

**Materia:** Gestión de Análisis y Diseño de Comercialización (COM145)

**Profesor:** Sarahí Aguilar González

**Fecha de entrega:** 24/05/2022

**Ciclo:** 1222

**Nombre del proyecto:** SteamStats

Miembros del Equipo		
ID	Nombre	Carrera
0219008	Doshner Galland, Ian	TI
0217308	Gonzalez Herrera, Bernardo	TI

Rúbricas				
ID	2-social		7-knowledge	
	D	C	A	JI

**Abstract**—Video games are many things, a means of entertainment, art, product and others. Thanks to their different interpretations, video games have gained fame and popularity exponentially. Thanks to this fame, the way in which they are acquired has diversified, one of the ways in which they can be acquired is through platforms; One of them is Steam. Steam has become one of the largest video game digital distribution platforms in the world, where thousands of video game developers make their creations known. Today there is a wide range of genres, however, not all of them are carried out strategically, that is, despite the fact that popular genre games are in trend, if the developer does not correctly carry out the strategy or the necessary study, it will not be in the bestseller list. That is why it seeks to develop a data analysis and a tool by which relevant information for developers can be consulted in this way, enabling the possibility of creating and launching a game with the greatest possible success.

**Palabras Clave**—Videojuegos, Steam, Dataset, Data Science, Machine Learning.

## I. INTRODUCCIÓN

Los videojuegos son muchas cosas, un medio de entretenimiento, arte, producto y demás. Gracias a sus diferentes interpretaciones los videojuegos se han ganado fama y popularidad exponencialmente. Gracias a esta fama se ha diversificado la manera en la que se adquieren, una de las maneras en las que se pueden adquirir son por medio de plataformas; Una de ellas es Steam. Steam se ha vuelto una de las plataformas de distribución digital de videojuegos más grande del mundo, en donde miles de desarrolladores de videojuegos dan a conocer sus creaciones. Hoy en día existe una gran extensión de géneros sin embargo no todos ellos están realizados de manera estratégica, es decir, a pesar de que juegos de género popular estén en tendencia si el desarrollador no realiza correctamente la estrategia o el estudio necesario este no estará en la lista de los más vendidos. Es por ello que se busca desarrollar un análisis de datos y una herramienta por el cuál se podrá consultar información relevante para los desarrolladores de esta manera habilitando la posibilidad de crear y lanzar un juego con el mayor éxito posible.

## Antecedentes

Nuestro proyecto se ha dividido en 3 Partes:

- Data Science
- Databases
- Machine Learning

### Data Science

Esta fue la primera parte de nuestro proyecto, en la cual pudimos observar todos nuestros datos puros de manera que pudimos sacar toda la información que necesitábamos o no. En esta parte fue algo tedioso, ya que muchas veces la información llegaba a ser innecesaria o incluso incompleta. En esta parte pudimos determinar que hay ciertos elementos que nos pueden servir para más adelante hacer predicciones que podemos venderles a los usuarios de paga de nuestra página.

### Databases

Esta fue la segunda fase de nuestro proyecto, aquí hicimos toda la infraestructura de la página, tanto con sus roles como con su contenido. Al principio tuvimos que normalizar todos los datos de manera que los soportará una bases de datos, después empezamos a hacer la pagina la cual (por ahora) consta de 3 roles:

- **Administrador:** Que se encarga de que todo esté en orden dentro de la página, que los roles de gente de paga y no paga estén bien asignados y en caso que sea necesario; Manipular los datos a mano pero eso sería algo a lo que no se quiere llegar.
- **Usuario de Paga:** Estos serán nuestros clientes a los cuales les brindaremos información necesaria para que puedan darse una idea que poder desarrollar o apoyar sus ideas de desarrollo con lo que dice la información de steam.
- **Usuario de No Paga:** Son usuarios potencialmente clientes que les damos una pequeña visualización sobre lo que pueden obtener si pagan los servicios de la aplicación. Este usuario más que es tratar de convencer al cliente de que lo que la información que ofrecemos vale la pena pagar.

