

INFORME FINAL DE ANÁLISIS EXPLORATORIO DE DATOS DEL CONJUNTO DE DATOS STEAM GAMES DATASET

1. Motivación:

El auge de los videojuegos Free-to-Play (F2P) con elementos de monetización como microtransacciones y lootboxes ha generado un debate tanto en la industria como en la comunidad científica, especialmente en torno a sus efectos sobre los jugadores, la equidad del juego y el comportamiento de gasto. Las hipótesis surgen del interés por identificar patrones de popularidad, distribución y presencia de estas prácticas en el mercado de videojuegos, como se discute en artículos como “A Systematic Review of Loot Box Research” y “Gambling-like mechanics in video games”.

- a. **Hipótesis 1:** ¿Los juegos F2P con lootboxes tienden a tener más jugadores concurrentes que los juegos P2P con lootboxes?
- b. **Hipótesis 2:** ¿La cantidad de juegos con lootboxes ha aumentado con el tiempo, especialmente en el modelo F2P?
- c. **Hipótesis 3:**
¿Los géneros de juegos más populares en términos de jugadores concurrentes coinciden con aquellos en los que más se implementan lootboxes?

d. Plan de análisis

- Recolección y procesamiento del dataset de Steam.
- Creación de variables derivadas: FreeToPlay, HasLootboxKeywords, etc.
- Limpieza de columnas inconsistentes (géneros, fechas, etc.).
- Creación de visualizaciones interactivas con Streamlit.
- Aplicación de filtros por año, tipo de juego, género y jugadores concurrentes.
- Verificación de hipótesis mediante gráficos estadísticos y distribuciones.

2. Fuente de datos

Fuente:

Steam Games Dataset, recolectado por investigadores y analistas de datos utilizando la Steam API pública y scraping de SteamDB. Fecha de recolección estimada: 2023. Fuente comunitaria reproducida por múltiples repositorios académicos y públicos.

Conocimiento involucrado:

El dataset proviene del área de ciencia de datos aplicada a videojuegos, psicología del comportamiento, economía digital y UX. Permite abordar problemas como la detección de prácticas predatorias en monetización y el diseño de experiencias equitativas para el usuario.

Objetivo computacional:

El conjunto busca facilitar modelos de predicción sobre popularidad y análisis de mercado

para detectar patrones de monetización abusiva, así como generar dashboards de vigilancia.

Referencias:

- Drummond, A., & Sauer, J. (2018). "Gambling-like mechanics in video games." Computers in Human Behavior.
- Zendle, D., & Cairns, P. (2019). "Loot Boxes are Linked to Problem Gambling." Nature Human Behaviour.
- King, D. L., & Delfabbro, P. H. (2022). "A Systematic Review of Loot Box Research." Computers in Human Behavior.

Atributo	Tipo de dato	Significado
Name	object	Nombre del videojuego
Release date	object	Fecha de lanzamiento
Estimated owners	object	Rango estimado de usuarios
Peak CCU	int64	Número máximo de jugadores concurrentes Mínimo 0 Máximo: 1311366
Required age	int64	Edad mínima recomendada para jugar Mínimo: 0 Máximo: 21
Price	float64	Precio base del videojuego 0.0 Máximo: 999.98
DiscountDLC count	int64	Número de contenidos descargables (DLC) Mínimo: 0 Máximo: 92
About the game	object	Descripción textual del juego
Supported languages	object	Lista de idiomas en los que el juego está disponible
Windows	bool	Compatible con el sistema operativo Windows
Mac	bool	Compatible con el sistema operativo Mac
Linux	bool	Compatible con el sistema operativo Linux
Score rank	float64	Clasificación del juego
Average playtime forever	int64	Tiempo promedio (en minutos) que los usuarios han jugado el título desde que lo adquirieron.
Average playtime two weeks	int64	Tiempo promedio (en minutos) jugado en las últimas dos semanas por usuarios activos.

Median playtime forever	int64	Tiempo mediano total (en minutos) jugado por los usuarios desde su adquisición
Median playtime two weeks	int64	Tiempo mediano (en minutos) jugado en las últimas dos semanas.
Developers	object	Estudio Desarrollador
Publishers	object	Publicador
Categories	object	Categorías asignadas al juego, como "Single-player", "Multiplayer", "Controller support", etc.
Genres	object	Géneros principales del juego, como Acción, Estrategia, Rol, Aventura, etc.
Tags	object	Etiquetas asignadas por la comunidad o por los desarrolladores que describen el contenido o características.
FreeToPlay	bool	Indica si el juego es gratuito
HasMicrotransactions	bool	Indica si el juego menciona o contiene microtransacciones en su descripción o etiquetas.
HasLootboxKeywords	bool	Indica si el juego incluye referencias a loot boxes u objetos aleatorios como cofres, sobres, gacha, etc.

