



Escuela de Administración
Escuela de Ingeniería, Ciencia y Tecnología

Maestría en Business Analytics

Optimización de Ingresos, Costos y Gastos de un operador hotelero en Bogotá, a través del fortalecimiento de su estrategia analítica.

Presentado por:

Ana María Mosquera, Linda López, Sara Martínez, & Gustavo Rodríguez

Bogotá, D.C. 25 de mayo de 2025



Escuela de Administración
Escuela de Ingeniería, Ciencia y Tecnología

Maestría en Business Analytics

Optimización de Ingresos, Costos y Gastos de un operador hotelero en Bogotá, a través del fortalecimiento de su estrategia analítica.

Presentado por:

Ana María Mosquera, Linda López, Sara Martínez, & Gustavo Rodríguez

Bajo la dirección
de: Andrés Posada

Bogotá, D.C. 25 de mayo de 2025

Tabla de contenido

Lista de figuras	7
Lista de tablas.....	10
Anexos	12
Resumen Ejecutivo.....	13
Palabras Clave.....	13
Abstract	14
Keywords	14
1. Introducción	15
2. Objetivos	18
2.1. Objetivo General	18
2.2. Objetivos Específicos	18
3. Marco Teórico	19
3.1. Analítica en la Industria Hotelera.....	19
3.1.1. Teoría de la Visualización en la Industria Hotelera	20
3.1.2. Análisis Envolvente de Datos en la Industria Hotelera.....	23
3.2. Metodología Microsoft Data Science Lifecycle.....	24
3.3. Series de Tiempo, Regresión Lineal y Random Forest.....	27
4. Metodología	29
5. Alcance del Proyecto Aplicado.....	34

5.1.	Gestión de Riesgos	35
6.	Cronograma.....	36
7.	Entendimiento del Negocio.....	37
7.1.	Mapeo de Procesos Actuales.....	38
7.2.	Diccionario de Datos.....	40
8.	Entendimiento de los Datos	43
8.1.	Análisis Componentes de Calidad de Datos	43
8.1.1.	Tabulación de Variables Presentes.....	43
8.1.2.	Cálculo y Revisión de Calidad de Datos.....	43
8.2.	Análisis de Factores de Riesgo Latentes	44
8.3.	Análisis Descriptivo de las Fuentes de Información	47
8.3.1.	Datos de Ocupación, Ingresos Mensuales.....	48
8.4.	Análisis de Estados Financieros, PyG.....	52
8.4.1.	Metodología Basada en Análisis Vertical	59
9.	Preparación de Base de Datos	61
9.1.	Datos Internos del Hotel.....	61
9.2.	Definición de Variables Exógenas	61
10.	Exploración de Modelos para Predicción de Ingresos	64
10.1.	Regresión Lineal	64
10.2.	Random Forest	64

10.3.	Modelos de Series de Tiempo	70
10.3.1.	Prophet	72
10.3.2.	ARIMA	73
10.3.3.	ARIMAX	74
10.3.4.	SARIMA	77
10.4.	Comparación de Modelos	83
11.	Modelo Definitivo de Predicción de Ingresos: SARIMAX	85
11.1.	Análisis Exploratorio Inicial	88
11.2.	Análisis de Causalidad de Granger	89
11.3.	Prueba de Estacionariedad.....	90
11.4.	Autocorrelación (ACF) y Autocorrelación Parcial (PACF).....	91
11.5.	Entrenamiento del Modelo, Ajuste del Modelo y Pruebas de Validación	92
11.6.	Puesta en Producción del Modelo.....	93
11.7.	Conclusiones del Modelo SARIMAX	93
12.	Modelo de Anomalías para Costos y Gastos.....	93
12.1.	Identificación de Modelos de Detección de Anomalías Útiles para este Caso de Estudio.....	94
12.2.	Identificación de Anomalías en Costos y Gastos Totales del Operador Hotelero....	98
12.3.	Análisis de Causalidad Anomalías Detectadas por Cada Línea de Negocio.....	99
12.4.	Integración con el Modelo Predictivo de Ingresos	100

13.	Tablero de Control	105
14.	Conclusiones y Recomendaciones	103
	Referencias Bibliográficas.....	106
	Declaración de Originalidad y Autenticidad	107
	Declaración de Exoneración de Responsabilidad.....	108
	Anexos	109
	Anexo 1. Validación de Estrategia de PE a través de LEGO Serious Play	109
	Anexo 2. Organigrama Corporativo Operador Hotelero.....	110
	Anexo 3. Entrevista Semiestructurada: Mapeo de Procesos Operador Hotelero	111

Lista de figuras

Figura 1 Ejemplo de dashboard financiero hecho con power pivot.....	21
Figura 2 Dashboard financiero elaborado con Power View	21
Figura 3 Fases del proyecto según Microsoft Data Science Lifecycle	29
Figura 4 Mapeo de procesos y reportes	39
Figura 5 Gráfico de caja y bigotes para ocupación de habitaciones	49
Figura 6 Índice de personas alojadas	50
Figura 7 Relación porcentaje de ocupación y RevPAR.....	51
Figura 8 Relación #pax alojadas vs. ocupación	51
Figura 9 Evolución mensual del margen neto.....	54
Figura 10 Distribución de ingresos por línea	54
Figura 11 Mapa de Calor Correlaciones entre variables - Regresión Lineal	65
Figura 12 Variables con alta correlación con "Ingresos Totales" – Regresión Lineal.....	66
Figura 13 Ejemplo de árboles - Random Forest.....	68
Figura 14 Predicciones vs Valores Reales - Random Forest	69
Figura 15 Serie Observada Ingresos Totales - Modelo Prophet.....	71
Figura 16 Predicción de Ingresos Totales con Prophet	71

Figura 17 Residuos Modelo Prophet.....	73
Figura 18 Ingresos base de entrenamiento – ARIMA.....	74
Figura 19 Autocorrelación – ARIMA.....	75
Figura 20 Prueba autocorrelación parcial – ARIMA	76
Figura 21 Predicción ingresos – ARIMA.....	76
Figura 22 Correlación de variables con los ingresos totales	78
Figura 23 Causalidad de granger.....	79
Figura 24 Predicción de ingresos – ARIMAX.....	79
Figura 25 Predicción ingresos – SARIMA	82
Figura 26 Estacionalidad promedio Mensual Ingresos	88
Figura 27 Correlación de variables frente al ingreso	89
Figura 28 Causalidad de granger.....	90
Figura 29 Prueba ACF.....	91
Figura 30 Prueba PACF	89
Figura 31 Pronóstico modelo SARIMAX.....	90
Figura 32 Resultado estadísticos de evaluación modelo SARIMAX	93
Figura 33 Resultado Modelo de anomalías Costos Totales	98

