

INSTITUTO TECNOLÓGICO AUTÓNOMO DE MÉXICO

---



# Modelo de Predicción de Demanda y Maximización de Ingresos para Hoteles

TESIS

QUE PARA OBTENER EL TÍTULO DE  
Maestro en Ciencia de Datos

PRESENTA

Oscar Daniel Ezequiel Camarena Gómez

ASESOR  
David Lampón

MÉXICO, D.F.

2018

“Con fundamento en los artículos 21 y 27 de la Ley Federal de Derecho de Autor y como titular de los derechos moral y patrimonial de la obra titulada **“Modelo de Predicción de Demanda y Maximización de Ingresos para Hoteles”**, otorgo de manera gratuita y permanente al Instituto Tecnológico Autónomo de México y a la biblioteca Raúl Baillères Jr., autorización para que fijen la obra en cualquier medio, incluido el electrónico, y la divulguen entre sus usuarios, profesores, estudiantes o terceras personas, sin que pueda percibir por tal divulgación una prestación”

Oscar Daniel Ezequiel Camarena Gómez

---

Fecha

---

Firma

## Resumen

## Abstract

# Índice general

1. Introducción	1
2. Marco teórico	6
References	22

# Índice de figuras

# Índice de tablas

2.1. Modelos de pronosótico por categoría . . . . .	15
---	----

# Capítulo 1

## Introducción

Una práctica común dentro de la industria de la hotelería es la maximización del ingreso, mejor conocida como *revenue management*. Esta surge de la necesidad de poder obtener vender cada uno de los cuartos dentro del inventario al mayor precio posible y de esta manera poder maximizar el ingreso percibido por la propiedad derivado de la venta de cuartos (El-Gayar N, et al. 2008). La técnica de *revenue management* tiene como base el análisis de información generada por la misma propiedad así como información ajena a ella para poder vender el producto adecuado, al cliente adecuado, en el momento y precio adecuados. Estas estrategias ayudan al hotelero en su gestión diaria de diversas maneras. Le permite establecer una estrategia coherente de precios a todos los niveles, controlar la distribución multicanal, realizar previsiones de la demanda, conocer mejor las necesidades y patrones de reserva de los diversos segmentos de clientes y disponer de más información sobre el mercado y el entorno que lo envuelve. Los sistemas de *revenue management* aplican principios básicos de economía para configurar los precios y el control del inventario que se pretende vender, de hecho, hay tres categorías básicas de administración de demanda que son utilizadas dentro del *revenue management*:

- *Decisiones Estructurales*: Se decide qué formato de venta se va a utilizar; a qué segmento de mercado se va a dirigir la venta; cuáles serán las condiciones de la venta que se ofrecerán.



## CAPÍTULO 1: INTRODUCCIÓN

- *Decisiones de Precio:* Cómo se asignará el precio para los elementos dentro del inventario; diferentes categorías de precios en el mismo inventario; cambios en los precios del inventario en el tiempo.
- *Decisiones de Cantidad de Inventario:* Se decide si se acepta o se rechaza una oferta de compra; cuanto inventario se destina a cada uno de los segmentos de mercado; cuando retirar un producto del mercado para venderlo posteriormente.

Tradicionalmente los hoteles utilizan los controles de decisión de cantidad de inventario como su estrategia por "default" (Aziz, et al. 2011), sin embargo la tecnología ha estado cambiando el mercado y la manera en que un hotel vende su inventario, ya que hoy en día los costos asociados al cambio de precio del inventario y el tiempo para distribuirlos ha bajado considerablemente gracias a la adopción de un ecosistema totalmente electrónico (Bitran, Caldentey 2003). Antes del auge del *revenue management* la aerolínea *British Airways* experimentaba aplicando promociones en distintos productos dentro de su inventario para generar demanda en los asientos que sin esta promoción hubieran volado vacíos. Posteriormente la dirección de la aerolínea *American Airlines* dió un paso más adelante y crearon una practica llamada *yield management*, la cual se enfoca principalmente en maximizar los ingresos de cada ruta vendida mediante un control del inventario basado en análisis de la información histórica disponible. *American Airlines* hizo fuertes inversiones en lo que ellos llamaron *yield management* y lograron tener modelos de pronóstico de demanda robustos, así como fuertes políticas de control de inventario y estrategias de sobreventa. Estas inversiones realizadas aunado al acto de desregularización de aerolíneas (1978) llevó a que *American Airlines* tuviera tarifas muy competitivas disponibles para el mercado, promoviendo una sana competencia entre las diferentes aerolíneas de la época. Luego del éxito de la implementación de las técnicas de *yield management* en las aerolíneas, otras industrias comenzaron a implementar este tipo de técnicas. Una de estas industrias fue la hotelera, la cuál cuenta con una problemática similar a la presentada en las aerolíneas: inventario perecedero, clientes reservando cuartos con tiempo de anticipación, competencia ofreciendo tarifas de bajo costo, y grandes esfuerzos para poder balancear la oferta y la demanda. Sin embargo, el problema presentado en la industria de hoteles es

