

# Análisis de reservas de hoteles



**Data Analytics**

**Comisión 66120**

Jhustyn Estévez

## Tabla de contenido

1.	Descripción de la temática de los datos .....	3
2.	Hipótesis de trabajo.....	3
3.	Objetivos del proyecto .....	3
4.	Alcance .....	3
5.	Usuario final y nivel de aplicación del análisis .....	3
6.	Dataset.....	4
7.	Diagrama Entidad-Relación.....	5
8.	Listado de Tablas .....	6
9.	Listado de columnas de tablas .....	6
10.	Transformación de datos.....	11
	Tabla dataset (alimenta el conjunto de datos inicial de tabla reservas) .....	11
	Tabla reservas .....	12
	Tabla calendario.....	12
11.	Medidas calculadas .....	13
12.	Modelo relacional de Power BI .....	14

## **1. Descripción de la temática de los datos**

El conjunto de datos se centra en la información de reservas para un hotel urbano y un hotel de resort. Contiene detalles como la fecha de la reserva, duración de la estadía, cantidad de adultos, niños y/o bebés, así como el número de espacios de estacionamiento disponibles, entre otros aspectos relevantes. Es importante destacar que toda la información identificable personalmente ha sido eliminada del conjunto de datos.

## **2. Hipótesis de trabajo**

- Existen estacionalidades en las reservas de hoteles.
- Existe una temporada específica del año en la que las tarifas son más favorables, posiblemente debido a la demanda estacional.
- Algunos atributos específicos de las reservas, como la duración de la estadía o la presencia de niños, pueden aumentar la probabilidad de recibir solicitudes especiales.

## **3. Objetivos del proyecto**

- Identificar patrones estacionales en las reservas de hotel.
- Determinar la relación entre la duración de la estadía y las tarifas diarias.

## **4. Alcance**

El proyecto se enfocará en analizar y visualizar datos de reservas de hoteles para extraer patrones significativos. El alcance incluirá explorar la temporalidad de las reservas, la duración de la estadía y posiblemente la influencia de otros factores en la tarifa diaria. No se abordarán aspectos relacionados con la identificación personal.

## **5. Usuario final y nivel de aplicación del análisis**

Usuario final: Analistas de datos, científicos de datos, y profesionales en la industria hotelera.

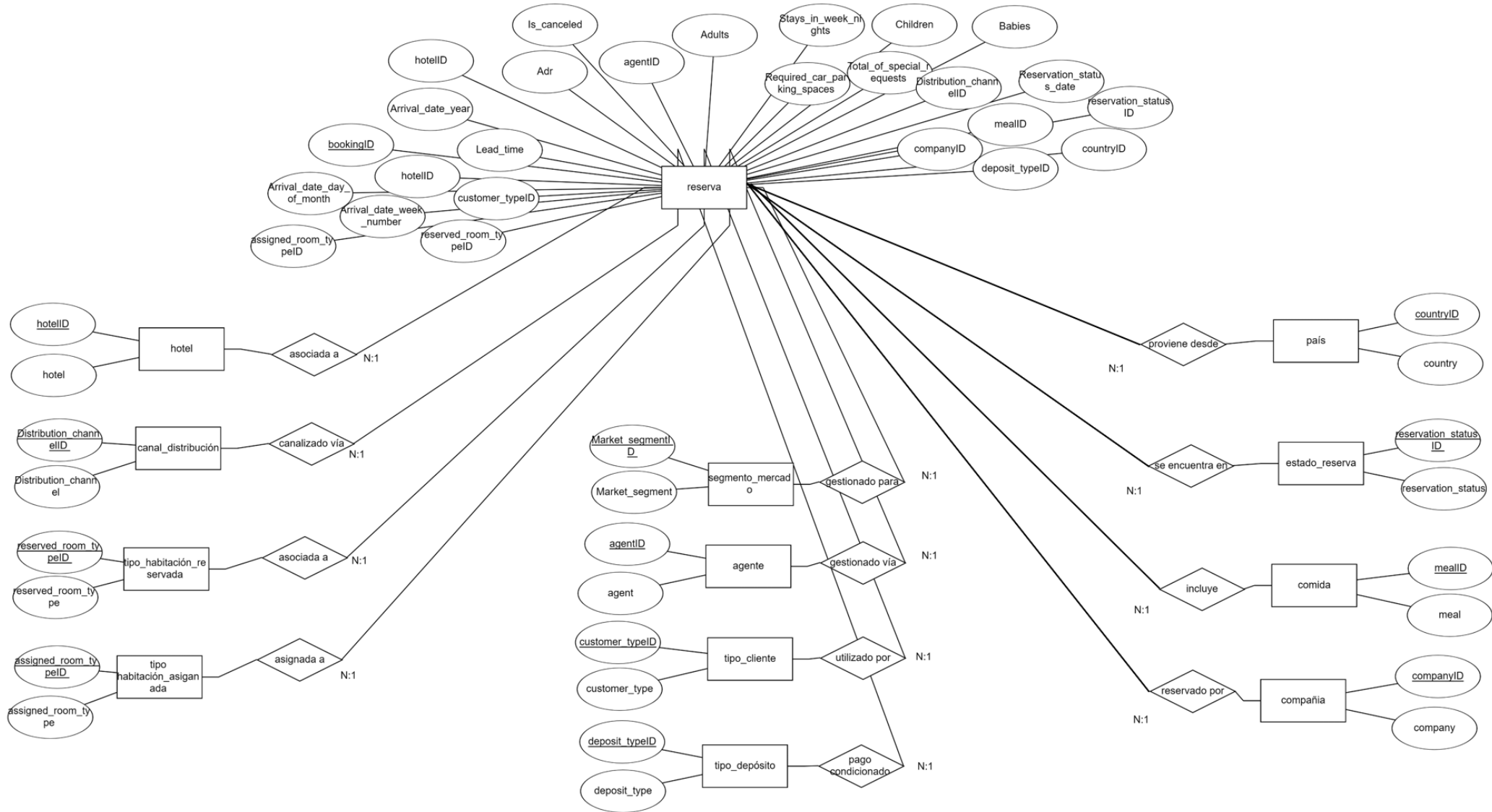
Nivel de aplicación: Exploratory Data Analysis (EDA) para comprender tendencias y patrones, y posiblemente, el desarrollo de modelos predictivos para la gestión hotelera. Este análisis puede ayudar a tomar decisiones informadas sobre estrategias de precios y planificación operativa.

