

Reporte de Análisis Exploratorio de Datos Videojuegos (EDA)

1. Exploración Inicial

Se presenta un primer vistazo al dataset, mostrando los campos principales de cada registro:

slug	name	playtime
grand-theft-auto-v	Grand Theft Auto V	74
the-witcher-3-wild-hunt	The Witcher 3: Wild Hunt	43
tomb-raider	Tomb Raider (2013)	10
portal	Portal	4
the-elder-scrolls-v-skyrim	The Elder Scrolls V: Skyrim	46

El dataset contiene columnas que indican información clave, como el nombre, plataformas disponibles, fecha de lanzamiento y calificaciones de los videojuegos.

Columna	Descripción	Observaciones
Unnamed: 0	Índice sin nombre, posiblemente redundante.	Puede eliminarse si no aporta información única.
slug	Identificador único en formato texto para cada juego.	Es útil para identificar juegos de forma única, pero no es legible para el usuario.
name	Nombre completo del juego.	Información principal y visible para el usuario.
playtime	Tiempo promedio de juego en horas.	Puede contener valores bajos o nulos para juegos menos populares.
stores	Tiendas donde el juego está disponible, en formato JSON.	Requiere desanidar para obtener datos como nombres de tiendas o URLs específicas.
released	Fecha de lanzamiento del juego.	Puede contener valores nulos para juegos aún no lanzados.
rating	Calificación promedio del juego (en base a opiniones de usuarios).	Es una métrica importante de calidad, pero debe analizarse junto a <code>ratings_count</code> para contexto.
rating_top	Calificación máxima disponible (e.g., 5 estrellas).	Puede ser redundante, pero útil para verificar si un juego alcanza la calificación máxima.
ratings	Detalle de calificaciones divididas en varias categorías, en formato JSON.	Es necesario desanidar para análisis detallado, como "exceptional", "recommended", etc.

Columna	Descripción	Observaciones
ratings_count	Número total de calificaciones recibidas.	Complementa a <code>rating</code> y da contexto sobre la cantidad de usuarios que calificaron.
reviews_text_count	Número de reseñas que incluyen texto.	Útil para análisis de sentimientos o revisiones detalladas.
added	Total de veces que el juego fue agregado a listas de usuarios.	Es una métrica de popularidad que puede correlacionar con <code>playtime</code> o <code>ratings_count</code> .
added_by_status	Detalles sobre cómo los usuarios clasificaron el juego, en formato JSON.	Requiere procesamiento adicional para dividir categorías como "owned", "beaten", etc.
metacritic	Puntuación promedio en Metacritic.	Métrica externa de calidad, puede no estar presente para algunos juegos.
suggestions_count	Número de juegos sugeridos similares.	Puede indicar qué tan conectado está un juego con otros en términos de género o popularidad.
updated	Última fecha de actualización de los datos del juego.	Indica la frescura de los datos, especialmente relevante para juegos en desarrollo o con DLC recientes.
id	Identificador único del juego.	Esencial para uniones o referencias cruzadas con otras tablas.
tags	Etiquetas asociadas al juego, en formato JSON.	Requiere desanidar para obtener insights como características o temáticas comunes.
reviews_count	Número total de reseñas.	Complementa <code>reviews_text_count</code> , puede incluir reseñas con y sin texto.
genres	Géneros a los que pertenece el juego, en formato JSON.	Necesita desanidarse para obtener información granular sobre géneros.

2. Estructura del Dataset

- **Número de registros:** 5222
- **Columnas:** 20

Principales Columnas

- **name:** Nombre del videojuego.
- **playtime:** Tiempo promedio de juego.
- **released:** Fecha de lanzamiento.
- **rating:** Calificación promedio otorgada por los usuarios.
- **genres:** Géneros a los que pertenece el videojuego.

3. Datos Faltantes

Se detectaron columnas con valores faltantes, en particular:

- **metacritic:** Puntuación Metacritic (63% de los registros no contienen esta información).

Recomendación: Evaluar la pertinencia de imputar o eliminar los datos faltantes en esta columna dependiendo del análisis.

