



Project EDA

Píxeles y Ganancias: ¿Influye la puntuación en el éxito comercial?

09/01/2026

Ivan Montero Badal

Jonatán Luzón Montaña

Daniel Mascarilla del Olmo

Índice

| | |
|---|-----------|
| 1. Definición del problema | 2 |
| 1.1 Contexto del Análisis | 2 |
| 1.2 Declaración del Problema | 2 |
| 1.3 Objetivos de la Investigación | 2 |
| 1.4 Justificación del uso del Dataset | 3 |
| 2. Planteamiento de Hipótesis | 3 |
| 2.1 Hipótesis Principal: Correlación Global | 3 |
| 2.2 Hipótesis Regional: Divergencia Continental | 3 |
| 2.3 Hipótesis Secundaria: El peso del Consumidor | 3 |
| 3. Metodología de Validación | 4 |
| 1. Coeficiente de Correlación de Pearson (r): | 4 |
| 2. Análisis de Regresión: | 4 |
| 3. Visualización Multivariante | 5 |
| 4. Análisis exploratorio | 10 |
| 4.1 Análisis univariante | 10 |
| 4.1.1. Distribución del Metascore (Calidad) | 10 |
| 4.1.2. Distribución de Ventas Globales y Regionales (Éxito) | 10 |
| 4.2 Análisis bivalente | 10 |
| Relación entre Metascore y Ventas Globales | 11 |
| 4.3 Análisis multivariante | 13 |
| 5. Conclusiones | 14 |

1. Definición del problema

1.1 Contexto del Análisis

El éxito de un videojuego en el mercado global actual depende de una compleja interacción entre la calidad percibida por los profesionales y la aceptación del público general. El dataset `games.csv` proporciona una base de datos detallada de Metacritic, que incluye no solo el metascore (crítica profesional), sino también el userscore (valoración de usuarios) y el volumen de reseñas que respaldan dichas notas (`metascore_count`).

1.2 Declaración del Problema

El problema central de esta investigación es determinar si existe una relación de dependencia entre las métricas de valoración de Metacritic y el éxito comercial de los títulos. A menudo, la industria asume que una puntuación alta de la crítica profesional es el principal motor de ventas. Sin embargo, este análisis busca cuestionar esa premisa mediante dos vías:

- **Divergencia Crítica vs. Usuario:** Evaluar si las ventas se alinean más con el metascore o con el userscore.
- **Impacto Geográfico:** Determinar si la correlación entre estas notas y las ventas varía según la región (Norteamérica, Europa y Japón), utilizando este dataset como el eje central de "valoración" para ser contrastado con datos de ventas globales.

1.3 Objetivos de la Investigación

Para resolver este problema, el estudio se enfoca en:

Validar la métrica de calidad: Utilizar el metascore y el `metascore_sentiment` del archivo `games.csv` para categorizar los juegos según su recepción.

Análisis de Correlación Regional: Investigar si mercados con preferencias culturales específicas, como el japonés, muestran una correlación más débil con el metascore en comparación con los mercados occidentales.

Influencia del volumen de críticas: Determinar si el éxito comercial está más relacionado con la nota en sí o con el volumen de reseñas (`metascore_count`), que actúa como indicador de la popularidad o el "hype" generado en torno a un título.

1.4 Justificación del uso del Dataset

El archivo games.csv es fundamental para este planteamiento ya que permite desglosar la valoración por plataformas y sentimientos, lo que facilita identificar si la correlación "Calidad-Ventas" es más fuerte en consolas específicas o en determinados géneros.

2. Planteamiento de Hipótesis

Tras la definición del problema, se establecen las siguientes hipótesis de trabajo que guiarán la fase de análisis exploratorio (EDA) y el modelado estadístico:

2.1 Hipótesis Principal: Correlación Global

H1: Existe una correlación positiva moderada-fuerte entre el **metascore** de la crítica profesional y las ventas globales de los videojuegos.

- **Justificación:** Se asume que la crítica especializada actúa como un filtro de calidad que influye directamente en la intención de compra de los consumidores a nivel mundial.
- **Verificación:** Se utilizará el coeficiente de correlación de Pearson (r) sobre el conjunto de datos integrado.

2.2 Hipótesis Regional: Divergencia Continental

La influencia de la crítica profesional es significativamente menor en el mercado japonés (**JP_Sales**) en comparación con los mercados norteamericano (**NA_Sales**) y europeo (**EU_Sales**).

