

Escuela de Negocios

Tipo de documento: Tesis de maestría



Master in Management + Analytics

Revenue Management para hoteles ubicados fuera de núcleos turísticos

Autoría: Mireнна, Mateo

Año: 2025

¿Cómo citar este trabajo?

Mireнна, M. (2025) "*Revenue Management para hoteles ubicados fuera de núcleos turísticos*". [Tesis de maestría. Universidad Torcuato Di Tella]. Repositorio Digital Universidad Torcuato Di Tella

<https://repositorio.utdt.edu/handle/20.500.13098/13745>

El presente documento se encuentra alojado en el **Repositorio Digital de la Universidad Torcuato Di Tella** bajo una licencia Creative Commons Atribución-No Comercial-Compartir Igual 4.0 Internacional

Dirección: <https://repositorio.utdt.edu>



UNIVERSIDAD
TORCUATO DI TELLA

MASTER IN MANAGEMENT + ANALYTICS

Revenue Management para hoteles ubicados fuera de núcleos turísticos

TESIS

Mateo Mirena

Tutor: Julián Regatky

Mayo 2025

Resumen

La gestión de precios en hoteles ubicados fuera de los principales centros turísticos presenta desafíos particulares, dada la variabilidad de la demanda y la ausencia de estrategias avanzadas de revenue management. En este trabajo se desarrolla un modelo predictivo orientado a optimizar los ingresos de un hotel en Pilar, Buenos Aires, Argentina, enfocado en el segmento de huéspedes individuales durante fines de semana y feriados.

Se utilizaron datos históricos de estadias, tarifas, eventos y feriados para construir modelos de predicción de demanda. Inicialmente, se aplicó una regresión lineal múltiple con método Bootstrap para robustecer las estimaciones; posteriormente, se reemplazó por un modelo de Random Forest, optimizado mediante Grid Search.

La evaluación de los modelos se realizó utilizando datos no vistos de 2023 y 2024. Los resultados muestran que, al aplicar el modelo de Random Forest optimizado, el hotel podría haber generado ingresos significativamente superiores en comparación con la política actual de precios. En particular, se estimó que un aumento más moderado de tarifas (alrededor del 12%) hubiera maximizado el revenue total, superando ampliamente al incremento fijo del 30% que se implementa actualmente. Esta estrategia habría permitido captar una mayor cantidad de huéspedes, generando un incremento del **34% en los ingresos por habitaciones en 2023** y del **21% en 2024**, en comparación con los ingresos obtenidos bajo la política vigente. Este impacto económico se amplifica si se consideran los ingresos asociados a servicios complementarios como Alimentos y Bebidas, Health & Spa y otros consumos vinculados a una mayor ocupación.

El análisis reveló que la política actual de aumentos de precios basada en umbrales de ocupación podría ser subóptima. Modelos más sofisticados permiten identificar estrategias de precios dinámicos que capturan mejor la elasticidad de la demanda, generando mayores ingresos tanto por alojamiento como por consumos asociados (Alimentos y Bebidas, Health & Spa, entre otros).

Los resultados obtenidos refuerzan el potencial del uso de técnicas analíticas avanzadas en la industria hotelera, incluso en establecimientos de menor escala, y plantean recomendaciones concretas de mejora en la política de precios y en la gestión estratégica de la información.

Para Virginia y Fernando, mis grandes maestros.

Para Delfina, mi sostén y luz.

Abstract

Price management in hotels located outside major tourist centers poses particular challenges due to demand variability and the lack of advanced revenue management strategies. This work develops a predictive model aimed at optimizing the revenue of a hotel in Pilar, Buenos Aires, Argentina, focusing on the individual guest segment during weekends and holidays. Historical data on stays, rates, events, and holidays were used to build demand prediction models. Initially, a multiple linear regression with the Bootstrap method was applied to enhance estimation robustness; later, it was replaced by a Random Forest model, optimized through Grid Search.

The models were evaluated using unseen data from 2023 and 2024. Results show that, when applying the optimized Random Forest model, the hotel could have generated significantly higher revenues compared to the current pricing strategy. Specifically, a more moderate price increase (around 12%) would have maximized total revenue, clearly outperforming the fixed 30% hike currently applied. This alternative pricing approach would have attracted more guests, resulting in a **34% increase in room revenue for 2023** and a **21% increase for 2024** compared to actual performance under the current policy. This economic impact is further amplified when accounting for additional spending on services such as Food & Beverage, Health & Spa, and other guest-related amenities.

The analysis revealed that the current price increase policy based on occupancy thresholds may be suboptimal. More sophisticated models enable the identification of dynamic pricing strategies that better capture demand elasticity, generating higher revenues from both room sales and additional services (such as Food and Beverage, Health & Spa, among others). The findings highlight the potential of applying advanced analytical techniques in the hotel industry, even for smaller-scale establishments, and propose concrete recommendations for improving pricing policies and strategic information management.

Contenidos

Resumen.....	i
Abstract.....	iii
Contenidos.....	iv
 Introducción	 1
1.1. Contexto y Problema.....	2
1.2. Objetivo y Enfoque	2
 El Hotel.....	 5
2.1. Descripción general del hotel y sus servicios	5
2.1.1. Acomodaciones.....	5
2.1.2. Actividades y recreación.....	6
2.1.3. Gastronomía y restaurantes.....	6
2.1.4. Eventos sociales y corporativos	7
2.1.5. Ubicación y entorno.....	7
2.2. Modelo de ingresos del hotel	7
 Revenue Management.....	 13
3.1. Fundamentos.....	13
3.2. Modelos de precios dinámicos en la industria hotelera.....	14
3.2. Aplicación de modelos predictivos y análisis de demanda	15
3.3. Modelos de regresión y Random Forest en la predicción de demanda.....	15
3.3.1. Modelos de regresión lineal	16
3.3.2. Bootstrap como técnica de modelado y robustez en predicciones.....	18
3.3.3. Random Forest.....	19
3.3.4. Optimización de hiperparámetros mediante Grid Search para Random Forest.....	21
 Datos y fuentes.....	 23
4.1. Estructura.....	23
4.2. Exploración.....	25
4.3. Calidad de los datos.....	25
4.3.1. Cantidad de noches (nights).....	26
4.3.2. Años (year).....	28
4.3.3. Estadías complementary: Tarifas (rate_amount)	29
4.3.4. Moneda (currency).....	30
4.3.5. Tipo de habitación (room_type).....	31
4.3.6. Tarifa en USD (rate_usd).....	33
4.4. Preparación de los datos y ETL	35

Metodología	37
5.1. Modelado de la demanda	37
5.2. Aplicación del modelo de regresión lineal con Bootstrap	39
5.3. Reemplazo por modelo de Random Forest	40
5.4. Validación del modelo	41
Resultados	42
6.1. Análisis de la demanda predicha.....	42
6.2. Evaluación del modelo	45
6.3. Comparación de los modelos	49
Conclusiones	52
7.1. Logros alcanzados	52
7.2. Recomendaciones para el hotel	56
7.2.1. Política de precios.....	56
7.2.2. Estrategia de gestión de datos	57
7.3. Limitaciones del modelo	58
7.3.1. Limitaciones estructurales de los modelos utilizados.....	58
7.3.2. Factores críticos no incorporados al modelo	59
Apéndice A. Lista de variables para cada dataset.....	61
Apéndice B. Estadías dataset Profiling Report.....	64
Apéndice C. Detalle de room_type.....	66
Apéndice D. Rate USD Profiling Report	67
Referencias.....	69

Capítulo 1

Introducción

La gestión de Revenue Management (RM) en la industria hotelera gira en torno a un objetivo tan claro como complejo: optimizar el uso de un inventario de habitaciones limitado y perecedero. Estas habitaciones, particularmente en contextos de alta demanda como fines de semana o feriados, representan uno de los principales generadores de ingresos para los hoteles. Una gestión inadecuada puede derivar en oportunidades de ingresos perdidas, sobreventa, o estrategias de precios poco eficientes que afectan directamente la rentabilidad.

En efecto, la estimación de la demanda y la asignación de capacidad son pilares fundamentales del RM hotelero. Modelos mal calibrados pueden conducir a habitaciones vacías en momentos clave o a precios que no capturan correctamente la disposición a pagar de los distintos segmentos de clientes. Esto se traduce en pérdidas económicas inmediatas y, a largo plazo, puede deteriorar el posicionamiento competitivo del hotel.

A diferencia de las habitaciones, los datos no son un recurso escaso. Los sistemas de gestión hotelera recopilan de manera constante información sobre reservas, precios, anticipación de las compras, comportamiento de los huéspedes y patrones estacionales. Sin embargo, en muchos casos, estos datos son subutilizados o empleados de forma reactiva en lugar de estratégica.

¿Puede esta abundante disponibilidad de datos utilizarse para mejorar los resultados de ingresos y optimizar la asignación de habitaciones frente a la incertidumbre de la demanda? Mi hipótesis es que, sin lugar a duda, se puede.

1.1. Contexto y Problema

El sector hotelero enfrenta crecientes desafíos para maximizar su rentabilidad en un entorno cada vez más competitivo y cambiante. En este marco, el RM se consolida como una herramienta clave para optimizar ingresos mediante la gestión dinámica de precios y capacidades, en función de la demanda y el comportamiento de los consumidores (Kimes, 1989; Smith et al., 2009). Si bien esta práctica ha sido ampliamente adoptada por cadenas hoteleras internacionales con gran volumen de habitaciones y alta demanda, su implementación en establecimientos de menor escala o ubicados fuera de los principales centros turísticos aún presenta oportunidades de desarrollo (Choi et al., 2004).

Este trabajo se enfoca en un hotel ubicado en la localidad de Pilar, provincia de Buenos Aires, Argentina. Se trata de un establecimiento orientado a dos segmentos claramente definidos: por un lado, el segmento corporativo, vinculado a eventos, convenciones y jornadas empresariales, que representa aproximadamente un 45 % de los ingresos totales; por otro, el segmento individual, compuesto por público local que utiliza el hotel principalmente durante fines de semana y feriados, ya sea para alojarse o realizar eventos sociales, representando el 55 % restante de los ingresos. En ambos casos, los porcentajes incluyen tanto los ingresos por habitaciones como aquellos generados por servicios adicionales (Health & Spa, Alimentos y Bebidas, Estacionamiento, entre otros).

Actualmente, el hotel aplica una estrategia básica de RM para el segmento individual, que consiste en incrementar las tarifas en un porcentaje determinado cuando la ocupación para una noche específica (viernes, sábado o víspera de feriado) supera el 50 %. Esta política, si bien sencilla de implementar, no contempla adecuadamente la variabilidad de la demanda, ni se apoya en herramientas analíticas que permitan una toma de decisiones informada y rentable (Kimes & Wirtz, 2003).

1.2. Objetivo y Enfoque

El objetivo principal de esta tesis es desarrollar un modelo de RM basado en el análisis de datos históricos y en el uso de herramientas analíticas avanzadas, con un enfoque centrado en el

segmento individual del hotel. El propósito es optimizar la estrategia de precios y maximizar los ingresos en noches de alta demanda, particularmente durante fines de semana, feriados y fechas especiales, cuando la elasticidad de la demanda permite capturar mayor valor por habitación sin comprometer la experiencia del huésped (Kimes, 1989; Elmaghraby & Keskinocak, 2003).

Dentro del modelo general de ingresos del hotel, el análisis se enfocará específicamente en las Habitaciones Transeúntes, es decir, aquellas comercializadas de forma individual y por estadías cortas, no asociadas a eventos corporativos previamente contratados. Este tipo de habitaciones representa una porción estratégica del negocio, ya que su disponibilidad y precios pueden ajustarse dinámicamente según la demanda prevista, lo que las convierte en un componente clave de las decisiones de RM (Talluri & Van Ryzin, 2004).

Actualmente, las ventas de habitaciones se distribuyen en partes iguales entre el público individual y el corporativo (50 % cada uno). Sin embargo, al analizar el ingreso generado por cada segmento, se observa que el 49 % de los ingresos por habitaciones proviene del segmento individual, mientras que el 51 % corresponde al corporativo. Particularmente, el segmento individual presenta una elevada rentabilidad marginal: cada habitación vendida genera, en promedio, ingresos incrementales de USD 220, considerando consumos de servicios como Alimentos y Bebidas, Health & Spa y actividades recreativas. La tarifa promedio de hospedaje ronda los USD 200, lo que sugiere un margen atractivo para una estrategia de precios optimizada.

Dado este contexto, la tesis propone revisar las limitaciones del esquema actual de precios — basado en incrementos porcentuales fijos una vez superado cierto umbral de ocupación— y reemplazarlo por un modelo más robusto que incorpore:

- Técnicas de análisis de demanda.
- Identificación de patrones estacionales y de comportamiento.
- Herramientas de predicción de ocupación futura.

El objetivo final es evaluar la validez de la política de precios actual y, en base a evidencia empírica, proponer una estrategia alternativa que permita maximizar el ingreso por habitación

disponible (RevPAR) y mejorar la eficiencia en la captura del valor de mercado, manteniendo la competitividad frente a la amplia oferta hotelera de la zona.

Es importante señalar que el hotel no contempla, al momento de esta investigación, modificar la metodología de aumento fijo por umbral de ocupación. Por lo tanto, este trabajo analizará si la política vigente es adecuada o si existen oportunidades de mejora dentro de la misma lógica operativa.

Capítulo 2

El Hotel

Para poder comprender el tipo de análisis realizado y el modelo implementado, es fundamental comprender las características del hotel. A continuación, comentaré los principales aspectos a considerar:

2.1. Descripción general del hotel y sus servicios

El hotel es un establecimiento de primer nivel ubicado en la localidad de Pilar, Provincia de Buenos Aires. Gracias a su cercanía a centros empresariales, polos industriales, complejos recreativos, salones de eventos y un rápido acceso a la Ciudad de Buenos Aires, el hotel se presenta como una opción estratégica tanto para turistas locales como para público corporativo.

El hotel cuenta con una amplia gama de servicios y comodidades orientadas al confort y la satisfacción de sus huéspedes. Entre los principales servicios se destacan la pileta al aire libre, spa, gimnasio, jardines, salones para eventos y espacios de convenciones completamente equipados.

2.1.1. Acomodaciones

Las habitaciones están diseñadas para ofrecer una experiencia de descanso y relax, combinando elegancia con funcionalidad. Disponen de aire acondicionado, televisión por cable, caja de

seguridad, escritorio, conexión Wi-Fi y servicio a la habitación. Existen diversas categorías de habitaciones y suites que se adaptan a las necesidades del huésped, ya sea que viaje por ocio o negocios. Para ser más específico, hay cuatro tipos de habitaciones:

- Habitación clásica con vistas al frente
- Habitación de lujo con vistas al jardín
- Habitaciones ejecutivas
- Suites

Generalmente, las tarifas de las dos primeras categorías son similares, mientras que las habitaciones ejecutivas y suites presentan precios considerablemente más altos, lo cual se detallará más adelante.

2.1.2. Actividades y recreación

El hotel brinda un entorno ideal para el descanso y el bienestar, combinando naturaleza y opciones recreativas. Sus amplios jardines invitan a relajarse al aire libre, caminar, descansar bajo los árboles o disfrutar de actividades como fútbol, tenis y juegos infantiles. Las piscinas exteriores y áreas verdes completan la oferta al aire libre.

En el interior, se ofrecen juegos para todas las edades y un completo Health & Spa que incluye piscina climatizada, sauna, baño de vapor, sala de relax y tratamientos para el bienestar integral.

2.1.3. Gastronomía y restaurantes

El hotel ofrece una propuesta gastronómica variada, pensada para distintos momentos del día. Desde cocina de autor con inspiración mediterránea hasta un restaurante buffet internacional, las opciones se adaptan tanto a cenas elegantes como a comidas informales. Durante el verano, se suma un restaurante tipo parrilla al aire libre con especialidades argentinas.

El servicio de cafetería y coctelería está disponible todo el día, con alternativas ligeras para almuerzos o cenas. Todas las propuestas están abiertas al público general, posicionando al hotel como un referente gastronómico en la zona.

2.1.4. Eventos sociales y corporativos

Los eventos representan una dimensión central en la actividad del hotel, tanto por su frecuencia como por su impacto operativo y en la generación de ingresos. A lo largo del año, se realizan diversas celebraciones sociales (bodas, fiestas de 15, aniversarios, bar y bat mitzvá) y eventos corporativos (reuniones de negocios, convenciones, capacitaciones, lanzamientos de productos y actividades de team building).

Para su desarrollo, el hotel cuenta con salones versátiles que se adaptan a diferentes formatos y cantidades de invitados, incluyendo eventos de gran escala con más de 700 cubiertos. Estas instalaciones, junto con un servicio gastronómico de alta calidad y un equipo profesional experimentado, consolidan al hotel como una opción de referencia en el segmento de eventos.

Además, en fechas especiales como Navidad y Año Nuevo, el hotel organiza propuestas temáticas con cenas, espectáculos y actividades familiares. También se ofrecen veladas exclusivas en formato cena-show con artistas reconocidos, un producto distintivo en la zona.

2.1.5. Ubicación y entorno

Pilar es una localidad del norte del Gran Buenos Aires que combina zonas residenciales, centros comerciales y amplios espacios verdes. En las inmediaciones del hotel se encuentran shoppings, centros médicos, clubes de campo, parques industriales y accesos a rutas nacionales, lo que refuerza su posición como una opción integral tanto para el turismo como para las actividades empresariales.

2.2. Modelo de ingresos del hotel

El modelo de ingresos del hotel se basa en una estructura diversificada que combina alojamiento con actividades complementarias de alto valor agregado. La generación de ingresos se sustenta principalmente en dos pilares: los eventos (sociales y corporativos) y el alojamiento (corporativo e individual de corta estadía).

En términos generales, los ingresos pueden segmentarse en las siguientes categorías:

- **Eventos Corporativos:** Abarca reuniones empresariales, convenciones, lanzamientos de productos y otras actividades de negocios. Estos eventos no solo generan ingresos por el alquiler de espacios, sino también por servicios complementarios como gastronomía, audiovisuales y alojamiento para los asistentes.
- **Alimentos y Bebidas:** Comprende todos los consumos gastronómicos no corporativos ni de eventos sociales dentro del hotel, tales como restaurantes, lobby bar y room service.
- **Habitaciones Transeúntes:** Incluye las reservas de alojamiento realizadas de manera individual, ya sea por turismo, escapadas de fin de semana o fechas especiales. Este segmento se caracteriza por una tarifa promedio que ronda los USD 200 por habitación. En promedio, los huéspedes alojados suelen generar consumos adicionales de USD 220 vinculados a alimentos y bebidas, Health & Spa, y otros.
- **Health & Spa:** Comprende el uso de las instalaciones de spa, como piscina climatizada, hidromasaje, sauna, tratamientos faciales y corporales. Estos servicios representan un ingreso complementario importante, especialmente en fines de semana largos y temporadas de alta ocupación.
- **Eventos Sociales:** Incluye bodas, fiestas de 15, celebraciones familiares y festividades especiales. Al igual que los eventos corporativos, combinan ingresos de múltiples fuentes: alquiler de salones, gastronomía y servicios de ambientación.
- **Otros:** Engloba ingresos por actividades recreativas, boutique, estacionamiento, Wi-Fi, entre otros servicios.

