

Reporte Fase 1: Descubrimiento del Proyecto

1. VALIDACIÓN DEL MODELO DE RENTABILIDAD Y HOUSE EDGE EN UN CASINO EN LINEA

2. Objetivo del Proyecto

Verificar que el modelo de juego de la plataforma garantice un ingreso operativo positivo y predecible para cada ronda, independientemente de los resultados individuales de los usuarios

3. Descripción del Problema

En la industria de los casinos en línea, el principio fundamental que garantiza la sostenibilidad financiera es el llamado “House Edge” o ventaja de la casa. Se trata de una ventaja matemática implementada en el diseño de cada juego, que asegura que a largo plazo y en volúmenes grandes de apuestas, el casino o “La casa” siempre tendrá el casino deberá SIEMPRE obtener un beneficio y ser predecible.

La ventaja para el casino es un hecho matemático estático. En este casino, la ventaja es un resultado emergente y dinámico de la interacción entre el algoritmo de la ruleta y el perfil de riesgo de los apostadores en cada ronda.

Ahora bien, el modelo de este casino donde los jugadores apuestan sobre un coeficiente en ascenso y las ganancias están dadas en base a la predicción, la ventaja no parece ser un porcentaje fijo y evidente. En su lugar esta ventaja es dinámica y está determinada por el algoritmo interno. Esto genera un problema para verificar la verdadera rentabilidad del negocio, debido a que no es posible confiar únicamente en modelos teóricos, ya que la ventaja podría estar influenciada por variables inesperadas, sesgos en el comportamiento de los usuarios o incluso por la participación de cuentas internas como lo son los moderadores.

El proyecto busca descubrir, promediando miles de rondas, cuál es ese porcentaje de ventaja real que el casino logra imponer, así obteniendo la rentabilidad financiera de

este negocio y confirmar que efectivamente el casino genera un ingreso operativo positivo y predecible para como cualquier casino para que este pueda ser viable

4. Recursos Disponibles

Tecnología y Herramientas

- **Jupyter Notebook:** Se utilizará como el **entorno de desarrollo** principal. Permite la integración inmediata de visualizaciones con los resultados además de la creación de un documento “dinámico” facilitando el análisis
- **Python:** Se empleará como el lenguaje de programación central para todo el flujo de trabajo, manipulación de datos.
- **Pandas:** Esta biblioteca será la columna vertebral para la **manipulación, limpieza y transformación** de los datos transaccionales. Su uso será crítico para:
 - Cargar la base de datos o dataset en estructuras de Data Frames
 - Filtrar agrupar y agregar los datos para realizar análisis según sea la columna
- **Matplotlib:** Estas bibliotecas se utilizarán para la **visualización de datos y el análisis exploratorio**. Su uso me permitirá:
 - Generar histogramas y gráficos de densidad para entender la distribución del House Edge y el volumen de apuestas.
 - Hacer diagramas de caja (boxplots) para identificar valores atípicos y comparar las métricas clave entre juegos con y sin moderadores.
 - Visualizar correlaciones y patrones ocultos en los datos que sustenten las conclusiones cuantitativas.

Datos

- **ID:** Identificador único para cada ronda de juego, que servirá como llave primaria para el análisis.
- **Players:** Número total de usuarios participantes en una ronda específica, un indicador directo de la actividad y “engagement” o retención.
- **Items :** Representan, el volumen de artículos virtuales apostados y su **equivalente** monetario en dólares estadounidenses (USD).
- **Money:** Variable clave que representa el Ingreso Bruto en Riesgo o el total apostado en la ronda.
- **Ticks:** El coeficiente o multiplicador final determinado por el algoritmo de la casa para esa ronda. Es la variable central que define el resultado financiero.
- **PeopleWin:** El monto total (en USD) de ganancias netas para los jugadores en la ronda
- **PeopleLost:** El monto total (en USD) que los jugadores perdieron. Esta variable es sinónimo del Ingreso Bruto del Casino proveniente de las apuestas perdedoras.
- **Outpay:** El total de dinero (en USD) que el casino desembolsó a los ganadores. Representa las Obligaciones de Pago Totales o el costo de ventas para el casino.
- **Time:** Marca temporal precisa (formato YY-MM-DD hh-mm) que permite realizar análisis de series de tiempo, tendencias y estacionalidad.
- **Moderator:** Variable binaria que indica la participación de un moderador en la ronda, actuando como un control para evaluar la integridad y normalidad de los resultados.

