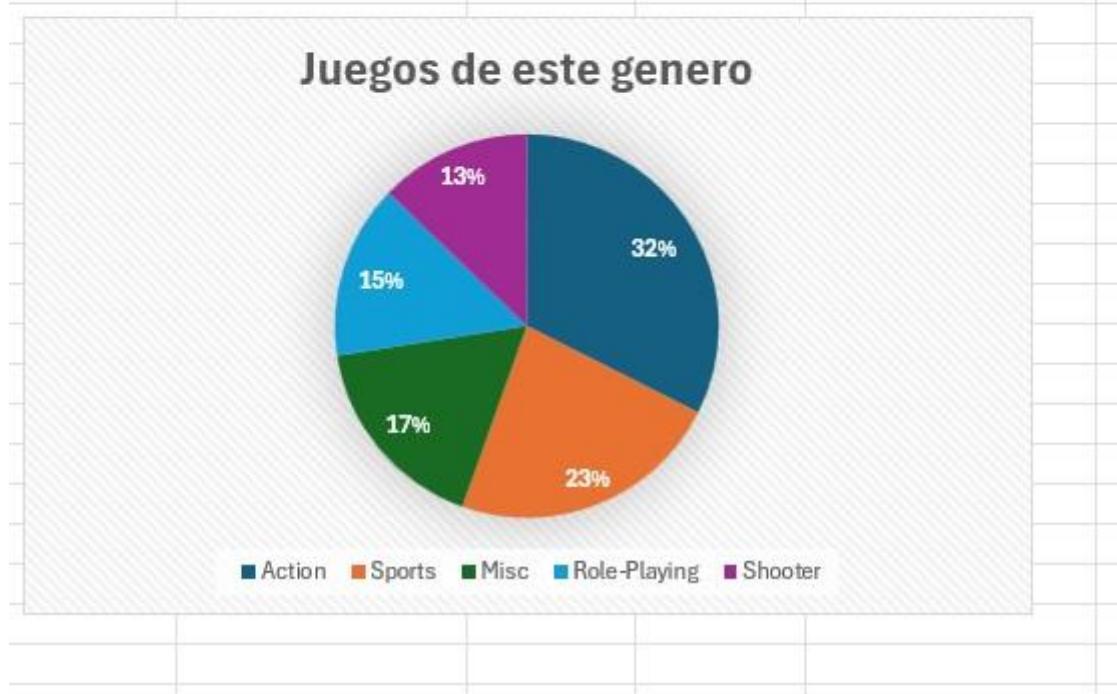
Repeticiones totales	Nombre de la empresa
1	Publisher
696	Nintendo
189	Microsoft Game Studios
412	Take-Two Interactive
682	Sony Computer Entertainment
966	Activision
918	Ubisoft
69	Bethesda Softworks
1339	Electronic Arts
632	Sega
52	SquareSoft
347	Atari
192	505 Games
376	Capcom
45	GT Interactive
823	Konami Digital Entertainment
15	Sony Computer Entertainment Europe
231	Square Enix
89	LucasArts
62	Virgin Interactive
217	Warner Bros. Interactive Entertainment
22	Universal Interactive
196	Eidos Interactive
4	RedOctane
161	Vivendi Games
30	Enix Corporation
928	Namco Bandai Games
1	Palcom
16	Hasbro Interactive
712	THQ
8	Fox Interactive
184	Acclaim Entertainment
40	MTV Games
214	Disney Interactive Studios
90	Majesco Entertainment
150	Codemasters
2	Red Orb
27	Level 5
2	Arena Entertainment
196	Midway Games
8	JVC
121	Deep Silver
14	989 Studios
6	NCSoft
1	UEP Systems
7	Parker Bros.
3	Maxis
4	Imagic

Repeticiones totales	Nombre de la empresa
1339	Electronic Arts
966	Activision
928	Namco Bandai Games
918	Ubisoft
823	Konami Digital Entertainment



Áre

Genero	Juegos de este genero	Genero	Juegos de este genero
Sports	2304	Action	3253
Platform	876	Sports	2304
Racing	1226	Misc	1710
Role-Playing	1471	Role-Playing	1471
Puzzle	571	Shooter	1282
Misc	1710		
Shooter	1282		
Simulation	851		
Action	3253		
Fighting	836		
Adventure	1276		
Strategy	671		





ETAPA 3 (Preparación de datos)

Se prepararon los datos haciendo unos cambios por rango y generalizando los datos para poderlos agrupar de una mejor manera todos como el año fue por grupo y por ejemplo el género cada género es un número.