La Figura 2.2.1 ilustra la distribución de ingresos según estas categorías:

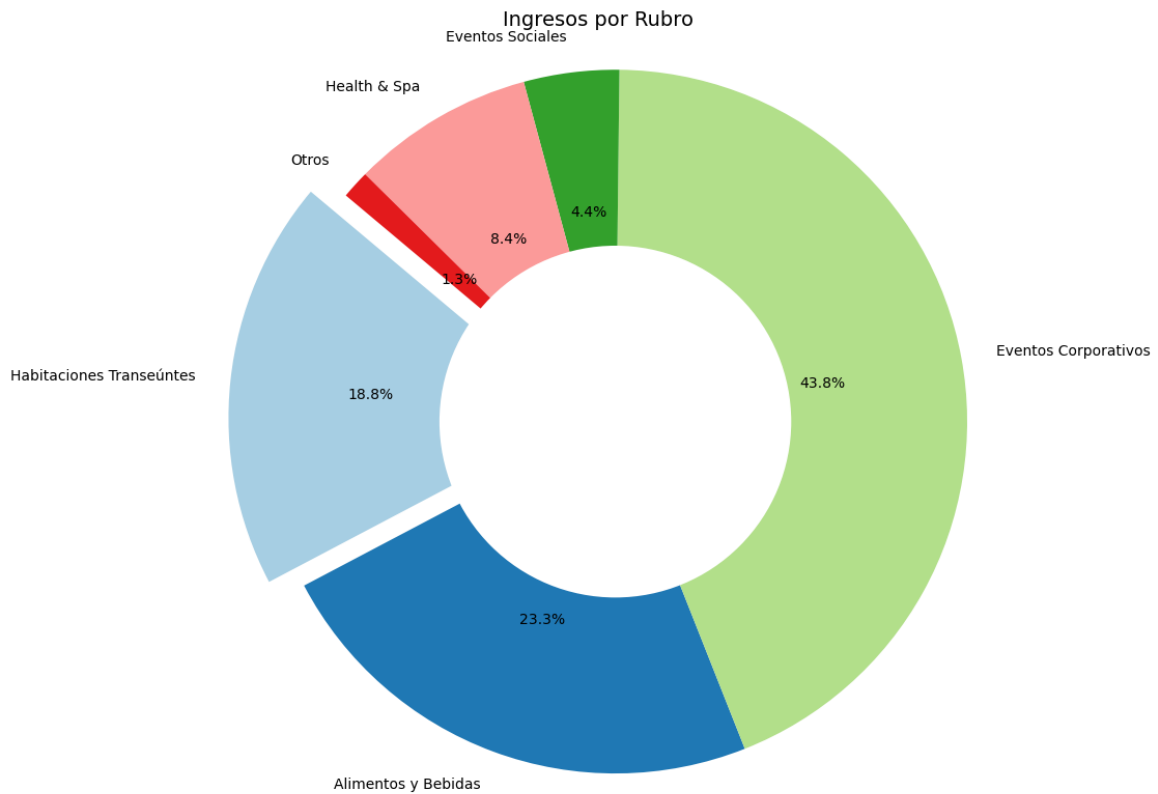


Figura 2.2.1: Mix de ingresos por rubro

Esta segmentación pone en evidencia el peso que tienen los eventos corporativos en el mix de ingresos del establecimiento, así como la relevancia del alojamiento individual como fuente recurrente de ventas.

En el análisis de comportamiento del público individual, resulta de especial interés comprender cómo se distribuyen las reservas a lo largo de la semana, ya que esto permite identificar patrones de demanda y posibles oportunidades de optimización tarifaria. Para ello, se realizó un estudio de habitaciones ocupadas por día de la semana.

A continuación, se presenta la figura 2.2.2 que ilustra la distribución porcentual de habitaciones ocupadas por día de la semana:

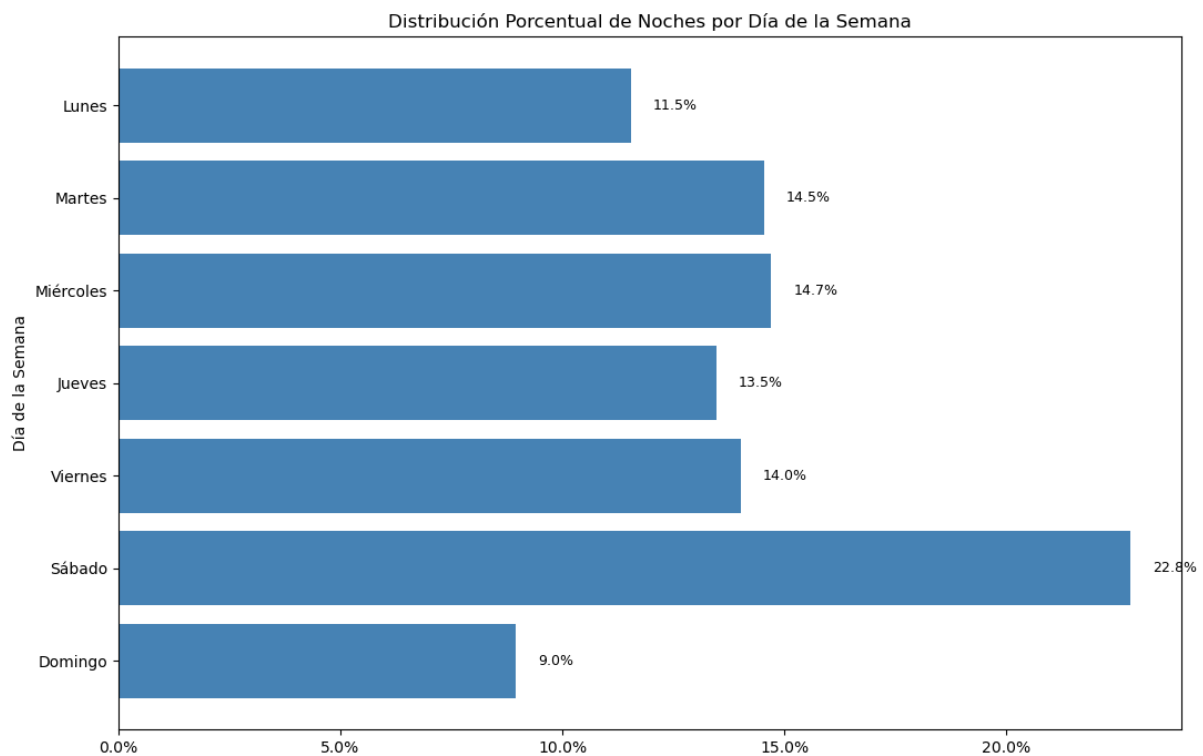


Figura 2.2.2: Mix de habitaciones por día de la semana

En la figura se observa que el sábado es el día de mayor contratación de habitaciones, reflejando la preferencia por escapadas de fin de semana. Los restantes días presentan una ocupación más homogénea, aunque claramente inferior, y en general responden a la demanda del público corporativo.

Esta información es clave para diseñar estrategias de precios dinámicos que consideren el ciclo semanal de la demanda, permitiendo maximizar el RevPAR en días de alta demanda y estimular las reservas en los días de menor actividad.

Otra dimensión relevante es la estacionalidad mensual. El análisis de habitaciones ocupadas por mes permite identificar los periodos de mayor y menor actividad. En el caso del hotel, se destacan los meses de marzo y diciembre.

A continuación, se presenta la figura 2.2.3, donde gráficamente se puede ver la distribución de habitaciones ocupadas a lo largo de los distintos meses del año:

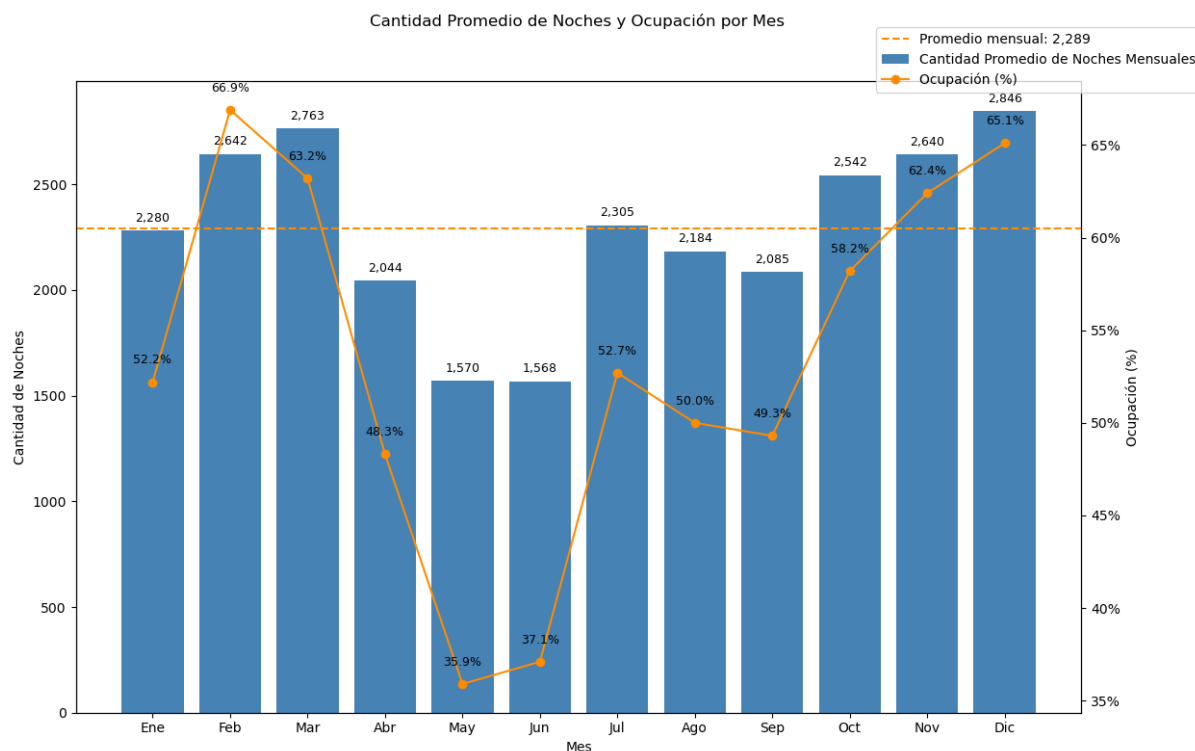


Figura 2.2.3: Cantidad de habitaciones mensuales promedio y ocupación porcentual

El mes de marzo se destaca particularmente por el alto volumen de eventos corporativos que suelen marcar el inicio del calendario anual para muchas organizaciones. Esta demanda estructural convierte a marzo en un mes clave para la facturación del segmento corporativo. Por otro lado, diciembre presenta también una ocupación significativamente elevada, impulsada por la realización de eventos sociales, como fiestas de fin de año, bodas, cenas de empresas y celebraciones familiares, así como por las propuestas especiales que el hotel ofrece en torno a las festividades navideñas y de Año Nuevo. Esta estacionalidad mencionada, se repite año a año, teniendo en cuenta los aspectos del calendario comercial que suele repetirse año a año. Esta estacionalidad reafirma la importancia de una planificación tarifaria que considere las particularidades de cada mes y permita capitalizar las oportunidades de mayor demanda.

Cabe recordar que las ventas de habitaciones se distribuyen equitativamente entre los segmentos individual y corporativo, representando cada uno el 50% tanto en cantidad de noches vendidas como en ingresos generados.

Por último, en el marco de la política comercial actual, el hotel dispone de 128 habitaciones estándar para la venta al público individual, excluyendo Suites y Ejecutivas (las cuales suman 13). La estrategia tarifaria implementada se basa en el siguiente esquema: las primeras 64

habitaciones se ofrecen a una tarifa base fija, mientras que a partir de la 65^a habitación vendida, se aplica un aumento porcentual también fijo al precio. Esta política, de tipo umbral, busca capitalizar la alta demanda en fechas puntuales —como fines de semana largos o eventos especiales—, ajustando el precio únicamente cuando se supera dicho nivel de ocupación. Si bien esta estrategia permite capturar valor en picos de demanda, su estructura rígida y basada en un incremento porcentual arbitrario, no contempla las particularidades de cada fecha ni la elasticidad real del mercado, como se explorará más adelante en la sección metodológica.

Capítulo 3

Revenue Management

En la siguiente sección, se comentará teóricamente los fundamentos, metodologías y técnicas relacionadas al RM, en especial, ligadas a la industria hotelera. Algunas técnicas, serán desarrolladas en las siguientes secciones.

3.1. Fundamentos

El Revenue Management (RM) es una estrategia empresarial orientada a maximizar los ingresos mediante la gestión dinámica de precios y la optimización de la capacidad, basada en el análisis de la demanda y el comportamiento de los clientes. Originalmente, el RM surgió en la industria aérea, pero rápidamente se expandió a otros sectores, como la hotelería, los alquileres de automóviles y las compañías de cruceros (Kimes, 1989), entre otros. El objetivo del RM es vender el producto adecuado al precio adecuado, en el momento adecuado, y al cliente adecuado.

Los pilares del RM incluyen:

- **Segmentación de clientes:** Identificar diferentes grupos de clientes con diferentes disposiciones a pagar, lo que permite aplicar precios diferenciados según las características del cliente (Smith et al., 2009).
- **Gestión de inventario:** Limitar la disponibilidad de productos en función de la demanda anticipada y el precio. En el contexto hotelero, esto puede implicar asignar un

número limitado de habitaciones a diferentes segmentos de clientes. Estos segmentos, suelen estar asociados a plataformas de turismo online (OTAs), venta directa y walk ins, donde en cada uno tiene su propia demanda y precios asociados.

- **Optimización de precios:** Ajustar los precios de forma dinámica según la demanda y las condiciones del mercado. Este aspecto se centra en maximizar los ingresos de cada unidad disponible (por ejemplo, cada habitación de hotel). Generalmente, la estrategia consiste en incrementar los precios a medida que aumenta la ocupación del hotel.
- **Predicción de demanda:** Estimar la cantidad de demanda para optimizar las decisiones de precios y la asignación de inventario. Esto puede implicar técnicas de modelado estadístico y machine learning.

Sobre estos últimos dos pilares, se van a ir ahondando en las siguientes secciones.

3.2. Modelos de precios dinámicos en la industria hotelera

Los modelos de precios dinámicos son una de las herramientas más poderosas en el Revenue Management, especialmente en la industria hotelera, donde la demanda es volátil y depende de factores como la estacionalidad, eventos locales, precios de la competencia, y condiciones económicas, entre otras. Estos modelos ajustan los precios de manera continua, basándose en la oferta y la demanda en tiempo real.

Existen diversos enfoques para el desarrollo de modelos de precios dinámicos, entre los que se incluyen:

- **Modelos de optimización estocástica:** Utilizan la teoría de optimización para determinar el precio que maximiza los ingresos, considerando la incertidumbre en la demanda y la capacidad. Por ejemplo, se pueden aplicar técnicas de simulación para modelar diferentes escenarios de demanda y precios (Bitran & Caldentey, 2003).
- **Modelos de subasta y precios de equilibrio:** En algunos contextos, los precios se determinan mediante subastas o plataformas de reserva online, donde la demanda se encuentra con el precio en tiempo real (Elmaghraby & Keskinocak, 2003).

- **Algoritmos de predicción de demanda:** Los modelos de predicción de demanda, como los modelos ARIMA o de machine learning, pueden predecir la cantidad de habitaciones demandadas a lo largo del tiempo y ajustar los precios dinámicamente en consecuencia.

3.2. Aplicación de modelos predictivos y análisis de demanda

La predicción de la demanda es fundamental para las estrategias de RM, ya que permite anticipar las fluctuaciones en la demanda y ajustar los precios y la disponibilidad en consecuencia. Los modelos predictivos pueden basarse en una amplia variedad de enfoques, desde modelos estadísticos clásicos hasta técnicas avanzadas de machine learning.

- **Modelos estadísticos:** Los modelos tradicionales como regresión lineal y ARIMA (AutoRegressive Integrated Moving Average) son utilizados para predecir la demanda futura en función de variables históricas como las tarifas pasadas, la ocupación y los factores estacionales (Choi et al., 2004).
- **Machine Learning:** El uso de algoritmos de aprendizaje automático, como Random Forest, permite modelar la demanda de manera no lineal, capturando patrones complejos que los modelos tradicionales no pueden identificar (Bertsimas et al., 2013).

Estos modelos no sólo pueden predecir la demanda de habitaciones, sino también el comportamiento del cliente, lo que puede incluir la probabilidad de que un cliente realice una reserva en función de diversas características (como tipo de habitación, precio y temporada).

3.3. Modelos de regresión y Random Forest en la predicción de demanda

A continuación, se presentarán teóricamente los modelos elegidos para este trabajo.

3.3.1. Modelos de regresión lineal

Los modelos de regresión han sido utilizados en la industria hotelera para predecir la demanda de habitaciones en función de diversas variables explicativas. Estos modelos pueden ser tanto lineales como no lineales, dependiendo de la naturaleza de los datos y de las relaciones entre las variables. Diversos estudios han demostrado su eficacia para la predicción de demanda hotelera y otros aspectos relacionados con el RM, incluyendo la comparación entre métodos tradicionales como la regresión lineal y modelos más complejos como Random Forest (Moreno-Izquierdo et al., 2019; García et al., 2018; Köseoglu et al., 2021).

Uno de los enfoques más tradicionales es la regresión lineal múltiple, donde se modela la relación entre una variable dependiente continua —como el número de habitaciones reservadas— y un conjunto de variables independientes, tales como el precio promedio, el día de la semana, la estacionalidad o eventos especiales, entre otras. El modelo tiene la siguiente estructura general:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \varepsilon \quad (1)$$

Donde:

- y representa la variable dependiente (por ejemplo, cantidad de habitaciones demandadas),
- β_0 es el intercepto,
- $\beta_1, \beta_2, \dots, \beta_p$ son los coeficientes de regresión que miden el impacto de cada variable explicativa,
- ε es el término de error aleatorio.

La estimación de los parámetros β se realiza comúnmente mediante el método de mínimos cuadrados ordinarios (MCO), buscando minimizar la suma de los errores cuadráticos entre los valores observados y los predichos:

$$\min_{\beta} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2)$$

Una ventaja fundamental de la regresión lineal es su simplicidad interpretativa, ya que permite entender de forma directa el efecto de cada variable sobre la demanda. Sin embargo, presenta limitaciones cuando las relaciones entre las variables no son estrictamente lineales, una situación frecuente en el ámbito hotelero donde influyen múltiples factores de manera compleja (Kimes & Wirtz, 2003).

Para evaluar la calidad predictiva del modelo, se utilizará el Root Mean Squared Error (RMSE) como métrica principal. El RMSE mide la magnitud promedio de los errores de predicción, penalizando más severamente los errores grandes, y se define como:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (3)$$

Donde:

- y_i son los valores observados,
- $\hat{y}_i$ son los valores predichos,
- n es el número total de observaciones.

El RMSE tiene la ventaja de mantener las mismas unidades que la variable dependiente, facilitando su interpretación. Un RMSE más bajo implica una mejor capacidad predictiva del modelo. Esta métrica es utilizada en problemas de forecasting de demanda (Hyndman & Athanasopoulos, 2018), y permite evaluar de forma intuitiva la precisión del modelo.

Finalmente, diversos estudios han encontrado que, si bien la regresión lineal continúa siendo una herramienta válida para establecer una línea base de predicción, técnicas de machine learning como Random Forest suelen mejorar el desempeño predictivo, especialmente en presencia de no linealidades y relaciones de interacción entre variables (García et al., 2018; Moreno-Izquierdo et al., 2019; Cheng & Edwards, 2019; Köseoglu et al., 2021).

3.3.2. Bootstrap como técnica de modelado y robustez en predicciones

El método bootstrap, introducido por Efron (1979), es una técnica de remuestreo que permite aproximar la distribución de un estimador mediante la generación de múltiples muestras aleatorias (con reemplazo) a partir de los datos originales. Este enfoque resulta particularmente útil cuando se desea evaluar la variabilidad de un modelo, mejorar su robustez o construir intervalos de confianza sin requerir fuertes supuestos paramétricos.

Desde el punto de vista metodológico, el procedimiento consiste en:

1. Tomar B muestras aleatorias con reemplazo del conjunto de datos original.
2. Ajustar el modelo en cada una de las B muestras.
3. Evaluar el desempeño del modelo y/o combinar las predicciones resultantes.

Matemáticamente, si $D = \{(x_i, y_i)\}_{i=1}^n$ representa el conjunto de datos original, se generan B conjuntos bootstrap D_b^* (con $b = 1, 2, \dots, B$), sobre los cuales se ajustan modelos independientes. La predicción final para una nueva observación puede calcularse como el promedio de las predicciones individuales:

$$\hat{y} = \frac{1}{B} \sum_{b=1}^B \hat{y}_b(x) \quad (4)$$

Una ventaja adicional del enfoque bootstrap es que, al obtener múltiples estimaciones para una misma predicción, también es posible construir intervalos de confianza empíricos. Por ejemplo, si se desea un intervalo de confianza del 90% para la predicción de una observación, se pueden ordenar las B predicciones correspondientes y tomar los percentiles 5 y 95 como límites inferior y superior. Esto permite cuantificar la incertidumbre asociada a cada estimación puntual, lo cual es especialmente valioso en contextos donde se toman decisiones operativas o estratégicas basadas en modelos predictivos.

