**Universidad de los Andes - Maestría en Ingeniería de la Información**

**Profesor**: Fabian Camilo Peña

**Fecha entrega:** 21 de septiembre de 2025

**Estudiantes**: Juan Carlos Tovar Orjuela – José Jorge Geles Carrillo

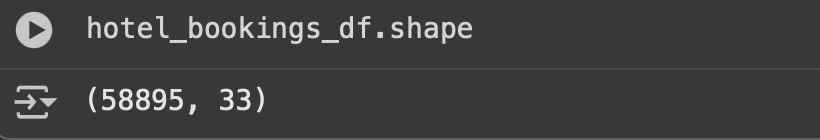
**Taller 1 – Ciencia de Datos**

1. **Reporte de Entendimiento Inicial de Datos**

Este reporte presenta un análisis del dataset de reservas hoteleras, con el fin de encontrar y optimizar estrategias de ocupación. En una fase inicial el objetivo es comprender la estructura y contenido del conjunto de datos, identificando los atributos clave para el análisis, esto nos permite entender los datos e identificar posibles soluciones al problema.

* **Dimensiones del Dataset**

El dataset hotel\_bookings\_df, tiene 58,895 registros (filas) y 33 atributos (columnas).



* **Tipos de datos**

**int64 y float64**: Estos datos son utilizados para análisis cuantitativos, representan atributos numéricos para las variables:

* is\_canceled
* lead\_time
* arrival\_date\_year
* arrival\_date\_week\_number
* arrival\_date\_day\_of\_month
* stays\_in\_weekend\_nights
* stays\_in\_week\_nights
* adults
* children
* babies
* is\_repeated\_guest
* previous\_cancellations
* previous\_bookings\_not\_canceled
* booking\_changes
* agent
* days\_in\_waiting\_list
* adr
* required\_car\_parking\_spaces
* total\_of\_special\_requests
* kids

**object**: Estas variables son utilizadas para entender las clasificaciones y características de las reservas, representan atributos categóricos o de texto

* hotel
* arrival\_date\_month
* meal
* country
* market\_segment
* distribution\_channel
* reserved\_room\_type
* assigned\_room\_type
* deposit\_type
* company
* customer\_type
* reservation\_status
* reservation\_status\_date
* **Top 5 de Atributos más Importantes para el Análisis**

**1. *is\_canceled* (Estado de la Reserva)**

Predecir la cancelación. Es una variable binaria (int64) con valores de **0** (reserva no cancelada) y **1** (reserva cancelada). La distribución de este atributo es útil para entender la tasa de cancelación general.

**2. *lead\_time* (Tiempo de Anticipación)**

Este atributo numérico (int64) mide la cantidad de días transcurridos entre la fecha de la reserva y la fecha de llegada. Es un variable directa en la planificación de la ocupación y en la probabilidad de cancelación.

**3. *hotel* (Tipo de Hotel)**

Esta variable categórica (object) indica si la reserva corresponde a un **"Resort Hotel"** o a un **"City Hotel"**. Es importante entender la distribución, ya que las políticas de precios, los patrones de reserva y la demanda pueden variar significativamente entre ambos tipos de establecimientos.

**4. *adults* (Número de Adultos)**

Este atributo numérico (int64) indica el número de adultos en la reserva. Es una variable esencial para la asignación de habitaciones y la estimación de ingresos.

