UNIVERSIDAD PANAMERICANA

UNIVERSIDAD PANAMERICANA
Campus México

Class: Marketing Analysis and Design Management (COM145)

Profesor: Sarahí Aguilar González

Date: 24/05/2022

Cicle: 1222

Project Name: SteamStats

Members		
ID	**Name**	**Carreer**
0219008	Doshner Galland, Ian	TI
0217308	Gonzalez Herrera, Bernardo	TI

Rubrics				
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.

Key Words—*Videojuegos, Steam, Dataset, Data Science, Machine Learning.*

I. INTRODUCTIONN

Video games are many things, a medium of entertainment, art, product and so on. 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 are through platforms; One of them is Steam it has become one of the largest digital video game distribution platforms in the world, where thousands of video game developers publicize their creations. Nowadays there is a large extension of genres however not all of them are made strategically, that is, despite the fact that popular genre games are in trend if the developer does not perform correctly the strategy or study needed will not be on the bestseller list. That is why it seeks to develop a data analysis and a tool by which relevant information can be consulted for developers in this way enabling the possibility of creating and launching a game with the greatest possible success.

Antecedentes

Our project has been divided into 3 Parts:

- Data Science
- Databases
- Machine Learning

Data Science

This was the first part of our project, in which we were able to observe all our pure data so that we could get all the information we needed or not.
In this part it was somewhat tedious, since many times the information became unnecessary or even incomplete.
In this part we were able to determine that there are certain elements that can help us to later make predictions that we can sell to the paid users of our page.

Databases

This was the second phase of our project, here we did all the infrastructure of the page, both with its roles and with its content. At first we had to normalize all the data so that it will be supported by a database, then we started making the page which (for now) consists of 3 roles:

- Administrator: That it takes care that everything is in order within the page, that the roles of paying and non-paying people are well assigned and if necessary; Manipulate the data by hand but that would be something you don't want to get to.
- Usuario de Paga: These will be our customers to whom we will provide necessary information so that they can give themselves an idea that they can develop or support their development ideas with what the steam information says.
- Usuario de No Paga: They are potentially customer users who give them a little insight into what they can get if they pay for the app's services. This user more than is trying to convince the customer that what the information we offer is worth paying for..

Machine Learning

Now that we are in the third phase of our program, we come to the following conclusion. All the data that we have been using and cleaning can give them an economic approach with the Data Science section. On our page, having a paid membership now, we want our customers to have a tool which they can with the "gender" be able to determine only with that data an estimated rating that that future game could have. Also, we want to create a tool that by entering the name of a game, can recommend games of the same style, since this will help developers to know if a game they were inspired to create their own game, what would be the other titles with which it can compete.

II. Desarrollo

The project began with an extensive analysis of several datasets, some related to steam, others of video games in general. The focus of the analysis is the creation and

identification of tools and data to facilitate the development of video games. With the data identified, it was decided to develop a platform for video game developers in which it could be consulted through inputs that requirements are necessary to develop a successful game.

The project to make the machine learning model was aimed at fitting with this background, the same datasets were used and was directed with the vision of helping developers, making a simulator on the platform to make predictions about the game to be launched, in which the possible User Score to be achieved and sales in certain regions are shown.

1. Problem

In the world of video game development there are trends that can set the course of the market. For example, in certain years one genre of video games is sold more than others; Or certain old games regain popularity depending on social networks. Nowadays that for a developer is a bit frustrating, since he does not know that if his game he has in mind he develops it in a certain way he may have profits or not. And another issue is that Steam, which is the number one platform in video games, has certain algorithms to recommend games, the algorithm is based on "Genre" and "tags". Depending on what genre and tags the game has, the steam algorithm can more or less recommend your game and also has a lot to do with the trends of games that you have in social networks.

Therefore, a video game developer faces several disadvantages and/or things that he takes into account to develop a game.

2. Solución actual / Visión General

The solution is to analyze a group of datasets and extract important data, applying a little data analysis. These will go through a filter with the aim of simplifying the process and modeling.

