



# **Análisis de Juegos de la Plataforma Steam**

**Presentación de Entrega Final – Coder House**

**Profesor:** Dario Ceballos

**Tutor:** Martín Segovia

**Comisión:** 32645

**Año:** 2023

Última actualización 10/03/2023  
**Asdrubal Alejandro Vilera Vivas**

# Índice

Introducción.	Página 3
Descripción de la temática de los datos.	Página 3
Alcance	Página 4
Tablas Originales y Organización	Página 4
Diagrama Entidad-Relación	Página 7
Transformación de Tablas	Página 8
Medidas Calculadas	Página 10
Análisis Funcional del Tablero de Power BI	Página 12
- Diagrama Entidad-Relación de Power BI	Página 14
- Análisis Funcional por Solapa	Página 15
Conclusiones	Página 24

## Introducción

Este proyecto está orientado a mostrar estadísticas y valoraciones de los juegos de la plataforma de Steam desarrollados y activos hasta 2022. Se procederá a analizar el promedio de calificaciones obtenidas por juegos creados por los diversos desarrolladores en dicha plataforma, así como también su relación con el género, popularidad (recomendaciones) y costo.

## Descripción de la temática de los datos

Esta base de datos cuenta con una extensa lista de videojuegos, indicando en cada uno su ID en plataforma, valoración, recomendaciones, fecha de lanzamiento, género, categoría, jugabilidad, sistemas operativos compatibles y muchos otros datos que ayudaran con los análisis correspondientes.

Cada uno de estos videojuegos se encuentra en la plataforma Steam, creada por la empresa Valve en el año 2003.

### **Objetivos:**

- Implementación de un Dashboard interactivo que enseñe los juegos mejor valorados, clasificándolos por géneros, teniendo en cuenta la relación precio-calidad.
- Evaluar el comportamiento general de usuarios en cuanto a la utilización de la plataforma, de esta manera saber cuándo implementar los cambios más importantes y estrategias afines al resto de objetivos.
- Teniendo en cuenta los juegos con mayor cantidad de reseñas negativas o positivas, ayudará para establecer estrategias de la plataforma en cuanto a su base de juegos.
- Establecer un patrón con carencias y fortalezas en cuanto a nuevos proyectos o proveedores que deseen ofrecer sus juegos en la plataforma.

### **Tipo de análisis:**

- ✓ **Descriptivo:** análisis de los géneros más populares en la plataforma.
- ✓ **Diagnóstico:** implementación de un Dashboard interactivo que muestre los juegos mejor valorados, clasificándolos por géneros, teniendo en cuenta la relación precio-calidad.
- ✓ **Predictivo:** se crea un modelo predictivo de la cantidad de recomendaciones de los nuevos lanzamientos, para así identificar cuáles clasificaciones de juegos son más demandadas por la comunidad de la plataforma.
- ✓ **Prescriptivo:** con el análisis de los hechos y el modelo predictivo se puede establecer un patrón de fracasos de los proveedores de la plataforma y así formar una estrategia de ofrecimiento de nuevos productos o proveedores.

## **Alcance**

Se realizará el análisis del proyecto en un nivel táctico, el cual permita elaborar planes operativos y generar estrategias a futuro, que inspiren y orienten al cumplimiento de los objetivos generales y específicos de la empresa a mediano y largo plazo.

## **Tablas Originales y Organización**

### **Juegos:**

Indica toda la información general de cada juego publicado en la plataforma. Esta tabla contiene como llave foránea la App ID y llave primaria la Header Image.

### **Clasificación:**

Son columnas dentro del CSV de la Tabla 1, pero que contienen información relevante para clasificar cada juego de la plataforma con su género y categoría. Utilizaría como llave primaria la App ID. Se separó en un archivo CSV diferente de la tabla original.

### **Desarrolladores:**

Igualmente son columnas dentro del CSV de la Tabla 1, simplemente indica quien es el encargado de gestionar el desarrollo y publicación de cada juego.

Utilizaría como llave primaria la App ID. Se separó en un archivo CSV diferente de la tabla original.

### **Valoraciones:**

Columnas en la Tabla 1 que informan sobre las calificaciones dadas por cada usuario de la plataforma que haya utilizado el juego, positivas o negativas. La llave primaria que utilizaría es el Name (nombre) del juego. Se separó en un archivo CSV diferente de la tabla original.

### **Cantidad de Jugadores:**

Esta tabla posee la información cuantitativa de jugadores activos, promedios e incrementos de cada juego, utilizando como llave primaria la App ID. Es la Tabla 2 a utilizar.

**Tabla 1 - Juegos**

<b>Campo</b>	<b>Tipo de Dato</b>	<b>Tipo de Clave</b>
App_ID	INT	FK
Name	VARCHAR(50)	FK
Short_description	VARCHAR(200)	
Owners	VARCHAR(30)	
Price	INT	
Initial_price	INT	
Discount	INT	
CCU	INT	
Languages	TEXT(200)	
Platforms	TEXT(30)	
Release_dates	DATE	
Required_age	INT	
Website	VARCHAR(300)	
Header_image	VARCHAR(300)	PK

**Tabla 2 – Cantidad de Jugadores**

<b>Campo</b>	<b>Tipo de Dato</b>	<b>Tipo de Clave</b>
Month	VARCHAR(20)	
Avg_players	FLOAT	
Gain	FLOAT	
Porc_gain	VARCHAR(15)	
Peak_players	INT	
App_ID	INT	PK
Games	VARCHAR(50)	

**Tabla 3 – Clasificación**

<b>Campo</b>	<b>Tipo de Dato</b>	<b>Tipo de Clave</b>
App_ID	INT	PK
Categories	TEXT(200)	
Genre	TEXT(50)	
Tags	VARCHAR(200)	
Type	TEXT(8)	

