

CODERHOUSE

Comisión 37755

Data Analytics

Equipo:

- Francesco Centarti Maestu
- Pablo Maestu

1ra Pre-entrega

>> Objetivos Generales:

Definir la temática del proyecto

Iniciamos este trabajo mediante la búsqueda y recopilación de bases de datos relativas a las visualizaciones de videojuegos en la plataforma Twitch, contando actualmente con la totalidad de las mismas en el lapso de tiempo comprendido entre el año 2016 y 2021.

Además, disponemos de los premios otorgados por cada juego en sus distintas competiciones desde el año 1998 a 2021.

De esta forma, nos centraremos en el análisis de la relación y/o comparación entre la audiencia en Twitch y las ganancias monetarias por cada juego en torneos, cuestión que consideramos es de gran utilidad para empresas desarrolladoras de videojuegos. Todo esto, en el lapso comprendido entre los años 1998 y 2021.

>> Objetivos Específicos:

Definir un objetivo de manera SMART.

S Determinar el comportamiento a nivel económico de los videojuegos en relación al caudal de vistas en la plataforma Twitch.

M Serán medidos los premios otorgados en torneos y la cantidad de horas visualizadas de cada Videojuego en Twitch.

A El realismo y posibilidad de aplicación de nuestro objetivo se basa en que disponemos de bases de datos completas referidas a la cantidad de visualizaciones en Twitch como también los montos de dinero otorgados por cada juego en un rango de tiempo determinado.

R Nuestro análisis es relevante y aplicable al contexto socioeconómico actual dado el aumento de la utilización de plataformas de stream online y el consumo de videojuegos, siendo de gran utilidad para empresas desarrolladoras a la hora de iniciar nuevos proyectos.

T Se estima que el presente trabajo estará terminado para el mes de Septiembre del año 2022, considerándose de aplicación inmediata para empresas desarrolladoras de videojuegos dada la contemporaneidad de los datos con los que contamos.

Describir una hipótesis de trabajo.

Comprobar si existe relación entre la audiencia generada por los videojuegos en la plataforma Twitch con las ganancias de los torneos de videojuegos, con el objetivo de que las empresas desarrolladoras sepan dónde convendría invertir y la importancia de su participación en plataforma Twitch.

Identificar las fuentes de datos necesarias.

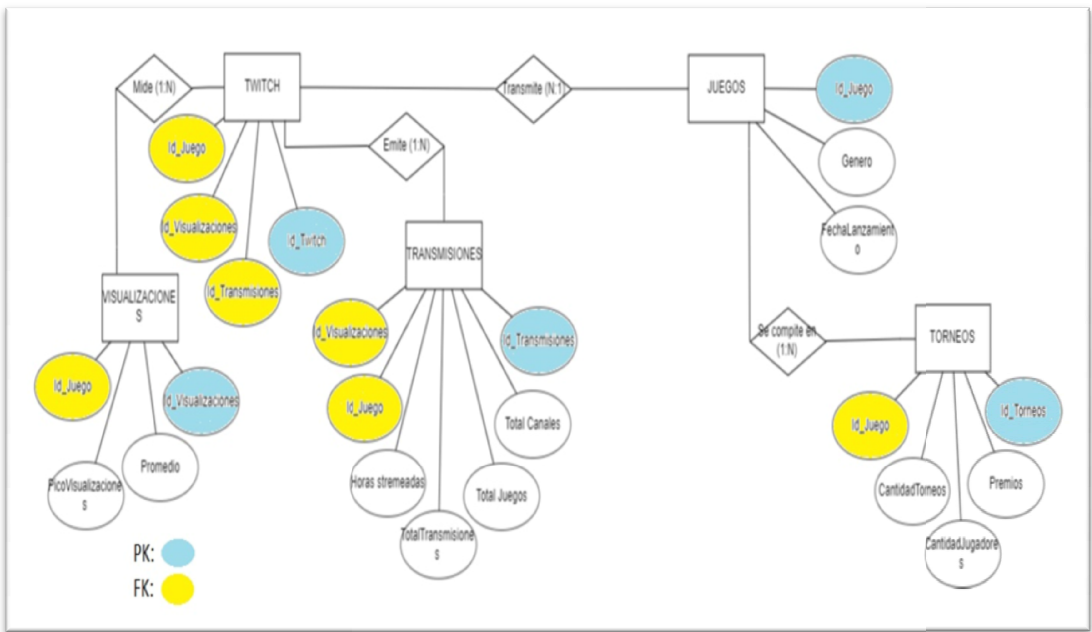
Dentro de la carpeta “DataSet” se encuentran las bases de utilizadas en el presente trabajo.

Links de bases de datos:

<https://www.kaggle.com/datasets/rankirsh/esports-earnings>

<https://www.kaggle.com/datasets/rankirsh/evolution-of-top-games-on-twitch>

Diagrama entidad-relación (incorporado a la carpeta de entrega en formato .jpg):



A continuación se realiza una descripción del contenido de las tablas utilizadas en el proyecto.

Tabla 1: TWITCH

Contiene la relación entre transmisiones y visualizaciones totales de cada juego.