Figura 34 Resultado Modelo de anomalías Gastos Totales98

Figura 35 Resultado Modelo de anomalías líneas de negocio100

Lista de tablas

Tabla 1. Cronograma fases del proyecto.....	36
Tabla 2 Evaluación de componentes calidad de datos EEFF Operador Hotelero.....	46
Tala 3 Evaluación de componentes calidad de datos Estadísticas KPI.....	47
Tabla 4 Estadísticos descriptivos Ingresos Mensuales OCC.....	48
Tabla 5 Ingresos Totales	55
Tabla 6 Estadísticos descriptivos total ingresos	55
Tabla 7 Total ingresos por trimestre	55
Tabla 8 Gastos de administración.....	56
Tabla 9 Estadísticos Gastos de administración.....	56
Tabla 10 Variación gastos de administración.....	57
Tabla 11 Gastos de Ventas	57
Tabla 12 Estadísticos Gastos de Ventas	57
Tabla 13 Variación Gastos de Ventas	57
Tabla 14 Gastos Servicios Públicos	58
Tabla 15 Estadísticos Gastos Servicios Públicos	58

Tabla 16 Variación Gastos Servicios Públicos.....	58
Tabla 17 Gastos Mantenimiento.....	59
Tabla 18 Estadísticos Gastos Mantenimiento.....	59
Tabla 19 Variación Gastos de Mantenimiento	59
Tabla 20 Análisis Vertical.....	60
Tabla 21 Resultados Regresión lineal	66
Tabla 22 Comparativo de desempeño de modelos predictivos de ingresos	84

Anexos

Anexo 1. Validación de estrategia de PE a través de LEGO Serious

Play Anexo 2. Organigrama Corporativo Hotel

Anexo 3. Entrevista semiestructurada Finance Controller

Resumen ejecutivo

Este proyecto de grado tiene como propósito diseñar una herramienta que respalde la toma de decisiones financieras de un operador hotelero, mediante el análisis de sus estados financieros históricos (2021-2024), indicadores del sector y variables exógenas relevantes. Para ello, se realizó un análisis descriptivo detallado sus cifras, complementado con indicadores clave de la industria, lo que permitió entender la evolución de los ingresos, costos y gastos. Luego, se integraron variables externas en un modelo de predicción de ingresos basado en técnicas de series de tiempo SARIMAX, generando proyecciones mensuales alineadas con la dinámica del entorno y de la operación interna. Además, se proyectaron los costos y gastos a partir de un análisis vertical de los estados financieros, ofreciendo una visión integral de la rentabilidad esperada.

De forma complementaria, se desarrolló un modelo de detección de anomalías para identificar desviaciones significativas en los principales rubros de costos y gastos, lo cual facilita una gestión proactiva del margen del negocio. Finalmente, se diseñó un tablero de control interactivo que consolida el análisis financiero, los indicadores clave, las variables exógenas y las proyecciones, incluyendo alertas de anomalías. Esta herramienta busca proporcionar al operador hotelero una solución ágil y visual para el monitoreo constante y la toma de decisiones estratégicas basadas en datos.

Palabras Clave

Industria hotelera, analítica predictiva, costos, gastos, SARIMAX, Tablero de visualización.

Abstract

This thesis project aims to design a decision-support tool for a hotel operator by analyzing historical financial statements (2021–2024), industry performance indicators, and relevant external variables. A detailed analysis of the financial statements was conducted, complemented by key industry indicators to understand the behavior of revenues, costs, and expenses. External variables were then integrated into a revenue forecasting model using SARIMAX time series techniques, generating monthly projections aligned with both market dynamics and internal operations. Additionally, costs and expenses were projected through vertical analysis of the financial statements, offering a comprehensive view of expected medium-term profitability.

As a complementary component, an anomaly detection model was developed to identify significant deviations in key cost and expense lines, enabling proactive margin management. Finally, an interactive dashboard was designed to consolidate financial analysis, key performance indicators, external variables, and forecasts, including anomaly alerts. This tool provides the hotel operator with an agile and visual solution for continuous monitoring and data-driven strategic decision-making.

Keywords

Hotel, predictive analytics, costs, expenses, SARIMAX, BI dashboard.

1. Introducción

La industria hotelera ha estado recuperándose durante los últimos años, luego del impacto sufrido en 2020 debido a los efectos de la pandemia de COVID – 19. Según datos de la Organización Mundial del Turismo (OMT), que mide el movimiento de turistas alrededor del mundo, se reportó una disminución de turistas en América del 67% en el año 2020, durante la pandemia, y desde entonces ha seguido una tendencia a la baja; aunque en el año 2023 la disminución es solo del 9% frente al año inmediatamente anterior, se mantiene una disminución de turistas en la región (OMT, 2022).

En Bogotá, la industria durante las últimas décadas ha presentado un crecimiento significativo consolidando a la ciudad como uno de los principales destinos para el turismo y los negocios (Banco de la República, 2023). En medio de este desarrollo ha traído una alta competitividad en el mercado hotelero donde no solo se busca cada día brindar experiencias únicas a los clientes nacionales e internacionales, sino también desafíos financieros como optimizar en el mayor alcance posible sus operaciones con el fin de mantener la rentabilidad del negocio. En este contexto resulta crucial para la estrategia de cualquier empresa que haga parte de la industria del turismo tener una clara identificación de costos, gastos e ingresos para llevar a cabo una estrategia sólida que le permita no sólo recuperarse sino crecer y fortalecerse enfrentando variaciones estacionales en la demanda como fluctuaciones de costos en la operación, aspectos que afectan directamente su margen de ganancia. En un entorno altamente caracterizado por la incertidumbre económica, la capacidad de prever y controlar los datos es fundamental. La industria hotelera, aía, se apoya en la invaluable experiencia y profundo conocimiento del sector para la toma de decisiones.