## CAPÍTULO 1: INTRODUCCIÓN

más complejo aún ya que se debe tener en cuenta que un cliente puede permanecer en la propiedad más de una noche, afectando la oferta del hotel en días posteriores a las llegadas a sus huéspedes. Las técnicas de *revenue management* / *yield management* han sido estudiadas y explotadas de manera exitosa en la industria de la aerolínea, sin embargo hay mucho trabajo por hacer aún en otras industrias, incluyendo la hotelera ya que en la medida en que esta técnica se afine los ingresos que este sector perciba por concepto de renta de habitaciones puede crecer sustancialmente. Los sistemas existentes de *revenue management* tienen dos componentes principales: un módulo que se encarga de pronósticar la ocupación del hotel en días futuros y un módulo que toma este pronóstico como base para recomendar precios para el inventario de tal forma que se maximice el ingreso total de la propiedad. Existen dos técnicas que son comúnmente utilizadas para hacer pronósticos de ocupación:

- **Análisis de indicadores históricos:** Se construyen modelos de series de tiempo con los principales indicadores del hotel (habitaciones ocupadas, reservaciones recibidas, etc) que ayudan a entender la temporalidad del hotel y el comportamiento de la demanda de esa propiedad en días específicos. Este tipo de análisis suele tener errores muy altos.
- **Regresiones generalizadas, y modelos de *machine learning*:** En los casos en dónde se cuenta con información desagregada del hotel (información de cada una de las reservaciones) es posible aplicar modelos más sofisticados como regresiones generalizadas o redes neuronales, en las cuales se puede modelar el comportamiento de la demanda del hotel utilizando toda la información disponible de la propiedad reduciendo los errores de los pronósticos arrojados por el modelo.

Es importante mencionar que una mejora del 10 % en la exactitud del pronóstico de la ocupación puede llevar a un incremento de entre el 0.5 % y 3.0 % en el ingreso del hotel (Weatherford, et al. 2003), es por ello que resulta de gran interés la mejora continua de los modelos de pronóstico de ocupación.

### Objetivo

El objetivo de el presente trabajo es presentar una propuesta de modelo de pronóstico de ocupación y maximización de ingresos para hoteles, mismo que pudiera ser encapsulado dentro de un sistema de *revenue management*. El enfoque que se utilizó dentro de esta investigación para construir el modelo de pronóstico de ocupación fue la construcción de un modelo lineal generalizado (gracias a que se contaba con la información detallada de todas las reservaciones) cuyo resultado alimentaba a un segundo modelo que ofrece precios para tarifas y tipos de cuarto que maximizan el ingreso del hotel para una serie de fechas en el futuro.

Durante esta investigación se tuvo acceso a la base de datos de reservaciones para un hotel de negocios ubicado al sur de la Ciudad de México con información a partir del 01 de enero de 2015 hasta el 01 de agosto del 2018. También se utilizaron series de tiempo con el tipo de cambio del peso frente al dólar, una matriz de la calidad vs el precio de la propiedad y su set competitivo y un calendario de eventos de la plaza en la cuál se encuentra ubicado el hotel.

La primera etapa de la investigación consistió en realizar un análisis exploratorio de los datos, donde se detectó el nivel de limpieza, y comportamiento de las distintas variables de interés. Una vez concluido el análisis se generó un modelo de pronóstico de ocupación, el cual genera curvas de *pickup* de reservaciones y ajusta una regresión Poisson, la cual es un modelo lineal generalizado usado comúnmente en problemas de conteos de datos y tablas de contingencia. La regresión de Poisson asume que la variable de respuesta  $y$  tiene una distribución de Poisson y también asume que el logaritmo de su valor esperado puede ser modelado por una combinación lineal de los parámetros desconocidos. Una vez obtenido el pronóstico de ocupación se alimenta al modelo de maximización de ingresos, el cuál resuelve un problema de optimización con restricciones teniendo como función objetivo la suma de los precios por habitación multiplicado por el porcentaje de ocupación pronosticado.

El modelo implementado en este trabajo de investigación tuvo un buen desempeño, se ejecutaron casos de validación y las curvas ajustadas por el modelo de pronóstico presentaron *pseudo r cuadradas* desde XX hasta XX y el *MAPE* observado fue de XX. Para el caso del modelo de maximización de ingresos se

## CAPÍTULO 1: INTRODUCCIÓN

tuvo como resultado una matriz de precios contra días de anticipación a la fecha objetivo para un tipo de producto dentro del inventario. Esta matriz tiene como objetivo presentar diferentes precios en el tiempo desde X días antes de una fecha objetivo, de tal forma que mientras mayor sea la antelación con la que se reserva un cuarto, menor será el precio que se tenga que pagar por él. Durante las pruebas realizadas los precios que se obtuvieron estuvieron dentro del mercado competitivo de la propiedad estudiada.

En este trabajo de investigación se presentarán los resultados de ambos modelos y los detalles sobre las limitantes así como los siguientes pasos para poder robustecer aún más este modelo.

### Justificación

Si bien los modelos de pronóstico de ocupación y los sistemas de *revenue management* han madurado a lo largo de más de 30 años de haber sido implementados por primera vez en distintas industrias, aún hay mucho trabajo por hacer ya que el poder de cómputo y la capacidad de procesamiento ha aumentado significativamente a lo largo del tiempo abriendo nuevas oportunidades y facilitando la implementación de técnicas sofisticadas para el manejo de datos y el modelado estadístico que puede ser de gran ayuda durante la implementación de este tipo de modelos.

El mercado hotelero es altamente dinámico, genera inmensas cantidades de información día con día, misma que es almacenada en servidores y pocas veces es utilizada para poder obtener una ventaja competitiva. De ser exitoso el modelo aquí implementado se podría aprovechar la investigación realizada para generar un marco de trabajo que ayude a este sector en particular a aprovechar la información generada detrás de sus datos, de tal suerte que puedan obtener mayores ingresos por los servicios ofrecidos generando un ambiente de competencia sana entre los distintos jugadores de este mercado.

