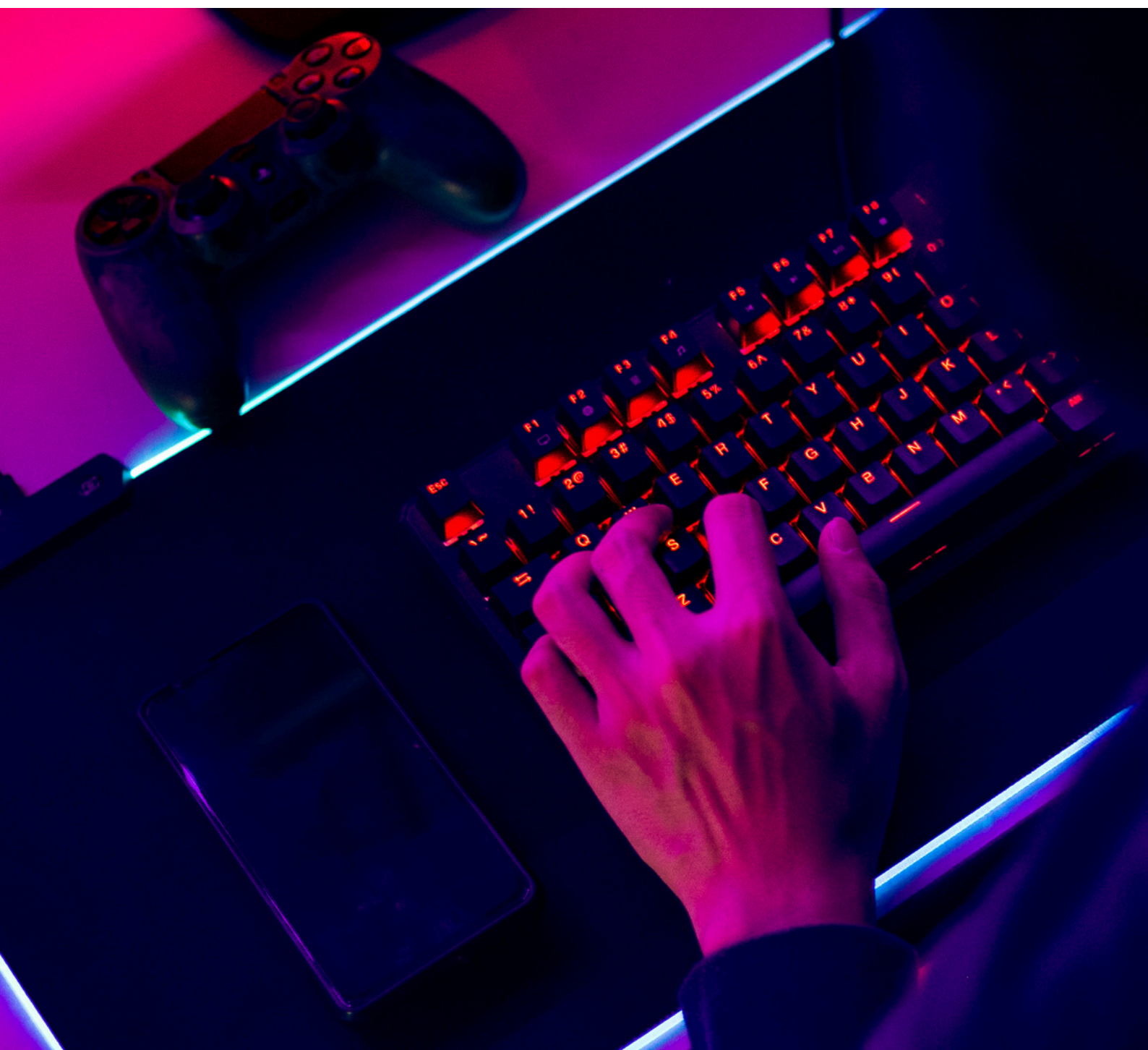


Análisis Exploratorio de Datos (EDA): El comportamiento del jugador en juegos online



ÍNDICE

01 Introducción

01.A Contexto

01.B Hipótesis

02 Exploración de los datos

03 Análisis Global de los datos

03.A Análisis Univariante, Bivariante y
Multivariante

04 Análisis específico

05 Conclusiones

01 Introducción

En este proyecto vamos a analizar el comportamiento del jugador en el ámbito de los juegos online. Para ello hemos tomado como referencia una muestra de datos de Kaggle.

