# Universidad Tecnológica de Morelia

# Asignatura:

Extracción de conocimientos de bases de datos

### Profesor:

Jose Luis Cendejas Valdez

### Alumnos:

Carina Ponce Osornio - UTM221037TI Alejandro Govea Martinez - UTM180078TI

Grupo: 9 A TIC

### Actividad:

"Aplicación deL modelo CRISP el la encuesta 'El impacto de los vídeo-juegos en la vida personal' "

Morelia Michoacán, a 13 de Agosto de 2025



# **ÍNDICE**

Fase 1. Entendimiento del negocio

Objetivo general

Objetivos del negocio

Fase 2. Entendimiento de los datos

Fase 3. Preparación de los datos

Limpieza de datos

Análisis de confiabilidad (Alpha de Cronbach)

Descripción de los resultados

Fase 4. Modelado

Aplicación del estudio de correlaciones (Bivariada de Pearson)

Fase 5. Evaluación

Fase 6. Despliegue

Selección del DataSet / Encuesta

Aplicación de etapa 1 y 2 de CRISP-DM

Pre-procesamiento de datos

Confiabilidad: Alfa de Cronbach

Correlaciones (Bivariada de Pearson) Y Tabla de predicciones

# 1.Descripción del modelo CRISP-DM y sus etapas

**CRISP-DM** (Cross-Industry Standard Process for Data Mining) es un marco en 6 fases, iterativo:

- 1. **Entendimiento del negocio**: definir objetivos, preguntas y criterios de éxito.
- 2. **Entendimiento de los datos**: recopilar, describir, explorar la calidad/estructura de los datos.
- 3. **Preparación de los datos**: limpieza, transformación, codificación, selección de variables.
- 4. **Modelado**: elegir técnicas (correlaciones, clustering, regresión, etc.) y ajustar parámetros.
- 5. **Evaluación**: verificar si los resultados responden al objetivo del negocio; revisar supuestos/limitaciones.
- 6. **Despliegue**: comunicar resultados, planes de acción y (cuando aplique) automatizar.

# Fase 1. Entendimiento del negocio

### **Objetivo general**

Conocer los hábitos, consumo y percepciones de los jugadores de videojuegos para identificar patrones de uso, impacto en salud/vida social y comportamiento de consumo (gasto, piratería, participación en comunidades). Estos hallazgos pueden usarse para: recomendaciones de salud, estrategias de marketing para la industria, o propuestas educativas.

### Objetivos del negocio

- **Identificar los hábitos de juego:** ¿Quiénes juegan, con qué frecuencia y en qué plataformas?
- Evaluar el impacto: ¿Cómo influyen los videojuegos en la salud, vida social y otros aspectos de la vida de los jugadores?
- Analizar el comportamiento y percepciones: ¿Qué piensan los jugadores sobre la industria, la piratería y el sistema de clasificación de juegos?
- Comprender el consumo: ¿Cuánto dinero gastan los jugadores en videojuegos y productos relacionados?

# Fase 2. Entendimiento de los datos

La encuesta cuenta con 86 registros de jugadores con 22 variables por registro

#### Descripción de los datos

#### Variables disponibles:

- 1. grupo\_edad: Grupo de edad (<18, 18-24, 25-34, 35-44, 45+).
- 2. genero: Género (Masculino, Femenino, Otro/Prefiere no decir).
- 3. edad inicio juego: Edad de inicio en videojuegos (<5, 6-12, 13-18, 19+).
- 4. horas\_juego\_diarias: Horas promedio de juego por día (<1, 1-2, 3-4, 5+).
- 5. dispositivos de juego: Opción múltiple(PC, consola, móvil, nube)
- 6. gasto\_mensual: Gasto mensual en videojuegos (\$0, \$1-\$50, \$51-\$100, >\$100).
- 7. gasto\_total\_vida: Gasto acumulado en videojuegos en la vida (<\$500, \$500-\$5,000, \$5,001-\$20,000, >\$20,000).
- 8. impacto\_salud: Impacto percibido en la salud (1–5 Likert).
- 9. impacto\_social: Impacto percibido en la vida social (1–5 Likert).
- 10. horas sueno: Horas promedio de sueño al día (<5, 5-7, 7-9, >9).
- 11. fatiga\_visual: Frecuencia de fatiga visual o dolor de cabeza (Nunca, Rara vez, A veces, Frecuentemente, Siempre).
- 12. modalidad\_juego: Modalidad de juego principal (Solo, Multijugador, Ambos por igual).
- 13. tiene\_hobbies: Si tiene hobbies o actividades extracurriculares (Sí/No).
- 14. cree\_violencia: Creencia de que los videojuegos incitan a la violencia (1–5 Likert).
- 15. abandono\_actividades: Frecuencia de abandono de actividades importantes por jugar (Nunca, Rara vez, A veces, Casi siempre, Siempre).

