

# Venta de Videojuegos



## Integrantes:

- Leandro Marin
- Leonel Colloca

***CODER HOUSE***

# *Entrega final*

## Índice:

1. Contexto.
2. Hipótesis/Preguntas de Interés.
3. Metadata y Análisis Exploratorio.
4. Insights.
5. Feature Analysis.
6. Feature Engineering.
7. Conclusión Final.

## Contexto comercial

*Debido al constante cambio tecnológico que vivimos, se ha producido un enorme desarrollo en los videojuegos, los cuales están próximos a simular nuestra realidad, lo cual nos beneficiaría a la hora de tomar decisiones, realizar ensayos o pruebas en dispositivos, entrenar a un piloto o una IA, diseñar y modificar estructuras previo a fabricarlas, etc permitiendo ahorrar costos y tiempo. Es por eso que las empresas más importantes del mundo están en la búsqueda de tecnologías como plataformas de videojuegos y simuladores que les permitan encontrar nuevas soluciones a sus servicios de forma más inmersiva.*

## **Definición del objetivo:**

*Con los datos históricos que se posee acerca de los videojuegos, géneros, plataformas y ventas, buscamos obtener indicios que demuestren qué tipos de plataformas serán las más utilizadas a futuro y qué género tendrá mayor relevancia permitiendo desarrollar nueva información sobre futuras ventas.*

## ***Problema comercial:***

Se pretende contar con un modelo predictivo que ayude a tomar las decisiones correctas sobre el futuro de las plataformas y géneros más relevantes de videojuegos.

## ***Contexto analítico:***

*Este conjunto de datos contiene una lista de videojuegos con ventas superiores a 100.000 copias.*

### **Los campos incluyen:**

- Rank: Ranking de ventas totales.
- Name: Nombre del juego.
- Plataform: Plataforma del juego.
- Year: Año en el cual el juego se lanzó.
- Genre: Género del juego.
- Publisher: Editor del juego.
- NA\_Sales: Ventas en Estados Unidos (Millones)
- EU\_Sales: Ventas en Europa (Millones)
- JP\_Sales: Ventas en Japón (Millones)
- Other\_Sales: Otras ventas (Millones)
- Global\_Sales: Total de ventas.

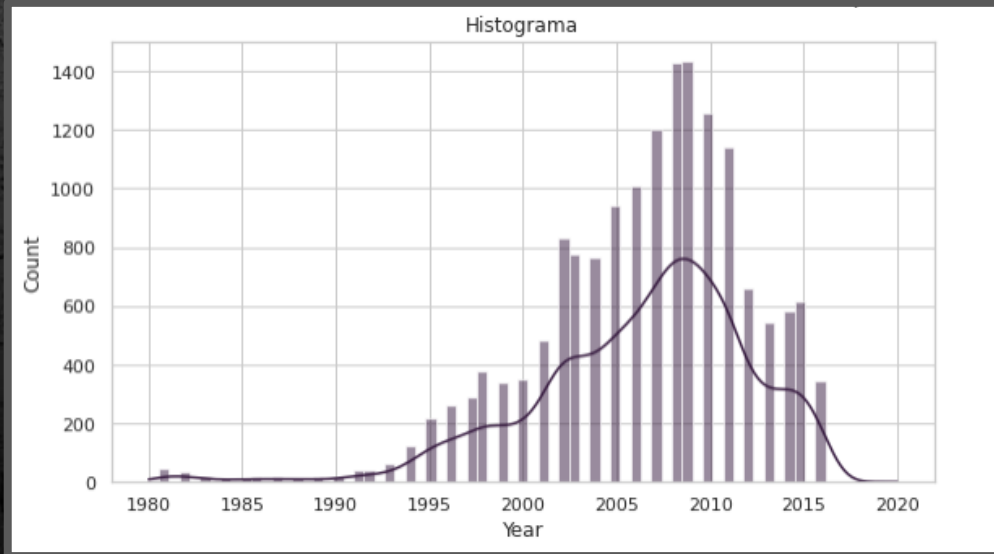


## *Serie de preguntas:*

- ¿Cuál fue la relación entre las ventas en Estados Unidos y el resto del mundo?
- ¿Cuál fue la relación entre las ventas en Europa y el resto del mundo?
- ¿Cuál fue la relación entre las ventas en Japón y el resto del mundo?
- ¿Cuáles son los juegos más vendidos?
- ¿Cuáles son los géneros más vendidos?
- ¿Cuál fue el formato de consola más utilizado?
- ¿Cuál fue el desarrollador que obtuvo más ventas?
- ¿Cuál fue la evolución entre las plataformas a lo largo de los años?
- ¿Cuánto ha variado la producción de juegos a lo largo de los años?

# Metadata y Análisis Exploratorio:

## PRODUCCIÓN DE JUEGOS ANUAL:



Se logra visualizar que existe un aumento en la cantidad de ventas de juegos con el paso de los años, sin embargo, también se logra identificar una caída en las ventas después del 2010.

Posibles factores:

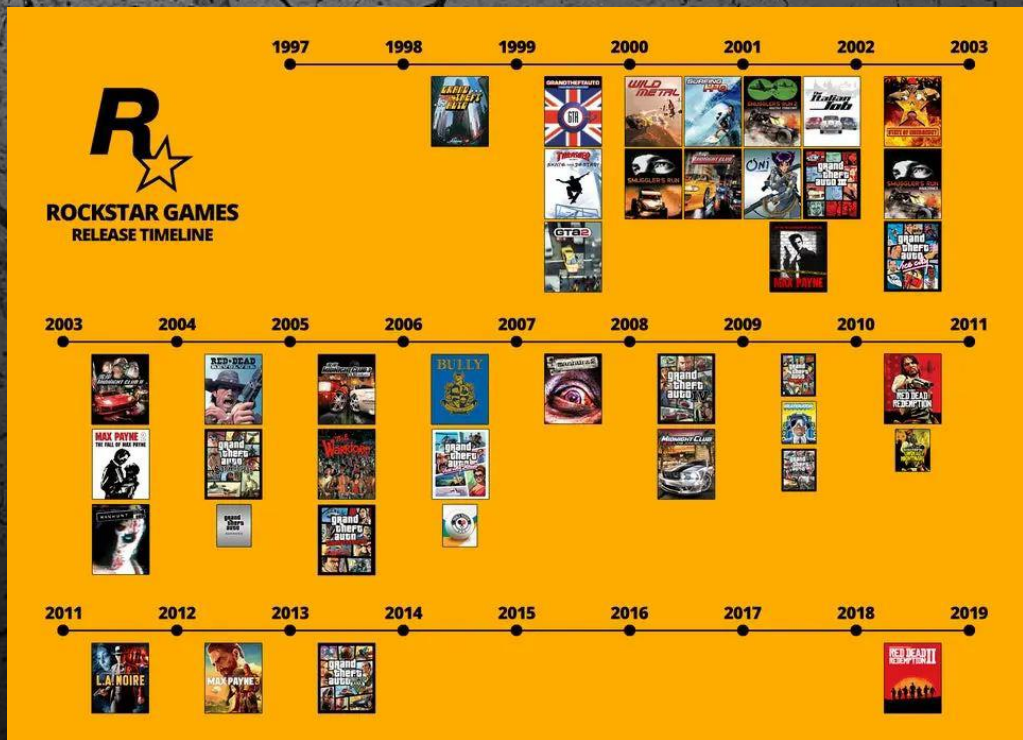
- Aumento de precios de juegos y/o consolas;
- Salida de servicios en la nube y dispositivos móviles que han tomado popularidad;
- Un cambio en el modelo de negocio actual, donde se busca explotar a los juegos por más tiempo con la entrega de actualizaciones y DLCs.

## Ejemplo de cambio de modelo de negocio: “Grand Theft Auto V”:

Un caso de ejemplo fue el Grand Theft Auto V, el cual se lanzó al mercado en septiembre del 2013 (aún pleno auge de producción de juegos).

Para octubre del mismo año sufrió una actualización que permitió a los usuarios jugar en modo Online.