Fuente:

Rabie El Kharoua. (2024).

 Predict Online Gaming Behavior Dataset [Data set].

Kaggle. <https://doi.org/10.34740/KAGGLE/DSV/8742674>.

<https://www.kaggle.com/datasets/rabieelkharoua/predict-online-gaming-behavior-dataset>

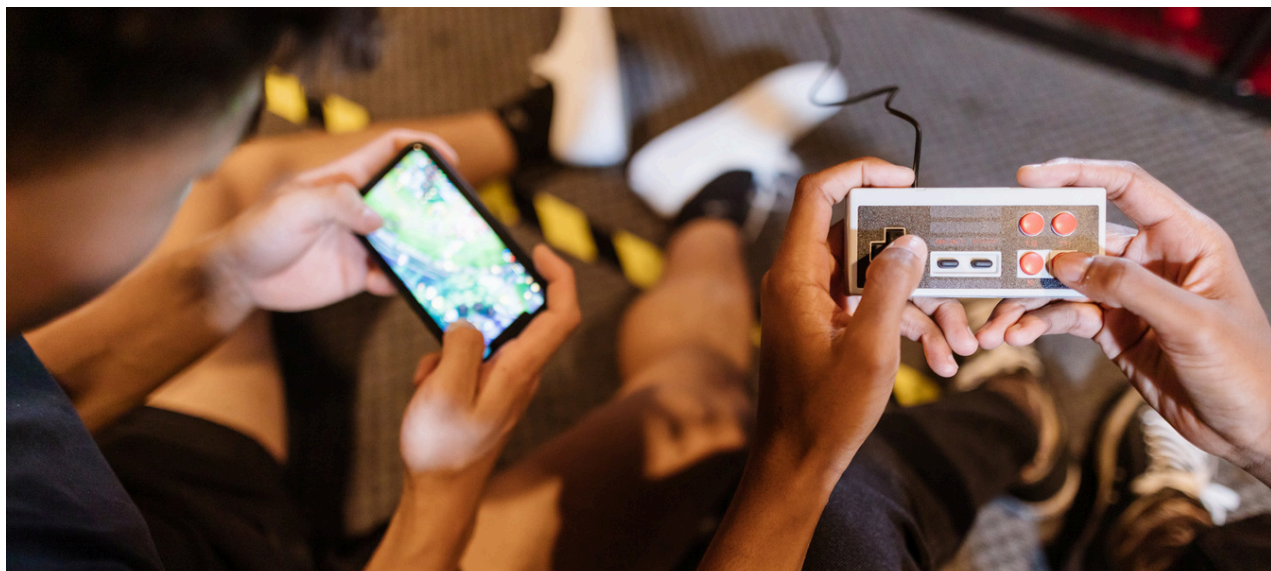
01.A Contexto

Se parte de este dataset (online_gaming_behavior_dataset.csv) que contiene una muestra de 40034 entradas y 13 columnas. Aparentemente el dataset ya ha sido tratado y acondicionado por el autor, listo para realizar análisis a partir del mismo. En un punto más avanzado del análisis se filtra el dataframe y se elimina la columna "PlayerID".

El estudio realizado comienza por un análisis global de los datos en pos de encontrar información acerca de cómo influyen las variables recogidas en el comportamiento de compra y nivel de engagement del jugador. Encontramos datos muy homogéneos y condensados, lo que ha dificultado encontrar diferencias significativas a la hora de relacionar variables. Por este motivo, se decide hacer una análisis más específico aplicando filtros al dataset y segmentando la muestra, con el objetivo de poder encontrar relaciones significativas y respuestas a nuestras preguntas e hipótesis.