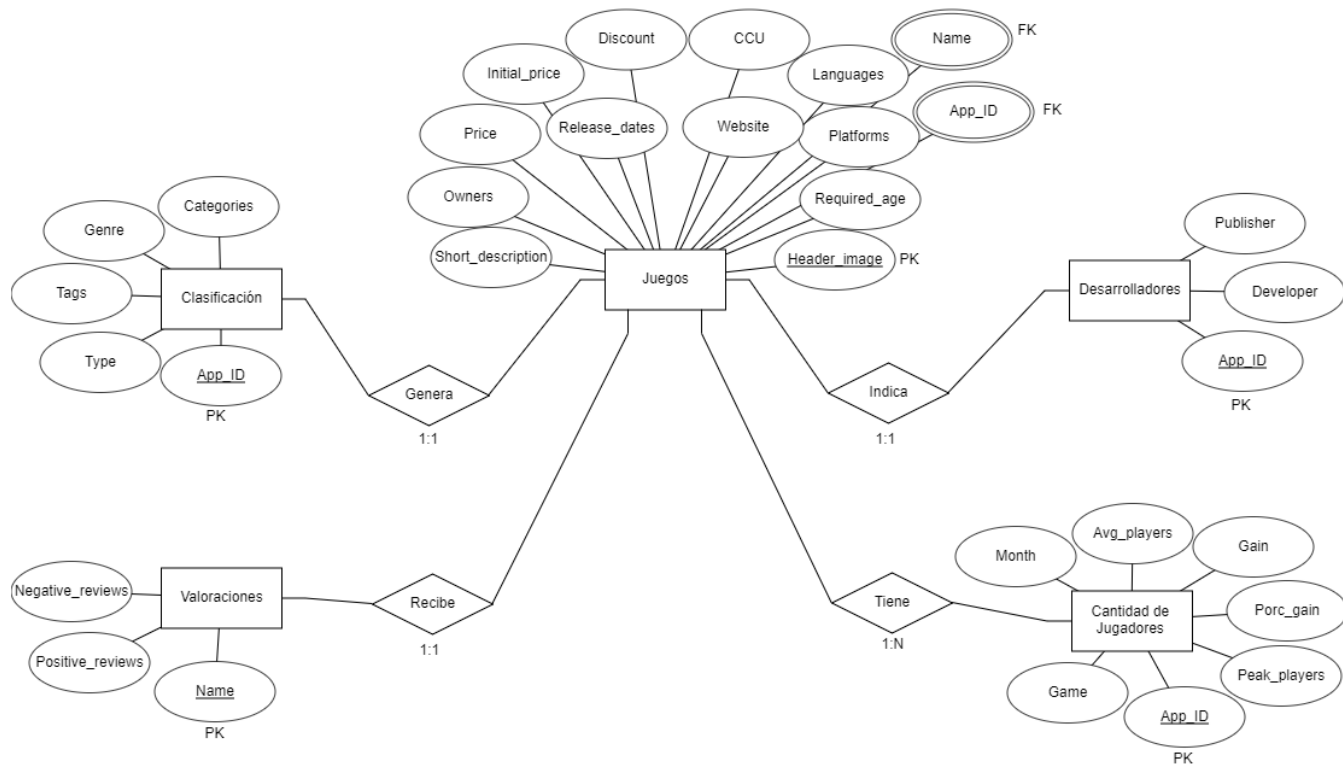
**Tabla 4 - Desarrolladores**

<b>Campo</b>	<b>Tipo de Dato</b>	<b>Tipo de Clave</b>
App_ID	INT	PK
Developer	VARCHAR(20)	
Publisher	VARCHAR(20)	

**Tabla 5 - Valoraciones**

<b>Campo</b>	<b>Tipo de Dato</b>	<b>Tipo de Clave</b>
Name	VARCHAR(50)	PK
Positive_reviews	INT	
Negative_reviews	INT	

# Diagrama Entidad-Relación



## **Transformaciones de Tablas**

De la base de datos original se eliminaron todos los campos sin información, pasando de aproximadamente 60.000 datos a 55.000 datos finales. Se ajustaron los valores numéricos a “No resumir” para evitar inconvenientes al momento de procesar los datos en el tablero.

### **Tabla 1 – Juegos:**

- Se removieron los errores indicados en el programa.
- Se eliminaron duplicados en la columna App\_ID.
- Se eliminaron duplicados en la columna Name.
- Se cambia formato de Price a INT para poder manejar los precios en los análisis.
- Se crea una nueva columna Category of Price para clasificar los diferentes precios de la columna Price.

Category of Price = IF('Tabla 1 - Juegos'[Price] < 500, "Low Cost", IF('Tabla 1 - Juegos'[Price] < 2000, "Medium Cost", "High Cost"))

### **Tabla 2 – Cantidad de Jugadores:**

- Se intenta cambiar el tipo de columna Month en formato Fecha pero presenta algunos errores.
- Se procede a reemplazar los valores “may-YY” por “May YYYY” de la columna Month.
- Se procede a reemplazar los valores “Last 30 days” por “July 2022”, al ser la última fecha de actualización de datos.
- Se cambia el formato de la columna Month a Fecha de forma exitosa.
- Se cambia el nombre de la columna Month a “Estimated\_date”.
- Se crea la columna Month en relación a la columna Estimated\_date.  
Month = MONTH('Tabla 2 - Cantidad de Jugadores'[Estimated\_date])
- Se crea la columna Year en relación a la columna Estimated\_date.  
Year = YEAR('Tabla 2 - Cantidad de Jugadores'[Estimated\_date])

### **Tabla 3 – Clasificación:**

- Se divide la columna Genre por delimitador de coma. Esto para tener en cuenta el Género principal de cada juego para los análisis. Ya que cada juego posee en distinta medida diferentes Géneros, pero uno solo principal.
- Se cambia el formato de Genre.1 y Genre.2 a Texto.
- Se cambia el nombre de la columna Genre.1 a Main Genre.
- Se cambia el nombre de la columna Genre.2 a Genres.
- Se procede a cambiar los valores nulos o en blanco a “Ninguno”.
- Se procede a cambiar los valores “Ninguno” por “None”.



**Tabla 4 – Desarrolladores:**

- Se colocan los valores en blanco o nulos de las columnas Developer y Publisher como None.
- Luego se modifica que todos los valores de Developer y Publisher como None sea reemplazado por Unknow.

**Tabla 5 – Valoraciones:**

- Se verifican y rectifican los formatos de cada columna a utilizar pero no hubo necesidad de realizar modificaciones en esta tabla.

**Creación de Tabla Medidas Calculadas:**

Se explicaran en otro apartado, pero se realiza creación de la tabla para mejor organización y distribución de los datos. Se le da el nombre de Medidas DAX.

**Creación de Tabla Calendario:**

Se creó la tabla calendario para poder tener una interacción temporal entre la fecha mínima y máxima que se encuentra en la base de datos, principalmente desde Tabla 1 – Juegos. Se crean las columnas de Day, Month, Year y Quarter para cualquier uso o cálculo posible, ya sea en el tablero actual o con modificaciones a futuro.

## Medidas Calculadas

Se encuentran todas posicionadas y registradas en la tabla Medidas DAX.

**-Cantidad de juegos por desarrollador:** se utiliza para saber la cantidad exacta de cada juego por desarrollador en la tabla 4.