	A	B	C	D	E	F	G	H
1	Plataform	Year	Genre	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
2	1	3	1	41	29	4	8	83
3	2	1	2	29	4	7	1	40
4	1	3	3	16	13	4	3	36
5	1	3	1	16	11	3	3	33
6	3	2	3	11	9	10	1	31
7	3	1	4	23	2	4	1	30
8	4	3	2	11	9	7	3	30
9	1	3	5	14	9	3	3	29
10	1	3	2	15	7	5	2	29
11	2	1	6	27	1	0	0	28
12	4	3	7	9	11	2	3	25
13	4	3	3	10	8	4	2	23
14	3	2	3	9	6	7	1	23
15	1	3	1	9	8	4	2	23
16	1	3	1	9	9	3	2	22
17	5	4	5	15	5	0	2	22
18	6	4	8	7	9	1	4	21
19	7	3	8	9	0	0	11	21
20	8	2	2	13	4	4	1	21
21	4	3	5	5	9	4	2	20
22	4	3	3	6	5	6	1	18
23	3	1	2	11	3	4	0	18
24	2	1	2	10	3	4	0	17
25	5	4	8	10	5	0	1	16
26	7	3	8	8	5	0	2	16
27	9	3	3	6	4	5	1	16
28	4	4	3	6	3	6	1	15
29	4	3	4	3	5	5	1	15
30	7	3	3	7	5	2	1	15
31	5	4	6	9	4	0	1	15
32	3	2	3	6	5	3	1	15
33	5	4	6	10	4	0	1	15
34	10	4	3	5	4	4	1	14
35	11	4	6	6	6	0	2	14
36	6	4	6	5	6	1	3	14
37	5	4	6	8	4	0	1	14
38	5	3	6	9	4	0	1	14

ETAPA 4 (Modelado)

Alpha de Cronbach.

Con la ayuda de jamovi se verifico cual era la fiabilidad de los resultados dando el siguiente resultado

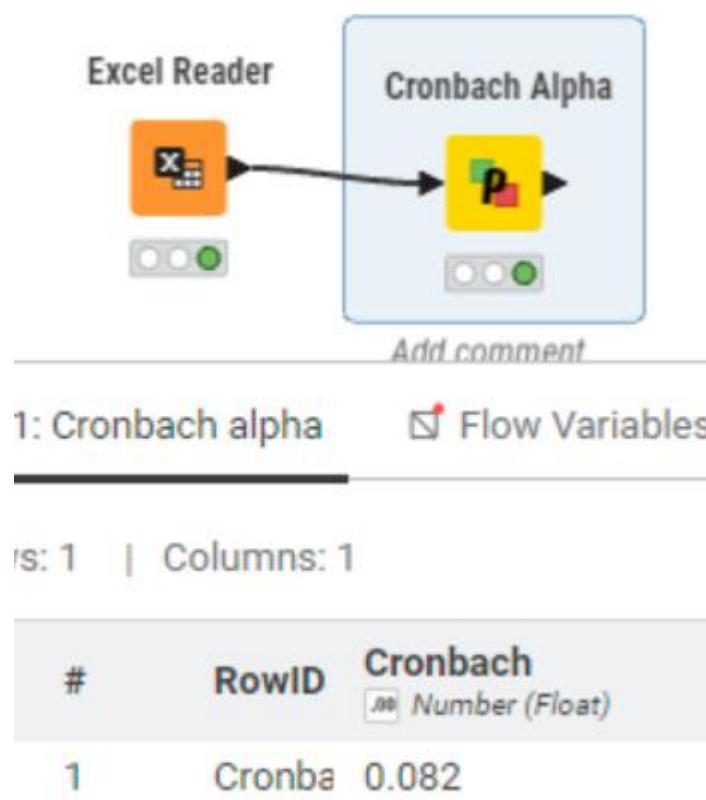
Análisis de Fiabilidad

Estadísticas de Fiabilidad de Escala

Alfa de Cronbach

escala 0.0865

Y con la ayuda de KNIME se volvieron a confirmar la fiabilidad de los resultados dando casi el mismo resultado solo con una ligera diferencia



Correlaciones:

Las 10 correlaciones más fuertes igualmente se analizaron con la ayuda de jamovi y fueron los siguientes resultados.

MATRIZ de correlación

Matriz de correlación

		A	B	C	D	E	F	G	H
A	r de Pearson	—							
	gl	—							
	valor p	—							
B	r de Pearson	-0.068 ***	—						
	gl	16325	—						
	valor p	< .001	—						
C	r de Pearson	0.064 ***	0.095 ***	—					
	gl	16325	16325	—					
	valor p	< .001	< .001	—					
D	r de Pearson	-0.075 ***	-0.075 ***	-0.052 ***	—				
	gl	16325	16325	16325	—				
	valor p	< .001	< .001	< .001	—				
E	r de Pearson	-0.067 ***	0.009	-0.039 ***	0.739 ***	—			
	gl	16325	16325	16325	16325	—			
	valor p	< .001	0.266	< .001	< .001	—			
F	r de Pearson	-0.060 ***	-0.139 ***	-0.061 ***	0.440 ***	0.422 ***	—		
	gl	16325	16325	16325	16325	16325	—		
	valor p	< .001	< .001	< .001	< .001	< .001	—		
G	r de Pearson	-0.047 ***	0.013	-0.022 **	0.538 ***	0.614 ***	0.261 ***	—	
	gl	16325	16325	16325	16325	16325	16325	—	
	valor p	< .001	0.090	0.005	< .001	< .001	< .001	—	
H	r de Pearson	-0.086 ***	-0.063 ***	-0.062 ***	0.918 ***	0.868 ***	0.590 ***	0.652 ***	—
	gl	16325	16325	16325	16325	16325	16325	16325	—
	valor p	< .001	< .001	< .001	< .001	< .001	< .001	< .001	—

Donde cada columna tiene el siguiente significado:

A	B	C	D	E	F	G	H
Plataform	Year	Genre	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales

Por lo que las 5 mas altas fueron las siguientes.

