



Ciencia de datos aplicada

Taller 1: Demanda de ocupación hotelera

Francisco Alejandro Santamaría
Dept. de Ingeniería Eléctrica y Electrónica
Universidad de los Andes
Bogotá D.C, Colombia
f.santmaria@uniandes.edu.co

I. PREÁMBULO. DEFINICIÓN DEL CONTEXTO DE NEGOCIO.

Como consultor científico de datos especializado en el sector hotelero, el objetivo de estos entregables es ayudar a cadenas de hoteles a optimizar estrategias de ocupación, enfrentando la creciente competencia y la variabilidad en la demanda para tomar decisiones acertadas sobre precios, cancelaciones y reservas, comprendiendo además los factores que determinan la duración de la estadía, la probabilidad de cancelación y la influencia de variables como el tipo de hotel, la composición del grupo de huéspedes o el canal de reservación, todo ello clave para garantizar ingresos sostenibles y el éxito del negocio.

Este entregable se divide en cuatro secciones, siendo éstas:

1. Reporte Inicial: Elemento que resume dimensiones y tipos de datos del dataset, destacando los 5 atributos más relevantes con un análisis univariado básico de su comportamiento o distribución.
2. Estrategia de análisis: Descripción breve de la estrategia de análisis y técnicas justificadas, desde estadísticos básicos hasta visualizaciones multivariadas, para reducir cancelaciones y mejorar la ocupación hotelera.
3. Desarrollo de la estrategia: Implementar la estrategia definida mostrando el procesamiento de datos, las técnicas estadísticas o de visualización aplicadas y los insights obtenidos de los resultados.
4. Generación de resultados: Redactar un informe ejecutivo o presentación breve con recomendaciones basadas en datos sobre factores clave que permitan reducir cancelaciones y mejorar la ocupación hotelera.

II. REPORTE INICIAL

En una primera aproximación de los datos, podemos decir lo siguiente:

1. Dimensiones del dataset: 58895x33

2. Tipos de datos: object, int y float
3. Número total de cancelaciones: 24, 229
4. Datos relevantes:

- a) hotel: diferencia entre urbano y resort.
- b) is-canceled: variable clave para medir cancelaciones.
- c) lead-time: anticipación de la reserva, influye en la probabilidad de cancelación.
- d) stays-in-weekend-nights: distingue viajes de ocio vs negocios.
- e) stays-in-week-nights: distingue viajes de ocio vs negocios.

Los criterios bajo los cuales se escogieron los cinco elementos más interesantes fueron su relevancia para el negocio hotelero, su capacidad de explicar patrones de ocupación y cancelación, y en mantener un balance entre variables categóricas y numéricas.

Algunos resultados obtenidos, en relación a gráficos univariados, fueron los siguientes:



Figura 1: Gráfico variable "hotel"

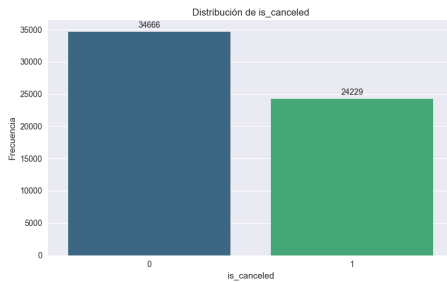


Figura 2: Gráfico variable "is-canceled"

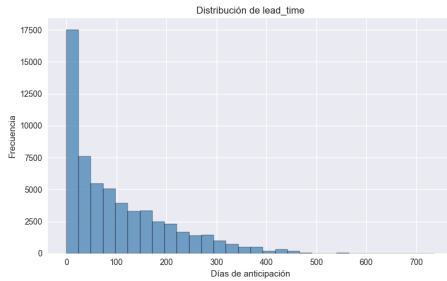


Figura 3: Gráfico variable "lead-time"

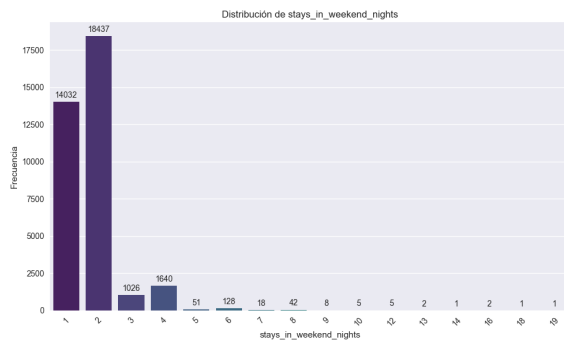


Figura 4: Gráfico variable "stays-in-weekend-nights"