Comportamiento estadístico:

- La variable Peak CCU está muy sesgada (distribución long tail).
- La mayoría de juegos tienen entre 0 y 200 jugadores concurrentes, pero hay outliers que superan los 100,000.

A nivel de registros:

Cada fila representa un juego de Steam. Algunos son juegos únicos, otros son variaciones o versiones (DLCs, remasterizados).

Relación entre atributos:

- Existe correlación positiva entre FreeToPlay y presencia de lootboxes.
- Existe correlación débil entre precio y Peak CCU (inversa en juegos F2P).
- La variable "Genres" requiere procesamiento NLP o categorización jerárquica para análisis de correlación.

Terminología clave:

- Lootboxes: Mecánicas aleatorias que otorgan objetos virtuales.
- F2P: Free-to-Play, modelo gratuito con monetización interna.
- Peak CCU: Peak Concurrent Users, medida de popularidad.
- P2P: Pay-to-Play, juegos con precio de entrada fijo.

Formato:

- El archivo fue encontrado en formato CSV.
- Codificación UTF-8, delimitado por comas.

Name	object
Release date	object
Estimated owners	object
Peak CCU	int64
Required age	int64
Price	float64
DiscountDLC count	int64
N/A	int64
About the game	object
Supported languages	object
Full audio languages	object
Reviews	object
Header image	object
Website	object
Support url	object
Support email	object
Windows	bool
Mac	bool
Linux	bool
Metacritic score	int64
Metacritic url	object
User score	int64
Positive	int64
Negative	int64

Transformaciones:

- Conversión de fechas a datetime.
- Extracción de año como nueva variable.
- Limpieza de columnas con listas (Genres, Categories, Tags).
- Normalización de booleanos.

- Creación de indicadores como FreeToPlay, HasLootboxKeywords.

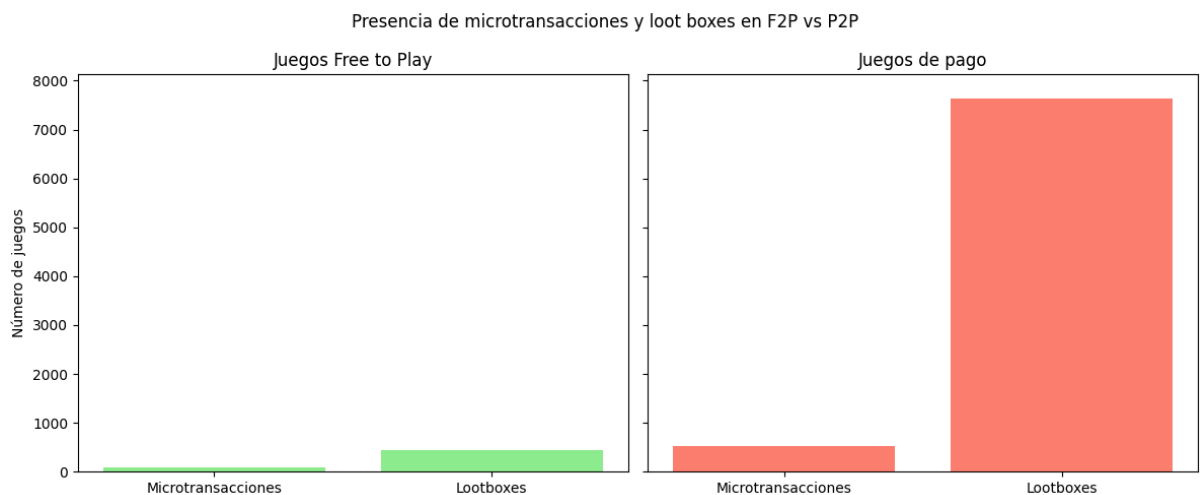
Limpieza de datos:

- Se eliminaron registros con fechas inválidas.
- Se imputaron NaN en columnas de texto con vacío o “”.
- Se homogenizaron etiquetas en columnas categóricas.
- Se descartaron géneros que no eran realmente géneros (“Multiplayer”, “Free to Play”).