Cantidad de juegos por desarrollador = COUNT('Tabla 4 - Desarrolladores'[Developer])

**-Cantidad de juegos por publicador:** se utiliza para saber la cantidad exacta de cada juego por publicador en la tabla 4.

Cantidad de juegos por publicador = COUNT('Tabla 4 - Desarrolladores'[Publisher])

**-Juegos Totales Registrados:** utilizada para conocer la cantidad total de juegos en la base de datos de registros de Steam.

Juegos Totales Registrados = COUNT('Tabla 1 - Juegos'[App\_ID])

**-Mediana de precios:** conocer la mediana de costos de juegos.

Mediana de precios = MEDIAN('Tabla 1 - Juegos'[Price])

**-Promedio de precios:** conocer el promedio de costos de juegos

Promedio de precios = AVERAGE('Tabla 1 - Juegos'[Price])

**-Puntuación real de valoración:** utilizada para conocer la puntuación o valoración real de cada juego. Restándole la cantidad negativa a las positivas para crear una valoración más precisa de cada juego.

Puntuacion real de valoracion = [Valoraciones positivas] - [Valoraciones negativas]

**-Valoraciones negativas:** suma de valoraciones negativas.

Valoraciones negativas = SUM('Tabla 5 - Valoraciones'[Negative\_reviews])

**-Valoraciones positivas:** suma de valoraciones positivas.

Valoraciones positivas = SUM('Tabla 5 - Valoraciones'[Positive\_reviews])

**-Valoraciones totales:** se utiliza para conocer el alcance de cantidad de valoraciones de cada juego.

Valoraciones totales = [Valoraciones positivas] + [Valoraciones negativas]

**-Porcentaje de juegos Low Cost:** se utiliza para conocer el porcentaje de juegos dentro de la plataforma que sean clasificados como Low Cost.

Porcentaje de juegos Low Cost =

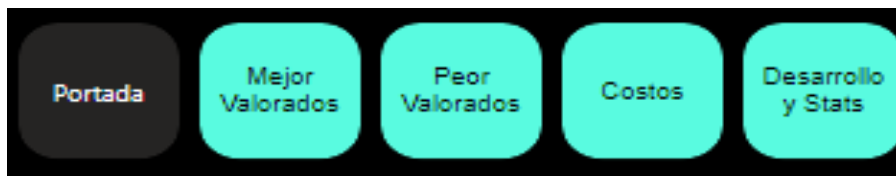
VAR Cantidad\_juegos = COUNT('Tabla 1 - Juegos'[App\_ID])

VAR Cantidad\_juegos\_Low\_Cost = CALCULATE(COUNT('Tabla 1 - Juegos'[App\_ID]),  
'Tabla 1 - Juegos'[Category of Price] = "Low Cost")

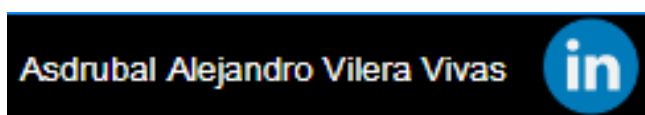
RETURN DIVIDE(Cantidad\_juegos\_Low\_Cost \* 100, Cantidad\_juegos)

## Análisis Funcional del Tablero de Power BI

La letra utilizada en todas las solapas es de Arial en Negrita. Los tamaños suelen variar dependiendo del tipo de encabezado, título o valor indicado en 10, 11, 16 y 40. En todas las solapas del tablero se presentan los botones interactivos para desplazarse por cualquier solapa deseada.



En todas las solapas se deja referencia del creador de la presentación del tablero y un ícono con acceso directo al URL del perfil de LinkedIn.



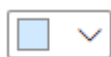
En todas las solapas se deja en evidencia el logo central de la empresa, respetando los colores utilizados por la plataforma y página web principal.



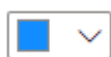
La paleta de colores utilizada para la presentación, en general, es de los siguientes colores con sus respectivos códigos hexadecimales.



Código: #59FBE0



Código: #D1E9FF



Código: #118DFF



Código: #000000

Para los gráficos en los que se necesitó mayor interacción de colores se agregaron los siguientes.

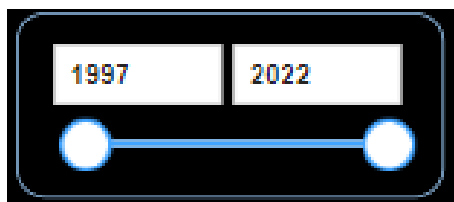
 Código: #118DFF

 Código: #12239E

 Código: #D64550

 Código: #A666B0

Luego de la primera solapa introductoria se implementó en el resto de solapas un segmentador en la parte superior para poder evaluar de forma interactiva la comprensión entre años y poder encontrar relaciones interesantes para las evaluaciones del proyecto. Se deja desde 1997 ya que esta fecha fue la creación del juego más antiguo de la plataforma (incluso sin haber existido la plataforma en ese momento).



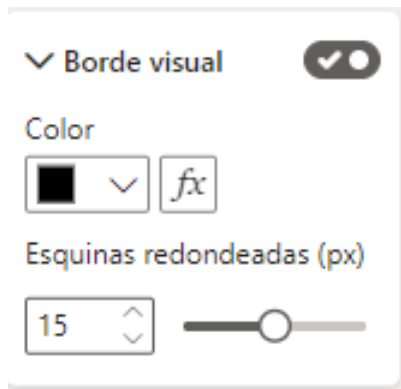
Apreciamos que desde la solapa 2 a la 5 tenemos diferentes segmentadores, lo cual permite apreciar la información desde distintas perspectivas. Son los segmentadores Géneros (selección múltiple), Publicadores y Desarrolladores (ambos últimos con lista desplegable). Los cuales permiten filtrar la información que se visualiza para mejores evaluaciones.