Es llamativo que desde entonces Rockstar solo ha comercializado apenas el Red Dead Redemption 2 (lanzado en 2018) y el Borderlands 3 (lanzado en 2019).





## Ejemplo de cambio de modelo de negocio: “Grand Theft Auto V”:

El nuevo modo online del GTAV fue tal éxito que su desarrollador Rockstar North ha provisto al juego de numerosos DLCs para que siga siendo atractivo con los años y mucho más personalizable.

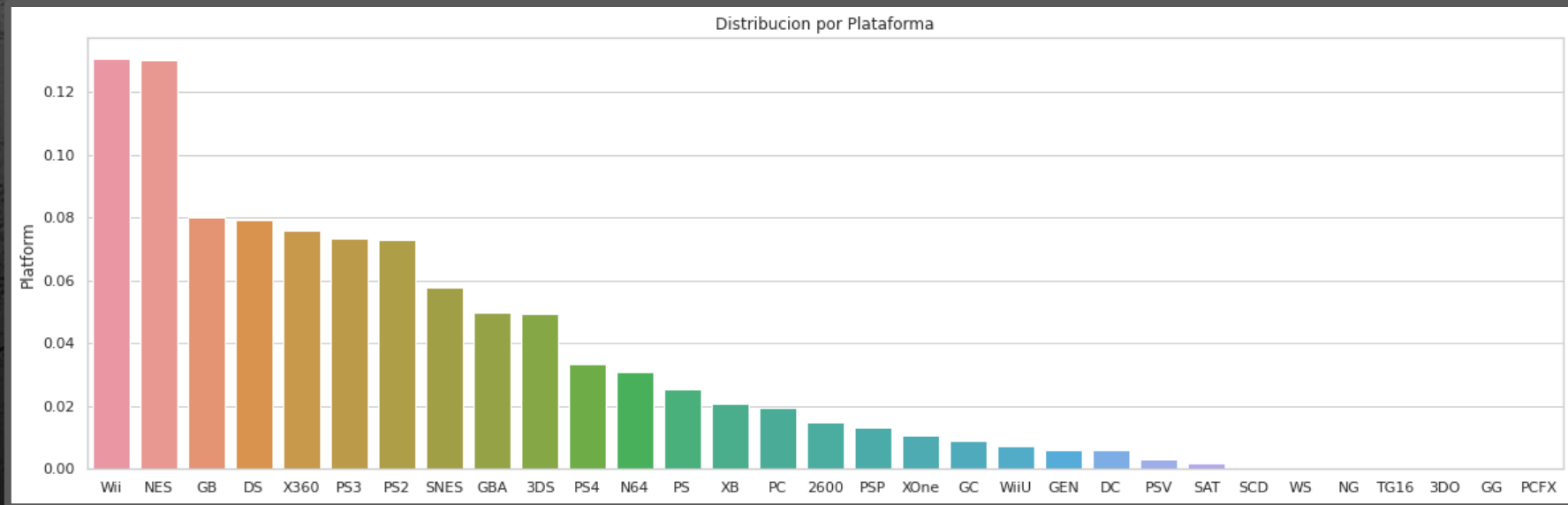
Con la salida de nuevas consolas se ha remasterizado el juego para mejorar los gráficos.

Esto ha permitido que hoy en día siga estando vigente, alcanzando 160 millones de ventas para el 2021.





## PROPORCIÓN DE CONSOLAS CON MÁS VENTAS:

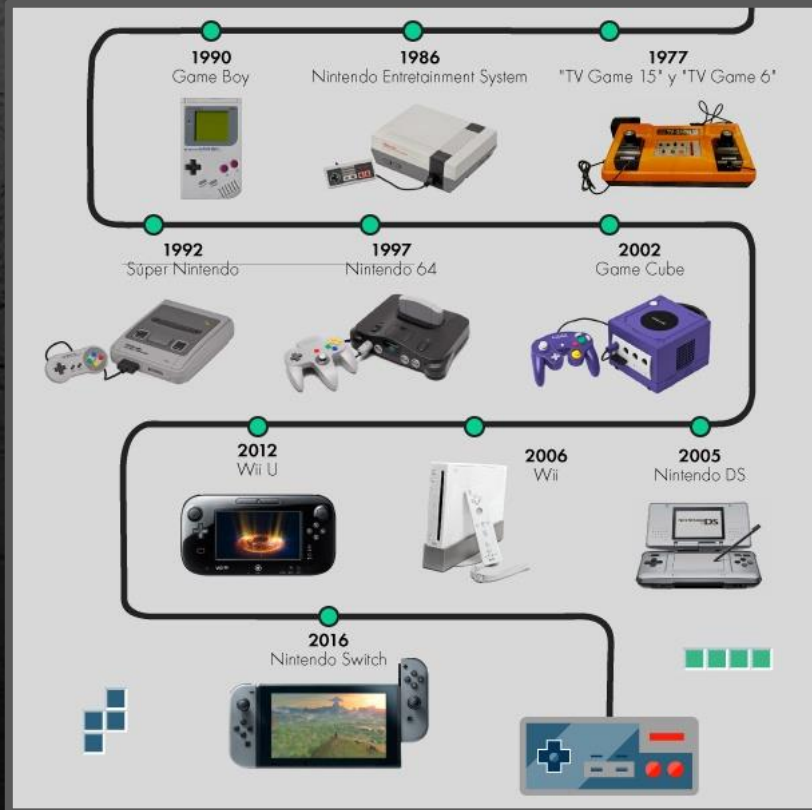


Se puede observar que las plataformas con un porcentaje considerable de ventas a lo largo de la historia fueron aquellas desarrolladas por Nintendo:

- la Nintendo Wii;
- la Nintendo NES;
- la GameBoy;
- la Nintendo DS.

Luego aparecen la Xbox 360 desarrollada por Microsoft, y las conocidas Playstation 2 y 3 desarrolladas por Sony.

## PROPORCIÓN DE CONSOLAS CON MÁS VENTAS:

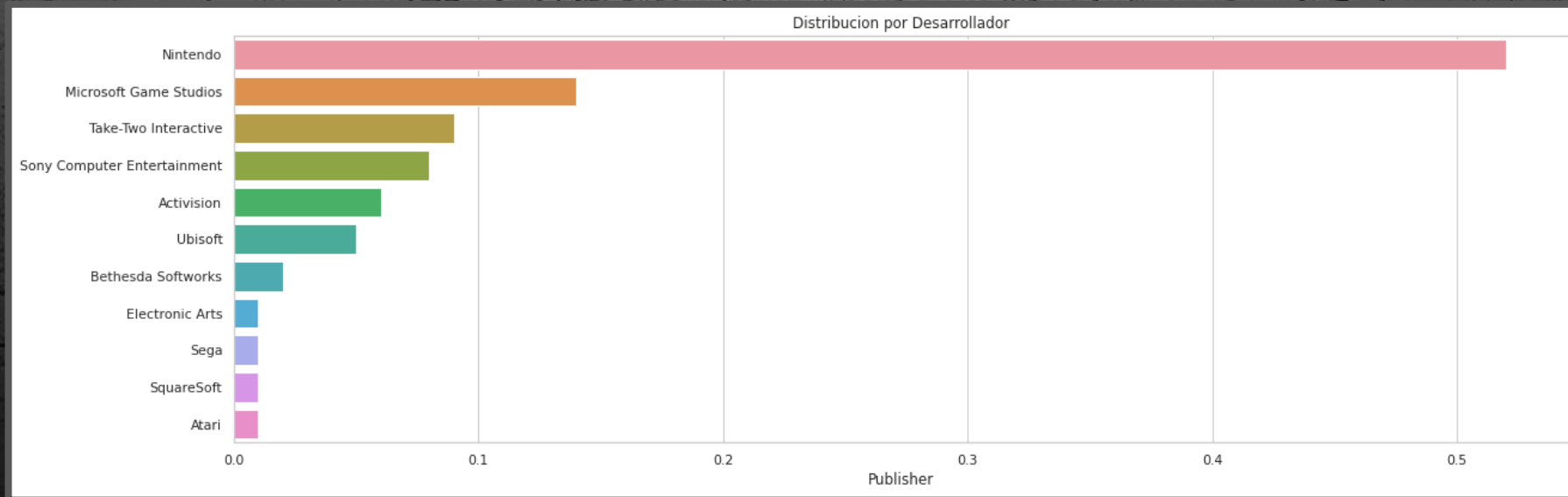


Es notable el éxito que tuvieron históricamente las consolas de Nintendo.

