



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
DEPARTAMENTO DE CIENCIA DE LA COMPUTACIÓN
IMT2200 - INTRODUCCIÓN A LA CIENCIA DE DATOS

Propuesta de Proyecto - IMT2200

Carlos R. Matías V. Benjamín U. Brian M.

11 de septiembre de 2025

”Análisis descriptivo y predictivo de los juegos de casino, haciendo énfasis en el jugador promedio”

Contexto y Motivación

Se hará un análisis exploratorio de los datos, pasando por la búsqueda de clustering y patrones, como también por la predicción de éstos mismos, todo esto tomando como foco principal el ”jugador promedio” de los diferentes juegos de los casinos que se estudiarán en los datasets presentados posteriormente. La idea central es dar poder de acción a potenciales o actuales jugadores a través del conocimiento del comportamiento de los datos presentados, siempre basándose en un fuerte sustento estadístico que lo respalde. A pesar de que el proyecto estará basado en la idea central del ”jugador promedio”, se hace también hincapié en capturar y generar información relevante para cualquier interesado en el sector. Se contempla, el uso del crecimiento económico del sector analizado, ingresos fiscales, dinámicas de consumo, aporte de impuestos, entre otros patrones y relaciones que surgen del análisis sistemático de la oferta del sector de juegos de Casinos.

Como motivación, se ha investigado en diversos artículos y un estudio estadístico realizado por la Asociación Chilena de Casinos de juego (”Radiografía de la apuesta en Chile al 2023”, Caracterización del jugador online), debidamente referenciadas al final de este documento. Lo que nos da indicios del creciente mercado de Casinos en Chile, tanto online como presenciales, y sus diferencias en la regulación de ganancias.

La audiencia objetivo incluye: **Jugadores ocasionales o asiduos**, que buscan información para realizar una mejor inversión de sus interacciones con los casinos de juegos, **Entidades gubernamentales**, para velar por una adecuada regularización y recaudación, **Trabajadores**

del sector, que tengan interés en el estudio para fines de, por ejemplo, optimización, **Analistas económicos**, con interés en los patrones que surgen del estudio.

Objetivos

1. **Descriptivo:** Caracterizar la oferta de juegos, los ingresos y los impuestos asociados.
2. **Predictivo:** Construir modelos de regresión para anticipar ingresos y retornos de máquinas en función de la oferta y las visitas.
3. **Clustering:** Identificar grupos de establecimientos o periodos con comportamientos similares en ingresos, impuestos y visitas.
4. **Analítico:** Explorar la relación entre la oferta de juegos, la demanda de usuarios y la recaudación fiscal.
5. **Accionable:** Proponer indicadores que permitan a operadores y reguladores tomar decisiones estratégicas.

Como se ha mencionado antes, nuestro proyecto contempla una variedad de potenciales interesados en aprovechar la información entregada, para poder tomar decisiones informadas basadas en resultados estadísticos de los datos.

Datos

Se han utilizado los datasets "Boletín estadístico de la Superintendencia de Casinos de Juego a noviembre de 2012", cuyo enlace está referenciado al final de este documento, que consisten en 8 ficheros en formato CSV, principalmente con variables de tipo numéricas y de tipo string, como también strings categóricos, así que no haremos una distinción fuerte entre strings y strings categóricos:

- **OfertaJuegos.csv:** Contiene variables como el nombre del casino, comuna, número de mesas (ruleta, cartas, dados), número de máquinas de azar, número de posiciones de bingo. Los tipos de dato son categóricos y numéricos (enteros)

Contiene la ofertas de juegos por categoría en los casinos en operación hasta Noviembre del año 2012, el problema es que solo es de ciertos casinos en particular y no de todos, hace falta una interpretación de una columna en la que se debería de especificar que significa cada int (las ultimas 2 columnas). También se especifica el juego y se le asigna valores, además se especifica la región en la que se ubica el casino.

- **IngresosBrutosJuego.csv:**

Contiene el ingreso bruto de cada casino registrado hasta la ultima actualización. Necesita limpieza y remover algunas columnas.

Los tipos de datos son categóricos y también contiene unas columnas de datos numéricos (float)

-
- **Impuestos.csv:** Contiene cada casino registrado en Chile hasta la última actualización (25 de enero del 2013), por cada mes del año 2012, falta limpieza, corregir errores y eliminar filas repetidas.

Los tipos de datos son categóricos y numéricos (float).

- **PosicionesJuego.csv:** posiciones de juego por categoría, con el "WIN diario por posición de juego (\$)" según categoría hasta noviembre de 2012, en pesos chilenos y dólares. Requiere limpieza y mejor orden.

Contiene tipos de datos numéricos float, como también entero, y strings categóricos.

- **ResumenIndustria.csv:** Un dataset que contiene un poco de todo, contiene el ingreso bruto de los juegos, Impuesto específico al juego (20%), Iva al juego, Total de visitas, Impuesto por entradas, gasto promedio por visita, Retorno real a clientes respecto a máquinas de azar, y los mismos datos en dólares. Esta ordenada por mes del año 2012. Es el resumen de lo más importante de los csv, juntando toda la información de los casinos.