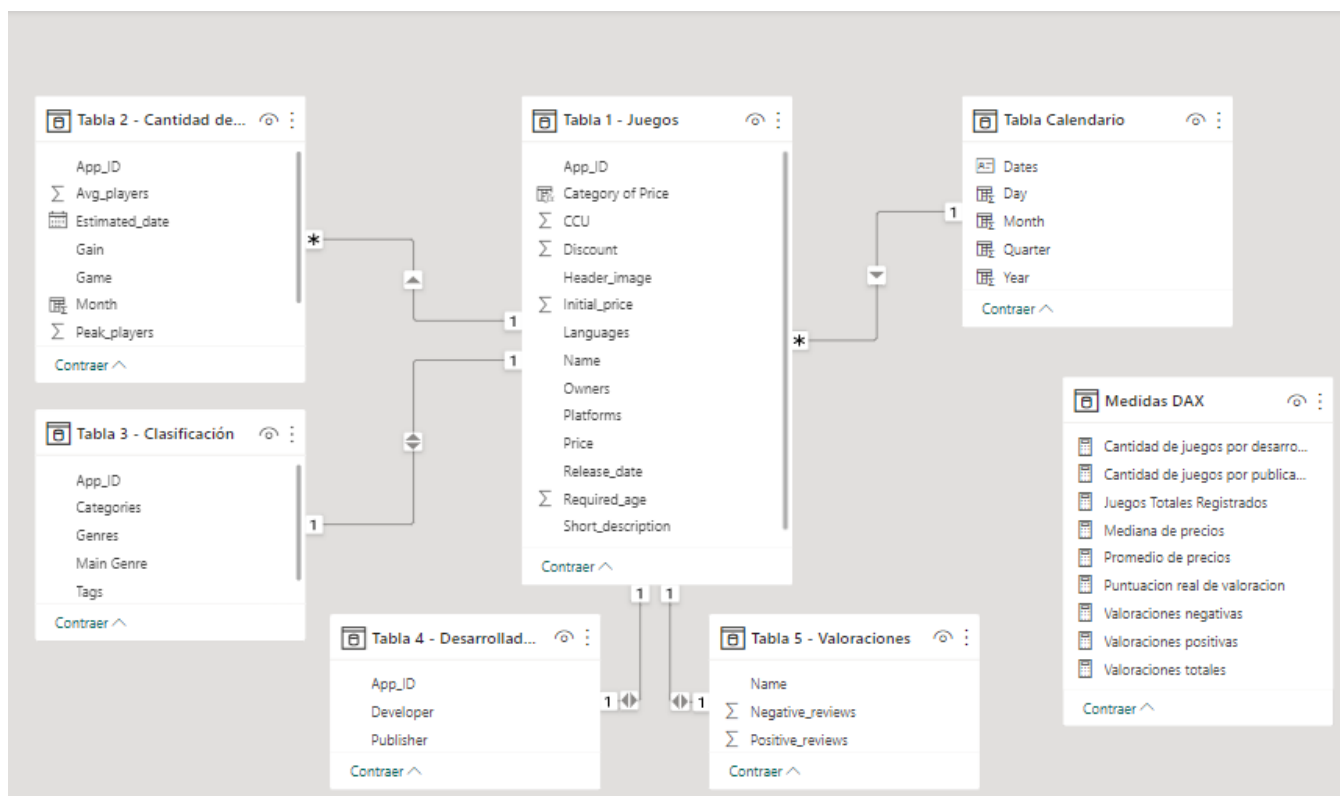
También se implementó, en el resto de solapas, un ícono interactivo en la parte inferior que permite retroceder a la solapa que se estaba visualizando anteriormente. Así como también el ícono de quitar los filtros de la página que se esté visualizando.



Siempre se aplicó la modalidad de recuadro con esquinas redondeadas para los objetos visuales centrales y laterales.



## Diagrama Entidad-Relación de Power BI:



## Análisis Funcional por Solapa

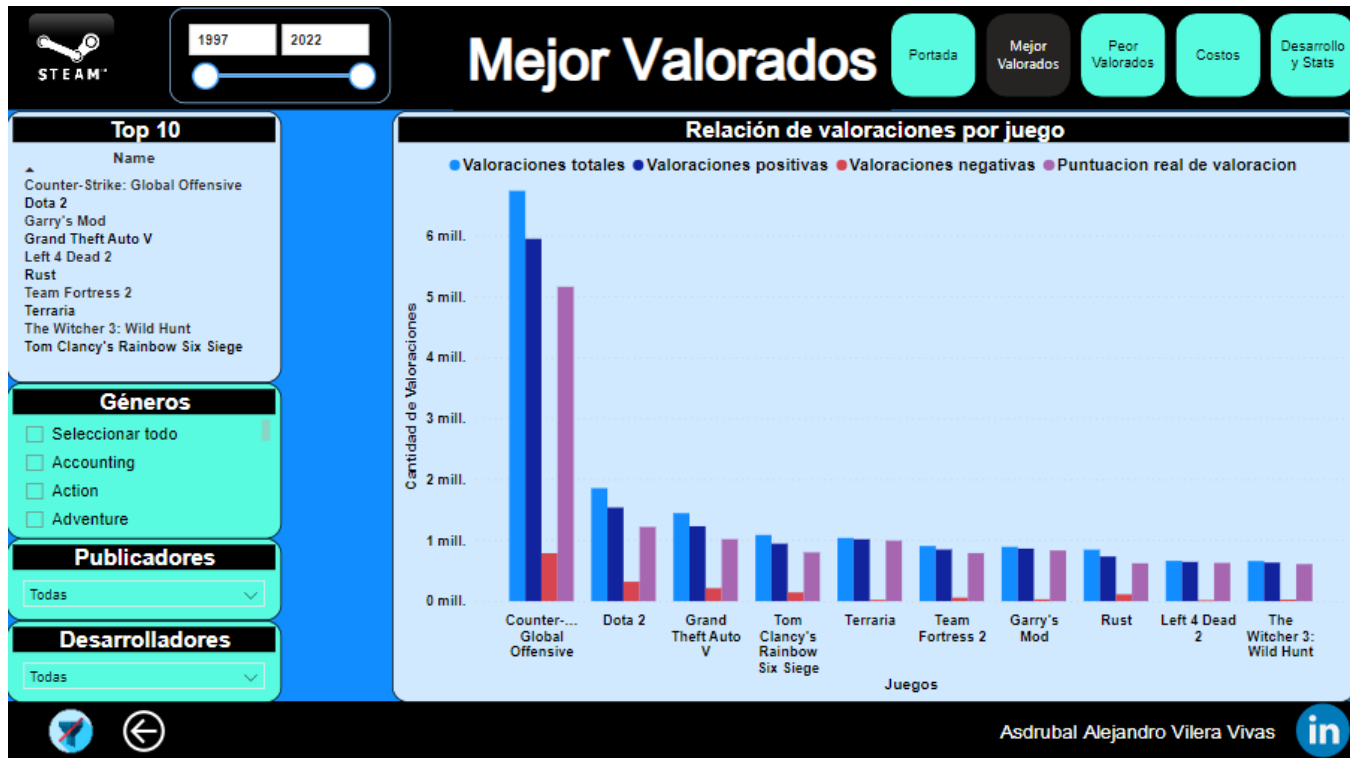
### Solapa 1 – Portada:



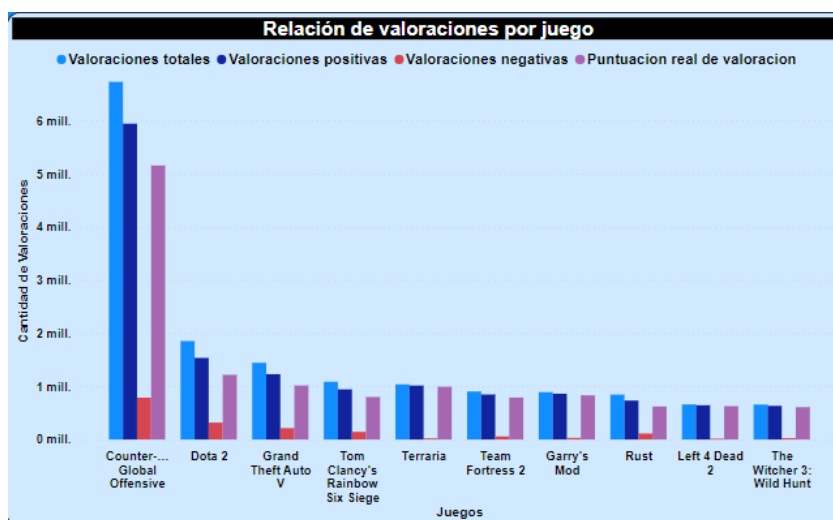
Por medio de la portada, presentamos la temática del proyecto y colocamos centrado y en la sección inferior un ícono interactivo para redirigir a la URL principal de la plataforma de Steam.

Se aprecia en la sección superior derecha los botones interactivos para desplazarnos por todo el tablero interactivo.

## Solapa 2 – Mejor valorados:



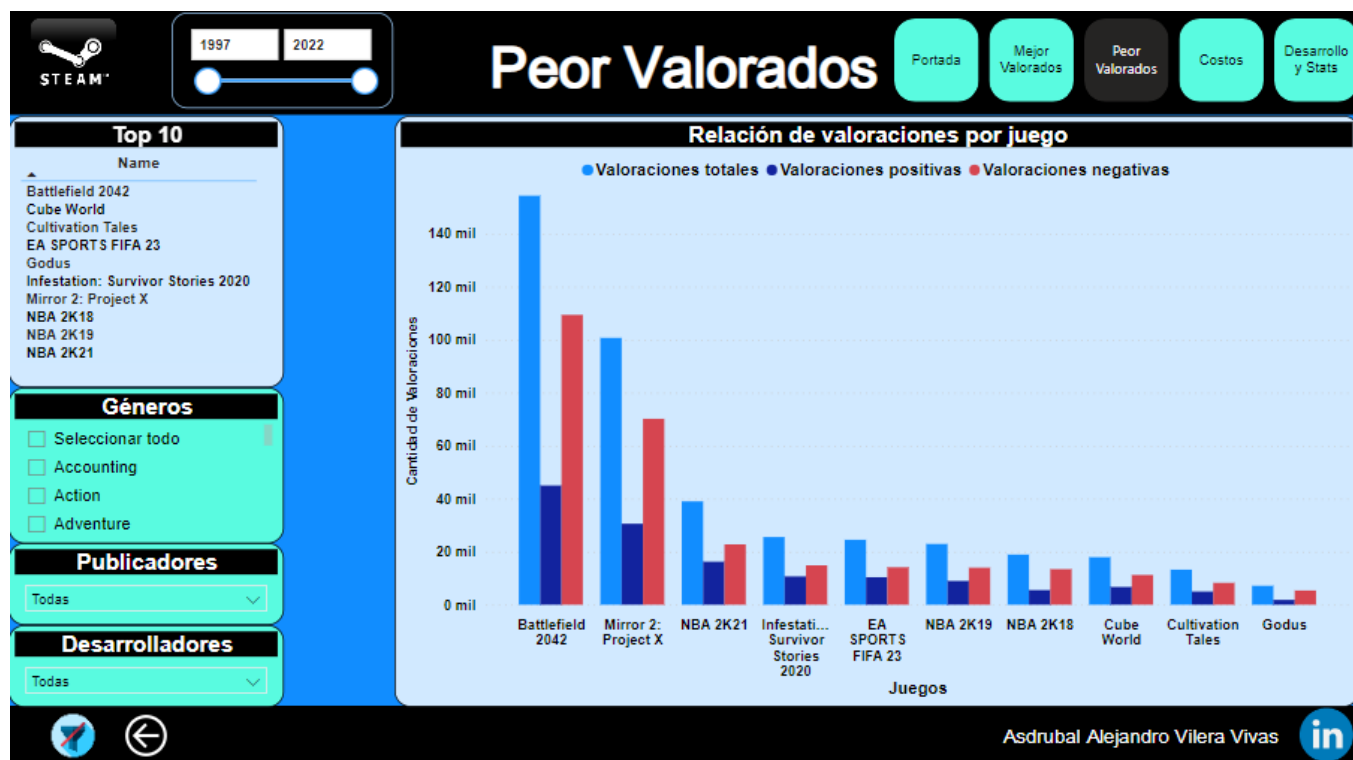
En esta solapa podemos visualizar siempre el Top 10 de juegos mejor valorados. Del lado izquierdo se aprecia el Top 10 en forma de lista para poder tenerlos presentes en cualquier evaluación en orden alfabético. Inferior a este último podemos apreciar tres diferentes segmentadores: Géneros, permite filtrar los top 10 entre los diferentes géneros de la plataforma; Publicadores, permite filtrar por los principales clientes publicadores de la plataforma; Desarrolladores, permite filtrar por los principales clientes desarrolladores de la plataforma.



