



**UNIVERSIDAD PERUANA DE CIENCIAS APLICADAS**

**SECCIÓN: CC52**

**GRUPO: 1**

**CURSO: Fundamentos de Data Science**

**PROFESOR(A): Nérida Isabel Manrique Tunque**

**TÍTULO: Análisis de demanda de hotelería**

El presente trabajo ha sido realizado por:

Nombre 1	Alfredo Mauricio Aragón Ovalle (u202210494)
Nombre 2	Tomás Edgar Ninán Melo (u201625763)
Nombre 3	José Guillermo Melgar Puertas (u202111660)
Nombre 4	Rody Sebastian Vilchez Marin (u202216562)

2024-01

En el artículo proporcionado, se puede observar que los datos utilizados provienen de las bases de datos de los sistemas de gestión de propiedades (PMS) de dos hoteles que están ubicados en Portugal. Como autores del artículo se encuentran Nuno Antonio, Ana de Almeida y Luis Nunes, estas tres personas están afiliadas al Instituto Universitario de Lisboa (ISCTE-IUL), Instituto de Telecomunicaciones, CISUC y ISTAR-IUL en Lisboa, Portugal. Por otra parte, podemos encontrar fechas importantes como por ejemplo la publicación de este artículo, el cuál fue recibido el 5 de octubre de 2018, aceptado el 26 de noviembre de 2018 y publicado en línea el 29 de noviembre de 2018.

Como se comentó en primera instancia, los datos utilizados provienen de las bases de datos de dos hoteles, el primer hotel se encuentra ubicado en la región turística de Algarve y el segundo hotel se encuentra situado en la ciudad de Lisboa. Estos datos se proporcionaron para el uso de análisis descriptivo, investigación en predicción de cancelaciones de reservas, segmentación de clientes, etc.

Existen varios grupos de interés los cuáles demandan del análisis de datos de reservas de hoteles, de igual manera, existe una amplia gama de casos de uso aplicables. Algunos de los posibles beneficiarios y casos se encuentran:

- **Profesionales de la industria hotelera:** Este grupo de personas pueden beneficiarse mediante el análisis de estos datos de manera que se entiende mejor el comportamiento de los clientes al momento de reservar, predecir si las reservas podrían cancelarse, segmentar los clientes y mejorar su satisfacción.
- **Investigadores en Machine Learning y Data Mining:** Las personas que se encargan del aprendizaje automático y minería de datos podrían utilizar estos datos para probar y/o desarrollar modelos de predicción respecto a las cancelaciones de reservas de hotel. De esta forma, les permitiría comparar y mejorar algoritmos de clasificación y/o segmentación.
- **Educadores en Gestión:** Las personas encargadas de educar podrían utilizar estos datos para enseñar los conceptos de análisis de datos y segmentación de clientes y diferentes aplicaciones que se relacionen con la gestión hotelera.

Gracias a estos posibles casos de aplicación se podrían brindar nuevas oportunidades para la mejora de toma de decisiones, eficiencia operativa y mejorar la satisfacción del usuario o cliente de la industria hotelera.

## Descripción del Dataset

### Nombre del Dataset: Hotel Bookings

Para describir nuestro conjunto de datos, llevamos a cabo un análisis exhaustivo de cada variable disponible. Para garantizar una presentación ordenada y precisa, empleamos la librería knitr, la cual nos permitió crear una tabla que contiene las variables, sus tipos y una breve descripción de cada una. La tabla se estructura en las siguientes columnas:

Variable: Nombre de la variable en los datos.

Tipo: Tipo de datos de la variable.

Descripción: Breve descripción de la variable y su significado.

```
> library(knitr)
> kable(estructura_datos)
```

	Variable	Tipo	Descripción
-----	-----	-----	-----
hotel	hotel	character	Tipo de hotel
is_canceled	is_canceled	integer	Indicador de cancelación
lead_time	lead_time	integer	Tiempo de antelación
arrival_date_year	arrival_date_year	factor	Año de llegada
arrival_date_month	arrival_date_month	character	Mes de llegada
arrival_date_week_number	arrival_date_week_number	integer	Número de semana de llegada
arrival_date_day_of_month	arrival_date_day_of_month	integer	Día del mes de llegada
stays_in_weekend_nights	stays_in_weekend_nights	integer	Noches de fin de semana
stays_in_week_nights	stays_in_week_nights	integer	Noches de semana
adults	adults	integer	Adultos
children	children	integer	Niños
babies	babies	integer	Bebés
meal	meal	character	Comida
country	country	character	País
market_segment	market_segment	character	Segmento de mercado
distribution_channel	distribution_channel	character	Canal de distribución
is_repeated_guest	is_repeated_guest	integer	Cliente repetido
previous_cancellations	previous_cancellations	integer	Cancelaciones previas
previous_bookings_not_canceled	previous_bookings_not_canceled	integer	Reservas no canceladas previas
reserved_room_type	reserved_room_type	character	Tipo de habitación reservada
assigned_room_type	assigned_room_type	character	Tipo de habitación asignada
booking_changes	booking_changes	integer	Cambios en la reserva
deposit_type	deposit_type	character	Tipo de depósito
agent	agent	character	Agente
company	company	character	Empresa
days_in_waiting_list	days_in_waiting_list	integer	Días en lista de espera
customer_type	customer_type	character	Tipo de cliente
adr	adr	numeric	ADR
required_car_parking_spaces	required_car_parking_spaces	integer	Espacios de estacionamiento requeridos
total_of_special_requests	total_of_special_requests	integer	Solicitudes especiales totales
reservation_status	reservation_status	character	Estado de la reserva
reservation_status_date	reservation_status_date	character	Fecha de estado de la reserva

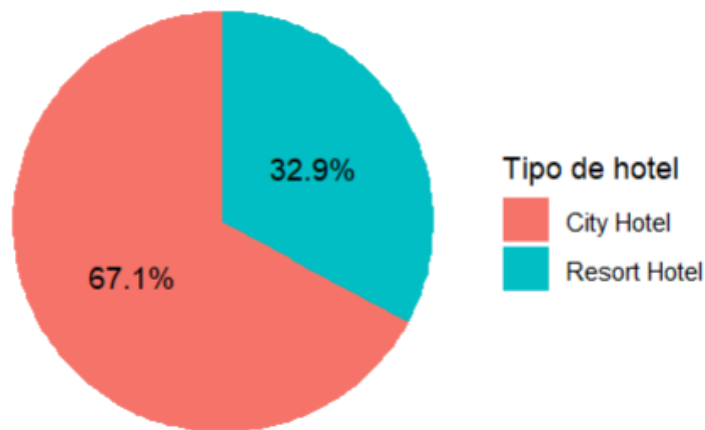
## 1. ¿Cuántas reservas se realizan por tipo de hotel?

