

Proyecto Visualización de Datos

Juan Carreño, Kevin Calderón & Felipe Rubiano
Mayo 2024.

Pontificia Universidad Javeriana
Departamento de ingeniería de sistemas
Visualización de Datos

Introducción:

El presente proyecto se enmarca en el dinámico sector del entretenimiento digital, específicamente en el mercado de videojuegos, una industria multimillonaria en constante evolución. La empresa para la cual se desarrolla este trabajo busca fortalecer su toma de decisiones estratégicas mediante la analítica y visualización de datos relacionados con videojuegos. Para ello, hemos recopilado datos de tres conjuntos obtenidos de Kaggle, un reconocido repositorio en el campo del análisis de datos. Estos datasets nos proporcionan una visión integral del comportamiento de los videojuegos en el mercado, cubriendo aspectos críticos como popularidad, preferencias de plataforma, rendimiento de ventas y evaluaciones de usuarios.

El análisis de estos datos permitirá identificar patrones y tendencias significativas, proporcionando insights valiosos para la toma de decisiones en áreas clave como inversiones, desarrollo y marketing. La preparación y limpieza de los datos fue esencial para asegurar la calidad y utilidad de la información analizada. Herramientas como Excel y scripts de Python fueron empleadas para la transformación y modelado de los datos, garantizando que las visualizaciones resultantes sean precisas y accionables.

Adicionalmente, se diseñaron modelos analíticos, incluyendo el modelo Entidad-Relación (E/R) y el modelo dimensional, que nos permitieron estructurar los datos de manera que optimicen el rendimiento en consultas analíticas y reportes. Estos modelos los consideramos esenciales para estructurar los datos para apoyar el análisis y la visualización, facilitando la identificación de tendencias y el descubrimiento de oportunidades de mejora.

En resumen, este proyecto busca entender profundamente el mercado de videojuegos a través de datos, proporcionando una base sólida para tomar decisiones estratégicas que impulsen el crecimiento y éxito de la empresa en este competitivo sector.

Objetivos de Analítica:

1. **Impacto de la Crítica y Usuarios en Ventas:** Hallar cómo las calificaciones de críticos y usuarios afectan las ventas globales y regionales, para entender el peso de las reseñas en el éxito comercial de los videojuegos.
2. **Análisis de Lanzamientos por Mes:** Determinar las tendencias de lanzamientos de las empresas publicadoras, y como estas pueden llegar a afectar las ventas de los juegos.
3. **Evaluación del Desarrollador:** Analizar la relación entre el desarrollador y el éxito del juego, tanto en ventas como en calificaciones, para identificar desarrolladores con los que asociarse o invertir y en que géneros especializarse.
4. **Tendencias de Género y Jugadores:** Explorar la popularidad de los diferentes géneros de juegos y la cantidad de jugadores que tienen, para informar el desarrollo de juegos que cumplen las preferencias actuales de los usuarios.
5. **Perfil de Juegos Populares:** Identificar las características comunes de los juegos más vendidos y mejor calificados (como género, plataforma y desarrollador), para dirigir futuras inversiones y estrategias de marketing.

Objetivos de Visualización:

1. **Dashboard de Métricas Clave:** Construir un Dashboard dinámico en el que consolidemos métricas clave como ventas totales, calificaciones promedio y cantidad de jugadores por juego, permitiendo filtrar por diversos criterios para proceder con un análisis más profundo.
2. **Juegos Publicados por Mes y Ventas por zona:** Crear una visualización combinada que muestre la cantidad de juegos publicados por mes, que permita analizar tendencias de publicación, así como una comparativa de ventas globas y por zona del mundo.
3. **Ranking de Desarrolladores:** Generar una clasificación visual de desarrolladores basada en ventas y puntuaciones, para reconocer rápidamente a los líderes de la industria.
4. **Diagrama de Dispersión de Calificaciones vs. Ventas:** Diseñar un diagrama de dispersión que compare las ventas y calificaciones (tanto de críticos como de usuarios) para cada juego, ilustrando la relación entre la percepción de calidad y el éxito comercial.
5. **Treemap de Popularidad por Género y Plataforma:** Crear un treemap que nos muestre la popularidad de los juegos según su clasificación (ranking) por género y plataforma, ofreciendo una visualización amplia e inmediata de las áreas de alto rendimiento.

Objetivos de negocio:

1. Identificación de Oportunidades de Inversión:

- **Objetivo:** Determinar las características comunes de los juegos más vendidos y mejor calificados para dirigir futuras inversiones y estrategias de marketing.
- **Identificador:** Global_Sales y Score (Metascore)

2. Optimización de la Distribución en Plataformas:

- **Objetivo:** Identificar qué plataformas tienen los juegos mejor recibidos y más vendidos para centrar los esfuerzos de desarrollo y marketing en estas plataformas.
- **Identificador:** Platform y Global_Sales

3. Evaluación de Desarrolladores:

- **Objetivo:** Analizar la relación entre el desarrollador y el éxito del juego para identificar desarrolladores clave con los que asociarse o invertir y en qué géneros especializarse.
- **Identificador:** Developer y Global_Sales

4. Análisis de Tendencias de Género y Preferencias de Jugadores:

- **Objetivo:** Explorar la popularidad de los diferentes géneros de juegos y la cantidad de jugadores para informar el desarrollo de juegos que cumplan las preferencias actuales del mercado.
- **Identificador:** Genre y Players

5. Impacto de Reseñas en Ventas:

- **Objetivo:** Determinar cómo las calificaciones de críticos y usuarios afectan las ventas para entender el peso de las reseñas en el éxito comercial de los videojuegos.
- **Identificador:** Score (Metascore) y User Score

Contexto:

El presente proyecto se centra en el dinámico sector del entretenimiento digital, específicamente en el mercado de videojuegos, que constituye una industria multimillonaria en constante evolución. Nuestro enfoque se dirige a la analítica y visualización de datos relacionados con videojuegos para una empresa que opera en este sector y busca fortalecer su toma de decisiones estratégicas.