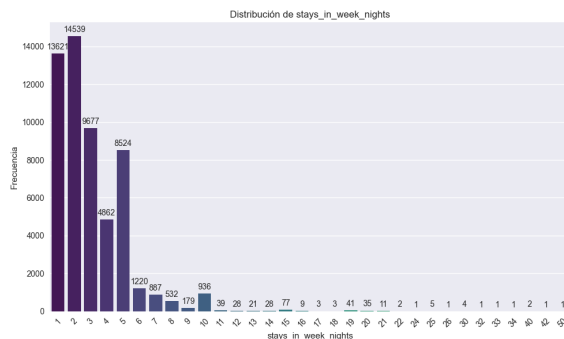


Figura 5: Gráfico variable "stays-in-week-nights"

III. ESTRATEGIA DE ANÁLISIS

Para abordar el problema de cancelaciones y ocupación hotelera, la estrategia de análisis comenzará con la preparación

de los datos, garantizando la correcta interpretación de variables categóricas y numéricas, así como la detección de valores atípicos y datos faltantes. Posteriormente, se realizará un análisis descriptivo mediante estadísticos básicos y visualizaciones que permitan identificar tendencias generales, como la relación entre el tipo de hotel, la anticipación de la reserva (lead-time) y la duración de la estadía. Estas primeras exploraciones proporcionarán un panorama de los factores que más inciden en las cancelaciones y en el comportamiento de la demanda. Vale mencionar que no se hará una depuración o remoción de datos, según lo recomendado por el profesor de la clase.

En la siguiente etapa se aplicarán técnicas de análisis bivariado y multivariado para explorar correlaciones entre atributos clave. Se emplearán gráficos de barras y de pie para comparar distribuciones, así como tablas cruzadas y análisis de correlación para identificar relaciones significativas. Esta estrategia permitirá descubrir patrones útiles para sugerir acciones concretas, como diseñar incentivos para reservas con mayor probabilidad de cancelación, con el fin de reducir cancelaciones y optimizar la ocupación.

IV. DESARROLLO DE LA ESTRATEGIA. PARTE 1

A continuación, aplicaremos la estrategia de análisis por varios puntos, desarrollando cada uno específicamente. La secuencia del desarrollo se hará de la siguiente manera:

1. Se realizarán estudios bi-variable, buscando la relación entre las cancelaciones y:
 - a) El tipo de estadía escogida
 - b) El tipo de cliente que reserva
 - c) El número de días de fin de semana reservados
 - d) El número de días entre semana reservados
2. Se harán revisiones multivariable, según los hallazgos del punto anterior, buscando algún patrón estructurado que pueda explicar las cancelaciones de los usuarios y como podría reducirse dicho comportamiento.

IV-A. Análisis del ratio de cancelaciones según el tipo de estadía

Para esta revisión, tomaremos el conjunto de datos *is - canceled*, que solo otorga 0 y 1, como variable dependiente; y compararlo con el conjunto de datos *hotel*, que nos indica el tipo de hotel.

Ahora bien, calculamos el promedio de *is - canceled* por cada tipo de hotel, obteniendo el siguiente resultado:



Figura 6: Ratio de cancelaciones según tipo de estancia

Notemos que la mayor proporción de cancelaciones corresponde a los hoteles en la urbe.

IV-B. Análisis del ratio de cancelaciones según el tipo de cliente

A continuación, pasaremos a hacer el mismo análisis del subcapítulo anterior, pero ahora con respecto al tipo de cliente. Recordemos que, gracias a la base de datos, podemos encontrar cuatro tipos de clientes:

1. Transient: Huéspedes ocasionales (reservas individuales).
2. Transient-Party: grupos de huéspedes reservando juntos.
3. Contract: Reservas bajo contrato corporativo o acuerdos con empresas/agencias.
4. Group: Reservas de grupos grandes.

En este sentido, extraemos para cada tipo de cliente su ratio de cancelaciones, con el deseo de observar qué tipo de cliente cancela más. Los resultados obtenidos fueron los siguientes:

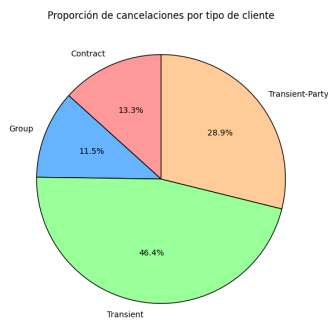


Figura 7: Ratio de cancelaciones según tipo de cliente

Notemos que la mayor proporción de cancelaciones corresponde a los huéspedes ocasionales.