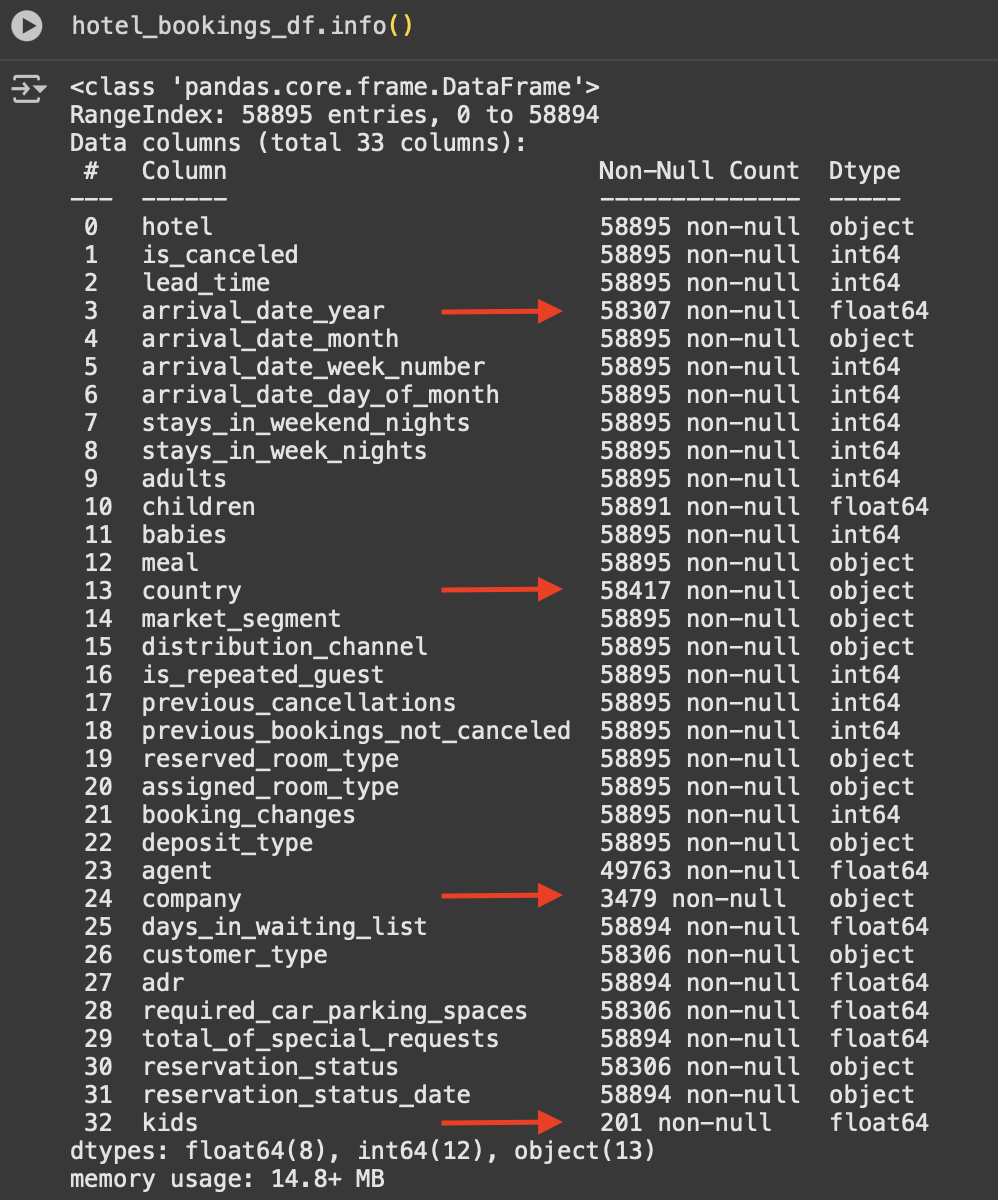
**5. *country* (País de Origen del Huésped)**

Esta variable categórica (object) proporciona el país de origen del cliente. Este atributo nos ayuda a generar la segmentación de mercado y también se utiliza para la creación de estrategias de precios o promociones específicas para diferentes mercados.

1. **Estrategia de Análisis para Optimización Hotelera**

En primera instancia se va a calcular la tasa de cancelación y compararla entre el hotel de ciudad y el resort. Esto nos permite identificar de forma general en dónde está el problema más grande. Luego, se va a usar la visualización de datos para entender los factores que más influyen. Por ejemplo, con gráficos, se podrá observar si las reservas hechas con mucha anticipación o las que usan ciertos canales (como agencias de viaje) tienen más probabilidades de ser canceladas. En segunda instancia se van a cruzar los datos para ver cómo variables como el tipo de habitación, el número de huéspedes y el tipo de depósito (sin depósito, reembolsable, etc.) afectan la cancelación. Finalmente, sé debe identificar qué segmentos de mercado o países son los más rentables y cuáles tienen la tasa más alta de no-shows, con esta información los hoteles puedan adaptar sus precios y promociones.