Observamos que el número de reservas varía significativamente según el tipo de hotel.

Así mismo, la representación porcentual nos revela que los hoteles tipo City cuentan con más de  $\frac{2}{3}$  de las reservas totales, superando ampliamente las reservas realizadas de los del tipo resort.

Finalmente es conveniente concluir que los hoteles tipo City satisfacen más del doble de la demanda hotelera que los hoteles tipo Resort.

## Reservas por tipo de hotel



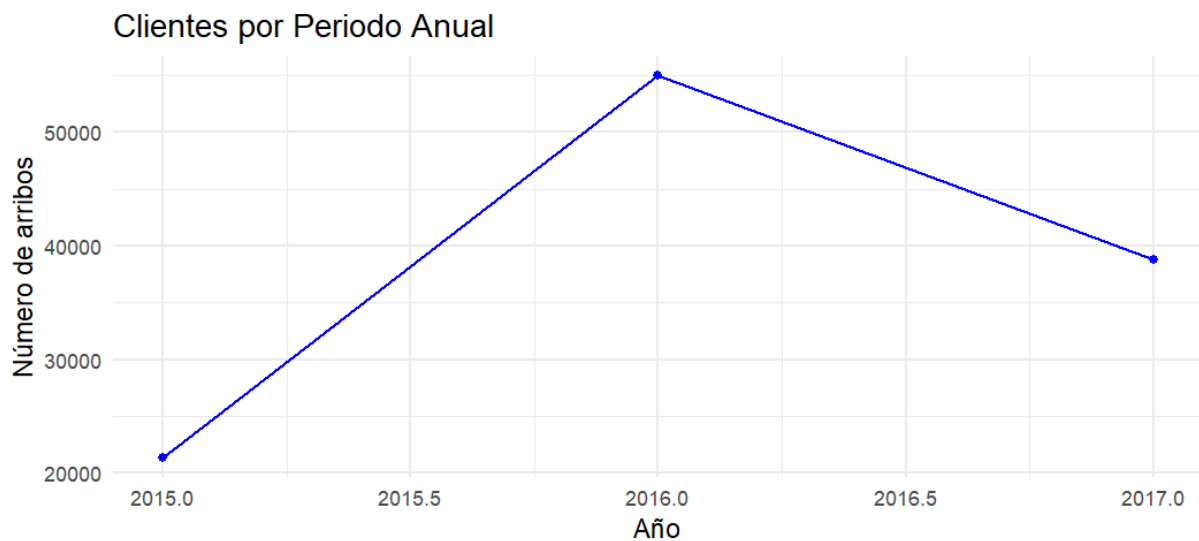
Detalle de número de reservas por hotel:

1	City Hotel	<u>77232</u>
2	Resort Hotel	<u>37851</u>

## 2. ¿Está aumentando la demanda con el tiempo?

Se realizó un análisis usando la fecha de ingreso del cliente dada su vinculación directa con la demanda de la industria hotelera (demanda efectiva). En el punto c) los registros de la demanda efectiva se mostraron limitados para el análisis, por ello se analizó también las reservas (demanda esperada) por contener mayor amplitud de fechas.

### a) Reservas por periodo anual



Número de arribos de clientes por año

A detalle en al lado del texto:

Al respecto de los límites del análisis, el primer cliente registrado llegó el : [1] "2015-07-01". Y, la fecha más antigua de reserva registrada: [1] "2014-10-17". Las más recientes no presentan mayor diferencia [1] "2017-08-31" , [1] "2017-09-10", respectivamente.

**Se evidencia una diferencia de casi un año** entre la reserva y el arribo más antiguo, y un **registro temporal más amplio en “Reservas”**

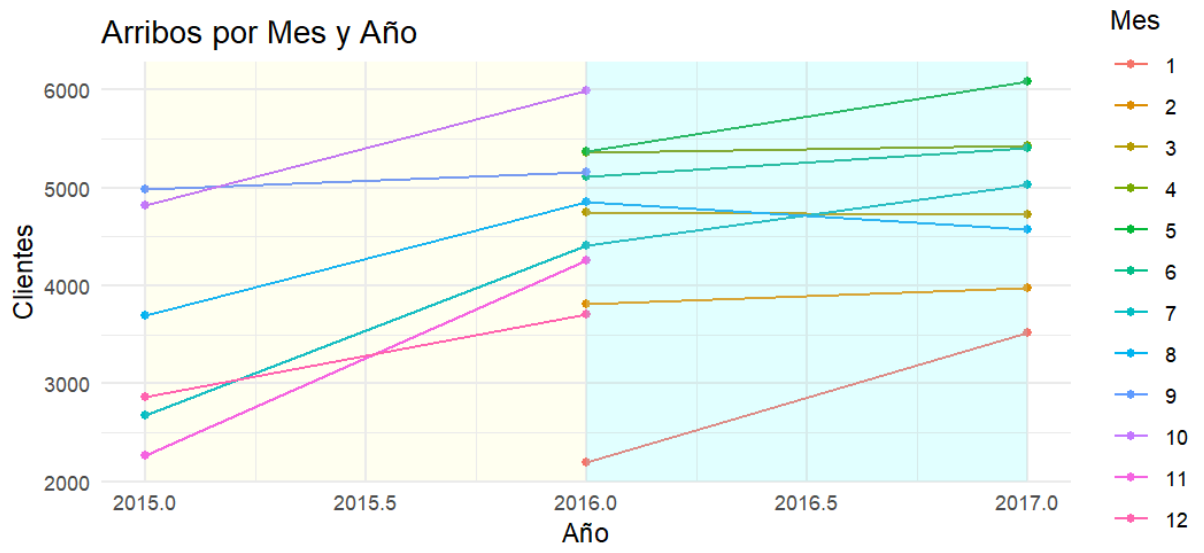
En un inicio **los rangos sirvieron para** minimizar los efectos de la incompletud en periodos ordinarios, en los años 2014 y 2017 (Reservas), pasando de entre  $\frac{1}{4}$  y un  $\frac{1}{3}$  de data ausente (2014, 2015) a un  $\frac{1}{12}$ (aprox.). **Sin embargo, el impacto en Clientes es apenas apreciable, así que no fueron aplicados en el presente grafico.**