5. Hipótesis Iniciales

1. **Hipótesis 1:** El House Edge (margen de la casa) no es constante, sino que varía de forma predecible en función del comportamiento agregado de los apostadores. Específicamente, en rondas con un volumen total de apuestas (money) más alto, el House Edge efectivo será mayor. La lógica de negocio es que

el algoritmo puede estar diseñado para capitalizar la mayor disposición a apostar de los jugadores en momentos de alta actividad.

2. **Hipótesis 2:** Las rondas en las que participa un moderador presentan un House Edge significativamente diferente al de las rondas normales. La dirección de esta diferencia (mayor o menor) revelaría si el rol del moderador es el de estimular la actividad (con un edge más bajo para hacer ganar a los jugadores) o el de garantizar la rentabilidad (con un edge más alto para corregir pérdidas atípicas del casino).
 3. **Hipótesis 3:** Existe una preferencia conductual sistemática por parte de los jugadores hacia coeficientes de apuesta elevados (ej: >1.5). Esta sobrevaloración de las probabilidades de ganar grandes premios conlleva una tasa de pérdida significativamente mayor para estas apuestas, lo que resulta en un House Edge efectivo más alto para el casino en las rondas donde predominan las apuestas de alto riesgo."
- La **Hipótesis 1** Busca validar la **eficacia del modelo** y su capacidad para maximizar ganancias cuando la demanda es alta.
 - La **Hipótesis 2** Evalúa la integridad y el control del sistema, verificando si una variable interna altera el campo de juego.
 - La **Hipótesis 3** Si se confirma el casino no debería modificar activamente el algoritmo

6. Definición de Stakeholders Clave

Partes Interesadas en el Proyecto y su Rol

1. Finanzas y Auditoría Interna

- **Rol Principal:** Garantizar la integridad financiera y la precisión del modelo.
- **Interés en el Proyecto:**

- Validar que el House Edge calculado se alinea con las proyecciones financieras y presupuestos.
- Confirmar que no existen anomalías o desviaciones en los ingresos que puedan indicar fallos en el sistema o prácticas fraudulentas.
- Utilizar el House Edge verificado para mejorar la precisión de los pronósticos de ingresos futuros.

2. Equipo de Desarrollo de Producto y Juego

- **Rol Principal:** Diseñar, mantener y mejorar la plataforma y la mecánica del juego.
- **Interés en el Proyecto:**
 - Comprender cómo el comportamiento de los usuarios (ej: apostar a coeficientes altos) impacta en la rentabilidad.
 - Recibir datos para equilibrar el juego: si el House Edge es demasiado alto, los jugadores pueden dejar la plataforma; si es demasiado bajo, el negocio no es sostenible.
 - Investigar la influencia del "moderador" en el sistema para asegurar su correcto funcionamiento.

3. Equipo de Marketing y Adquisición de Usuarios

- **Rol Principal:** Atraer nuevos jugadores y retener a los existentes.
- **Interés en el Proyecto:**
 - Identificar los perfiles de jugadores más valiosos (ej: los que apuestan a coeficientes altos, los que participan frecuentemente).
 - Diseñar campañas de retención dirigidas a jugadores que experimentan pérdidas (persecución de pérdidas), ofreciendo bonos o incentivos para mantenerlos en la plataforma.

- Utilizar el conocimiento sobre la estacionalidad y horarios pico para optimizar el timing y el presupuesto de las campañas publicitarias.