Para responder a las cuestiones planteadas inicialmente y generar un storytelling para nuestra presentación, nos planteamos el hipotético caso en el que se busca orientar la toma de decisiones para crear un videojuego online, que integre compras dentro del mismo.

A partir de aquí surgen las siguientes preguntas e hipótesis.



01.B Hipótesis

“Los jugadores con mayor engagement tienden a realizar más compras dentro del juego”

- ¿Qué variables influyen en la decisión de compra?
- ¿Existe algún género que concentre más compras?

“El nivel de engagement influye en el tiempo de juego dedicado”

- ¿Qué géneros de juego generan mayor engagement?
- ¿En qué género de juego se concentra un mayor número de inicios de sesión a la semana?
- ¿Son más populares los juegos que tienen mayor engagement?

Además de estas preguntas e hipótesis surgen otras durante el análisis y se abren nuevos caminos de estudio.

02 Exploración de los datos

Aplicando los métodos pertinentes vemos de forma general la naturaleza del dataset.

- Información sobre datos nulos. No presenta.
- Frecuencia de aparición de los valores y valores únicos. Se observa que las frecuencias de aparición de los valores de algunas variables son muy similares. En las variables que consideramos numéricas hay pocos valores únicos en comparación con el total.

Tipificación de los datos y clasificación de las variables:

- Clasificamos las variables teniendo en cuenta su cardinalidad, porcentaje de cardinalidad y tipo de dato.
- Se vuelve a observar que la cardinalidad de las variables numéricas no es muy elevada.
- Decidimos ignorar la variable PlayerID dado que presenta una cardinalidad del 100%. Prestamos atención a la variable PlayTimeHours también por tener una cardinalidad del 100%.
- Marcamos las variables target, importantes e interesantes sobre las que llevar a cabo nuestro análisis.

Medidas de tendencia central:

- Analizamos la media, la moda y la mediana en el set de datos.
- Obtenemos una aproximación global del perfil medio del jugador y el tiempo que dedica a esta actividad.
- Observamos que no hay outliers en el dataset.

03 Análisis global de los datos

03.A Análisis Univariante, Bivariante y Multivariante

Realizamos el análisis univariante de las variables categóricas y numéricas.

Primero nos centramos en realizar un análisis bivariante sobre la variable directora **"InGamePurchases"** con el fin de encontrar relaciones de comportamiento de esta variable con respecto al resto. Tras analizar su relación y realizar los tests correspondientes (Test Chi-cuadrado y Test ANOVA) podemos descartar nuestra hipótesis inicial y confirmar que aparentemente las compras dentro del juego no se ven influenciadas por ninguna otra variable, es completamente independiente. Por tanto, interpretamos que los jugadores con mayor engagement no realizan más compras.

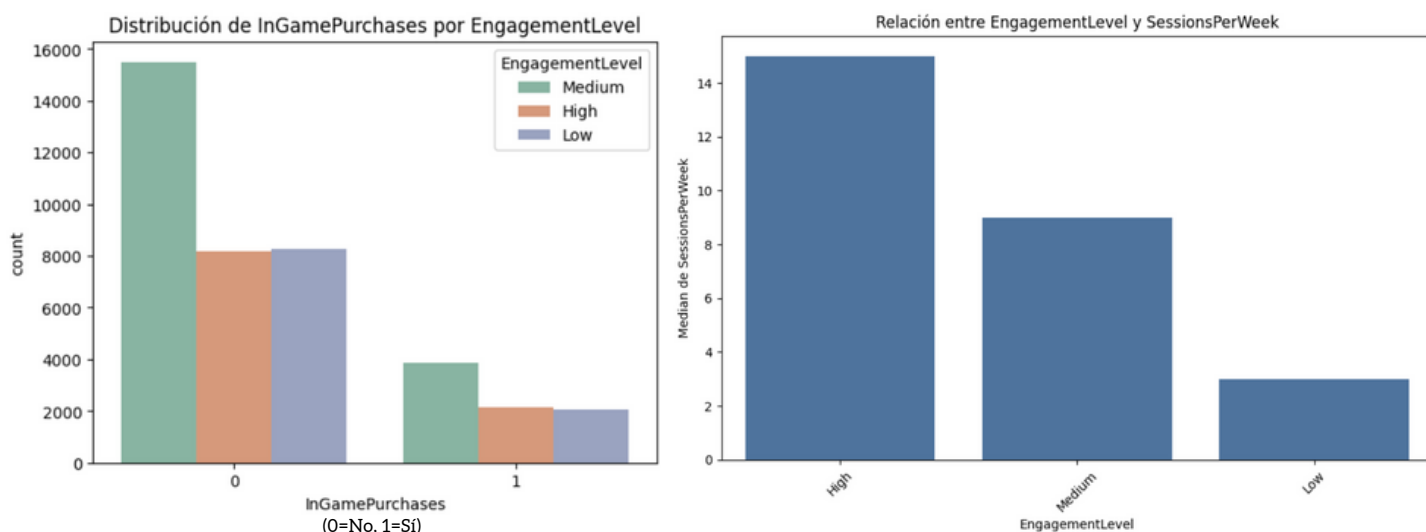
Esto nos abre un camino de actuación para realizar planes de marketing orientados a incentivar las compras.