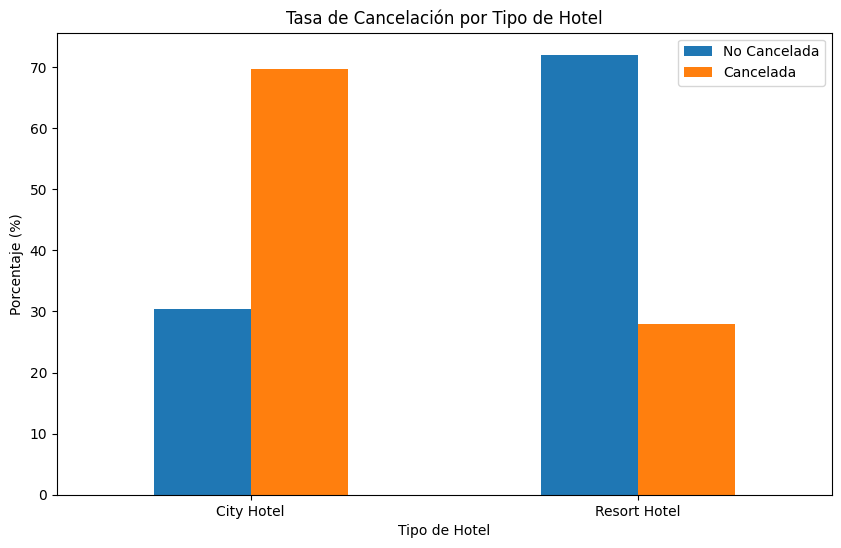
1. **Desarrollo de la Estrategia**
   1. *Limpieza de datos*

****

* **arrival\_date\_year**: Tiene 58,307 valores no nulos, lo que indica que hay 588 valores faltantes.
* **country**: Tiene 58,417 valores no nulos, es decir, hay 478 nulos.
* **agent:** Tiene solo 49,763 valores no nulos, lo que indica que hay gran cantidad de reservas sin intermediario.
* **company:** Tiene muy pocos valores no nulos (solo 3,479), por lo que podríamos considerar que esta columna no es muy útil para el análisis.
* **kids:** Esta columna tiene muy pocos valores no nulos (solo 201).

**3.2 Implementación de las Estrategias**

***Tasa de Cancelación por Tipo de Hotel*** - se calcula la tasa de cancelación para cada tipo de hotel. Esto nos dirá cuál de los dos tiene el mayor problema.



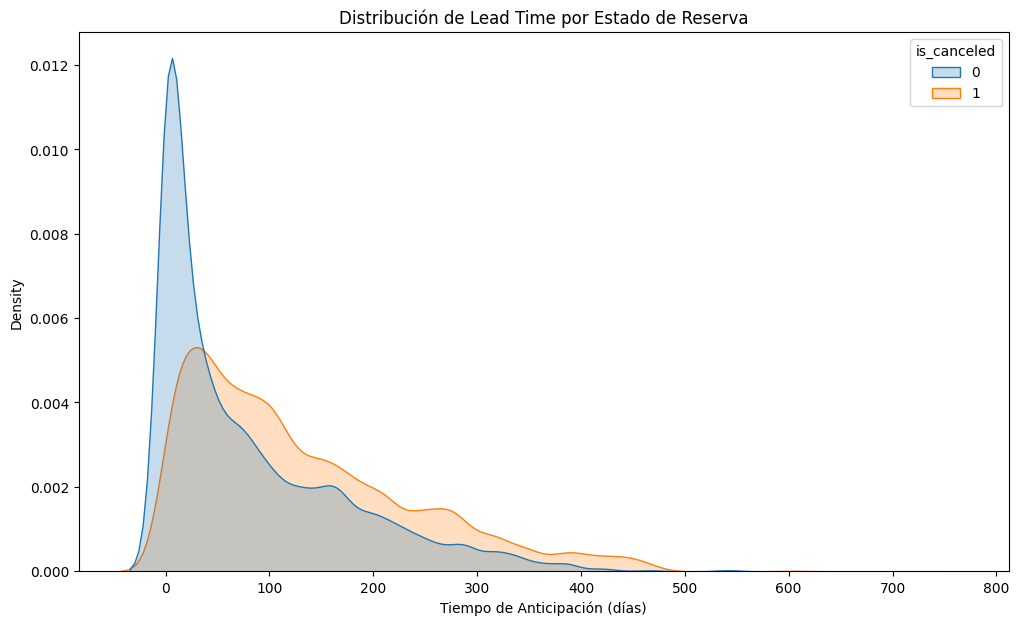
*City Hotel:* La barra naranja, representa las reservas canceladas, es significativamente más alta. Aproximadamente el 70% de las reservas en los hoteles de ciudad son canceladas, mientras que solo un 30% no lo son.