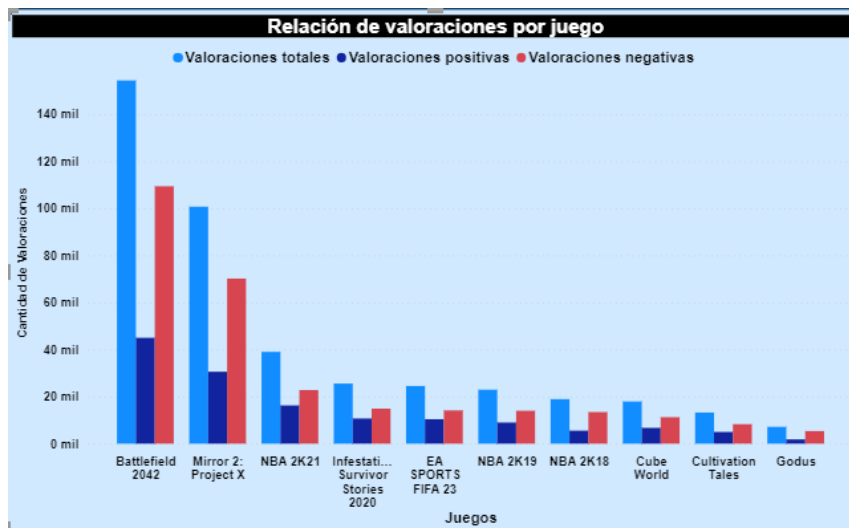


Con el gráfico de columnas agrupadas podemos evaluar la relación existente entre la cantidad de valoraciones realizadas por la comunidad de usuarios de la plataforma a cada juego. Por lo que permite visualizar cuáles han tenido más éxito entre sus usuarios y separar los diferentes tipos de valoraciones entre cada juego. De esta forma, también permite filtrar con las herramientas anteriormente mencionadas y tener una amplia evaluación de los principales juegos y géneros de la plataforma. El gráfico utiliza las medidas calculadas Valoraciones totales, Valoraciones positivas, Valoraciones negativas y Puntuación real de valoración.

### Solapa 3 – Peor Valorados:

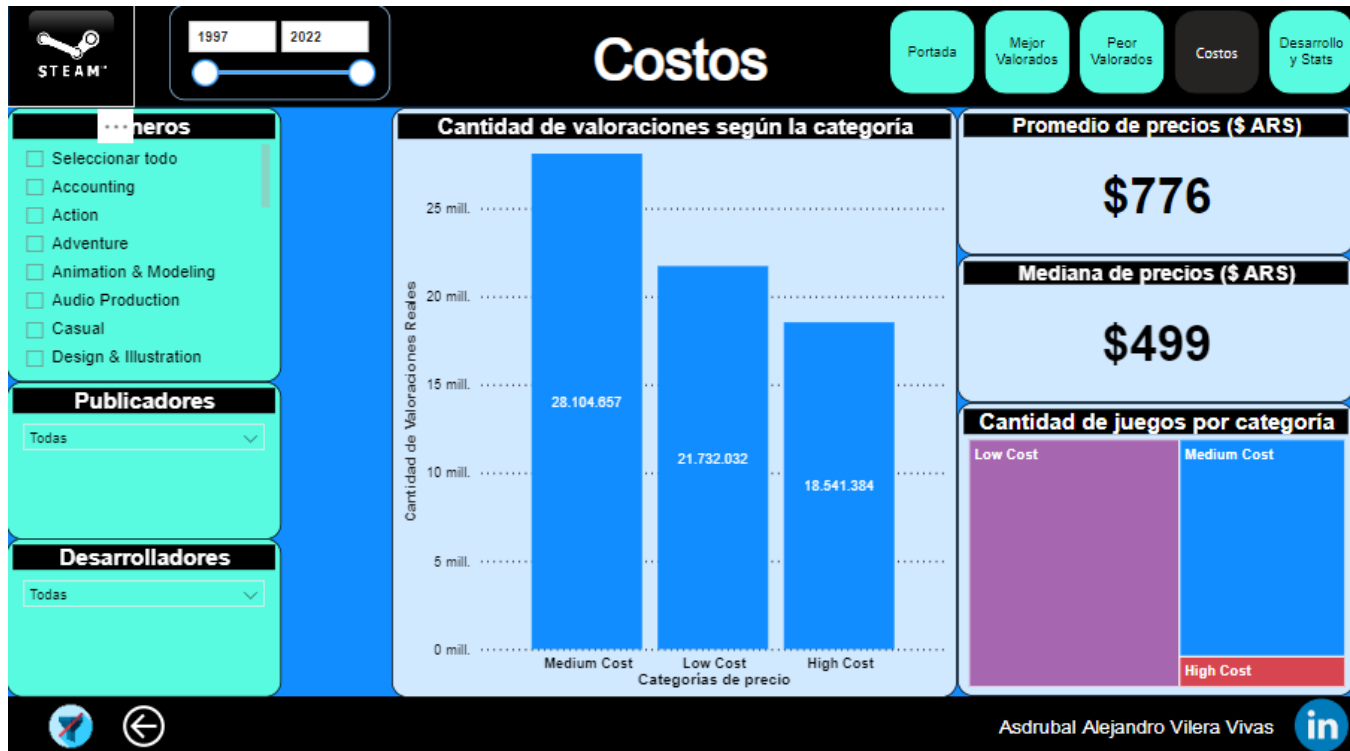


En esta solapa podemos visualizar siempre el Top 10 de juegos peor valorados. Del lado izquierdo se aprecia el Top 10 en forma de lista para poder tenerlos presentes en cualquier evaluación en orden alfabético. Inferior a este último podemos apreciar tres diferentes segmentadores: Géneros, permite filtrar los top 10 entre los diferentes géneros de la plataforma; Publicadores, permite filtrar por los principales clientes publicadores de la plataforma; Desarrolladores, permite filtrar por los principales clientes desarrolladores de la plataforma.



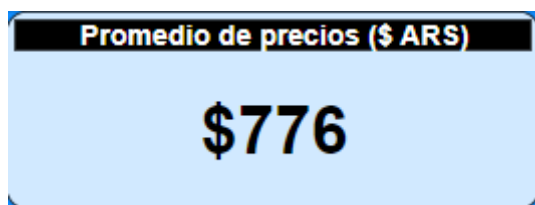
Con el gráfico de columnas agrupadas podemos evaluar la relación existente entre la cantidad de valoraciones realizadas por la comunidad de usuarios de la plataforma a cada juego. Por lo que permite visualizar cuáles han tenido menos éxito entre sus usuarios y separar los diferentes tipos de valoraciones entre cada juego. De esta forma, también permite filtrar con las herramientas anteriormente mencionadas y tener una amplia evaluación de los principales juegos y géneros de la plataforma. El gráfico utiliza las medidas calculadas Valoraciones totales, Valoraciones positivas, Valoraciones negativas. A comparación con el gráfico similar de la solapa anterior, podemos apreciar que no se realiza en conjunto con la medida Puntuación real de valoración, ya que al tener más cantidad de valoraciones negativas que positivas resulta imposible mostrar una valoración real al respecto.

