DATA-SET ANALISIS JUEGOS DE STEAM

Leonel Moises Chauca Maydana

INF - 354

1. INTRODUCCION

En la actualidad, la industria de los videojuegos es una de las más dinámicas y rentables a nivel mundial. La plataforma **Steam**, que es uno de los principales distribuidores digitales de juegos, ha revolucionado la forma en que los usuarios adquieren y disfrutan de los videojuegos. Con millones de usuarios activos y una vasta biblioteca de juegos, Steam se ha convertido en un referente dentro del ecosistema digital, no solo por su alcance, sino también por la cantidad de datos que genera a partir de la interacción de los usuarios con los juegos.

Este artículo se centra en el análisis de un dataset obtenido de Steam, que contiene información relevante sobre una amplia variedad de juegos, incluyendo aspectos como las **reseñas** de los usuarios, precios, horas de juego, géneros y descuentos. Este conjunto de datos ofrece una rica fuente de información que puede ser utilizada obtener insights para valiosos sobre las tendencias y comportamientos del mercado de videojuegos, así como para explorar el uso de inteligencia artificial (IA) en la predicción de variables clave, como la popularidad de los juegos o la optimización de precios.

El objetivo principal de este artículo es realizar un análisis exhaustivo de este dataset, desde su exploración inicial hasta la implementación de modelos predictivos y sistemas de recomendación basados IA. en Además, se abordarán los desafíos que enfrenta este tipo de análisis, tales como el preprocesamiento de los datos y el manejo de valores nulos o columnas irrelevantes, con el fin de construir un conjunto de limpio y apto datos para la construcción de modelos robustos.

A través de este análisis, se destacará cómo el uso de datos y algoritmos avanzados puede transformar la forma en que los desarrolladores, distribuidores y jugadores interactúan Steam, con la plataforma optimizando la experiencia de compra, mejorando las recomendaciones personalizadas y, en última instancia, aumentando el éxito de los juegos en el mercado. Además, se explorarán oportunidades futuras para mejorar este análisis y las aplicaciones que pueden surgir a medida que los datos continúan evolucionando.

2. STEAM QUE ES?

2.1. HISTORIA

Orígenes y Creación (2003): Steam fue lanzado en septiembre de 2003 por Valve Corporation, fundada por Gabe Newell v Mike Harrington, quienes anteriormente habían trabajado en Microsoft. Valve ya había tenido un gran éxito con juegos como Half-Life, pero se dieron cuenta de que la distribución física de videojuegos estaba quedando obsoleta. El objetivo original de Steam era permitir la actualización automática de sus juegos, ofreciendo parches de manera sencilla sin la necesidad de que los jugadores tuvieran que buscar e instalar manualmente.

Expansión y Crecimiento (2004-2007): En 2004, Steam comenzó a ampliar su oferta, añadiendo juegos de otros desarrolladores y convirtiéndose en una plataforma de distribución más abierta. Valve Durante este período, consolidó su posición con el lanzamiento de Half-Life 2, que fue uno de los primeros grandes éxitos en Steam. La plataforma creció rápidamente, especialmente gracias a su modelo de compra digital y descarga instantánea, lo que la convirtió en una alternativa atractiva para los jugadores.

Innovaciones (2008-2012): Entre 2008 y 2012, Steam introdujo numerosas características innovadoras como Steam Cloud (para guardar partidas en línea), Steam Workshop (una plataforma para mods y contenido generado por los usuarios) y Big Picture Mode (para usar Steam en televisores grandes). También expandió su catálogo de juegos y software, lo que ayudó a atraer a más usuarios y desarrolladores.

Dominio del Mercado (2013-Presente): Hoy en día, Steam sigue siendo la plataforma dominante de distribución digital de juegos. Además de ofrecer títulos AAA, también alberga indie games, aplicaciones de software, v herramientas de desarrollo. En 2013, Valve lanzó SteamOS, un sistema operativo basado en Linux, y más tarde, la Steam Machine y el Steam Controller para mejorar la experiencia en consola. Además, el lanzamiento de la Steam Deck en 2021, una consola portátil basada en Linux, reafirmó el compromiso de Valve con el acceso a los juegos en cualquier lugar.