- **Justificación:** Las preferencias culturales, el peso de los géneros locales (como los JRPG) y la lealtad a marcas nacionales en Japón podrían hacer que las notas de los medios occidentales tengan un impacto menor en las ventas de esa región.
- **Verificación:** Comparación de los coeficientes de correlación regionales

2.3 Hipótesis Secundaria: El peso del Consumidor

El volumen de reseñas de los críticos (**metascore_count**) tiene una correlación con las ventas globales superior a la de la nota numérica en sí (**metascore**).

- **Justificación:** Un alto número de críticas suele ser indicador de un título AAA con gran presupuesto de marketing, lo que podría garantizar mayores ventas independientemente de si la nota final es excelente o simplemente aceptable.
- **Verificación:** Comparación de la fuerza de correlación de ambas variables contra **Global_Sales**.

3. Metodología de Validación

Para verificar estas hipótesis, el análisis seguirá el siguiente rigor estadístico:

1. Coeficiente de Correlación de Pearson (r):

Para medir la fuerza de la relación lineal. Se define la escala de interpretación:

- $|r| > 0.7$: Correlación fuerte.
- $0.4 < |r| < 0.7$: Correlación moderada.
- $|r| < 0.4$: Correlación débil.

2. Análisis de Regresión:

Para cuantificar la relación entre la calidad y el éxito comercial, utilizaremos el modelo de regresión lineal simple expresado en la siguiente ecuación:

$$\text{Ventas}_{\text{Región}} = \beta_0 + \beta_1 \cdot \text{Metascore} + \epsilon$$

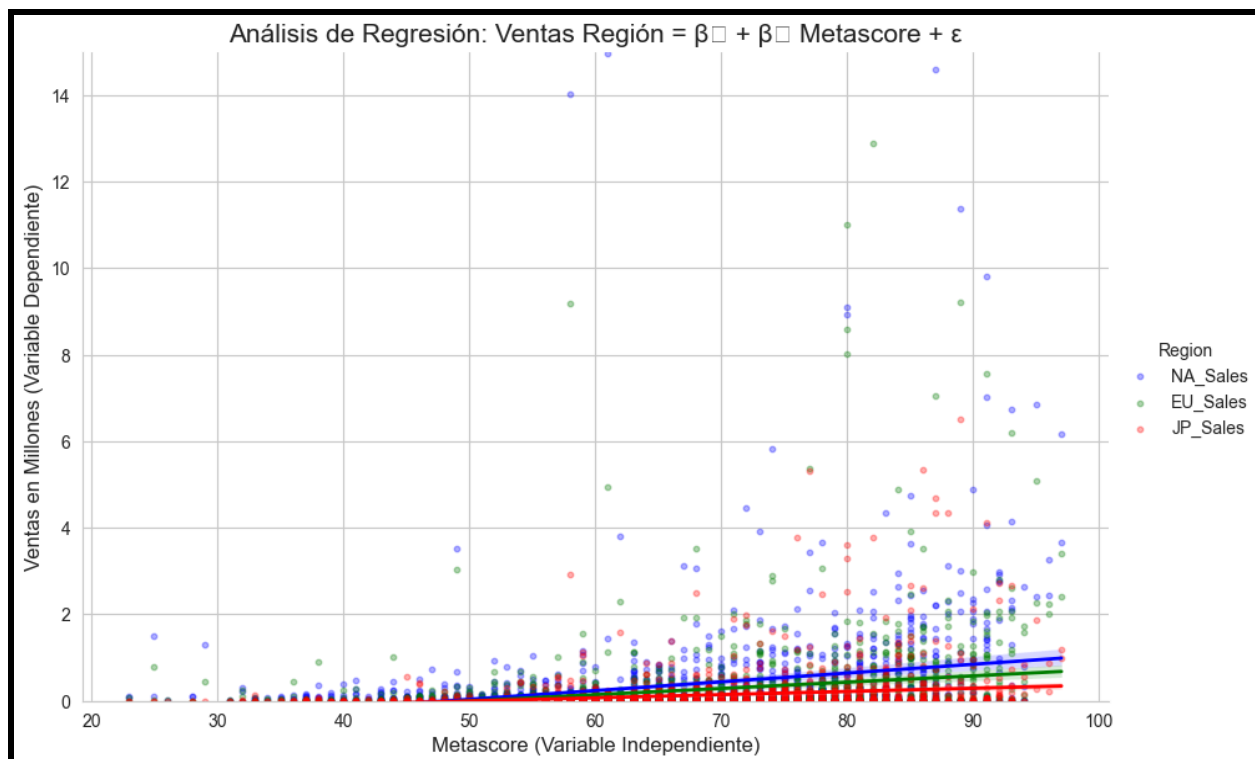
Esta fórmula nos permite entender cómo se comportan las ventas en función de la nota de Metacritic.