Tipo de clave	Campo	Tipo de Campo
PK- Index	id_Twitch	int
FK	id_Visualizaciones	int
FK	id_Transmisiones	int
FK	id_Juego	nvarchar(500)

Tabla 2: VISUALIZACIONES

Contiene los datos de visualizaciones totales, pico y promedio de cada juego.

Tipo de clave	Campo	Tipo de Campo
PK- Index	id_Visualizaciones	int
FK	id_Juego	nvarchar(500)
	PicoVisualizaciones	int
	Promedio	int

Tabla 3: TRANSMISIONES

Contiene los datos de transmisiones de cada juego, teniendo en cuenta las horas streameadas y la cantidad de canales que transmiten.

Tipo de clave	Campo	Tipo de Campo
PK- Index	id_Transmisiones	int
FK	id_Juego	nvarchar(500)
FK	id_Visualizaciones	int
	HorasStreameadas	nvarchar(500)
	TotalJuegos	int
	TotalCanales	int
	TotalTransmisiones	int

Tabla 4: JUEGOS

Contiene el nombre, género y fecha de lanzamiento de cada juego.

Tipo de clave	Campo	Tipo de Campo
PK- Index	id_Juego	nvarchar(500)
	genero	nvarchar(500)
	FechaLanzamiento	int

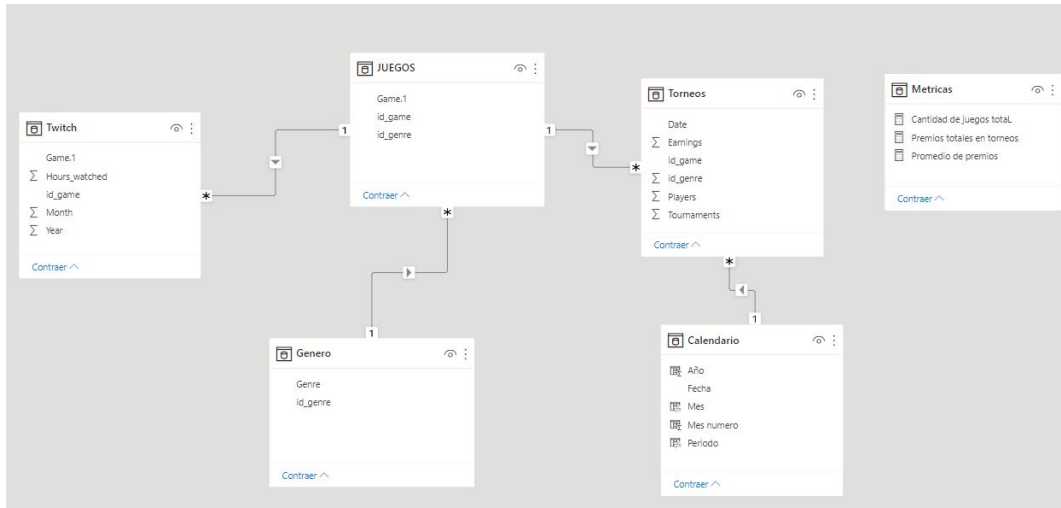
Tabla 5: TORNEOS

Contiene la cantidad de jugadores y torneos de cada juego, incluyendo el monto de los premios otorgados.

Tipo de clave	Campo	Tipo de Campo
PK- Index	id_Torneos	int
FK	id_Juego	nvarchar(500)
	CantidadTorneos	
	CantidadJugadores	
	Premios	

POWER BI

Para la utilización de Power BI modificamos el diagrama entidad relación usado anteriormente por un diagrama estrella.



A continuación hacemos un recuento de Transformaciones realizadas, medidas calculadas generadas con sus fórmulas y de las segmentaciones elegidas.

Se creó el parámetro Ruta para especificar ruta de acceso de archivo (los archivos en un primer momento antes de las modificaciones se llamaban “GeneralSportData” e “HistoricalEsportData”).

- Se duplicó la tabla GeneralEsportData que pasó a denominarse “JUEGOS”. Se creó en la tabla el índice de juegos y se le puso el nombre id_juegos. Se combinó la tabla JUEGOS con el índice de juegos creados (id_juegos) y se quitaron las columnas nombre de juego, fecha de lanzamiento y juegos.
- Se modificó el tipo de dato en las ganancias totales a \$.
- Se duplicó Games y se la denominó Generos. Se trajo a la consulta géneros y quitaron los duplicados. Se creó el índice de géneros.
- Se quitaron las columnas game, release y genre de JUEGOS.
- Se combinó en HistoricalEsportData game con id_game y se eliminó game de la consulta (HistoricalEsportData pasó a denominarse Twitch).
- Se combinó id_games de historical con id_games de general y nos quedamos con el id_genre