El análisis lo fundamentamos en tres conjuntos de datos obtenidos de Kaggle, un repositorio de alto reconocimiento en el campo del análisis de datos. Estos Dataset proporcionan una visión integral del comportamiento de los videojuegos en el mercado, cubriendo aspectos críticos como popularidad, preferencias de plataforma, rendimiento de ventas y evaluaciones de usuarios.

Los conjuntos de datos se describen de la siguiente manera:

Dataset “Videogames”:

Nombre: “Juegos de Video”

Origen: Repositorio Kaggle:

<https://www.kaggle.com/datasets/sagayaabinesh/videogames>

Propósito de recopilación: Provee un ranking basado en las ventas globales de videojuegos. Es crucial para comprender qué títulos dominan el mercado y cómo se distribuyen las ventas a nivel internacional.

Métodos de Recopilación: No especificado

Características:

Tamaño: Se presentan 11.493 valores únicos sin una frecuencia de actualización especificada. Aunque contiene información vital, como las ventas en diferentes regiones, la falta de especificación de procedencia limita su calidad, limitándola a media.

Relevancia: Este Dataset es una parte importante de nuestros datos, permitiéndonos abordar suposiciones y escenarios relacionados con perspectivas de mercado y ventas. Además, nos proporciona información adicional que enriquece nuestro análisis y nos ayuda a comprender mejor los diferentes aspectos del negocio.

Dataset “Metacritics Best Video Games of All Time 2022”:

Nombre: “Metacritics Los mejores videojuegos de todos los tiempos 2022”

Origen: Repositorio Kaggle con datos sacados de “Metacritic”. Sitio web que recopila información de consumo.

<https://www.kaggle.com/datasets/caiquerezende/metacritics-best-video-games-of-all-time-2021>

Propósito de recopilación: Generar un ranking actualizado de videojuegos destacados basado en calificaciones profesionales y opiniones de usuarios, para identificar estándares de excelencia y preferencias del mercado.

Métodos de Recopilación: Extracción del sitio web “Metacritics”, mediante el uso de un script de Python:

<https://github.com/caiques121/Portfolio/tree/main/metacritic>

Características:

Tamaño: Es una colección de 19.047 valores únicos, con frecuencia de actualización anual, catalogamos su calidad como de carácter alto, siendo uno de los más confiables, pero llegando a ser infalible sobre todo en lo relacionado a los gustos personales

Relevancia: Este conjunto de datos es crítico para entender las tendencias de éxito y las expectativas de calidad que influyen en la percepción de los consumidores y las estrategias de desarrollo de juegos.

Dataset “Metacritic video-games data”:

Nombre: “Datos meta críticos de videojuegos”

Origen: Repositorio Kaggle con datos extraídos de “Metacritics”.

<https://www.kaggle.com/datasets/brunovr/metacritic-videogames-data>

Propósito de recopilación: Crear un sistema de recomendación basado en contenido, utilizando calificaciones y metadata para personalizar la oferta de juegos a los usuarios.

Métodos de Recopilación:

“Crear un diccionario 'pages' que contendrá los objetos DataFrame de todas las páginas. Cada entrada es un DataFrame de pandas con datos de los juegos en cada página del sitio. Actualmente, hay 180 páginas de juegos calificados.

Para cada página, crear un diccionario 'data_page' de listas vacías para llenar con los datos de cada juego. Como cada página muestra 100 juegos, cada una de estas listas debería contener 100 elementos (excepto la última página).

Utilizar requests para acceder a la url de cada página y BeautifulSoup para analizar el archivo html.

Recorrer todos los juegos en cada página y extraer los datos relevantes.”

Tomado de: <https://github.com/BrunoBVR/projectGames/blob/main/scraper.ipynb>

Características:

Tamaño: Contiene 11.762 valores únicos, su recopilación se dio el noviembre de 2022, no se contempla ninguna actualización, y consideramos que es alta validez, ya que contiene rígidos sistemas de filtrado.

Relevancia: Esencial para la personalización y la mejora de plataformas de distribución de juegos, impactando directamente en la experiencia del usuario y la retención de clientes.

Recolección y comprensión de los datos:

En el proyecto hemos recopilado datos de tres fuentes principales en Kaggle, enfocándose en la integridad y relevancia para el análisis de los videojuegos. Se realizaron pruebas y validaciones para determinar la consistencia de los datos, resultando en dos datasets finales. El primer set de datos se transformó y modeló usando Excel y la función “buscarv” para enlazar nombres y desarrolladores. El segundo set se procesó mediante scripts de Python, seleccionando registros basados en nombres de juegos y eliminando los no coincidentes. Este procedimiento meticuloso asegura que los datos están listos para análisis y visualización avanzados, con la calidad y la estructura necesaria para respaldar las futuras decisiones de negocio.

La meticulosa compilación de los datos nos garantiza una base sólida para el análisis, permitiendo identificar patrones y tendencias significativas en el comportamiento de los usuarios y la popularidad de los videojuegos. La integridad y calidad de los datos compilados son esenciales para nosotros por la credibilidad y utilidad de los insights que generaremos, resultando en visualizaciones que eventualmente apoyen estratégicamente las decisiones empresariales.