## **6. Dataset**

El dataset consta de 32 atributos numéricos, binarios, categóricos y de fecha. Para acceder al xls puede hacerse mediante el enlace:

<https://drive.google.com/file/d/1zo0BZK8OshxGJea4w2y8dVZZMhUnDyBw/view?usp=sharing>

## 7. Diagrama Entidad-Relación



## 8. Listado de Tablas

- a) hotel
- b) comida
- c) reservas
- d) país
- e) segmento\_mercado
- f) canal\_distribución
- g) tipo\_habitación\_reservada
- h) tipo\_habitación\_asignada
- i) tipo\_depósito
- j) tipo\_cliente
- k) estado\_reserva
- l) agente
- m) compañía

## 9. Listado de columnas de tablas

- a) Hotel

Tipo de clave	Campo	Tipo de campo	Detalle
PK	hotelID	INT	Identificador único
	Hotel	VARCHAR	El tipo de hotel, ya sea "Hotel de ciudad" o "Hotel Resort".

- b) Comida

Tipo de clave	Campo	Tipo de campo	Detalle
PK	mealID	INT	Identificador único
	meal	VARCHAR	Tipo de comida reservada.

c) Reservas

Tipo de clave	Campo	Tipo de campo	Detalle
PK	bookingID	INT	Identificador único
FK	hotelID	INT	Indentificador
	Is_canceled	BINARY	Valor binario que indica si la reserva fue cancelada (1) o no (0).
	Lead_time	INT	Número de días entre la reserva y la llegada.
	Arrival_date_year	INT	Año de fecha de llegada.
	Arrival_date_week_number	INT	Número de semana de la fecha de llegada.
	Arrival_date_day_of_month	VARCHAR	Día del mes de fecha de llegada.
	Stays_in_week_nights	INT	Número de noches entre semana (de lunes a viernes) que se aloja el huésped.
	Adults	INT	Número de adultos.
	Children	INT	Número de hijos.

Tipo de clave	Campo	Tipo de campo	Detalle
	Babies	INT	Número de bebés.
	Adr	INT	Tarifa promedio diaria.
	Required_car_parking_spaces	BIT	Número de plazas de aparcamiento necesarias.
	Total_of_special_requests	BIT	Número de solicitudes especiales realizadas.
	Reservation_status_date	DATE	Fecha del último estado.
FK	Distribution_channelID	INT	Identificador
FK	hotelID	INT	Identificador
FK	mealID	INT	Identificador
FK	countryID	INT	Identificador
FK	reserved_room_typeID	INT	Identificador
FK	assigned_room_typeID	INT	Identificador
FK	deposit_typeID	INT	Identificador
FK	customer_typeID	INT	Identificador
FK	reservation_statusID	INT	Identificador
FK	agentID	INT	Identificador
FK	companyID	INT	Identificador

d) País

Tipo de clave	Campo	Tipo de campo	Detalle
PK	countryID	INT	Identificador único
	country	VARCHAR	País de origen.