4. Equipo de Cumplimiento Normativo y Riesgos

- **Rol Principal:** Asegurar que la plataforma opera dentro del marco legal y gestionar riesgos operativos.
- **Interés en el Proyecto:**
 - Verificar que la ventaja de la casa, aunque dinámica, es transparente y no constituye una práctica engañosa.
 - Auditar la participación de los moderadores para asegurar que no se está manipulando el juego de manera fraudulenta.
 - Gestionar el riesgo reputacional asociado a un House Edge excesivamente alto que podría ser percibido como injusto por la comunidad.

Se busca que:

- Finanzas: pronostique mejor.
- Producto y juego: optimice el juego.
- Marketing: retenga mejor.
- Cumplimiento: mitigue riesgos.
- La Dirección: dirija con estrategia.

7. Preguntas Clave

Categoría 1: Rentabilidad y House Edge

1. ¿Cuál es el House Edge promedio global de la plataforma y cómo se distribuye (media, mediana, desviación estándar)?
2. ¿El House Edge es consistentemente positivo en todas las rondas, confirmando la viabilidad del modelo de negocio?
3. ¿Existe una tendencia temporal (por hora, día, mes) en la evolución del House Edge?

4. ¿Cuál es el valor esperado (EV) por dólar apostado para el jugador promedio en esta plataforma?

Categoría 2: Comportamiento del Jugador

5. ¿El volumen total de apuestas en una ronda está correlacionado con el House Edge resultante?
6. ¿Existe una relación entre el número de jugadores y la rentabilidad de la ronda?
7. ¿Las rondas con un coeficiente de apuesta promedio más alto (inferido) generan un House Edge mayor, evidenciando el sesgo hacia el riesgo alto?
8. ¿Se observa el patrón de "persecución de pérdidas", donde un pico en el House Edge de una ronda es seguido por un aumento en el volumen de apuestas en la siguiente?

Categoría 3: Mecánica del Juego

9. ¿La participación de un **moderador** influye de manera estadísticamente significativa en el House Edge o el volumen de apuestas?
10. ¿El coeficiente final de la casa (ticks) muestra una distribución aleatoria o existen valores donde se detiene con más frecuencia?
11. ¿Existe una relación entre el ticks de la ronda y el volumen de apuestas o el número de jugadores de la ronda anterior?

Categoría 4: Análisis de Segmentación y Valor

12. ¿Un pequeño porcentaje de rondas de alto volumen concentra la mayoría de las ganancias del casino?
13. ¿Se pueden identificar "eventos extremos" (rondas con pérdidas excepcionales para la casa) y cuáles son sus características comunes?

14. ¿Las rondas que ocurren en horarios pico (noches, fines de semana) tienen un perfil de rentabilidad diferente al de los horarios valle?

Categoría 5: Pronóstico y Optimización del Negocio

15. Con el House Edge histórico, ¿cuál es el ingreso predecible para un volumen de apuestas proyectado?
16. ¿Qué variables (players, money, moderator, etc.) son los mejores predictores del House Edge de una ronda?
17. ¿Podemos construir un modelo que clasifique las rondas en "de alta rentabilidad" y "de baja rentabilidad" antes de que ocurran?
18. ¿Cuál es la frecuencia y el impacto financiero de las rondas donde el casino tiene un House Edge negativo (pierde dinero)?
19. ¿Cuál es el House Edge "ideal" que maximiza las ganancias sin disuadir a los jugadores?
20. ¿Qué métricas en tiempo real se podrían monitorear para alertar sobre desviaciones en la rentabilidad esperada del sistema?

8. Fuentes de Datos Identificadas

Fuente Primaria Principal:

1. Dataset Transaccional del Casino: Registros históricos completos de todas las rondas de juego, obtenidos directamente del socket en línea del casino. Esta es la fuente de verdad única y contiene todas las variables necesarias para el análisis cuantitativo.