Diccionario de Datos:

El Diccionario de Datos es una herramienta esencial para comprender los atributos que forman nuestra base de datos, proporcionando una guía detallada sobre cada columna y su relevancia en el análisis y visualización de la información. Con este recurso detallamos la naturaleza, el tipo de datos, las posibles restricciones y la importancia de cada variable, facilitando así su correcta interpretación y utilización en el desarrollo de nuestro proyecto. A continuación, enumeramos y describimos las veinte características principales de nuestro Dataset, las cuales han sido seleccionadas y compiladas con el objetivo de ofrecer una panorámica integral del comportamiento de los videojuegos dentro del mercado global.

1. **Rank:** Clasificación global del juego basada en ventas; tipo entero (int).
2. **Name:** Título del juego; cadena de texto (string), llave primaria para referencias cruzadas.
3. **Platform:** Sistema o consola en la que está disponible el juego; cadena de texto (string).
4. **Year:** Año de publicación del juego; tipo entero (int), relevante para tendencias históricas.
5. **Genre:** Género lúdico del juego; cadena de texto (string), útil para segmentación del mercado.
6. **Publisher:** Empresa que publica el juego; cadena de texto (string), para análisis de rendimiento por editor.
7. **NA_Sales:** Ventas en Norteamérica; número flotante (float), refleja éxito en este mercado.
8. **EU_Sales:** Ventas en Europa; número flotante (float), indica penetración en mercado europeo.
9. **JP_Sales:** Ventas en Japón; número flotante (float), importante para comprender el mercado asiático.
10. **Other_Sales:** Ventas en otras regiones; número flotante (float), para análisis de mercados emergentes.
11. **Global_Sales:** Total de ventas a nivel mundial; número flotante (float), indicador clave de éxito global.
12. **r-date (Release Date):** Fecha de lanzamiento del juego; cadena de texto (string) o fecha y hora (datetime), esencial para análisis temporal.
13. **Score (Metascore):** Calificación promedio del juego por críticos; número flotante (float) o entero (int), para evaluaciones de calidad.
14. **User Score:** Puntuación promedio por usuarios; número flotante (float) o entero (int), muestra la recepción del juego por el público.
15. **Developer:** Estudio o individuo que desarrolla el juego; cadena de texto (string), para asociar juegos con creadores.
16. **Genre2:** Segundo género del juego, si aplica; cadena de texto (string), para juegos con múltiples categorías.
17. **Players:** Número máximo de jugadores soportados; cadena de texto (string) o entero (int), para dinámicas de juego.
18. **Critics:** Número de críticas recibidas; número entero (int), útil para análisis de impacto mediático.
19. **Users:** Cantidad de usuarios que han jugado; número entero (int), para medir popularidad y engagement.

20. **Ranking:** Posición en listados específicos; número entero (int), para comparativas y benchmarking.

Preparación de Datos:

La preparación de los datos es una fase crítica en cualquier proyecto de análisis y visualización, ya que garantiza la calidad y la utilidad de la información utilizada para tomar decisiones. En este proyecto, hemos empleado un enfoque profundo y cuidadoso para asegurar que nuestros conjuntos de datos estén limpios, completos y listos para el análisis. A continuación, detallamos el proceso realizado:

1. Recolección y Selección de Datos:

Selección de Fuentes: Los datos fueron recolectados de repositorios reconocidos como Kaggle, asegurándonos de la relevancia y la confiabilidad de la información.

Filtrado de Variables: Se realizó una evaluación profunda de las variables disponibles, prescindiendo de aquellas repetidas o que no contribuyen directamente a nuestros objetivos analíticos o que no se alinean con nuestra visión del proyecto.

2. Limpieza de Datos:

- **Verificación de Integridad:** Se revisaron los datos para detectar valores faltantes o inconsistentes. Los registros incompletos se han imputado o eliminado según la naturaleza de los datos.
- **Estandarización de Formatos:** Se aseguró que todos los campos como fechas, nombres y categorías sigan un formato uniforme para facilitar su análisis.

3. Transformación y Modelado:

- **Excel - Búsqueda V:** Se utilizó la función “BUSCARV” para enlazar los datos de diferentes tablas basándose en el nombre del juego como llave primaria y el desarrollador como llave secundaria.
- **Script de Python:** Se desarrolló un script para filtrar y organizar los datos según el nombre del juego, eliminando registros duplicados o irrelevantes para el análisis propuesto.

4. Validación de Datos:

- Se realizaron pruebas de consistencia para verificar que las transformaciones y las imputaciones que hemos realizado no alteraran significativamente las tendencias o conclusiones generales de nuestro conjunto de datos.

Indicadores:

Los indicadores seleccionados para este proyecto son fundamentales para medir el éxito y evaluar el rendimiento en relación con los objetivos de negocio planteados. Estos indicadores nos proporcionan una visión clara y cuantificable del progreso hacia nuestras metas y nos ayudan a identificar áreas de oportunidad para mejorar. A continuación, se describen los principales indicadores utilizados:

Ventas Globales (Global_Sales):

Descripción: Total de ventas de cada videojuego a nivel mundial, medido en millones de unidades.

Relevancia: Este indicador es clave para evaluar el éxito comercial de los videojuegos y permite identificar los títulos más populares y rentables.

Fuente de Datos: Dataset “Videogames” de Kaggle.

Ventas por Región (NA_Sales, EU_Sales, JP_Sales, Other_Sales):

Descripción: Ventas de cada videojuego desglosadas por regiones (Norteamérica, Europa, Japón y otras regiones), medidas en millones de unidades.

Relevancia: Estos indicadores permiten analizar el desempeño regional de los videojuegos y detectar mercados fuertes o emergentes.