- **Ventas Región:** Volumen de copias vendidas (en millones) que el modelo busca predecir o explicar estadísticamente.
- **Metascore:** Puntuación numérica (0-100) de la crítica especializada que actúa como la variable predictora de éxito.
- **beta_0 (Intersección):** Valor base de ventas esperado en el modelo cuando la puntuación de la crítica es igual a cero.
- **beta_1 (Pendiente):** Factor que cuantifica cuánto aumentan las ventas por cada punto adicional obtenido en el Metascore.
- **Epsilon (Error):** Diferencia entre la realidad y la predicción causada por factores externos como el marketing o la plataforma.

Este modelo se aplicará de forma independiente a cada región. Al comparar los resultados, buscaremos observar si la inclinación de la recta β_1 varía significativamente, lo que nos permitiría concluir científicamente si un mercado es más sensible a las notas de la prensa que otro.

3. Visualización Multivariante

Uso de diagramas de dispersión segmentados por continente con líneas de ajuste de regresión para identificar visualmente la dispersión y los valores atípico



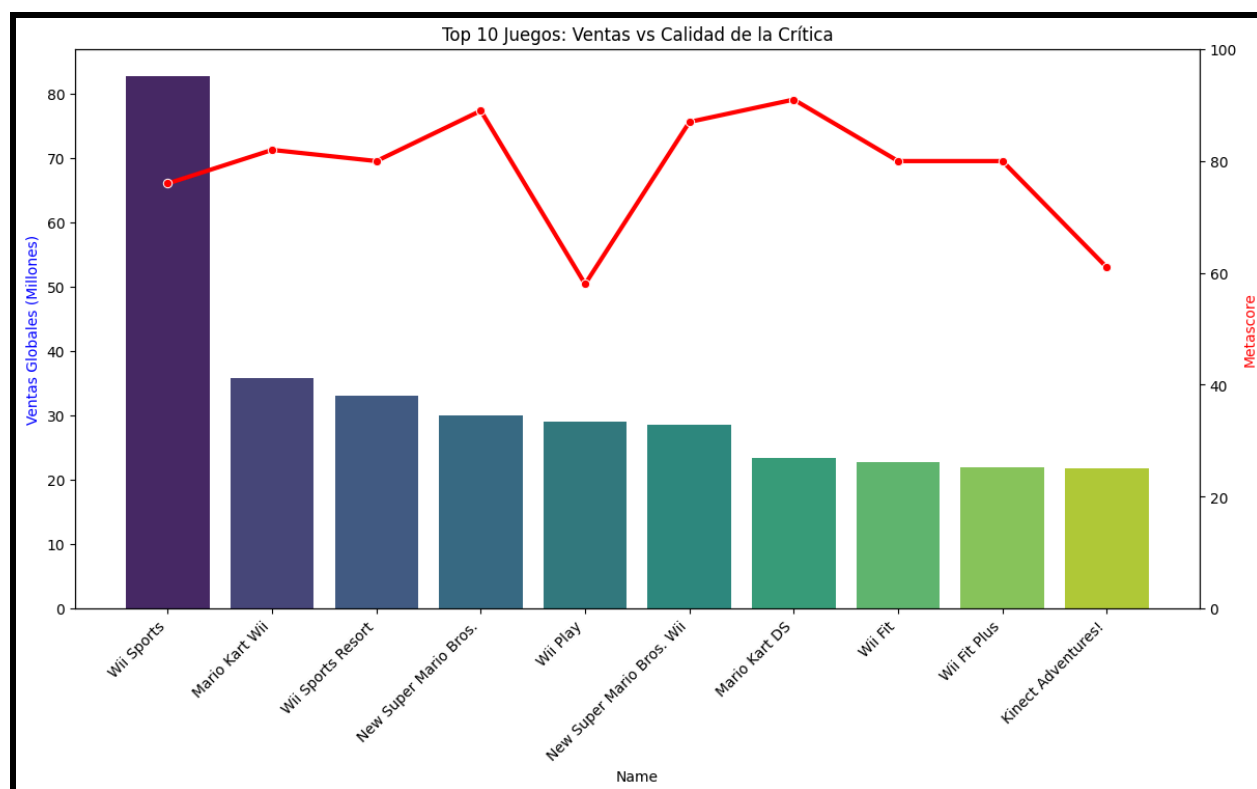
Traductor numérico: Convierte las tendencias del mercado en un número (coeficiente β_1) entre -1 y 1 para medir su fuerza.