## Capítulo 2

# Marco teórico

En este capítulo se introducirá y discutirá el trabajo previo realizado y los conceptos necesarios para comprender las distintas técnicas y enfoques utilizados para la construcción de un modelo de predicción de demanda y maximización de ingresos así como el impacto que los resultados tienen dentro de la industria hotelera. Se dará una breve explicación del proceso de gestión de una propiedad mencionando cuales son los principales indicadores de control del proceso, así como las variables estudiadas para la toma de decisiones.

### Gestion de una propiedad

El proceso de gestión de propiedades o de hoteles comprende una serie de actividades que tienen como objetivo el garantizar la rentabilidad de un proyecto de arrendamiento. Dentro de las actividades que se llevan a cabo están las siguientes:

- Gestión de *canales de venta*
- Gestión de *segmentos de mercado*
- Gestión de *precios por tarifa*
- Análisis del *comportamiento de la plaza*

## CAPÍTULO 2: MARCO TEÓRICO

- Análisis de *indicadores principales*

A continuación profundizaremos en cada una de las actividades que forman parte de este proceso.

### Gestión de canales de venta

El equipo administrativo del hotel o propiedad, deben analizar diariamente el comportamiento de sus canales de venta o distribución. Típicamente un hotel tiene un catálogo estándar de canales de distribución:

- Sitio Web
- Call Center / FrontDesk
- Agencias de viajes en línea
- Globalizadores Agencias de viajes

Cada uno de los canales de venta van dirigidos a un segmento de mercado en específico y lo que el equipo de administración debe hacer es asegurar que la demanda del hotel se mantenga lo suficientemente alta para poder asegurar la continuidad operativa del negocio. Para lograr esto se genera una estrategia de distribución, en la cuál se designa parte del inventario a cada uno de los canales y se fija un precio a cada una de las tarifas a ofertar. Los precios para un producto en específico tienden a variar entre los diferentes canales y esto tiene que ver con el costo asociado a la operación de cada uno de los canales. Por ejemplo, el sitio web típicamente es un canal de reservación propio del hotel, lo que significa que las ventas generadas por este canal son libres de comisiones; por el contrario, las agencias de viajes en línea tienden a cobrar una comisión por la venta generada por esos canales.

Lo que la administración del hotel debe conseguir es poder tener la mezcla de ventas entre sus canales que maximice el ingreso, de tal forma que los costos generados por las ventas no *canibalicen* las ganancias de la propiedad. Es con esta restricción donde se descartaría una asignación total del inventario a las agencias de viajes en línea, ya que la comisión de este canal puede ser de hasta el 25 % del total de la venta, lo cual podría representar una pérdida de hasta  $\frac{1}{4}$  de

las ganancias que se pudieron haber obtenido. Si por el contrario asignáramos todo el inventario disponible a los canales propios del hotel, si bien no se paga una comisión, la demanda generada por estos canales es inferior a la demanda generada por las agencias de viajes en línea.

### Gestión de segmentos de mercado

La gestión de los segmentos de mercado es similar a la gestión de los canales de venta ya que estos están relacionados a los canales de distribución.

Las propiedades típicamente distinguen y segmentan a sus clientes en las siguientes categorías:

- *Clientes directos*: Aquellos clientes que reservan en algún canal directo del hotel
- *Clientes de negocios*: Aquellos clientes que pertenecen a una empresa que tiene alguna tarifa convenida con el hotel. Típicamente esta tarifa está sujeta a una producción de cuartos durante un periodo
- *Clientes de mayoreo*: Aquellos clientes que reservan utilizando un canal mayorista; puede ser una agencia de viajes en línea o mediante un agente de viajes.
- *Grupos*: Aquellos clientes que llegan a la propiedad como parte de un grupo.

Para gestionar eficientemente los segmentos de mercado se debe crear una estrategia de asignación de precios a las diferentes tarifas que van dirigidas a cada uno de ellos. Por ejemplo, un *cliente directo* que llega al hotel sin reservación, es propenso a pagar un precio más alto por una habitación disponible, al contrario de un cliente que viene como parte de un grupo que reservó con mayor tiempo de anticipación con una tarifa mucho más baja. El equipo administrativo debe cuidar los segmentos de mercado que ocupan la propiedad ya que el tener un grupo muy grande ocupándola significa muchas veces tener un hotel lleno a una tarifa muy baja y con mayor desgaste en las habitaciones, y por el otro lado, no siempre es posible llenar un hotel con clientes que llegan sin reservación los cuales están dispuestos a comprar una habitación a un precio más elevado.

## Gestión de precios por tarifa

Una de las actividades mas importantes del proceso de gestión de una propiedad es la gestión de los precios por tarifa ya que generalmente las tarifas disponibles en un hotel están asociadas a un segmento de mercado y a un canal de reservación. Para poder realizar esta actividad se debe establecer una estrategia en la cual se define si se quiere incrementar la ocupación de la propiedad o si se quiere incrementar el *RevPAR*(revenue per available room). Una vez establecida la estrategia inicial se deben modificar los precios para las tarifas que impactan los canales y segmentos donde se quiere fomentar la demanda o incrementar el precio. Si el objetivo es el primero, lo mas probable es que el cambio implique una baja en las tarifas o creación de promociones sujetas a diversas restricciones, en cambio, si el objetivo es el segundo muy probablemente el cambio implique una alza en algunas tarifas y el cierre de algunas tarifas en donde este incremento no se pueda llevar a cabo por temas contractuales, por ejemplo las tarifas convenidas con otras empresas o clientes.