Cabe destacar a la Nintendo NES, que se ha mantenido como un éxito de ventas desde su salida en 1986.

Por otro lado, es llamativo que la Playstation 4 y la Xbox One no presentan un gran porcentaje de ventas comparado con sus antecesoras.

## PROPORCIÓN DE DESARROLLADORES CON MÁS VENTAS:



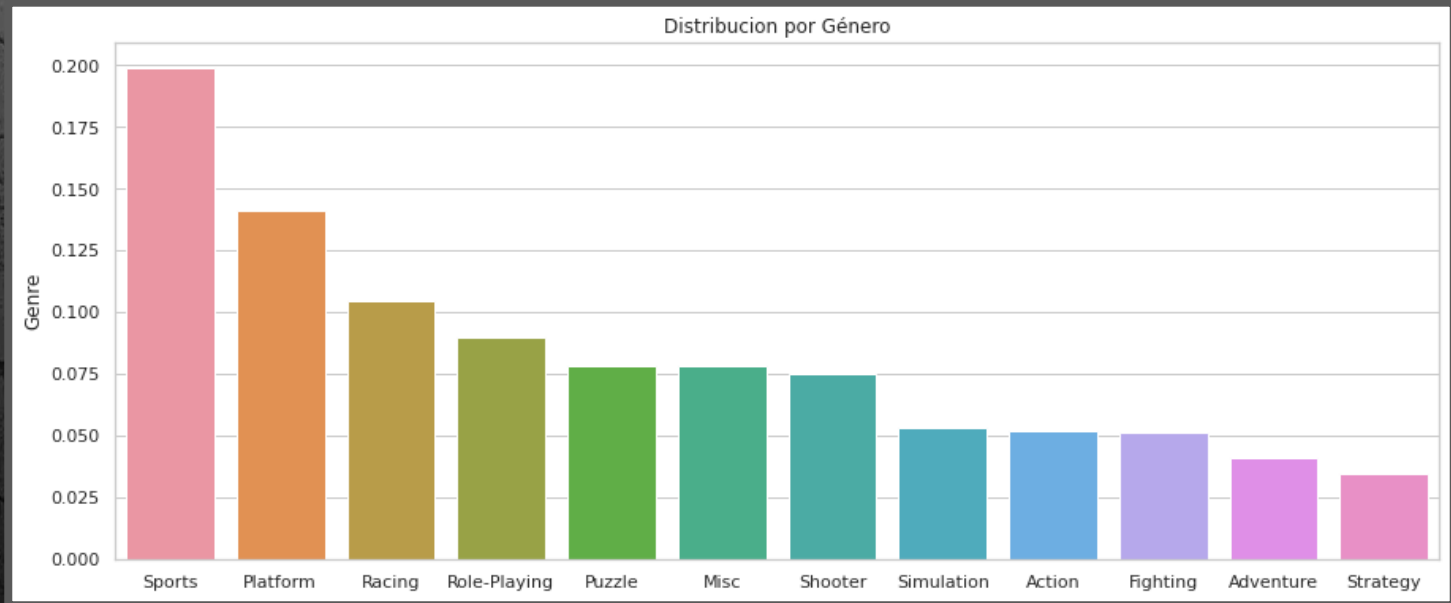
Se puede observar que el desarrollador que más juegos ha vendido a lo largo de la historia fue Nintendo (Así como también comprobamos anteriormente que es el que más plataformas ha ofrecido) con casi el 50% del mercado.

Es seguido por desarrolladores americanos como Microsoft Game Studios con el 15% de las ventas, y en menor porcentaje Take-Two (propietario de Rockstar Games), Sony, Activision y Ubisoft.





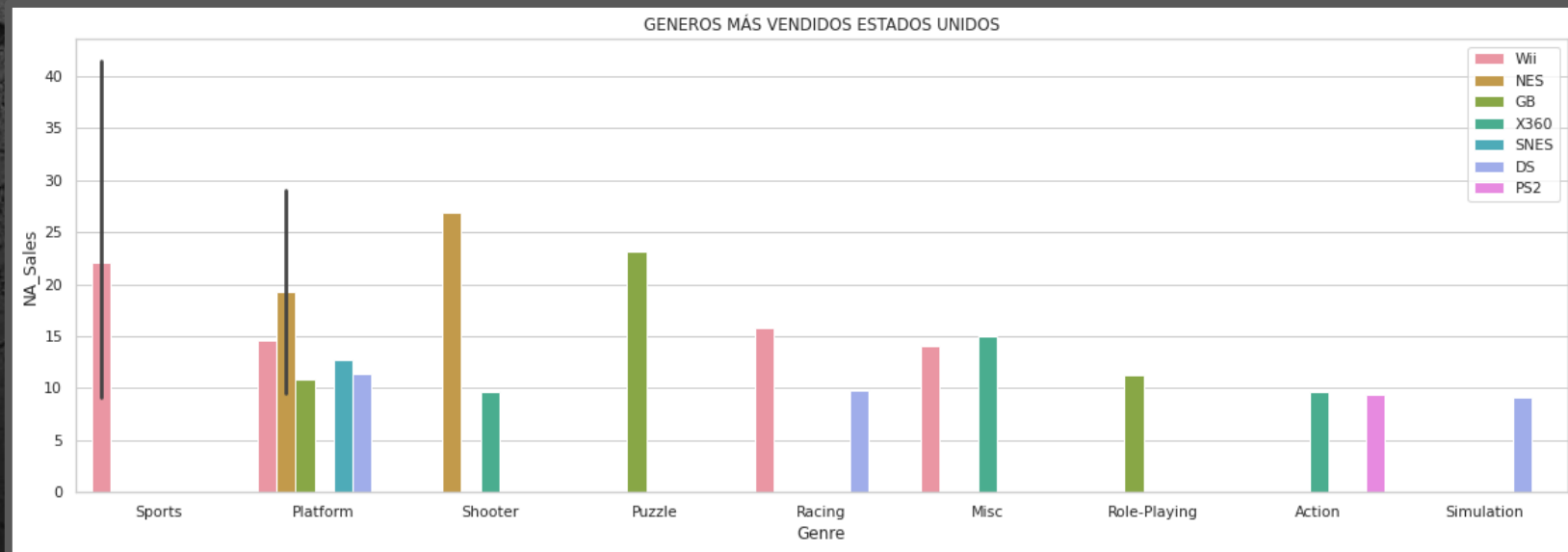
## PROPORCIÓN DE GÉNEROS CON MÁS VENTAS:



Por un lado se puede observar que el género más consumido es el de Deportes alcanzando casi el 20% del mercado, mientras lo siguen los juegos de Plataformas, Carreras y Role-Playing.

Es notable la baja popularidad de los juegos de Aventura y Estrategia.