Variables involucradas	Correlación
Other_Sales y global_Sales	0.652

NA_Sales y Other_Sales	0.538
JP_Sales y NA_Sales	0.422
JP_Sales y Other_Sales	0.261
EU_Sales y Other_Sales	0.164
NA_Sales y EU_Sales	0.739
EU_Sales y Other_Sales	0.614
NA_Sales y Other_Sales	0.538
JP_Sales y Global_Sales	0.590
Other_Sales y Global_Sales	0.652

FASE 5 (Evaluación)

Predicciones:

DESCRIPCIÓN DE VARIABLE	VALOR	NIVEL	PREDICCIÓN	KPIS
Other_Sales y Global_Sales	0.652	Moderado	Existe una correlación positiva fuerte entre las ventas en otras regiones y las ventas globales. Esto indica que los mercados fuera de América del Norte, Europa y Japón tienen un impacto	Incrementar ventas en mercados emergentes en un 15% para títulos de acción/deportes. $(Other_Sales / Global_Sales) \times 100$

			considerable en el total global, especialmente para géneros populares como acción y deportes.	
NA_Sales y Other_Sales	0.538	Moderado	Las ventas en Norteamérica están moderadamente relacionadas con las ventas en otras regiones. Esto sugiere que los títulos exitosos en NA tienden a tener buen desempeño internacional, aunque no necesariamente en Europa o Japón.	Lanzar campañas localizadas para los 5 títulos más vendidos en NA en "Otras regiones". $(Other_Sales / NA_Sales) \times 100$
JP_Sales y NA_Sales	0.422	Moderado	Aunque los mercados de Japón y Norteamérica tienen diferencias culturales, ciertos géneros como RPG y aventura muestran una correlación significativa en sus ventas entre ambas regiones.	Desarrollar 2-3 RPGs/anuales con temática transcultural. $JP_Sales / (JP_Sales + NA_Sales)) \times 100$
JP_Sales y Other_Sales	0.261	Baja	La relación entre Japón y otras regiones es débil pero positiva. Esto puede reflejar que algunos títulos japoneses logran penetrar mercados menos tradicionales, como Latinoamérica o Australia.	Aumentar inversión en marketing en LatAm/Oceanía para franquicias japonesas en un 25%. $(Other_Sales / JP_Sales) \times 100$
EU_Sales y Other_Sales	0.164	Baja	Las ventas en Europa y otras regiones tienen una correlación baja, lo que indica que el comportamiento del consumidor puede variar considerablemente entre estos mercados, especialmente en	Realizar estudio de mercado para entender divergencias y desarrollar estrategias regionales diferenciadas. $EU_Sales / Other_Sales \times 100$

			géneros como simulación o estrategia.	
NA_Sales y EU_Sales	0.739	Alta	Existe una correlación positiva muy fuerte entre las ventas en Norteamérica y Europa. Esto refleja gustos similares entre los mercados occidentales y estrategias de marketing global efectivas.	Sincronizar lanzamientos entre NA y EU Fórmula: $(EU_Sales / NA_Sales) \times 100$
EU_Sales y G Other_Sales	0.614	Moderado-Alta	Las ventas en Europa tienen una correlación importante con las ventas en otras regiones, sugiriendo que los éxitos europeos tienen buena recepción en mercados emergentes.	Extender éxito europeo a otras regiones Fórmula: $(Other_Sales / EU_Sales) \times 100$
NA_Sales y Other_Sales	0.538	Moderado	Las ventas en Norteamérica están moderadamente relacionadas con las ventas en otras regiones, indicando que algunos títulos exitosos en NA tienen potencial en mercados emergentes.	Capitalizar éxitos de NA en otras regiones Fórmula: $(Other_Sales / NA_Sales) \times 100$
JP_Sales y Global_Sales	0.590	Moderado-Alta	Aunque Japón es un mercado más nicho, los juegos exitosos allí tienden a tener también buen desempeño global, especialmente franquicias japonesas populares.	Optimizar impacto global de títulos japoneses Fórmula: $(JP_Sales / Global_Sales) \times 100$
Other_Sales y Global_Sales	0.652	Moderado-Alta	Las ventas en otras regiones tienen una correlación importante con las ventas globales, mostrando que mercados emergentes contribuyen significativamente al éxito total.	Aumentar contribución de otras regiones al total Fórmula: $(Other_Sales / Global_Sales) \times 100$