**Al analizar el contenido del gráfico**, observamos un gran aumento para el segundo periodo y una caída para el 3er periodo. A detalle:

1	2015	<u>21306</u>
2	2016	<u>55004</u>
3	2017	<u>38773</u>

**Concluimos en** un amplio crecimiento en el primer periodo y un decrecimiento en el segundo periodo de la demanda hotelera anual. Y un registro temporal más amplio en reservas que en clientes.

#### b) Clientes por periodo anual



Al tratar de realizar un vistazo más detallado (mes/año) observamos un gráfico complicado de interpretar. Para ello nos enfocaremos en secciones.

En la **sección amarilla** apreciamos un **crecimiento en todos los meses** que cuentan con registros en 2015 y 2016, ello indica un alto crecimiento bastante consistente entre los meses medidos.

En la **sección celeste** apreciamos un único mes decreciendo,

	Ano	Mes	Clientes
	<int>	<int>	<int>
1	<u>2015</u>	7	<u>2677</u>
2	<u>2015</u>	8	<u>3701</u>
3	<u>2015</u>	9	<u>4982</u>
4	<u>2015</u>	10	<u>4825</u>
5	<u>2015</u>	11	<u>2262</u>
#	i 21 more rows		

julio (7), contando con un sorprendente crecimiento en enero (1), y uno leve en junio (6).

Al ponderar, usando la escala lineal del gráfico y el mantenimiento aparente en los demás meses compatibles. Consideramos que el gráfico indica un moderado incremento de la demanda.

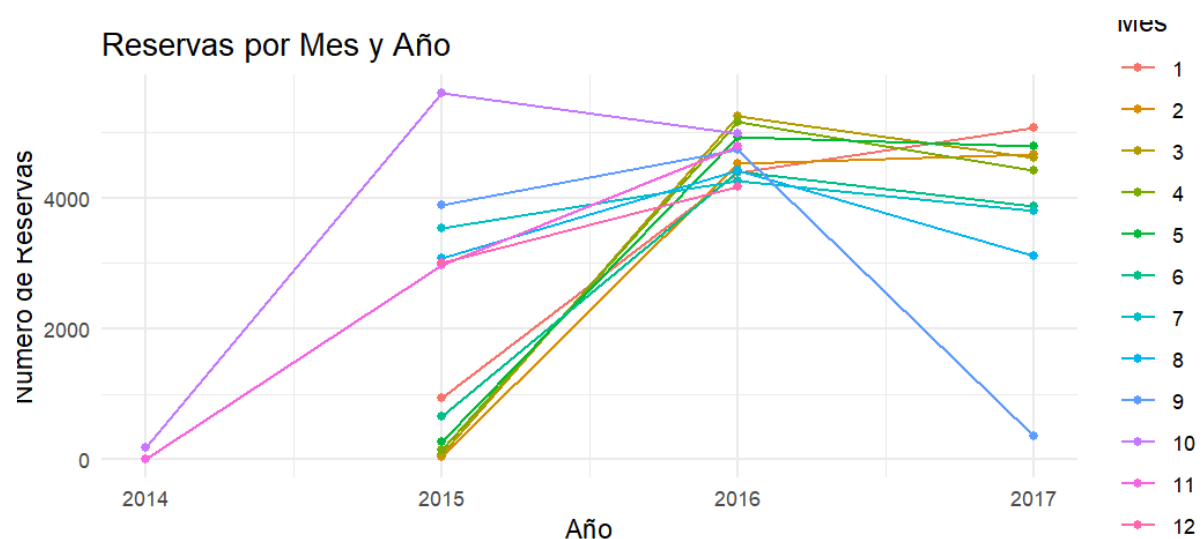
Concluimos, un crecimiento acelerado en el primer periodo y un crecimiento desacelerado en el segundo.

De este modo, al realizar un análisis más detallado **contradecimos nuestra primera conclusión** “Concluimos en [...] y un decrecimiento en el segundo periodo de la demanda hotelera anual”.

En síntesis, la demanda hotelera aumentó con el tiempo, pero ha empezado a desacelerarse

### c) Reservas de clientes por año (analítica)

La **siguiente pregunta** a responder sería si la **demanda seguirá aumentando** con el tiempo, entrando en el campo analítico.



Como nos indicaba el punto (a), contamos con mayor amplitud. Observamos crecimiento de la demanda esperada en los dos primeros periodos, sin embargo **en el tercero vemos una disminución prominente de la demanda esperada**. Sin embargo, un detalle que podríamos pasar fácilmente por alto es la relación existente en este último periodo, cuanto más actual es el mes mayor es el decrecimiento. La hipótesis de que la causa es el fin de los datos, cobra sentido **parcialmente** si consideramos la

	Ano	Mes	Reservas
	<dbl>	<dbl>	<int>
1	2014	10	180
2	2014	11	1
3	2015	1	946
4	2015	2	42
5	2015	3	83
#	i 30 more rows		

**fecha de corte de registro** de datos como la fecha más actual en la que arribó un cliente (a) "2017-08-31". Si bien la demanda decreció de forma preocupante en los últimos 6 meses el prominente decrecimiento del mes de septiembre (9) es explicado por la fecha de corte y debe ser descartado.