e) segmento\_mercado



Tipo de clave	Campo	Tipo de campo	Detalle
PK	Market_segmentID	INT	Identificador único
	Market_segment	VARCHAR	Designación del segmento de mercado.

f) canal\_distribución

Tipo de clave	Campo	Tipo de campo	Detalle
PK	Distribution_channelID	INT	Identificador único
	Distribution_channel	VARCHAR	Canal de distribución de reservas.

g) tipo\_habitación\_reservada

Tipo de clave	Campo	Tipo de campo	Detalle
PK	reserved_room_typeID	INT	Identificador único
	reserved_room_type	VARCHAR	Tipo de habitación reservada.

h) tipo\_habitación\_asignada

Tipo de clave	Campo	Tipo de campo	Detalle
PK	assigned_room_typeID	INT	Identificador único
	assigned_room_type	VARCHAR	Tipo de habitación asignado en el check-in.

i) tipo\_depósito

Tipo de clave	Campo	Tipo de campo	Detalle
PK	deposit_typeID	INT	Identificador único
	deposit_type	VARCHAR	Tipo de depósito realizado.

j) tipo\_cliente

Tipo de clave	Campo	Tipo de campo	Detalle
PK	customer_typeID	INT	Identificador único
	customer_type	VARCHAR	Tipo de reserva.

k) estado\_reserva

Tipo de clave	Campo	Tipo de campo	Detalle
PK	reservation_statusID	INT	Identificador único
	reservation_status	VARCHAR	Último estado de la reserva.

l) agente

Tipo de clave	Campo	Tipo de campo	Detalle
PK	agentID	INT	Identificador único
	agent	VARCHAR	DNI de la agencia de viajes.

m) compañía

Tipo de clave	Campo	Tipo de campo	Detalle
PK	companyID	INT	Identificador único
	company	VARCHAR	DNI de la empresa.

## 10. Transformación de datos

### Tabla dataset (alimenta el conjunto de datos inicial de tabla reservas)

```
let
Source = Excel.Workbook(File.Contents("C:\Users\jhust\OneDrive\Coder\Entrega 3\hotel_bookings_entrega3.xlsx"), null, true),
data_sheet = Source[[Item"data",Kind="Sheet"]][Data],
#Changed Type = Table.TransformColumnTypes(data_sheet,{{("Column1", type any), ("Column2", type any), ("Column3", type any), ("Column4", type any), ("Column5", type text), ("Column6", type any), ("Column7", type any), ("Column8", type any), ("Column9", type any), ("Column10", type any), ("Column11", type any), ("Column12", type any), ("Column13", type any), ("Column14", type any), ("Column15", type any), ("Column16", type any), ("Column17", type any), ("Column18", type any), ("Column19", type any), ("Column20", type any), ("Column21", type any), ("Column22", type any), ("Column23", type any), ("Column24", type any), ("Column25", type any), ("Column26", type any), ("Column27", type any), ("Column28", type any), ("Column29", type any), ("Column30", type any), ("Column31", type any), ("Column32", type any)}}),
#Removed Top Rows = Table.Skip(#Changed Type,1),
#Promoted Headers = Table.PromoteHeaders(#Removed Top Rows, {PromoteAllScalars=true}),
#Changed Type1 = Table.TransformColumnTypes(#Promoted Headers,{{("hotel", type text), ("is_cancelled", Int64.Type), ("lead_time", Int64.Type), ("arrival_date_year", Int64.Type), ("arrival_date_month", type text), ("arrival_date_week_number", Int64.Type), ("arrival_date_day_of_month", Int64.Type), ("stays_in_weekend_nights", Int64.Type), ("stays_in_week_nights", Int64.Type), ("adults", Int64.Type), ("children", Int64.Type), ("babies", Int64.Type), ("meal", type text), ("country", type text), ("market_segment", type text), ("distribution_channel", type text), ("is_repeated_guest", Int64.Type), ("previous_cancellations", Int64.Type), ("previous_bookings_not_cancelled", Int64.Type), ("reserved_room_type", type text), ("assigned_room_type", type text), ("booking_changes", Int64.Type), ("deposit_type", type text), ("agent", type any), ("company", type any), ("days_in_waiting_list", Int64.Type), ("customer_type", type text), ("adr", type number), ("required_car_parking_spaces", Int64.Type), ("total_of_special_requests", Int64.Type), ("reservation_status", type text), ("reservation_status_date", type date)}}),
#Replaced Errors = Table.ReplaceErrorValues(#Changed Type1,{{("children", 0)}}),
#Added Index = Table.AddColumn(#Replaced Errors, "Index", 123456, 1, Int64.Type),
#Changed Type2 = Table.TransformColumnTypes(#Added Index,{{("Index", type text)}}),
#Renamed Columns = Table.RenameColumns(#Changed Type2,{{("Index", "bookingID"}})),
#Reordered Columns = Table.ReorderColumns(#Renamed Columns,{"bookingID","hotel","is_cancelled","lead_time","arrival_date_year","arrival_date_month","arrival_date_week_number","arrival_date_day_of_month","stays_in_weekend_nights","stays_in_week_nights","adults","children","babies","meal","country","market_segment","distribution_channel","is_repeated_guest","previous_cancellations","previous_bookings_not_cancelled","reserved_room_type","assigned_room_type","booking_changes","deposit_type","agent","company","days_in_waiting_list","customer_type","adr","required_car_parking_spaces","total_of_special_requests","reservation_status","reservation_status_date"}),
#Renamed Column1 = Table.RenameColumns(#Reordered Columns,{{("hotel", "hotelID"}}),
#Replaced Value1 = Table.ReplaceValue(#Renamed Column1,"Resort Hotel","1",Replacer.ReplaceText,{"hotelID"}),
#Replaced Value1 = Table.ReplaceValue(#Replaced Value1,"City Hotel","2",Replacer.ReplaceText,{"hotelID"}),
#Renamed Column2 = Table.RenameColumns(#Replaced Value1,{{("meal","mealID"), ("country","countryID"), ("reserved_room_type","reserved_room_typeID"), ("assigned_room_type","assigned_room_typeID"}}),
#Renamed Column2
```