Sin embargo, al integrar un manejo detallado y sistemático de los datos, es posible robustecer significativamente las proyecciones. Esta aproximación no solo permite una visualización más precisa y sistemática de los costos, sino que también facilita la identificación de estacionalidades clave. Al reconocer estos patrones, las empresas hoteleras pueden optimizar sus niveles de gastos y costos de operación, transformando así las predicciones en decisiones estratégicas basadas en información contundente y adaptable a las realidades cambiantes del mercado.

Una de estas empresas clave en la industria del turismo son los hoteles, y por ende esta necesidad de fortalecerse y crecer está de manera clara identificada actualmente por parte del operador hotelero en el que se desarrolla el presente trabajo. Este operador hotelero cuenta con más de 10 años de experiencia y está presente en Colombia y Ecuador. En el marco del desarrollo de la maestría de Business Analytics, el operador hotelero busca en los estudiantes de la maestría un apoyo para crear una estrategia de analítica de datos que le permita predecir y visualizar los ingresos, costos y gastos de los hoteles en tiempo real, para poder así no solo tomar acciones correctivas, sino también tener acciones preventivas que soporten una estrategia organizacional más informada.

En el presente trabajo se buscará atender esta necesidad del operador hotelero a través de la exploración de sus datos financieros e índices de ocupación, con el fin de realizar un análisis descriptivo de estos datos históricos y actuales. Este análisis descriptivo será la base para determinar no solo el estado actual, sino también cuáles serán los indicadores y métricas claves que se deben tomar en cuenta para generar el modelo predictivo que se busca.

Según Mayenberger (2017) la base de la planificación estratégica en la industria

hotelera se encuentra en utilizar un sistema de información para el control operacional y el análisis financiero, haciendo uso de técnicas informáticas que permitan realizar análisis de los datos de manera ágil. Por esto el objetivo final de este proyecto, será poder integrar tanto el análisis descriptivo como el predictivo en un tablero de visualización que permita guiar las decisiones estratégicas de los hoteles. Además de poder predecir el comportamiento de costos y gastos, también se buscará identificar anomalías en estos, a través de un modelo de detección y control que permita ser más eficientes financieramente, para que así el operador hotelero tenga la capacidad de estructurar planes de contingencia que les permitan no ver afectados sus márgenes de rentabilidad una vez se materialicen dichas anomalías.

La implementación de estos modelos beneficiará al operador hotelero en términos de previsión y control de gastos, al reducir la incertidumbre en la gestión de costos, se espera que se pueda adoptar una postura más proactiva frente a la variabilidad económica y se mantenga una estructura de costos adaptada a las necesidades del mercado, promoviendo así la sostenibilidad y competitividad del hotel.

2. Objetivos

2.1. Objetivo General

Fortalecer la estrategia analítica de un operador hotelero en Bogotá, para optimizar la estructura de ingresos, costos y gastos, y facilitar así la toma de decisiones.

2.2. Objetivos Específicos

- Determinar el estado de los ingresos, costos y gastos de los hoteles para el periodo comprendido entre enero de 2021 y diciembre de 2024, a través de un análisis descriptivo de los estados financieros y los índices de ocupación.
- Identificar oportunidades de mejora en la arquitectura analítica del hotel, a partir de su nivel de madurez en el uso de analítica dentro de sus procesos.
- Proyectar la estructura de costos y gastos del hotel, a partir de un modelo predictivo aplicado a los ingresos del hotel, el cual permita apoyar el control presupuestal.
- Identificar comportamientos inusuales según las estacionalidades del hotel que impacten de manera significativa su estructura de costos y gastos, a través de un modelo de detección de anomalías.
- Respalda la toma de decisiones del hotel en términos de ingresos, costos y gastos, creando un tablero de visualización que refleje el estado actual, las

proyecciones y los indicadores.

3. Marco Teórico

3.1. Analítica En La Industria Hotelera

El sector hotelero en Colombia se ha caracterizado por un notable dinamismo y crecimiento, impulsado por factores como el aumento de la inversión extranjera, el desarrollo de nuevos competidores y las reformas tributarias implementadas para atraer inversionistas. Sin embargo, la entrada de nuevas cadenas internacionales ha generado turbulencia empresarial, un fenómeno descrito como un entorno de cambios rápidos e impredecibles que genera incertidumbre, según estudios previos, se origina por el crecimiento económico, avances tecnológicos y cambios regulatorios. En este sentido, Ansoff (1979) destacó que la turbulencia se convierte en un desafío para los gerentes, quienes deben gestionar sorpresas y discontinuidades en el entorno. Esta turbulencia se traduce en un dinamismo constante dentro del sector, como lo evidencian los cambios en la demanda, los modelos de negocio y las fluctuaciones en las tarifas hoteleras.

Además, se enfrentan a desafíos relacionados con la incertidumbre económica y política, lo que exige un análisis constante del entorno para ajustar las estrategias de manera efectiva. La gestión de riesgos se convierte en un aspecto clave, ya que las empresas deben tomar decisiones que no sólo respondan a las condiciones actuales del mercado, sino que también anticipen futuros cambios (Ansoff, 1979).

Con base en lo anterior, y para enfrentar la incertidumbre económica mencionada, además de la relevancia del turismo en la industria colombiana, Pardo & Navarro (2020) mencionan la importancia para la industria hotelera de estar siempre a la vanguardia en el

uso de nuevas tecnologías como lo es análisis de datos ya que este puede traer consigo distintas ventajas competitivas. Una de las ventajas competitivas resaltada por los autores, es el lograr establecer criterios de control contable dentro de la gestión empresarial, a través de informes, automatización e integración de datos, investigación de mercado, y análisis comparativo, entre otros.

Se exploraron algunos casos de uso de analítica en empresas del sector hotelero que serán revisados a continuación.