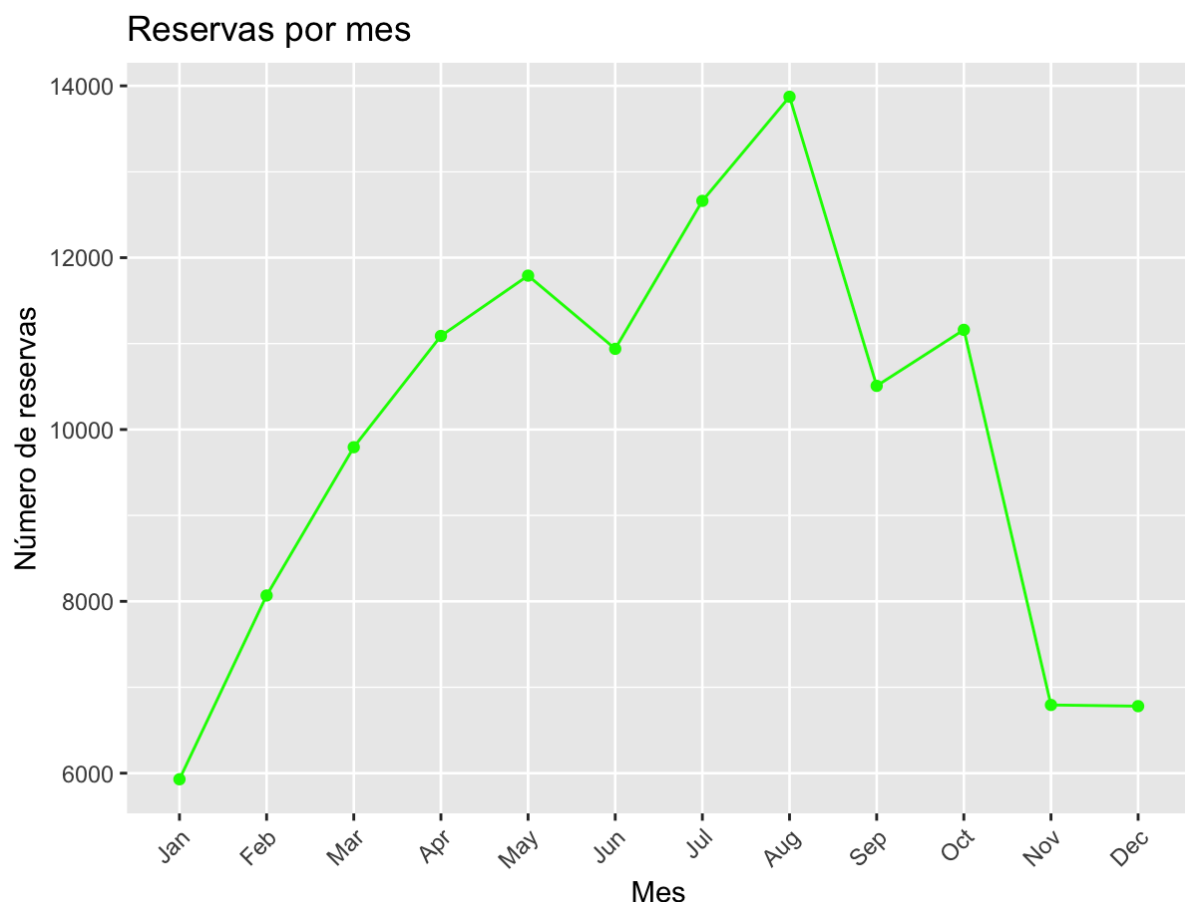
Ello nos indica que, **la demanda esperada ha decrecido moderadamente y en aumento durante los 5 últimos meses considerables.**

Teniendo en cuenta la desaceleración del crecimiento de la demanda efectiva (b) y el preocupante decrecimiento moderado de la demanda esperada en el último semestre.

**Consideramos** un futuro agudizamiento del decrecimiento de la demanda hotelera para los hoteles portugueses materia de análisis.

### 3. ¿Cuándo se producen las temporadas de reservas: alta, media y baja?

Las reservas por mes se pueden dividir en 3 temporadas las cuales son denominadas alta, media y baja, siendo la temporada alta aquel período de tiempo en el que existe una mayor cantidad de reservas, temporada baja aquel periodo de tiempo con la menor cantidad de reservas y finalmente la temporada media la cuál se ubica en medio de estas 2.



Como se puede observar en el gráfico generado en la herramienta RStudio, la temporada baja corresponde a aquél periodo de tiempo el cuál comprende los meses de:

- Noviembre
- Diciembre
- Enero
- Febrero

Los 2 primeros contando con aproximadamente 6800 reservas, Enero con menos de 6000 y Febrero con un poco más, con 8068

Siguiendo con la temporada media, podemos observar que hay 2 periodos de tiempo que podrían entrar en esta categoría:

- Marzo, Abril, Mayo, Junio

Y

- Septiembre, Octubre

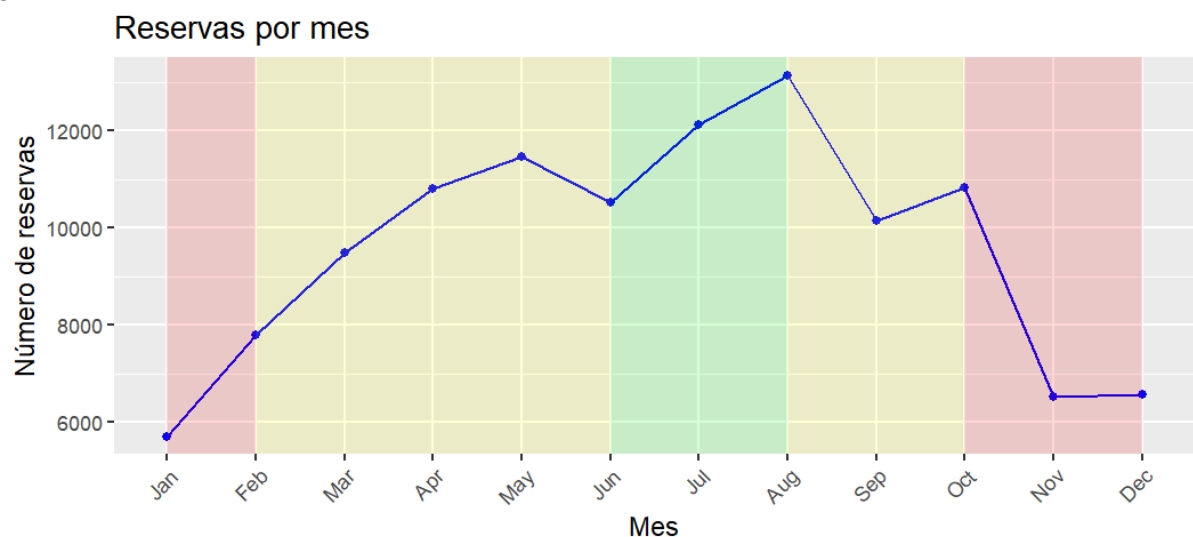
Esto ya que, durante esos meses, las reservas por mes varían de 9794 hasta 11791.