Es importante mencionar que la ejecución de esta tarea dentro del proceso de gestión depende del análisis continuo de la información histórica generada por el hotel (comportamiento de la propiedad en años anteriores en la misma fecha sujeta al estudio), conocimiento de la plaza en donde se ubica la propiedad (depende de la experiencia del equipo administrativo del hotel), información ajena al hotel (situación marco económica, eventos cercanos a la plaza, desempeño de la competencia cercana a la propiedad, etc). Muchas veces el conocer esta información implica un gran esfuerzo por parte del equipo ya que la información tiende a estar distribuída en diversos sistemas o en ocasiones deberán recabarla físicamente (visitando a su competencia o haciendo llamadas telefónicas) dejando un riesgo latente de recibir información que no es precisa.

Otra de las fuertes dependencias para la ejecución efectiva de esta tarea es contar con sistemas que distribuyan los nuevos precios en todos los canales disponibles en poco tiempo y con poco esfuerzo ya que de lo contrario no tendria sentido realizar esta labor si al final los precios no pueden ser cambiados dentro de un tiempo en el que el mercado pueda responder de acuerdo a lo planeado durante el análisis de la información. Una vez que los precios son alterados, el equipo deberá analizar nuevamente la información para asegurar que los cambios tuvieron los



efectos deseados, de lo contrario, se deberá cambiar la estrategia lo antes posible para evitar afectar el desempeño de la propiedad.

## Análisis del comportamiento de la plaza

En el sector de la hotelería en México, se le llama plaza a la ubicación geográfica en dónde se encuentra una propiedad y típicamente agrupa a propiedades independientes o de otras marcas (set competitivo), oficinas, corporativos, centros comerciales, y lugares de interés turístico. La plaza genera gran cantidad de información que es de interés para el equipo administrativo de una propiedad, ya que el desempeño financiero de esta depende mucho de la situación que presenta la plaza en un momento determinado.

Los hoteles que comparten una plaza generalmente están dispuestos a compartir información de su propiedad con otros, de esa forma pueden tener un conocimiento mas amplio sobre la realidad que se vive en ella. Si una propiedad requiere obtener datos de su competencia, basta con una simple llamada telefónica a la propiedad de interés para solicitar los datos de interés, por ejemplo: *cuartos ocupados*, *cuartos disponibles*, *cuartos fuera de servicio*, *ingresos generados por la venta de habitaciones*. Con esta información se pueden calcular algunos indicadores que son indispensables para el analisis del comportamiento de la plaza:

- % de Ocupación:  $\frac{\text{cuartos ocupados}}{\text{cuartos disponibles}}$
- ADR (Average Daily Rate o Tarifa Promedio):  $\frac{\text{ingresos}}{\text{cuartos ocupados}}$
- RevPAR (Revenue per available room o Tarifa Efectiva):  $\frac{\text{ingresos}}{\text{cuartos disponibles}}$

Para poder conocer la situación de una propiedad frente a la competencia en una plaza se generan otros indicadores que facilitan este tipo de análisis.

- Penetración de ocupación:  $\frac{\% \text{ ocupacion}}{\% \text{ ocupacion plaza}}$
- Penetración de tarifa:  $\frac{\text{tarifa promedio}}{\text{tarifa promedio plaza}}$

El hotel o propiedad que cuente con los valores mas altos para estos indicadores será el hotel con mayor demanda y mayores ingresos dentro de la plaza, sin

## CAPÍTULO 2: MARCO TEÓRICO

embargo, para que el análisis sea efectivo se debe escoger al set competitivo de manera adecuada, de tal forma que los hoteles contenidos dentro de este set sean de la misma gama y vayan dirigidos al mismo público objetivo.

### **Análisis de indicadores principales**

Como se mencionó anteriormente, existen indicadores de desempeño de una propiedad u hotel que deben ser estudiados de manera diaria para poder tomar decisiones que mejoren el desempeño de esta, asegurando una continuidad operativa y la rentabilidad del inmueble:

- % ocupación
- ADR
- REvpar

Típicamente durante el análisis realizado, se compara el desempeño de cada uno de estos indicadores de manera diaria, mensual y anual, comparando los valores contra el año anterior y buscando siempre una variación positiva entre años. En caso de tener variaciones negativas se debe buscar la causa de esta variación. Se puede tratar de una baja en la demanda general de la plaza, un periodo de recesión financiera generalizada, un nuevo competidor en la plaza, etc. Una vez identificado la causa se deben tomar las acciones necesarias para generar demanda a la propiedad en cuestión. Típicamente la forma más efectiva de hacerlo es modificar los precios de tal forma que el mercado responda favorablemente sin afectar la rentabilidad del inmueble.

### **Antecedentes: Modelos de Revenue Management**

Podemos definir la práctica de *revenue management* como la aplicación de técnicas analíticas que intentan predecir el comportamiento de un consumidor al nivel de un micro mercado, optimizando la disponibilidad de productos y sus precios de tal forma que maximicen el crecimiento de los ingresos. El principal objetivo de esta práctica es vender el producto indicado al cliente indicado al tiempo indicado en el precio indicado. La esencia de esta práctica radica en entender

## CAPÍTULO 2: MARCO TEÓRICO

la percepción del producto que tienen los clientes y alinear los precios de estos productos de acuerdo a esta haciendo una correcta segmentación de los clientes objetivos.