3.1.1. Teoría De La Visualización En La Industria Hotelera

La gestión financiera en la industria hotelera desempeña un papel crucial en la toma de decisiones estratégicas y operativas. En este contexto, el uso de herramientas informáticas, como los tableros de control, ha demostrado ser esencial para el monitoreo efectivo del rendimiento financiero. Según un estudio presentado en la 14ª Conferencia Ibérica sobre Sistemas y Tecnologías de la Información (CISTI, 2019), el desarrollo de estos tableros facilita la visualización y análisis de Indicadores Clave de Desempeño (KPIs), permitiendo a los gerentes evaluar el cumplimiento de metas financieras y optimizar la gestión de recursos, la capacidad de analizar datos históricos y proyectar tendencias futuras permite a los gerentes financieros tomar decisiones informadas. Este proceso se ve fortalecido por la implementación de tableros de control interactivos, que consolidan información crítica y presentan métricas clave de manera comprensible y accesible.

El estudio previamente mencionado (CISTI, 2019) propone dos tipos principales

de tableros para la gestión financiera hotelera:

1. **Tablero Power Pivot:** Incluye 10 KPIs distribuidos en ocho gráficos, proporcionando comparaciones mensuales y anuales. Permite una navegación flexible a través de diferentes periodos, facilitando el análisis detallado del desempeño financiero.

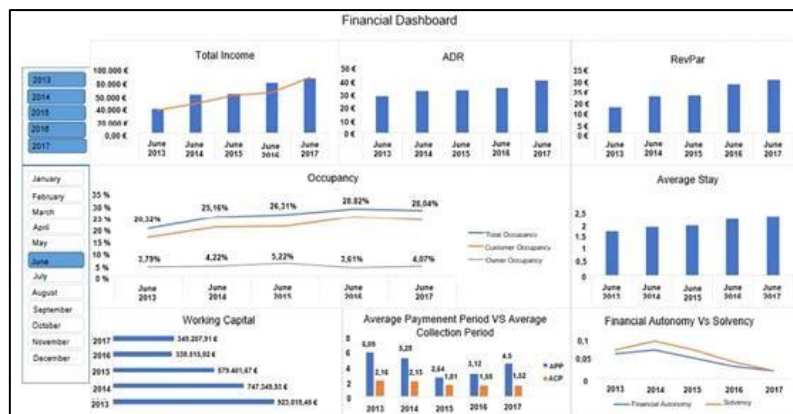


Figura 1 Ejemplo de dashboard financiero hecho con power pivot

Fuente: (CISTI, 2019)

2. **Tablero Power View:** Se enfoca en nueve KPIs representados en tres tipos de gráficos. Este tablero consolida métricas como la Tarifa Promedio Diaria (ADR), los Ingresos por Habitación Disponible (RevPAR) y las tasas de ocupación, proporcionando una visión anual integral.



Figura 2 Dashboard financiero elaborado con Power View

Fuente: (CISTI, 2019)

Ambos tableros son diseñados para ofrecer una experiencia de usuario intuitiva, con funcionalidades interactivas que mejoran la comprensión de los datos y respaldan la toma de decisiones estratégicas.

Es importante enfatizar en la necesidad de seleccionar indicadores financieros y no financieros bajo criterios específicos. Entre estos se destacan la selectividad, representatividad, simplicidad, costo reducido, estabilidad, enfoque experimental, comparación externa, accesibilidad y mejora continua. Los KPIs seleccionados deben alinearse con los objetivos estratégicos de la organización, proporcionando información valiosa para la gestión financiera y operativa.

Cada KPI incluye componentes clave como:

Objetivo: Define las metas estratégicas a alcanzar.

Fórmula de cálculo: Permite medir cuantitativamente el rendimiento.

Resultado y meta: Evalúan el progreso hacia los objetivos.

Desviación y desempeño: Analizan las diferencias entre las metas y los resultados reales.

Evaluación cualitativa: Proporciona un contexto adicional sobre el rendimiento observado.

En conclusión, los tableros de control contribuyen significativamente a la eficiencia operativa y a la mejora continua en la industria hotelera. Al proporcionar una plataforma

centralizada para el análisis de KPIs, permiten a los hoteles optimizar sus recursos y alcanzar sus objetivos financieros, destacándose como una solución innovadora y práctica para la gestión financiera en el sector.

3.1.2. *Análisis Envolvente De Datos En La Industria Hotelera*

Carrillo (2017), realizó un estudio pionero en Bucaramanga, Colombia, aplicando el Análisis Envolvente de Datos (DEA) para evaluar la eficiencia del sector hotelero local. En un contexto de creciente competencia, impulsado por políticas gubernamentales favorables y proyectos de infraestructura, este análisis resulta crucial para identificar oportunidades de mejora en la gestión hotelera. El DEA, una herramienta analítica robusta, permitió comparar la eficiencia de los hoteles al relacionar sus insumos (como personal y energía) con sus productos (ingresos y ocupación).

El DEA con su metodología no paramétrica de programación lineal multiobjetivo ha ganado considerable relevancia en los últimos años como herramienta para evaluar la eficiencia relativa de un conjunto de unidades de decisión (DMUs). A diferencia de los métodos estadísticos tradicionales, el DEA no requiere supuestos previos sobre la forma funcional de la relación entre los insumos y los productos, lo que lo convierte en una herramienta versátil y flexible para analizar una amplia variedad de situaciones (Carrillo, 2017).

El principio fundamental del DEA consiste en construir una frontera eficiente que representa el mejor desempeño posible para las DMUs evaluadas, considerando los insumos utilizados y los productos obtenidos. Esta frontera se construye a partir de las propias DMUs, sin necesidad de información externa. Las DMUs que se encuentran en esta

frontera se consideran eficientes, ya que están utilizando sus recursos de manera óptima. Aquellas que se encuentran por debajo de la frontera se consideran ineficientes, lo que indica que podrían obtener los mismos resultados con menos recursos o mayores resultados con los mismos recursos.

Sin embargo, también tiene limitaciones como son: sensibilización a datos atípicos, no proporciona información casual, y requiere un número suficiente de DMUs para obtener resultados confiables. Los resultados de este estudio subrayan la importancia del DEA como un instrumento para establecer benchmarks y diseñar estrategias de mejora en la gestión hotelera.