### Machine Learning

Ahora que estamos en la tercera fase de nuestro programa, llegamos a la siguiente conclusión. Todos los datos que hemos estado utilizando y limpiando les podremos dar un enfoque económico con el apartado de Ciencia de datos. En nuestra página al tener ahora una membresía de pago, queremos que nuestros clientes tengan una herramienta la cual puedan con el “género” poder determinar solo con esos datos una calificación estimada que podría tener ese futuro juego.

También, queremos crear una herramienta que al introducir el nombre de un juego, pueda recomendar juegos del mismo estilo, ya que esto le ayudará a los desarrolladores a saber si un juego en el que se inspiraron para crear su propio juego, cuáles serían los demás títulos con los que puede competir.

## II. Desarrollo

El proyecto inició con un análisis extenso de varios datasets, unos relacionados con steam, otros de videojuegos en general. El enfoque del análisis es la creación e identificación de herramientas y datos para facilitar el desarrollo de videojuegos. Con los datos identificados se decidió desarrollar una plataforma para los desarrolladores de videojuegos en la cual se pudiera consultar a través de inputs que requerimientos son necesarios para desarrollar un juego exitoso.

El proyecto para realizar el modelo de machine learning se encaminó a encajar con estos antecedentes, se utilizaron los mismos datasets y se encaminó con la visión de ayudar a los desarrolladores, realizar un simulador en la plataforma para realizar predicciones sobre el juego a lanzar, en el que se muestre el posible User Score a conseguir y las ventas en ciertas regiones.

### 1. Problemática

En el mundo del desarrollo de los videojuegos existen las tendencias que pueden marcar el rumbo del mercado. Por ejemplo, en ciertos años se vende más un género de videojuegos que otros; O ciertos juegos viejos retoman popularidad dependiendo de las redes sociales. Hoy en día eso para un desarrollador es un poco frustrante, ya que no sabe que si su juego que tiene en mente lo desarrolla de cierta manera puede que tenga ganancias o no.

Y otra cuestión es que Steam, que es la plataforma número uno en videojuegos, tiene ciertos algoritmos para recomendar juegos, el algoritmo se basa en "Género" y "tags". Dependiendo de qué género y que tags tenga el juego, el algoritmo de steam podrá recomendar más o menos tu juego y también tiene que ver mucho con las tendencias de juegos que se tenga en redes sociales.

Por lo tanto, un desarrollador de videojuegos se enfrenta a varias desventajas y/o cosas que tiene en cuenta para desarrollar un juego.

### 2. Solución actual / Visión General

La solución consiste en analizar un grupo de datasets y extraer datos importantes, aplicando un poco de análisis de datos. Estos, pasarán por un filtro con el objetivo de simplificar el proceso y modelado.

Entrenar un modelo con los datos filtrados y generar predicciones para evaluar la precisión de dicho modelo.

Se complementará la solución con visualizaciones de los datos para una comprensión más amigable.

### 3. Limitaciones actuales de la solución

Al empezar el proyecto una de las principales limitaciones fue la limpieza de datos, los datasets que tomamos tienen mucha información útil que nos podría servir para el proyecto, pero tuvimos que hacer una pequeña utilizando técnicas de data science para poder obtener lo que necesitábamos.

Una vez eso, nos dimos cuenta que en la manera de poder hacer el modelo lógico de la base de datos había un problema. Ya que no se concretaban los datos de ninguna manera, así que tuvimos que encontrar un argumento que pudiera estar en cualquiera de las tablas y que no afectará la estructura de toda la base de datos. Una solución fue utilizar el id único de las aplicaciones en steam, así podríamos conectar todo mediante los ids. Pero al momento de hacer los queries, tuvimos que limpiar la base de datos otra vez, ya que al agregar una columna en cada tabla, muchos datos (como vienen de un csv), se empezaron a mover y tuvimos que arreglar esos problemas.