*Resort Hotel:* La barra azul ("No Cancelada") domina, con alrededor del 72% de las reservas concretándose, frente a solo un 28% de cancelaciones. Esto demuestra que los hoteles Resort tienen una tasa de ocupación mucho más estable.

**Insight**

La principal conclusión es que el "City Hotel" es el tipo de hotel con mayor riesgo de cancelación. Las estrategias para reducir cancelaciones deben enfocarse principalmente en este segmento del negocio.

***Impacto del Tiempo de Anticipación (lead\_time)***

****

*Reservas no canceladas (curva azul):* La gran mayoría de estas reservas se concentran en un lead\_time bajo, con un pico muy pronunciado cerca de los 0 días. Esto significa que los clientes que reservan con poca antelación rara vez cancelan.

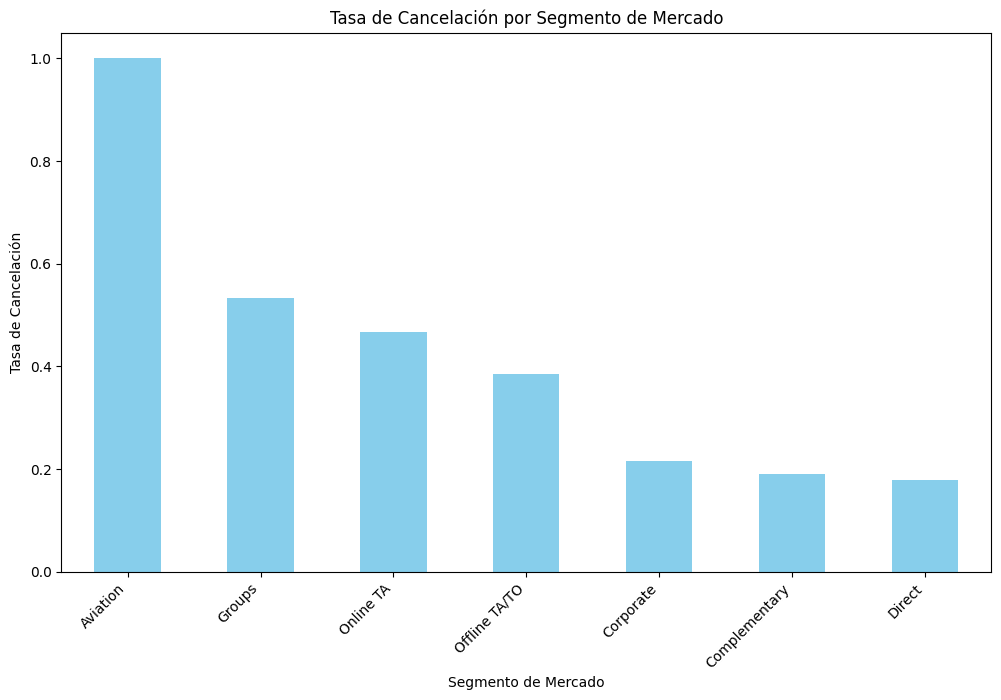
*Reservas canceladas (curva naranja):* La distribución es mucho más dispersa y se extiende a lo largo de los días. La curva es más plana, lo que indica que las cancelaciones ocurren con mayor frecuencia cuando el lead\_time es largo.