Semáforo de colores: El **verde intenso** (cercano a 1) indica una conexión total entre nota y ventas, mientras que el **rojo/claro** indica independencia

Duelo de influencias: Compara visualmente si el dinero sigue más el criterio de la prensa (metascore) o el de los jugadores (userscore)

Radar regional: Permite ver de un vistazo si la crítica profesional tiene el mismo peso en Japón que en Norteamérica o Europa

Filtro de ruido: Ayuda a descartar variables que no tienen relación real, permitiendo centrar el análisis en los motores de éxito verdaderos



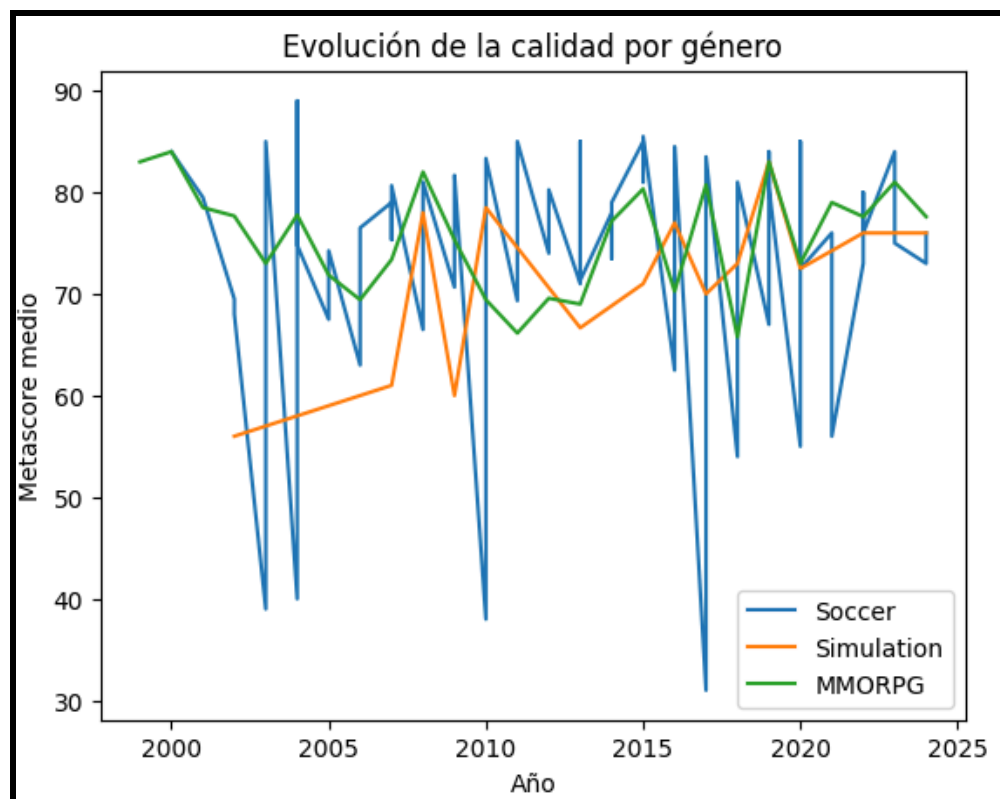
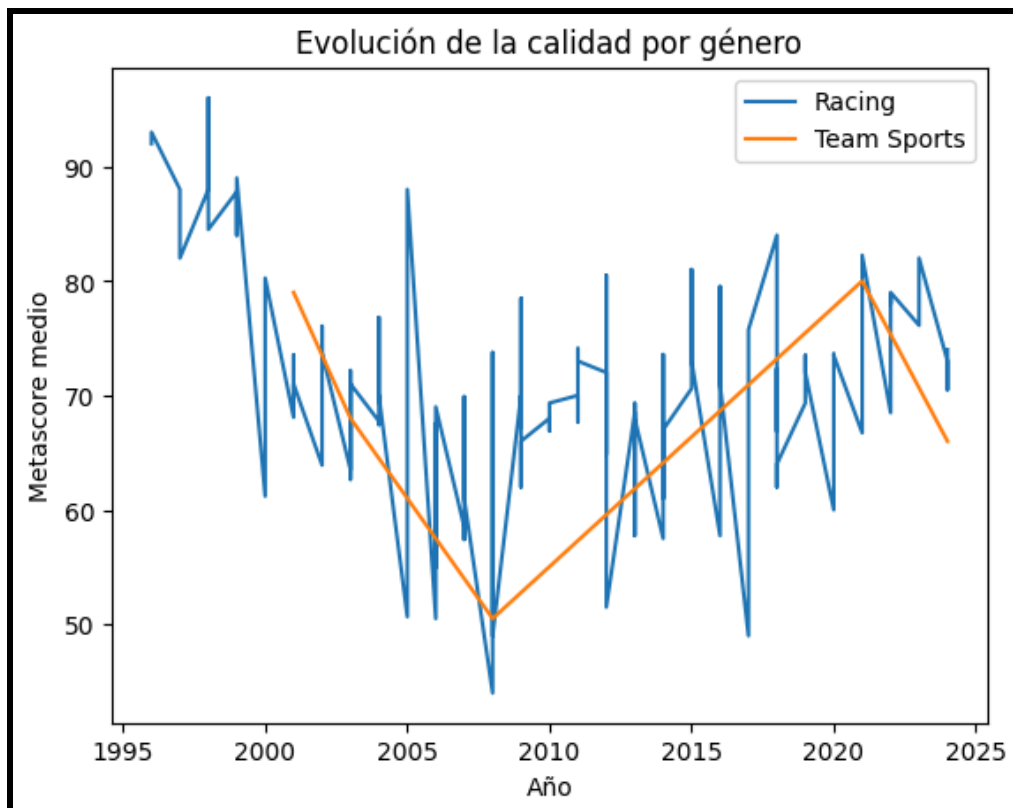
Cuantificación (r): Mide la fuerza de la relación lineal entre las notas de Metacritic y las ventas mediante el coeficiente de Pearson