Datos Específicos por variables:

- **Cálculo del House Edge y Valor Esperado:**
 - **Variable:** money (Volumen total apostado) y outpay (Total pagado a ganadores).
- **Análisis de Relación Volumen-Rentabilidad:**
 - **Variable:** money (Volumen de apuestas) y el House Edge calculado.
- **Análisis del Perfil de Riesgo del Apostador:**
 - **Variable:** outpay (Total pagado) y peopleLost (Ingreso bruto del casino).
- **Auditoría de la Influencia de Moderadores:**
 - **Variable:** moderator (Indicador de participación) y el House Edge calculado.
- **Análisis de Series de Tiempo y Estacionalidad:**
 - **Variable:** time (Marca temporal) y el House Edge calculado.
- **Detección de Patrones Conductuales Secuenciales:**
 - **Variable:** time (para ordenar secuencialmente las rondas), money y el House Edge calculado.

Nota: Este proyecto se basa exclusivamente en los datos transaccionales internos del casino. No se utilizan fuentes externas como encuestas o datos de mercado, lo que convierte este análisis en una auditoría interna basada puramente en el rendimiento operativo registrado. La limpieza y preprocesamiento de esta fuente única será una etapa crítica del proyecto.

9. Justificación del Proyecto

Importancia del Análisis

La operación de un casino en línea, como cualquier negocio, se sustenta en un modelo de rentabilidad claro y sostenible. En este contexto, el House Edge (ventaja de la casa) no es un concepto abstracto, sino el principal motor financiero del negocio. Sin embargo, cuando este modelo se basa en una mecánica de juego no tradicional y dinámica, la incertidumbre sobre su funcionamiento real introduce riesgos operativos y estratégicos significativos.

Este análisis cuantitativo es de vital importancia porque convierte una suposición teórica en una certeza medible. Validar empíricamente el House Edge permite a la organización transitar de la especulación a la gestión basada en datos. En la práctica, esto se traduce en tres impactos críticos:

1. **Verificación del Modelo de Negocio:** Confirma que la plataforma es financieramente viable al demostrar que genera un ingreso operativo positivo y predecible en el largo plazo. Esto es fundamental para justificar inversiones futuras y asegurar la sostenibilidad de la empresa.
2. **Optimización de la Estrategia Operativa:** Al identificar los factores que influyen en la rentabilidad como el comportamiento de los jugadores, los horarios pico o el efecto de funciones internas como los moderadores el análisis proporciona una hoja de ruta para maximizar los ingresos. El departamento de producto puede ajustar la mecánica del juego, y el equipo de marketing puede diseñar campañas más efectivas dirigidas a los segmentos de jugadores más valiosos.
3. **Gestión Proactiva de Riesgos:** El estudio sirve como una auditoría interna que mitiga riesgos. Permite detectar si existen anomalías, sesgos inesperados o posibles vulnerabilidades en el algoritmo que podrían estar regalando ganancias a los jugadores o, por el contrario, exponiendo a la compañía a riesgos reputacionales por una ventaja percibida como injusta.

En esencia, este proyecto no solo responde a la pregunta académica de "¿Cuál es la ventaja de la casa?", sino que sienta las bases para una toma de decisiones estratégica informada, asegurando que el casino no solo funcione, sino que prospere de manera eficiente y controlada.

10.- ¿Cuántos datos y que tipo son?

El análisis se sustentará en un conjunto de datos masivo, compuesto por aproximadamente 55,000 registros. Cada registro representa una transacción financiera individual, correspondiente a una ronda de juego completada en la plataforma del casino.

En términos de clasificación, se trabajará con datos estructurados de naturaleza mixta:

1. **Datos Numéricos:**
 - a. **Enteros:** ID, players, ticks.

- b. **De punto flotante (continuos):** items, money, peopleWin, peopleLost, outpay.

2. Datos Temporales:

- a. La variable time permite transformar el dataset en una serie temporal para análisis diarios, horarios y de estacionalidad.

3. Datos Categóricos/Nominales:

- a. La variable moderator actúa como un indicador binario (presencia o ausencia), fundamental para los análisis comparativos.