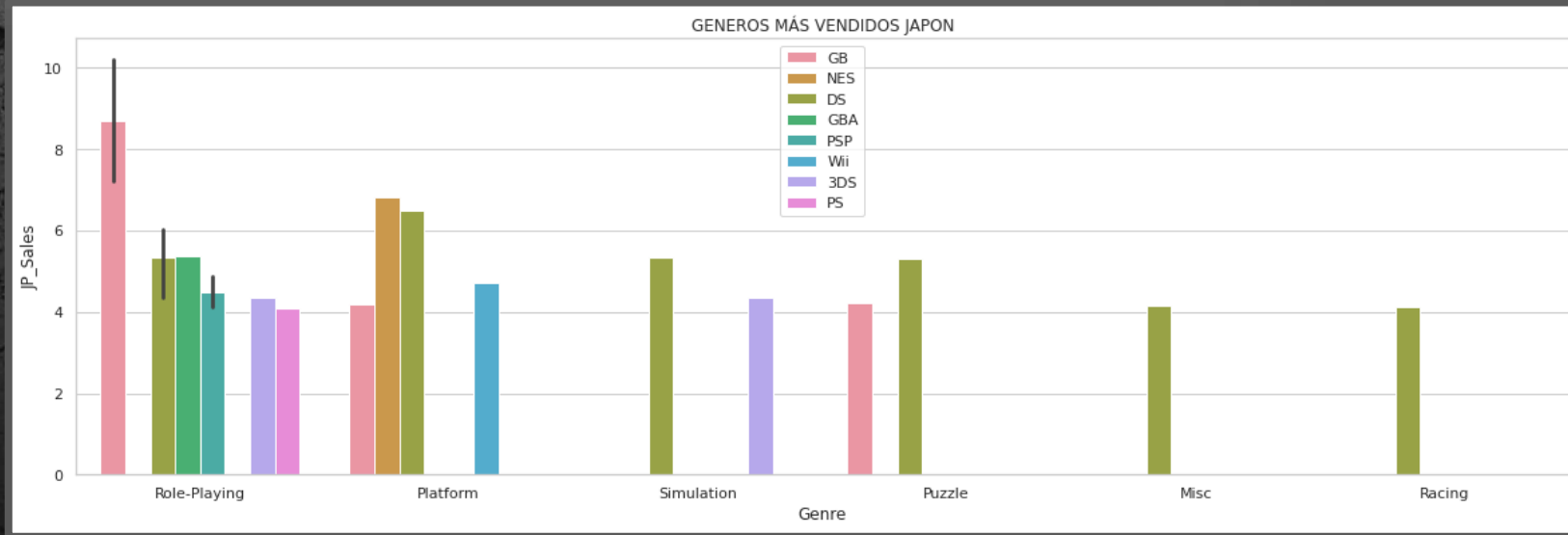
# POPULARIDAD DE GÉNEROS Y CONSOLAS DE ESTADOS UNIDOS



Podemos observar que los americanos han tenido un consumo preferente por juegos y consolas de Nintendo, seguido de desarrollos propios del país producidos por Microsoft Games y Activision, acompañados por plataformas como Xbox.

Vemos que en Estados Unidos existe un consumo de juegos de género muy variado, liderando las ventas los juegos de Deporte, Plataforma y Shooter, que inclusive se ofertan para distintas consolas.

# POPULARIDAD DE GÉNEROS Y CONSOLAS DE JAPÓN

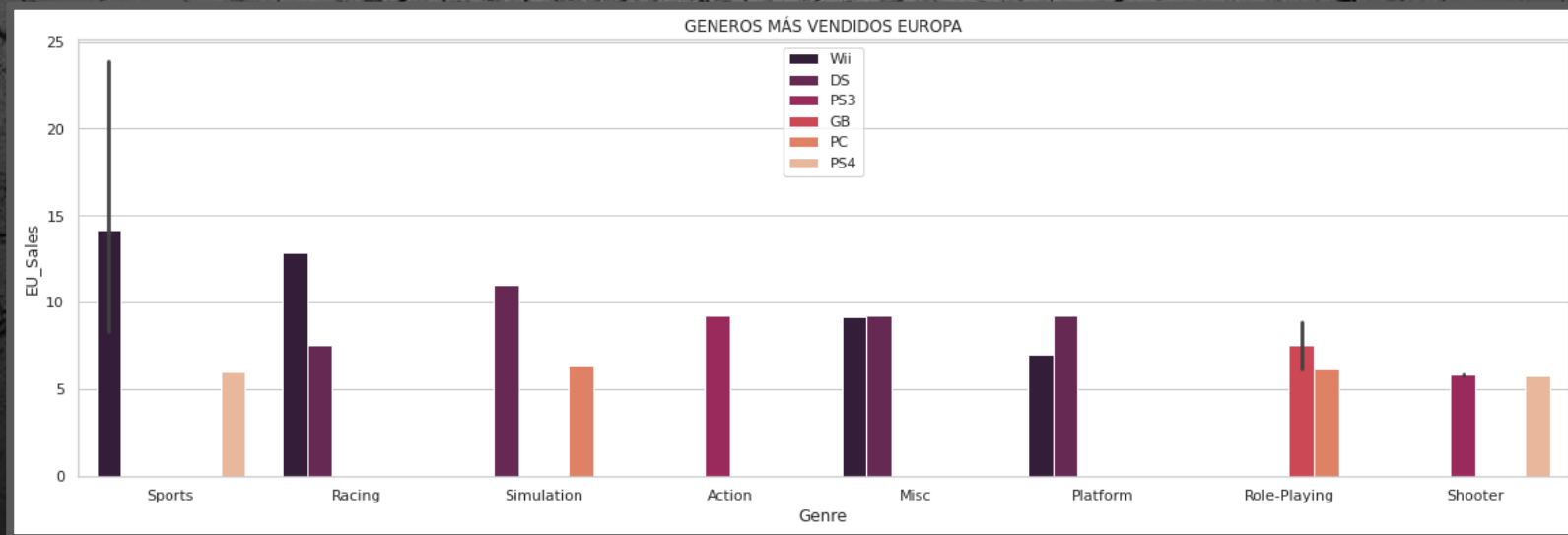


Vemos que en Japón existe un consumo de juegos de género no tan variado, liderando las ventas los juegos de RPG, para todas las consolas disponibles de Nintendo.

Es interesante ver que en Japón tienen una menor diversidad de consumos de géneros, y cómo las plataformas de Estados Unidos han tenido apenas relevancia con ventas de la Playstation Portátil y la Playstation Vita, lo cual deja en evidencia que Japón tiene preferencia por plataformas portátiles.



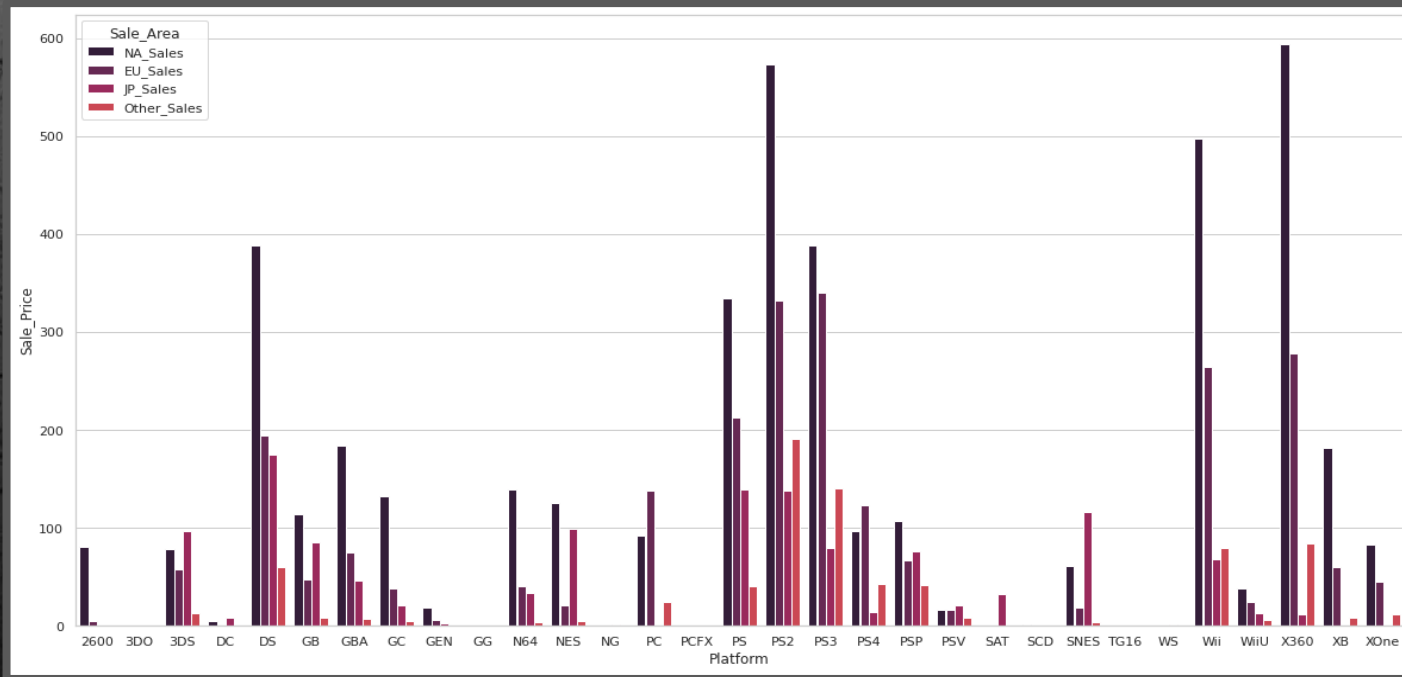
# POPULARIDAD DE GÉNEROS Y CONSOLAS DE EUROPA



Vemos que en Europa existe un consumo bastante parecido al de Estados Unidos, pero con géneros muy variados, liderando las ventas los juegos de Deportes, Carreras, Simulación y Acción, en donde compiten plataformas como Nintendo y Playstation.