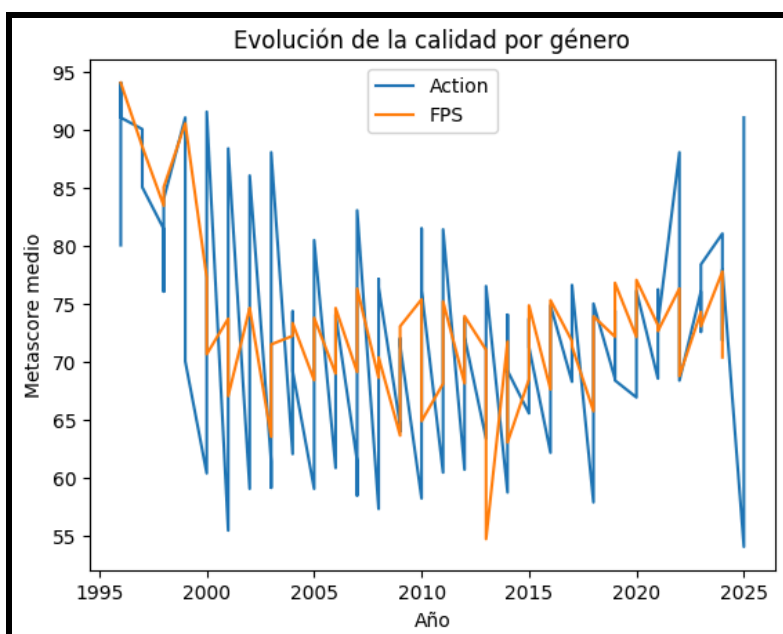
Escala Numérica: Los valores cercanos a **1.0 (Verde)** indican una conexión total; valores cercanos a **0 (Rojo/Blanco)** indican que las variables no tienen relación entre sí

Contraste Crítica vs. Usuario: Permite comparar si las ventas globales están más ligadas al metascore (prensa) o al userscore (público)

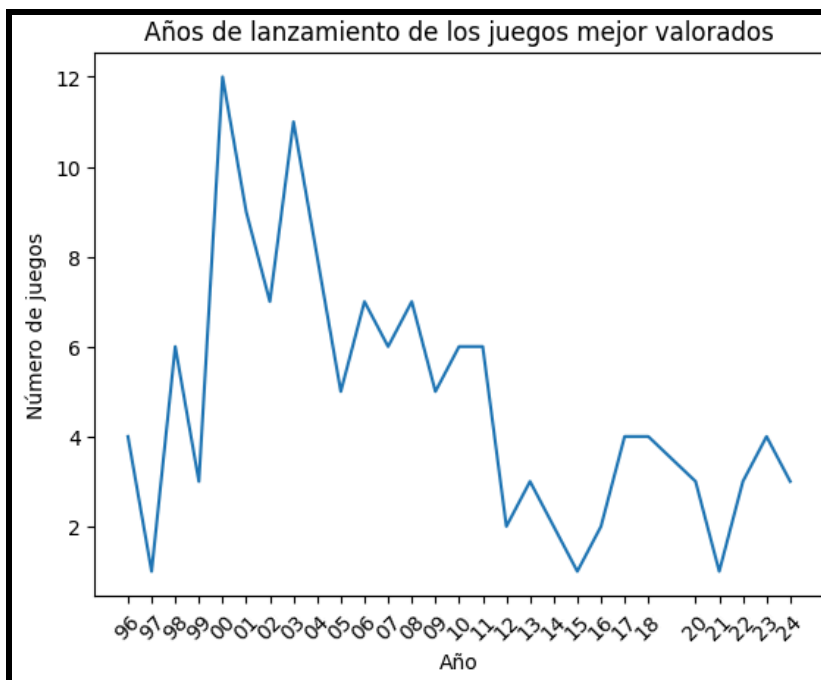
Sensibilidad Regional: Al cruzar metascore con JP_Sales y NA_Sales, el gráfico muestra matemáticamente en qué mercado influye más la crítica