- 16. descarga\_pirata: Ha descargado juegos de manera ilegal/pirata (Sí/No/Prefiere no responder).
- 17. conoce\_esrb\_pegi: Conocimiento del sistema ESRB/PEGI (Sí/No).
- 18. considera\_esrb\_pegi: Considera la clasificación ESRB/PEGI al comprar/regalar un juego (Siempre, A veces, Nunca).
- 19. compra\_mercancia: Compra de mercancía oficial de videojuegos (Sí/No).
- 20. participa\_comunidades: Participación en comunidades en línea sobre videojuegos (Activo, Ocasional, No).
- 21. percepcion\_industria: Percepción sobre el futuro de la industria (1–5 Likert).
- 22. industria\_empleos: Acuerdo con que la industria genera empleos significativos (1–5 Likert).

En la siguiente tabla se presentan las preguntas que se mencionaron anteriormente y se usaron en la encuesta, así como su categoría y tipo de dato para un mejor entendimiento.

Categoría	Variables clave	Tipo de dato
Demográficos	Edad, Género, Edad inicio	Numérico/categórico
Hábitos	Horas/día, Plataforma, Gastos	Numérico
Impacto	Salud, Vida social, Sueño	Escala Likert
Conocimiento/Percepción	ESRB, Piratería, Violencia	Binario/Likert

# Fase 3. Preparación de los datos

### Limpieza de datos

Corrección de inconsistencias y eliminar redundantes

#### Conversión de respuestas:

Las escalas tipo Likert se cuantificaron con valores del 1 al 5, donde 1 representa la valoración negativa y 5 representa el mayor grado de acuerdo.

En las preguntas con tres opciones de respuesta, se asignaron los valores 1, 3 y 5.

Para las preguntas con dos opciones de respuesta, se utilizaron los valores 1 y 5

#### Ejemplo:

Variables clave:

Horas/día (numérica).
Gasto mensual (numérica).
Impacto en salud (Likert 1-5).
¿Piratería? (binaria: 1=No, 5=Sí).

#### Limpieza:

Eliminar columnas irrelevantes (marca temporal, totales).

Verificar valores atípicos (ej. horas > 24).

# Análisis de confiabilidad (Alpha de Cronbach)

El alfa de Cronbach es un coeficiente que se utiliza para medir la fiabilidad o consistencia interna de un conjunto de ítems en una escala o prueba. En otras palabras, evalúa si los diferentes ítems de una prueba o cuestionario miden el mismo constructo de manera consistente. Un valor alto de alfa de Cronbach indica que los ítems están altamente relacionados entre sí y, por lo tanto, la escala es más fiable.

Resultados al aplicar el alpha de cronbach:

K = Número de ítems

Vi = Varianza independiente

Vt = Varianza total

a= Alfa de cronbach

K	22	
Vi	34.858	
Vt	69.000	
Sección 1	1.048	
Sección 2	0.495	
Abosuluto S2	0.495	
а	0.518	

# Descripción de los resultados

Como se muestra en la imagen anterior, el resultado de nuestro análisis de confiabilidad es de 0.518, esto quiere decir que la consistencia interna es pobre.