En el presente trabajo, el método bootstrap fue utilizado para adaptar el modelo de regresión lineal múltiple, entrenando múltiples regresiones sobre muestras bootstrapeadas del conjunto de datos de entrenamiento. Esta estrategia permitió:

- Estimar la variabilidad de las predicciones.
- Robustecer el modelo frente a potenciales anomalías o sobreajuste al conjunto de datos original.
- Obtener una medida más estable de desempeño promedio (evaluada mediante el RMSE). Cada modelo de regresión fue ajustado sobre una muestra diferente, y las predicciones se agregaron para obtener una estimación final más confiable.

3.3.3. Random Forest

El Random Forest es un modelo de aprendizaje automático basado en conjuntos de árboles de decisión, desarrollado originalmente por Breiman (2001). Su popularidad en aplicaciones predictivas radica en su capacidad para manejar relaciones no lineales complejas entre variables, su resistencia al sobreajuste y su versatilidad en distintos tipos de datos (estructurados y no estructurados). A diferencia de modelos lineales que imponen una estructura funcional fija, Random Forest crea múltiples árboles de decisión de forma aleatoria y combina sus resultados para mejorar la precisión y reducir la varianza del modelo.

Técnicamente, el algoritmo construye n árboles de decisión $T_1, T_2, \dots, T_n$, a partir de subconjuntos bootstrap del conjunto de entrenamiento como lo mencionado en el punto anterior. Esta aleatorización no solo permite que los árboles individuales sean diversos entre sí, sino que también reduce el riesgo de sobreajuste característico de los árboles individuales, aumentando la generalización del modelo (Breiman, 2001).

Para cada división de un nodo, Random Forest selecciona aleatoriamente un subconjunto de m variables candidatas entre todas las variables disponibles, y escoge la mejor división basada en esas m variables. La predicción final para un valor continuo (regresión) se obtiene promediando las predicciones individuales de todos los árboles:

$$\hat{y} = \frac{1}{n} \sum_{i=1}^n T_i(x) \quad (3)$$

Entre sus principales ventajas metodológicas se destacan:

-
- Capacidad para modelar relaciones altamente no lineales y complejas entre variables (Breiman, 2001; Liaw & Wiener, 2002).
 - Robustez frente a outliers y ruido en los datos.
 - Manejo eficiente de variables categóricas, datos faltantes y gran cantidad de predictores sin necesidad de extensivo preprocesamiento.
 - Estimación de la importancia de variables, facilitando la interpretación de cuáles factores influyen más en la variable objetivo.

Desde el punto de vista de la evaluación del modelo, en problemas de regresión como el forecasting de demanda hotelera, el desempeño se mide habitualmente con métricas como el Root Mean Squared Error (RMSE), permitiendo comparar directamente los resultados con los obtenidos por modelos tradicionales como la regresión lineal.

En cuanto a su aplicación en la industria hotelera, el modelo Random Forest ha demostrado ser especialmente útil para predecir la demanda de habitaciones en función de factores como precios dinámicos, variaciones estacionales, eventos especiales y condiciones climáticas locales. Por ejemplo, Moreno-Izquierdo et al. (2019) y García et al. (2018) comparan la performance de Random Forest frente a regresiones tradicionales en predicción de ocupación hotelera, encontrando mejoras significativas en la capacidad predictiva. Asimismo, Zhang et al. (2017) utilizaron Random Forest para modelar la demanda de hoteles considerando tanto datos internos (historial de reservas, tarifas) como externos (eventos, feriados, condiciones climáticas), mostrando que la incorporación de múltiples fuentes de información es fundamental para optimizar las predicciones en este sector.

Así, Random Forest se posiciona como una herramienta poderosa en el contexto de RM hotelero, donde las decisiones de pricing, inventario y estrategia comercial dependen críticamente de una correcta estimación de la demanda futura.

3.3.4. Optimización de hiperparámetros mediante Grid Search para Random Forest

La optimización de hiperparámetros es un paso fundamental en la construcción de modelos basados en árboles de decisión, como los algoritmos de Random Forest, ya que su desempeño puede variar significativamente en función de parámetros como la profundidad máxima del árbol, el número mínimo de muestras por nodo, o la cantidad total de árboles en el ensamble. Una de las técnicas más utilizadas para este fin es el Grid Search, que consiste en definir un conjunto discreto de valores posibles para cada hiperparámetro y evaluar el desempeño del modelo para cada combinación posible de esos valores.

Metodológicamente, el proceso de Grid Search sigue los siguientes pasos:

1. Se establece una grilla (grid) de combinaciones posibles de valores de hiperparámetros.
2. Para cada combinación, se entrena un modelo nuevo sobre el conjunto de entrenamiento.
3. Se evalúa el desempeño de cada modelo mediante una métrica de validación apropiada (por ejemplo, el RMSE para problemas de regresión).
4. Se selecciona la combinación de hiperparámetros que optimiza la métrica de interés.

En el caso particular de modelos de árboles de decisión y Random Forest, algunos de los hiperparámetros más relevantes para ajustar mediante Grid Search incluyen:

- **max_depth**: profundidad máxima permitida para cada árbol, controlando el grado de complejidad.
- **min_samples_split**: número mínimo de muestras requeridas para dividir un nodo interno.
- **min_samples_leaf**: número mínimo de muestras que debe tener un nodo hoja.
- **n_estimators** (en Random Forest): cantidad de árboles en el ensamble.
- **max_features**: número máximo de características consideradas al buscar la mejor división.

El Grid Search puede complementarse con técnicas de validación cruzada para obtener una estimación más robusta del desempeño del modelo en datos no vistos, evitando así problemas de overfitting durante el ajuste de los hiperparámetros.

Diversos estudios han demostrado que la optimización de hiperparámetros mediante Grid Search puede mejorar significativamente la capacidad predictiva de modelos basados en árboles de decisión (García et al., 2018; Moreno-Izquierdo et al., 2019), incluyendo contextos como la predicción de demanda hotelera donde las relaciones entre variables suelen ser no lineales y complejas.

Capítulo 4

Datos y fuentes

Los datos utilizados para el desarrollo del proyecto fueron proporcionados por uno de los hoteles más relevantes de Argentina, bajo un acuerdo de confidencialidad. Se concedió acceso completo a la base de datos del hotel. No obstante, no toda la información disponible resultaba relevante para el problema abordado. A continuación, voy a detallar la información relevante y utilizada para realizar el modelo.

4.1. Estructura

Los datos que analizamos y aprovechamos en el marco de este proyecto incluyen:

- **Estadías de clientes**

Este conjunto de datos contiene información correspondiente a todas las estadías registradas en el hotel. Cada fila representa una persona alojada, identificada por un identificador único de estadía y un identificador de cliente (que se mantiene constante a lo largo de las diferentes visitas realizadas por una misma persona). Adicionalmente, se cuenta con el monto de la tarifa por noche de cada persona alojada, junto con la moneda correspondiente (pesos o dólares como alternativas).

Para cada estadía se incluye información demográfica del huésped, detalles sobre la estadía, condiciones del alojamiento, entre otros datos relevantes. El período considerado abarca desde el año 2019 hasta el 2024 inclusive.

- **Tipo de cambio**

Este conjunto de datos fue incorporado con el fin de unificar el criterio de monedas, ya que las tarifas por noche incluidas en el conjunto de datos principal pueden estar expresadas tanto en pesos argentinos como en dólares estadounidenses. La información de tipo de cambio permitió normalizar los valores y facilitar un análisis homogéneo de los ingresos y precios. El tipo de cambio utilizado fue el dólar CCL.

- **Feriados**

Se incluyó un listado completo de los feriados nacionales y de las noches previas a feriados correspondientes al período analizado. Este dataset clasifica cada fecha según el tipo de fin de semana en que se encuadra (por ejemplo, feriados simples, fines de semana largos y otros). También, se especificó que feriado estaba en cuestión, por ejemplo, si el mismo correspondía a semana santa, navidad o cualquier otro que se dé a lo largo de los años. Esta información resultó clave para identificar patrones de demanda asociados a fechas especiales y para segmentar el análisis en función del comportamiento de los huéspedes en contextos de mayor o menor actividad.

- **Eventos**

Este conjunto de datos contiene el detalle de todos los eventos sociales realizados en el hotel, registrados día por día durante el período analizado. Para cada evento se incluye información sobre el tipo de celebración (como casamientos, cumpleaños, entre otros), el espacio del hotel en el que se llevó a cabo (salones, jardines, restaurantes, etc.) y la cantidad estimada de asistentes.

Este dataset permitió incorporar al análisis una dimensión adicional vinculada a la ocupación de espacios comunes y al impacto que este tipo de actividades puede tener sobre la demanda de alojamiento, servicios gastronómicos y otras áreas del hotel.

La lista entera de variables está incluida en el **Apéndice A**, al igual que un par de ejemplos correspondiente a Estadías de clientes.

4.2. Exploración

La exploración de datos es una parte clave tanto de la fase de comprensión de los datos como de preparación de los datos y sus principales objetivos son dos:

- Detectar problemas de calidad en los datos que puedan afectar negativamente el desempeño del modelo, como valores faltantes, outliers o variables categóricas con cardinalidad irregular. En el contexto de RM, estos problemas pueden distorsionar las estimaciones de demanda y llevar a decisiones subóptimas en la política de precios.
- Conocer los datos en profundidad, es decir, comprender las características de las variables relevantes para la predicción de la demanda. Qué tipo de valores pueden tomar, en qué rangos se distribuyen y cómo se comportan esas distribuciones. Esta comprensión es esencial para ajustar correctamente los parámetros de los modelos de estimación de demanda definidos, y así tomar decisiones óptimas de capacidad o precios que maximicen el ingreso esperado considerando la incertidumbre.

Dentro de la exploración de los datos, se realizó un ProfileReport sobre el conjunto de estadías para entender y detectar cualquier anomalía o insight necesario para continuar con el modelo y maximizar la significancia estadística.

El ProfileReport de Estadías de clientes está incluido en el **Apéndice B**.

4.3. Calidad de los datos

La calidad de los datos es un aspecto crítico en el desarrollo de modelos predictivos y en la toma de decisiones informadas. La existencia de errores, outliers y datos inconsistentes puede distorsionar las predicciones y afectar la precisión de los modelos. En este caso, se identificaron varios problemas con el conjunto de datos de estadías de clientes. A continuación, detallo cada uno de los puntos de dolor relacionados con los datos, las acciones tomadas para abordarlos y otras consideraciones importantes:

4.3.1. Cantidad de noches (**nights**)

La variable **nights** refleja la cantidad de noches que un huésped permaneció en el hotel. En esta, se detectaron estadías con duraciones anómalas. Por ejemplo, algunos huéspedes habían permanecido en el hotel por más de 2 años consecutivos. Teniendo en cuenta esto, se conversó con el hotel para entender si efectivamente las personas habían permanecido ese tiempo alojados, y efectivamente eso no fue así. Estos errores se debieron a fallas en los procedimientos de auditoría diaria, ya que, en lugar de reflejar las noches reales de estadía, la variable registra el número de días que tardó en completarse estos procesos, más allá de las noches efectivamente tomadas.

También, en varios casos, se registraron valores de **nights** iguales a 0, lo que es completamente incoherente con la lógica del problema, ya que no es posible que un huésped haya tenido una estadía de 0 noches.

Para mitigar estas situaciones fuera de lo comun, se identificaron y eliminaron los registros con noches iguales a 0. Tambien, al validar con el hotel la duración máxima habitual de una estadía, se estableció un límite máximo de 7 noches consecutivas. Este criterio permitió delimitar un umbral razonable para depurar outliers extremos.

A continuación, se presentan dos gráficos que ilustran el efecto de la limpieza de datos.

La Figura 4.3.1.1 muestra la distribución original de la variable **nights**, donde se evidencian valores atípicos que alcanzan estadías de más de 700 noches, claramente inconsistentes con la operación real del hotel.

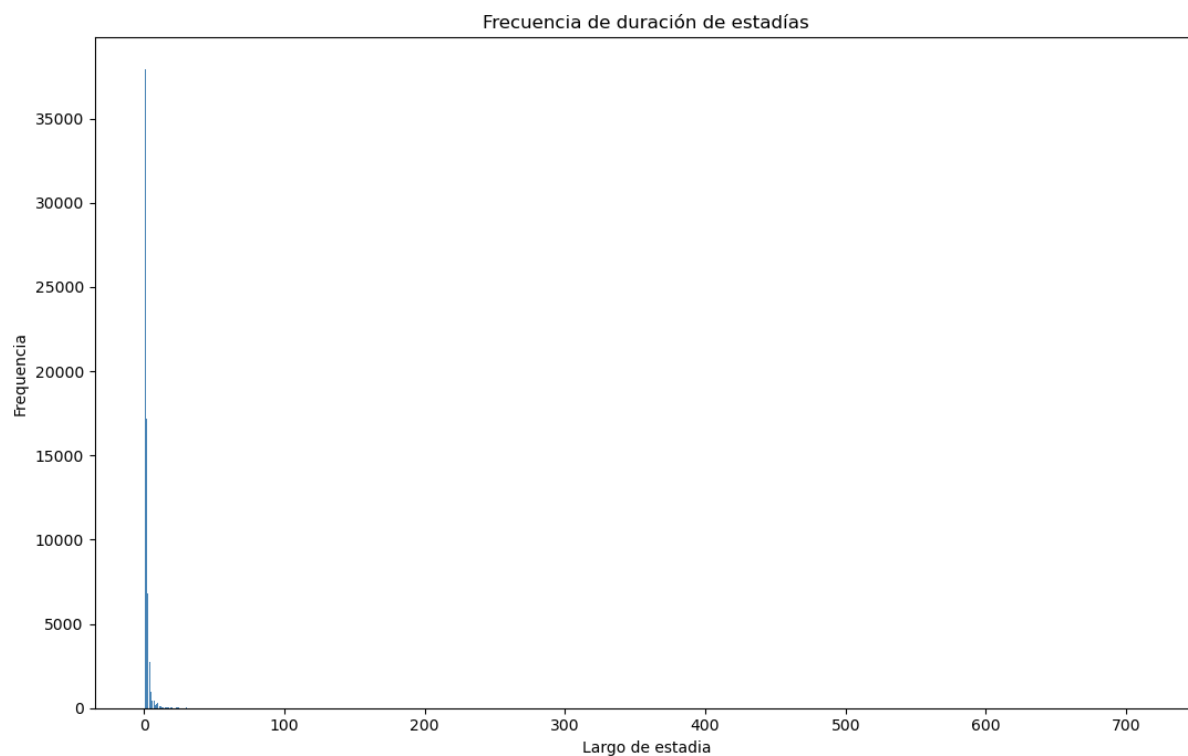


Figura 4.3.1.1: Frecuencia de cantidad de noches

En contraste, la Figura 4.3.1.2 refleja la distribución luego del proceso de depuración, considerando únicamente estadías de hasta 7 noches consecutivas. En esta versión depurada, se observa que la mayoría de las estadías se concentra en 1 o 2 noches, lo cual es consistente con el comportamiento esperado para huéspedes individuales durante fines de semana y feriados.

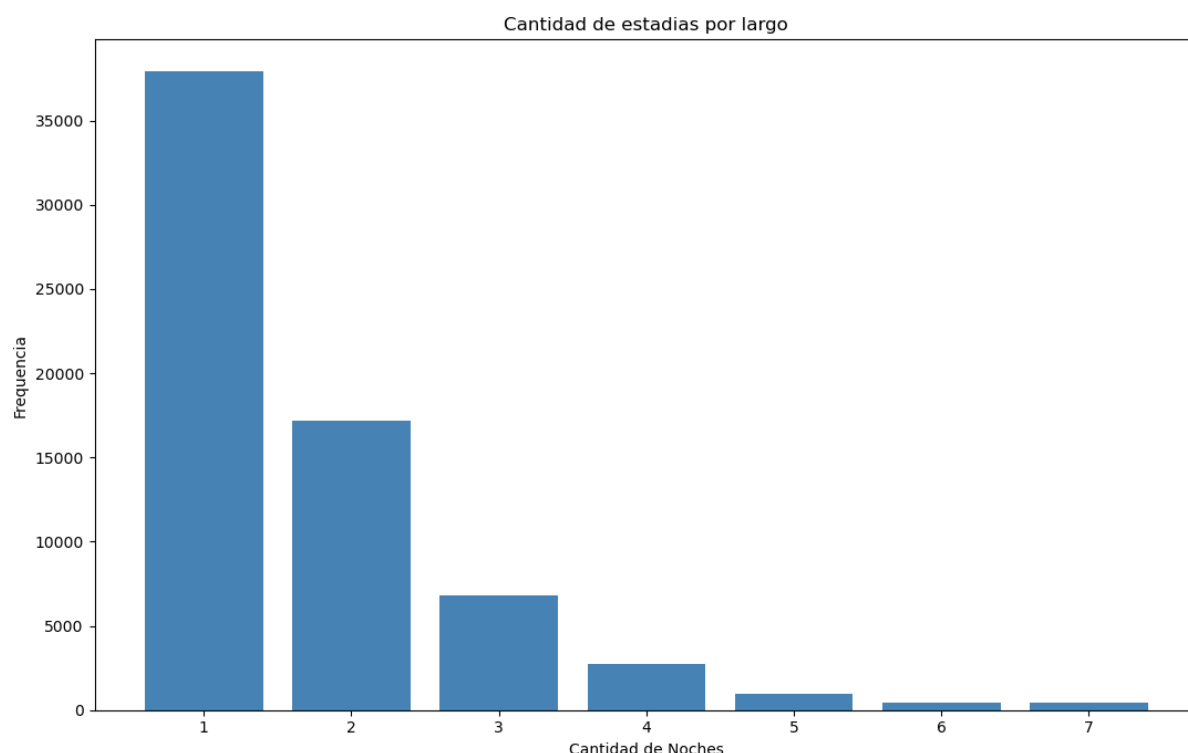


Figura 4.3.1.2: Frecuencia de cantidad de noches hasta 7 noches consecutivas

La eliminación de estos registros erróneos mejora significativamente la calidad de la variable **nights**, asegurando que represente con mayor precisión la duración real de las estadias y evitando distorsiones en los resultados del modelo.

4.3.2. Años (year)

Dentro de la variable **year**, se partió de la base que los años disponibles en el conjunto de datos abarcan el período comprendido entre 2019 y 2024, ambos inclusive. No obstante, es importante destacar que la pandemia de COVID-19, de público conocimiento, tuvo un impacto disruptivo sobre la operación del hotel. En particular, durante 2020 el establecimiento permaneció cerrado desde marzo hasta diciembre, sin registrar actividad comercial ni ocupación de habitaciones, en línea con las restricciones sanitarias impuestas a nivel nacional.

El año 2021, si bien implicó una reapertura parcial, estuvo caracterizado por una baja ocupación y severas limitaciones operativas, producto de las restricciones de movilidad, el aforo reducido en espacios comunes, y la baja demanda tanto del turismo interno como del segmento

corporativo. Este contexto generó una distorsión significativa en los datos de ocupación, lo cual debe ser considerado al momento de interpretar las series históricas.