Realizamos el mismo estudio poniendo el foco en la otra variable marcada como target **"EngagementLevel"**. Para este caso ponemos el foco en estudiar cómo se relaciona con el resto de variables que más nos interesan para responder a las preguntas planteadas, **"GameGenre"** y **"SessionsPerWeek"**.

No encontramos relación representativa entre el nivel de engagement y el género de juego, pero sí podemos comprobar que el número de inicios de sesión aumenta según el nivel de engagement. Es decir, los perfiles con un mayor nivel de engagement inician sesión con más frecuencia a la semana.

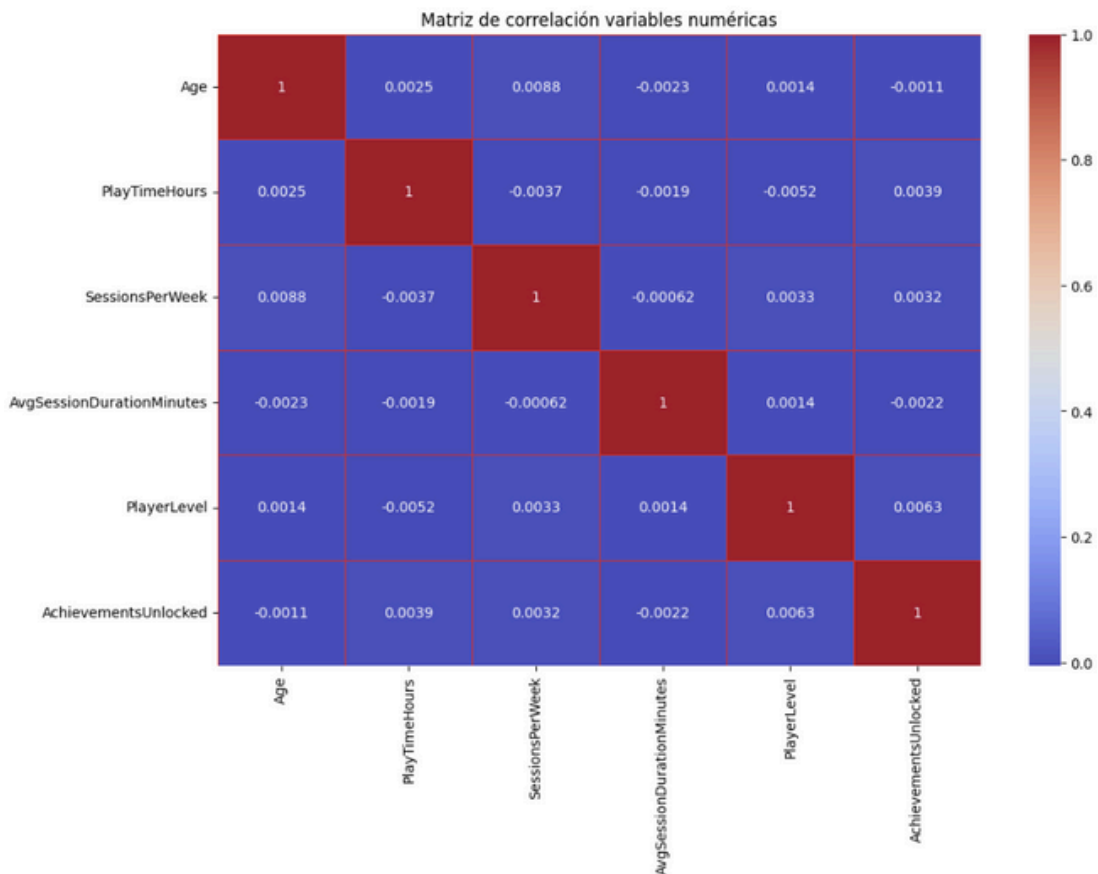
También comprobamos que aparentemente no existe una estrecha relación entre el nivel de engagement y el género de juego. Esto puede deberse a la cantidad de jugadores tan homogénea.