Fuente de Datos: Dataset "Videogames" de Kaggle.

Calificación de Críticos (Score):

Descripción: Calificación promedio otorgada por críticos, medida en una escala de 0 a 100.

Relevancia: Este indicador ayuda a entender la percepción de calidad de los videojuegos por parte de expertos, lo cual puede influir en las decisiones de compra de los consumidores.

Fuente de Datos: Dataset "Metacritics Best Video Games of All Time 2022" de Kaggle.

Calificación de Usuarios (User Score):

Descripción: Calificación promedio otorgada por los usuarios, medida en una escala de 0 a 10.

Relevancia: Este indicador refleja la satisfacción del usuario y es crucial para evaluar la aceptación de los videojuegos en el mercado.

Fuente de Datos: Dataset "Metacritic video-games data" de Kaggle.

Número de Críticas Recibidas (Critics):

Descripción: Número total de críticas recibidas por cada videojuego.

Relevancia: Este indicador permite medir el impacto mediático y la visibilidad de los videojuegos.

Fuente de Datos: Dataset "Metacritic video-games data" de Kaggle.

Número de Usuarios que Han Jugado (Users):

Descripción: Cantidad total de usuarios que han jugado cada videojuego.

Relevancia: Este indicador es importante para medir la popularidad y el engagement de los videojuegos.

Fuente de Datos: Dataset "Metacritic video-games data" de Kaggle.

Plataforma de Juego (Platform):

Descripción: Sistema o consola en la que está disponible cada videojuego.

Relevancia: Este indicador es esencial para analizar la distribución y accesibilidad de los videojuegos en diferentes plataformas.

Fuente de Datos: Dataset "Videogames" de Kaggle.

Género del Videojuego (Genre):

Descripción: Categoría lúdica del videojuego, como acción, aventura, deportes, etc.

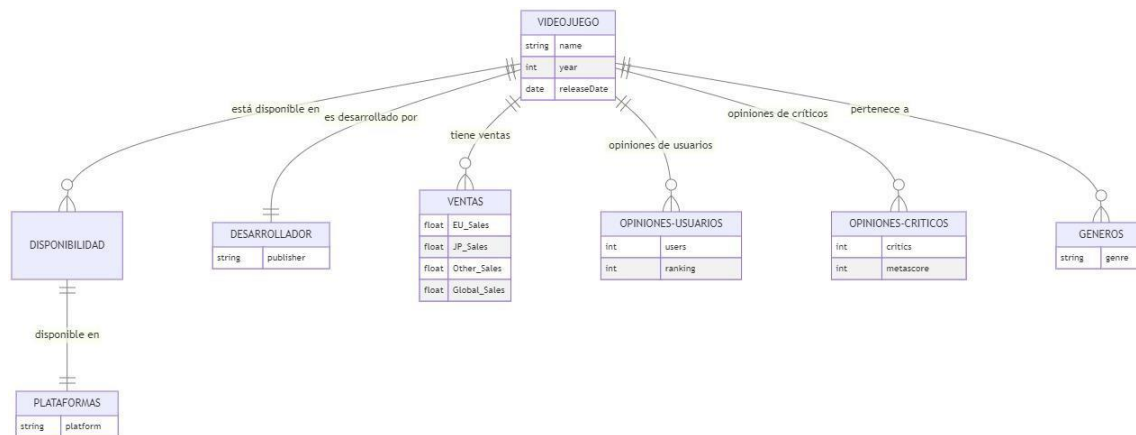
Relevancia: Este indicador facilita la segmentación del mercado y permite identificar las preferencias de los usuarios.

Fuente de Datos: Dataset "Videogames" de Kaggle.

Modelo Analítico:

Para la construcción del modelo analítico para este proyecto nos basamos en dos enfoques fundamentales: el modelo Entidad-Relación (E/R) y el modelo dimensional. Estos modelos son esenciales para estructurar los datos de manera que apoyen eficazmente el análisis y la visualización.

Modelo Entidad/Relación:



El Modelo Entidad-Relación (E/R) es fundamental en el diseño de bases de datos relacionales y se centra en representar los datos en términos de entidades y sus interrelaciones. Este modelo facilita el entendimiento estructural del dominio de datos, permitiendo una clara organización y una base sólida para operaciones de mantenimiento y consulta de datos.

- **Videojuego:** Centraliza la información crítica de cada título, incluyendo su identificador único (ID), nombre, y año de lanzamiento. Es el eje para el análisis de tendencias y la popularidad en el mercado.
- **Disponibilidad:** Indica en qué plataformas está disponible cada juego, proporcionando datos fundamentales para el análisis de distribución y accesibilidad del mercado.
- **Desarrollador:** Refleja quién ha creado el juego. Esta entidad es vital para establecer correlaciones entre desarrolladores específicos y el éxito de sus títulos.
- **Ventas:** Agrupa las métricas de ventas, tanto a nivel regional como global, permitiendo análisis económicos y de mercado.
- **Opiniones Usuarios:** Captura las valoraciones y reseñas de los usuarios, lo cual es crucial para entender la aceptación del mercado y la satisfacción del consumidor.
- **Opiniones Críticos:** Compila las valoraciones de expertos, que son fundamentales para comprender el impacto de las críticas en las ventas y la popularidad de los juegos.
- **Géneros:** Clasifica los juegos en uno o más géneros, proporcionando una dimensión esencial para analizar las preferencias del usuario y la diversidad del catálogo de juegos.