A continuación, se presenta la figura 4.3.2, que resume la evolución de la ocupación total anual (en cantidad de noches) y la ocupación porcentual anual (porcentaje sobre la capacidad total disponible) en el período 2019–2024, lo que permite visualizar el impacto de la pandemia y la posterior recuperación progresiva del nivel de actividad del hotel:

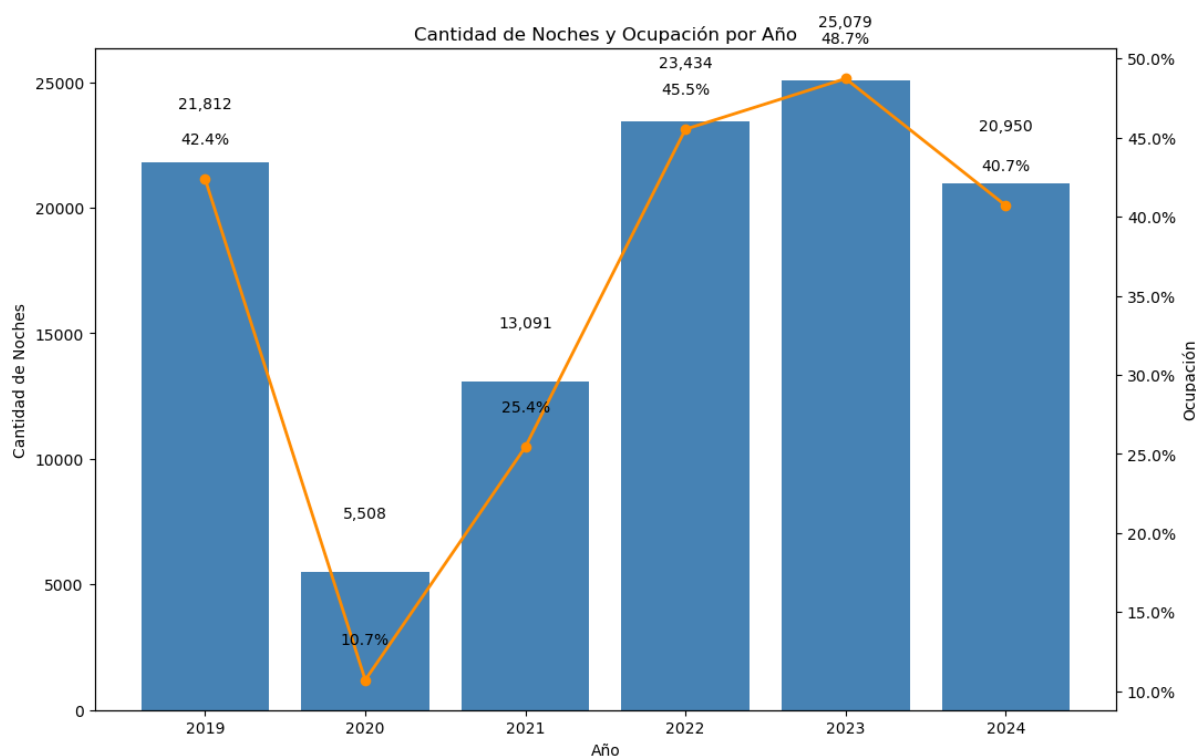


Figura 4.3.2: Ocupación nominal y porcentual anual

Con el objetivo de evitar sesgos en el entrenamiento de los modelos, se decidió excluir los años 2020 y 2021 del análisis, dado su carácter atípico. Los modelos resultantes se detallan en las próximas secciones.

4.3.3. Estadías complementary: Tarifas (`rate_amount`)

En relación con la variable `rate_amount`, se identificaron registros con tarifas igual a 0. Tras validarlo con el management del hotel, se confirmó que corresponden a cortesías otorgadas a periodistas, influencers y otras celebridades, en el marco de acciones publicitarias

orientadas a posicionar y difundir las distintas propuestas del hotel. Si bien estas tarifas forman parte de una estrategia comercial válida, no reflejan el comportamiento de precios y demanda estándar aplicado al público general. Estos registros podrían distorsionar el modelo de predicción de demanda, dado que estos valores atípicos afectan la precisión de las estimaciones, especialmente en modelos sensibles como la regresión.

Dicho esto, se excluyeron del análisis todos los registros con **rate_amount** igual a 0, considerando que no representan tarifas pagadas por clientes habituales del hotel, buscando mejora la calidad y la representatividad de los datos utilizados, permitiendo que los modelos de predicción trabajen con tarifas reales de mercado y optimicen de forma más precisa las estrategias de precios.

4.3.4. Moneda (currency)

En el conjunto de datos de estadías, una de las variables contenidas era **currency**. Haciendo un análisis exploratorio de la misma, se vio como las tarifas estaban expresadas tanto en dólares como en pesos argentinos. Esta inconsistencia en la moneda de los datos podría dificultar la comparación de precios y la estimación de la demanda. A continuación, se puede ver la distribución porcentual de tarifas expresadas en USD y ARS:

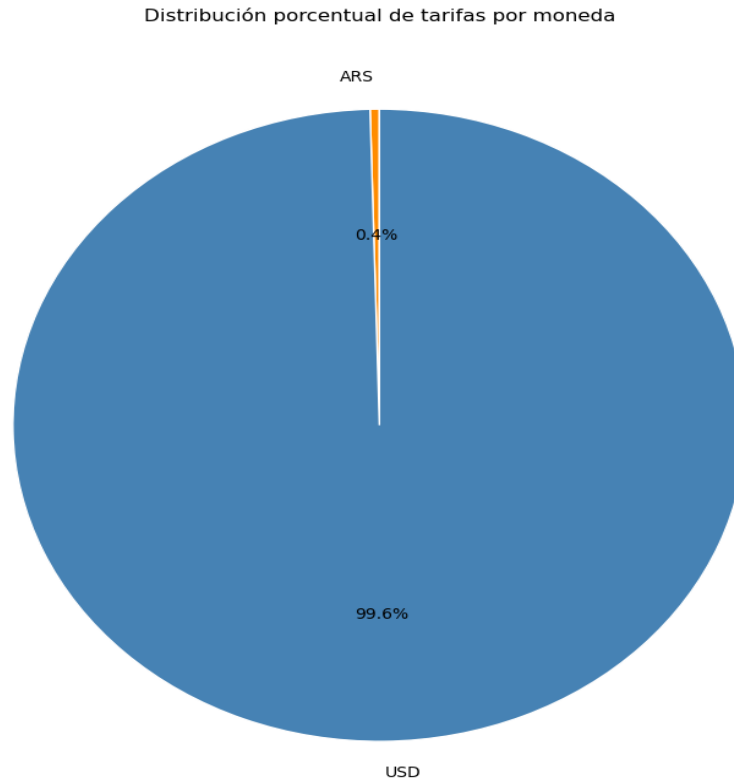


Figura 4.3.4: Distribución porcentual de tarifas expresadas en USD y ARS

Aunque la mayoría de las tarifas ya estaban expresadas en dólares estadounidenses (USD), se decidió unificar todos los valores en esta moneda para evitar sesgos en los análisis de precios y predicción de demanda debido a diferencias de escala. Esta conversión cobra especial relevancia en un contexto de alta volatilidad cambiaria como el argentino.

Para ello, se utilizaron los valores históricos del tipo de cambio Dólar CCL correspondiente a la fecha de cada estadía, obtenidos de una fuente confiable (Ámbito Financiero, s.f., Dólar CCL histórico, <https://www.ambito.com/contenidos/dolar-cl-historico.html>). La conversión se aplicó a todas las tarifas originalmente expresadas en pesos argentinos (ARS), permitiendo así una comparación homogénea y robusta en el análisis temporal.

4.3.5. Tipo de habitación (`room_type`)

Dentro de la variable `room_type`, se identificaron códigos de habitación que no pertenecen a habitaciones estándar, sino que corresponden a alojamientos futuros, “no show” o cargos de

servicios que no están directamente relacionados con hospedajes. Estos códigos erróneos, generan confusión en el análisis, ya que no representaban tipos de habitación reales.

Por otro lado, hay 3 tipos de habitaciones las cuales tienen una tarifa promedio muy superior al promedio del resto de las habitaciones. Las mismas son las habitaciones Suite y Ejecutivas. El promedio entre estas últimas y el resto puede verse en la figura 4.3.5:

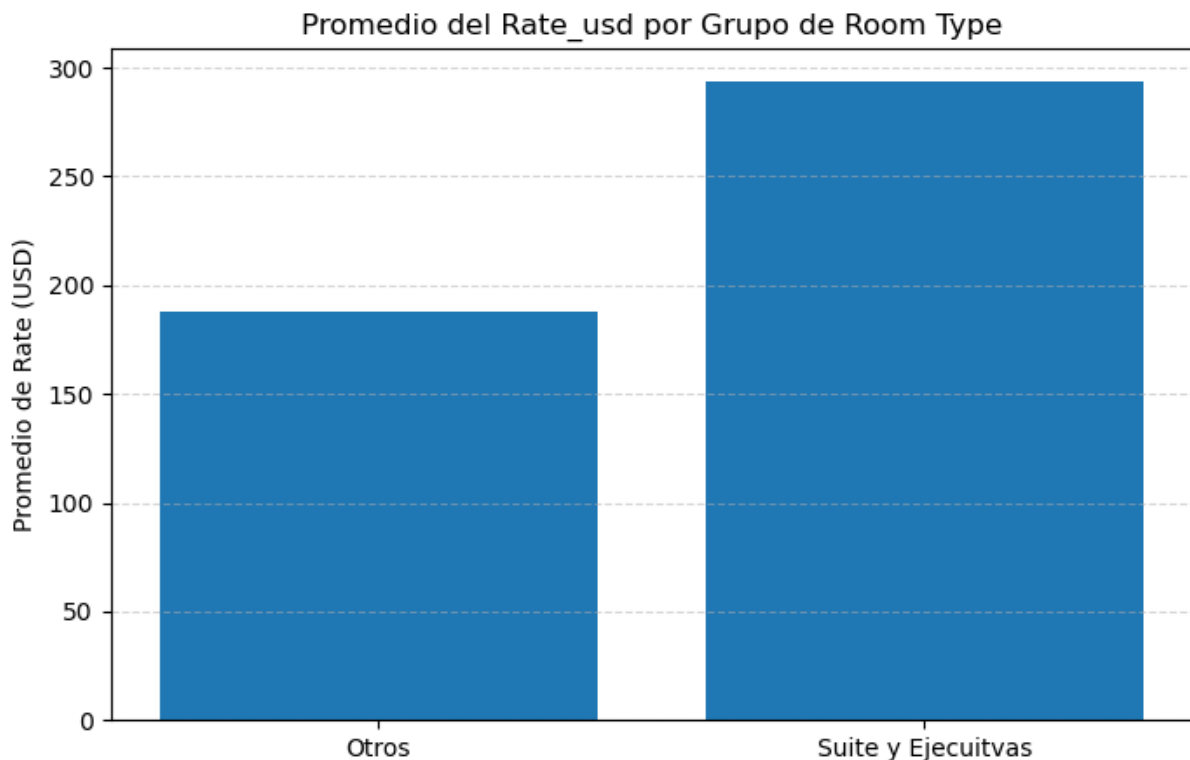


Figura 4.3.5: Rate promedio para habitaciones Suite y Ejecutivas vs. el resto

Estos registros distorsionaban la segmentación de las habitaciones, lo que podrían haber afectado el análisis de demanda y el ajuste de precios. Esto se menciona principalmente a que este tipo de habitaciones Suite y Ejecutivas, son la gran minoría entre el total de habitaciones. Para ser más preciso, solo 13 de las 141 habitaciones (9%), son de la clase mencionada. Adicionalmente, al analizar el revenue generado por este tipo de habitaciones, se observa que el mismo cae al 7% sobre el revenue total.

Con esto en mente, se excluyeron estos registros de tipo de habitación con códigos erróneos y se retuvieron solo aquellos que representaban habitaciones estándar ofrecidas por el hotel. También, se verificaron los registros de "no show" y "futuros alojamientos" y se aseguraron de que no se incluyeran en el conjunto de datos final, para evitar distorsionar la segmentación de la demanda y la optimización de precios.

Por último, se excluyeron las habitaciones de tipo Suite y Ejecutivas para no distorsionar las tarifas comúnmente contratadas y, por ende, trabajar con aumentos de precios más realistas.

El detalle de códigos de **room_type** están incluidos en el **Apéndice C**.

4.3.6. Tarifa en USD (**rate_usd**)

A través de un nuevo Profiling Report realizado posteriormente a la limpieza de las secciones anteriores, se identificaron tarifas extremas que sesgan el análisis, generando confusión en el mismo, ya que no representaban tarifas habituales.

Estos registros distorsionaban los precios ofrecidos, y sobre todo, el aumento de precios entre las primeras 64 habitaciones y las siguientes, punto el cual será ahondado más adelante (**price_diff**).

Teniendo en cuenta esto, se definió un valor mínimo y un valor máximo para el **rate_usd**, determinado por el percentil 5 y 95 respectivamente, encontrado en el Profiling Report, excluyendo los valores por fuera de los límites del dataset.

El profile report de **rate_usd** está incluido en el **Apéndice C**.

Con la limpieza de los registros extremos del **rate_usd**, se asegura que solo se analicen las tarifas habituales, lo que mejora la precisión de las estimaciones de demanda, los cálculos de ocupación y el impacto del aumento de precios producto de la política de RM actual.

Más allá de filtrar tarifas que no reflejan precios habituales de mercado, es igualmente importante asegurarse de que la variable **rate_usd** conserve un nivel suficiente de variabilidad. Esto se debe a que la relación entre precio y demanda es un componente central en los modelos de predicción aplicados a Revenue Management. Si los precios observados no muestran suficiente dispersión, el modelo pierde capacidad para aprender cómo varía la demanda frente a distintos niveles tarifarios, lo que limita seriamente su poder predictivo. Por lo tanto, garantizar una adecuada variabilidad en los precios no solo mejora la calidad del análisis, sino que resulta clave para construir estimaciones confiables y útiles para la toma de decisiones.

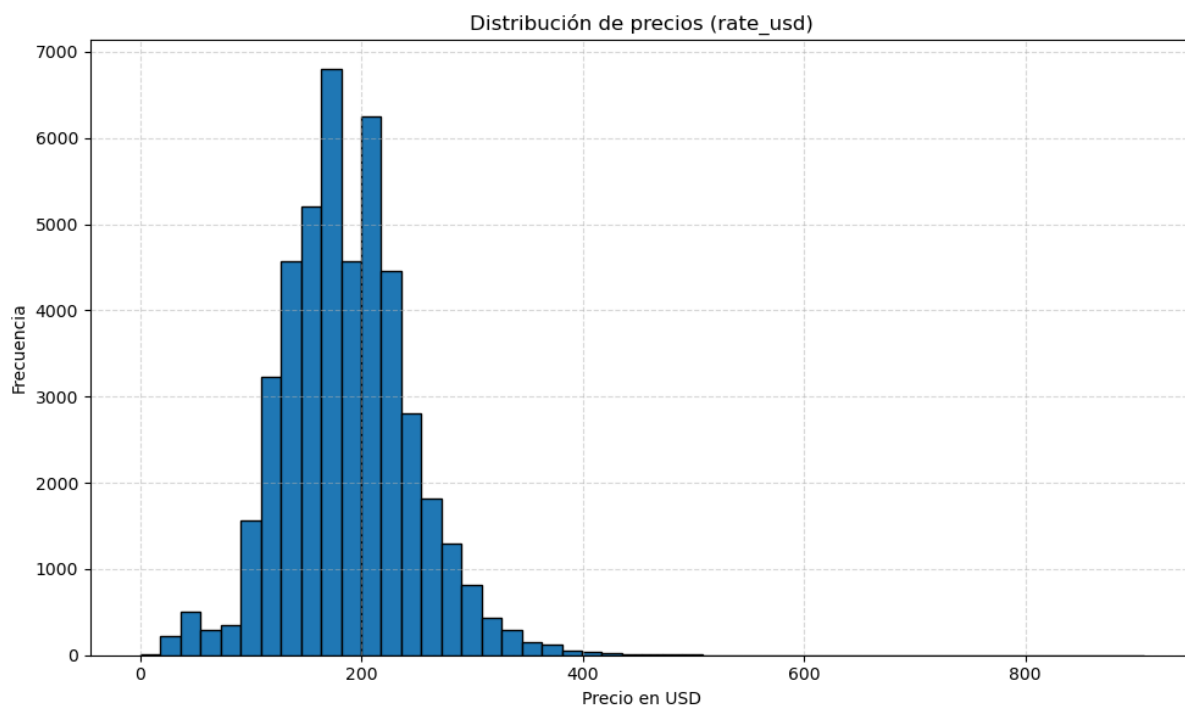


Figura 4.3.6.1: Distribución de **rate_usd**

Como se observa en la Figura 4.3.6.1, la distribución de **rate_usd** presenta una dispersión significativa, lo que respalda su utilidad como predictor.

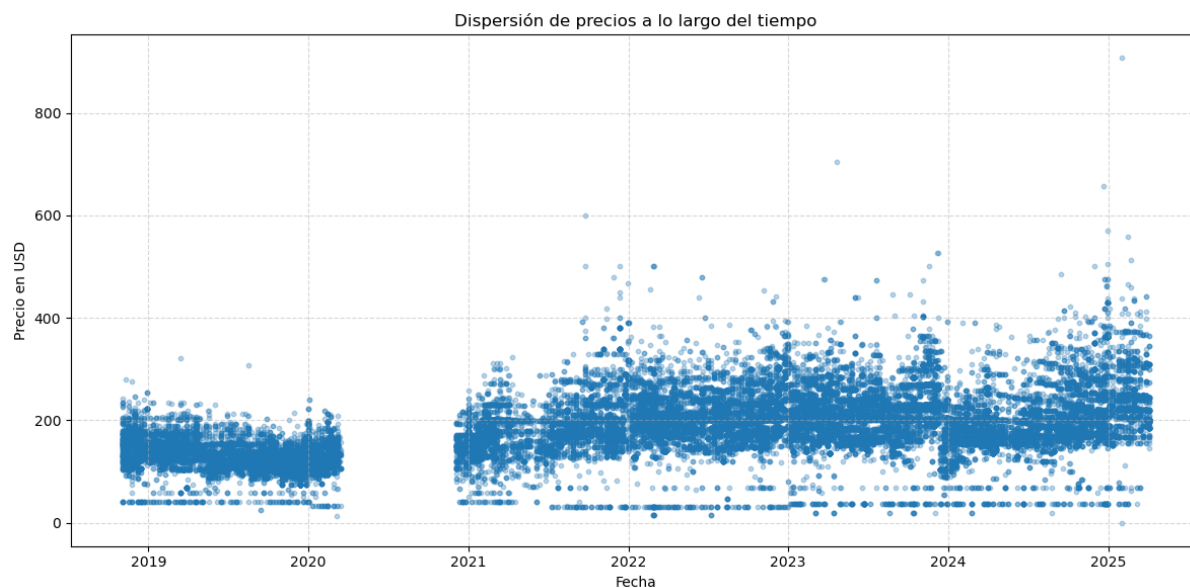


Figura 4.3.6.2: Dispersión de **rate_usd** a lo largo del tiempo

Por otra parte, la Figura 4.3.6.2 muestra cómo esta tarifa varió a lo largo del tiempo, reflejando dinámicas estacionales, promociones y ajustes de precio realizados por el hotel. Esta

variabilidad es clave para que el modelo pueda generar contrafactuales confiables sobre escenarios de precios alternativos.

Para asegurar la coherencia metodológica, el análisis de dispersión de la variable `rate_usd` presentado a continuación incluye únicamente observaciones correspondientes a los años considerados válidos para el modelado: 2019, 2022, 2023 y 2024. Se excluyeron explícitamente los años 2020 y 2021 debido a lo mencionado en la sección 4.3.2.. Este criterio de selección fue aplicado de forma consistente a lo largo de todo el trabajo, y se refleja también en los datos utilizados para entrenar y validar los modelos predictivos. La Figura 4.3.6.3. a continuación muestra la dispersión de `rate_usd` únicamente para los años incluidos en el análisis:

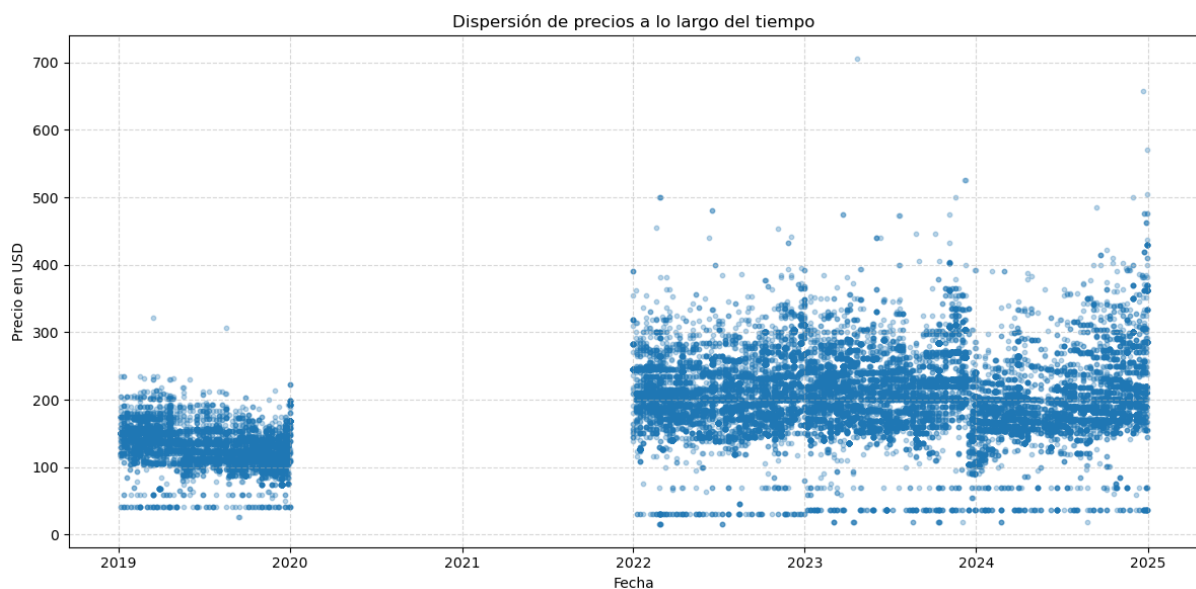


Figura 4.3.6.2: Dispersión de `rate_usd` a lo largo del tiempo de los años incluidos en el modelo

4.4. Preparación de los datos y ETL

En general, la calidad de los datos en este conjunto fue un desafío debido a los errores en las auditorías, la presencia de valores atípicos, y la inconsistencia en las monedas y categorías de habitaciones. Sin embargo, se tomaron medidas adecuadas para limpiar y normalizar los datos:

1. Exclusión de registros erróneos (como noches igual a 0 y tarifas de empleados).
2. Unificación de las tarifas a una sola moneda para facilitar la comparación.