2.2. QUE ES STEAM

Steam es una plataforma digital desarrollada por Valve Corporation, que ofrece distribución de videojuegos, gestión de contenido digital, y

servicios sociales para los jugadores. Lanzada en 2003, Steam se ha convertido en uno de los servicios más grandes y populares para la compra, descarga gestión У de videojuegos en PC, Mac, y Linux, y más recientemente, en dispositivos portátiles consolas. Steam se distingue solo por su extensa biblioteca de juegos, sino también por su ecosistema que permite a los jugadores interactuar, compartir contenido y participar en la comunidad de manera más integrada que muchas otras plataformas.

2.3. COMPONENTES PRINCIPALES DE STEAM Tienda Digital:

Steam permite a los usuarios comprar videojuegos de una enorme variedad de géneros, desde grandes títulos AAA hasta juegos independientes Además (indie). videojuegos, la plataforma también ofrece aplicaciones software, películas contenidos adicionales como **DLCs** (contenidos descargables), que mejoran la experiencia de los juegos.

Biblioteca de Juegos: Todos los juegos comprados en Steam se almacenan en una biblioteca digital en línea. Los usuarios pueden descargar e instalar los juegos en cualquier momento, además de gestionar sus títulos, agruparlos por categorías, crear colecciones, y gestionar la configuración del sistema.

Comunidad Social: Steam ha creado un espacio para la interacción entre jugadores. Los usuarios pueden tener amigos en la plataforma, enviar y recibir mensajes, participar en chats de texto y voz, compartir logros y capturas de pantalla, y organizar grupos. Además, las funcionalidades como Steam Workshop permiten a los jugadores compartir y descargar mods (modificaciones del juego) que extienden o personalizan la experiencia de juego.

Actualizaciones Automáticas:

Una característica fundamental de Steam es la gestión automática de las actualizaciones de los juegos. Los usuarios no tienen que preocuparse por descargar parches o nuevas versiones manualmente, ya que Steam actualiza los juegos de forma automática en segundo plano.

Funcionalidades Avanzadas:

Steam ofrece opciones adicionales como el Steam Cloud (que guarda las partidas y configuraciones en la nube), Big Picture Mode (para disfrutar de los juegos en pantallas grandes o televisores), y soporte para realidad virtual con SteamVR.

2.4. ¿COMO FUNCIONA STEAM?

Steam es, en esencia, un mercado digital y un gestor de juegos. Los usuarios compran juegos y software a través de la tienda de Steam, que se descarga directamente a sus computadoras o dispositivos compatibles. Después de la compra, el contenido digital se añade a la biblioteca del usuario, y puede ser instalado en cualquier momento.

Además de las compras, Steam también fomenta la interacción social y la creación de contenido. Los jugadores pueden acceder a las funciones sociales de Steam, como los foros, los comentarios de usuarios, y las reseñas de productos, para obtener retroalimentación y recomendaciones de otros jugadores.

La plataforma también permite a los desarrolladores y creadores de contenido subir sus juegos y actualizaciones, gestionando todo el proceso de venta y distribución digital sin la necesidad de intermediarios o distribuidores.

2.5. PROCESO DE COMPRA EN STEAM

El proceso de compra en Steam es sencillo y completamente digital. A continuación, te explico paso a paso cómo se realiza una compra en la plataforma:

1. Creación de una cuenta en Steam:

Para poder comprar en Steam, el primer paso es crear una cuenta en la plataforma. Este proceso es gratuito y se realiza a través de la página web o la aplicación de Steam.

2. Navegar por la tienda:

Una vez que se ha iniciado sesión, el usuario puede explorar la tienda de Steam. Los juegos están organizados en categorías como "Tendencias", "Ofertas Especiales", "Juegos Populares" o "Géneros", entre otros. También existe un

buscador para encontrar juegos específicos.

- 3. Añadir juegos al carrito:
 Cuando se encuentra un juego
 que le interesa, el usuario
 puede hacer clic en el botón
 "Añadir al carrito". Aquí, puede
 elegir otros juegos o contenido
 adicional para agregar a su
 compra.
- 4. Verificar y realizar el pago: El siguiente paso es verificar el contenido del carrito y proceder al pago. Steam ofrece diversas formas de pago, como tarjetas de crédito, PayPal, tarjetas de Steam Wallet y, en algunas regiones, incluso pagos mediante transferencia bancaria.
- 5. Confirmación y descarga:
 Una vez realizada la compra,
 Steam confirmará la
 transacción y el juego será
 automáticamente agregado a
 la biblioteca del usuario.
 Desde allí, se podrá descargar
 e instalar directamente.
- 6. Acceso inmediato: Después de completar la descarga, el usuario puede comenzar a jugar el título adquirido de inmediato. Steam también maneja las actualizaciones y parches para asegurarse de que el juego esté siempre al día.

