**Proyecto Integrado Video Juegos**

**1. Descripción de los datos.**

Se nos presentan cuatro tablas:

a) Indicadores del desarrollo humano mundial del Banco Mundial: En esta tabla encontramos una gran variedad de indicadores distribuidos por año (2000 – 2019) y en cuatro países/regiones (USA, UE, JPN, WORLD), que van desde el INB per cápita, la esperanza de vida, prevalencia de uso de anticonceptivos, hasta el área selvática. En orden para analizar los datos de los videojuegos, es necesario escoger los indicadores que puedan proporcionarnos una visión general del estado de la economía de diferentes regiones y del mundo. A continuación, se presenta un breve análisis de los indicadores escogidos.

*INB Per Cápita*: Es es el valor de todos los bienes y servicios producidos por los residentes de un país durante un período determinado (generalmente un año). Este indicador nos puede dar una señal del poder adquisitivo de las personas de un determinado país/región.

*Crecimiento del PIB (%anual) y Índice de Deflación del PIB (% anual):* Estos indicadores nos pueden ayudar a tener una percepción cómo avanza la economía región y si existe una disminución del poder adquisitivo o una posible inestabilidad económica en el país/región.

*Población total entre 0 y 14 años de edad, Población total entre 15 y 64 años de edad, Población total de 65 años de edad y más:* Estos indicadores nos darán una distribución de las edades por país/región, lo cual será importante en el análisis ya que en los indicadores de los juegos se encuentra el indicador de edad recomendada para los juegos.

Dado que en los datos de video juegos tenemos datos para Japón, la Unión Europea y el mundo, se tendrán en cuenta estos datos en la tabla Indicadores del desarrollo humano mundial del Banco Mundial; se incluye USA por ser referencia también.

b) Tablas Juegos en Steam y Video Game Sales:

Mientras que Juegos Steam reporta estadísticas sacadas desde esta plataforma para diferentes video juegos, Video Games Sales tiene información sobre ventas de los videos juegos. Al proporcionarnos información complementaria, tiene sentido unirlas por nombre de video juego y tratarlas como una sola tabla.

Cabe recalcar que, a pesar de que hay información realmente complementaria, hay otra información que es repetida; por ejemplo, la columna de género está en las dos tablas. En este caso específico es que utilizaremos la información de una columna para rellenar datos faltantes de la otra.

**2. Metas principales del análisis.**

En las tablas proporcionadas existen un millón de análisis que se pueden realizar; sin embargo, el realizar un análisis completo de todos los datos tomaría mucho más tiempo del que se dispone para este ejercicio académico. Teniendo en cuenta que el objetivo principal de este ejercicio es realizar un proceso ETL y, además, las preguntas de este ejercicio propuesto, se reduce el análisis a los siguientes propósitos:

1. Investigar la evolución del INB, crecimiento del PIB y inflación.
2. Revisar la distribución de ranking, rango etario, precio y ventas entre los diferentes géneros.

Según estos objetivos, estas son las siguientes columnas de estas tablas que se conservarán para el análisis:

*Indicadores del desarrollo humano mundial del Banco Mundial:* INB Per Cápita, Crecimiento del PIB (%anual), Índice de Deflación del PIB (% anual), Población total entre 0 y 14 años de edad, Población total entre 15 y 64 años de edad, Población total de 65 años de edad y más.

*Tablas Juegos en Steam y Video Game Sales:* Game Name/name, Release\_date/ Year\_of\_realease, Platform, Genre /genres, positive\_ratings, negative\_ratings, EU\_Sales, JP\_Sales, Global\_Sales, Critic\_Score, User\_Score, Rating.

**3. Transformaciones.**

Transformación en tabla Indicadores del desarrollo humano mundial del Banco Mundial:

* Cambiar el nombre de la tabla a Indicadores Banco Mundial.
* Trasponer.
* Cambiar nombre de las columnas y filas.
* Colocar el tipo de dato correcto para todas las columnas
* Dejar las columnas y filas necesarias para el análisis.

Transformación en tabla Tablas Juegos en Steam + Video Game Sales:

* Unir las tablas por los campos Game Name/name. Hay muchos juegos repetidos por lo que se quitan duplicados.
* Se realiza una transformación para que Release\_Date quede como una fecha únicamente de año. Después de esto, se analiza que Year\_of\_realease tiene un total de 65% de datos vacíos, por lo que se decide eliminar esta columna. Idealmente se debería hacer un trabajo estadístico como, por ejemplo, ver que tanto coinciden los datos de las dos columnas, pero este análisis no se realiza por falta de tiempo.
* Podemos ver que genres de Juegos Steam clasifica los juegos en varias categorías, mientras que en la tabla Video Games Sales [Genre] sólo se clasifica en una. Por razones de simplificación, la idea sería etiquetar los juegos con la menor cantidad de categorías posibles.