3. Filtrado de registros irrelevantes relacionados con "no show" y otros cargos no representativos.

Estas acciones mejoraron la calidad de los datos y aseguraron que los modelos predictivos, como la regresión lineal y Random Forest, operaran sobre datos consistentes y representativos, lo que permitió obtener resultados más precisos y útiles para la optimización de precios y la gestión de la demanda en el hotel.

Luego de aplicar todos los criterios de limpieza y depuración de datos detallados en las secciones anteriores, el tamaño de la base se redujo de 94.141 filas y 12 columnas a 32.317 filas y 21 columnas. Esta poda fue necesaria para asegurar la calidad y coherencia de la información utilizada en los modelos predictivos. Cabe destacar que el recorte también responde al foco específico de la tesis: la demanda del público individual durante fines de semana, feriados y fines de semana largos. Para dimensionar esto, basta con señalar que aproximadamente el 30% de los días del año corresponden a viernes, sábados o vísperas de feriados, lo que implica que el modelo trabaja sobre un subconjunto naturalmente acotado del total de fechas posibles, en línea con el objetivo estratégico del estudio.

En la siguiente sección, se comentará sobre las variables de interés generadas en el proceso de ETL, para lograr la metodología desarrollada.

Capítulo 5

Metodología

En la presente sección, se detallará en profundidad la metodología implementada y su paso a paso, hasta lograr el modelo.

5.1. Modelado de la demanda

El modelado de la demanda en este contexto tiene como objetivo principal construir un contrafactual robusto que permita estimar cuál habría sido la ocupación del hotel en escenarios alternativos de precios. Para ello, se busca estimar la cantidad de noches reservadas por los clientes en función de distintas variables explicativas como el tipo de habitación, el precio, las características estacionales, los fines de semana largos y los días festivos. Este enfoque permite simular cómo podría haberse comportado la demanda bajo distintas estrategias de precios.

Para llevar a cabo este modelado, se realizó un proceso de extracción, transformación y carga (ETL) de datos que incluyó diversas fases:

1. **Preprocesamiento y Limpieza de Datos:** Se filtraron y limpiaron los datos para eliminar registros erróneos (por ejemplo, registros de habitaciones no válidas o tarifas preferenciales). Se excluyeron registros con precios excesivamente altos o estancias superiores a una semana.

2. **Expansión de Datos:** Se expandieron las fechas de cada reserva para cada noche de estadía. Esto permitió tener un conjunto de datos donde cada fila representa una habitación ocupada en una fecha específica.
3. **Variables de Interés:** A partir de la fecha específica, se calculó información como la estación del año (verano, otoño, invierno, primavera) y si la noche correspondía a un fin de semana o feriado. También se calculó el tipo de cambio para ajustar las tarifas a una moneda común (USD).
4. **Agregación de Datos:** Se agregaron columnas que representaban características clave para evaluar la política de Revenue Management actual. Entre ellas, se incluyó el promedio de tarifa (**avg_first_64**) correspondiente a las primeras 64 habitaciones vendidas en cada noche, reflejando el precio aplicado bajo la política base. A partir de la 65.^a habitación, el hotel aplica un aumento porcentual fijo —generalmente del 30%— respecto al precio anterior. Para capturar este cambio, se calculó también el promedio de tarifa de las habitaciones vendidas por encima del umbral y el cambio porcentual entre ambos niveles de precio (**price_diff**). Estas variables permiten analizar cómo responde la demanda ante ese salto tarifario y si la política implementada efectivamente maximiza el ingreso.

El modelo se enfoca en predecir la demanda de habitaciones para aquellas fechas en las cuales la ocupación superó un umbral de 64 habitaciones, utilizando como variables explicativas características como el precio de la habitación, la estacionalidad, el tipo de noche y la presencia de eventos. Para dar contexto, de las 478 noches identificadas como viernes, sábados o vísperas de feriados —es decir, las fechas relevantes para el público individual—, en 262 noches se superó efectivamente el umbral de 64 habitaciones ocupadas. Esto representa aproximadamente el 55% de los casos, y delimita el universo concreto sobre el cual se aplica el modelo. Las métricas generadas a partir de la agrupación por noche, como la cantidad de noches ocupadas, el precio promedio (**avg_first_64**) y el cambio porcentual de tarifa (**price_diff**), fueron fundamentales para el análisis posterior y la validación de la política de precios implementada.

Con estas métricas generadas, el objetivo es evaluar críticamente la política actual de Revenue Management del hotel. Más específicamente, se busca determinar si el aumento tarifario del 30% aplicado a partir de la 65.^a habitación vendida es efectivamente el que maximiza los

ingresos por habitaciones transeúntes, o si existe margen de mejora dentro de la misma lógica operativa. La hipótesis es que la política actual resulta subóptima, en tanto el aumento implementado excede el punto de maximización de ingresos, generando una pérdida potencial de demanda e ingresos adicionales, tanto por alojamiento como por consumos complementarios.

La elección de utilizar el promedio de tarifa de las primeras 64 habitaciones (**avg_first_64**) y compararlo con el promedio de las tarifas de las unidades vendidas por encima de ese umbral responde a que esa es exactamente la estrategia comercial vigente: se mantiene una tarifa fija hasta alcanzar un 50% de ocupación, y luego se aplica un salto tarifario único del 30%. Este modelo refleja con precisión el funcionamiento real del sistema de precios del hotel, y por solicitud del management, no se permite modificar la estructura del umbral ni explorar variantes más complejas como aumentos progresivos o escalonados. Por lo tanto, el análisis se concentra en cuantificar si el aumento actual es el adecuado, manteniendo constante la lógica operativa, y no en proponer alternativas que modifiquen la regla de negocio base.

5.2. Aplicación del modelo de regresión lineal con Bootstrap

El primer paso para predecir la demanda fue el uso de un modelo de regresión lineal, que permitió modelar la relación entre el número de noches ocupadas (**cantidad_noches**) y las variables predictoras (como las tarifas de las habitaciones y las características de la fecha, como la estacionalidad). Sin embargo, dado que se quería incorporar la variabilidad inherente a las predicciones de la demanda, se utilizó Bootstrap, una técnica estadística que permite obtener estimaciones robustas de la incertidumbre mediante el re-muestreo con reemplazo. Una ventaja adicional del enfoque bootstrap es que, al obtener múltiples estimaciones para una misma predicción, también es posible construir intervalos de confianza empíricos.

1. **Entrenamiento del Modelo:** Se dividió el conjunto de datos en dos partes: entrenamiento (con datos hasta 2022) y test (2023).

2. **Estimación de la Incertidumbre:** Se entrenaron 1000 modelos con diferentes muestras Bootstrap y se calculó la media de las predicciones generadas por los modelos. Esto permitió capturar la variabilidad en las predicciones y estimar intervalos de confianza, delimitando el rango dentro del cual se espera que se ubique la demanda real con un 90% de confianza.
3. **Evaluación del Modelo:** El desempeño del modelo se evaluó utilizando el error cuadrático medio (MSE) y en particular, la raíz del error cuadrático medio (RMSE) sobre el conjunto de prueba. Estas métricas permitieron medir la capacidad del modelo para predecir de manera precisa el número de noches ocupadas, a pesar de la variabilidad inherente de la demanda.

5.3. Reemplazo por modelo de Random Forest

Aunque el modelo de regresión lineal con Bootstrap proporcionó predicciones razonables, los resultados no fueron satisfactorios desde el punto de vista de las métricas de evaluación y la precisión de las predicciones en comparación con las expectativas. Debido a la naturaleza no lineal de la demanda de habitaciones y la complejidad de las relaciones entre las variables, se decidió reemplazar el modelo de regresión lineal por un modelo de Random Forest.

1. **Entrenamiento de Random Forest:** Se usó el modelo de Random Forest, que es un conjunto de árboles de decisión.
2. **Bootstrap:** Similar al modelo anterior, se entrenaron 1000 modelos de Random Forest con muestras aleatorias del conjunto de entrenamiento. Esto no solo permitió incorporar la variabilidad de los datos, sino también, y más importante, fue posible estimar intervalos de confianza al 90%.
3. **Evaluación del Modelo:** Al igual que con la regresión lineal, el modelo de Random Forest se evaluó utilizando el MSE y RMSE sobre el conjunto de prueba. Este modelo, mostró una mejor capacidad predictiva que la regresión lineal, al capturar relaciones complejas entre variables.

5.4. Validación del modelo

La validación del modelo es un paso crucial para evaluar su capacidad de generalizar a nuevos datos. En este caso, el modelo de Random Forest se validó utilizando datos del año 2024, para comprobar si las predicciones realizadas con el modelo de 2023 eran consistentes en el futuro.

1. **Simulación de aumentos de precios:** Se generaron predicciones en base a los modelos entrenado, variando el **price_diff** (el cambio porcentual en la tarifa de las habitaciones) entre 0% y 45%. Esto permitió explorar cómo diferentes estrategias de precios afectarían la demanda y, por ende, los ingresos del hotel.
2. **Comparación de resultados:** Se compararon los resultados de 2023 con los obtenidos para 2024. El rendimiento del modelo se evaluó al identificar el precio y el **price_diff** que maximizaban los ingresos del hotel en cada año, y se compararon con el escenario original de precios, identificando si la estrategia implementada era la óptima.
3. **Evaluación del revenue:** Finalmente, se generaron gráficos mostrando cómo los ingresos anuales variaban en función del aumento de precios. Esto permitió visualizar los puntos de máxima rentabilidad y compararlos con los ingresos reales obtenidos con la estrategia de precios existente. También se analizó la demanda estimada en los extremos de los intervalos de confianza para visualizar empíricamente las limitaciones del modelo.

En resumen, la validación consistió en extrapolar los resultados al año 2024 para comprobar que el modelo podía predecir el comportamiento de la demanda bajo diferentes condiciones de precio. Esto permitió evaluar la eficiencia de la estrategia de precios actual, considerando la variabilidad en la demanda.

Capítulo 6

Resultados

A continuación, vamos a analizar en profundidad los resultados estadísticos asociados a los modelos.

6.1. Análisis de la demanda predicha

El hotel bajo análisis cuenta con una capacidad total de 141 habitaciones. No obstante, este estudio se focalizó exclusivamente en las 128 habitaciones correspondientes a las categorías estándar, excluyendo las habitaciones tipo Suite y Ejecutivas (13 unidades), como se detalló en el apartado 4.3.4. Esta decisión responde al objetivo de construir un modelo más representativo del comportamiento de la demanda del segmento más sensible al precio, que constituye el núcleo de la estrategia de RM del establecimiento.

Actualmente, el sistema de pricing del hotel se basa en una lógica reactiva, donde los precios se ajustan semana a semana a partir de una regla simple: cuando y si la ocupación supera el 50%, se aplica un aumento porcentual en las tarifas. Este enfoque, si bien funcional, carece de una base analítica sólida y no contempla estimaciones precisas de demanda.

En este contexto, el modelo propuesto se orienta a predecir la demanda diaria de habitaciones en aquellos días donde se espera superar el umbral del 50% de ocupación, lo que equivale a 64 habitaciones ocupadas sobre las 128 consideradas. Esta definición delimita claramente el ámbito de aplicación del modelo, centrado en escenarios de alta demanda donde las decisiones tarifarias tienen un mayor impacto económico.

Para alimentar el modelo predictivo, se construyeron variables independientes relacionadas con el pricing. En particular, se calculó la tarifa promedio de las primeras 64 habitaciones vendidas, así como la tarifa promedio del segundo bloque de habitaciones (a partir de la 65° unidad). A partir de estas dos variables, se derivó una tercera: el porcentaje de aumento ambos bloques.

Una vez entrenado el modelo con datos históricos, se estimó la demanda diaria para el año 2023, no incluido en el entrenamiento. Esta predicción fue evaluada con la métrica RMSE (Root Mean Squared Error), explicada en la sección 3.3.1.

En este caso, un RMSE bajo indica que el modelo logra capturar adecuadamente los niveles de ocupación diaria en los días de mayor demanda, lo que valida su uso como herramienta de apoyo en decisiones de Revenue Management. En secciones posteriores se comparará esta métrica entre distintos enfoques (regresión lineal y Random Forest), así como su impacto en los ingresos.

Dado que la demanda hotelera se organiza como una serie temporal —donde cada observación representa habitaciones ocupadas por noche—, se evaluó si existía autocorrelación entre días consecutivos. Este análisis permite validar la elección del modelo estadístico, considerando que enfoques como ARIMA requieren estructura autocorrelada.

Para ello, se confeccionó un autocorrelograma (ACF) de la variable objetivo **cantidad_noches**, presentado en la Figura 6.1. El eje X representa los rezagos en días, mientras que el área sombreada en celeste corresponde al intervalo de confianza del 95%.

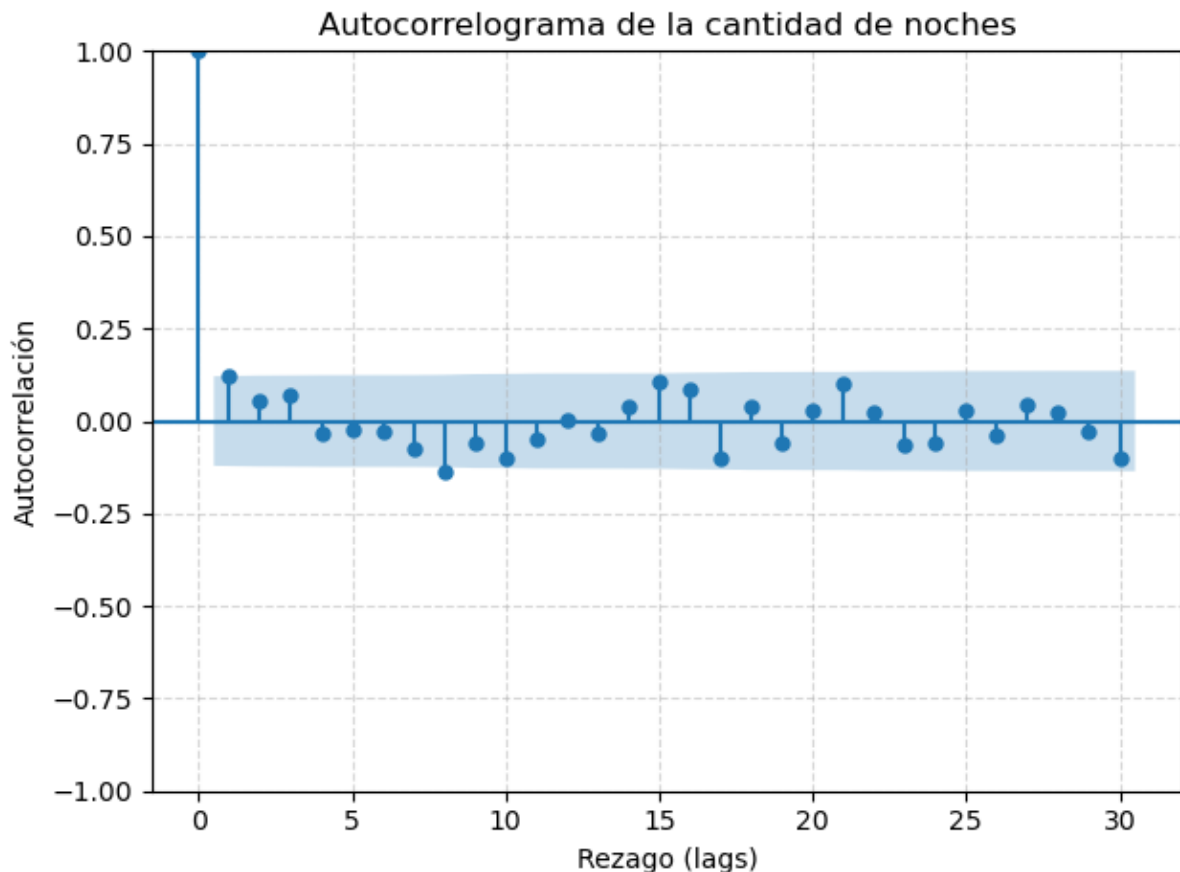


Figura 6.1. Autocorrelograma de la variable `cantidad_noches`.

Como puede observarse, salvo el rezago 0 (autocorrelación perfecta), el resto no muestra valores estadísticamente significativos que excedan el nivel de confianza definido. Esto indica que no existe una dependencia temporal fuerte entre días consecutivos y, por lo tanto, no se justifica aplicar un modelo autorregresivo como ARIMA

La elección de un modelo de regresión lineal resulta, en este sentido, adecuada, ya que los datos no presentan autocorrelación significativa. Si bien existen patrones estacionales —como estacionalidad semanal y mensual, analizados en las figuras 2.2.2 y 2.2.3— estos pueden ser capturados con variables categóricas como día de la semana, mes o temporada. Además, la selección de este modelo respondió a un objetivo metodológico específico: establecer un modelo baseline simple, transparente y de fácil interpretación, sobre la cual comparar el desempeño de modelos más complejos. Por lo tanto, su simplicidad no constituye una limitación, sino una decisión deliberada en el proceso de validación y construcción incremental del modelo.

6.2. Evaluación del modelo

La evaluación empírica de los modelos desarrollados se realizó inicialmente sobre el año 2023, considerado como conjunto de prueba. Este período fue seleccionado por no haber sido incluido en el entrenamiento, lo que permite observar el desempeño del modelo frente a datos "nuevos", simulando una situación real de pronóstico. Posteriormente, los modelos también fueron aplicados sobre datos de 2024 a modo de validación externa, cuyos resultados se discutirán más adelante.

Se calcularon los valores de RMSE para cada uno de los modelos aplicados sobre las predicciones de 2023. Esta métrica, ya introducida en la sección 6.1, brinda una primera comparación cuantitativa del desempeño entre los enfoques de regresión lineal y Random Forest.

Modelo	RMSE
Regresión lineal	15.06
Random Forest	12.83

Figura 6.2.1: RMSE por modelo

Como se observa en la figura 6.2.1, el modelo de Random Forest presentó un 15% menor RMSE, lo que indica una mayor precisión general en las predicciones realizadas sobre los días de alta demanda. Esta diferencia sugiere que la naturaleza no lineal y la capacidad del modelo de Random Forest para capturar interacciones complejas entre variables le permitió ajustarse mejor al comportamiento real de la demanda en contextos variados.

Con el objetivo de visualizar el desempeño a lo largo del tiempo, se construyeron gráficos mensuales que comparan la demanda real con la estimada por cada modelo, considerando únicamente las habitaciones demandadas por encima del umbral de 64 unidades vendidas por noche. Adicionalmente, y no menor, se incluyeron los intervalos de confianza de las habitaciones estimadas. Las siguientes figuras, permiten detectar sesgos, sobreestimaciones o subestimaciones sistemáticas en ciertos períodos.