Estas observaciones nos ayudarán posteriormente a segmentar el dataset y realizar comprobaciones más específicas.



Analizamos las variables numéricas y generamos un mapa de calor con el objetivo de encontrar alguna correlación.

Observamos que los valores de correlación están muy alejados de -1 y 1 lo que dificulta poder relacionarlas entre sí.

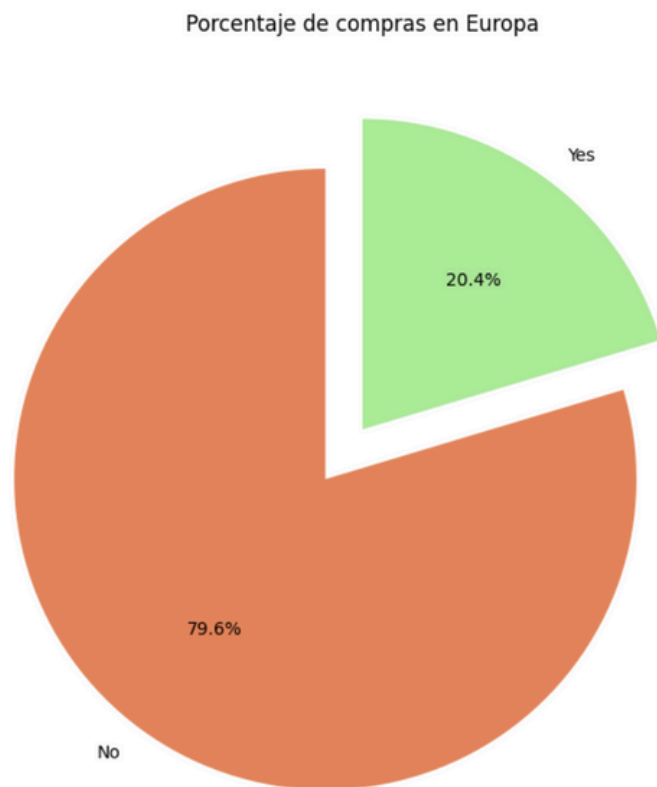
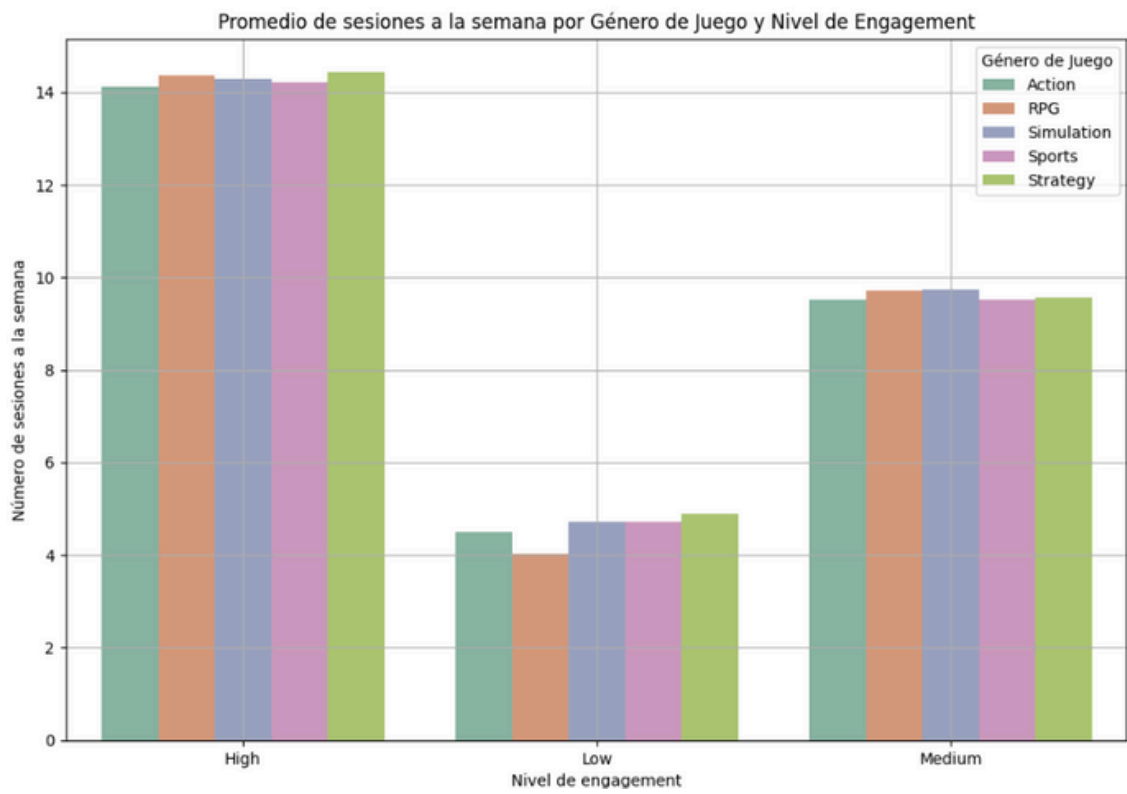