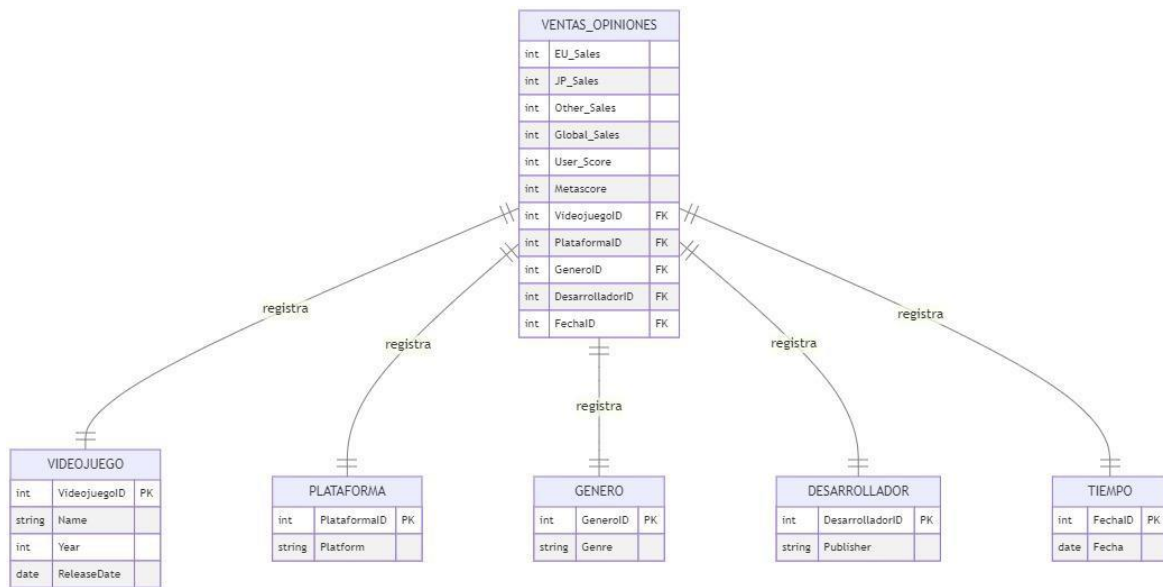
Al diseñar este modelo, se ha tenido en cuenta las relaciones naturales del mercado de videojuegos para reflejar la realidad del sector:

- **Relación de Disponibilidad:** Un videojuego puede estar disponible en múltiples plataformas. Esta relación inherente del mercado es fundamental para analizar estrategias de distribución y la presencia del juego en diferentes ecosistemas.
- **Relaciones de Opiniones:** Un título puede generar múltiples opiniones de usuarios y críticos. Capturar esta diversidad de feedback es esencial para entender la recepción del juego y su impacto en las ventas.
- **Relación con Desarrolladores:** Un desarrollador puede trabajar en varios juegos y distintos géneros, lo que influye en la percepción de su marca y el seguimiento de los fans.

- **Relaciones de Ventas:** Las ventas se rastrean por región y globalmente, ya que un juego puede tener un rendimiento dispar en diferentes mercados, lo cual es clave para la toma de decisiones de inversión y marketing.

Modelo Dimensional:

El Modelo Dimensional es un enfoque diseñado para optimizar el rendimiento en consultas analíticas y reportes en bases de datos de data warehouse. Este modelo estructura los datos en tablas de hechos y dimensiones, lo que facilita el análisis rápido y eficiente a través de diferentes perspectivas y dimensiones del negocio.



- **Tabla de Hechos:** Centraliza las métricas clave como las ventas en diferentes regiones, las puntuaciones globales y las fechas, funcionando como el punto de unión para las diversas dimensiones que informan el análisis.
- **Dimensiones:**
 - **Videojuego:** Contiene atributos descriptivos de los juegos que permiten un análisis detallado de cada título.
 - **Plataforma:** Ofrece un análisis por plataforma, permitiendo evaluar el rendimiento y la popularidad por cada sistema.
 - **Género:** Facilita la segmentación del análisis por categorías de juegos, como acción o aventura.
 - **Desarrollador:** Permite ver el rendimiento y las tendencias asociadas a cada desarrollador.
 - **Tiempo:** Permite el análisis temporal y la identificación de tendencias estacionales o de lanzamientos.

Indicadores:

1. Perfil de Juegos Populares

Indicadores:

- **Género más vendido:** Cantidad de juegos vendidos por género.
- **Plataforma más exitosa:** Ventas promedio por plataforma.
- **Desarrollador con mejores calificaciones:** Promedio de calificaciones por desarrollador.

2. Análisis de Desempeño por Plataforma

Indicadores:

- **Ventas por plataforma:** Número total de ventas en cada plataforma.
- **Calificación promedio por plataforma:** Promedio de calificaciones de juegos en cada plataforma.
- **Tasa de crecimiento de ventas por plataforma:** Incremento porcentual de ventas por plataforma en un periodo de tiempo determinado.

3. Evaluación del Desarrollador

Indicadores:

- **Ventas promedio por desarrollador:** Número promedio de ventas de juegos por desarrollador.
- **Calificación promedio de juegos por desarrollador:** Promedio de calificaciones de juegos desarrollados por cada desarrollador.
- **Ratio de juegos exitosos por desarrollador:** Porcentaje de juegos de un desarrollador que alcanzan un cierto umbral de ventas o calificaciones.

4. Tendencias de Género y Jugadores

Indicadores:

- **Popularidad de géneros:** Número de juegos vendidos por género.
- **Cantidad de jugadores por género:** Número de jugadores activos por género.
- **Tendencias de crecimiento por género:** Incremento porcentual de ventas o jugadores en cada género.