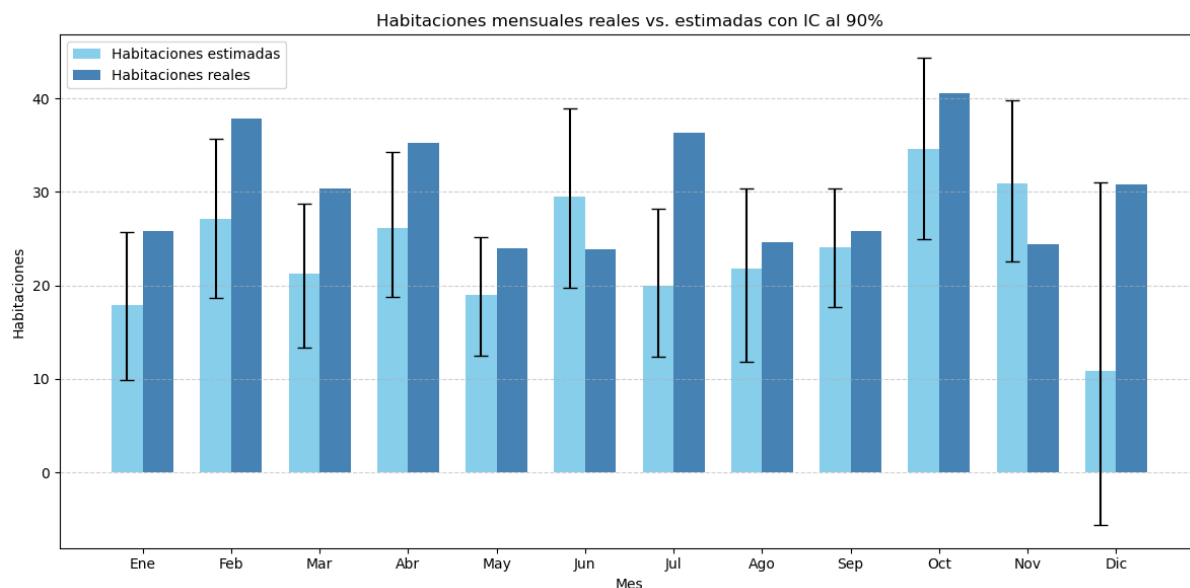


Figura 6.2.2: Demanda de habitaciones 2023 real vs estimada con Regresión Lineal incluyendo intervalos de confianza

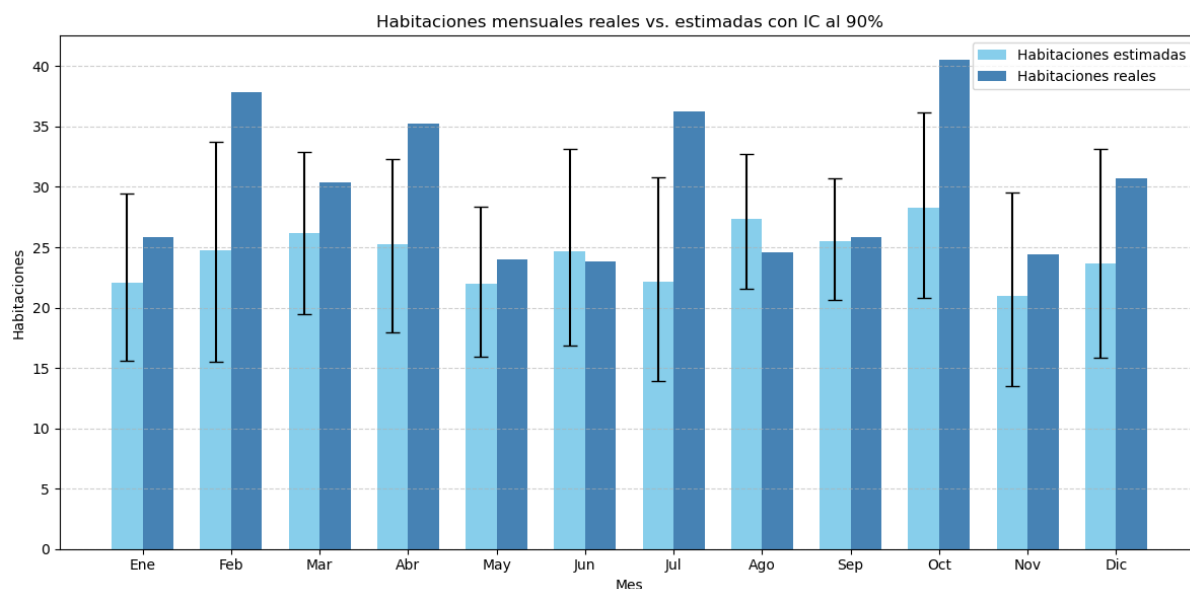


Figura 6.2.3: Demanda de habitaciones 2023 real vs estimada con Random Forest incluyendo intervalos de confianza

En ambos gráficos (figuras 6.2.2 y 6.2.3), se aprecia cómo se comportó cada modelo en relación con la demanda observada. El modelo de Random Forest muestra una mayor capacidad de adaptación, logrando seguir más de cerca las fluctuaciones reales mes a mes. No obstante, en ambos modelos se tiende a subestimar la demanda en los picos de mayor ocupación, como los meses de febrero, abril, julio y octubre. A pesar de esta tendencia general a la subestimación, en la mayoría de los casos la demanda real sí quedó comprendida dentro del margen de error estimado, especialmente en el modelo de Random Forest.

Estos resultados refuerzan la idea de que los métodos basados en árboles ofrecen ventajas importantes cuando se trata de capturar comportamientos no lineales en los datos, algo frecuente en la industria hotelera donde la demanda está influida por múltiples factores externos y efectos calendario complejos.

Con el objetivo de seguir mejorando el desempeño, se realizó una optimización de hiperparámetros del Random Forest mediante Grid Search. Se evaluaron combinaciones de parámetros críticos como **max_depth**, **min_samples_leaf** y **n_estimators**.

Luego de aplicar Grid Search y seleccionar los mejores hiperparámetros en base a la minimización del error de predicción, se reentrenó el modelo sobre el conjunto de entrenamiento y se evaluó nuevamente su desempeño sobre los datos de prueba de 2023.

El modelo optimizado logró alcanzar un RMSE de 12,05, mejorando respecto al modelo original (12,83). Esta reducción adicional del error, aunque más sutil, confirma que el ajuste fino de hiperparámetros mejora la capacidad predictiva al capturar con mayor precisión las variaciones en la demanda hotelera. En términos relativos, esta mejora representa una reducción del 6 % respecto al Random Forest sin ajuste, y del 20 % respecto a la regresión lineal, consolidando su robustez.

A continuación, podemos ver la comparación de la demanda real y estimada, posterior a la optimización de hiperparámetros:

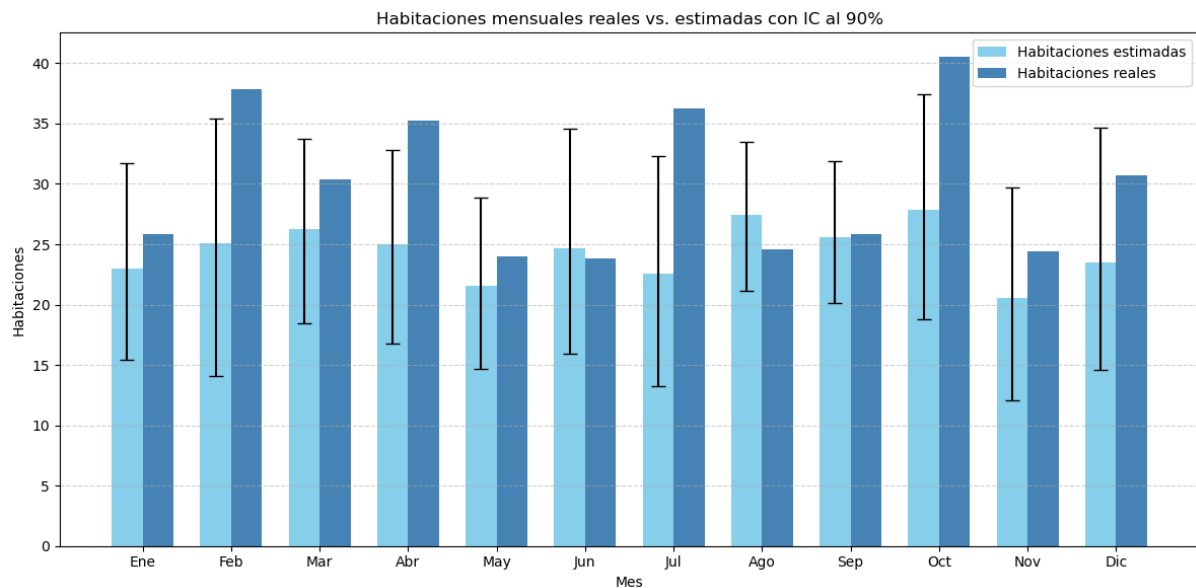


Figura 6.2.4: Demanda de habitaciones 2023 real vs estimada con Random Forest y GridSearch incluyendo intervalos de confianza

Ahora bien, para comprender mejor el alcance del error reportado, es importante poner en contexto el valor del RMSE obtenido. El modelo optimizado presenta un valor de 12,05, lo que significa que, en promedio, el modelo puede equivocarse por unas 12 habitaciones al estimar la demanda por encima del umbral de 64 habitaciones vendidas por noche. Dado que este modelo se aplica exclusivamente a noches con alta ocupación (donde puede haber hasta 64 habitaciones adicionales en juego), este error equivale a un desvío relativo del 19% sobre el total de habitaciones posibles en ese tramo. Desde una perspectiva operativa, este margen de error se considera razonable, ya que permite orientar decisiones sin exigir una precisión exacta que sería inalcanzable en la práctica.

Por último, considerando que cada habitación representa ingresos no solo por alojamiento sino también por servicios complementarios (estimados en USD 220 por huésped), una mejora parcial en la precisión puede traducirse en un impacto monetario significativo. Por lo tanto, el RMSE obtenido no solo es teóricamente sólido, sino también aceptable y útil para el management del hotel a la hora de ajustar políticas de precios y mejorar la eficiencia comercial en períodos de alta demanda.

6.3. Comparación de los modelos

La comparación entre los modelos implementados —regresión lineal y Random Forest con GridSearch— se realizó a partir de una evaluación empírica utilizando datos reales del año 2023. Ambos modelos fueron entrenados con información histórica y evaluados sobre un conjunto común de fechas, lo que permite establecer una comparación directa de su capacidad predictiva y su aplicabilidad en el contexto de RM hotelero.

Tal como se explicó en la sección 6.2, el Random Forest optimizado presentó el menor RMSE, demostrando mayor precisión en la predicción para días de alta ocupación. Esto se debe a su capacidad para modelar relaciones no lineales e interacciones complejas sin necesidad de transformar explícitamente los datos.

En cambio, la regresión lineal, aunque más simple e interpretable, mostró limitaciones ante patrones de demanda volátiles, especialmente en contextos con saltos de precio, feriados o fines de semana largos.

Una de las ventajas del Random Forest es que permite medir la importancia relativa de las variables predictoras, lo cual brinda información útil para la toma de decisiones.

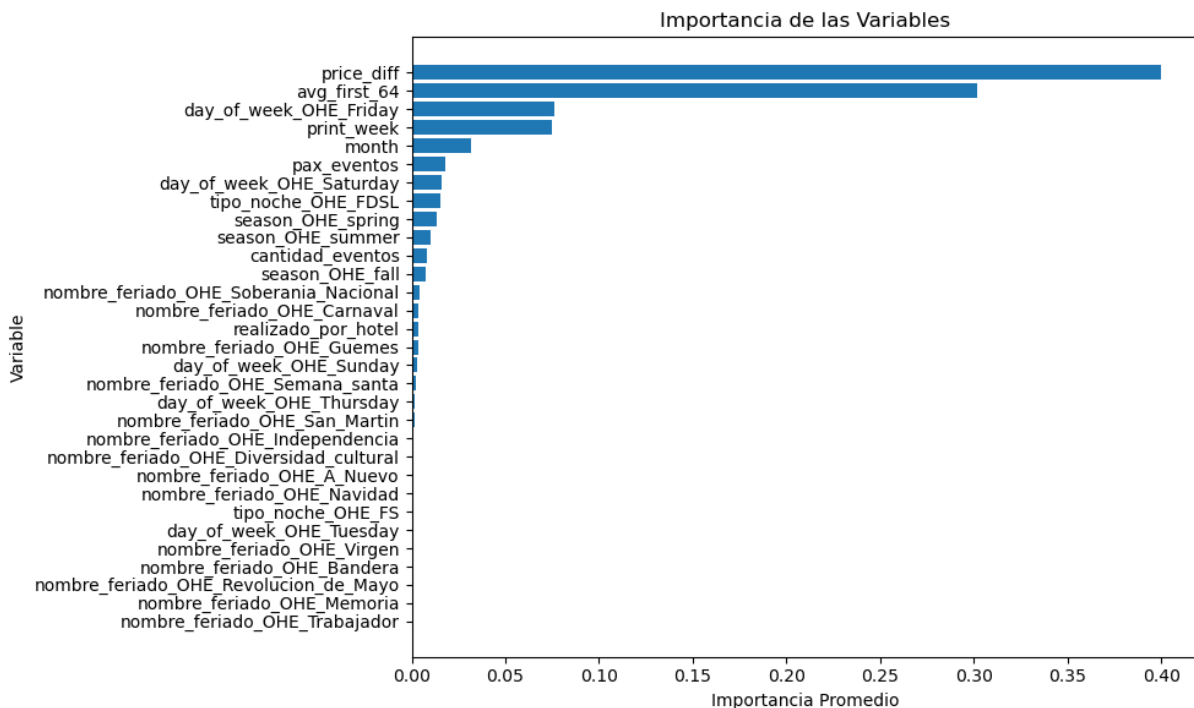


Figura 6.3.1: Feature importance del modelo de Random Forest

El gráfico de Feature Importance (figura 6.3.1) generado a partir del modelo entrenado, revela que la variable más influyente a la hora de estimar la demanda es la diferencia porcentual de precios entre bloques de habitaciones, es decir, el incremento de tarifa aplicado a medida que se agota el inventario más económico. Esto sugiere que los clientes son altamente sensibles a los aumentos tarifarios.

En segundo lugar aparece el precio promedio de las primeras 64 habitaciones, lo que muestra que el nivel de tarifa base es un determinante clave en el comportamiento de compra. Este punto va muy en línea con lo mencionado en la sección 4.3.6, acerca de la variabilidad en los precios.

Este análisis confirma empíricamente que los precios —y, en particular, su estructura de escalonamiento— juegan un rol central en la configuración de la demanda. Además, valida la hipótesis de que los aumentos tarifarios desproporcionados, pueden implicar grandes cambios en la demanda de habitaciones.

Para interpretar mejor la regresión lineal, se calcularon los coeficientes promedio a partir de los 1.000 modelos bootstrap. En la figura 6.3.2, se observa que **price_diff** posee el coeficiente de mayor magnitud absoluta, reafirmando su rol central. Esto coincide con los resultados del Random Forest.

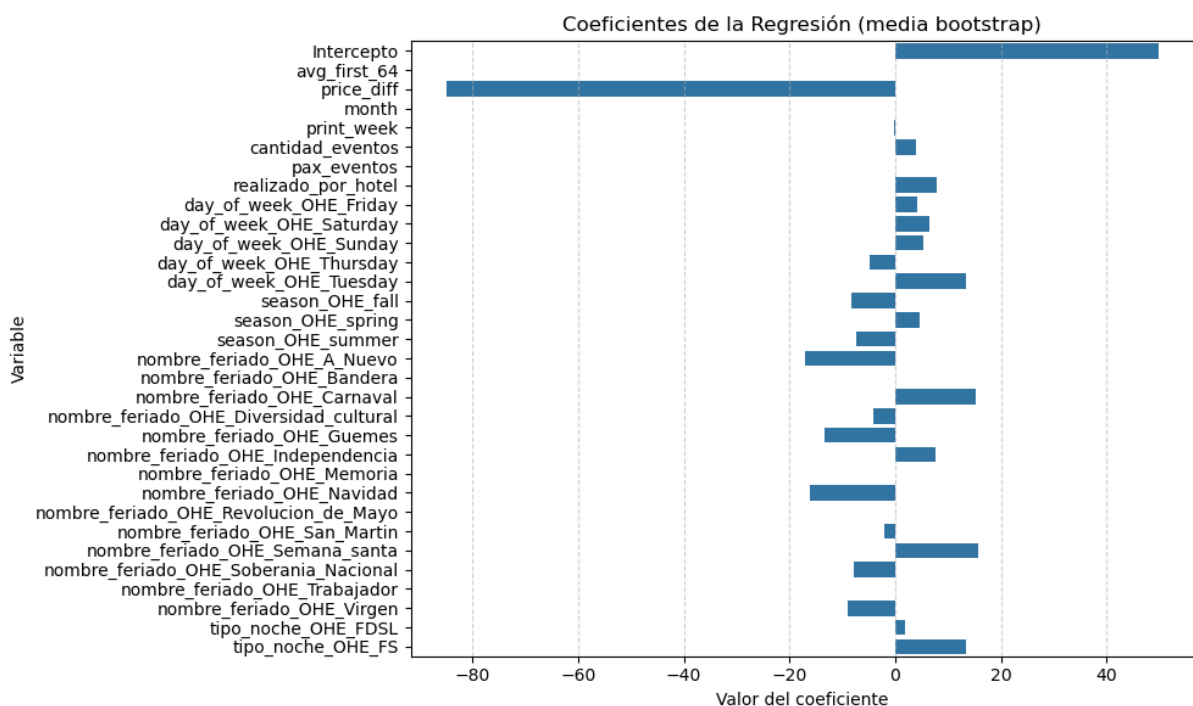


Figura 6.3.2: Coeficientes estimados por el modelo de regresión (media bootstrap)

En cambio, el coeficiente asociado a **avg_first_64** resulta cercano a cero, lo cual indicaría que su efecto marginal sobre la demanda es bajo o nulo cuando se mantiene constante el resto de las variables. Sin embargo, en el análisis de feature importance del Random Forest ambas variables (**price_diff** y **avg_first_64**) presentan una importancia relativa similar. Esta diferencia puede explicarse porque el modelo de regresión lineal asume una relación estrictamente lineal y aislada entre cada variable y la variable dependiente, mientras que el Random Forest capta interacciones y efectos no lineales que pueden no ser visibles en el marco de un modelo lineal tradicional. En otras palabras, aunque **avg_first_64** no tenga un efecto directo fuerte en forma lineal, sí podría estar involucrada en combinaciones de variables que explican bien la demanda.

Desde una perspectiva operativa, ambos modelos ofrecen ventajas distintas. La regresión lineal presenta una estructura simple, fácil de implementar y con bajo costo computacional. Es adecuada como una primera aproximación o para escenarios donde se prioriza la interpretabilidad.

En contraste, el modelo de Random Forest —aunque más complejo— ofrece una mayor capacidad de ajuste y predicción. Si bien demanda mayor potencia de cómputo y una calibración cuidadosa, especialmente cuando no se implementa *out-of-the-box*, sus resultados justifican el esfuerzo en escenarios de alta demanda y sensibilidad al precio, donde maximizar los ingresos es prioritario.

En función de los resultados, el Random Forest optimizado se consolida como la herramienta más robusta para predecir demanda en contextos dinámicos y sensibles al precio, y representa un aporte significativo para el diseño de estrategias de RM basadas en evidencia.

Capítulo 7

Conclusiones

A continuación, se comentarán los resultados monetarios y las recomendaciones finales para el management del hotel.

7.1. Logros alcanzados

Los resultados obtenidos a través del modelo de regresión lineal para el año 2023 muestran que, bajo los precios realmente observados, el aumento tarifario del 30% aplicado una vez alcanzado el umbral de 64 habitaciones no habría maximizado los ingresos por alojamiento. Por el contrario, el modelo sugiere que mantener el precio base (es decir, un aumento del 0%) habría generado mayores ingresos debido a una mayor ocupación. Este hallazgo no implica que aumentar precios sea intrínsecamente incorrecto, sino que, en este caso particular, el precio de partida podría haber sido demasiado elevado, restringiendo la captación de clientes con menor disposición a pagar. En este sentido, se refuerza una premisa clave del Revenue Management: maximizar ingresos implica ajustar los precios no solo en función de la ocupación, sino también en función de la heterogeneidad del mercado y su disposición a pagar.