Finalmente, la temporada alta se podría considerar durante los meses de:

- Julio
- Agosto

Esto ya que ambos meses cuentan con la mayor afluencia de clientes por mes, contando con 12661 y 13873 reservas respectivamente.

De este modo nuestro análisis de las temporadas hoteleras, se resume en el siguiente grafico:

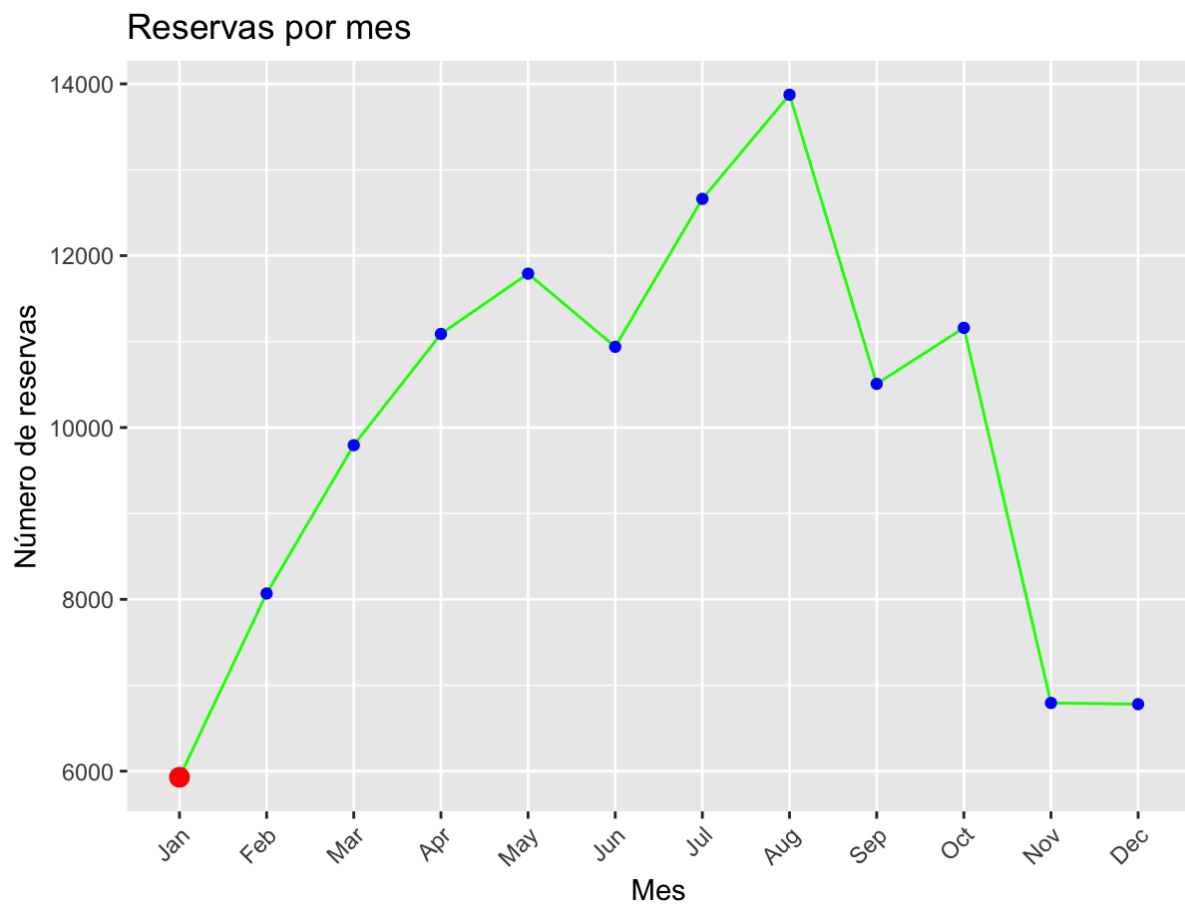




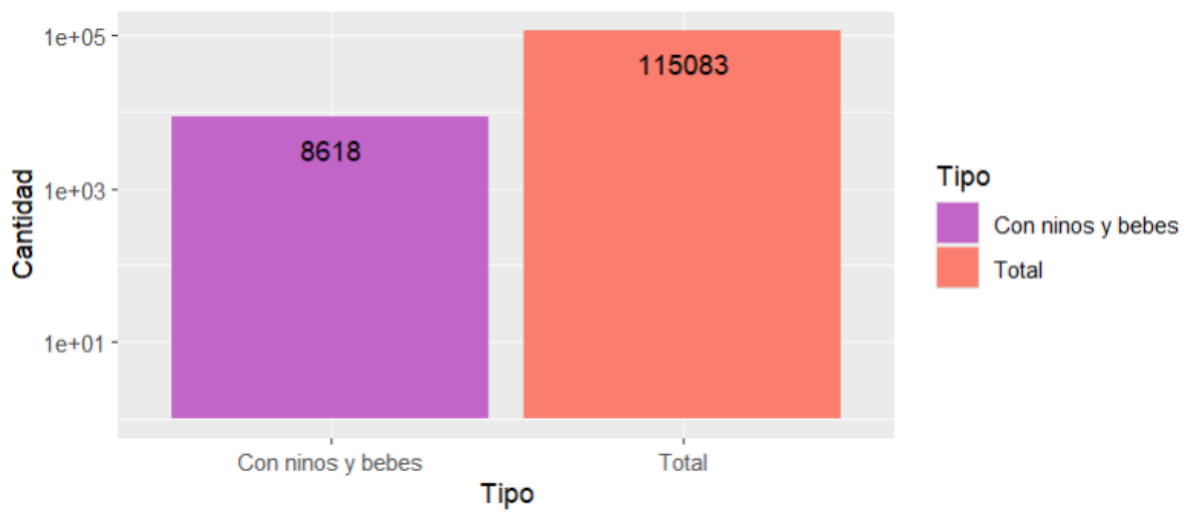
#### 4. ¿Cuándo es menor la demanda de reservas?