Cabe destacar que ha aparecido la PC como plataforma de juegos, permaneciendo como alternativa a las plataformas antes mencionadas, gracias a la posibilidad de permitir jugar juegos de cualquier desarrollador sea de Nintendo, Sony o Microsoft.

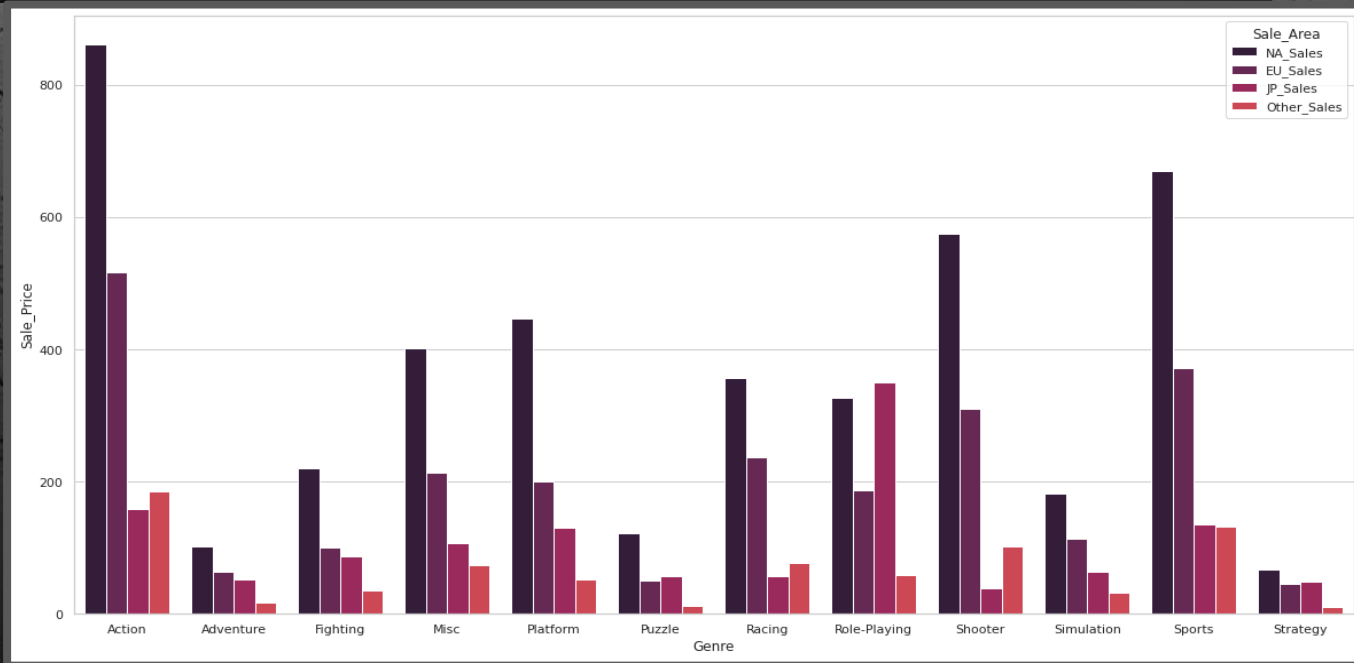
# VENTAS DE JUEGOS POR PLATAFORMA A NIVEL MUNDIAL



Está claro que la Xbox360, la Nintendo Wii, las Playstation 1, 2 y 3 y la Nintendo DS son los líderes del mercado, con Estados Unidos como el máximo consumidor de estas plataformas.

Ninguna de ellas es la consola de última generación, lo que se puede deber a dos principales características, el tiempo que las consolas tienen en el mercado es superior a las consolas de última generación como la PS4, y la accesibilidad económica es superior a estas también.

# VENTAS DE JUEGOS POR GÉNERO A NIVEL MUNDIAL



Podemos ver cómo los juegos RPG (o Role-Playing) son los únicos en los cuales las ventas en Japón superan a Estados Unidos. En los géneros Acción, Shooter y Racing las ventas alrededor del mundo fueron aún mayores que en Japón.

Es notoria la popularidad que tienen juegos de Acción, Deportes y Shooter en el mercado de Estados Unidos y Europa. Si se quisiera desarrollar un nuevo juego con posibilidad de tener un gran mercado para insertarse, sería ideal que pertenezca a alguno de los géneros antes mencionados.



## Insight - Venta de Juegos en Estados Unidos

- Se destaca el género de deportes de la Nintendo Wii junto a Wii Sports.
- Fuerte presencia del mercado local, con la presencia de Microsoft Games y Activision destacables junto a la plataforma Xbox.
- Se ve un consumo de géneros muy variado liderando los juegos de Action, Shooter y Racing.
- Es el mercado con mayor volumen en ventas y, además, lideran los juegos y consolas japonesas en cuanto a consumo.

## Insight - Venta de Juegos en Japón

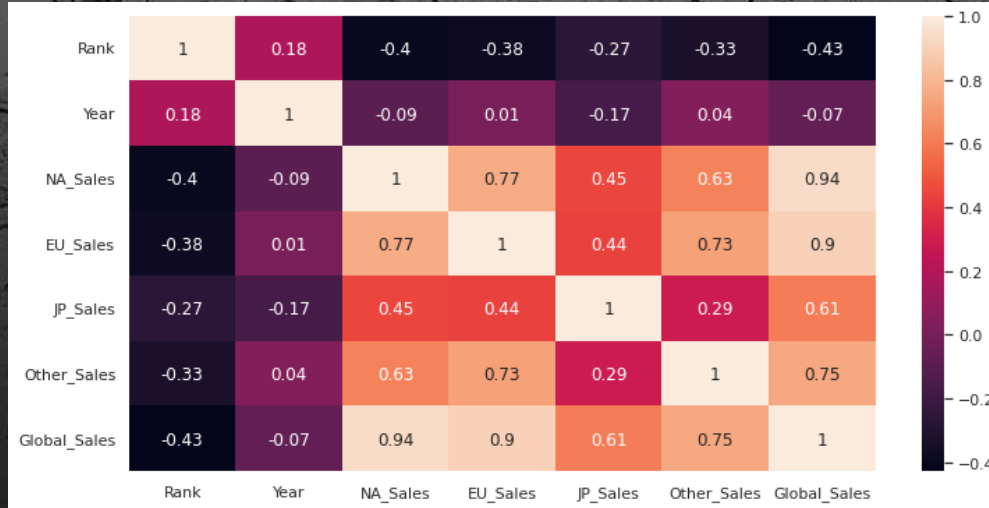
- El récord de ventas se lo lleva el género de Role-Playing o RPG, con los juegos Pokémon Red y Pokemon Blue.
- Se destaca su consumo tanto de géneros, plataformas y consolas locales, producido en mayor proporción por Nintendo.
- Cabe resaltar a la Nintendo DS como consola perseverante, que si bien no supera en ventas a las Nintendo NES y GameBoy, permaneció en el mercado como plataforma para géneros más populares.
- Existen preferencias por consolas portátiles como la PS Vita.
- Tiene un consumo muy regional.