# Fase 4. Modelado

Aplicación del estudio de correlaciones (Bivariada de Pearson)

Al aplicar la bivariada de pearson, se detectó que mayormente los datos tienen una reacción entre 0.5 y 0.3, lo que nos dice que es una correlación moderada. A continuación se muestra una parte de la matriz de correlaciones.

		Α	В	С	D	E	F	G	н	1	J	К	L I	M M	١ ٥	Р
		A .	ь			-	-			<u>'</u>	,	κ		IVI I		r
Α	R de Pearson	_														
	gl															
	valor p															
В	R de Pearson	-0.041	_													
		84	_													
	gl	0.710														
	valor p															
С	R de Pearson	-0.241	0.227	_												
	gl	84	84	_												
	valor p	0.025	0.036	_												
D	R de Pearson	0.137	-0.165	-0.358	_											
	gl	84	84	84	_											
	valor p	0.207	0.129	<.001	_											
E	R de Pearson	-0.340	0.248	0.481	-0.349	_										
	gl	84	84	84	84	_										
	valor p	0.001	0.021	<.001	0.001	_										
						0.247										
F	R de Pearson	0.288	-0.235	-0.208	0.289	-0.217	_									
	gl	84	84	84	84	84	_									
	valor p	0.007	0.030	0.055	0.007	0.044	_									
G	R de	0.281	-0.315	-0.341	0.310	-0.370	0.608	_								
	Pearson	0.201	0.515	0.5-11	0.510	0.570	0.000									
	gl	84	84	84	84	84	84	_								
	valor p	0.009	0.003	0.001	0.004	<.001	<.001	_								
н	R de	0.110	-0.162	-0.199	0.174	-0.067	0.131	0.214	_							
	Pearson															
	gl	84	84	84	84	84	84	84	_							
	valor p	0.312	0.135	0.067	0.109	0.542	0.229	0.047	_							
1	R de	0.053	-0.125	-0.140	0.200	-0.053	0.015	0.124	0.409	_						
	Pearson															
	gl	84	84	84	84	84	84	84	84	_						
	valor p	0.626	0.251	0.199	0.065	0.625	0.888	0.256	<.001	_						
J	R de	-0.270	0.016	0.032	-0.023	0.079	0.079	-0.052	-0.225	-0.025	_					
	Pearson															
	gl	84	84	84	84	84	84	84	84	84	_					
	valor p	0.012	0.880	0.770	0.834	0.471	0.472	0.637	0.037	0.823	_					
K	R de	-0.034	0.007	-0.046	-0.058	-0.049	0.088	0.074	-0.001	-0.147	-0.103	_				
	Pearson			_		_										
	gl	84	84	84	84	84	84	84	84	84	84	_				
	valor p	0.758	0.952	0.677	0.597	0.651	0.423	0.500	0.995	0.177	0.346	_				
L		0.055	-0.182	-0.116	0.269	-0.305	0.402	0.266	0.046	0.136	-0.020	0.041	_			
	Pearson															
	gl	84	84	84	84	84	84		84	84	84	84	_			
	valor p	0.615	0.093	0.289	0.012	0.004	<.001	0.013	0.674	0.211	0.854	0.711	_			
M	R de	0.112	0.032	-0.238	-0.213	-0.095	-0.082	0.003	-0.118	-0.193	-0.016	0.031	-0.014	_		
	Pearson															
	gl	84	84	84	84	84	84		84	84	84	84	84	_		
	valor p	0.304	0.770	0.027	0.049	0.385	0.452	0.978	0.279	0.074	0.882	0.775	0.901	_		
N	R de	-0.141	0.210	0.343	-0.368	0.290	-0.152	-0.213	-0.136	0.009	-0.040	0.021	-0.079	0.016	_	
	Pearson															
	gl	84	84	84	84	84	84		84	84	84	84	84	84	_	
	valor p	0.195	0.052	0.001	<.001	0.007	0.164	0.049	0.212	0.931	0.713	0.848	0.469	0.887	_	
0	R de	0.067	-0.008	-0.161	0.156	-0.073	0.007	0.131	-0.033	-0.098	-0.157	0.474	0.009	0.175	-0.046	_
	Pearson															
	gl	84	84	84	84	84	84	84	84	84	84	84	84	84	84	_
	valor p	0.541	0.945	0.138	0.151	0.505	0.948	0.230	0.765	0.369	0.148	<.001	0.933	0.107	0.677	_

### Fase 5. Evaluación

En esta fase se revisan los resultados obtenidos para determinar si cumplen con los objetivos del negocio y de la investigación.