IV-C. Análisis del ratio de cancelaciones según la cantidad de días de fin de semana reservados

Ahora bien, pasaremos al estudio de la cancelación presentada según los días de fin de semana y los días entre semana, empezando por los días de fines de semana. Para

ello, nos volvemos a apoyar en el dato binario *is - canceled*, obteniendo los siguientes resultados:

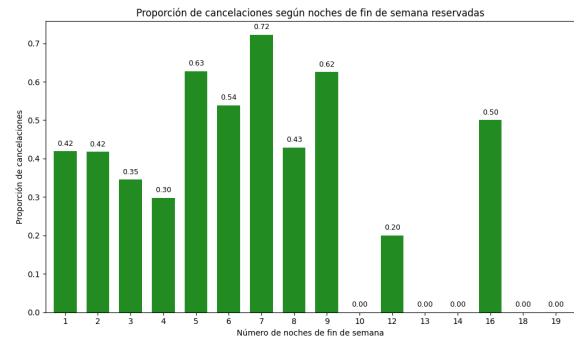
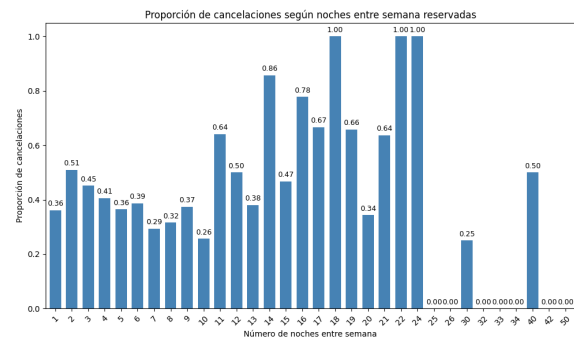


Figura 8: Ratio de cancelaciones según tipo de cliente

Para este escenario, podemos observar que los valores de entre 5 y 9 días reservados, el ratio de cancelación es el más alto, en comparación con sus homónimos.

IV-D. Análisis del ratio de cancelaciones según la cantidad de días entre semana reservados

En consonancia con los dos análisis anteriores, ahora veremos cuál es el ratio de cancelaciones con respecto a la cantidad de días entre semana reservados. Usando la misma lógica, obtuvimos los siguientes resultados por cada número de día:



En este sentido, lo más apropiado sería abordar los comportamientos encontrados, haciéndolo de la siguiente manera:

1. Realizar un análisis entre las cancelaciones, el tipo City Hotel y:
 - a) El tipo de cliente Transient
 - b) Las reservas de días en fines de semana
 - c) Las reservas de días entre semana
2. Realizar un analisis entre las cancelaciones, tipo de cliente Transient y:
 - a) Las reservas de días en fines de semana
 - b) Las reservas de días entre semana

V-A. Análisis de las cancelaciones de City Hotel

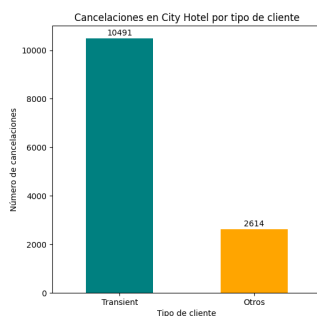


Figura 10: Análisis de las cancelaciones de City Hotel, enfocándose en el cliente Transient

Para este punto, notamos que el cliente Transient es definitivamente el tipo de usuario al que debemos de abordar si deseamos reducir la cantidad de cancelaciones del tipo de hotel mas cancelado.

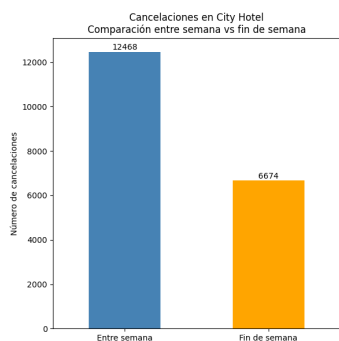


Figura 11: Análisis de las cancelaciones de City Hotel, según el tipo de día

Este apartado nos muestra que, aunque cancelaciones entre semana están por encima de las cancelaciones de días de fines de semana, siguen siendo valores enormes en de cancelaciones. Podría ser prudente abordar ambos apartados.