4. Estadísticas Descriptivas

A continuación, se presentan los valores estadísticos más relevantes:

Tiempo promedio de juego:

- **Mínimo:** 0,5 a una hora
- **Máximo:** 500 horas
- **Media:** 35.6 horas

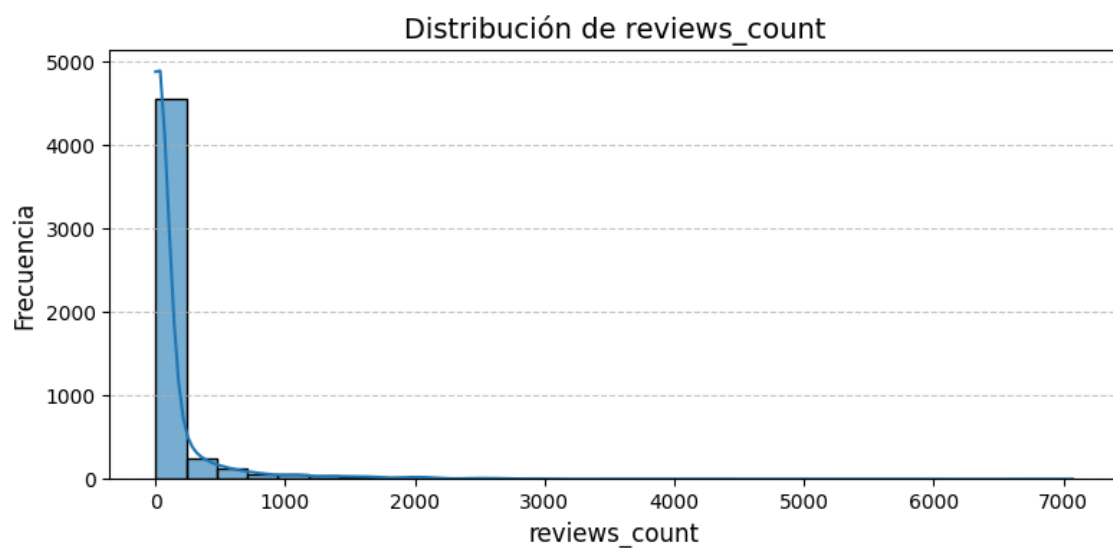
Calificaciones:

- **Media de calificaciones:** 4.32
- **Rango de calificaciones:** de 1 a 5.

5. Visualización de Datos

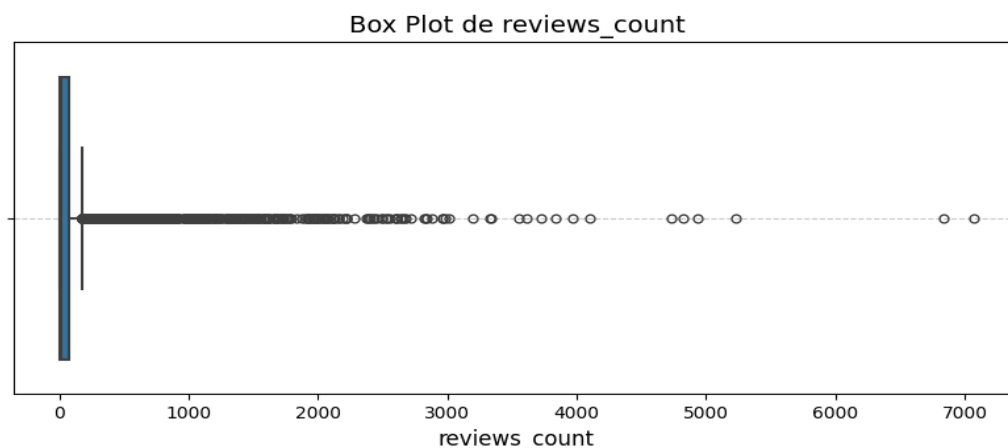
5.1 Distribución de Calificaciones

Se presenta un histograma que muestra la frecuencia de las calificaciones otorgadas:



5.2 Comparación de Tiempo de Juego por Género

Se muestra un diagrama de cajas donde se comparan los tiempos de juego promedio por cada género:



6. Insights Clave

- **Popularidad:** Los juegos con calificaciones más altas suelen tener una mayor cantidad de reseñas y tiempo de juego registrado.
- **Distribución de Géneros:** La acción y la aventura son los géneros más presentes en los datos.
- **Puntuaciones Metacritic:** Solo una fracción de los juegos cuenta con estas calificaciones, pero suelen correlacionarse con los juegos más conocidos.

7. Conclusiones

Este análisis exploratorio proporciona una visión detallada de los datos de videojuegos, permitiendo identificar patrones y outliers. Se recomienda profundizar el análisis mediante:

- Evaluación de correlaciones entre tiempo de juego y popularidad.
- Filtro de los juegos según la disponibilidad de plataformas.
- La decisión entre el usuario y la crítica.
- El mejor juego hasta la fecha.

Esto puede ser clave para futuras decisiones de mercado y estrategias de desarrollo.