2.6. IMPACTO DE STEAM EN LA INDUSTRIA DEL VIDEOJUEGO

Steam transformado ha completamente la industria de los videojuegos. Antes de su aparición, los desarrolladores y editores dependían distribuidores físicos para llevar sus juegos al mercado. Steam permitió que desarrolladores independientes accedieran a una audiencia global sin los costos asociados con distribución física.

La plataforma también ha jugado un papel clave en la popularización de los juegos independientes (indie games), ayudando a que títulos de bajo presupuesto como Hollow Knight, Celeste y Undertale encuentren un gran éxito.

2.7. COMUNIDAD DE STEAM IMPACTO Y CARACTERISTICAS SOCIALES

Una de las mayores fortalezas de Steam no solo es su catálogo, sino también su enfoque en la **comunidad** y en la **interacción social**. Estas son algunas de las características que definen a Steam como una plataforma social para jugadores:

Grupos y Foros: Los usuarios de Steam pueden unirse a grupos relacionados con sus juegos favoritos, intereses comunes o creadores de contenido. Además. los **foros** permiten а los usuarios discutir temas. compartir consejos, hacer simplemente preguntas interactuar con otros jugadores.

Logros y Steam Cards: Los **logros** son una característica social que alienta a los jugadores a completar ciertas tareas dentro de los juegos. También están las **Steam** Trading Cards, que los jugadores pueden recolectar, intercambiar y usar para obtener insignias que mejoran la personalización de su perfil de usuario.

Marketplaces y Steam Points
Shop: Steam permite a los
jugadores comprar y
vender cosméticos, trading
cards, e incluso skins de
juegos en su mercado (Steam
Market). El Steam Points
Shop es otra característica que
permite a los usuarios ganar
puntos por sus compras y
canjearlos por decoraciones
de perfil, emoticonos y otros
elementos exclusivos.

Transmisiones en vivo y Steam Broadcasting:

Steam ofrece la posibilidad de hacer **transmisiones en vivo** de tus juegos, lo que facilita la creación de contenido y la interacción con la comunidad.

Steam Broadcasting permite que otros usuarios te sigan mientras juegas, proporcionando una experiencia social en tiempo real

3. EXPLORACION INICIAL DEL DATASET

El primer paso en cualquier análisis es la exploración inicial, que permite conocer el conjunto de datos con el que se trabajará, identificar patrones generales y entender la naturaleza de las variables presentes. En este caso, estamos analizando un dataset extraído de Steam, la plataforma distribución digital de de videojuegos más popular. Este conjunto de datos contiene información sobre una amplia variedad de juegos, desde las reseñas de los usuarios hasta detalles sobre precios, géneros, tiempo de juego y más.

4. DESCRIPCIÓN GENERAL DEL DATASET

El dataset contiene información relacionada con juegos disponibles en la plataforma Steam, y está compuesto por varias columnas que representan

diferentes características de cada juego. Al explorar este dataset, nos enfrentamos a un conjunto de registros (filas) que corresponden a juegos individuales, y cada columna contiene una variable que describe algún aspecto de cada juego.

Algunas de las **columnas clave** incluyen:

- name: El nombre del juego.
- developer: El desarrollador del juego.
- publisher: La entidad que publica el juego.
- genres: Los géneros o categorías que describen el tipo de juego (por ejemplo, acción, aventura, RPG, etc.).
- price: El precio del juego (en dólares, por ejemplo).
- score_rank: La clasificación de las reseñas de los usuarios.
- **owners:** La cantidad de propietarios del juego.
- average_playtime: El tiempo promedio de juego por usuario.
- review_score: El puntaje de reseña promedio del juego.
- release_date: Fecha de lanzamiento del juego.

La cantidad de registros en el dataset refleja la magnitud de Steam como plataforma, con miles de juegos listados. En términos de dimensiones del dataset, estas se componen tanto de un alto número de filas (juegos) como de columnas (atributos asociados a cada juego).