- Acceso al archivo Excel ubicado en la ruta especificada.
- Selección de la hoja de datos.
- Transformación de los tipos de datos de las columnas según la especificación.
- Omisión de las primeras filas del conjunto de datos.
- Promoción de la primera fila como encabezados de columnas.
- Transformación adicional de los tipos de datos de las columnas.
- Reemplazo de valores erróneos (en la columna "children") con cero.
- Agregación de un índice único para cada fila del conjunto de datos.
- Transformación del tipo de datos de la columna de índice a texto.
- Renombramiento de la columna de índice a "bookingID" para que sirva como identificador único de las reservas.
- Reordenamiento de las columnas para una estructura más coherente.
- Renombramiento de la columna "hotel" a "hotelID".
- Reemplazo de valores en la columna "hotelID" para codificar "Resort Hotel" como "1" y "City Hotel" como "2".
- Renombramiento de algunas columnas para mejorar la claridad y consistencia en la nomenclatura.

## Tabla reservas

```
let
Source = data,
#"Extracted First Characters" = Table.TransformColumns(Source, ({"arrival_date_month", each Text.Start(_, 3), type text})),
#"Inserted Merged Column" = Table.AddColumn(#"Extracted First Characters", "fecha_llegada", each Text.Combine({Text.From([arrival_date_year], "en-US"), [arrival_date_month], Text.From([arrival_date_day_of_month], "en-US")}, "-"), type text),
#"Changed Type" = Table.TransformColumnTypes(#"Inserted Merged Column",({{"fecha_llegada", type date}})),
#"Inserted Addition" = Table.AddColumn(#"Changed Type", "Addition", each [stays_in_weekend_nights] * [stays_in_week_nights], Int64.Type),
#"Renamed Columns" = Table.RenameColumns(#"Inserted Addition",({{"Addition", "dias_estadia"}})),
#"Added Custom" = Table.AddColumn(#"Renamed Columns", "fecha_salida", each if [reservation_status] = "Check-Out" then Date.AddDays([fecha_llegada], [dias_estadia]) else null),
#"Filtered Rows" = Table.SelectRows(#"Added Custom", each ([bookingID] < "123993" and [bookingID] < "125483" and [bookingID] < "125889")),
#"Inserted Sum" = Table.AddColumn(#"Filtered Rows", "Addition", each List.Sum([adults], [children], [babies]), Int64.Type),
#"Renamed Columns1" = Table.RenameColumns(#"Inserted Sum",({{"Addition", "cantidad de huéspedes"}})),
#"Filtered Rows1" = Table.SelectRows(#"Renamed Columns1", each ([cantidad de huéspedes] < 0)),
#"Added Custom1" = Table.AddColumn(#"Filtered Rows1", "tipo_reserva_interna", each if [adults] <= 2 and [children] = 0 and [babies] = 0 then "Solo adulto(s)"
else if [adults] <= 2 and [children] > 0 and [babies] > 0 then "Familiar"
else if [adults] > 2 and [children] > 0 and [babies] > 0 then "Grupal"
else if [adults] = 0 and [children] > 0 and [babies] > 0 then "Infantil"
else "Undefined")
in
#"Added Custom1"
```