### 4. Propósito, uso y alcance

El propósito de este proyecto fue crear un modelo de machine learning para complementar los antecedentes mencionados. El propósito de dicho modelo es predecir/simular el futuro de un juego en desarrollo, proporcionando los datos de "stars" o "ratings" de usuario que podría recibir dicho juego y la posibilidad de aproximar cuantas millones de unidades se pueden vender en ciertas zonas.

### 5. Covid

Los datos que pudimos recopilar entre las fechas de 2020-2022 que fue la pandemia, pudimos ver que la salida de juegos en steam no aumentaron como tal, sino que disminuyó. Pero la cantidad de usuarios que entraron a la plataforma creció de gran manera.

De tal manera, las reviews de juegos, número de ventas creció de manera drástica. Creemos que este tiempo de pandemia benefició mucho a los desarrolladores de juegos, ya que la gente empezó a volver a jugar ya que en sus tiempos libres pudieron retomar esta actividad.

A pesar de todo lo anterior, no creció de manera positiva la cantidad de juegos subidos a la plataforma. Ya que Steam cuenta con una de las comunidades más grandes de desarrolladores Indies y 3th party en el mundo de los videojuegos. Pero como normalmente estos desarrolladores son empresas pequeñas o incluso un grupo de personas

haciendo videojuegos. La pandemia del Covid-19 seguramente les afectó en el desarrollo de sus juegos y creemos que en un futuro se podrán reponer para poder seguir trabajando en sus proyectos pendientes. Claro no se puede negar que los videojuegos pueden afectar de manera negativa a las personas, los videojuegos pueden llegar a ser muy entretenidos al punto de simular una adicción provocando que el usuario no se cuide o descuide su bienestar, no comiendo, ni durmiendo lo suficiente, descuidando la actividad física pero los estudios apoyando la negatividad de los videojuegos son muy escasos y poco verídicos.

## 6. Desarrollo para la industria

Creemos que todo esto puede beneficiar de un lado positivo a la industria de los videojuegos, ya que podremos determinar los gustos de los consumidores de una manera fácil de interpretar. Tiene mucho potencial a largo plazo, ya que los desarrolladores podrán ver que si su juego tiene ciertos géneros y tags, podrán hacer un análisis rápido para saber si el juego pueda ser bueno o pueda ser malo. Claro, si el juego en si no es un juego que sea “divertido” o “bueno” no podras tener la calificación deseada. Este modelo solo se basa en que si tu juego es bueno o en realidad es un juego que potencialmente le puede gustar a la gente y que puedas tener tu “fandom” como desarrollador; Solo te servirá para poder posicionarlo bien en steam, ya que si no lo logras posicionar, por más bueno que sea el juego el mismo algoritmo de la pagina no lo recomendaría y podría pasar que un muy buen juego no se venda o no tenga las notas deseadas porque al principio no pusiste bien los géneros y tags dentro de la aplicación para poder que el juego se posicione donde se merece.

## III.Implementación

### Preparación de datos

Los datos se obtuvieron a través del API de Steam y el uso de Kaggle. Se utilizaron herramientas de big data y de cloud computing para la extracción, reducción y análisis de datos. Posteriormente se realizó un modelo de base de datos que se adecuara con los datos que se extrajeron. Finalmente se realizará una aplicación web con el fin de unificar el análisis de datos, el modelado y el machine learning para que esto sea un producto tangible y presentable.