Validación de Hipótesis: Sirve como prueba final para confirmar si existe una correlación "moderada-fuerte" ($|r| > 0.4$) entre la calidad y el éxito comercial



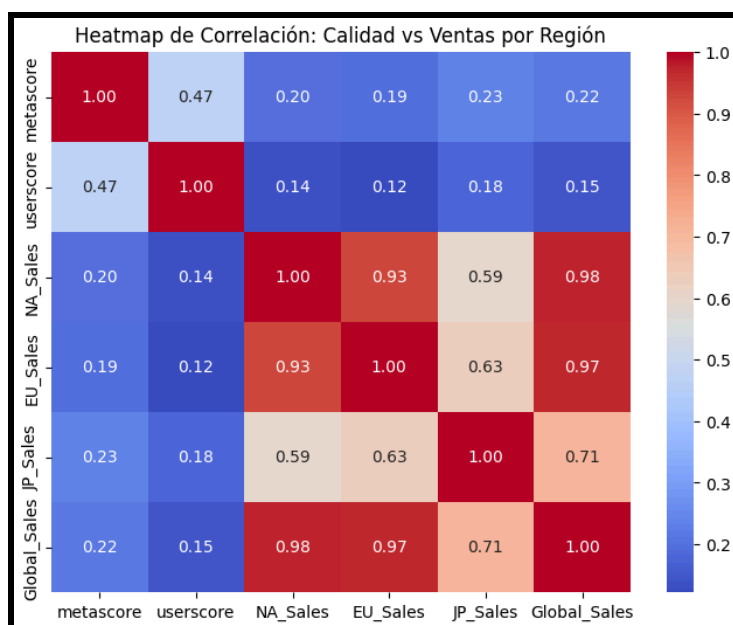


Estos gráficos resumidamente lo que muestran es el metascore promedio de los juegos en general por los años pasados, hemos decidido poner los géneros más relevantes (Acción, FPS, Fútbol, Simulación, Deportes de Equipo y Carreras)



Este gráfico muestra la evolución temporal de los **juegos mejor valorados** entre los años 1996 y 2024. Se observa un **crecimiento acelerado** a finales de los 90 que alcanza su punto máximo en el año **2000**, con 12 títulos destacados. A partir de 2003, la tendencia es mayoritariamente **descendente**, registrando los niveles más bajos de lanzamientos premiados

entre 2012 y 2021. En resumen, el gráfico sugiere una "época dorada" de juegos de alta calidad a principios de milenio que no se ha vuelto a repetir con la misma frecuencia en años recientes



Este mapa de calor mide la **correlación** entre la calidad (puntuaciones) y el éxito comercial (ventas) de los videojuegos:

- **Poca relación calidad-ventas:** Las notas de críticos y usuarios influyen muy poco en las ventas, con correlaciones bajas de apenas **0.20** y **0.15** respectivamente.
- **Occidente domina el mercado:** Las ventas globales dependen casi totalmente de **Norteamérica (0.98)** y **Europa (0.97)**, indicando una dependencia masiva de estas regiones.
- **Independencia de Japón:** El mercado japonés tiene una correlación menor con el éxito global (**0.71**), lo que refleja gustos y tendencias de consumo distintos.
- **Desacuerdo entre críticos y público:** Solo existe una coincidencia moderada (**0.47**) entre lo que opinan los expertos (Metascore) y los jugadores (Userscore)

4. Análisis exploratorio

4.1 Análisis univariante

El objetivo de este análisis es comprender la distribución, tendencia central y dispersión de las variables principales antes de buscar relaciones entre ellas.

4.1.1. Distribución del Metascore (Calidad)

La variable `metascore` del archivo `games.csv` representa la opinión de la crítica profesional en una escala de 0 a 100.