Contiene tipos de dato numérico entero, específicamente usado para valores monetarios en tipo de moneda CLP y USD, también datos de tipo numérico decimal, y strings categóricos.

- **RetornoMaquinas.csv:** Es el monto total apostado en las máquinas por casino y por mes del año 2012, tanto en porcentaje como en CLP, con info adicional. No necesita limpieza pero sí orden y conciencia de cómo están separados los datos y donde inicia cada tabla (Así como el resto de csv's).

- **Visitas.csv:** Contiene tanto la cantidad de visitas por casino como el impuesto por entradas y el gasto promedio de cada visita por mes del año 2012.

Contiene tipos de datos numéricos enteros y punto flotante, como también strings categóricos.

- **ParqueMaquinas.csv:** Número de máquinas de azar por fabricante y procedencia hasta Noviembre del 2012.

Contiene datos string de tipo categóricos, como también enteros y float.

Cada dataset incluye series temporales y variables de interés económico. La integración permitirá un análisis completo de la industria.

Preguntas de Investigación

1. **¿Cuál es el retorno real promedio a clientes en máquinas de azar?** Averiguamos la duda principal de cualquier cliente tentado a apostar, cuanto dinero es retornado en base a los montos apostados.
2. **¿Cuál es el monto total apostado en juegos de azar?** Averiguamos los montos apostados por los clientes mediante los meses.
3. **¿Cuáles son los ingresos brutos de juegos de azar?** Averiguamos los ingresos brutos en juegos de los casinos en Chile.
4. **¿Cuál es la oferta de juegos distribuida por casinos?** Averiguamos la capacidad de los juegos de cada casino por categoría.
5. **¿Cuál es el resumen de la industria de casinos de juegos en el periodo de estudio?** Averiguaremos el resumen que incorpora la ganancias en bruto por juego, impuestos específicos y de valor agregado, gasto promedio por visitas, entre otros valores que demuestren las ganancias de los casinos mes a mes.
6. **¿Que similitudes presentan algunos casinos?** Finalmente podemos identificar patrones que se repitan entre los diferentes casinos, comparar su capacidad, ingresos y distribución de juegos. (Extra)
7. **¿Origen de máquinas de azar?** Como extra podemos averiguar la procedencia mediante fabricante y país de origen, de las máquinas de azar. (Extra)

Estas preguntas se desarrollarán enfocándose en el diseño tentativo explicado más abajo, que incluye relaciones entre variables, un modelo predictivo y descriptivo, así como también la clasificación, agrupación (clusters), y la búsqueda de patrones relacionados con las preguntas formuladas.

Diseño Tentativo

- **Análisis exploratorio:** basado en visualización de los datos con el uso de herramientas estadísticas descriptivas.
- **Clustering (k-means):** para identificar agrupamiento de patrones de comportamiento, de tipos de juego y/o establecimientos. También se abre la oportunidad para utilizar otro tipo de algoritmo de clustering, según los tipos de datos lo requieran.
- **Machine Learning, Regresión lineal/polinómica:** para predecir posibles patrones de comportamiento en el jugador promedio, como también de ingreso y ganancias, entre otros.
- **Correlaciones entre los datos:** relación entre la afluencia de visita a ciertos juegos, el volumen de flujo de dinero, ingresos e impuestos, entre otros.
- **Métricas de evaluación:** métricas como R^2 , MAE, MSE y análisis de varianza, para poder sustentar estadísticamente los modelos de predicción.

1. Referencias

1. Repositorio del proyecto en GitHub. <https://github.com/ReToxiE/Proyecto-IMT-2200-2-Grupo-5>
2. Boletín estadístico de la Superintendencia de Casinos de Juego (*a noviembre de 2012*). <https://datos.gob.cl/dataset/3985/resource/febd82fb-2b87-4cef-aebf-46ede878a769>
3. Wikipedia - Casino. <https://es.wikipedia.org/wiki/Casino>
4. Artículo 1 - Los ingresos de los casinos de Chile crecen un 2% en mayo. <https://www.gamingintelligence.com/es/finanzas/balances/68865-chile-mayo/>
5. Artículo 2 - Crece el mercado del juego en Chile. <https://www.theclinic.cl/2024/10/16/poker-con-dinero-real-en-su-version-online/>
6. Estudio - “Radiografía de la apuesta en Chile al 2023, ”Caracterización del jugador online”. <https://accj.cl/wp-content/uploads/2023/06/ppt-estudio-radiografia-apuesta-online-2023.pdf>
7. GitHub - Matías V. <https://github.com/aesmatias>
8. GitHub - Carlos R. <https://github.com/Ex-0m>
9. GitHub - Benjamín U. <https://github.com/ReToxiE>
10. GitHub - Brian M. <https://github.com/brianmunozaylwin>