3.2. Metodología Microsoft Data Science Lifecycle

El desarrollo del presente proyecto de grado se llevó a cabo siguiendo un enfoque estructurado basado en las mejores prácticas de la ciencia de datos moderna (Rautenstrauch, 2024), adaptadas a la realidad del operador hotelero. La metodología empleada comprende las siguientes etapas:

Entendimiento Del negocio

Como punto de partida, se realizó una caracterización detallada del modelo operativo y financiero del operador hotelero, con el fin de comprender las dinámicas de la industria, sus principales líneas de ingresos, costos y gastos, así como los factores que impactan su rentabilidad. En esta etapa se definió el problema central a abordar: la necesidad de contar con una herramienta de apoyo a la toma de decisiones financieras que permita proyectar los ingresos, costos y gastos de manera precisa y detectar posibles

desviaciones operativas de forma proactiva. Esta definición se alineó con el objetivo del operador hotelero, orientados al fortalecimiento del control financiero y la optimización del margen del negocio.

Preparación y entendimiento de los datos

En esta etapa se realizó la recolección, limpieza y análisis exploratorio de los datos históricos proporcionados por el operador, correspondientes a los estados financieros mensuales y los KPIs entre enero de 2021 y mayo de 2024. Asimismo, se recopilaron variables exógenas del entorno turístico, macroeconómico y del sector hotelero que pudieran tener impacto en el desempeño financiero. La exploración de datos permitió identificar patrones, tendencias y posibles relaciones entre las variables internas y externas, facilitando la selección de aquellas con mayor potencial explicativo para la ejecución de los modelos. La robustez del comportamiento financiero en el sector hotelero se sustenta en gran medida en la relación intrínseca y directa entre los ingresos generados y los costos y gastos operacionales en los que se incurre mensualmente. Esta interdependencia es un pilar fundamental para comprender la estructura de rentabilidad del operador hotelero, consecuentemente, para la toma de decisiones estratégicas. Analizamos los estados financieros dispuestos por el operador en donde encontraremos una estabilidad notable en la proporción de los costos y gastos respecto a los ingresos totales a lo largo del tiempo. Esta coherencia subraya la aplicabilidad de metodologías como el análisis vertical, una herramienta financiera clave que permite desglosar la composición de los estados financieros, expresando cada partida como un porcentaje de un total base, en este caso, los ingresos totales.

Esta aproximación es particularmente relevante en la industria hotelera debido a factores inherentes a su operación. Por un lado, existen costos variables directamente vinculados a la ocupación y los ingresos, como los suministros de alimentos y bebidas, los servicios de limpieza y el personal temporal. Un incremento en la ocupación se traduce naturalmente en un aumento de estos costos, estableciendo un vínculo causal directo. Por otro lado, los costos fijos o semifijos, como el alquiler, los salarios administrativos y ciertos servicios públicos, si bien son menos volátiles, suelen mantener una proporción constante con los ingresos totales, contribuyendo a la estabilidad general de la relación gasto-ingreso. Finalmente, las prácticas internas de gestión de ingresos y presupuesto de los hoteles ya sean explícitas o implícitas, suelen apuntar a mantener ciertas relaciones costo-ingreso, reforzando así esta proporcionalidad. La comprensión y cuantificación de estas relaciones son esenciales para desarrollar modelos predictivos precisos y para optimizar la gestión financiera del hotel, como se profundizará en secciones posteriores.

Desarrollo de modelos

Con base en los datos depurados y comprendidos, se desarrolló un modelo predictivo de ingresos empleando técnicas de series de tiempo con enfoque en SARIMAX, integrando las variables exógenas seleccionadas previamente. Posteriormente, para complementar el modelo ya mencionado y facilitar una proyección integral de los estados financieros, se aplicó un análisis vertical detallado de los Estados de Resultado. Este análisis permitió entender la estructura relativa de costos y gastos frente a los ingresos, y demostrar su estabilidad a lo largo del tiempo. Adicionalmente, se implementó un modelo de detección de anomalías, orientado a identificar posibles desviaciones significativas en las

principales líneas de costos y gastos, permitiendo una gestión proactiva cuidado en margen del negocio.

Implementación

Como resultado del proceso, se diseñó y desplegó un tablero de control interactivo que integra las proyecciones financieras, el modelo de anomalías, los KPI operativos y las principales variables exógenas, brindando al operador una herramienta ágil y visualmente amigable para el seguimiento y la toma de decisiones estratégicas.

Aplicar la metodología Microsoft Data Science Lifecycle permitió estructurar el proceso analítico, desarrollando modelos que optimizarán la gestión financiera, minimizando riesgos y maximizando la eficiencia operativa en un entorno competitivo.

3.3. Series De Tiempo, Regresión Lineal y Random Forest

Como se ha mencionado hasta acá lo que se busca es mejorar el control presupuestal del operador hotelero al hacer una predicción de los ingresos para así con base en los ingresos pronosticados poder presupuestar los costos y gastos del hotel, ya que estos son determinados por un porcentaje de los ingresos para alcanzar el margen esperado. Así las cosas, para poder predecir dichos ingresos existen diferentes técnicas a nivel estadístico, las series de tiempo, la regresión lineal y el algoritmo random forest, que se explicaran a continuación.

El análisis de series de tiempo es fundamental en diversas disciplinas, ya que permite modelar y predecir fenómenos que evolucionan a lo largo del tiempo. Una serie temporal se define como una secuencia de observaciones ordenadas cronológicamente, y

su estudio implica comprender patrones como tendencias, estacionalidades y ciclos. Estos patrones son esenciales para realizar pronósticos precisos y tomar decisiones informadas en diferentes campos (Jiménez, 2012). En el caso de este proyecto, el ciclo estacional del hotel es de 12 meses, y a través del uso de series de tiempo, se busca identificar estacionalidades (picos y valles constantes) dentro del periodo comprendido entre enero de 2021 y diciembre de 2024.

Por otro lado, está la técnica de regresión, esta es una técnica estadística que busca modelar la relación entre una variable dependiente y una o más variables independientes. En este contexto la regresión se utiliza para capturar la dependencia temporal y prever valores futuros.

Sin embargo, los modelos de regresión tradicionales pueden enfrentar limitaciones cuando se trata de datos con estructuras complejas o no lineales (Roldán, 2020). Esto permite prever que la regresión no sea la mejor técnica para el presente proyecto, sin embargo, se hará uso de esta técnica para acompañar la exploración y el entendimiento de los datos.

Finalmente, tenemos el algoritmo Random Forest. Este algoritmo, como menciona Breiman (2001), es un método de ensamblado que construye múltiples árboles de decisión y combina sus resultados. Random Forest es especialmente útil en problemas de regresión y clasificación, y ha demostrado ser eficaz en el análisis de series de tiempo, donde puede capturar relaciones no lineales y manejar grandes volúmenes de datos (Ortega & Antón, 2021; Funcas, 2021).