Train a model with the filtered data and generate predictions to evaluate the accuracy of that model.

The solution will be complemented with visualizations of the data for a friendlier understanding.

3. Limitations

At the beginning of the project one of the main limitations was the data cleaning, the datasets we took have a lot of useful information that could be useful for the project, but we had to make a small one using data science techniques to be able to obtain what we needed.
Once that, we realized that in the way we could make the logical model of the database there was a problem.
Since the data was not specified in any way, so we had to find an argument that could be in any of the tables and that will not affect the structure of the entire database. One solution was to use the unique id of the applications on steam, so we could connect everything through the ids. But at the time of making the querys, we had to clean the database again, since by adding a column in each table, a lot of data (as it comes from a csv), began to move and we had to fix those problems.

4. Porpuse Use and reach

The purpose of this project was to create a machine learning model to complement the aforementioned background. The purpose of this model is to predict/simulate the future of a game in development, providing the data of "stars" or "ratings" of user that could receive said game and the possibility of approximating how many millions of units can be sold in certain areas.

5. Covid

The data that we were able to collect between the dates of 2020-2022 that was the pandemic, we could see that the output of games on steam did not increase as such, but decreased. But the number of users who entered the platform grew greatly.
In such a way, the reviews of games, number of sales grew drastically. We believe that this time of pandemic benefited game developers a lot, since people began to play again since in their free time they were able to resume this activity.
Despite all the above, the number of games uploaded to the platform did not grow positively. Since Steam has one of the largest communities of Indies and 3rd party developers in the world of video games. But as usually these developers are small companies or even a group of people making video games. The Covid-19 pandemic surely affected them in the development of their games and we believe that in the future they can be replenished to be able to continue working on their pending projects.
Of course there is no denying that video games can negatively affect people, video games can be very entertaining to the point of simulating an addiction causing the user not to take care of themselves or neglect their well-being, not eating, or sleeping enough, neglecting physical activity but studies supporting the negativity of video games are very scarce and not very true.

6. Industry

We believe that all this can benefit the video game industry on a positive side, since we will be able to determine the tastes of consumers in an easy way to interpret. It has a lot of long-term potential, since the developers will be able to

see that if their game has certain genres and tags, they will be able to do a quick analysis to know if the game can be good or it can be bad.
Of course, if the game itself is not a game that is "fun" or "good" you will not be able to have the desired rating. This model is only based on whether your game is good or actually a game that people can potentially like and that you can have your "fandom" as a developer; It will only serve you to be able to position it well on steam, since if you do not manage to position it, no matter how good the game is, the same algorithm of the page would not recommend it and it could happen that a very good game does not sell or does not have the desired notes because at the beginning you did not put the genres and tags well within the application to be able to position the game where it deserves.

III.Implementation

Data Cooking

The data was obtained through the Steam API and the use of Kaggle. Big data and cloud computing tools were used for data extraction, reduction and analysis. Subsequently, a database model was made that was adapted with the data that was extracted. Finally, a web application will be made in order to unify data analysis, modeling and machine learning so that this is a tangible and presentable product.

ERD

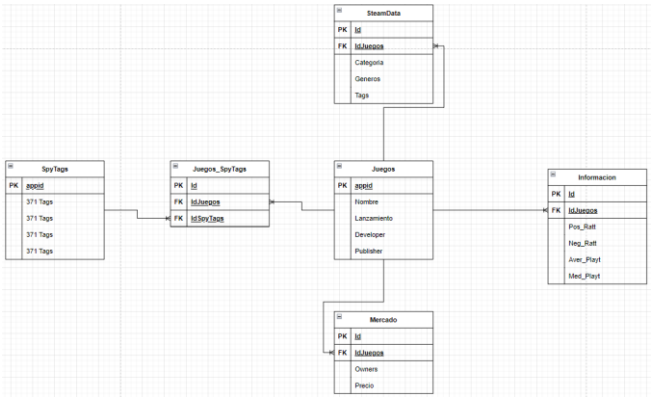

Figure IA1: ERD

Data Base

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: Tables

IV.Testing

During the development of the tests we obtained various results that help the user to have a more accurate analysis of which video games and genres have impacted more currently.

