



Benemérita Universidad Autónoma de Puebla

Facultad de ciencias de la computación (FCC)

Análisis exploratorio sobre las tendencias populares en la industria de los videojuegos

Introducción a la ciencia de datos

Responsable: Manzanarez Peña Victor Hugo

Docente: Jaime Alejandro Romero Sierra

Índice

| | |
|--|-------------|
| Objetivo del análisis..... | Pag. 3 |
| Descripción del problema..... | Pag. 3 |
| Tecnologías y herramientas..... | Pag. 4 |
| Descripción de los datos..... | Pag. 4 |
| Hipótesis Iniciales..... | Pag. 5 |
| Definición de Stakeholders Clave..... | Pag. 5-6 |
| Preguntas clave..... | Pag. 6-7 |
| Fuentes de datos identificadas..... | Pag. 8-9-10 |
| Justificación del proyecto..... | Pag. 10-11 |
| ¿Cuántos datos y de qué tipo son?..... | Pag. 11-12 |

Objetivo del análisis: Identificar que géneros, plataformas y regiones son más rentables en la industria de los videojuegos. De esta forma las distribuidoras y desarrolladoras tienen un panorama más claro sobre qué tipo de juegos producir, en que plataformas sacarlos y en que mercados pueden llegar a tener mayor éxito.

Descripción del problema: Las desarrolladoras e industrias no siempre cuentan con la opción de lanzar un videojuego para todas las plataformas y de forma global. De igual forma no siempre se cuenta con la información adecuada sobre el público al que uno piensa que el juego va dirigido, arriesgándose así a perder un mercado clave, no conseguir la meta de ventas, que el género del juego no sea del agrado del mercado, que el juego se lance en un momento inoportuno o que en el peor de los casos, este sea un fracaso. Con este análisis se busca apoyar a la industria y a las desarrolladoras a tener un enfoque más claro sobre en qué plataforma es más rentable lanzar el videojuego, en que regiones es más rentable lanzar el videojuego o si conviene lanzarlo de manera global, el momento adecuado tanto para su anuncio como para su lanzamiento, que público es más propenso a gustar del contenido creado y por ende en la zona geográfica precisa en la cual este podría tener más éxito, maximizando así las ganancias y optimizando la inversión hecha en cosas como publicidad, distribución de copias físicas, permisos y licencias, entre otras.

Tecnología y herramientas:

Se cuenta con las siguientes tecnologías para realizar el análisis de los datos:

- Lenguajes y entorno de programación: Python, Visual Studio Code, Jupyter Notebook, google colab.
- Librerías: Pandas, matplotlib, seaborn, numphy, plotly, scikit-learn, xgboost.

Disclaimer: Esta es una lista preliminar, pueden o no hacerse uso de las tecnologías mencionadas aquí en un futuro y de la misma forma, hacer uso de otras tecnologías y herramientas no mencionadas.

Descripción de los datos:

- Rango (Rank): Clasificación de las ventas totales
- Nombre (Name): Nombre del videojuego
- Plataforma (Platform): Plataforma(s) en la que se lanzó el videojuego
- Año (Year): Año en el que se lanzó el videojuego
- Género (Genre): Genero del videojuego
- Editora (Publisher): Editora del videojuego
- Ventas en NA (NA sales): Ventas del videojuego en Norteamérica en millones
- Ventas en EU (EU Sales): Ventas en europa en millones
- Ventas en JP (JP Sales): Ventas en japon en millones
- Otras ventas (Other Sales): Ventas en el resto del mundo
- Ventas globales (Global Sales): Ventas mundiales totales

Hipótesis iniciales

1.- Los videojuegos lanzados en consolas tienen mayor probabilidad de éxito que los lanzados en consolas portátiles o PC

En la industria del gaming, mayormente se priorizan las consolas de sobremesa ya que históricamente han sido las que más impacto generan tanto por su innovación como por sus campañas publicitarias.

2.-Los videojuegos de genero Acción y Deportes contienen la mayor parte de las ventas globales.

Por popularidad y ser más atractivos para el público joven, los géneros más comunes son los que suelen atraer más jugadores nuevos y por ende mayores ventas

3.-La región de Norteamérica representa el mercado más grande en ventas de videojuegos

La economía y la cultura gamer en Norteamérica suelen ser mayores que en otras regiones.