04 Análisis específico

Con el objetivo de responder de una forma algo más precisa a nuestras preguntas iniciales, vamos a centrarnos en el mercado Europeo y segmentar el dataset por un rango de edad en el que se encuentre la edad media del jugador.

Realizamos una copia del dataset, eliminamos la columna "PlayerID" y renombramos los valores de la variable "InGamePurchases" para que se muestren como No y Yes.

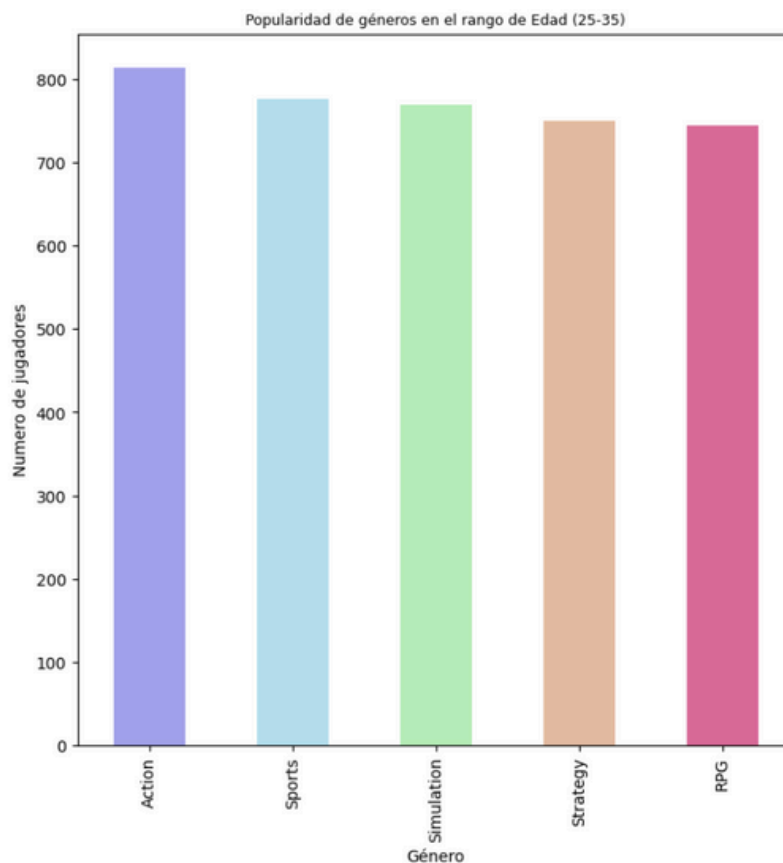
- Visualizamos el porcentaje global de compras en Europa comprobando que sólo el 20% de jugadores realiza compras dentro del juego.
- Comparamos número de inicio de sesiones, nivel de engagement y categoría de juego. Observamos que existe un mayor número de inicio de sesiones a la semana en perfiles con un alto nivel de engagement. Los géneros de juego en los que se presenta son estrategia y RPG.
- Posteriormente realizamos la misma comparación teniendo en cuenta el total de inicios de sesión semanales. Se puede apreciar que la mayoría de inicios de sesión se realiza en perfiles con un nivel de engagement medio.