4.1. REVISION DE VALORES NULOS

crucial Un aspecto en la exploración inicial de un dataset es la identificación de valores nulos o faltantes. Los valores nulos pueden tener un impacto significativo en el análisis, ya que pueden sesgar los resultados o impedir la construcción modelos predictivos precisos. En este paso, se debe revisar cada columna del dataset para detectar la presencia de datos faltantes.

Por ejemplo:

Las columnas como review_score o price pueden contener valores nulos si el juego no tiene reseñas o si no está disponible a la venta.

La columna release_date podría tener valores nulos en juegos que no tienen una fecha definida de lanzamiento.

Es importante tener una estrategia para lidiar con estos valores, ya sea mediante la imputación de los mismos (rellenándolos con un valor promedio o utilizando otras técnicas) o eliminándolos si representan una proporción insignificante del total de los datos.

4.2. TIPOS DE VARIABLES Y DISTRIBUCION DE DATOS

Una vez que tengamos claro el contenido y los valores del dataset, fundamental hacer un análisis descriptivo de cada una de las variables para comprender su distribución y el tipo de datos que contienen. Este paso proporciona información sobre la naturaleza de cada atributo y nos permite identificar patrones clave.

4.2.1. VARIABLES CATEGORICAS

genres: Contiene categorías que representan los géneros de los juegos. En esta columna. podríamos encontrar una variedad de géneros, como "Acción", "Aventura". "Simulación". etc. Εl análisis de estas categorías permitirá identificar nos qué tipos de iuegos predominan en la plataforma y si ciertos géneros están más relacionados con juegos exitosos o bien valorados.

4.2.2. VARIABLES NUMERICAS

price: Esta variable contiene los precios de los juegos. Un análisis descriptivo de esta columna puede revelar si existe una relación entre el

precio y la popularidad de los juegos, o incluso si los descuentos afectan las compras.

owners: El número de propietarios de un juego nos da una idea de su popularidad y alcance. Este dato puede estar relacionado con la calificación y las reseñas de los usuarios.

4.2.3. VARIABLES DE TEXTO

name: El nombre de los juegos es una variable textual que no tiene un valor numérico directo. Sin embargo, puede ser útil para agrupaciones o análisis de similitudes.

developer y publisher: Al igual que los nombres de los juegos, estas columnas contienen nombres entidades. Aunque no son tipo numérico, de pueden usar para identificar patrones relacionados con los desarrolladores la ٧ influencia de ciertas compañías en el mercado.

4.2.4. IDENTIFICACION DE TENDENCIAS GENERALES Y OUTLIERS

Es fundamental detectar patrones y outliers (valores atípicos) en los datos, especialmente en columnas como price,

owners y average_playtime. Los outliers pueden indicar anomalías o valores extremos que podrían ser importantes (por ejemplo, juegos con precios excesivamente altos o bajos) o errores en los datos.

Un análisis visual de estas distribuciones mediante histogramas o boxplots ayudará a entender mejor la variabilidad de los datos y si es necesario realizar alguna transformación antes de continuar con el análisis.

5. ANALISIS DETALLADO DEL DATASET

En esta sección. nos adentramos en los detalles del dataset proporcionado por SteamSpy, el cual contiene información sobre los juegos disponibles en la plataforma de Steam. Este análisis busca entender cada columna, la naturaleza de los datos que contiene y cómo se puede interpretar la información sobre los videojuegos en la plataforma más popular de distribución digital de juegos para PC.

5.1. COLUMNAS

1. appid: Identificador único del juego

Descripción: La columna appid contiene un número único que sirve como identificador para cada juego en la plataforma de Steam.

Propósito y Relevancia:

El appid es clave para hacer un seguimiento de los juegos en Steam, ya que cada juego tiene asignado un identificador único que lo distingue de otros títulos en la base de datos de Steam. Aunque por sí mismo no proporciona información sobre el juego, es fundamental para realizar consultas o acceder a más detalles del juego en plataformas como SteamDB o la API de Steam.



2. name: Nombre del juego

Descripción: La columna name contiene el nombre del juego.

Propósito y Relevancia: Este campo es crucial, ya que es la forma más común identificar un juego. El nombre del juego está directamente relacionado con reconocimiento de marca y su popularidad en la plataforma. Un análisis de esta columna puede ayudar a identificar qué títulos dominan el mercado y cuáles tienen menos visibilidad.



3. developer: Desarrollador del juego

Descripción: El nombre del estudio o individuo que desarrolló el juego.