Además, al considerar los ingresos marginales por huésped asociados a consumos adicionales dentro del hotel —como Alimentos y Bebidas, Health & Spa y otros servicios complementarios, estimados en un promedio de USD 220 por persona, según lo detallado en la sección 2.2—, resulta claro que la ocupación cobra un rol estratégico aún mayor. Aumentar el volumen de huéspedes, incluso a precios de alojamiento más bajos, puede resultar una estrategia más rentable cuando se contempla el revenue total del huésped. Por ello, se sugiere

como línea de investigación futura el desarrollo de un modelo integral de Revenue Management que considere tanto los ingresos por habitación como los ingresos por servicios complementarios a la hora de definir precios óptimos. Sobre este último punto, se ahondara en la sección 7.3.

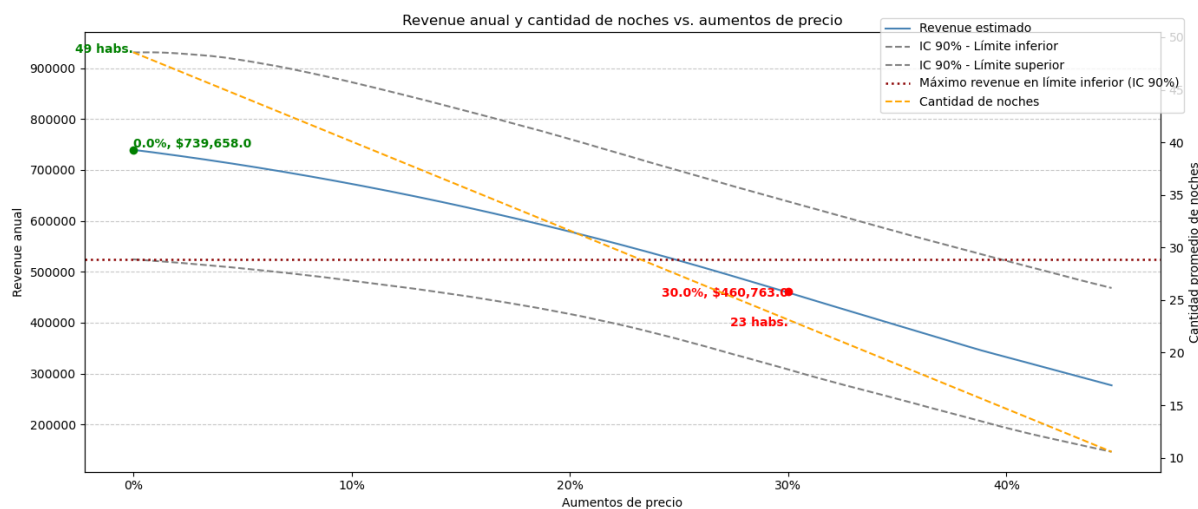


Figura 7.1.1: Revenue y cantidad estimada por medio de una regresión lineal para 2023

Como validación adicional de este hallazgo, se estimó el revenue asociado a la cota inferior del intervalo de confianza del modelo bootstrap, es decir, al escenario más pesimista dentro del rango de predicciones generadas. Incluso bajo este supuesto pesimista, el ingreso proyectado fue superior al revenue efectivamente obtenido con la política de aumento del 30%. Este resultado refuerza la solidez del modelo y pone de manifiesto que, aún considerando la incertidumbre inherente a las estimaciones, una política de precios menos agresiva habría generado un mejor desempeño económico. Por ende, no solo se valida la recomendación desde el punto de vista del valor esperado, sino también en términos de riesgo, mostrando que una estrategia más conservadora podría haber generado ingresos más altos en un amplio rango de escenarios posibles.

Este análisis también fue replicado para el año 2024, como se menciona en el punto 5.4. Al aplicar el modelo de regresión lineal a esos datos, los resultados fueron consistentes con los de 2023. Específicamente, se comprobó que mantener el precio base también habría sido la estrategia más óptima para maximizar los ingresos. Esto refuerza la conclusión de que la política actual de aumento de precios es subóptima. La figura 7.1.2 ilustra la estimación del revenue y la cantidad de habitaciones para 2024. Al comparar los gráficos de 2023 y 2024, se

observa una congruencia en los resultados, destacándose que el revenue asociado al límite inferior del intervalo de confianza sigue siendo superior al revenue real.

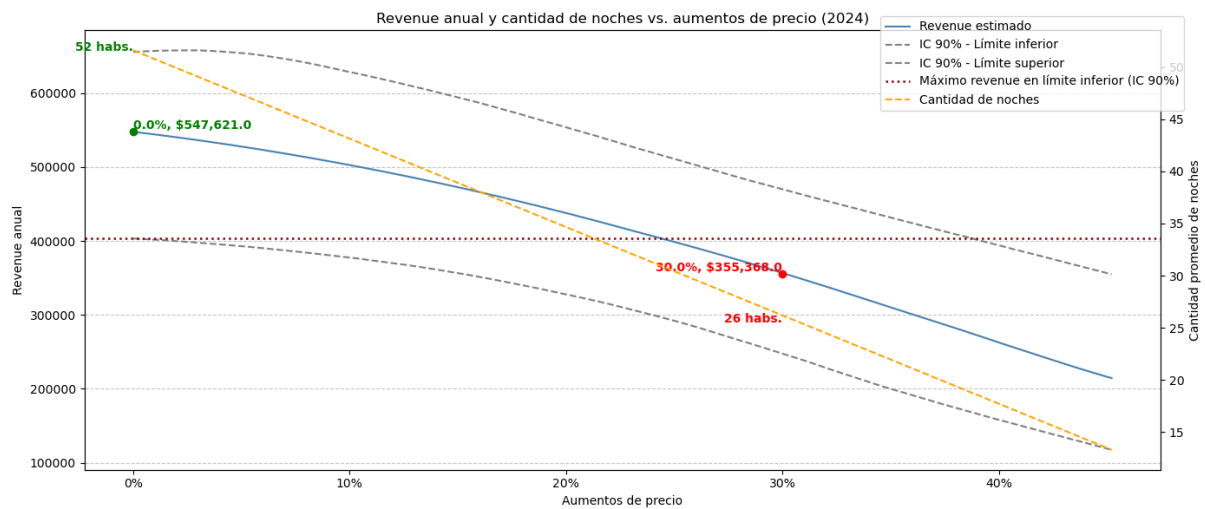


Figura 7.1.2: Revenue y cantidad estimada por medio de una regresión lineal para 2024

El análisis del modelo Random Forest para 2023 muestra que, en sintonía con los resultados obtenidos con el modelo de regresión lineal, los aumentos de precios fueron excesivos. Inclusive, observando el revenue asociado a la cota inferior del intervalo de confianza, el mismo es superior al revenue real.

Sin embargo, en este caso, los cambios en la demanda de habitaciones y el revenue asociado, son congruentes a las teorías tradicionales de RM, las que mencionan que un aumento en los precios, no conlleva una reducción de la demanda de igual magnitud. Según este modelo, el aumento óptimo de precios sería aproximadamente del 12% cuando la demanda supera el umbral de 64 habitaciones.

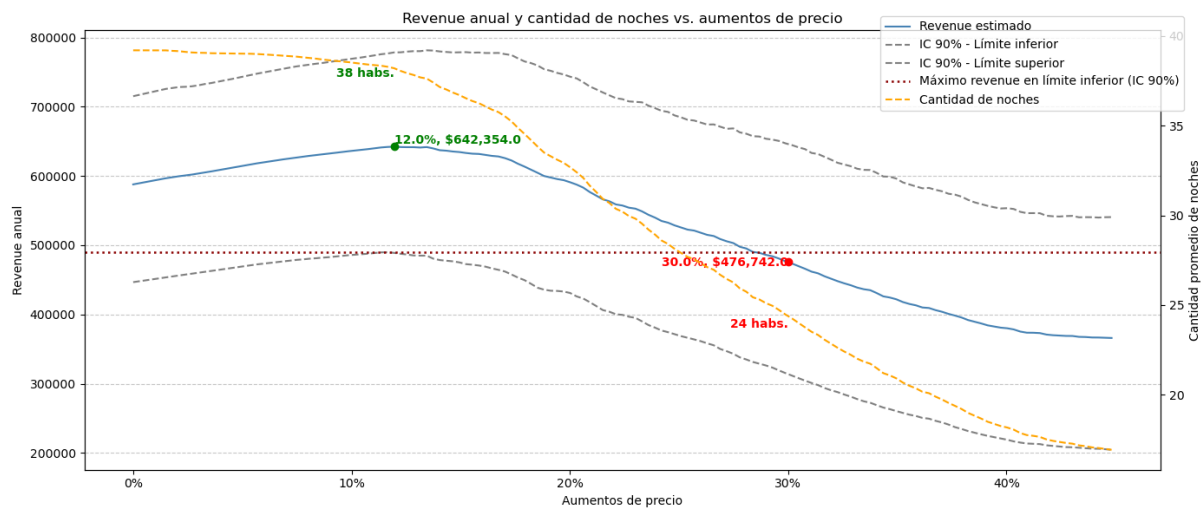


Figura 7.1.3: Revenue y cantidad estimada por medio de un modelo de Random forest para 2023

Este comportamiento coincide con estudios clave, como los de Kimes (1989) y Cross et al. (2009), que argumentan que maximizar ingresos no implica simplemente aumentar precios, sino ajustarlos según el comportamiento del consumidor y la elasticidad de la demanda.

Al volver a correr este modelo con los datos de 2024, se observó que el aumento de precios que maximiza los ingresos sería del 12,7%, lo que a nivel revenue es prácticamente igual que el asociado a un 12% de aumento. Esta similitud refuerza la estabilidad y consistencia de los resultados a lo largo del tiempo.

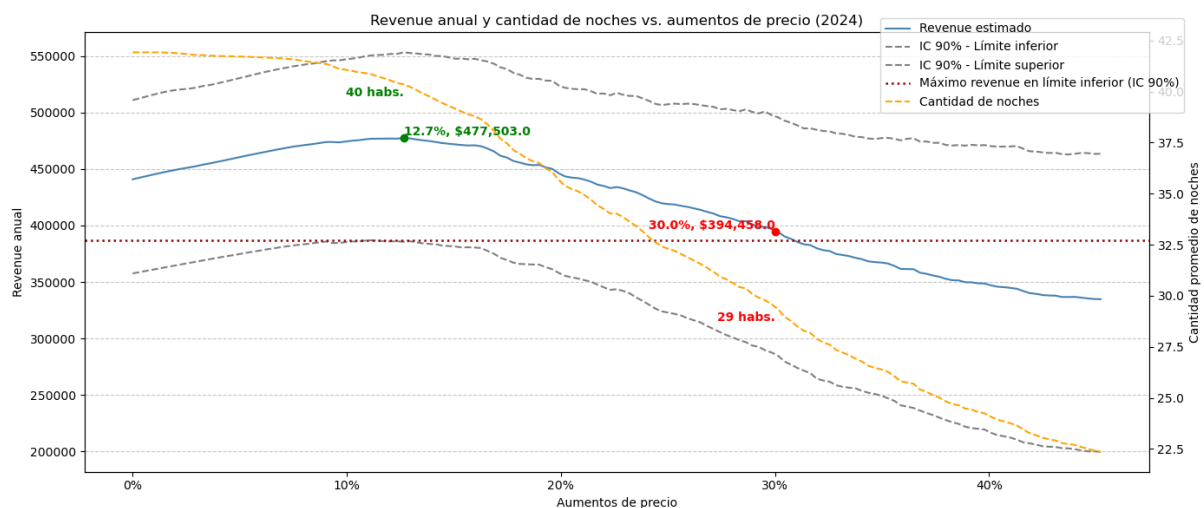


Figura 7.1.4: Revenue y cantidad estimada por medio de un modelo de Random forest para 2024

En cuanto a la cantidad de habitaciones, se mostró que un aumento en la demanda de habitaciones podría haber sido posible. Específicamente, un 52% para 2023 y un 38% para

2024. Este incremento es notable y, como se mencionó anteriormente, también incrementaría los ingresos marginales provenientes de los servicios complementarios del hotel.

Los resultados de ambos modelos son significativos y sugieren que, en este caso particular, una política de aumento más moderada habría generado mayores ingresos por alojamiento. Esto resalta la utilidad de modelos predictivos robustos como el Random Forest y refuerza la hipótesis de que la política tarifaria actual no es óptima.

7.2. Recomendaciones para el hotel

En base a los resultados obtenidos, este estudio propone dos líneas de acción fundamentales para optimizar la gestión de ingresos del hotel: (1) ajustar la política de precios y (2) mejorar la calidad y disponibilidad de los datos.

7.2.1. Política de precios

Una de las principales recomendaciones es reducir el porcentaje de aumento promedio, actualmente fijado en un 30%, a un rango más moderado de entre 10% y 15%, con el objetivo de maximizar el revenue sin perjudicar la ocupación. Esta recomendación tiene como objetivo maximizar el revenue al mismo tiempo que se optimiza la ocupación, particularmente en los meses donde la demanda es más sensible al precio. A continuación, se sugiere una aplicación diferenciada según el comportamiento estacional:

- **Meses con estacionalidad más baja (por ejemplo, los meses de invierno):** Durante estos períodos, la demanda suele ser más débil y sensible al precio. Por ello, se recomienda limitar los aumentos a un máximo del 10%. Mantener precios accesibles podría incentivar una mayor ocupación sin afectar significativamente los ingresos por habitación, al compensarse con volumen.
- **Meses con mayor estacionalidad (por ejemplo, en verano o en fechas especiales como Semana Santa, Navidad y Año Nuevo):** En estos casos, la demanda es más inelástica y sostenida. Se sugiere aplicar incrementos de hasta un 15%, especialmente

en momentos donde los atractivos del hotel, como jardines y piletas, representan una ventaja competitiva frente a otras opciones de alojamiento.

Esta política segmentada permitiría maximizar los ingresos durante los picos de demanda sin sacrificar la ocupación en períodos más sensibles al precio. No obstante, es importante aclarar que estas recomendaciones deben ser aplicadas con flexibilidad, considerando también variables exógenas que pueden modificar de manera abrupta el comportamiento de la demanda.

7.2.2. Estrategia de gestión de datos

Para mejorar la precisión de las estimaciones de demanda y el poder predictivo de los modelos en el futuro, es esencial mejorar los procesos de auditoría y recolección de datos, entre otras cosas. A continuación se detallan algunas acciones recomendadas:

- **Fortalecimiento de los procesos de auditoría:** Es esencial minimizar errores en variables clave, como la cantidad de noches (**nights**), mediante procedimientos de validación más estrictos que aseguren datos consistentes y completos.
- **Ampliación y diversificación de la base de datos:** El hotel debería incorporar nuevas fuentes de información, como precios de la competencia, registros de campañas de marketing y calendario de eventos relevantes en la zona. Estos factores externos pueden tener un impacto directo sobre la demanda, y su inclusión fortalecería el poder explicativo de los modelos.
- **Desarrollo evolutivo del modelo:** A medida que se acumulen más datos históricos en condiciones diversas, el modelo podrá ajustarse mejor a la dinámica del mercado. Esta mejora continua no solo incrementará la capacidad predictiva, sino que también permitirá establecer políticas de precios más eficientes y adaptativas.

Implementar estas recomendaciones contribuirá a una toma de decisiones más informada y dinámica, fortaleciendo tanto la estrategia de revenue management como la rentabilidad del hotel en el largo plazo.

7.3. Limitaciones del modelo

A continuación, se presentan las principales limitaciones identificadas en el modelo utilizado, tanto desde la perspectiva metodológica como desde la disponibilidad y calidad de los datos.

7.3.1. Limitaciones estructurales de los modelos utilizados

Una de las principales limitaciones de los modelos aplicados en esta investigación —regresión lineal y Random Forest— es su tendencia a subestimar la demanda en períodos de alta ocupación, especialmente en fechas clave para el público individual. Esta debilidad puede explicarse por la dificultad que presentan ambos enfoques para capturar fluctuaciones estacionales complejas o eventos atípicos que influyen de forma puntual en la demanda hotelera.

Asimismo, el enfoque adoptado por el modelo se restringe a los días en los que la ocupación supera un umbral predefinido (64 habitaciones), lo que deja fuera de análisis aquellos días con niveles de demanda menores. Esta segmentación limita la comprensión integral del comportamiento de la demanda a lo largo del año y reduce la capacidad del modelo para ajustar dinámicamente los precios o el propio umbral de ocupación, actualmente fijado en un 50%.

A futuro, se recomienda avanzar hacia un modelo más flexible y completo, que permita:

- Estimar la demanda en todos los días del año, no solo aquellos por encima de un umbral.
- Recalibrar automáticamente el umbral de ocupación, en función de la temporada, el comportamiento histórico o la anticipación de la reserva.
- Optimizar no solo los niveles de precios, sino también las decisiones sobre cuántas habitaciones ofrecer en cada tramo tarifario.

7.3.2. Factores críticos no incorporados al modelo

Además de las limitaciones estructurales abordadas previamente, existen tres grandes dimensiones que no fueron adecuadamente contempladas en el diseño y aplicación del modelo: los ingresos complementarios por huésped, la ausencia de variables externas relevantes y las limitaciones inherentes a la calidad y disponibilidad de los datos.

A continuación, se desarrollan cada una de ellas:

- **Exclusión de ingresos adicionales por servicios complementarios:** Uno de los principales sesgos del modelo actual de revenue management radica en su enfoque exclusivo en los ingresos por alojamiento, sin considerar el efecto indirecto de la ocupación sobre otras unidades de negocio del hotel. Tal como se analizó en la Sección 2.2, servicios como Alimentos y Bebidas, Health & Spa, y otros consumos complementarios generan en promedio USD 220 adicionales por huésped. Este monto, en muchos casos, puede superar al ingreso generado por la habitación misma. Por lo tanto, mantener precios más accesibles para incrementar la ocupación podría representar una estrategia más rentable en términos globales. Incorporar esta visión integral del huésped dentro del modelo permitiría tomar decisiones de pricing más alineadas con la rentabilidad total del cliente y no solamente con el margen unitario por habitación. Esta mejora constituye una línea de trabajo prioritaria para futuras versiones del modelo.
- **Ausencia de variables externas relevantes:** El modelo actual no incluye información externa relevante, lo que puede inducir a sesgos por variables omitidas. Entre los factores exógenos que deberían incorporarse para aumentar la capacidad explicativa del modelo, se destacan:
 - **Precios de la competencia:** fundamentales para modelar la elasticidad relativa de la demanda.
 - **Campañas de marketing e inversión publicitaria:** determinan la visibilidad del hotel y su capacidad de atracción.

- **Condiciones macroeconómicas** inflación, actividad económica, tipo de cambio e incertidumbre política, influyendo en el poder adquisitivo de los clientes.
- **Eventos especiales en la zona:** conciertos, congresos, ferias o eventos deportivos que generan incrementos puntuales y predecibles en la demanda.

La incorporación de estas variables permitiría capturar mejor los determinantes reales del comportamiento del consumidor y reducir errores de predicción derivados de shocks externos no modelizados.

- **Limitaciones de los datos disponibles:** En relación con la calidad y disponibilidad de los datos, el modelo enfrentó importantes restricciones que condicionaron su capacidad predictiva. Por un lado, se detectaron errores operativos en la base histórica, particularmente vinculados a procesos de auditoría que afectaron variables clave como la cantidad de noches (**nights**). Estos errores obligaron a descartar registros inconsistentes, reduciendo así el tamaño efectivo de la muestra utilizada para el entrenamiento del modelo.

Por otro lado, si bien se contó con datos desde el año 2019 hasta 2024, la utilidad de esta serie se vio comprometida por dos factores: el cierre del hotel durante 2020 y el bajo nivel de actividad observado en 2021 debido a la recuperación postpandemia. Esto impidió capturar patrones estacionales consistentes y tendencias de largo plazo.