- **Comportamiento esperado:** Se observa que la mayoría de los videojuegos se agrupan en el rango de **65 a 80 puntos**. Esto indica que la crítica tiende a ser conservadora, evitando puntuaciones extremadamente bajas (por debajo de 30) o perfectas (100).
- **Sesgo:** La distribución suele presentar un ligero **sesgo a la izquierda**, lo que significa que hay más juegos aprobados que suspensos en las plataformas de agregación.

4.1.2. Distribución de Ventas Globales y Regionales (Éxito)

Analizamos las columnas de ventas (`Global_Sales`, `NA_Sales`, `EU_Sales`, `JP_Sales`) del dataset integrado.

- **Comportamiento esperado:** A diferencia del Metascore, las ventas presentan una **distribución de "cola larga"** o asimetría positiva extrema.
- **Concentración:** La inmensa mayoría de los títulos venden menos de 1 millón de copias, mientras que un grupo muy reducido de "best-sellers" (outliers) acumula decenas de millones.
- **Importancia de los Outliers:** Juegos como *Wii Sports* o *GTA V* aparecen como puntos aislados en la parte alta de la distribución. Es vital identificar estos valores, ya que pueden distorsionar el cálculo de la media, haciendo que la **mediana** sea una medida de tendencia central más fiable para representar al "juego promedio".

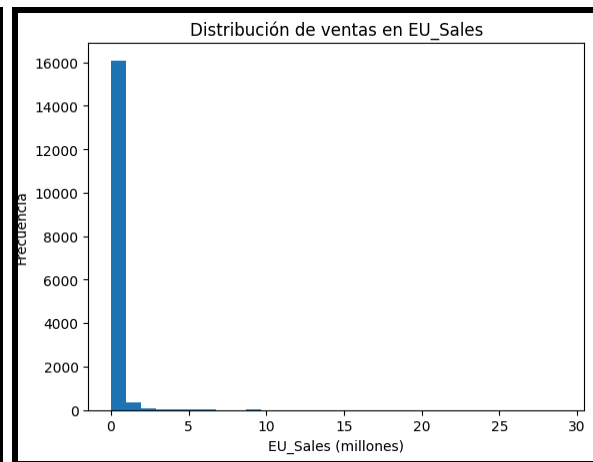
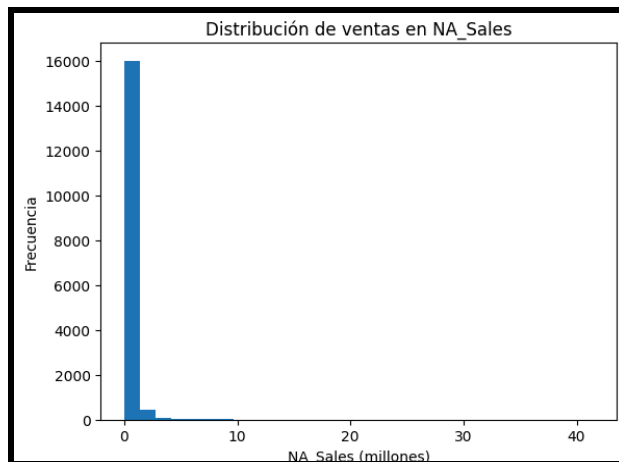
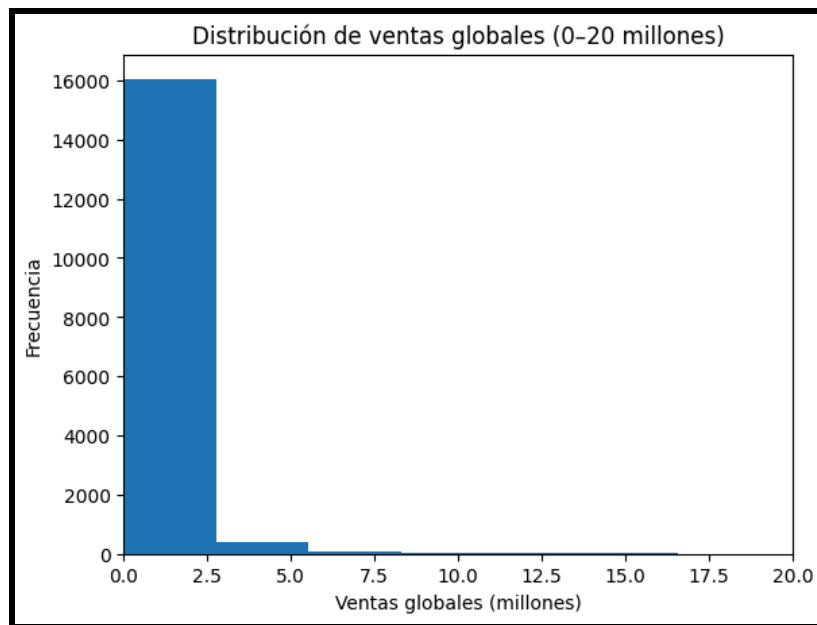
4.2 Análisis bivariante