## Insight - Venta de Juegos en Europa

- Se destaca el género de deportes de la Nintendo Wii junto a Wii Sports.
- Su consumo es similar al estadounidense, destacando Nintendo principalmente y luego desarrolladores americanos como Electronic Arts y Activision.
- Se destaca la aparición de la PC, que ha permanecido como alternativa de plataforma de juegos.
- Tiene un consumo diverso de géneros y plataformas, inclinándose a los géneros que se vinculan a la realidad como Sports, Racing y Simulation.
- Es el segundo mercado con mayores ventas.

# Feature Analysis:

## A) Variables numéricas:

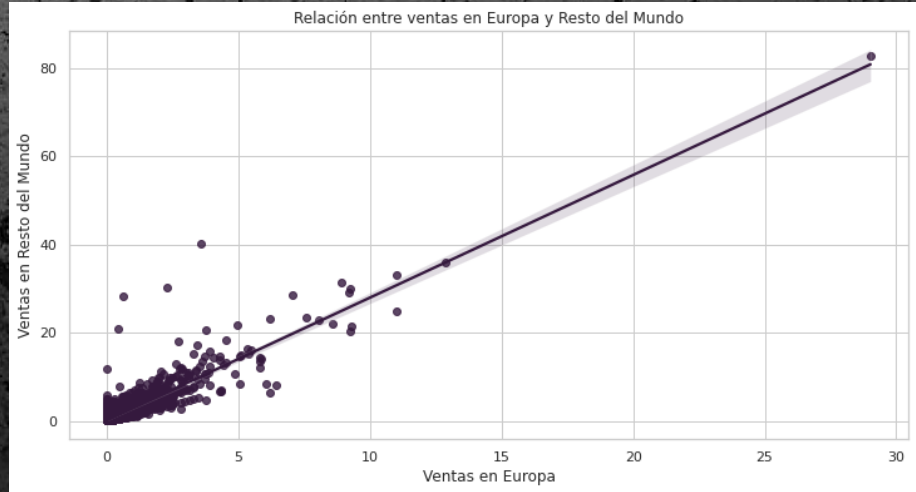
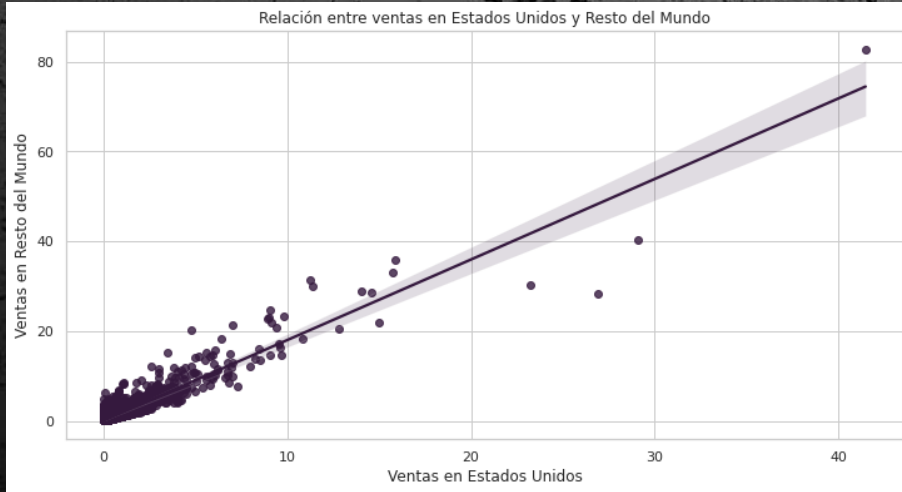


Las ventas en Estados Unidos, Europa y Japón influyen de manera significativa en el resto del mundo, con una correlación de 0.94, 0.90 y 0.61 respectivamente.

Sin embargo, no existe una gran correlación entre las ventas propias de Estados Unidos y Europa con respecto a Japón, seguramente porque los usuarios consumen distintos géneros o plataformas en esas regiones, obteniendo una correlación de 0.45 y 0.44.

# Feature Analysis:

## A) Variables numéricas:



Como en Estados Unidos y Europa existe mucha similitud con respecto a ventas, sólo evaluaremos la correlación de EEUU respecto al Resto del Mundo.

Existe una relación lineal con respecto a la cantidad de ventas en Estados Unidos y Japón, y el Resto del Mundo. Es decir, a mayor cantidad de ventas en una de las 2 regiones de mayor consumo, también crecen las ventas en el Resto del Mundo.



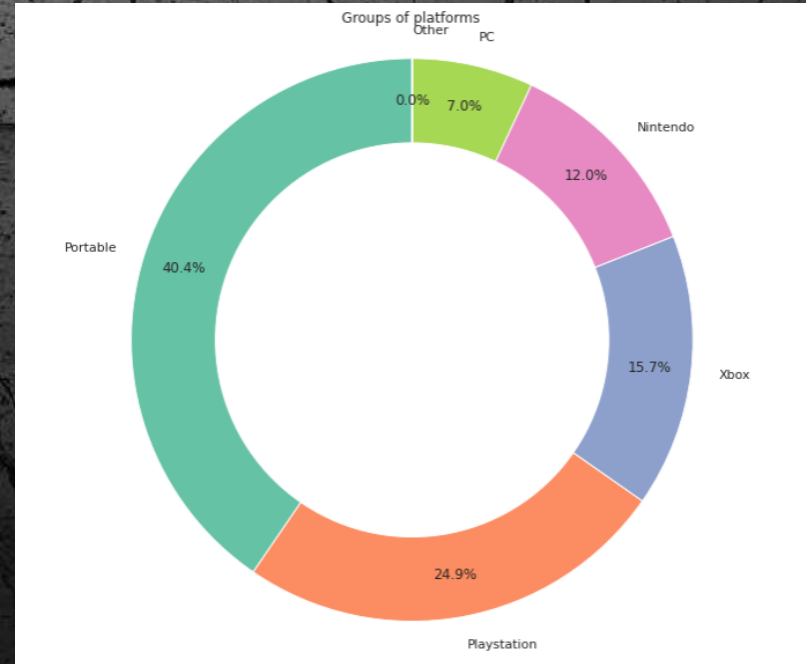
# Feature Analysis:

## B) Variables categóricas:

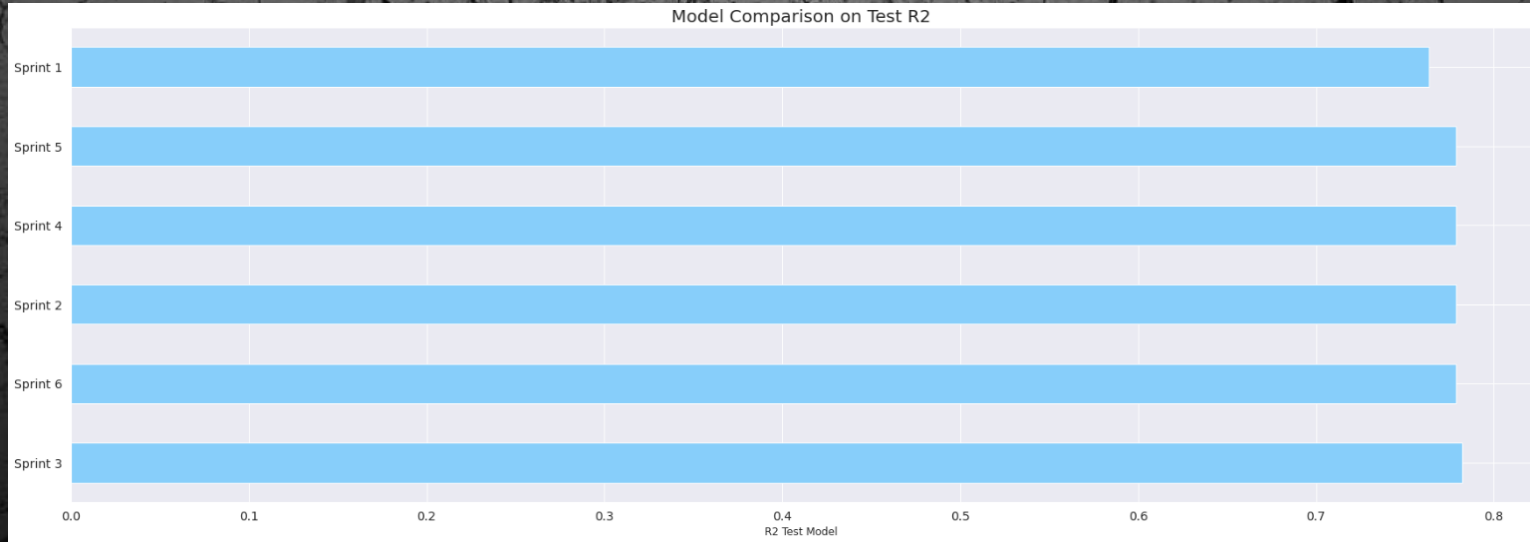
Debido a la gran variedad de consolas existentes en el mercado que resultan poco relevantes, decidimos generalizar dichas consolas en 5 grupos principales que tienen alcance de ventas en todas las regiones:

- Playstation
- Xbox
- PC
- Nintendo
- Portátiles

Que exista una mayor distribución de consolas portátiles no quiere decir que fueron éxito de ventas, sino que posiblemente haya más modelos de consolas en el mercado.



# Feature Engineering: Evaluación de modelo



Necesitamos reducir las features de nuestro dataset a sólo aquellas principales que determinarán qué características tienen los juegos que son récord de ventas. Para ello, realizaremos una serie de "Sprints" con distintos valores de k-features.

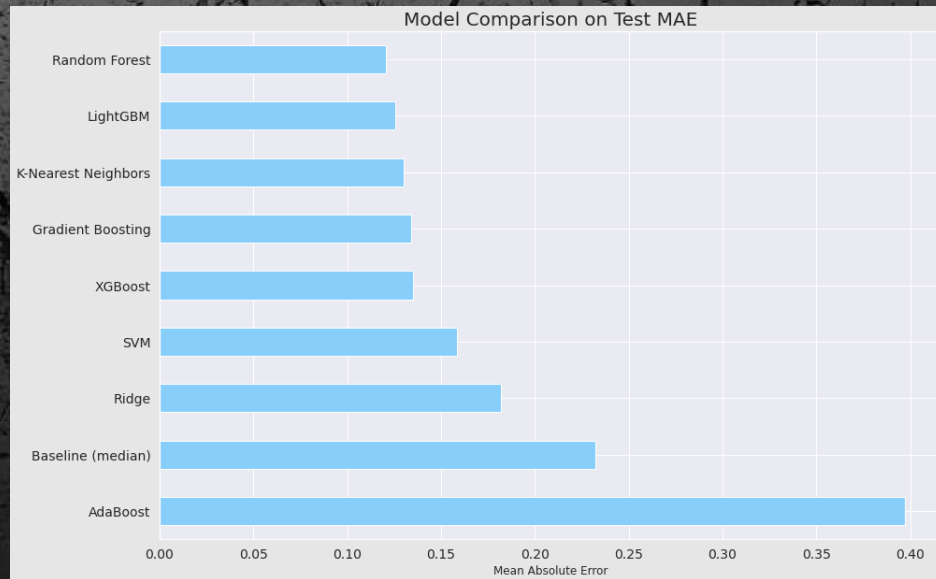
Podemos observar que el Sprint 3, el cual utilizó 7 features y una reducción mediante un Forward Selection presenta un mejor Coeficiente de determinación, y en consecuencia mucho menor Error Absoluto que los demás Sprints.

# Feature Engineering: Evaluación de modelo

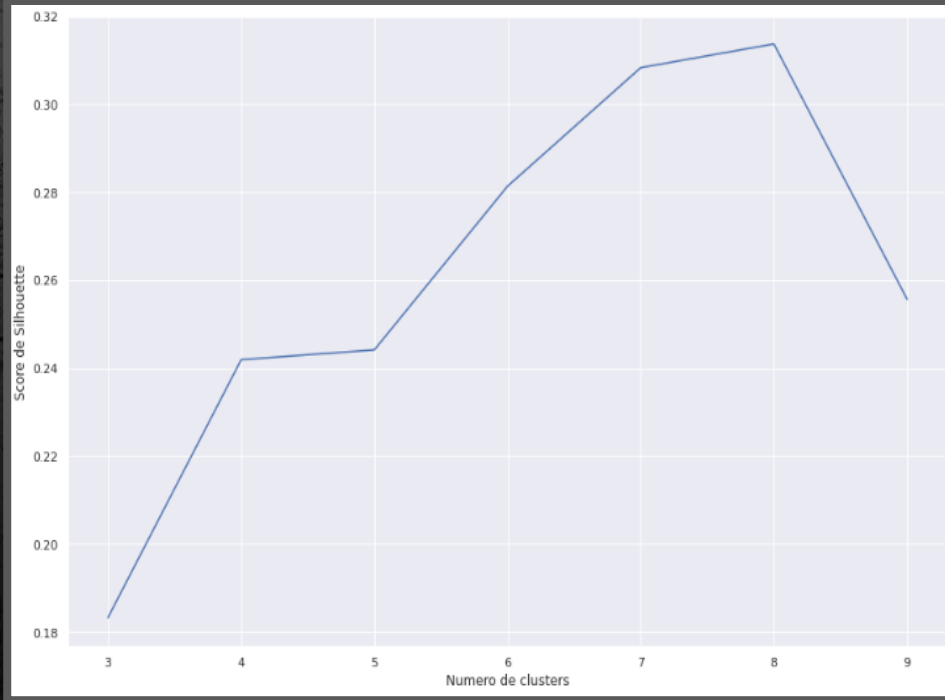
Ya con el dataset reducido a las features principales, compararemos distintos modelos de regresión evaluando aquellos que menos Error Absoluto presenten:

- Support Vector Machine
- Random Forest
- Gradient Boosting
- K-Neighbors
- Ridge
- AdaBoost
- XGBoost
- LightBoost

Decidimos elegir Random Forest, K-Neighbors, Gradient Boosting, XGBoost y SVM para nuestro modelo, los cuales optimizaremos con hiperparámetros y luego evaluaremos cuál es que posee mejor performance.



# Feature Engineering: Clustering

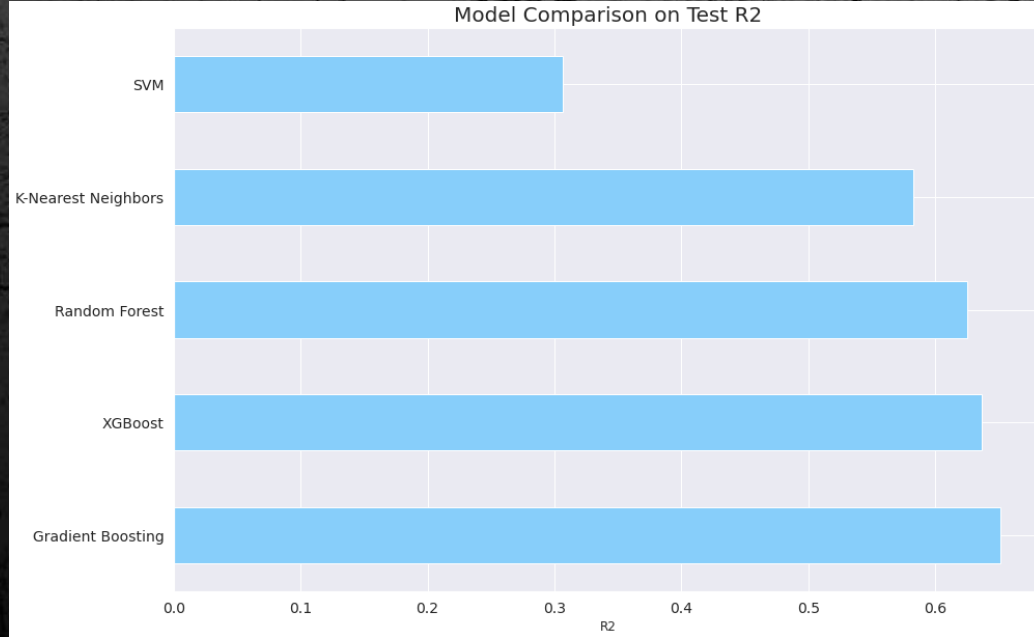


Podemos observar mediante el método de Silhouette que a partir de los 7 clústers la curva queda constante y luego descende, lo cual indica que el número ideal de clústers es de 7.