**Insight**

A mayor tiempo de anticipación en la reserva, mayor es la probabilidad de que el cliente cambie de opinión y la cancele. El hotel puede usar esta información para:

* Ajustar las políticas de cancelación: Para reservas con lead\_time altos, podrían requerir un depósito no reembolsable o aplicar cargos por cancelación más estrictos.
* Activar campañas de retención: Enviar recordatorios o promociones exclusivas a los clientes que reservaron con mucha antelación para incentivarlos a mantener su reserva.

***Segmento de Mercado y Cancelaciones***

****

*Aviation (Aviación):* Este segmento tiene una tasa de cancelación del 100%. Esto es un hallazgo crítico. Probablemente se trata de reservas "fantasma" o de pruebas, y no de reservas reales de clientes.

*Groups y Online TA:* Estos dos segmentos muestran tasas de cancelación muy altas, de aproximadamente 55% y 45% respectivamente. Esto porcentaje es muy alto, ya que las agencias de viaje en línea (Online TA) son un canal de reserva popular.

*Corporate, Complementary y Direct:* Estos segmentos tienen las tasas de cancelación más bajas, por debajo del 25%. Las reservas corporativas y las directas son, por lo tanto, las más confiables para la ocupación del hotel.

**Insight**

El análisis por segmento de mercado revela que no todos los canales de reserva son iguales. A las reservas por hechas por el canal de Aviation, y las agencias de viaje en línea (Online TA) y los grupos se deben implementar las políticas de cancelación más rigurosas, ya que representan un riesgo significativo. Por otro lado, la dirección debe enfocar sus esfuerzos en incentivar las reservas directas o corporativas, que son las más seguras.

1. **Informe Ejecutivo: Estrategias para Aumentar la Ocupación y Reducir Cancelaciones**

* ***El Problema:*** El Hotel de Ciudad es el Foco de la Crisis

Los datos muestran que el City Hotel enfrenta un problema de cancelaciones mucho mayor, con una tasa de alrededor del 70%. Esto contrasta con la estabilidad del Resort Hotel, que tiene una tasa de cancelación de solo un 28%. La estrategia debe priorizar la reducción de cancelaciones en los hoteles de ciudad.

* ***Las Causas Principales***

Identificación de factores que impulsan estas cancelaciones:

* Anticipación de la reserva (Lead Time): A mayor tiempo entre la reserva y la llegada, mayor es la probabilidad de que el cliente cancele. Las reservas hechas con más de 60 días de antelación son las más riesgosas.
* Segmento de Mercado: Las reservas del segmento Aviation tienen una tasa de cancelación del 100%, los canales Online TA (agencias de viaje online) y Groups también presentan tasas de cancelación muy altas 55% y 45% respectivamente.
* ***Recomendaciones y Estrategias Basadas en Datos***

Para atacar estos problemas, proponemos un conjunto de acciones concretas:

*Políticas de Depósito Flexibles por Lead Time:*

* Para reservas en City Hotels con un lead\_time superior a 60 días, se recomienda que implementen un depósito no reembolsable obligatorio. Esto servirá como un filtro para las reservas serias y mitigará el riesgo.
* Para reservas de último minuto (menos de 7 días), pueden ofrecer una política de cancelación más flexible para atraer a más clientes, ya que sabemos que estas reservas rara vez se cancelan.

*Optimizar la Estrategia por Canal de Reserva:*

* Prioricen las reservas directas: Ofrecer un descuento exclusivo o un beneficio adicional (como desayuno gratis) a los clientes que reserven directamente desde su página web. Nuestro análisis muestra que estas reservas son las más confiables.
* Revisar los acuerdos con Agencias de Viaje Online: Se recomienda analizar si las altas comisiones que pagan a los Online TA se justifican, dada su alta tasa de cancelación. se deben consideren negociar condiciones más estrictas o dirigir el gasto de marketing a otros canales más rentables.

*Gestión de la Duración de la Estadía:*

El análisis de la duración de la estadía en las variables (stays\_in\_week\_nights y stays\_in\_weekend\_nights) describe las reservas más cortas, para aumentar la estadía en las City Hotels, las cuales son más volátiles. Se recomienda ofrecer paquetes de descuento para estadías de 3 o más noches para incentivar estancias más largas y estables.