Es importante mencionar que la disciplina de *revenue management* se compone de dos partes principales:

1. Pronóstico de ocupación
2. Asignación de precios

En las siguientes secciones discutiremos acerca de los trabajos previos sobre los dos principales componentes de esta disciplina.

### Trabajo previo: modelos de pronóstico de ocupación

La base de una práctica efectiva de *revenue management* es un buen modelo de pronóstico de demanda, ya que el resultado de este será la entrada del modelo de optimización de precio de tal suerte que mientras menor sea el error en el pronóstico de la demanda, mejor será la optimización de los precios para cada uno de los productos sujetos al análisis.

Se tiene registro de trabajo previo para el desarrollo de modelos de pronóstico de demanda en diferentes industrias, algunas de ellas son: *líneas aéreas*, *retail*, *telecomunicaciones* y *hospitalidad*, siendo la primera la que cuenta con modelos de pronóstico más maduros, este hecho obedece a que en esta industria es en dónde a principios de los años 80's nace la disciplina de *revenue management*.

La venta de asientos en un vuelo es un caso dónde naturalmente se puede utilizar un modelo de optimización de ingresos ya que cuenta con ciertas características que hacen que el modelado del problema sea muy intuitivo:

- Se cuenta con un inventario reducido
- El inventario es perecedero (una vez que el avión despegue, los asientos vacíos no generan un ingreso)
- Se cuenta con un amplio segmento de mercado que genera demanda sobre el producto

## CAPÍTULO 2: MARCO TEÓRICO

- Los asientos se compran con distintos tiempos de anticipación (hay clientes que compran con un mayor tiempo de anticipación que otros)
- Se conoce el tiempo de estancia de cada uno de los clientes en la aeronave.

Las condiciones mencionadas anteriormente facilitan la construcción de un modelo de pronóstico de demanda. A continuación haremos una breve descripción de la evolución de este tipo de modelos a lo largo del tiempo.

Beckmann y Bobkowski (1958) presentaron modelos estadísticos que describen reservaciones de pasajeros, cancelación de reservaciones y *no shows*. En su trabajo los autores comparan modelos utilizando diferentes distribuciones probabilísticas: *Poisson*, *Binomial Negativa* y *Gamma* aplicadas en datos generados por aerolíneas con la finalidad de modelar las reservaciones efectivas y propone una condición óptima para un nivel de sobreventa para un vuelo en particular.

Posteriormente, otros autores continuaron explorando nuevas metodologías para poder construir modelos de predicción de demanda. Uno de los autores más citados en las investigaciones de modelos de pronóstico de demanda es Lee (1990) que presenta un artículo en donde explica dos enfoques para resolver el problema de pronóstico de ocupación en aerolíneas:

- Modelos de series de tiempo
- Modelos basados en información de reservaciones

Los modelos basados en series de tiempo consideran únicamente las series de tiempo con la información de llegadas o porcentajes de ocupación aplicando modelos de series de tiempo (como suavizamiento exponencial, ARIMA, etc). En estos casos no se utiliza la información propia de las reservaciones.

El enfoque basado en reservaciones hace uso de la información contenida en las reservaciones para pronosticar llegadas futuras. En este tipo de modelos típicamente se considera el concepto de *pick-up*. Esto quiere decir que dadas  $K$  reservaciones para un día futuro  $T$ , esperamos tener un *pick-up* de  $N$  reservaciones más desde este momento hasta el momento  $T$ . El pronóstico entonces será  $K + N$

Existen dos versiones del modelo de *pick-up* (Weatherford y Kimes, 2003). La versión aditiva del modelo en donde a las reservaciones registradas en el hotel se

## CAPÍTULO 2: MARCO TEÓRICO

le suman el promedio de reservaciones que típicamente llegan en un periodo entre la fecha actual y la fecha de llegada del huésped (tomando la temporalidad del hotel en cuenta), y la versión multiplicativa, la cual es muy similar, con la única diferencia de únicamente se toma en cuenta una fracción de las reservaciones actualmente registradas en el hotel.

Existen trabajos en los cuales se presenta un modelo avanzado en el cual se combinan modelos de series de tiempo y modelos basados en reservaciones, tal es el caso del artículo presentado por Flides y Ord (2002) en el cuál se concluye que la combinación de ambos modelos generalmente tienen un mejor desempeño en cuanto al pronóstico de la demanda. Otro trabajo que presenta un enfoque similar es el realizado por Sa (1987) en el cuál se utiliza una regresión multiple para desarrollar un modelo de pronóstico combinado, en este caso, la variable dependiente eran las reservaciones pendientes por llegar mientras la variable independiente incluía el número de resrvaciones confirmadas, un índice de temporalidad, índice semanal y un promedio del comportamiento histórico de las reservaciones pendientes. Esta regresión se corría para varios días antes de la llegada de los clientes ( $t = 7, 14, 21, \text{ y } 28$ ).