In this first graphic they present the years in which Steam has managed to sell more video games. In it you can see how from 2012 the sale of video games has had a rapid growth.

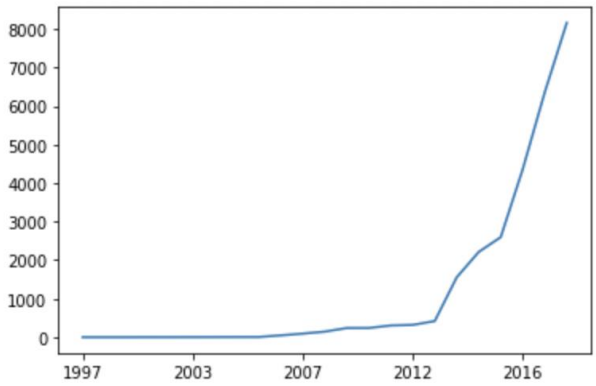

Figure PA1: What is the year in which more games were published?

In the same way, it was sought to obtain which developers have had the greatest impact within the Steam platform, helping the user to observe the different methodologies that each developer occupies. According to these results the user can obtain a basis of why these are in the first 10 developers based on their genre to be developed, most popular video games, vision of the company, among many other factors.

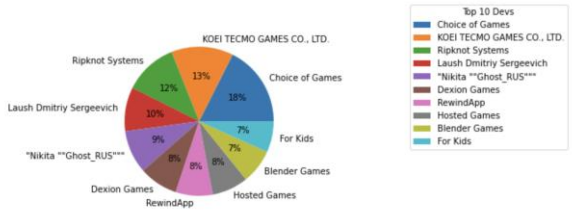

Figure PA2: ¿Which developer brings out the most games?

Finally it was decided to provide the user with the most searched tags, in the graph we can see how often these tags are searched. This graphic aims to show the most sought-after genres or game modes today. As we can see the most sought after genre is a casual action game therefore the user can be guided from it to define their game genre. Based on this same the developer can observe that the best thing would not be to develop a puzzle type game but that an action one would be more successful in the market.

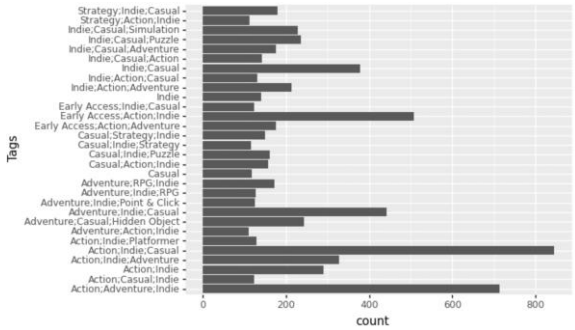

Figure PA3: ¿Qué spy_tags are used the most?

V. Conclusions

The sheer scale of Steam, both by number of games and by number of users, reveals the enormous diversity of behavior among players. This diversity of behaviors manifests itself as a continuous spectrum where all possible behaviors are embodied. Most gamers exhibit modest behaviors associated with casual gaming, with outliers having more extreme or unusual behavior. We found a variety of strong correlations indicating that social media shapes how users play. Datasets combined into a single dataset no matter how complex it becomes becomes a complete source of data which offers numerous follow-up research possibilities to collect qualitative data from individual users and correlate it with their quantified gambling habits.

We believe that we have been able to reach a satisfactory result, since we could predict a similar result with the criticisms that users on steam. The model leans heavily on game tags and this can help developers know which tags can promote in your game so that you can sell more or draw the attention of future buyers of your games.

On the part of the game recommender, we also believe that we have reached a satisfactory result, since by inserting the name of a game based on its tags and its same name it recommends games of the same style. We believe that it is also a tool that fulfills its function and could help to know where the general tastes of the public of such games are going.

REFERENCES

[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