Lo primero que hacemos es concatenar las columnas genres y Genre. A esta nueva columna le aplicamos un DISTINCT para ver cuántas combinaciones diferentes hay y en total son 588 combinaciones. Cabe recalcar que hay muchos géneros en la tabla genres que realmente no encajan en los géneros tradicionales (Accounting, Audio Production, Casual, Design &Illustration, Early Access, Free to Play, Photo Editing, Software Training, Video Production, Web Publishing ). Teniendo en cuenta que la columna genres tiene, como máximo, tres tags, separamos esta columna en tres diferentes y remplazamos todos estos géneros no tradicionales en cada una de estas columnas y, al hacer esto, vemos que la columna 3 tiene un 91% de datos vacíos, por lo que se descarta, teniendo en cuenta que la columna 1 y 2 tienen 9% y 65% vacío. Se hace lo mismo con la columna Genre, en este caso sacando los géneros Misc y Platform, dejando esta columna con un total de 70% vacío.

Todas estas columnas modificadas se unen y volvemos a aplicar un DISTINCT para ver que se ha reducido a 168 combinaciones. Separamos de nuevo y se hacen las siguientes transformaciones con el fin de generalizar géneros: RPG 🡪 Role-Playing, Gore 🡪 Violent. Hasta este punto, cuando unimos de nuevo, tenemos una totalidad de, máximo, tres géneros por juego, pero nos damos cuenta de que hay un problema de repeticiones como por ejemplo (Strategy Strategy Action) o (Role-Playing Role-Playing Role-Playing). Se crea, e incluye, un script de Python(Genres\_Scripts) donde se resuelve esto. Una vez hecho esto, por último, nos quedamos con los siguientes tipos repeticiones (‘Action,Indie’ | ‘Indie, Action’); se crea, e incluye, un script de Python(Genres\_Scripts2) donde se solventa esta situación.

Después de toda la transformación, se pasan de 588 géneros a 90, reduciendo esto un 85% de la transformación.

* Colocar el tipo de dato correcto para todas las columnas
* Se crea una columna %(Positive Ratings/Total Ratings) dividiendo (Positive Ratings\*100/ (Positive Ratings \* Negative Ratings))
* EU Sales, JP Sales, Global Sales: Aquí notamos que, en algunas filas, EU Sales + JP Sales > Global Sales, lo cual no es posible y, para realizar un buen análisis, no debemos tener en cuenta los datos con esta particularidad. Para esta transformación tomamos EU Sales/ Global Sales, después creamos una columna condicional donde todos los valores que sean mayores a 1 sean null y, por último, multiplicamos esta columna por Global Sales para tener la columna corregida; hacemos lo mismo con la columna JP Sales.

Para adivinar las unidades vemos que toda la franquicia Grand Theft Auto ha vendido aproximadamente 6148 millones. Revisando los datos, para alcanzar algo similar a esta cifra, serían (Millones de US$).

* User Score: Hay un dato ‘dtb’, lo remplazamos por un vacío y se multiplica por 10 para poder ser comparado con Critic Score.
* Price: Después de hacer una pequeña investigación por mercado libre, se puede ver que el precio medio de los juegos ronda los $US 40. Para que este precio tenga coherencia con los datos hay que dividirlo entre 10.
* Rating:

Las clasificaciones son E: para todos público, E+: mayores de 10 años, T: adolescentes M: Mayores de 17 años.

Para hacerlo comparable con las categorías de la tabla Indicadores del desarrollo humano mundial del Banco Mundial, los agrupamos así ( (E,E+,T): 0-15 años, M: 15 años o más). Se hace el respectivo cambio en la tabla de indicadores de desarrollo.

* Cambiar nombre de las columnas.
* Dejar las columnas y filas necesarias para el análisis.

**4. Análisis de Datos Económicos.**

Hacer el ejercicio de conocer el estado de una economía significa analizar una cuantiosa cantidad de indicadores, por lo que es bueno aclarar que, si bien este no es un ejercicio exhaustivo, revisar tres hitos importantes como los son el INB, crecimiento del PIB e inflación, nos puede al menos proporcionar una imagen general.