Es importante hacer las siguientes observaciones sobre los modelos comentados anteriormente:

- Los modelos de regresión lineal asumen que existe una correlación entre el número de reservaciones confirmadas y el número final de reservaciones (en el día 0), de tal forma que:  $\text{pronostico}_{dia0} = a + b * \text{reservaciones}_{diaN}$
- Los modelos de regresión logarítmica asumen que existe una correlación entre el número de reservaciones confirmadas y el número final de reservaciones (en el día 0), de tal forma que:  $\log(\text{pronostico}_{dia0}) = a + b * \log(\text{reservaciones}_{diaN})$
- El modelo aditivo suma las reservaciones actuales al promedio histórico del *pickup* de reservas desde el día de la lectura  $diaN$  hasta el día de la estancia  $dia0$ :  $\text{pronostico}_{dia0} = \text{reservas}_{diaN} + \sum_{i=0}^n \frac{\text{pickup}_i}{n}$
- El modelo multiplicativo multiplica las reservaciones actuales por la proporción del promedio histórico de reservaciones desde el  $diaN$  hasta el día de la estancia  $dia0$ :  $\text{pronostico}_{dia0} = \text{reservas}_{diaN} * \frac{\text{reservas}_n - \sum_{i=0}^n \frac{\text{pickup}_i}{n}}{\text{reservas}_n}$

Categoría	Modelo
Historico	1. Mismo día año anterior 2. Promedios móviles 3. Suvizamiento exponencial 4. ARIMA
Avanzado	1. Aditivo a. Pickup Clasico b. Pickup Avanzado 2. Multiplicativo a. Curva de reservaciones
Combinados	1. Promedios ponderados de modelo histórico y modelo avanzado 2. Regresión 3. Modelo de información completa

Tabla 2.1: Modelos de pronóstico por categoría

La tabla 2.1 resume los modelos más comunes utilizados para pronosticar ocupación, agrupados por categoría.

## Desempeño de modelos de pronóstico

Se han realizado estudios con el objetivo de comparar el desempeño de los modelos previamente presentados, a continuación presentaremos los resultados obtenidos.

Wickham (1995) estudió la efectividad de una variedad de modelos de pronóstico de ocupación utilizando datos de aerolíneas. Durante este trabajo se comparan modelos históricos de ocupación (utilizando promedios simples y promedios ponderados) y modelos de *pickup* (clasicos y avanzados) llegando a la conclusión de que los segundos son más efectivos, al menos para los datos aplicados durante su análisis.

Weahterford (1998) comparó los modelos aditivos frente a los modelos multiplicativos y modelos de regresión dentro del contexto de las aerolíneas y encontró que los modelos aditivos y de regresión tienen un mejor desempeño que los multiplicativos.

## Modelos de machine learning

Si bien las técnicas estadísticas anteriormente mencionadas pueden predecir efectivamente niveles de ocupación y demanda, se requieren de amplios conocimientos en estadística y de largos procedimientos para poderlas aplicar de manera tal que funcionen correctamente o de otra manera se puede optar por el uso de *software* comercial el cuál resulta muy costoso la mayoría de las veces. Para solventar estos hechos, algunos autores presentan un enfoque diferente para construir modelos de predicción de ocupación, tal es el caso de **Caicedo-Torres, Payares** (2016) quienes proponen el uso de algoritmos de *Machine Learning* para construir modelos predictivos de ocupación y demanda en la industria de la hospitalidad. La ventaja de estos modelos es que están listos para ser utilizados por el staff administrativo de la propiedad sin la necesidad de contar con amplios conocimientos de estadística. Además, estos modelos tienen la ventaja de poder ser empaquetados dentro de aplicaciones de bajo costo, o de ser ejecutados en una infraestructura basada en la nube, lo cual significan soluciones rentables para el sector de hospitalidad.

A continuación presentaremos a detalle algunos algoritmos de M.L. utilizados para la construcción de este tipo de modelos.

### Regresión de Ridge

La regresión de ridge es un algoritmo de regresión que incluye un término de regularización en su función de costos con el fin de lograr una mejor aproximación al momento de generalizar el modelo. La regresión de Ridge penaliza el algoritmo con un factor proporcional a la norma Euclidiana (L2) del vector de parámetros.

$$J(\theta) = \sum_{i=1}^n (\theta^T x_i - y_i)^2 + \alpha \|\theta\|^2$$

Dónde el factor de regularización  $\alpha$  controla la importancia relativa de la penalidad compleja. Para optimizar esta ecuación se debe encontrar la derivada de la función de costos e igualarla a cero, de tal forma que el vector de parámetros que minimiza el error es:

$$\theta = (X^T X + \lambda I)^{-1} X^T y$$

## CAPÍTULO 2: MARCO TEÓRICO

Las nuevas predicciones se pueden calcular como sigue:

$$h(X) = \theta^T x = \sum_{i=1}^n \theta_i x_i$$

### Ridge Regresión con Kernel

Este método consta de aplicar el truco del Kernel a la regresión previamente comentada, esto con la finalidad de proyectar los datos originales en un espectro de variables sin tener que explícitamente calcular las transformaciones. Para esto, se comienza con el vector de pesos optimizados para la Regresión de Ridge:

$$\theta = (X^T X + \lambda I)^{-1} X^T y$$

Se manipula algebraicamente para obtener:

$$\theta = X^T (X X^T + \lambda I)^{-1} y$$

De tal forma que una predicción puede ser obtenida por:

$$h(x') = \theta^T x' = y^T (X X^T + \lambda I)^{-1} X x'$$

Una función de Kernel evalúa a un producto punto de la forma

$$f(x_1, x_2) = \phi(x_1) \cdot \phi(x_2)$$

sin tener que explícitamente calcular la transformación  $\phi$ . Si definimos  $K_{i,j} = f(x_i, x_j)$  y  $k_i = f(x_i, x')$ , donde  $x_i$  es la  $i$ -ésima fila de la matriz  $X$ , entonces

$$h(x') = \theta^T x' = y^T (K + \lambda I)^{-1} k$$

Ajustamos un modelo en un espacio dimensional mayor inducido por  $\phi$  sin tener que incurrir en una complejidad computacional adicional. Se pueden utilizar varios Kernels, entre los más populares están el Kernel polinomial