- En un análisis de confiabilidad (alfa de Cronbach), esto significa interpretar el valor α: por ejemplo, 0.56 indica consistencia interna moderada-baja.
- Se verifican supuestos: ¿los ítems miden el mismo constructo?, ¿hay ítems que reducen la fiabilidad?, ¿se requiere depuración?
- Se comparan los resultados con criterios de éxito establecidos al inicio (por ejemplo,  $\alpha \ge 0.6$  para aceptar el instrumento en un estudio exploratorio).
- Se documentan limitaciones y posibles mejoras antes de pasar a la siguiente fase.

# Rangos del Alfa de Cronbach

Alfa de Cronbach	Consistencia Interna
α ≥ <b>0,9</b>	Excelente
0,8 ≤ α < 0,9	Buena
0,7 ≤ α < 0,8	Aceptable
0,6 ≤ α < 0,7	Cuestionable
0,5 ≤ α < 0,6	Pobre
α < <b>0,5</b>	Inaceptable

gplresearch.com

# Fase 6. Despliegue

Aquí se implementan o comunican los hallazgos para que generen valor.

- En tu caso, implica **presentar los resultados del alfa de Cronbach**, correlaciones y conclusiones a las partes interesadas (profesor, equipo de investigación, stakeholders).
- Proponer planes de acción basados en la fiabilidad obtenida:

- $\circ$  Si el  $\alpha$  es bajo, rediseñar ítems o dividir el cuestionario en subescalas más coherentes.
- Si es aceptable, proceder a usar la escala en modelos predictivos o correlacionales.
- Entregar **informes**, **visualizaciones y recomendaciones** para que se apliquen en decisiones o mejoras futuras.

### Selección del DataSet / Encuesta

Usaremos tu **encuesta de 22 ítems tipo Likert**, con 86 registros capturados (tras limpieza quedaron 79 respuestas válidas; ver #4). El objetivo es **identificar patrones de uso e impacto** de los videojuegos y relacionarlos con variables de consumo y percepción. Esto se alinea con los objetivos que ya planteaste en el documento.

# Aplicación de etapa 1 y 2 de CRISP-DM

#### Fase 1 - Entendimiento del negocio:

- **Objetivo general**: conocer hábitos, consumo y percepciones para detectar patrones de uso e impactos (salud/vida social) y generar recomendaciones/acciones.
- Objetivos específicos:
  - Quiénes juegan, frecuencia y plataformas.
  - o Impacto percibido en salud/vida social.
  - o Percepciones sobre industria, piratería y clasificaciones.
  - o Gasto mensual y acumulado.
- Criterios de éxito: obtener fiabilidad aceptable del instrumento (alfa ≥ 0.6–0.7 deseable), encontrar correlaciones interpretables y traducibles a acciones.

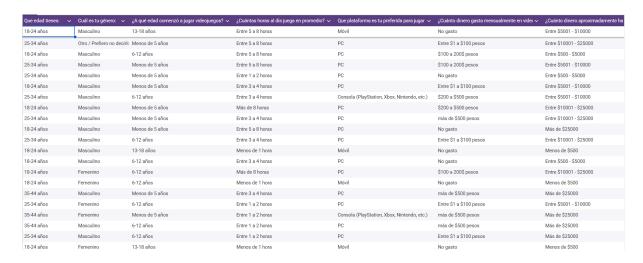
#### Fase 2 – Entendimiento de los datos (lo que observamos):

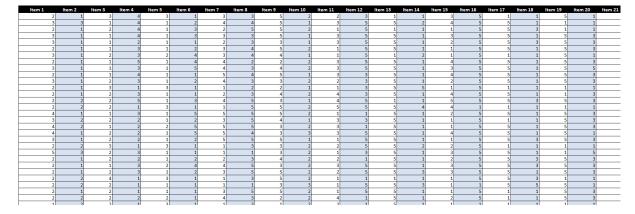
• **Estructura**: 22 ítems tipo Likert (1–5) + campos auxiliares (marca temporal, total).