Definición de StakeHolders Clave

Desarrolladores de videojuegos: Saber que géneros, plataformas y características generan más ventas para así alinearse a las preferencias del mercado al que quieran dirigirse

Publishers (Empresas distribuidoras): Identificar los mercados mas rentables por región y época del año en la cual es mas recomendable lanzar un videojuego

Plataformas de venta: Conocer cuáles son los títulos y géneros que deben tener en stock para mejorar la planeación del inventario y maximizar así sus ventas

Inversionistas de la industria: Identificar que mercados pueden explotar o por el contrario, cuáles van en declive. De esta forma saben en qué empresas y proyectos deben invertir

Departamentos de marketing y publicidad: Ajustar sus campañas de promoción en base a tendencias con mayor éxito, ajustando así su publicidad en base al género, región o plataforma

Preguntas clave

¿Qué géneros de videojuegos han generado más ventas globales a lo largo del tiempo?

¿Cuáles son las plataformas más rentables en términos de ventas totales?

¿Existe una relación entre el año de lanzamiento y la cantidad de ventas totales?

¿En qué región (Norteamérica, Europa, Japón, Otros) se concentra la mayor cantidad de ventas?

¿Las ventas de un videojuego están relacionadas a su Publisher?

¿Los géneros más populares son los mismos en las diferentes regiones?

¿Qué compañías han logrado con más frecuencia, posicionarse entre las mas vendedoras?

¿Las críticas positivas tienen alguna relación con mayores ventas?

¿Existe alguna relación entre las fechas en las que se lanzaron parte de los videojuegos más exitosos?'

¿Cuáles son los géneros que han decaído más a lo largo de los años?

Fuentes de datos identificadas

1.- Kaggle – Video Game Sales Dataset

Es la fuente de información y base de datos principal, contiene información histórica sobre ventas de videojuegos a nivel global.

```
import pandas as pd
df = pd.read_csv("vgsales.csv")
df
```

| | Rank | Name | Platform | Year | Genre | Publisher | NA_Sales | EU_Sales | JP_Sales | Other_Sales | Global_Sales |
|-------|-------|--|----------|--------|--------------|------------|----------|----------|----------|-------------|--------------|
| 0 | 1 | Wii Sports | Wii | 2006.0 | Sports | Nintendo | 41.49 | 29.02 | 3.77 | 8.46 | 82.74 |
| 1 | 2 | Super Mario Bros. | NES | 1985.0 | Platform | Nintendo | 29.08 | 3.58 | 6.81 | 0.77 | 40.24 |
| 2 | 3 | Mario Kart Wii | Wii | 2008.0 | Racing | Nintendo | 15.85 | 12.88 | 3.79 | 3.31 | 35.82 |
| 3 | 4 | Wii Sports Resort | Wii | 2009.0 | Sports | Nintendo | 15.75 | 11.01 | 3.28 | 2.96 | 33.00 |
| 4 | 5 | Pokemon Red/Pokemon Blue | GB | 1996.0 | Role-Playing | Nintendo | 11.27 | 8.89 | 10.22 | 1.00 | 31.37 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 16593 | 16596 | Woody Woodpecker in Crazy Castle 5 | GBA | 2002.0 | Platform | Kemco | 0.01 | 0.00 | 0.00 | 0.00 | 0.01 |
| 16594 | 16597 | Men in Black II: Alien Escape | GC | 2003.0 | Shooter | Infogrames | 0.01 | 0.00 | 0.00 | 0.00 | 0.01 |
| 16595 | 16598 | SCORE International Baja 1000: The Official Game | PS2 | 2008.0 | Racing | Activision | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 |
| 16596 | 16599 | Know How 2 | DS | 2010.0 | Puzzle | 7G//AMES | 0.00 | 0.01 | 0.00 | 0.00 | 0.01 |
| 16597 | 16600 | Spirits & Spells | GBA | 2003.0 | Platform | Wanadoo | 0.01 | 0.00 | 0.00 | 0.00 | 0.01 |

16598 rows × 11 columns


2.- Fuentes secundarias:

Entertainment Software Association (ESA): Provee estadísticas anuales del mercado en Norteamérica.



Newzoo Global Games Market Report: Analisis global de ventas, ingresos y tendencias.