Utilizaremos dicha cantidad de clústers para la optimización por hiperparámetros de nuestros modelos.



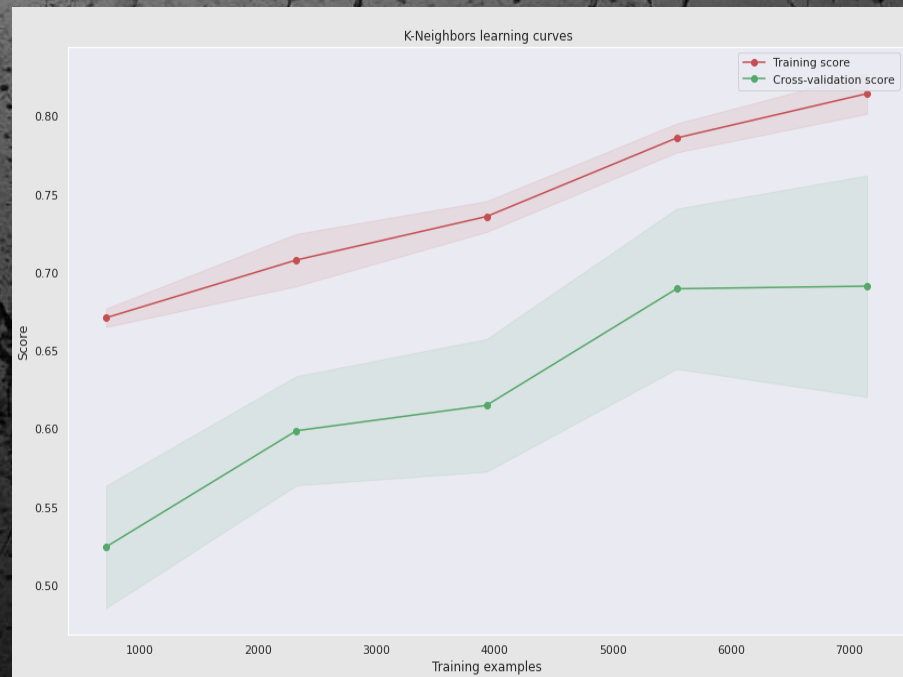
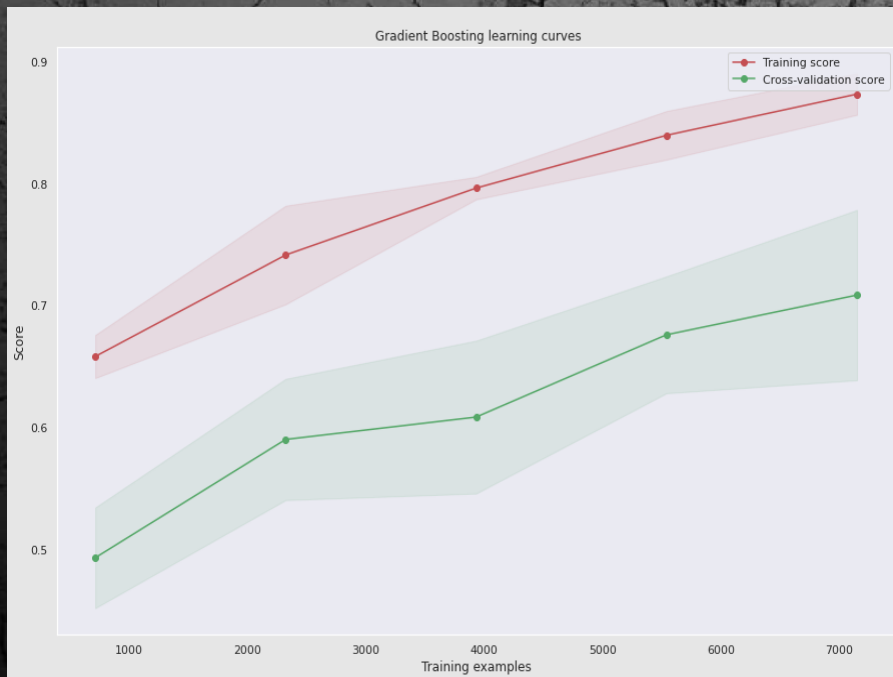
# Feature Engineering: Tuning de hiperparámetros



Realizamos una grilla de hiperparámetros para optimizar los distintos modelos de ML elegidos mediante RandomizedSearch, y luego mediante GridSearch.

Notamos que tanto Random Forest, XGBoost y GradientBoosting presentan los mejores Coeficientes de Determinación.

# Feature Engineering: Verificación de Overfitting



Observamos que el efecto de Overfitting en el entrenamiento y testeo de nuestros datos es menor en los modelos de Gradient Boosting y de K-Nearest Neighbors, visualizando una “sombra” en la curva de entrenamiento de datos, indicando una mayor generalización de los mismos.

# Conclusión final

En conclusión, elegimos el modelo de ensamble **Gradient Boosting** para predecir nuestros datos de ventas de video juegos para cada región, y considerando las principales features que determinan dichas ventas, según lo detectado en el Sprint 3 al realizar la reducción de dimensionalidad:

- Ventas en cada zona: Estados Unidos, Europa, Japón y Resto del Mundo
- Desarrolladores: Ignition Entertainment y Paradox Interactive
- Consola: Playstation
- Género: Si consideramos las principales ventas de los desarrolladores mencionados, deberíamos enfocarnos en juegos de Estrategia y Acción.

A partir de la optimización con hiperparámetros de nuestro modelo, se logra un Coeficiente de Determinación que permite explicar el 65% de los datos, y apenas tener un Error Absoluto Promedio del 12,6%.

Dicho modelo ha funcionado muy bien comparado con los demás modelos evaluados, y en algunos caso ha sido superado levemente en R2 o MAE por otros modelos como XGBoost, pero considerando que queremos evitar el sobreajuste de los datos, decidimos que Gradient Boosting es el modelo que mejor se adapta y responde a nuestros datos.

# Conclusión final

## Ignition Entertainment

- Enfocada en la distribución de juegos para PlayStation Portable (PSP) y Nintendo DS.
- También ha distribuido juegos en Norteamérica y Europa para Xbox 360, PlayStation 2 y 3 y PC.
- Tiene sedes de desarrollo y distribución en Japón y Estados Unidos.
- Los principales géneros desarrollados fueron RPG, Acción y Terror.

## Paradox Interactive

- Enfocada en la distribución de juegos de Estrategia.
- Trabaja principalmente con consolas Xbox que sean compatibles con sistemas de Microsoft, tales como Xbox One.
- Tiene sedes de distribución en Europa y Estados Unidos.
- Los principales géneros desarrollados fueron RPG, Estrategia y Simulación.
- Desde el 2021 se están ampliando a otras plataformas.

Desde un punto de vista personal, creemos haber logrado llegar a resultados similares a los descubiertos en el EDA, lo cual fue muy satisfactorio. Por otro lado, al buscar información sobre los desarrolladores, pudimos darnos cuenta del alcance global que tienen, así como también que nuestro modelo destacó las principales características que presentan los juegos que son récord en ventas globales, y no sólo regionales, algo que no tuvimos en cuenta en nuestro análisis exploratorio anterior, por ejemplo, donde estancan los juegos de RPG, Acción y Simulación.

Fuente:

[https://es.wikipedia.org/wiki/UTV\\_Ignition\\_Entertainment](https://es.wikipedia.org/wiki/UTV_Ignition_Entertainment)

[https://es.wikipedia.org/wiki/Paradox\\_Interactive](https://es.wikipedia.org/wiki/Paradox_Interactive)