- Calidad: valores completos en la mayoría de ítems; algunas filas no válidas (resúmenes/ marcadores) que se excluyeron en la limpieza.
- Tamaño útil: 79 respuestas válidas (tras limpieza).

# Pre-procesamiento de datos

Se revisó la base de datos para identificar y eliminar registros incompletos o filas con marcadores no válidos. Se seleccionaron únicamente las columnas correspondientes a los ítems del cuestionario y se convirtieron los valores a formato numérico. Este proceso aseguró que el análisis posterior de confiabilidad y correlaciones se realizará únicamente con respuestas válidas y consistentes.





# Confiabilidad: Alfa de Cronbach

Calculé el alfa de Cronbach a partir de los 22 ítems (n=79):

- $\alpha = 0.518$
- Interpretación: consistencia interna moderada-baja. Como referencia práctica: ≈0.6 "aceptable" en contextos exploratorios; ≥0.7 "aceptable/buena". Esto sugiere revisar redacción de ítems, invertir ítems si aplica, o segmentar el instrumento en subescalas coherentes para mejorar α.

К	22
Vi	34.858
Vt	69.000
Sección 1	1.048
Sección 2	0.495
Abosuluto S2	0.495
а	0.518

# Correlaciones (Bivariada de Pearson) Y Tabla de predicciones

• Top 10 correlaciones absolutas.

Top 10 correlaciones (|r|):

$$Var H \leftrightarrow Var G - r = 0.608$$

#### Predicción:

Si **G** aumenta, **H** tiende a ser mayor.

- Implementar campañas de **conciencia de gasto** para jugadores con gasto mensual alto, fomentando presupuestos responsables.
- Ofrecer herramientas de seguimiento y límites opcionales en compras.
- Monitorear cambios en gasto acumulado tras las intervenciones.



### $Var F \leftrightarrow Var D - r = 0.481$

#### Predicción:

Si D aumenta, F tiende a ser mayor.

- Diseñar actividades o contenidos que fortalezcan de forma positiva ambas variables, evitando efectos no deseados.
- Introducir incentivos cruzados que fomenten el desarrollo equilibrado de ambas conductas.
- Evaluar mediante encuestas y métricas de uso.

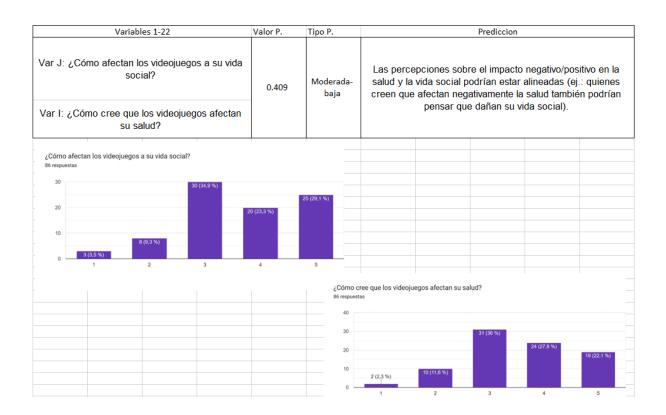
		1	
/ar F: Que plataforma es tu preferida para jugar?	0.481	Moderada	La plataforma preferida podría estar relacionada con la edad a la que comenzó a jugar. Por ejemplo, quienes empezaron más jóvenes podrían preferir consolas clásicas, mientras que
Var D: ¿A qué edad comenzó a jugar videojuegos?			quienes empezaron más tarde podrían inclinarse por PC/móviles.
Que plataforma es tu preferida para jugar 86 respuestas			¿A qué edad comenzó a jugar videojuegos?
25,6%			12,8%
52,3%			19,8%
● PC			
<ul> <li>Consola (PlayStation, Xbox, Nintendo, etc.)</li> </ul>			Menos de 5 años
Móvil			6-12 años
● En la nube (GeForce Now, Xbox Cloud,			● 13-18 años ● 19-25 años
etc. Otra			25 años o más

### $Var J \leftrightarrow Var I - r = 0.409$

Predicción:

Si I aumenta, J tiende a ser mayor.