5. Impacto de la Crítica y Usuarios en Ventas

Indicadores:

- **Relación calificación-ventas:** Correlación entre la calificación de usuarios/críticos y las ventas.

Incremento de ventas post-reviewIndicadores:

1. Perfil de Juegos Populares

Indicadores:

- **Género más vendido:** Cantidad de juegos vendidos por género.
- **Plataforma más exitosa:** Ventas promedio por plataforma.
- **Desarrollador con mejores calificaciones:** Promedio de calificaciones por desarrollador.

2. Análisis de Desempeño por Plataforma

Indicadores:

- **Ventas por plataforma:** Número total de ventas en cada plataforma.
- **Calificación promedio por plataforma:** Promedio de calificaciones de juegos en cada plataforma.
- **Tasa de crecimiento de ventas por plataforma:** Incremento porcentual de ventas por plataforma en un periodo de tiempo determinado.

3. Evaluación del Desarrollador

Indicadores:

- **Ventas promedio por desarrollador:** Número promedio de ventas de juegos por desarrollador.
- **Calificación promedio de juegos por desarrollador:** Promedio de calificaciones de juegos desarrollados por cada desarrollador.
- **Ratio de juegos exitosos por desarrollador:** Porcentaje de juegos de un desarrollador que alcanzan un cierto umbral de ventas o calificaciones.

4. Tendencias de Género y Jugadores

Indicadores:

- **Popularidad de géneros:** Número de juegos vendidos por género.
- **Cantidad de jugadores por género:** Número de jugadores activos por género.
- **Tendencias de crecimiento por género:** Incremento porcentual de ventas o jugadores en cada género.

5. Impacto de la Crítica y Usuarios en Ventas

Indicadores:

- **Relación calificación-ventas:** Correlación entre la calificación de usuarios/críticos y las ventas.
- **Incremento de ventas post-review:** Aumento de ventas después de recibir una calificación positiva de críticos o usuarios.

- **Promedio de calificaciones:** Promedio de calificaciones de críticos y usuarios para los juegos más vendidos.
- : Aumento de ventas después de recibir una calificación positiva de críticos o usuarios.
- **Promedio de calificaciones:** Promedio de calificaciones de críticos y usuarios para los juegos más vendidos.

Herramientas:

Para el desarrollo de este proyecto, hemos utilizado una variedad de herramientas que nos permitieron recolectar, limpiar, analizar y visualizar los datos de manera efectiva. A continuación, se describen las principales herramientas empleadas y su relevancia en el proyecto:

Microsoft Excel:

Descripción: Software de hoja de cálculo utilizado para la recolección inicial, limpieza y transformación de datos.

Aplicación en el Proyecto: Utilizamos Excel para realizar operaciones de búsqueda y vinculación de datos mediante la función BUSCARV, así como para la limpieza de datos y eliminación de registros duplicados. También se empleó para la creación de tablas dinámicas y gráficos básicos que facilitaron la exploración preliminar de los datos.

Python:

Descripción: Lenguaje de programación de alto nivel utilizado para la manipulación avanzada de datos, modelado y automatización de tareas.

Aplicación en el Proyecto: Desarrollamos scripts de Python para filtrar, organizar y transformar los datos según las necesidades del análisis. Utilizamos bibliotecas como pandas para el manejo de datos, numpy para operaciones numéricas, y matplotlib y seaborn para la creación de gráficos y visualizaciones.

Google Colab:

Descripción: Entorno basado en la nube para el desarrollo de cuadernos de Jupyter, que permite la colaboración y el uso de recursos computacionales avanzados.

Aplicación en el Proyecto: Empleamos Google Colab para ejecutar análisis de datos y visualizaciones cuando necesitábamos recursos computacionales adicionales. Esta herramienta nos permitió trabajar de manera colaborativa en tiempo real y acceder a bibliotecas y entornos preconfigurados.

Power BI:

Descripción: Herramienta de análisis de negocio que proporciona capacidades de visualización de datos y creación de informes interactivos.

Aplicación en el Proyecto: Power BI se utilizó para diseñar informes ejecutivos en formato PDF, consolidando las estadísticas de los indicadores y presentando los hallazgos de manera clara y profesional. La herramienta permitió integrar datos de diversas fuentes y aplicar filtros dinámicos para una mejor comprensión de los resultados.

Análisis de resultados encuesta:

Objetivo 1: Perfil de Juegos Populares

Descripción: Identificar las características comunes de los juegos más vendidos y mejor calificados.

Clasificación Promedio: 8.67

Análisis: La puntuación promedio que obtuvimos de 8.67 indica que los participantes consideraron que la visualización cumple de manera efectiva con el objetivo planteado. La mayoría de las respuestas se concentraron en los valores altos (7-9), lo que nos sugiere una alta satisfacción con la identificación de características comunes en los juegos populares.

Objetivo 2: Análisis de Desempeño por Plataforma

Descripción: Determinar qué plataformas tienen los juegos mejor recibidos y más vendidos.

Clasificación Promedio: 8.61

Análisis: Con una puntuación promedio de 8.61, los participantes también valoraron positivamente la visualización para este objetivo., indicando que la visualización proporciona información clara y útil sobre el desempeño por plataforma.

Objetivo 3: Evaluación del Desarrollador

Descripción: Analizar la relación entre el desarrollador y el éxito del juego.

Clasificación Promedio: 8.89

Análisis: Este objetivo obtuvo la puntuación promedio más alta (8.89), lo que refleja una percepción muy positiva sobre la capacidad de la visualización para mostrar la relación entre desarrolladores y el éxito de los juegos. La concentración de respuestas en el rango superior sugiere que los participantes encontraron esta visualización particularmente informativa y efectiva.