## Diagrama de Bases de Datos

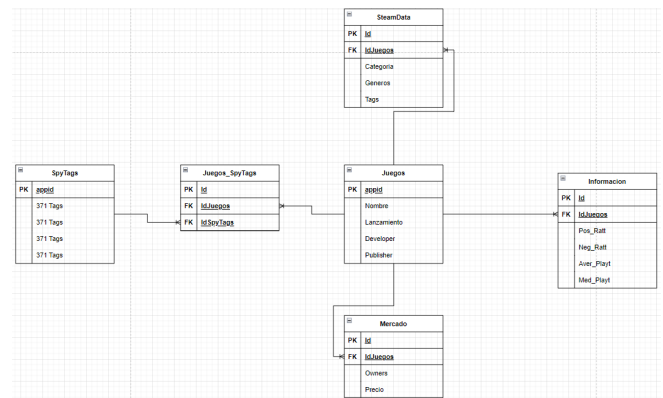


Figura IA1: DER

## Diseño de Base de Datos

Tabla	Propiedad	Tipo de Dato	Permite Nulos	Características
Juegos	Appid	Varchar(20)	NO	
Juegos	Nombre	Varchar(250)	NO	
Juegos	Lanzamiento	date	SI	
Juegos	Developer	Varchar(250)	NO	
Juegos	Publisher	Varchar(250)	NO	
Informacion	Id	Int	NO	Autoincremental
Informacion	idJuegos	Varchar(20)	NO	Referencia a Juegos
Informacion	Pos_Ratt	Int	NO	
Informacion	Neg_Ratt	Int	NO	
Informacion	Aver_Playt	Int	NO	
Informacion	Med_Playt	Int	NO	
Mercado	Id	Int	NO	Autoincremental
Mercado	idJuegos	Varchar(20)	NO	Referencia a Juegos
Mercado	Owners	Varchar(128)	NO	
Mercado	Precio	Money	NO	
SteamData	Id	Int	NO	Autoincremental
SteamData	idJuegos	Varchar(20)	NO	Referencia a Juegos
SteamData	Categoria	Varchar(300)	NO	
SteamData	Genero	Varchar(250)	NO	
SteamData	Tags	Varchar(150)	NO	
SpyTag	Appid	Varchar(20)	NO	
SpyTag	1980s	Varchar(50)	SI	
SpyTag	1990	Varchar(50)	SI	
SpyTag	2 5d	Varchar(50)	SI	
... 366 datos	...	Varchar(50)	SI	
SpyTag	Wresteling	Varchar(50)	SI	
SpyTag	Zombies	Varchar(50)	SI	
SpyTag	E_sports	Varchar(50)	SI	
Juegos_SpyTags	Id	Int	NO	Autoincremental
Juegos_SpyTags	idJuegos	Varchar(20)	NO	Referencia a Juegos
Juegos_SpyTags	IdSpyTags	Varchar(20)	NO	Referencia a SpyTags

Figura IA2: Tablas

## IV.Experimentos/Pruebas

Durante el desarrollo de las pruebas obtuvimos diversos resultados que ayudan al usuario a tener un análisis más certero de que videojuegos y géneros han impactado más actualmente.

En esta primera gráfica presentan los años en los que Steam ha logrado vender más videojuegos. En la misma se ve cómo a partir del año 2012 la venta de videojuegos ha tenido un crecimiento rápido.

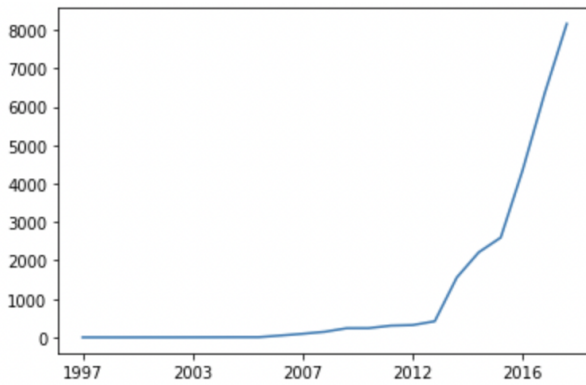


Figura PA1: ¿Cuál es el año en el que se publicaron más juegos?