- Crear módulos o experiencias donde el aumento de I (p. ej., uso de cierto tipo de juego o conocimiento) potencie J (comportamiento asociado).
- Monitorear correlación tras aplicar campañas informativas o de motivación

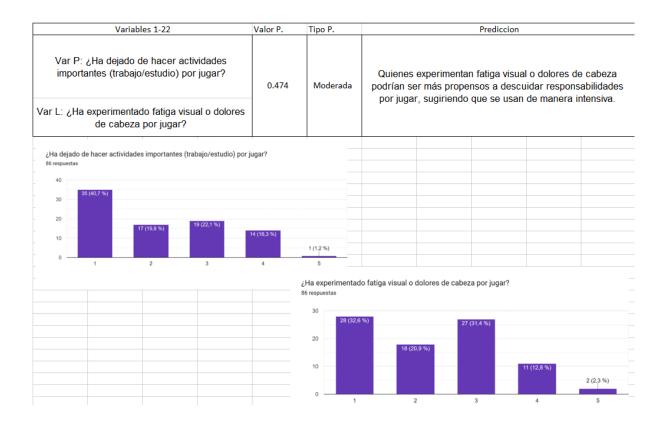


# $Var P \leftrightarrow Var L - r = 0.474$

#### Predicción:

Si L aumenta, P tiende a ser mayor.

- Integrar promociones o programas que fomenten simultáneamente ambas conductas.
- Crear "eventos especiales" que incentiven la participación en L para impulsar P.
- Medir cambios en la frecuencia conjunta.

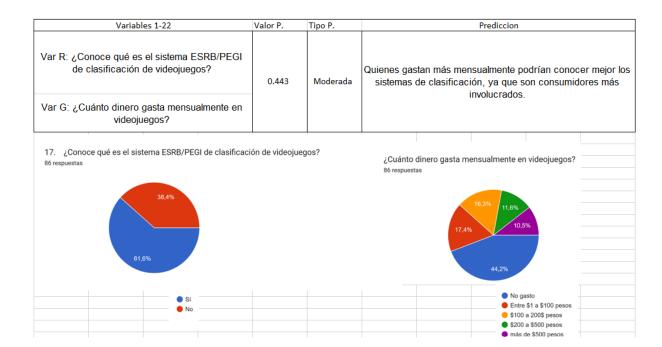


### $Var R \leftrightarrow Var G - r = 0.443$

Predicción:

Si G aumenta, R tiende a ser mayor.

- Ofrecer beneficios adicionales a jugadores con gasto mensual alto para reforzar la dimensión R (p. ej., participación en comunidades o eventos).
- Evaluar si esta relación puede aprovecharse para programas de fidelización.

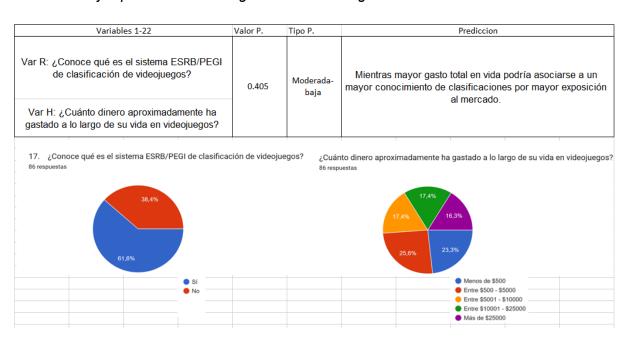


### $Var R \leftrightarrow Var H - r = 0.405$

#### Predicción:

Si H aumenta, R tiende a ser mayor.