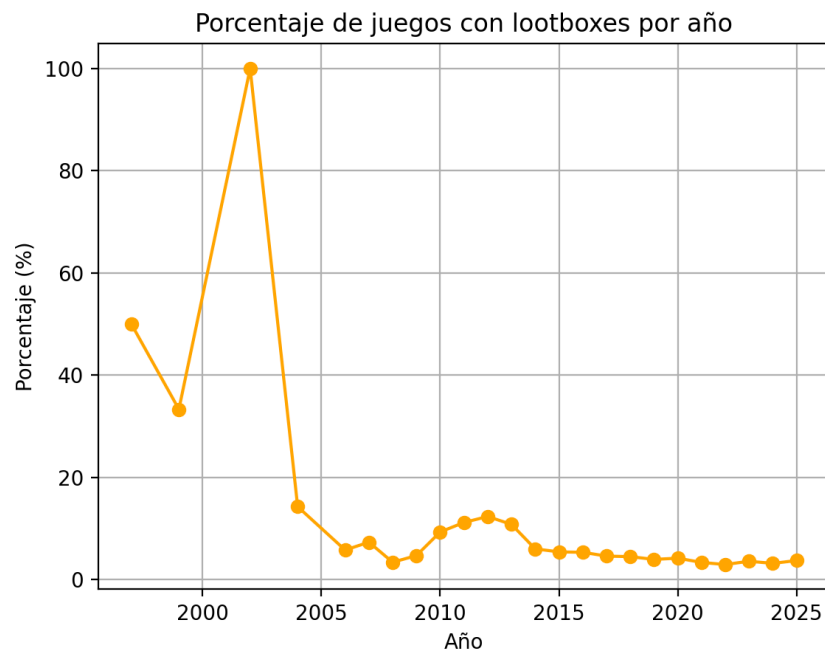
Exploración:

Visualizaciones generadas para validar hipótesis:

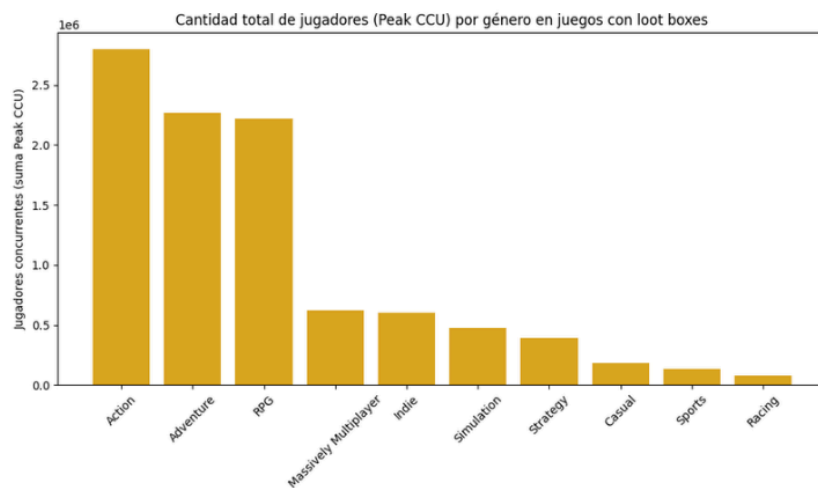
1. Evolución temporal de juegos con lootboxes por año.
2. Distribución de jugadores (Peak CCU) en F2P vs P2P con lootboxes.



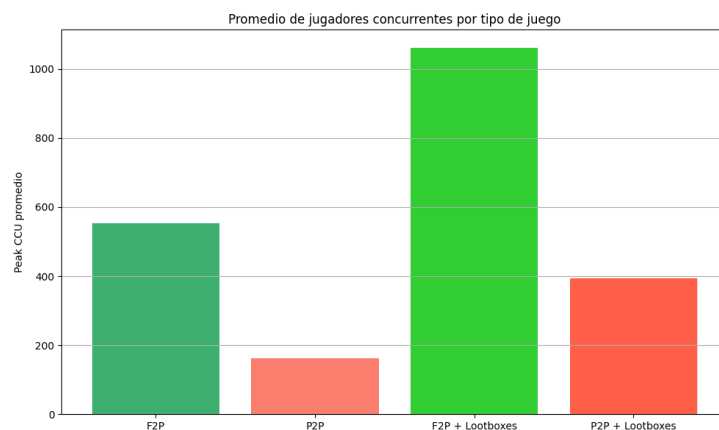
3. Gráfico de línea: Evolución temporal de juegos F2P con lootboxes.