$$k(x_1, x_2) = (\gamma x_1^T x_2 + \beta)^\delta$$

Y el Kernel Gaussiano

$$k(x_1, x_2) = \exp\left(-\frac{\|x_1 - x_2\|^2}{2\delta^2}\right)$$



### Red neuronal multicapa

Una red neuronal es un modelo de M.L. capaz de clasificar y realizar regresión. Esta compuesta por  $n$  capas totalmente conectadas de neuronas artificiales formando un grafo conectado (en el caso más simple se tiene una capa oculta y una capa de salida). En este modelo, la salida de cada capa de neuronas será la entrada de la siguiente capa. La función de activación es dada por la siguiente ecuación

$$y = f\left(\sum_i^n w_i x_i + b\right)$$

donde  $f$  es una función de activación. Entre las opciones más populares se encuentran la sigmoide  $y = (1 + \exp(-x))^{-1}$  y la función lineal  $y = x$ . La elección de la función de activación en la capa de salida define el tipo de red neuronal (si clasifica o realiza regresión)

Las redes neuronales son entrenadas mediante un algoritmo llamado *propagación hacia atrás* (*backpropagation*) que es una aplicación del algoritmo de descenso por gradiente que toma en cuenta la contribución de las neuronas de la capa oculta al error  $E$

### Red neuronal de base radial

Las redes neuronales de base radial (RBF) usualmente tienen una sola capa oculta que emplea una serie de funciones llamadas *funciones de base radial* como detectores de características. Las redes RBF pueden ser utilizadas para realizar regresiones o clasificaciones. En este modelo la capa oculta utiliza función gaussiana como función de activación  $\phi$ :

$$\phi_j(x) = \exp\left(-\frac{\|x - \mu_j\|^2}{2\sigma_j^2}\right)$$

Y como función de salida, se tiene la siguiente ecuación:

$$y(x) = \sum_{j=0}^M w_{kj} \phi_j(x)$$

donde  $w_{kj}$  es el vector de pesos que conecta la neurona  $j$  de la capa oculta con la neurona  $k$  de la capa de salida.

El entrenamiento de este tipo de modelos consiste en encontrar un conjunto de valores para los parámetros  $\mu_j$ ,  $\sigma_j$  y  $w_{jk}$  que maximizan el desempeño de la red.

### Desempeño de algoritmos de M.L.

Para realizar el entrenamiento, validación y pruebas de estos modelos se utilizó un proceso de *Cross Validation* en cada uno de los data sets. Los pronósticos se hicieron para un día posterior y para 7 días posteriores, midiendo el desempeño del modelo utilizando la medida *MAPE* (Mean Absolute Percentage Error) definida como:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{y_t - h_t}{y_t} \right|$$

Los resultados mostraron que la *regresión de Ridge* una transformación polinomial de grado dos tuvo el mejor desempeño con un *MAPE* del 8.2012 % durante la validación y un *MAPE* del 8.6561 % durante las pruebas. Durante el proceso de pruebas no se tuvo evidencia de sobreajuste del modelo.

Por otra parte el modelo la *regresión de Ridge con Kernel* tuvo un *MAPE* de 8.69622 % durante la validación, seguido por la *red neuronal* con un *MAPE* del 12.89174 % y finalmente la *red neuronal RBF* con un *MAPE* de 26.32086 %.

Los resultados obtenidos durante esta investigación fueron bastante prometedores sustentando el uso de modelos de M.L. como modelos de caja negra para estimar la ocupación futura de una propiedad, sin requerir conocimiento o experiencia estadística del estaff del hotel, permitiendo nuevos avances en las técnicas de *revenue management* en el sector de hospitalidad.

### Modelos de Big Data

En el sector de turismo y hospitalidad se han hecho esfuerzos para obtener modelos de pronóstico de llegadas y ocupación basados en información generada a partir de las búsquedas en páginas web y el tráfico generado en ellas. Si bien, los resultados obtenidos han confirmado la validéz de algunas fuentes de datos generados en línea, poco trabajo se ha hecho cruzando diferentes fuentes de datos, por ejemplo, utilizar los datos de las búsquedas realizadas en *Google* combinados con el tráfico generado en el sitio web del hotel, los cuales deben

## CAPÍTULO 2: MARCO TEÓRICO

estar altamente correlacionados (Tierney y Pan 2012).

Existen algunos trabajos en donde se adoptan dos métodos principales para obtener la ocupación de un hotel a partir de datos generados en línea. Los métodos utilizados son: *Promedios móviles integrados autoregresivos con variables externas* y *Modelo de regresión dinámica con cambios Markov (MS DR)*

En el estudio presentado por Pan y Yang (2017) se presentan dos modelos de series de tiempo en el que se incorporan tres fuentes de información generada en línea para predecir la ocupación de un hotel. Los datos de entrada contenían datos que describían:

- Tráfico generado en el sitio web de la propiedad
- Información relacionada con las peticiones realizadas a los motores de búsquedas (*Google*)
- Información del clima en el destino en dónde se ubica la propiedad

El primer modelo presentado es un modelo *ARMAX* en donde se utilizan variables generadas con *big data* como predictores directos de la ocupación. El modelo se definió de la siguiente manera:

$$y_t = \alpha + x_t\beta + \mu_t$$
$$\mu_t = \sum_{i=1}^m \rho_i \mu_{t-i} + \sum_{j=1}^n \theta_j \epsilon_{t-j} + \epsilon_t$$

donde  $y_t$  es la variable dependiente y  $x_t$  es un vector de variables exógenas independientes. Si  $\beta$  es igual a 0, el modelo se convierte en un modelo ARMA(m,n) estándar.

