



DATA SCIENCE APPLIED TO VIDEO GAME DEVELOPING

Análisis y modelamiento para proyectos
de desarrollo de videojuegos



AGENDA



ABSTRACTO

01

EXPLORATORY
DATA ANALYSIS

03

MODELAMIENTO Y
VALIDACIÓN
CRUZADA

05

02

RESUMEN
METADATA

04

SELECCIÓN DE
VARIABLES Y
ENCODING

06

CONCLUSIONES
FINALES



ABSTRACTO



La Investigación realizada busca dar una servir como **apoyo a los equipos desarrolladores de videojuegos**, con el fin de aumentar las probabilidades de éxito y las ventas futuras.

Otra de las motivaciones detrás de esta investigación es **dar apoyo en la asignación de presupuesto** eficiente para el desarrollo del proyecto considerando el éxito esperado de este.

Mediante la **utilización de data histórica pública** de distintos videojuegos, se presenta un **análisis gráfico y estadístico de las distintas variables** y cómo estas influyen entre sí para **contestar las preguntas e hipótesis** planteadas.

Finalmente, se presenta un modelo de **machine learning** que busca dar un parámetro para **estimar el éxito o fracaso de ventas** de un proyecto de desarrollo de videojuegos en base a un parámetro definido.





RESUMEN METADATA



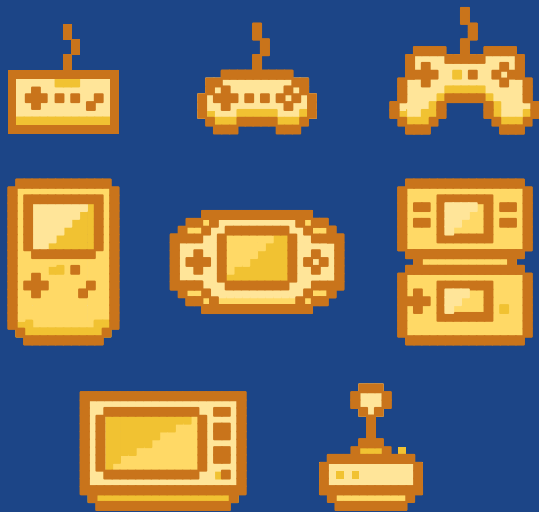
Dataset de más de 100.000 registros de videojuegos.

Variable	Descripción
Rank	Ranking de las ventas promedio
Name	Nombre del videojuego
Platform	Plataforma de lanzamiento del videojuego (PC, PS4, etc.)
Year	Año de lanzamiento del videojuego
Genre	Género del videojuego
Publisher	Editor del videojuego
Global_Sale	Ventas totales a nivel mundial.
Critic_score	Puntaje agregado del staff de Metacritic
User_score	Puntaje asignado por los suscriptores de Metacritic
Usercount	Número de usuarios que dieron puntaje de crítica
Developer	Desarrollador
Rating	Rating asignado por la ESRB



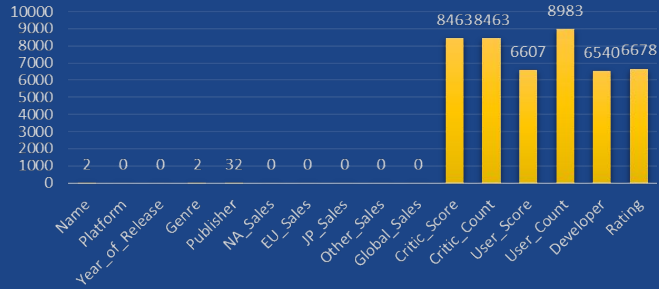
┐ PREGUNTAS HIPÓTESIS ┌

▲
¿Es posible determinar cierto nivel de ventas para un videojuego en base a ciertas características del mismo?