Objetivo 4: Tendencias de Género y Jugadores

Descripción: Explorar la popularidad de los diferentes géneros de juegos y la cantidad de jugadores.

Clasificación Promedio: 8.44

Análisis: La puntuación promedio de 8.44 indica una percepción positiva, aunque ligeramente menor en comparación con otros objetivos. Las respuestas fueron variadas, pero aún se concentran en los valores altos, lo que nos sugiere que la visualización aunque es efectiva, podría beneficiarse de mejoras adicionales para alcanzar el nivel de satisfacción de otros objetivos.

Objetivo 5: Impacto de la Crítica y Usuarios en Ventas

Descripción: Hallar cómo las calificaciones de críticos y usuarios afectan las ventas.

Clasificación Promedio: 8.67

Análisis: Con una puntuación de 8.67, este objetivo se alinea con el primer objetivo en términos de percepción positiva. Las respuestas reflejan una buena aceptación de la visualización en términos de su capacidad para mostrar el impacto de las calificaciones en las ventas.

1. Objetivo 1

Perfil de Juegos Populares: Identificar las características comunes de los juegos más vendidos y mejor calificados (como género, plataforma y desarrollador), para dirigir futuras inversiones y estrategias de marketing.

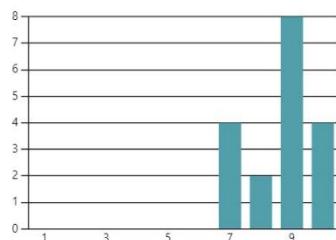
Teniendo en cuenta el objetivo y la pantalla de visualización mostrada. En una escala del 1 al 10, ¿Qué tanto siente que se cumple con lo planteado?

[Más detalles](#)

[Información](#)

8.67

Clasificación promedio



3. Objetivo 3

Evaluación del Desarrollador: Analizar la relación entre el desarrollador y el éxito del juego, tanto en ventas como en calificaciones, para identificar desarrolladores con los que asociarse o invertir y en que géneros especializarse.

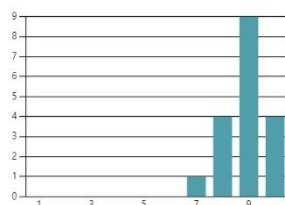
Teniendo en cuenta el objetivo y la pantalla de visualización mostrada. En una escala del 1 al 10, ¿Qué tanto siente que se cumple con lo planteado?

[Más detalles](#)

[Información](#)

8.89

Clasificación promedio



2. Objetivo 2

Análisis de Desempeño por Plataforma: Determinar qué plataformas tienen los juegos mejor recibidos y más vendidos, lo que puede guiar decisiones sobre en qué plataformas centrar los esfuerzos de desarrollo e identificar oportunidades en plataformas en las cuales se tienen carencias de interés y ventas.

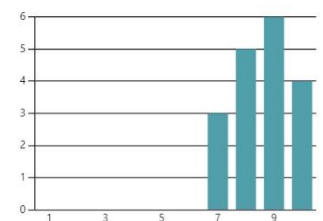
Teniendo en cuenta el objetivo y la pantalla de visualización mostrada. En una escala del 1 al 10, ¿Qué tanto siente que se cumple con lo planteado?

[Más detalles](#)

[Información](#)

8.61

Clasificación promedio



4. Objetivo 4

Tendencias de Género y Jugadores: Explorar la popularidad de los diferentes géneros de juegos y la cantidad de jugadores que tienen, para informar el desarrollo de juegos que cumplen las preferencias actuales de los usuarios.

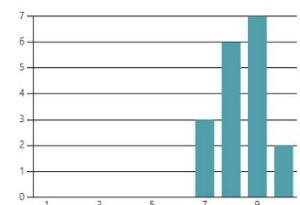
Teniendo en cuenta el objetivo y la pantalla de visualización mostrada. En una escala del 1 al 10, ¿Qué tanto siente que se cumple con lo planteado?

[Más detalles](#)

[Información](#)

8.44

Clasificación promedio



5. Objetivo 5

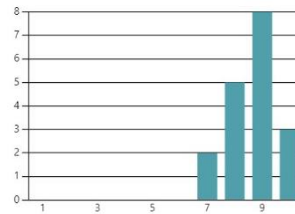
Impacto de la Crítica y Usuarios en Ventas: Hallar cómo las calificaciones de críticos y usuarios afectan las ventas globales y regionales, para entender el peso de las reseñas en el éxito comercial de los videojuegos.

Teniendo en cuenta el objetivo y la pantalla de visualización mostrada. En una escala del 1 al 10, ¿Qué tanto siente que se cumple con lo planteado?

[Más detalles](#)

[Información](#)

8.67
Clasificación promedio



Conclusiones:

A lo largo de este proyecto, hemos podido profundizar en el análisis y la visualización de datos relacionados con el mercado de videojuegos, una industria que presenta un dinamismo y crecimiento sorprendente y constante. Nuestro enfoque en la recopilación y análisis de datos nos ha permitido obtener insights valiosos que pueden informar y fortalecer la toma de decisiones estratégicas en la empresa.

Primero, hemos logrado identificar patrones significativos en las ventas y calificaciones de videojuegos, lo que nos ha proporcionado una visión clara de las características comunes entre los títulos más exitosos. Este análisis nos permitirá dirigir mejor las inversiones y estrategias de marketing hacia géneros, plataformas y desarrolladores que presentan un mayor potencial de éxito.