Cómo se mencionó en el punto anterior, la temporada baja incluye los meses de Noviembre, Diciembre, Enero y Febrero.

Siendo Enero el mes con la menor cantidad de reservas al contar únicamente con 5929, lo cuál equivale a aproximadamente el 4.97% de reservas anuales.

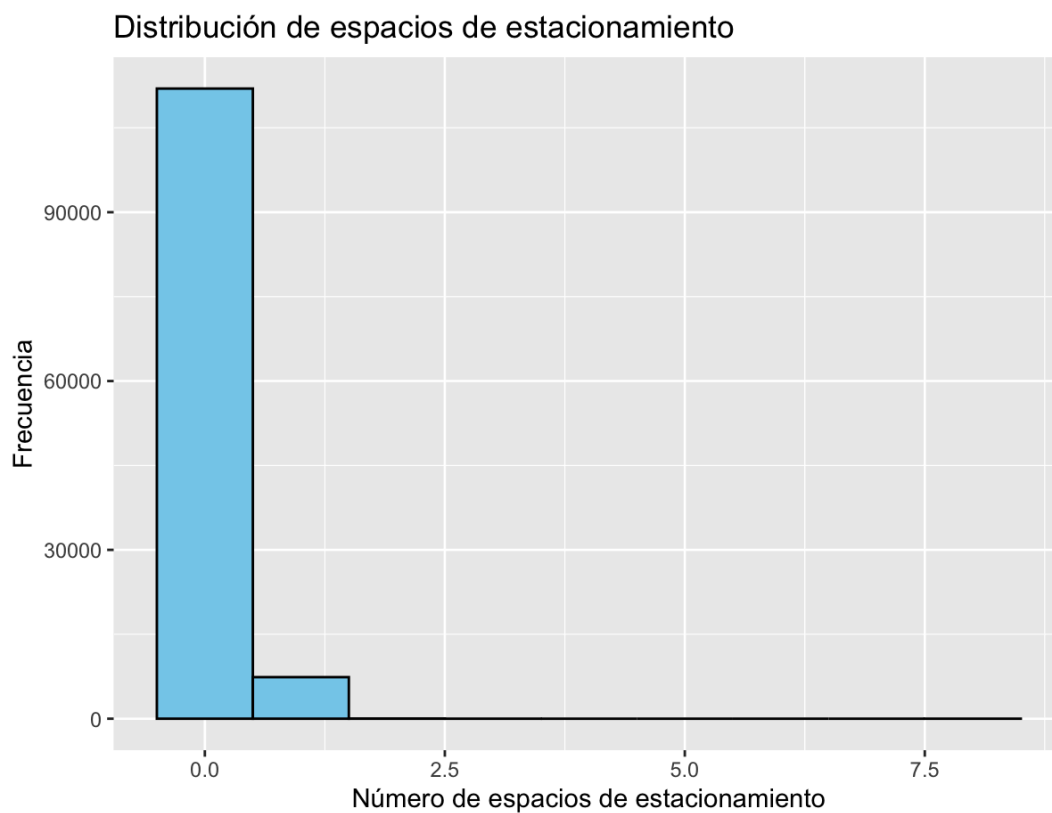


### 5. ¿Cuántas reservas incluyen niños y/o bebés?



De 115083 reservas, tan solo 8618 cuentan con algún niño y/o bebé, representando aproximadamente el 7.49% de reservas.

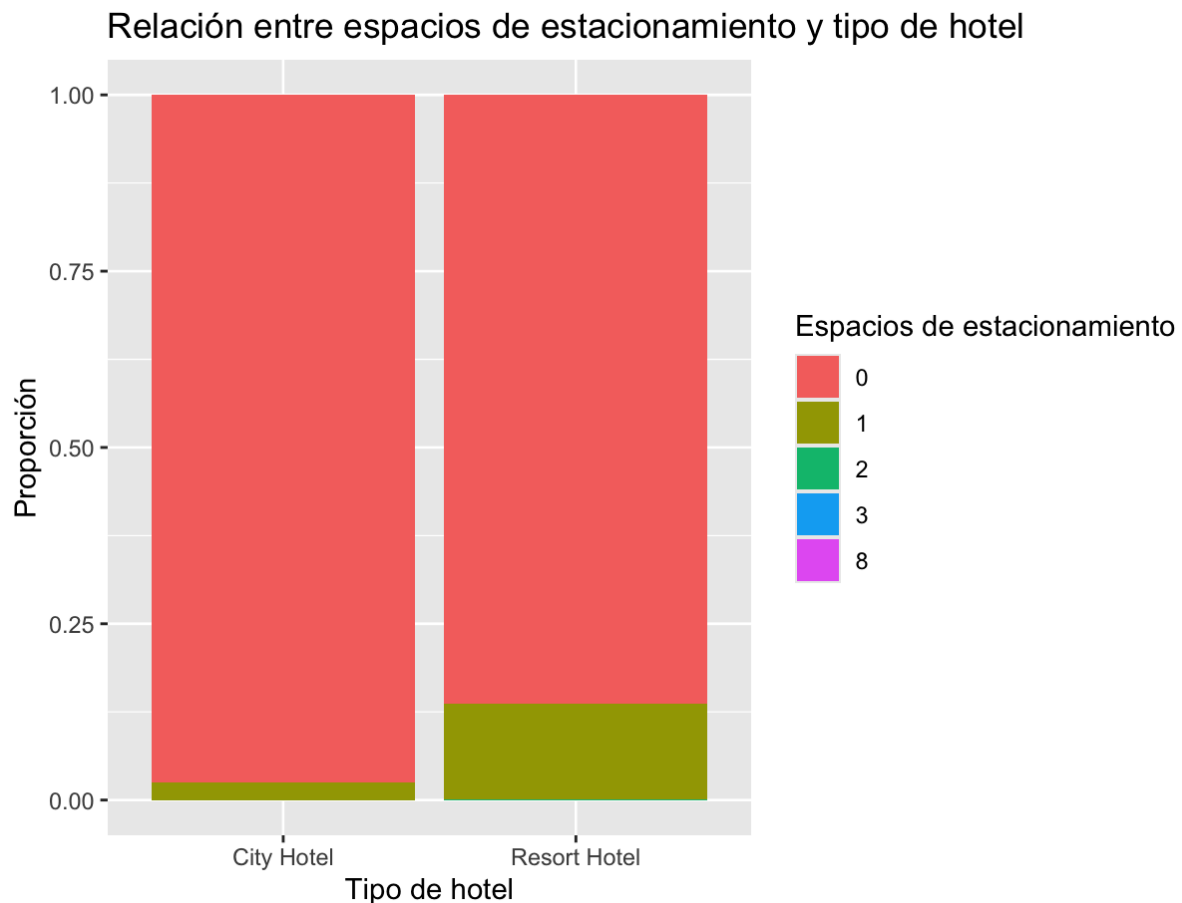
### 6. ¿Es importante contar con espacios de estacionamiento?