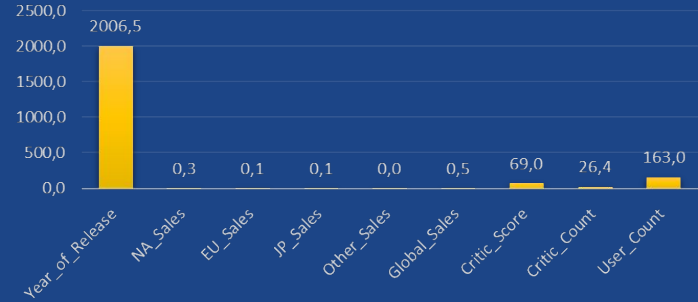


EXPLORATORY DATA ANALYSIS

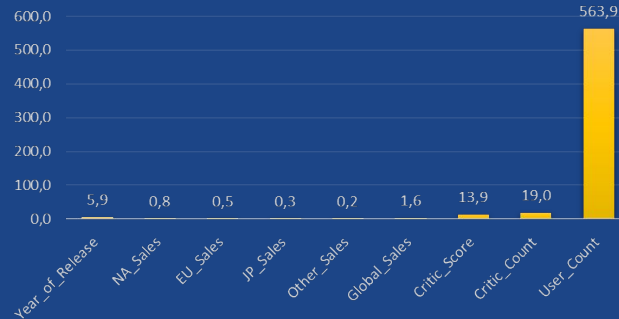
NaN por Variable



Media



Desviación Estándar



SELECCIÓN DE VARIABLES Y ENCODING



One-Hot Encoding:



Year

Se agrupó por
décadas con
columnas categóricas



Platform

Columnas categóricas
por plataforma



Genre

Columnas categóricas
por género

Scores



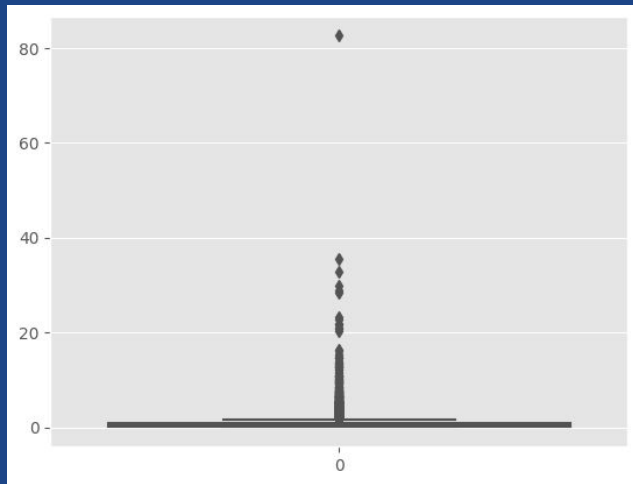
Normalización



SELECCIÓN DE VARIABLES Y ENCODING



Barplot de Global Sales:



Permite ver que la mayoría de las ventas globales de videojuegos se acumulan en cierto rango. Superar ese rango significa que pertenecen a **ventas mayoritariamente mejores que el común de la industria**. Dicho valor es muy cercano a la media.



MODELAMIENTO Y VALIDACIÓN CRUZADA



Modelo utilizado: Regresión Logística

Métricas de rendimiento obtenidas:

- Precisión: 0,72
- Recall: 0,41
- F1 Score: 0,53
- Accuracy: 0,81

El modelo tiene una precisión y accuracy decentes, lo que significa que **es capaz de clasificar correctamente la mayoría de las instancias**. Sin embargo, el recall es relativamente bajo, lo que indica que **el modelo puede perder algunas instancias positivas**. La puntuación F1 muestra un equilibrio razonable entre la precisión y el recall.



└ MODELAMIENTO Y VALIDACIÓN CRUZADA ─



Validación Cruzada

Puntaje Promedio Obtenido: 0,81

El puntaje indica que en promedio el **modelo tiene un rendimiento aceptable**. Por otro lado, la variabilidad de los puntajes obtenidos es baja, por lo que se puede considerar que **se puede seguir sosteniendo que el modelo tiene un buen rendimiento**.



✚ CONCLUSIONES FINALES ✚



Podemos concluir que el modelo resulta útil para poder determinar si un videojuego estará por sobre un nivel promedio de ventas o no. Sin embargo, es necesario recalcar que el número de videojuego que se encuentra cerca del promedio es bastante alto.

Dicho lo anterior, el cumplir con el propósito de justificar la asignación de presupuesto para un proyecto de desarrollo de videojuegos puede verse un poco facilitada, ya que hay características que escapan de este análisis.

Para futuros proyectos se podría considerar la posibilidad de determinar las ventas en base a un modelo de regresión lineal.