## Solapa 4 – Costos:



Para poder establecer una categoría en cuanto al precio se creó la medida calculada  $\text{Category of Price} = \text{IF}(\text{'Tabla 1 - Juegos' [Price]} < 500, \text{"Low Cost"}, \text{IF}(\text{'Tabla 1 - Juegos' [Price]} < 2000, \text{"Medium Cost"}, \text{"High Cost"}))$ . Lo cual permite establecer que cualquier juego que cueste menos de 500 ARS es Low Cost; menos de 2000 ARS sería Medium Cost y mayor a esta cifra sería clasificado como High Cost. Estas cifras se tomaron en cuenta por entrevistas realizadas a la empresa y opiniones de los usuarios de la plataforma.

Indicando ya la clasificación de costos, podemos apreciar que se distribuyen diferentes gráficos de tarjeta para explicar la situación general a nivel de costos.



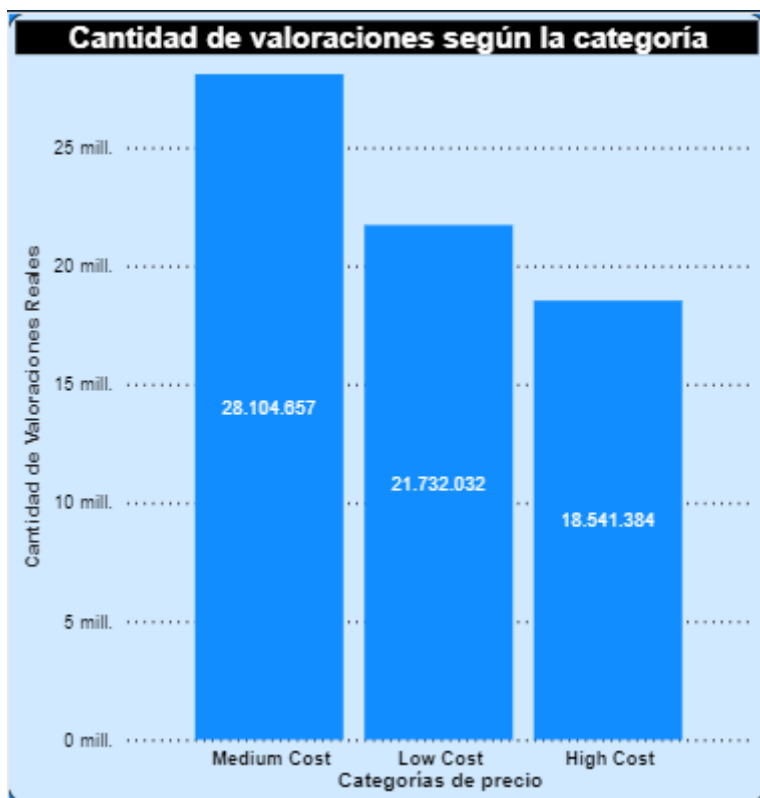
Utiliza la medida calculada Promedio de precios. Es un promedio de los costos.



Utiliza la medida calculada Mediana de precios. Es una mediana de los costos.



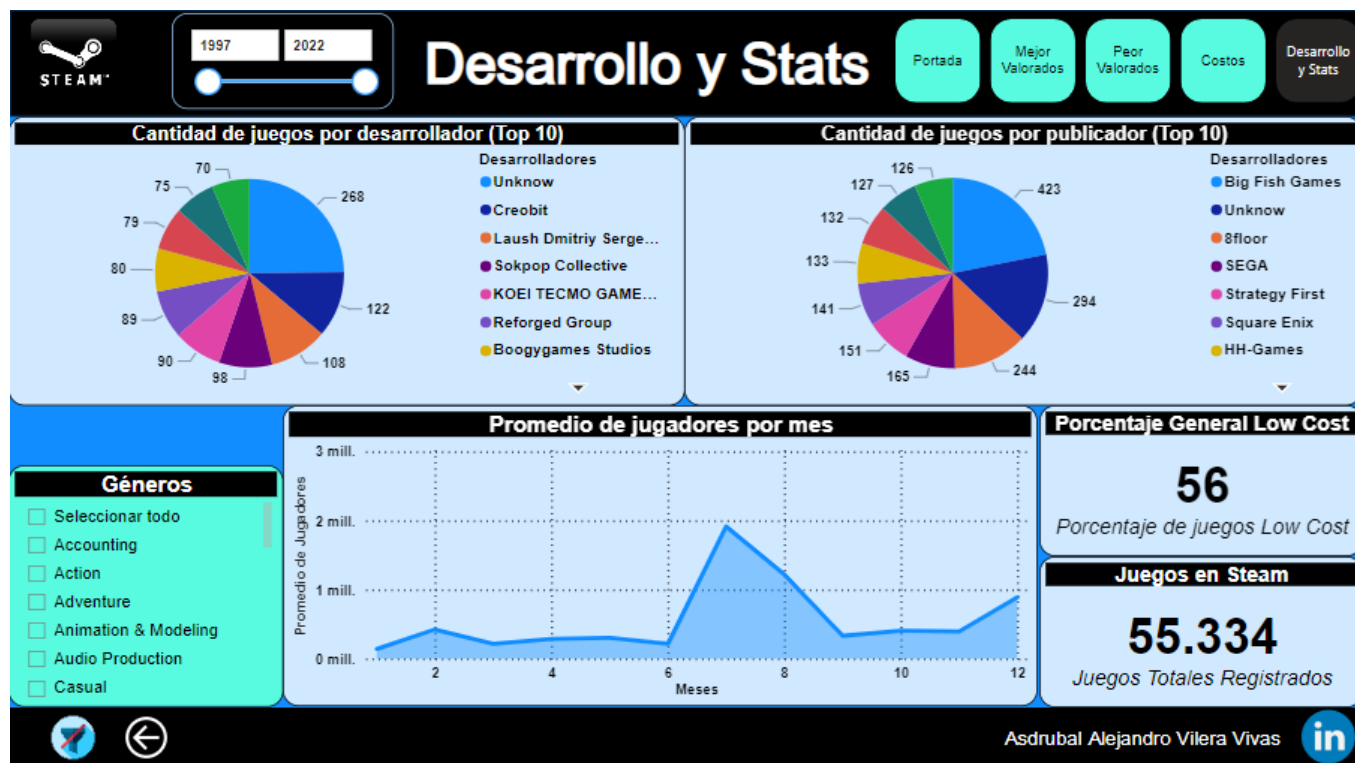
Se aprecia también un gráfico treemap para conocer, en proporción, la cantidad de juegos en la plataforma según su clasificación de costos. Utilizando como categoría la columna creada Category of Price.



Podemos observar un gráfico de columnas agrupadas que establece la relación precio-calidad y la cantidad de valoraciones reales que recibe cada una de las categorías de costo. Utiliza la nueva columna Category of Price y la medida generada Puntuación real de valoración.