- Diseñar encuestas o contenido especial para quienes han gastado más históricamente, buscando reforzar la conducta **R**.
- Incluir mensajes personalizados según el historial de gasto.



# $Var S \leftrightarrow Var R - r = 0.417$

Predicción:

Si R aumenta, S tiende a ser mayor.

- Integrar actividades que combinen los comportamientos R y S, fomentando la participación simultánea.
- Incentivar acciones en R que naturalmente lleven a incrementos en S.

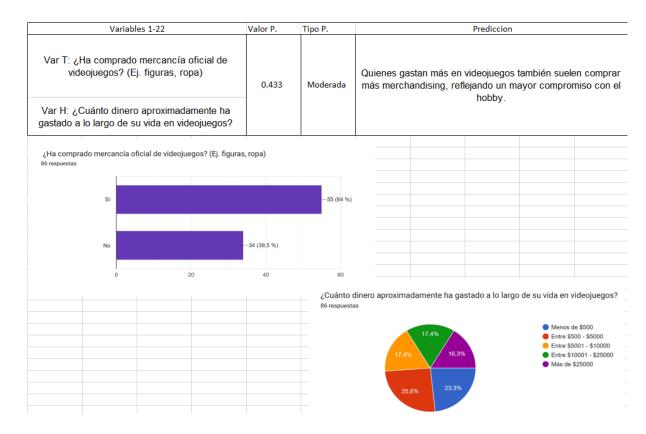
Variables 1-22	Valor P.	Tipo P.	Prediccion
Var S: ¿Toma en cuenta la clasificación ESRB/PEGI a la hora de comprar/regalar un juego?	0.417	Moderada- baja	Quienes conocen el sistema de clasificación son más propensos a usarlo al comprar/regalar juegos.
Var R: ¿Conoce qué es el sistema ESRB/PEGI de clasificación de videojuegos?			
¿Toma en cuenta la clasificación ESRB/PEGI a la hora de 86 respuestas	comprar/rega		¿Conoce qué es el sistema ESRB/PEGI de clasificación de videojuegos?
8,1%			38,4% 61,6%

### $Var T \leftrightarrow Var H - r = 0.433$

#### Predicción:

Si H aumenta, T tiende a ser mayor.

- Aprovechar el historial de gasto como criterio para campañas que promuevan la conducta T (p. ej., compra de cierto tipo de producto o participación en eventos).
- Segmentar y personalizar incentivos.

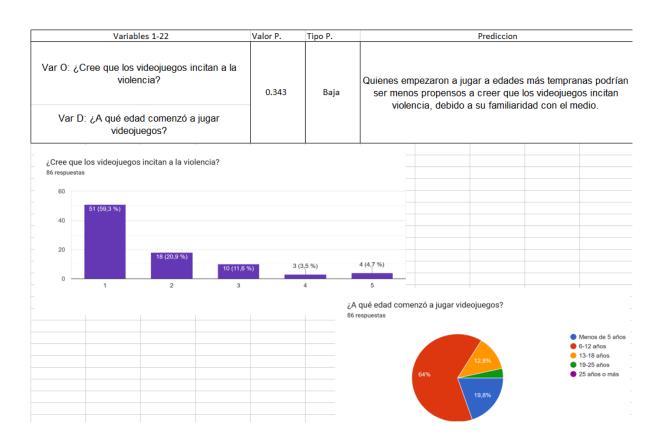


### $Var O \leftrightarrow Var D - r = 0.343$

#### Predicción:

Si D aumenta, O tiende a ser mayor.

- Diseñar intervenciones que, al reforzar D, estimulen también O.
- Verificar si esta relación es aprovechable para maximizar impacto en ambas variables de forma simultánea.



# $Var Q \leftrightarrow Var H - r = 0.324$

#### Predicción:

Si H aumenta, Q tiende a ser mayor.

- Ofrecer descuentos o promociones especiales a jugadores con alto gasto acumulado.
- Promocionar alternativas legales atractivas (suscripciones, demos, ofertas).
- Incluir mensajes educativos sobre los riesgos y beneficios de opciones legales en el punto de compra.