Como se mencionó previamente, el INB Per Cápita (US$) es el valor de todos los bienes y servicios producidos por los residentes de un país durante un período, en este caso, un año. Podemos ver que, tanto el INB de USA, EU como el de JP, se encuentran por arriba del general del mundo, así que, apegándonos a la definición del indicador, podríamos estar tentados a asumir que estos países han producido bienes y servicios en una mayor cantidad que el resto del mundo; sin embargo, cuando vemos el Aumento del PIB (%anual), al menos en lo que va desde el siglo XXI antes de la pandemia, la tendencia de estas naciones es equiparable a la del mundo.

Cabe recalcar que el Aumento del PIB de JP es siempre menor al mundial y que, en el caso del año 2008, donde el Aumento del PIB del mundo es negativo y, tanto el de JP, EU, como el de USA son, incluso, más negativos que este. Sólo un poco de historia, en el año 2008 la economía global dio un sacudón que la sumergió en una de las cinco peores crisis financieras debido al desplome del sector inmobiliario de los Estados Unidos.

Entonces, ¿por qué el crecimiento del INB per Cápita (US$) es mayor, tanto en Estados Unidos, EU y JP, con respecto al mundo? Ante esta pregunta, debemos considerar que el INB per Cápita es simplemente el PIB dividido por la cantidad de personas en la región especificada. El producto interno bruto, o PIB, es un indicador económico que refleja el valor monetario de todos los bienes y servicios finales producidos por un territorio en un determinado periodo de tiempo. Por un periodo de tiempo, y lo recalco, porque puede que estas naciones, en general, sí hayan creado más bienes y servicios que el resto de los países en el mundo, pero esta diferencia abismal entre el INB de estas naciones y el resto del mundo se debe principalmente, al menos en mi opinión y lo que puedo ver en los datos, a un esfuerzo abismal por parte de estas naciones durante años, más de los que se muestran en nuestros datos, por controlar la inflación.

La inflación es cuando se produce un incremento general de los precios al comparar dos periodos de tiempo. Afecta a bienes y servicios y, con el tiempo, reduce el valor de la moneda; es decir, con la misma cantidad de dinero se pueden pagar menos cosas. Es interesante ver en el gráfico de inflación que, en el caso de JP con respecto al mundo, existe deflación; por otra parte, USA y EU mantienen su inflación, como máximo, igual a la del mundo y, en su mayoría, por debajo de esta. ¿Y esto qué significa? Bueno, al ser la inflación una medida que, a muy grandes rasgos, cuantifica la disponibilidad de un bien y servicio con respecto a la demanda en una determinada región, significa que estas regiones manejan ese frágil equilibrio bastante bien y es este manejo de décadas lo que ha hecho que sus bienes y servicios valgan más con respecto a los del mundo.

De este análisis podemos concluir que las economías de JP, EU y USA, son economías fuertes, sanas y que saben cómo mantenerse en el tiempo; pero ¿esto realmente afecta en que podamos vender más juegos en estas regiones? No hay suficientes datos para saberlo; por ejemplo, si hubiera datos de países en desarrollo podríamos compararlos con los que tenemos de países desarrollados y realizar un análisis más detallado para saber en qué regiones enfocarse. A pesar de esto, revisando el gráfico de poblaciones, la población de JP, USA y EU con respecto al mundo es insignificante, siendo la de JP, en promedio, el 20% de la de EU; sin embargo, es interesante notar que el total de ventas de videojuegos de JP, al menos en los años revisados, es el 50% del total de ventas de EU; y, lo más interesante de todo, es que el INB de JP es mayor que el de EU. ¿Es esto una señal de que el INB puede influir en la venta de juegos o que simplemente los japoneses consumen más videojuegos que los residentes de la Unión Europea? De nuevo, no tenemos los datos suficientes para sacar una conclusión.

**Análisis por Género.**

La idea con este breve análisis es revisar algunos hitos de los diferentes géneros de videojuegos.

Como podemos ver, los géneros Action, Action Adventure, Action Shooter, Role-Playing, Action Role-Playing, Action Adventure son géneros que, comparados con el resto, generan mayores ingresos y, además, estos juegos generan mayores ganancias cuando son diseñados para un público adulto, a excepción de Action Adventure en la cual la distribución está divida, aproximadamente, mitad y mitad entre público adulto y niños/adolescentes. Por otro lado, géneros como Violent Shooter, Adventure Indie Puzzle y Action Sports son géneros que no generan una cantidad de ventas considerable. Cabe recalcar que hay géneros como Adventure Indie que están dedicados, casi en su totalidad, a menores de 14 años.