En la sección lateral izquierda apreciamos tres diferentes segmentadores: Géneros, permite filtrar entre los diferentes géneros de la plataforma; Publicadores, permite filtrar por los principales clientes publicadores de la plataforma; Desarrolladores, permite filtrar por los principales clientes desarrolladores de la plataforma.

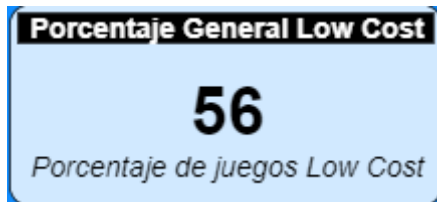
## Solapa 5 – Desarrollo y Stats:



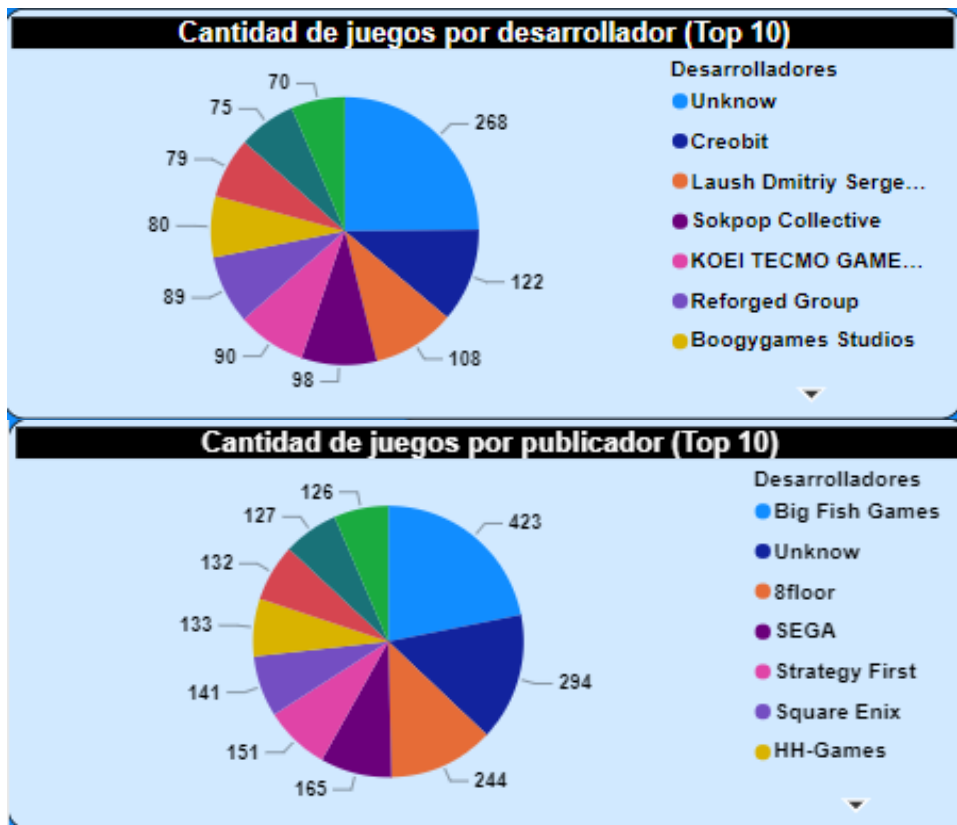
Para nuestra última solapa registramos información sobre los desarrolladores y publicadores, así como los registros principales de la plataforma. En la esquina inferior izquierda podemos apreciar el segmentador de Géneros, el cual permite filtrar entre los diferentes géneros de la plataforma para todos los gráficos y tarjetas de la solapa.



Utiliza la medida calculada Juegos Totales Registrados.



Utiliza la medida calculada Porcentaje de juegos Low Cost.



Ambos gráficos circulares indican la cantidad de juegos, ya sea por publicador o por desarrollador. Obviamente, indica sólo los principales 10 de cada empresa, ya que, para éste análisis, nos interesa conocer quiénes son los principales de la plataforma. Podemos apreciar que hay muchos juegos sin información de sus creadores y publicadores. Ambos gráficos, de desarrolladores y publicadores, utilizan las medidas

calculadas Cantidad de juegos por desarrollador y Cantidad de juegos por publicador, respectivamente.



Podemos visualizar en la parte inferior un gráfico de áreas apiladas que nos permite observar el promedio de jugadores por cada mes del año, e incluso, un proporcional de diferentes años comprendidos para observar un promedio de cada mes de forma individual. De esta forma podemos conocer el comportamiento de la plataforma en cuanto a cantidad de usuarios y conocer un patrón que ayude a implementar estrategias futuras. Utiliza la columna de Month de la Tabla Calendario.

## **Conclusiones**

Podemos observar que la plataforma cuenta con una cantidad de productos considerable. Sin embargo, es necesario brindarle relevancia a la forma de promocionar y mantener estos juegos a flote dentro de la plataforma. Una baja cantidad (pero considerable) de juegos terminan siendo de gran relevancia para la empresa.

En cuanto al comportamiento de usuarios en la plataforma, al menos considerando todo el plazo de tiempo del proyecto (2003-2022), apreciamos que el mes con mayores ingresos es en Julio y cercano a final de año, lo cual permite promocionar nuevos juegos y estrategias durante estas fechas de forma anual e implementar los cambios más relevantes y mantenimientos durante los principios de año.

Para los registros de cada juego subido a la plataforma, se puede pensar en formalizar un registro más estricto en cuanto a la información provista por los desarrolladores y publicadores de la plataforma. Esto se concluye por la alta cantidad de “Unknow” en gran parte de los juegos registrados.

Podemos ver que la mayoría de éxito de juegos fue de Medium Cost, por lo que se puede implementar un aumento masivo de los juegos (consultando con cada desarrollador y publicador al respecto), ya que el promedio de costos es muy bajo a comparación a la devaluación anual que presenta Argentina para totales de 2022. Esto va a permitir mejor distribución de clasificación y aumento de las ganancias al promocionar más juegos Medium dentro de la plataforma. Esto se toma en cuenta también analizando la alta cantidad de juegos Low Cost que forman parte de la plataforma contra los diferentes tipos de costos.

Para mantener los juegos ya exitosos, se sugiere promocionar nuevas ofertas y modalidades para seguir incrementando la cantidad de usuarios. En cuanto a los juegos peor valorados, esta observación nos permite calcular nuevas estrategias para mejorar la jugabilidad y temáticas de los juegos ya presentes y los nuevos por ingresar.