More Free Reports




Unlocking Games Revenue: Player Behavior and Payment Trends in the West

Free Report

August 12, 2025

Unlocking games revenue: player behavior and payment trends in the west | Newzoo x Tebex Whitepaper

Games Market | Gamer Research



Insights into the Japanese Gaming Market in 2025

Free report

July 30, 2025

Free Report

Japanese Gaming Market Report 2025

PC & Console | Games Market | Games Data



2025 Mapping IP Fandom with the Global Gamer Study

Free report

July 1, 2025

Free Report

Mapping IP Fandom with the Global Gamer Study in 2025

Gamer Research

Statista: Datos complementarios de consumo de videojuegos por región y género.

Mercados líderes en el sector de los videojuegos según ingresos en 2024

Principales mercados de **videojuegos** a nivel mundial en 2024, según ingresos (en millones de dólares)

 Topic | 4 de julio de 2025


Industria del videojuego en España - Datos estadísticos

Datos estadísticos sobre la industria del videojuego en España

 Topic | 18 de julio de 2025

Industria mundial del videojuego - Datos estadísticos

Datos estadísticos sobre la industria del videojuego en el mundo

 Topic | 30 de septiembre de 2024

Industria del videojuego en Estados Unidos - Datos estadísticos

Datos estadísticos sobre la industria del videojuego en Estados Unidos

Metacritic/OpenCritic: Calificaciones de críticos y usuarios, útiles para probar hipótesis relacionadas a la calidad percibida contra las ventas.

GAMES

Find your next captivating gaming moment

New Releases [SEE ALL](#)



Digimon Story Time ...

78 Generally Favorable



Ghost of Yotei

86 Generally Favorable



Super Mario Galaxy ...

88 Generally Favorable



LEGO Party!

80 Generally Favorable

Justificación del proyecto

La industria de los videojuegos en los últimos años se ha convertido en uno de los sectores de entretenimiento más rentables.

Comprender las tendencias de consumo en el mercado y el hecho de tener registros de éxitos históricos en ventas de videojuegos, amplía el panorama para las nuevas generaciones para poder tomar decisiones estratégicas e informadas. Tener acceso a un análisis amplio sobre ventas, géneros, fechas y editoriales, les permite a los desarrolladores tener una referencia sobre hacia donde quieren dirigir su proyecto. Pueden tomar decisiones más inteligentes en base a decisiones pasadas y de esta manera (quizá no asegurar) pero si mejorar en mayor medida el margen de éxito que podrían alcanzar. De la misma forma logra que Project Sponsors y diferentes inversionistas adquieran conocimiento sobre qué proyectos son más propensos a explotar y cuáles no. También ayuda a que desarrolladoras independientes puedan tomar ejemplo de éxitos previos para así encaminar de manera adecuada sus propios lanzamientos. Este análisis de igual manera resulta relevante ya que no solo permite entender con mayor facilidad el mercado global, si no también aporta una visión detallada por regiones, lo que es fundamental en un mercado tan diverso como lo es el de los videojuegos. Además, el análisis puede ayudar a responder preguntas críticas para el negocio, como qué factores determinan el éxito de un videojuego, cómo influyen los géneros y plataformas en la rentabilidad, qué Publisher dominan el mercado y por qué. A partir de esta información, se pueden tomar decisiones que contribuyan a reducir riesgos financieros en el desarrollo de nuevos títulos.

¿Cuántos datos y de qué tipo son?

El dataset cuenta actualmente con 16598 filas y 11 columnas. De los cuales se cuentan con 6 columnas con datos tipo float, 1 columna de datos tipo int y 4 columnas con datos tipo object.

Este número de datos es previo a la limpieza del dataset, no es el número final de datos con el que se contará para el análisis.

Los tipos de datos específicos de cada columna se muestran a continuación:

```
Data columns (total 11 columns):
#      Column      Non-Null Count  Dtype
---  -
0      Rank        16598 non-null    int64
1      Name         16598 non-null    object
2      Platform     16598 non-null    object
3      Year         16327 non-null    float64
4      Genre        16598 non-null    object
5      Publisher    16540 non-null    object
6      NA_Sales     16598 non-null    float64
7      EU_Sales     16598 non-null    float64
8      JP_Sales     16598 non-null    float64
9      Other_Sales  16598 non-null    float64
10     Global_Sales  16598 non-null    float64
dtypes: float64(6), int64(1), object(4)
memory usage: 1.4+ MB
```