Con el objetivo de ver datos más específicos, decidimos segmentar la muestra por la media de edad de los jugadores (32 años). Encontramos que el porcentaje con esta media de edad es muy pequeño con respecto al de la población total y por tanto nos centramos en un rango de edad de 25 a 35 años.

Para este rango, que supone el 32% de la muestra, buscamos los géneros más populares. Descubrimos que el género más popular es el de acción, seguido por muy poca diferencia de deportes.

Adicionalmente, calculamos el nivel de engagement para este género. El nivel medio de engagement es el más extendido en nuestro segmento. Con esta información podemos comparar con la tabla inicial, la frecuencia de inicios de sesión por semana.



05 Conclusiones

Respondiendo a nuestras hipótesis de partida:

- No existe una relación significativa entre las compras dentro del juego y el nivel de engagement del jugador.
- No se ha encontrado evidencia de que algunas de las variables de estudio influya positivamente en las compras.
- No se ha encontrado evidencia de que exista un mayor número de compras significativo según el género de juego.
- Ante estos resultados, podemos tomar una posición proactiva para aumentar las compras aprovechando la fortaleza de otras variables, por ejemplo midiendo la popularidad del juego, determinando el perfil medio del consumidor, tiempo invertido, etc.
- Aparentemente existe relación entre el nivel de engagement y el número de inicios de sesión a la semana.
- Los géneros de juego que generan mayor engagement y tienen mayor número de inicios de sesión a la semana son Estrategia y RPG, no necesariamente son los más populares. Se puede tomar como un nicho de mercado.
- Los géneros de juego más populares presentan un nivel de engagement medio. Deportes y Acción.
- Perfil de jugador medio:
 - 32 años.
 - Preferencia por juegos de Deporte y Acción.
 - 9,5 inicios de sesión a la semana.
 - Nivel de engagement medio.

Para un contexto hipotético en el que se busca optimizar la toma de decisiones para crear un videojuego online, que integre compras dentro del mismo, podemos decir lo siguiente:

- En general las compras no son representativas dentro de los juegos online, el porcentaje de jugadores que no compra es mucho mayor al que sí realiza compras.
- Para crear un juego que incentive a la compra y teniendo en cuenta que el engagement aparentemente no influye en la decisión de compra, podemos centrarnos en la popularidad del género más que en su alto nivel de engagement dado que será más probable que juegue un mayor número de personas.
- A partir de aquí podemos determinar el tipo de juego que queremos desarrollar, así como implementar estrategias de marketing efectivas que incentiven las compras dentro del juego.
- Por ejemplo:
 1. Tomando los inicios de sesión como oportunidades de impacto, se pueden implementar ofertas especiales como pases de batalla específicos o misiones y ventos especiales.
 2. Podemos segmentar mejor las campañas de marketing y orientarlas al jugador target.



Bootcamp Data Science
Blanca Novella Serrano
2024

Análisis Exploratorio de Datos (EDA):
El comportamiento del jugador en juegos online