La integración de técnicas tradicionales como la regresión con métodos modernos como Random Forest ofrece un enfoque robusto para el análisis y la predicción de series

temporales. Esta combinación permite aprovechar la interpretabilidad de los modelos estadísticos clásicos y la capacidad de los algoritmos de aprendizaje automático para manejar datos complejos y de alta dimensión.

4. Metodología

Dentro del ejercicio académico para poder aclarar la metodología del proyecto de grado se realizó un ejercicio bajo el concepto de la actividad Lego Serious Play (Ver Anexo 1) para cumplir los objetivos planteados, se desarrollará una metodología estructurada en varias fases, como se observa en la Figura 3, cada una enfocada en preparar y analizar los datos, así como en la implementación y evaluación de modelos predictivos. A continuación, se describen cada una de las etapas:

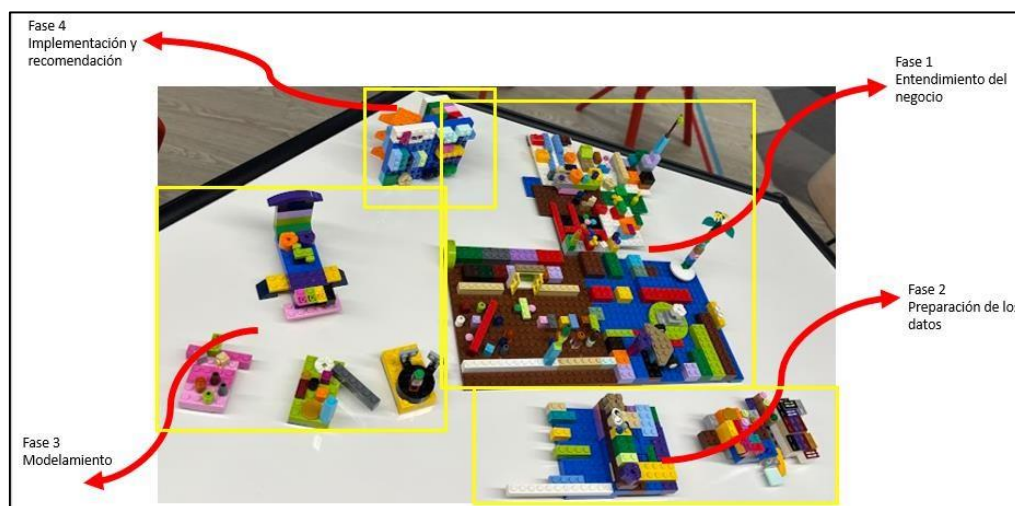


Figura 3 Fases del proyecto según Microsoft Data Science Lifecycle

Fuente: elaboración propia.

- 1. Entendimiento De Negocio:** La primera fase consistirá en un análisis exhaustivo de los datos proporcionados por el operador hotelero. Se comenzará con una revisión detallada del contexto del negocio, evaluando cómo los datos reflejan las dinámicas operativas de los hoteles y considerando el impacto de

variables exógenas, como fluctuaciones estacionales en la demanda y variaciones en los precios de insumos. Esto permitirá construir un marco conceptual en el que los datos se alineen con los objetivos de negocio, para esto se realiza una entrevista semi estructurada (ver Anexo 3) con el finance controller para detallar los procesos organizacionales que ocurren antes de obtener los datos que nos brindan.

También se generará un diccionario de datos que describa cada variable y sus características, facilitando la comprensión y la comunicación de los datos utilizados en el proyecto.

- 2. Preparación De Los Datos:** En esta etapa, se llevará a cabo un proceso de limpieza y normalización de los datos, en el que se identificarán y manejarán datos faltantes, valores atípicos y cualquier inconsistencia en la información. Además, se definirán umbrales para las variables clave, lo que permitirá detectar y corregir posibles anomalías. Una vez depurados, los datos se transformarán para estandarizar su formato y facilitar su análisis. También se considerará la creación de nuevas variables (feature engineering) que puedan enriquecer los modelos, como tasas de ocupación, costos promedio por categoría y variables temporales que representan la estacionalidad.

Como producto de esta fase se tendrá un análisis descriptivo de los datos que se presentará al operador hotelero, para continuar alimentando y validando la comprensión y uso de los datos antes de pasar a la fase de modelado.

- 3. Desarrollo De Modelos:** Con los datos listos para su análisis, se procederá a la selección e implementación de modelos estadísticos que permitan realizar

predicciones de los costos y gastos futuros, así como anomalías en los mismos.

Entre los modelos que se considerarán se incluyen:

- a) Modelo predictivo de costos y gastos

- b) Modelo de optimización de costos y gastos

De manera introductoria, es importante mencionar una breve descripción sobre cada uno de los modelos que serán objeto de análisis en este trabajo de grado. Así:

Prophet

El modelo Prophet es una herramienta de pronóstico desarrollada para manejar series temporales con fuertes componentes estacionales y con posibles cambios abruptos en las tendencias. Este modelo se basa en una descomposición aditiva que incluye tres componentes principales: tendencia, estacionalidad y días festivos o eventos especiales. Su estructura permite capturar patrones complejos de comportamiento temporal, lo que lo hace útil en contextos donde se espera que las series tengan variaciones cíclicas regulares o interrupciones repentinas. Prophet es especialmente práctico en aplicaciones donde se requiere generar pronósticos rápidos, precisos y con intervalos de confianza, incluso con mínimos ajustes por parte del analista.

ARIMA

El modelo ARIMA (AutoRegressive Integrated Moving Average) es un modelo

estadístico utilizado para analizar y predecir series temporales univariantes. Su estructura se basa en tres elementos: el componente autorregresivo (AR), que utiliza rezagos de la propia serie; el componente de media móvil (MA), que considera los errores residuales pasados; y el componente de integración (I), que transforma la serie en estacionaria a través de diferenciación. ARIMA es eficaz en situaciones donde no existe una estacionalidad marcada y donde la serie temporal puede ser representada de forma eficiente a través de sus propias dinámicas pasadas.