	Media	Mediana	Desviacion_Estandar
1	0.06112979	0	0.2395693

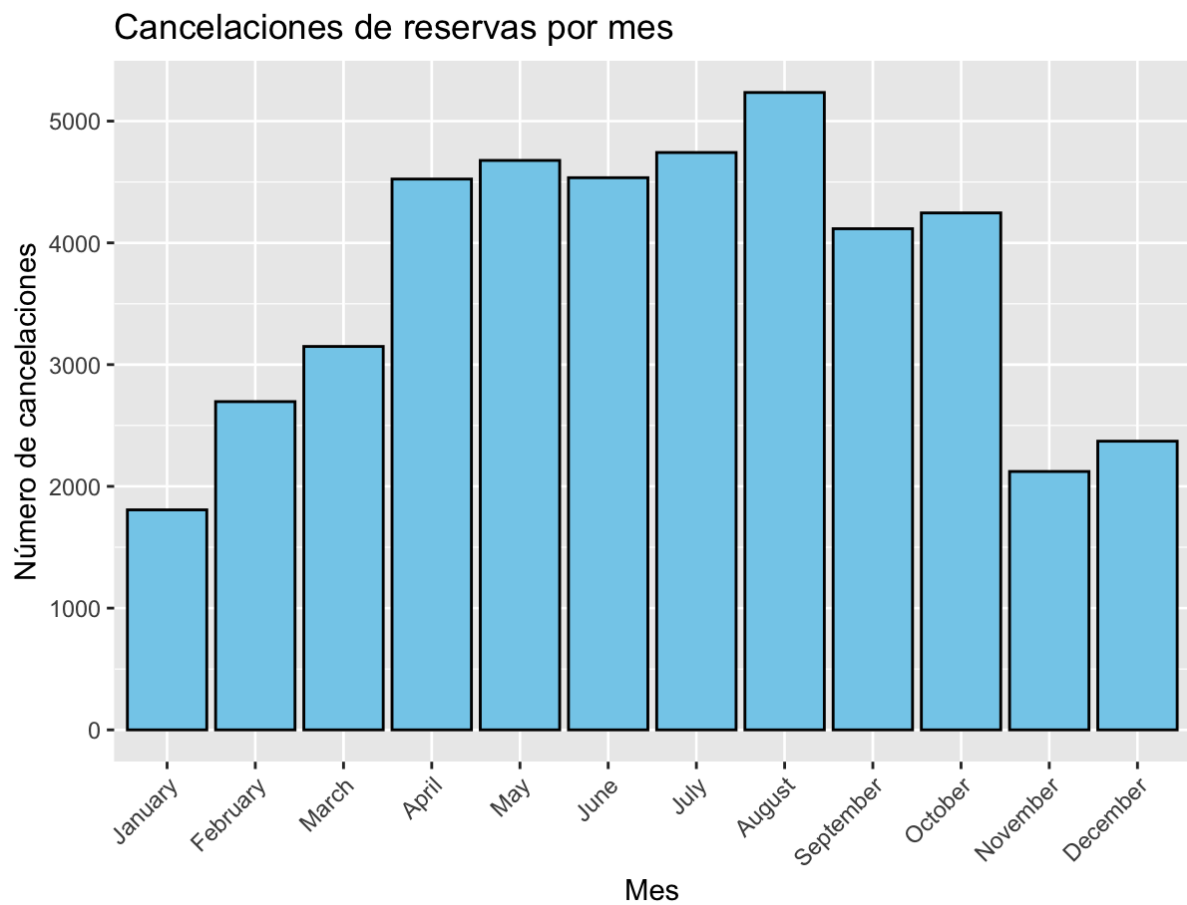
Del gráfico podemos apreciar que, si bien es cierto que algunas reservas pueden requerir de hasta 8 espacios de almacenamiento, esto no suele ser lo habitual y, de hecho, la gran mayoría de reservas no requieren del mismo. Esto lo podemos observar al ver tanto la distribución como la media de la misma, lo que nos indica que se requieren de media 0.06 espacios de estacionamiento por reserva.

Implicando que, si bien es un requisito para muchos, no es una necesidad para la gran mayoría.