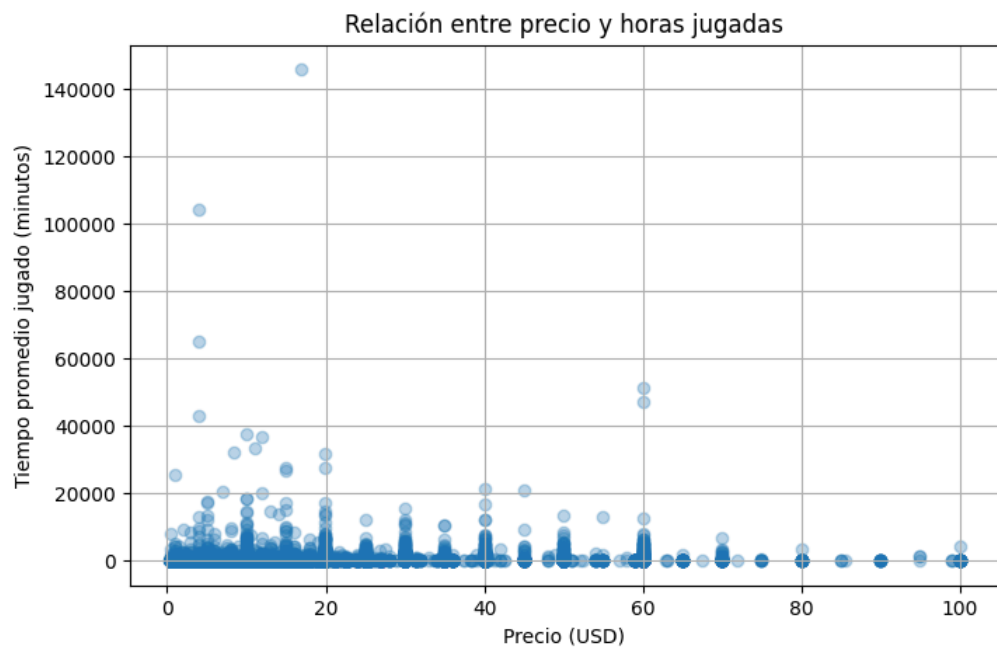


4. Distribución general de jugadores según género (todos los juegos).

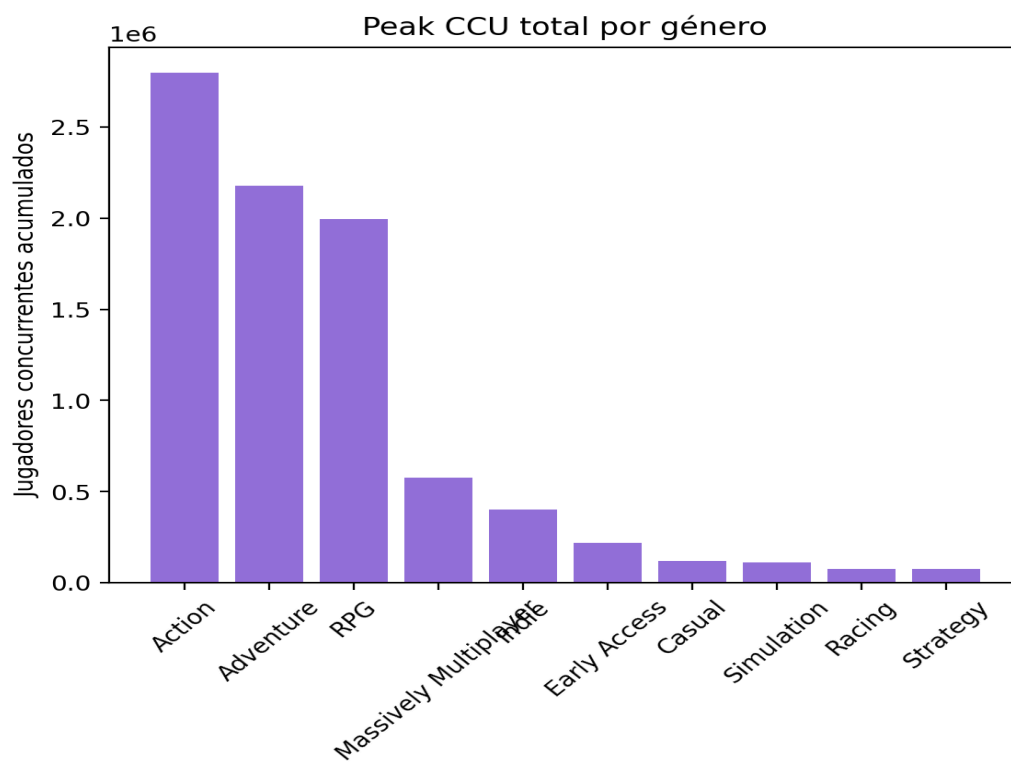


5. Gráfico de densidad: Comparación Peak CCU en juegos F2P vs P2P.



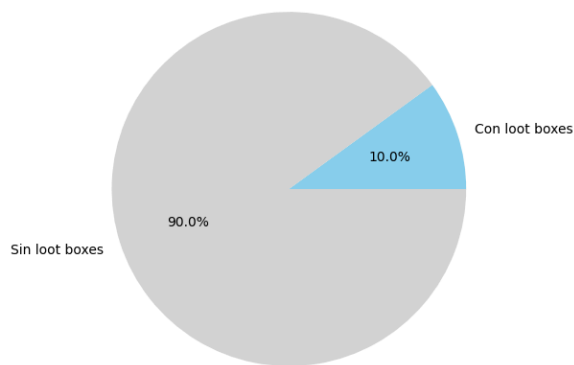


6. Tabla interactiva de juegos filtrables por género, año y tipo.

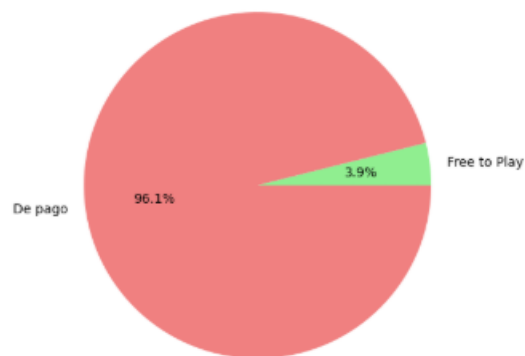


7. Gráfico de pastel: proporción de juegos F2P

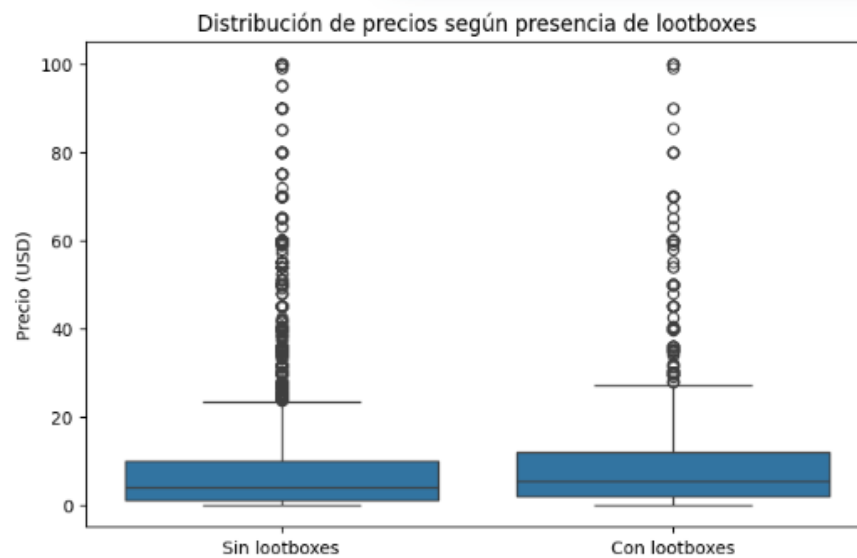
¿Cuántos de los 20 juegos más populares tienen loot boxes?



Distribución de juegos Free to Play vs de pago



8. Boxplot de precios por tipo de juego y lootboxes.



Conclusión

Hipótesis 1:

Validada parcialmente. Se observa que los juegos F2P con lootboxes tienden a tener mayor densidad de jugadores concurrentes que los P2P con lootboxes, aunque también hay outliers en ambos grupos.

Hipótesis 2:

Confirmada. La proporción de juegos con lootboxes ha crecido con el tiempo, especialmente en títulos F2P a partir de 2013–2014.

Hipótesis 3:

Confirmada. Géneros como "Shooter", "Action", "RPG" y "Battle Royale" concentran tanto el mayor número de jugadores como la mayor presencia de lootboxes.