ARIMAX

El modelo ARIMAX extiende la lógica del modelo ARIMA al incorporar variables

explicativas externas (exógenas), lo que permite capturar el efecto de factores externos sobre el comportamiento de la variable objetivo. Esta característica es especialmente útil cuando se conoce que el fenómeno a modelar está influenciado por variables adicionales, como indicadores económicos, precios, políticas o condiciones externas. ARIMAX es apropiado para situaciones en las que se desea mejorar el poder predictivo de la serie utilizando información complementaria, manteniendo al mismo tiempo la estructura temporal tradicional del ARIMA.

SARIMAX

El modelo SARIMAX (Seasonal AutoRegressive Integrated Moving Average

with eXogenous regressors) es una evolución del modelo ARIMAX que, además de incluir variables exógenas, incorpora componentes estacionales de manera explícita. Esta capacidad lo convierte en una herramienta potente para analizar y pronosticar series que presentan patrones recurrentes en ciclos definidos, como los mensuales o anuales. Una de sus principales ventajas es su flexibilidad para adaptarse a contextos donde los datos son limitados, ya que, al combinar múltiples fuentes de información y estructurar adecuadamente los componentes estacionales y autorregresivos, permite generar pronósticos robustos incluso con series cortas. Esto lo hace especialmente adecuado para aplicaciones como el pronóstico de ingresos mensuales de un hotel, donde puede no haber una larga serie histórica disponible.

Los modelos desarrollados serán evaluados en función de su capacidad predictiva y su ajuste a los datos. Para ello, se utilizarán métricas como el error cuadrático medio (MSE), el error absoluto medio (MAE) y el coeficiente de determinación (R^2). La eficacia de los modelos se medirá en términos de su precisión y su utilidad para predecir costos y gastos futuros, buscando un nivel de error bajo y un buen ajuste a los datos históricos. Además, se realizan pruebas con distintos parámetros con el fin de asegurar que los modelos sean robustos y generalizables a datos futuros, y lo óptimo para el funcionamiento real del operador hotelero.

- 4. Implementación y Recomendaciones:** Finalmente, los modelos que demuestren ser más efectivos serán documentados y sus resultados analizados para generar recomendaciones prácticas hacia el operador hotelero. Estas recomendaciones

están enfocadas en la optimización de costos y en la mejora de la planificación financiera, proporcionando estrategias específicas para una gestión proactiva de gastos. Además, se documentarán sugerencias para el uso continuo de estos modelos, permitiendo a la empresa ajustar las proyecciones en función de nuevas variables o cambios en el mercado.

5. Alcance Del Proyecto Aplicado

El presente proyecto tiene como objetivo principal fortalecer la capacidad de toma de decisiones del operador hotelero a través del desarrollo de un modelo analítico que permita comprender y predecir el comportamiento de los costos y gastos de sus establecimientos. Este análisis se centrará en uno de los hoteles de Bogotá, con el fin de optimizar la gestión presupuestaria y mejorar la eficiencia operativa.

A través de un análisis descriptivo de los datos históricos proporcionados por la cadena hotelera, se identificarán patrones y tendencias en los costos y gastos, así como su relación con variables como la demanda, la estacionalidad y el tipo de cliente. Con base en esta información, se desarrollará un modelo predictivo que permitirá simular diferentes escenarios y anticipar los posibles impactos de las decisiones de gestión en la estructura de costos.

Como resultado de este proyecto, se entregarán los siguientes productos:

Análisis Descriptivo Detallado: Un informe que presente de manera clara y concisa los hallazgos del análisis exploratorio de los datos, incluyendo visualizaciones que faciliten la comprensión de los resultados.

Modelo Predictivo Validado: Un modelo estadístico capaz de pronosticar los costos y gastos futuros del hotel, considerando las variables más relevantes.

Modelo De Detección y Control Validado: Un modelo estadístico capaz de identificar anomalías dentro de los costos y gastos del hotel, con el fin de

optimizar recursos.

Panel De Control Interactivo: Una herramienta visual que permita a los usuarios monitorear en tiempo real el desempeño financiero de los hoteles, comparar los resultados reales con las predicciones y evaluar el impacto de las iniciativas de reducción de costos.

Manual de usuario: Un documento que explique de manera sencilla cómo utilizar el panel de control y los modelos que se brinden, así como las limitaciones y consideraciones a tener en cuenta.

El alcance de este proyecto se centrará en el análisis de los datos históricos proporcionados por el operador hotelero. Los resultados obtenidos serán válidos dentro del contexto y las limitaciones de los datos disponibles. Es importante destacar que el modelo predictivo se basa en supuestos y patrones históricos, por lo que su precisión dependerá de la estabilidad del entorno y de la capacidad de los datos para capturar los cambios futuros.

5.1. Gestión De Riesgos

Para mitigar los riesgos asociados a la calidad de los datos y a la adopción de las herramientas analíticas por parte de los usuarios, se llevarán a cabo las siguientes acciones:

Validación de datos: Se realizará una rigurosa revisión de la calidad y consistencia de los datos antes de iniciar el análisis.

Comunicación y capacitación: Se diseñarán sesiones de capacitación para los usuarios clave, con el objetivo de familiarizarlos con las herramientas y los

resultados del proyecto.

6. Cronograma

El proyecto fue dividido en 4 fases para lograr su ejecución, que se pueden ver en la Tabla

1. Este cronograma fue estructurado de tal forma que permitiera organizar de manera secuencial y eficiente cada una de las actividades planteadas, garantizando el cumplimiento de los objetivos en los tiempos establecidos. Este cronograma contempló desde la fase de entendimiento del negocio y definición del problema, pasando por la recopilación y análisis de datos, construcción y validación de modelos, hasta el despliegue de la herramienta de visualización y la documentación final del proyecto.

Tabla 1. Cronograma fases del proyecto

	may-24	jun-24	jul-24	ago-24	sep-24	oct-24	nov-24	ene-25	feb-25	mar-25	abr-25	may-25	jun-25
Fase 1. Anteproyecto													
Entendimiento del negocio													
Definición y alcance del proyecto													
Entendimiento de los datos													
Exploración inicial de los datos													
Fase 2. Preparación de datos													
Preparación de los datos													
Recopilación de variables exógenas													
Análisis descriptivo													
Presentación de análisis descriptivo													
Fase 3. Modelado de datos													
Modelo predictivo													
Modelo de optimización													

Fase 4. Entrega y cierre														
Presentación de modelo predictivo														
Evaluación														
Entrega de manual de recomendaciones														

Fuente: Elaboración propia.