V-B. Análisis entre las cancelaciones y tipo de cliente Transient

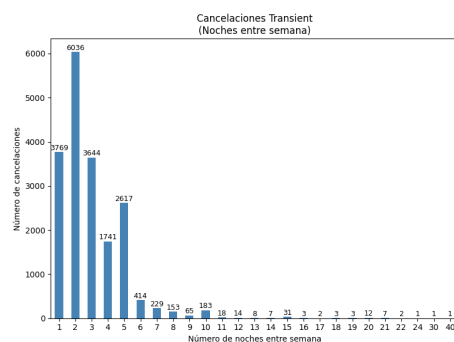


Figura 12: Análisis entre las cancelaciones y tipo de cliente Transient para días entre semana

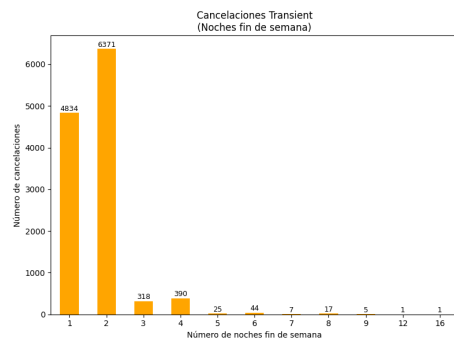


Figura 13: Análisis entre las cancelaciones y tipo de cliente Transient para días de fin de semana

Gracias a estos dos estudios podemos notar que las cancelaciones son muy elevadas en los valores de días más bajos. Asimismo, que la diferencia entre el numero de cancelaciones es muy marcada, pues hay diferencia de hasta un orden de magnitud.

VI. CIERRE Y CONCLUSIONES

Con el deseo de poder reducir las cancelaciones, y habiendo realizado todo el análisis mostrado en este documento, desde mi persona se proponen las siguientes acciones o correcciones a implementar:

1. Gracias a resultados obtenidos en el capítulo V-A, podría sugerir implementar un sistema de karma, donde los usuarios Transient empiezan teniendo una capacidad de reserva limitada (por ejemplo, máximo 1 noche entre semana y 1 de fin de semana), y que, a medida que van teniendo reservas exitosas, sin cancelaciones, este numero de capacidad de reservas va aumentando.
2. Gracias los resultados obtenidos en los apartados II, IV-C, IV-D y V-B; aconsejaría que el sistema de reservas, en especial para el usuario Transient, no permita reserva mas de 10 días entre semana y mas de dos de fines de semana. Esto principalmente porqué son el grupo de días

con el ratio de cancelación mas alto, y con una cantidad de reservas moderada.

Esto para poder, por lo menos, evitar reservaciones que ya cuentan con las mas altas tendencias de cancelaciones.

3. En cuanto a los resultados obtenidos en los apartados II y IV-A, está claro que debe de hacerse el hotel urbano mucho mas consistente, pues aunque cuenta con el número mas alto de reservas, también es el que mas cancelan entre los dos tipos de hoteles.

Para ello, y en relación con relación al sistema de karma propuesto, podría sugerir una política de reembolsos mas agresiva para los usuarios con niveles de karma nuevos o negativos. Es decir, no hacer una devolución total del valor de la reserva, sino un retorno del 70 %; valor que puede irse modificando a medida que el karma se hace favorable.

4. Con relación a los resultados obtenidos en el apartado IV-C, donde se evidenció que las cancelaciones en fines de semana aumentan entre los cinco y los nueve días reservados, sería apropiado implementar un sistema de incentivos directos por confirmación temprana. Por ejemplo, elevar positivamente el karma, ofrecer descuentos escalonados o créditos de consumo en restaurante/s-pa/bar que solo se habiliten si la reserva no es cancelada. De esta manera, el cliente percibe un beneficio inmediato que se perdería al cancelar, reduciendo la probabilidad de abandono en estancias largas de fin de semana.

REFERENCIAS

- [1] Cathy O'Neil and Rachel Schutt. Doing Data Science: Straight Talk from the Frontline. O'Reilly Media, Inc. 2013 [Último acceso: 10 08 2025]
- [2] Daniel Vaughan. Data Science: The Hard Parts. O'Reilly Media, Inc. 2024 [Último acceso: 10 08 2025]
- [3] Foster Provost and Tom Fawcett. Data Science for Business. O'Reilly Media, Inc., 2013 [Último acceso: 10 08 2025]
- [4] Doug Rose. Data Science: Create Teams That Ask the Right Questions and Deliver Real Value. APress. 2016. [Último acceso: 10 08 2025]