Propósito y Relevancia: Esta columna es útil para identificar las compañías o desarrolladores que dominan el mercado de Steam. Un juego desarrollado por una compañía de renombre (por ejemplo, Valve o Rockstar)

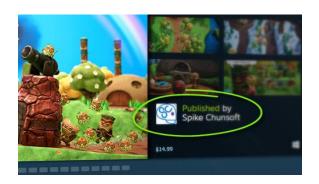
probablemente tendrá más visibilidad V ventas que aquellos desarrollados por estudios más pequeños o independientes. Analizar esta columna puede ayudar identificar tendencias de desarrollo y a evaluar el impacto de los desarrolladores en el éxito del juego.



4. publisher: Editor del juego

Descripción: El nombre de la compañía o entidad que publica el juego, que a menudo difiere del desarrollador.

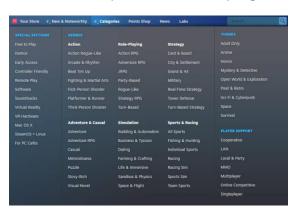
Propósito y Relevancia: Los editores juegan un papel fundamental en la comercialización У distribución de los juegos. Mientras algunos que desarrolladores también publican sus propios juegos, muchos prefieren asociarse con editores que tengan la experiencia en la distribución y promoción global. columna es relevante para estudiar cómo los editores influyen en el éxito de un título.



7. score_rank: Rango de calificación del juego

Descripción: Una clasificación numérica que refleja la calificación global del juego en la plataforma.

Propósito: Ayuda a evaluar la recepción general de los juegos. Un rango bajo puede indicar críticas negativas, mientras que un rango alto sugiere que los jugadores valoran positivamente el juego.



positive: Número de reseñas positivas

Descripción: La cantidad de reseñas positivas recibidas por el juego de parte de los usuarios de Steam.

Propósito: Este dato es crucial para evaluar la popularidad y la recepción positiva de un juego. Un alto número de reseñas positivas generalmente está correlacionado con un éxito en la plataforma.



negative: Número de reseñas negativas

Descripción: La cantidad de reseñas negativas recibidas por el juego.

Propósito: Junto con la columna de reseñas positivas, esta columna permite medir la polarización de la opinión de los jugadores y obtener un panorama completo sobre la recepción del juego.



11. userscore: Puntuación promedio del juego según los usuarios

Descripción: La puntuación promedio otorgada al juego por los usuarios, en una escala de 0 a 100.

Propósito: Proporciona una evaluación rápida del juego en función de las opiniones de los jugadores. Es una forma de medir el impacto de un juego desde la perspectiva del jugador.



owners: Rango estimado de jugadores que poseen el juego

Descripción: Estimación del número de jugadores que han comprado o poseen el juego en Steam, expresado en un rango (por ejemplo, "1M - 5M" significa que entre 1 y 5 millones de jugadores poseen el juego).

Propósito: Este dato es fundamental para medir el éxito de ventas de un juego.

También puede ayudar a identificar juegos que, aunque no tienen un número de reseñas tan alto, aún alcanzan una gran audiencia.



average_forever: Promedio total de tiempo jugado en todas las plataformas

- Descripción: El promedio de tiempo jugado por los jugadores de todos los juegos. Este dato está en minutos.
- Propósito: Proporciona una medida de la longevidad de un juego en términos de tiempo de juego. Juegos con altos promedios suelen tener una base de jugadores más comprometida.



average_2weeks: Promedio de tiempo jugado en las últimas dos semanas

- Descripción: Promedio de tiempo jugado en las últimas dos semanas.
 Similar a la columna anterior, pero más centrada en la actividad reciente.
- Propósito: Permite analizar la actividad reciente de los jugadores y detectar juegos que siguen siendo populares con el paso del tiempo o que pueden haber perdido interés.



median_forever: Mediana del tiempo jugado de todos los jugadores

Descripción: Mediana del tiempo jugado en todas las plataformas, en minutos.

Propósito: A diferencia del promedio, la mediana puede ser más útil para entender la "experiencia típica" de los jugadores, ya que no se ve afectada por valores extremos (como los jugadores que juegan miles de horas).



median_2weeks: Mediana del tiempo jugado en las últimas dos semanas

Descripción: Mediana del tiempo jugado en las últimas dos semanas, en minutos.