De igual manera se buscó obtener que desarrolladores han tenido mayor impacto dentro de la plataforma Steam, ayudando al usuario a observar las diferentes metodologías que ocupa cada desarrollador. Acorde a estos resultados el usuario puede obtener una base de por qué estos se encuentran en los primeros 10 desarrolladores basándose en su género a desarrollar, videojuegos más populares, visión de la empresa, entre muchos otros factores.

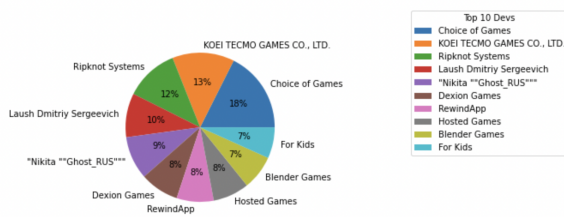


Figura PA2: ¿Qué developer saca más juegos?

Por último se decidió proporcionarle al usuario los tag más buscados, en la gráfica podemos observar con qué frecuencia son buscados estos tags. Esta gráfica tiene como objetivo mostrar los géneros o modos de juego más buscados en la actualidad. Como podemos observar el género más buscados es un juego casual de acción por lo tanto el usuario se puede guiar a partir de ello para definir su género de juego. En base a esta misma el desarrollador puede observar que lo mejor no sería desarrollar un juego de tipo puzzle sino que uno de acción tendría más éxito en el mercado.

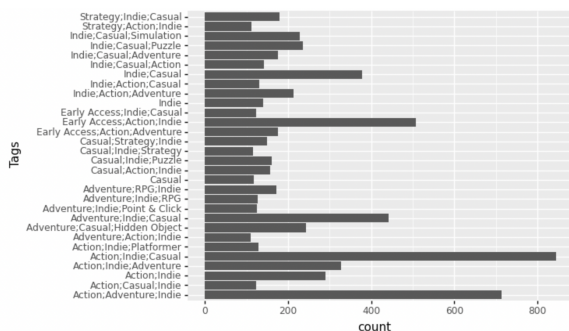


Figura PA3: ¿Qué spy\_tags son los más usados?

## V.Conclusiones

La gran escala de Steam, tanto por número de juegos y por número de usuarios, revela la enorme diversidad de comportamiento entre los jugadores. Esta diversidad de conductas se manifiesta como un espectro continuo donde se encarnan todas las conductas posibles. La mayoría de los jugadores exhiben comportamientos modestos asociados con los juegos casuales, con valores atípicos que tienen un comportamiento más extremo o inusual. Encontramos una variedad de fuertes correlaciones que indican que las redes sociales dan forma a cómo juegan los usuarios. Los conjuntos de datos combinados en un solo conjunto de datos por más complejo que se vuelva se convierte en una fuente completa de datos la cual ofrece numerosas posibilidades de investigación de seguimiento para recopilar datos cualitativos de usuarios individuales y correlacionarlos con sus hábitos de juego cuantificados.

Creemos que hemos podido llegar a un resultado satisfactorio, ya que si pudimos predecir un resultado parecido con las críticas que de los usuarios en steam. El modelo se apoya mucho sobre los tags de los juegos y esto puede ayudar a los desarrolladores a saber que tags pueden promover en su juego para que pueda vender más o le llame la atención a los futuros compradores de sus juegos. Por la parte del recomendador de juegos, también creemos que hemos llegado a un resultado satisfactorio, ya que al insertar el nombre de un juego a base de sus tags y de su mismo nombre te recomienda juegos del mismo estilo. Creemos que también es una herramienta que cumple con su función y podría ayudar a saber por dónde van los gustos generales del público de ese tipo de juegos.

## REFERENCIAS

[1]Data. (2022). GitHub.

<https://github.com/FoxuF/DataDataData>

[2]GAC. (2022). GitHub.

<https://github.com/FoxuF/GAC-Project>

[3]G.H. (2022). BigData. GitHub.

<https://github.com/FoxuF/BigData>

[4]Kaggle. (2020). Kaggle.

<https://www.kaggle.com/datasets/nikdavis/>

steam-store-raw