El que un videojuego genere mayores ganancias tiene que ver con dos variables: cantidad de videojuegos lanzados de un género específico y costo de cada copia. Al examinar las cantidades de copias vendidas, podemos ver que géneros como Action y Action Adventure son géneros muy utilizados al crear juegos, por lo que tiene sentido que estos géneros generen grandes ganancias, pero, por otra parte, géneros como Action Shooter, Role-Playing y Action Role-Playing son géneros que generan buenas ganancias pero que, al contrario de los anteriores, la cantidad de juegos lanzados de estos géneros es mucho menor. Esto presenta un mar azul para los inversores donde pueden entrar a un mercado de buenas ganancias sin tanta competencia. Existen otros géneros como Strategy, Action Fighting, Racing, Adventure Strategy que, si bien no generan una cantidad equiparable de dinero comparado con los géneros anteriormente mencionadis, siguen siendo mercados interesantes porque generan una cantidad de dinero considerable e, igualmente, la cantidad sacada de juegos relacionados a estos géneros es baja, es decir hay poca competencia. Géneros como Violent Shooter, Adventure Indie Puzzle y Action Sports además de generar pocas ganancias, también son géneros de los que no se sacan muchos videojuegos y, esto, aunque, en general, puede ser visto como algo malo, también puede significar que en estos géneros no se han sacado buenos juegos que atraigan la atención del público y, prueba de esto es que, al revisar el gráfico de calificaciones, ninguno de esto géneros sobrepasa el 75% en ninguna de las calificaciones, por lo que un juego bien diseñado, con una temática interesante que pueda interesarle a los usuarios, podría ser también una buena opción, aunque también arriesgada. Es interesante ver que géneros como Indie y Action Indie que ocupan un buen porcentaje de géneros utilizados para crear videojuegos, generen pocas ganancias.

Al revisar el promedio de precio por género, este nos podría dar una indicación de qué tan costoso es crear un videojuego de cierto género. Es interesante ver que los géneros de juegos que son más costoso, como Violent Simulation Sports o Racing Simulation Sports, son géneros que producen pocas ganancias y de los cuales pocos juegos se sacan; además sus críticas son menores al 60% por lo que son géneros que no gustan tanto al público y son costosos de realizar. Géneros como Action, y Action Adventure, se encuentran por debajo del promedio de costo mientras que Action Shooter, Role-Playing y Action Role-Playing son un poco más costosos que el promedio; el Score de estos géneros, en general, también se encuentran por debajo del 60%, lo cual es un poco contraintuitivo con el hecho que sean los géneros que más venden. Géneros que crean pocas ganancias como Violent Shooter, Adventure Indie Puzzle y Action Sports en general, tienen un costo igual o mayor al promedio. Lo géneros que tienen menor costo, como Strategy Shooter o Massively Multiplayer Simulation, son géneros de los cuales pocos juegos se sacan, generan pocas ganancias y, además su score es menor al 75%; sin embargo, géneros como Adventure Puzzle, comparte la mayoría de estas características, a excepción de que la crítica a este es bastante buena, por lo que puede ser un género interesante que gusta a los usuarios y que cuesta poco hacer. Hablando de géneros de pocas ganancias y pocos juegos generados relacionados a estos, es interesante ver géneros como Game-Development o Action Indie Role-Playing, son géneros que, en general, gustan al público, por lo que podría haber una demanda que no se encuentra cubierta en géneros como estos.

**Análisis Etario**

Como podemos ver en los gráficos, en general la tendencia alrededor del mundo es que la población “menor de 14 años” es mucho menor que la población de “14 años o más” alrededor del mundo, siendo que JPN y EU son regiones donde, comparadas con el mundo, la tasa de nacimiento es menor que el promedio.

Es curioso que la proporción de (venta videojuegos para “menores de 14 años”)/( venta videojuegos para “14 años o más”) sea similar a (población de “menores de 14 años”)/ (población de “14 años o más”). En el caso de JPN el enfoque es muchísimo mayor para vender videojuegos con clasificación “14 años o más”, siendo el de “menores de 14 años” apenas el 18% del total de ventas; por otro lado, esta misma variable en EU es del 30% y en el mundo es del 25%. El porcentaje de población de JPN menor de 14 años es del 13%, mientras que esta variable en EU es de15% y en el mundo del 27%. Es curioso notar como, tanto en JPN como en EU, el porcentaje de venta de videojuegos aptos para menores de 14 años es mayor que el porcentaje total de esta población en las diferentes regiones, por lo que, comparado con el mundo, nos puede dar una pista que en estas regiones se consume más este tipo de juegos que en el resto del mundo, donde el porcentaje de venta de videojuegos aptos para “menores de 14 años” es similar al porcentaje total de población “menor de 14 años”.