El segundo modelo presentado en este trabajo es el MSDR especificado como:

$$y_t = \tau_s + x_t\alpha + z_t\beta_s + \epsilon_s$$

dónde  $\tau_s$  es el estado de intercepción;  $x_t$  es un vector de variables exógenas con coeficientes independientes del estado;  $z_t$  es un vector de variables exógenas con coeficientes dependientes del estado  $\beta_s$ ;  $\epsilon_s$  es un error normal *i.i.d* con media 0 y una varianza dependiente del estado de  $\sigma^2$ . Este modelo permite a las variables en el vector  $z_t$  con parámetros dependientes del estado tomar diferentes

## CAPÍTULO 2: MARCO TEÓRICO

comportamientos basados en el estado que responden a un proceso de Markov con  $J$  estados y una probabilidad transicional  $p_{ij}$  especificada como sigue:

$$p_{ij} = \frac{1}{1 + \sum_{m=1}^k \exp(-q_{im})} \text{ si } j = k$$
$$p_{ij} = \frac{\exp(-q_{ij})}{1 + \sum_{m=1}^k \exp(-q_{im})} \text{ si } j \neq k$$

donde  $q_{ij}$  es el parámetro transformado definido como

$$q_{ij} = -\log\left(\frac{p_{ij}}{p_{ik}}\right)$$

Para empezar con la estimación del modelo se utiliza el algoritmo EM *expectation-maximization* el cuál arroja el valor inicial. Posteriormente se predicen las probabilidades de cada uno de los estados y actualizando la verosimilitud en cada paso. Un paso clave de esta estimación es el cálculo de la verosimilitud de los estados latentes la cual es obtenida mediante la iteración de la verosimilitud condicional.

Para medir el desempeño de ambos modelos se utilizó la medida *MAPE*

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{y_t - h_t}{y_t} \right|$$

Los resultados arrojan que el modelo ARIMAX tuvo un mejor desempeño con un *MAPE* del 3.804% comparado con el 4.714% obtenido por el modelo MS DR

Se puede concluir que los modelos basados en big data presentan una alternativa para poder generar pronósticos de ocupación y demanda utilizando información exógena a los hoteles, sin embargo se debe tomar en cuenta que el costo asociado a la configuración y puesta en marcha de este tipo de modelos es alto ya que se debe de contar con una infraestructura que pueda soportar el manejo de grandes volúmenes de información.

## Modelos de optimizacion de ingresos

### Resumen

# Bibliografía

- [1] El-Gayar N, Hendawi AT, Zakhary A, El-Shishiny H. *A proposed decision support model for hotel room revenue management*. ICGST Int J Artif Intell Mach Learn 2008.
- [2] Heba Abdel Aziz, Mohamed Saleh, Mohamed H. Rasmy, Hisham ElShishiny *Dynamic room pricing model for hotel revenue management systems*. Faculty of Computers and Information, Cairo University 2011.
- [3] Gabriel Bitran, Rene Caldentey *An Overview of Pricing Models for Revenue Management*. Sloan School of Management, MIT, Cambridge, Massachusetts 2011.
- [4] McGill, J. and van Ryzin, G *Revenue management: Research overview and Prospects*. Transportation Science 1999.
- [5] Cross, R. *Revenue Management: Hard-Core Tactics for Market Domination*. New York, NY: Broadway Books 1997.
- [6] Marriott, Jr., J. and Cross, R. *Room at the revenue inn. In Book of Management Wisdom: Classic Writings by Legendary Managers*. Peter Krass, New York, NY: Wiley 2000.
- [7] Weatherford, L. R., Kimes, S. E. *A comparison of forecasting methods for hotel revenue management*. Cornell University 2003.
- [8] M. J. Beckmann and F. Bobkowski *Airline Demand: An Analysis of Some Frequency Distributions*. Naval Res. Logistics 1958.

## BIBLIOGRAFÍA

- [9] Athanasius Zakharya, Amir F. Atiyab, Hisham El-Shishiny and Neamat E. Gayara *Forecasting hotel arrivals and occupancy using Monte Carlo simulation* Faculty of Computers and Information, Cairo University 2009.
- [10] Lee, A. *Airline Reservations Forecasting* Erewhon, NC: Prentice-Hall 1990
- [11] Flides, R, Ord, K. *Forecasting competitions: their role in improving forecasting practice and research* Oxford: Blackwell 2002
- [12] Ming Tse, Tony S., Tung Poon Yiu *Analyzing the use of an advance booking curve in forecasting hotel reservations* Taylor and Francis 2015
- [13] Wickham, R. R. *Evaluation of forecasting techniques for short term demand of air transportation* MIT Thesis: Fligt Transportation Lab 1995
- [14] Weatherford, L. R. *Forecasting issues in revenue management* Springs INFORMS Conference 1998
- [15] Caicedo-Torres, W., Payares, F. *A machine learning model for occupancy rates and demand forecasting in the hospitality industry* Universidad Tecnológica de Bolivar 2016
- [16] Tierney, Heather L.R, Bing Pan *A poisson regresion examination of the relationship between website traffic and search engine queries* Netnomics 2012
- [17] Pan,B., Yang, Y *Forecasting destination weekly hotel occupancy with big data* Journal of Travel Research, 2017