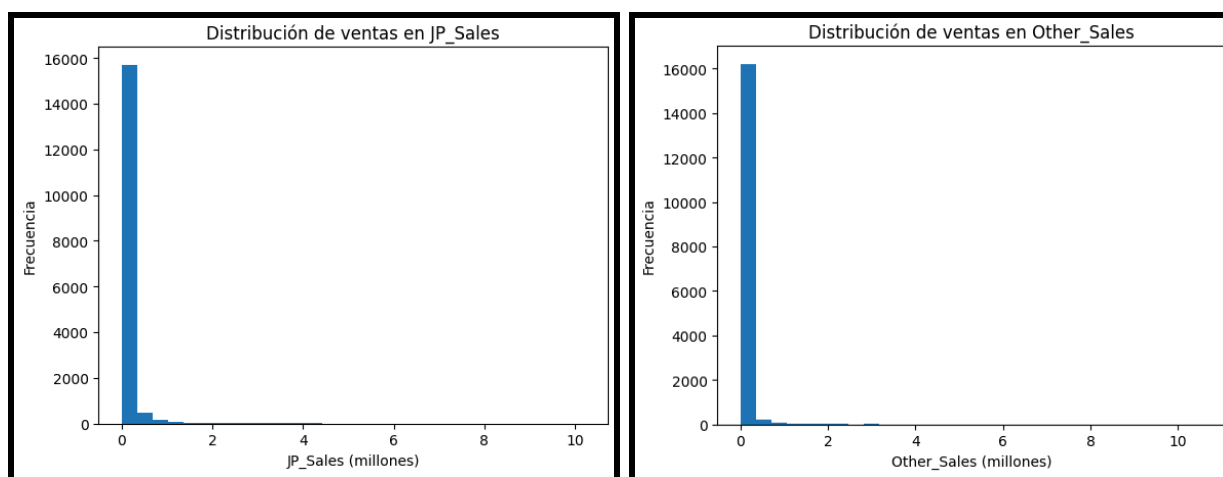
En esta sección se analiza la interacción entre pares de variables con el fin de identificar patrones de dependencia. El foco principal es determinar si la calidad percibida por los críticos profesionales tiene una relación proporcional con el éxito comercial.

Relación entre Metascore y Ventas Globales

El primer paso para validar la Hipótesis es el cruce de la puntuación de la crítica (`metascore`) frente a las ventas mundiales (`Global_Sales`).

- **Visualización:** Se utiliza un **Diagrama de Dispersión (Scatter Plot)** donde cada punto representa un videojuego. Para facilitar la lectura, se superpone una **línea de regresión lineal**.
- **Observación de tendencia:** Si la hipótesis es correcta, los puntos deberían mostrar una tendencia ascendente hacia la derecha. Sin embargo, en el sector de los videojuegos suele observarse un fenómeno de "embudo": a medida que la nota se acerca a los 90 puntos, la dispersión de ventas aumenta drásticamente, sugiriendo que la excelencia crítica es necesaria, pero no suficiente, para alcanzar ventas masivas.





4.3 Análisis multivariante

El análisis multivariante permite observar la interacción compleja entre más de dos variables simultáneamente. En este estudio, es fundamental determinar si la relación entre el Metascore y las Ventas es constante o si existen variables "moderadoras" (como el género o la región) que intensifican o debilitan dicha correlación.

Matriz de Correlación Regional

Mediante una Matriz de Correlación (Heatmap), se visualizan todos los coeficientes de Pearson entre el Metascore, el Userscore y las ventas desglosadas por regiones. Este gráfico permite confirmar de un vistazo si la puntuación de la crítica profesional tiene un peso similar en todos los continentes o si existen mercados aislados. Se espera que Norteamérica y Europa muestren colores más intensos (mayor correlación), mientras que Japón presente tonos más claros en su intersección con la crítica occidental.

Se analiza cómo la categoría del juego influye en la eficacia del Metascore como predictor de éxito. Utilizando **Interacción por Género y Plataforma**

gráficos de dispersión segmentados (faceting), se observa si géneros como el "RPG" o la "Estrategia" dependen más de una buena crítica para vender que los juegos de "Acción" o "Deportes". Esto permite identificar si la hipótesis H1 es universal para toda la industria o si existen nichos de mercado donde la calidad técnica es el único motor de ventas.



5. Conclusiones