- Acceso al conjunto de datos inicial.
- Extracción de los primeros tres caracteres del mes de llegada.
- Creación de una nueva columna combinando el año, mes y día para formar la fecha de llegada.
- Conversión del tipo de datos de la columna de fecha de llegada a tipo fecha.
- Creación de una nueva columna calculando la duración de la estadía en días.
- Renombramiento de la columna de duración de la estadía.
- Creación de una nueva columna para la fecha de salida basada en el estado de la reserva.
- Filtrado de filas con identificadores de reserva específicos que no tienen huéspedes.
- Creación de una nueva columna sumando el número de huéspedes.
- Renombramiento de la columna sumatoria de huéspedes.
- Filtrado de filas donde el número de huéspedes es cero. Esto para excluir las reservas sin huéspedes.
- Creación de una nueva columna clasificando el tipo de reserva según el número y tipo de huéspedes.

## Tabla calendario

```
let
Source = data,
#"Removed Other Columns" = Table.SelectColumns(Source, {"reservation_status_date"},
#"Removed Duplicates" = Table.Distinct(#"Removed Other Columns"),
#"Renamed Columns" = Table.RenameColumns(#"Removed Duplicates", ({{"reservation_status_date", "fecha_estado_reserva"})),
#"Inserted Year" = Table.AddColumn(#"Renamed Columns", "Year", each Date.Year([fecha_estado_reserva]), Int64.Type),
#"Inserted Month Name" = Table.AddColumn(#"Inserted Year", "Month Name", each Date.MonthName([fecha_estado_reserva]), type text),
#"Inserted Quarter" = Table.AddColumn(#"Inserted Month Name", "Quarter", each Date.QuarterOfYear([fecha_estado_reserva]), Int64.Type),
#"Inserted Day" = Table.AddColumn(#"Inserted Quarter", "Day", each Date.Day([fecha_estado_reserva]), Int64.Type)
in
#"Inserted Day"
```

- Acceso al conjunto de datos inicial.
- Selección de la columna "reservation\_status\_date".
- Eliminación de filas duplicadas en la columna seleccionada.
- Renombramiento de la columna a "fecha\_estado\_reserva".
- Creación de una nueva columna para extraer el año de la fecha de estado de reserva.

- f) Creación de una nueva columna para obtener el nombre del mes a partir de la fecha de estado de reserva.
- g) Creación de una nueva columna para determinar el trimestre del año basado en la fecha de estado de reserva.
- h) Creación de una nueva columna para obtener el día del mes desde la fecha de estado de reserva.

## 11. Medidas calculadas

- **Conteo\_reservas:** calcula el número total de reservas realizadas

`conteo_reservas = DISTINCTCOUNT(reservas[bookingID])`

- **Adr\_promedio:** calcula el promedio del ingreso diario obtenido por una habitación ocupada en el hotel durante un período de tiempo específico

`adr_promedio = AVERAGE(reservas[adr])`

- **Avg\_solicitudes\_especiales:** calcula el promedio de solicitudes especiales hechas por los huéspedes durante un período de tiempo determinado.

`avg_solicitudes_especiales = AVERAGE(reservas[total_of_special_requests])`

- **Estadía\_promedio\_week:** calcula la duración promedio de la estadía de los huéspedes durante los días laborables de la semana.

`estadía_promedio_week = AVERAGE(reservas[stays_in_week_nights])`

- **Estadía\_promedio\_weekend:** calcula la duración promedio de la estadía de los huéspedes durante los fines de semana.

`estadía_promedio_weekend = AVERAGE(reservas[stays_in_weekend_nights])`

- **Promedio\_estadía:** calcula el promedio de la duración de la estadía de los huéspedes en el hotel.

`promedio_estadía = AVERAGE(reservas[dias_estadia])`

- **Promedio\_lead\_time:** calcula el promedio de la cantidad de días entre la fecha de reserva y la fecha de llegada de los huéspedes al hotel

`promedio_lead_time = AVERAGE(reservas[lead_time])`

- **Suma\_huespedes:** calcula la suma total de adultos, niños y bebés incluidos en todas las reservas.

`suma_huespedes = SUMX(reservas,reservas[children]+reservas[adults]+reservas[babies])`

## 12. Modelo relacional de Power BI