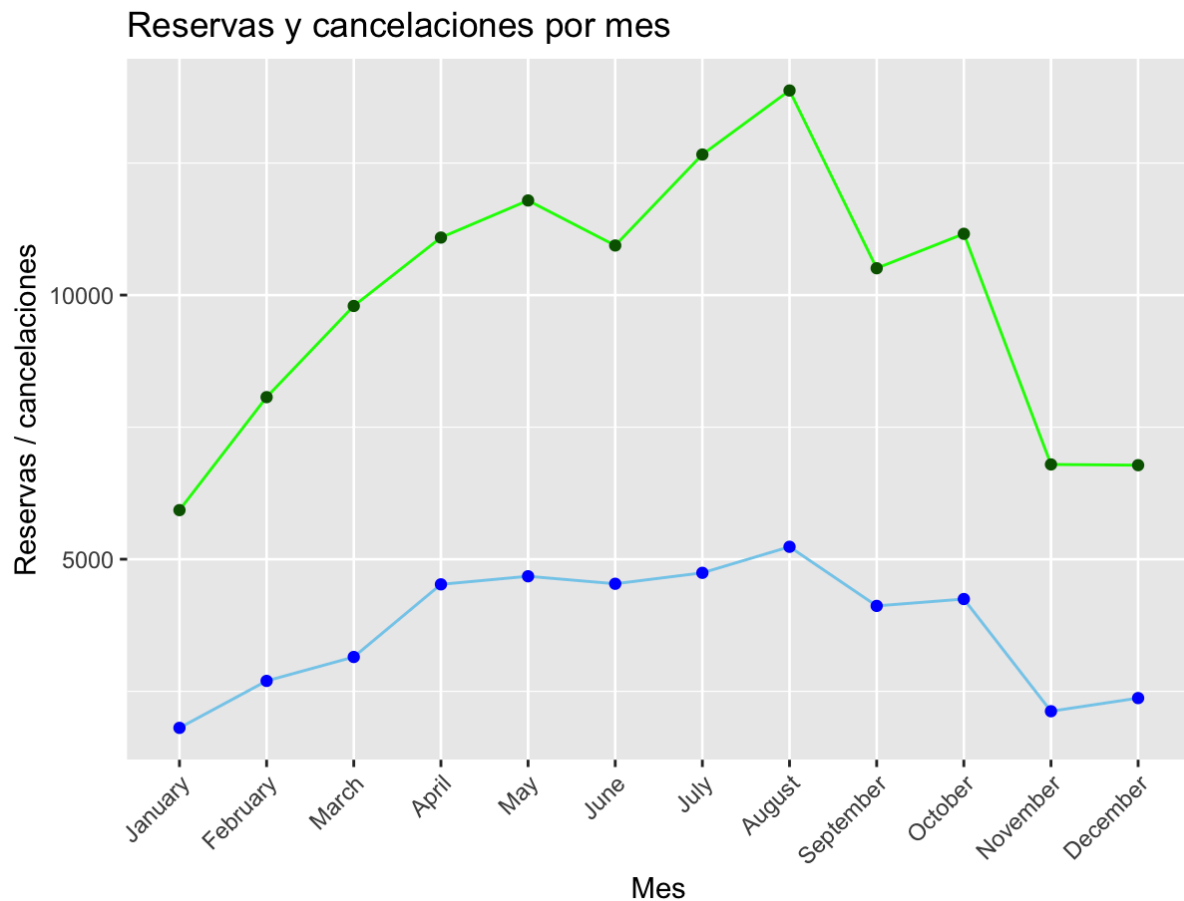


Así mismo, notamos una asimetría en la cantidad de espacios de estacionamiento requeridos por tipo de hotel, siendo aquellos hoteles Resort los cuales requieren de una mayor cantidad de espacios de estacionamiento tal como se puede observar en la segunda gráfica.

7. ¿En qué meses del año se producen más cancelaciones de reservas?



Como se puede observar en el gráfico, los meses del año en el que se producen más cancelaciones de reservas son desde Abril hasta Octubre, siendo Agosto el mes con la mayor cantidad de cancelaciones, contando con 5235.



Así mismo, podemos notar que este periodo de tiempo se ajusta perfectamente a la temporada media y alta, por lo que podemos observar claramente que existe una relación directamente proporcional entre la cantidad de reservas y las cancelaciones de las mismas y, por ende, podemos concluir que una mayor cantidad de reservas viene acompañada de una mayor cantidad de cancelaciones.

### Conclusiones:

- Respecto a las reservas por tipo de hotel, se observa que los hoteles tipo City tienen una demanda significativamente mayor que los hoteles tipo Resort, con más del doble de reservas realizadas. Este análisis revela una preferencia marcada por los hoteles urbanos sobre los resorts.
- Sobre la evolución de la demanda efectiva hasta la fecha, se evidenció un sorprendente crecimiento inicial al 2016, sin embargo el crecimiento se desaceleró constantemente hasta el 2017. De cara al futuro, los datos analizados auguran la agudización de la situación con una próxima disminución de la demanda efectiva.

- Durante los meses de Noviembre, Diciembre, Enero y Febrero, que conforman la temporada baja. Enero destaca como el mes con la menor cantidad de reservas, representando el 4.97% del total anual.
- De un total de 115,083 reservas analizadas, solo el 7.49% incluyen niños y/o bebés, ello sugiere que la mayoría de las reservas se realizan para adultos sin compañía de menores de edad.
- Se observa una variabilidad en la cantidad de espacios de estacionamiento requeridos según el tipo de hotel, teniendo los hoteles resort una predominante dominancia, a pesar de tener menos demanda.
- Existe una correlación entre la temporada de mayor demanda de reservas (temporada media y alta) y el aumento en la cantidad de cancelaciones. Los meses desde Abril hasta Octubre muestran el mayor número de cancelaciones, Agosto el mes con la mayor cantidad registrada (5235 cancelaciones), lo que sugiere una relación directamente proporcional entre la cantidad de reservas y las cancelaciones.

### Enlace a nuestro repositorio:

Para la entrega de nuestro proyecto hemos recopilado nuestros datos dentro de un repositorio en GitHub al cual se puede acceder con el siguiente enlace: <https://github.com/R0SEWT/CC216-TP-2024-1>