- Se creó la tabla calendario, en base a las fechas de torneos. Se creó una columna de año, una del número del mes, y otra con el mes a la que se le indicó que se ordene por “Mes Numero”. Además se creó una de periodo. Se desactivó desde la configuración de Power bi la fecha y hora automática de carga de datos.
- Cargamos la base de datos de Twitch. Quitamos columnas innecesarias. Duplicamos la misma tabla para crear la de juegos.
- Combinamos consulta entre Games y JUEGOS para crear la lista de juegos, con combinación interna solo de filas coincidentes y quitamos los duplicados.
- Quitamos la carga de la anterior tabla Games
- Sumamos a la tabla JUEGOS el id de cada género (desde GeneralEsportData)

Generación de medidas:

Cantidad de juegos total = `DISTINCTCOUNT(JUEGOS[id_game])`

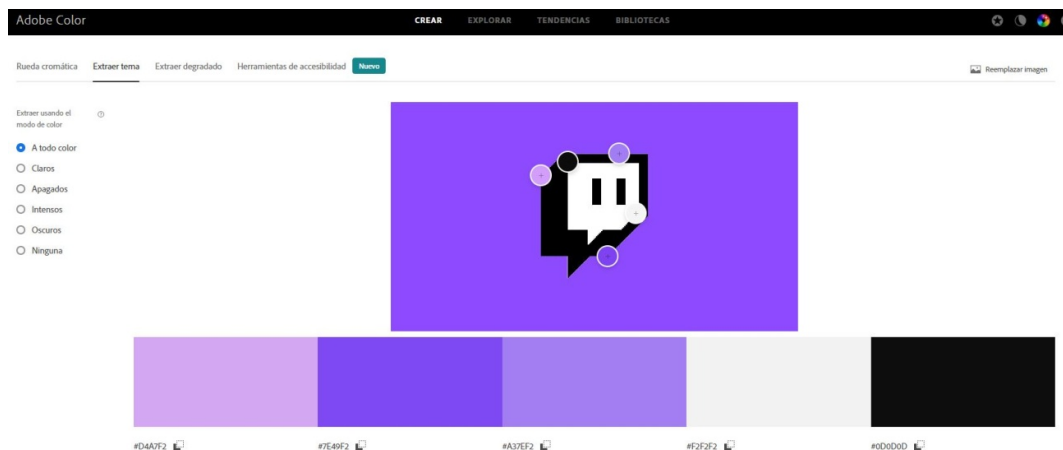
Cantidad total de horas visualizadas = `SUM(Twitch[Hours_watched])`

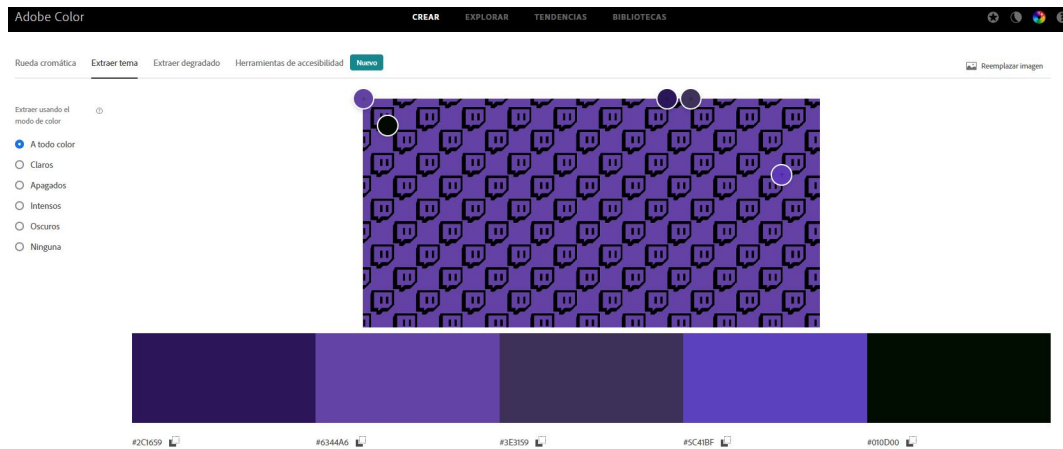
Cantidad total de torneos = `SUM(Torneos[Tournaments])`

Premios totales en torneos = `SUM(Torneos[Earnings])`

Dashboard – Visualización:

Se utilizó la paleta de colores correspondiente a uno de los factores fundamentales utilizados en el proyecto: Twitch desde <https://color.adobe.com/es/create/color-wheel>





Conclusión

En base al análisis realizado durante el proyecto, teniendo en cuenta la hipótesis de la que partimos, podemos observar que hay dos géneros que representan el 75,64% de los premios otorgados en torneos y el 68,16% de las visualizaciones en la plataforma online Twitch.

En primer lugar, el género Multiplayer online battle arena, que significa más de la mitad de premios totales en torneos, y se ubica primero también en la cantidad total de visualizaciones en Twitch, donde representa el 39,46%.

En segundo lugar, el género First-person shooter, con más de una cuarta parte de los premios en torneos (25,52%) y un 28,7% de visualizaciones en Twitch.

De esta manera, las empresas desarrolladoras de videojuegos deberían tener en cuenta principalmente estos géneros, no solamente en su aspecto competitivo sino también en lo que respecta a su difusión y participación activa en la principal plataforma de streaming online.