En consecuencia, la combinación de una base de datos limitada, con errores y con escasa profundidad histórica, restringió el desarrollo de modelos más complejos y redujo su capacidad para generalizar correctamente frente a condiciones nuevas o cambiantes del mercado.

Apéndice A. Lista de variables para cada dataset

A continuación, se detallan las variables de cada dataset:

- **Estadías de clientes**
 - `name_id`
 - `first_name`
 - `last_name`
 - `arrival`
 - `departure`
 - `nights`
 - `room_type`
 - `rate_amount`
 - `currency`
 - `total_amount`
 - `adults`
 - `childs`

También, se puede observar un par de ejemplos correspondientes a la base de datos en cuestión:

name_id	first_name	last_name	arrival	departure	nights	room_ type	rate_a mount	curren cy	total_a mount	adults	childs
11169166	Sohel	Afza	2019-05-06	2019-05-08	2	EXEC	172	USD	345	1	0
11342930	Neil	Pereira	2019-05-05	2019-05-08	3	KINS	139	USD	416	1	0
11343437	Chiaki	Kinjo	2019-05-06	2019-05-08	2	TGNS	139	USD	277	1	0

- **Tipo de cambio**

- **fecha**
- **tc**

- **Feriados**

- **fecha**
- **nombre_feriado**
- **feriado**
- **noche_feriada**
- **tipo_finde**
- **numero_de_finde**

- **Eventos**

- **night**
- **cliente**

- **pax**
- **tipo**
- **salon**
- **realizado_por_hotel**

Apéndice B. Estadias dataset Profiling Report

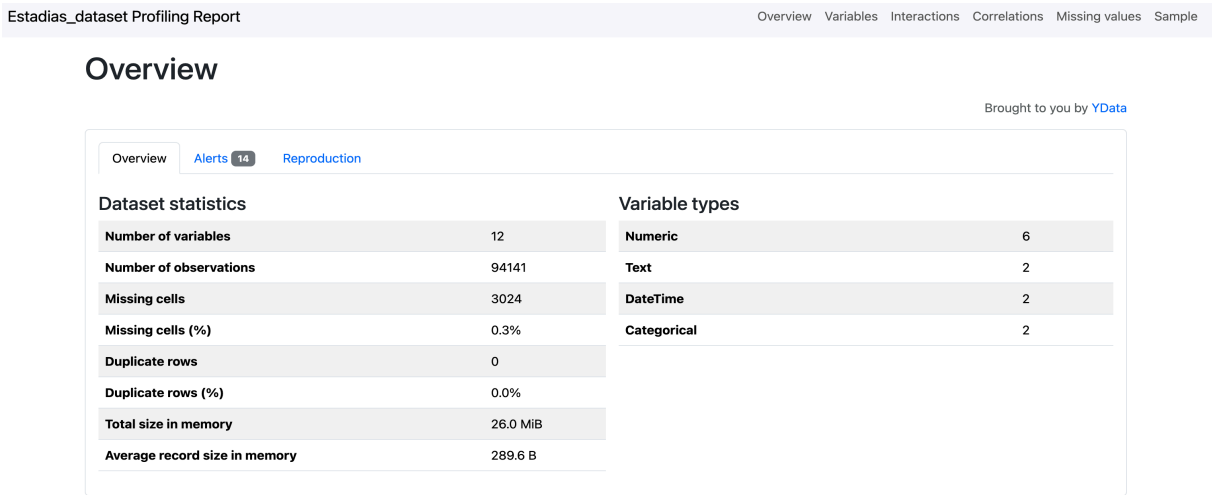


Figura B.1: Overview Profiling Report

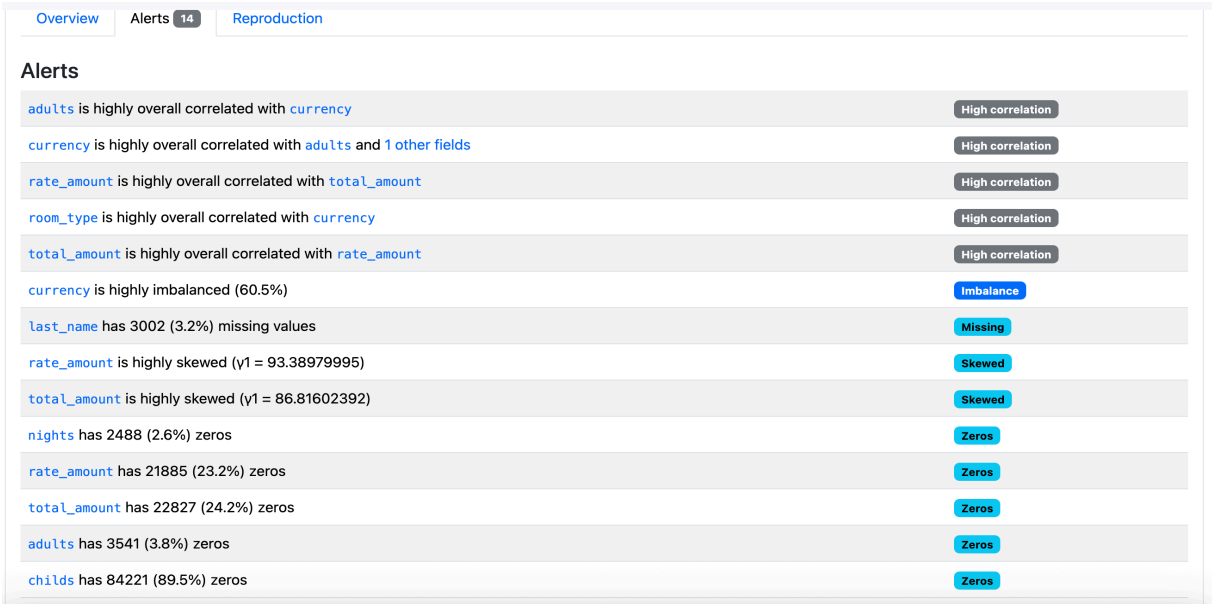


Figura B.2: Alerts Profiling Report

Variables

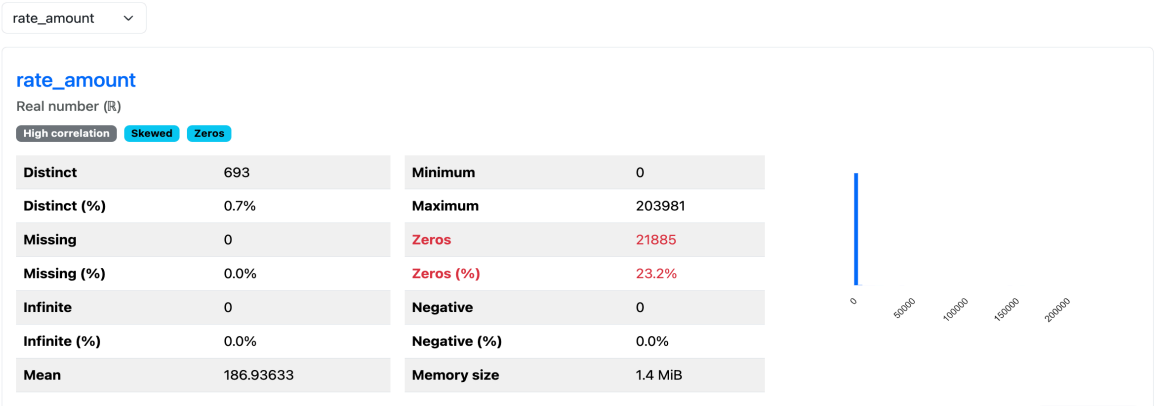


Figura B.3: **rate_amount** Profiling Report

Variables

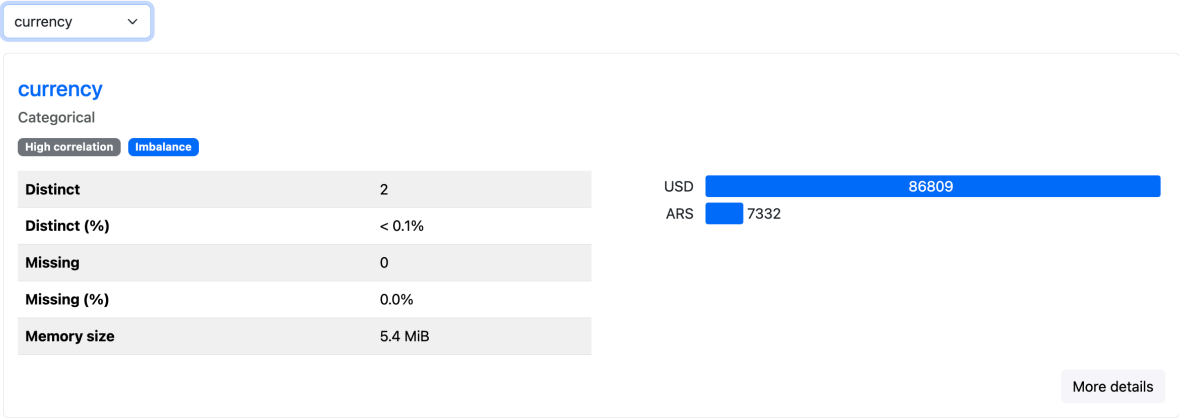


Figura B.4: **currency** Profiling Report

Correlations

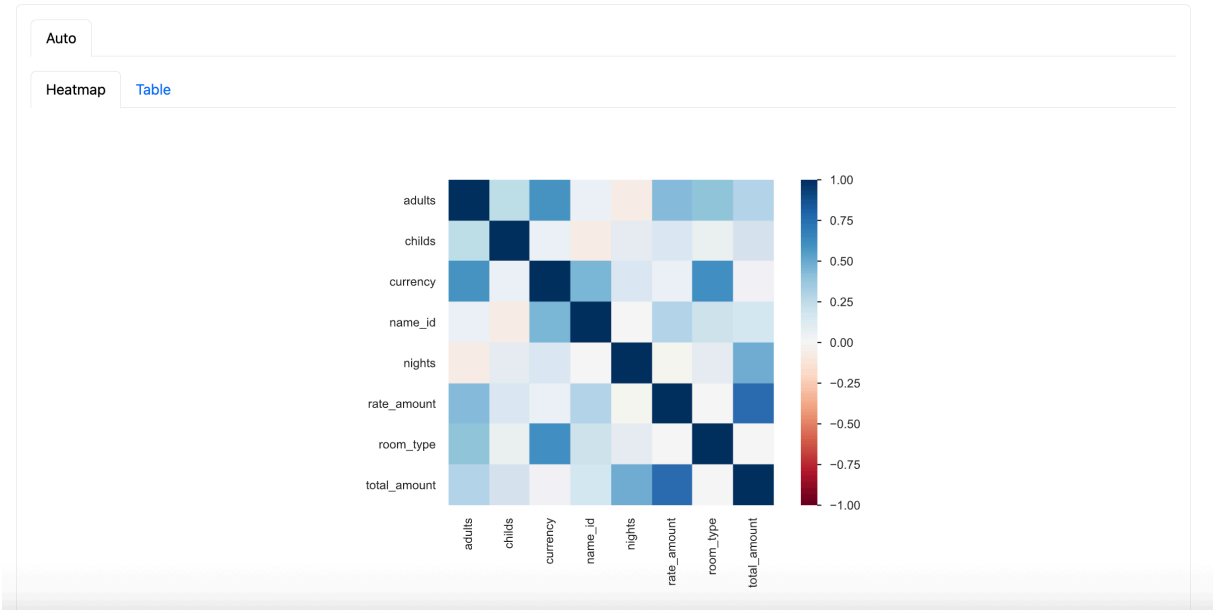


Figura B.5: Correlations Profiling Report

Apéndice C. Detalle de room_type

A continuación, se detallan todos los códigos de **room_type** y su correspondiente detalle:

- KINS: Habitación clásica con vistas al frente con cama king
- TWNS: Habitación clásica con vistas al frente con camas twin
- KGNS: Habitación de lujo con vistas al jardín con cama king
- TGNS: Habitación de lujo con vistas al jardín con cama twin
- KISK: Habitación clásica con vistas al frente con cama king
- PM: Cargos de servicios no relacionados con hospedajes
- SUIT: Habitación Suite de 2 ambientes con vista al jardín
- PT: alojamientos futuros
- EXEC: Habitación ejecutiva con vista al frente
- KGSK: Habitación de lujo con vistas al jardín con cama twin
- TWSM: Habitación clásica con vistas al frente con camas twin
- EKNG: Habitación ejecutiva con vista al jardín
- PF: Cargos de servicios no relacionados con hospedajes
- TGSM: Habitación de lujo con vistas al jardín con cama twin
- KIAC: Habitación accesible con vistas al frente con cama king
- KGAC: Habitación de lujo accesible con vistas al jardín con cama king
- PY: alojamientos futuros
- PO: no shows
- ROOM: alojamientos futuros
- PI: no shows

Apéndice D. Rate USD Profiling Report

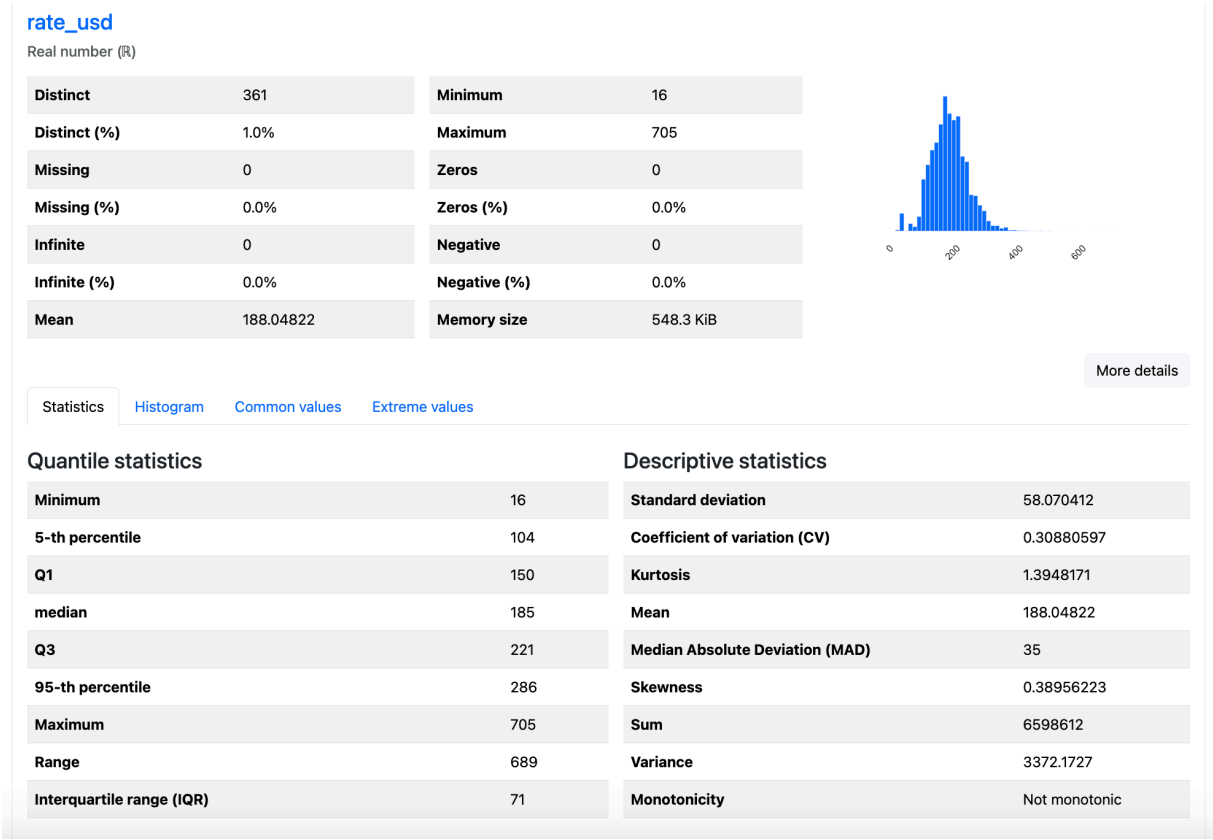


Figura D.1: rate_usd Profiling Report

Referencias

- [1] Kimes, S. E. (1989). The basics of yield management. *Cornell Hotel and Restaurant Administration Quarterly*.
- [2] Smith, B. A., Leimkuhler, J. F., & Darrow, R. L. (2009). Yield management at American Airlines. *Interfaces*, 39(2), 10-21.
- [3] Choi, T., Leung, R., & Okumura, T. (2004). A review of predictive models for hotel demand. *Journal of Hospitality & Tourism Research*, 28(1), 1-13.
- [4] Kimes, S. E., & Wirtz, J. (2003). Has revenue management become acceptable? *Journal of Service Research*, 6(2), 125–135.
- [5] Bitran, G. R., & Caldentey, R. (2003). An overview of pricing models for revenue management. *Manufacturing & Service Operations Management*, 5(3), 203-229.
- [6] Kimes, S. E. (1989). The fundamentals of revenue management in the hotel industry. *International Journal of Hospitality Management*, 8(1), 45-49.
- [7] Kimes, S. E., & Wirtz, J. (2003). Revenue management in the hotel industry. In *Handbook of Hospitality Operations and IT* (pp. 233-244). Elsevier.
- [8] Kimes, S. E. (1989). The basics of yield management. *Cornell Hotel and Restaurant Administration Quarterly*, 30(3), 14-19.
- [9] Elmaghraby, W., & Keskinocak, P. (2003). Dynamic pricing in the presence of inventory constraints. *Management Science*, 49(10), 1380-1397.
- [10] Talluri, K. T., & Van Ryzin, G. J. (2004). *The Theory and Practice of Revenue Management*. Springer.
- [11] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
<https://doi.org/10.1023/A:1010933404324>
- [12] Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R News*, 2(3), 18–22.

- [13] García, L., García-Murillo, M., & Moreno, M. N. (2018). Hotel occupancy prediction using machine learning techniques. *Procedia Computer Science*, 138, 791-798. <https://doi.org/10.1016/j.procs.2018.10.103>
- [14] Moreno-Izquierdo, L., Ramón-Rodríguez, A., & Such-Devesa, M. J. (2019). Forecasting tourist accommodation demand with machine learning models: The case of Valencia (Spain). *Tourism Management*, 75, 338–347. <https://doi.org/10.1016/j.tourman.2019.04.010>
- [15] Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1), 1–26. <https://doi.org/10.1214/aos/1176344552>
- [16] Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning* (2nd ed.). Springer.
- [17] Probst, P., Wright, M. N., & Boulesteix, A. L. (2019). Hyperparameters and tuning strategies for random forest. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 9(3), e1301. <https://doi.org/10.1002/widm.1301>
- [18] Zhang, Z., Wang, D., Xie, K. L., & Singh, A. (2017). Hotel room demand forecasting: A machine learning approach. *International Journal of Hospitality Management*, 70, 27–40. <https://doi.org/10.1016/j.ijhm.2017.01.012>
- [19] Cheng, M., & Edwards, D. (2019). Predicting hotel booking cancellations with machine learning models. *International Journal of Hospitality Management*, 77, 89–99. <https://doi.org/10.1016/j.ijhm.2018.06.015>
- [20] Köseoglu, M. A., et al. (2021). Machine learning approaches in hotel revenue management: A review and research agenda. *Tourism Management Perspectives*, 37, 100785. <https://doi.org/10.1016/j.tmp.2020.100785>
- [21] Bertsimas, D., & Shioda, R. (2013). Revenue management in a dynamic pricing environment: Predicting and optimizing demand. *Operations Research*, 61(5), 1092-1108.
- [22] Wang, Y., & Yang, X. (2006). The use of regression models in demand forecasting for the hotel industry. *International Journal of Hospitality Management*, 25(3), 429-439.

- [23] Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R news*, 2(3), 18-22.
- [24] Zhang, J., Zhao, X., & Zhang, R. (2017). Application of Random Forest in the prediction of hotel demand. *Journal of Hospitality & Tourism Research*, 41(5), 625-642.
- [25] Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice*. 2nd ed. OTexts. Disponible online: <https://otexts.com/fpp2/>
- [26] Cross, R. G., Higbie, J. A., & Cross, S. L. (2009). The impact of revenue management on the hotel industry: A 20-year retrospective. *Cornell Hospitality Quarterly*, 50(4), 465-474.