Propósito: Ayuda a obtener una imagen precisa del tiempo que los jugadores están invirtiendo en el juego recientemente.



price: Precio actual del juego

Descripción: El precio del juego en la tienda de Steam en el momento del análisis.

Propósito: Es un indicador de la política de precios del desarrollador y editor. Los juegos con precios bajos pueden atraer a más jugadores, mientras que los más caros pueden indicar una estrategia de juegos premium.



initialprice: **Precio inicial del juego**

Descripción: El precio original del juego cuando fue lanzado.

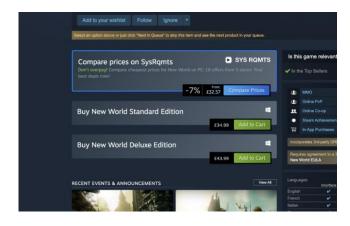
Propósito: Permite comparar el precio actual con el precio original y observar cómo los descuentos o rebajas de Steam impactan en la estrategia de precios de los desarrolladores.



discount: **Porcentaje de descuento aplicado**

Descripción: El porcentaje de descuento que se aplica al juego en comparación con el precio inicial.

Propósito: Esta columna es esencial para analizar las estrategias de precios y promociones, especialmente durante eventos como las rebajas de Steam.



languages:Idiomas disponibles en el juego

Descripción: Lista de idiomas en los que el juego está disponible.

Propósito: Indica la accesibilidad del juego a una audiencia global. Los juegos que ofrecen soporte multilingüe pueden atraer a un público más amplio.



genre: Género del juego

Descripción: El género o tipos de juegos a los que pertenece el título (por ejemplo, acción, aventura, rol, etc.).

Propósito: Ayuda a clasificar los juegos en categorías según su jugabilidad y temática. Los géneros populares pueden ofrecer indicaciones sobre las preferencias de los jugadores en la plataforma.



ccu: Usuarios concurrentes en el juego

Descripción: La cantidad de jugadores que están jugando el juego simultáneamente en un momento dado.

Propósito: Este dato es útil para analizar la actividad en tiempo real de los juegos. Los juegos con una alta cantidad de jugadores concurrentes suelen ser populares y tienen una base de jugadores activa.



tags: Etiquetas asociadas al juego

Descripción: Una lista de etiquetas que describen el juego, tales como géneros, modos de juego o características especiales (por ejemplo, multijugador, VR, acción, etc.).

Propósito: Estas etiquetas permiten categorizar y filtrar los juegos en la plataforma de Steam. Un análisis de las etiquetas puede ayudar a identificar tendencias y características comunes en los juegos más populares.



6. CONCLUSION

A lo largo de este artículo, hemos recorrido en detalle la plataforma Steam, un pilar esencial en la industria de los videojuegos desde su creación en 2003. Steam no solo ha transformado la forma en que los jugadores adquieren y disfrutan de los videojuegos, sino que ha redefinido las dinámicas sociales y económicas de la industria, convirtiéndose en un referente global para distribuidores y desarrolladores.

Steam, a través de sus innovadoras funcionalidades como el Steam Workshop, el Steam Cloud, las rebajas masivas, y su enfoque en el apoyo a los desarrolladores independientes,

ha creado un ecosistema único que conecta a millones jugadores y creadores en todo el mundo. La plataforma no solo facilita la compra de juegos, sino que también fomenta la creación y personalización de contenido, el intercambio social construcción de una comunidad que contribuye activa al crecimiento y la evolución del propio servicio.

En el dataset de SteamSpy que se exploró, encontramos una valiosa fuente de información sobre los juegos disponibles en Steam y su rendimiento en el mercado. Las columnas del dataset proporcionan análisis un detallado de cada juego, desde el número de ventas y el precio hasta las categorías y plataformas a las que pertenecen. Este dataset es crucial para los analistas de datos y desarrolladores interesados en entender mejor las dinámicas del mercado de Steam, identificar tendencias en las preferencias de jugadores descubrir oportunidades de crecimiento en el sector.

Al analizar este dataset, podemos identificar patrones clave, como la correlación entre el precio, la popularidad y las reseñas de los usuarios, lo que podría ser útil para predecir el éxito de un nuevo lanzamiento o para ajustar las estrategias de marketing de un juego. Además, este análisis

también ofrece una visión sobre las tendencias de compra y el comportamiento de los consumidores, permitiendo a los desarrolladores y editores ajustar sus estrategias de ventas y lanzamiento.