Segundo, nuestro análisis ha destacado la importancia de ciertas plataformas en el rendimiento de ventas de los videojuegos. Esto nos permitirá enfocar los esfuerzos de desarrollo y distribución en aquellas plataformas que demuestran ser más lucrativas y explorar oportunidades en plataformas emergentes.

Además, hemos podido evaluar la relación entre desarrolladores y el éxito de sus títulos. Este conocimiento nos permitirá identificar posibles asociaciones estratégicas con desarrolladores clave y orientar nuestros recursos hacia géneros específicos en los que estos desarrolladores han demostrado tener un mayor impacto.

La exploración de tendencias en géneros y preferencias de jugadores nos ha proporcionado información crucial para el desarrollo de juegos que se alineen con las expectativas y deseos del mercado actual. Esta información será vital para adaptar nuestras ofertas y mantenerse competitivos en un mercado en constante cambio.

Finalmente, hemos comprendido mejor el impacto de las reseñas de críticos y usuarios en las ventas de videojuegos. Este hallazgo subraya la importancia de mantener una alta calidad y receptividad a las opiniones del público, lo cual es fundamental para el éxito comercial de nuestros productos.

Este proyecto nos ha permitido consolidar una base sólida de datos y análisis que sería invaluable para la planificación estratégica futura. Las conclusiones obtenidas no solo refuerzan la importancia de la analítica y visualización de datos en la toma de decisiones, sino que también nos proporcionan herramientas prácticas para mejorar nuestras operaciones y estrategias en el competitivo mercado de los videojuegos.

Lecciones Aprendidas:

Durante el desarrollo de este proyecto, hemos adquirido valiosas experiencias y conocimientos que nos han permitido mejorar tanto nuestras habilidades técnicas como nuestra comprensión del proceso de visualización y análisis de datos en un contexto empresarial. A continuación, destacamos las principales lecciones aprendidas durante este proceso:

Importancia de la Comprensión del Negocio: Comprender a fondo el sector y las necesidades específicas de la empresa es fundamental para orientar correctamente el análisis de datos y la visualización. Esta comprensión inicial facilita la formulación de preguntas relevantes y la definición de objetivos alineados con las prioridades del negocio.

Relevancia de la Calidad de los Datos: La calidad de los datos recolectados es crucial para obtener resultados precisos y valiosos. Aprendimos la importancia de realizar un proceso riguroso de limpieza y validación de datos para asegurar su integridad y fiabilidad. Esto incluye la detección y corrección de valores atípicos, la imputación de datos faltantes y la estandarización de formatos.

Efectividad de las Técnicas de Modelado y Transformación de Datos: Emplear técnicas adecuadas de modelado y transformación de datos nos permitió estructurar la información de manera eficiente para el análisis. La utilización de herramientas como Excel y scripts de Python facilitó la preparación de los datos, permitiendo una integración fluida de diversas fuentes de información.

Diseño de Visualizaciones Interactivas: La creación de visualizaciones dinámicas y dashboards interactivos resultó ser una herramienta poderosa para presentar los hallazgos de manera clara y comprensible. Las visualizaciones efectivas no solo mejoran la interpretación de los datos, sino que también permiten a los usuarios explorar la información desde diferentes perspectivas y tomar decisiones informadas.

Colaboración y Trabajo en Equipo: El trabajo en equipo fue esencial para el éxito del proyecto. La colaboración y comunicación constante entre los miembros del grupo nos permitió abordar los desafíos de manera más eficiente y aprovechar al máximo las habilidades individuales de cada integrante. La división de tareas y la asignación de responsabilidades claras fueron clave para mantener el proyecto organizado y cumplir con los plazos establecidos.

Flexibilidad y Adaptabilidad: Durante el desarrollo del proyecto, enfrentamos varios desafíos imprevistos que requirieron adaptabilidad y flexibilidad. Aprendimos a ajustar nuestras estrategias y métodos de trabajo en función de las circunstancias, lo que nos permitió superar obstáculos y continuar avanzando hacia nuestros objetivos.

Valor de las Pruebas de Aceptación: Validar los reportes y visualizaciones con usuarios finales a través de encuestas y pruebas de aceptación fue fundamental para asegurar que nuestros productos finales cumplieran con las expectativas y necesidades del cliente. Este proceso nos proporcionó retroalimentación valiosa que nos permitió realizar ajustes y mejoras antes de la presentación final.

Lista de referencias

- Intarex. (n.d.). *La analítica de datos: Clave en la toma de decisiones*. Recuperado de [Intarex](#)
- InLab FIB. (2017). *¿Qué es la visualización de datos (DataViz)?*. Recuperado de [InLab FIB](#)
- The Data Privacy Group. (n.d.). *What is a Data Dictionary?*. Recuperado de [The Data Privacy Group](#)
- Alteryx. (n.d.). *Data Preparation*. Recuperado de [Alteryx](#)
- Universidad Europea. (n.d.). *Modelo entidad-relación*. Recuperado de [Universidad Europea](#)
- IBM. (n.d.). *Dimension tables and entities*. Recuperado de [IBM](#)
- Data Science Dojo. (2019). *Data Cleaning Techniques*. Recuperado de Data Science Dojo
- Towards Data Science. (2018). *The Importance of Data Quality*. Recuperado de Towards Data Science
- Analytics Vidhya. (2020). *A Comprehensive Guide to Data Preparation*. Recuperado de Analytics Vidhya
- Practical Data Science. (2021). *Best Practices for Data Visualization*. Recuperado de Practical Data Science
- Real Python. (n.d.). *Data Analysis with Python*. Recuperado de Real Python
- Power BI Documentation. (n.d.). *Creating Interactive Dashboards*. Recuperado de [Power BI Documentation](#)