7. Entendimiento Del Negocio

Para lograr una mejor aproximación y por ende un mejor desarrollo del proyecto, es necesario comprender el negocio. El operador hotelero cuenta con 32 años de experiencia en desarrollo y comercialización de proyectos inmobiliarios y hoteleros, y con más de 10 años de operaciones hoteleras. Actualmente están ubicados en Colombia y Ecuador, y para efecto de este proyecto se tendrá en cuenta uno de los dos hoteles ubicados en Bogotá, el Wyndham Bogotá (Ac. 24 #51-40) ubicado estratégicamente cerca del Aeropuerto El Dorado y en zonas empresariales clave como Ciudad Salitre, lo que las convierte en opciones ideales para ejecutivos y eventos corporativos, en un sector competitivo donde predominan cadenas hoteleras de 3 y 4 estrellas. Su enfoque o core de negocio está orientado hacia viajeros de negocios, deportivos o de eventos empresariales. La competencia en Bogotá exige no solo que los hoteles ofrezcan tarifas atractivas, sino que también gestionen su inventario de forma eficaz y optimizando la ocupación con el fin de garantizar ingresos, incluso en períodos de menor demanda. Además, la integración de tecnologías avanzadas como dashboards de gestión hotelera, facilita la toma de decisiones estratégicas basadas en datos en tiempo real. (conversación con el hotel)

El operador hotelero se enfrenta a un entorno con baja variabilidad en sus costos y gastos operativos, un reto que limita su flexibilidad para reaccionar a cambios significativos en la demanda. Sin embargo, esta estabilidad financiera también puede ser una ventaja, ya que facilita la previsión de resultados a largo plazo. La optimización del inventario y la capacidad a través de técnicas de Revenue Management, como el análisis predictivo y el ajuste dinámico de precios, resulta crucial para maximizar ingresos y

mantener una alta ocupación. Con el uso adecuado de la analítica descriptiva y predictiva, el operador puede identificar tendencias en la ocupación y ajustar sus servicios para maximizar ingresos sin incrementar significativamente los costos.

El operador está dividido en tres vicepresidencias: vicepresidencia de operaciones y administración, vicepresidencia financiera & vicepresidencia de desarrollo (Ver Anexo 2). Los datos que serán analizados para este proyecto serán datos financieros y de ocupación que involucran a las primera dos vicepresidencias nombradas, y por consiguiente, los análisis que se realizaran en este trabajo serán soporte para las decisiones de estas dos vicepresidencias. Así, para ampliar el entendimiento del negocio se realizó una visita presencial a los hoteles, en la cual fue posible conversar con los dos vicepresidentes, y conocer mejor la necesidad de estos, como producto de estas visitas se realizó un mapeo de los procesos actuales del hotel que se detalla a continuación.

7.1. Mapeo de Procesos Actuales

El operador hotelero cuenta con tres líneas de negocio principales: la ocupación de habitaciones, la venta de alimentos y bebidas, y una tercera denominada “otros ingresos” que incluye el alquiler de espacios para eventos, entre otros. Estas líneas de negocio alimentan sistemas de información que dejan registro de los ingresos de cada línea.

Para este registro de los procesos de las distintas líneas se usan dentro del hotel dos sistemas de información de Oracle: Opera y Symphony. Por un lado, Symphony es un sistema de punto de venta, diseñado para la gestión completa de restaurantes, por lo que se usa para la línea de alimentos y bebidas, y se tiene dentro de este un ambiente para cada punto de venta (ej. “room service”, “restaurante”, etc..), con sus respectivos bloqueos de

seguridad y accesos para que la información no pueda ser manipulada y modificada. Por otro lado, el sistema Opera es un sistema de gestión de propiedades (PMS) para hoteles, y proveedores de alojamiento, que proporciona herramientas para el personal del hotel, como:

- Gestión de reservas
- Registro de entrada y salida de clientes
- Asignación de habitaciones
- Gestión del inventario de habitaciones
- Satisfacción de las necesidades de los huéspedes
- Funciones de contabilidad y facturación

Estos dos sistemas son auditados por el auditor de ingresos quien revisa que los datos en sistema correspondan a los datos facturados. Una vez auditados, son consultados por el finance controller para tomar la información de los ingresos, costos y gastos del hotel, y crear así reportes con indicadores como el índice de ocupación, el RevPAR, total revenue, entre otros que se describirán más adelante. En la figura 4 se muestra gráficamente cómo se alimentan estos reportes.

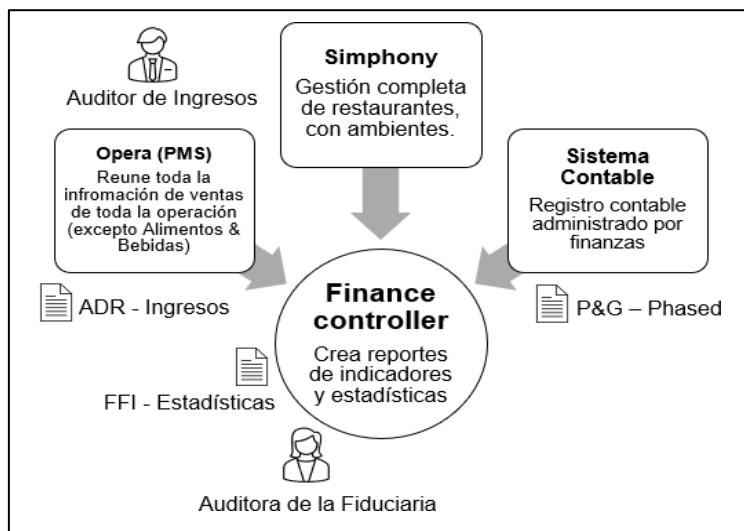


Figura 4 Mapeo de procesos y reportes

Fuente: elaboración propia.

Para cumplir con el objetivo de fortalecer la estrategia analítica del operador, se toman las fuentes de información que se observan en la figura 4 con el icono de archivo.

7.2. Diccionario de Datos

- **Habitaciones en el Hotel:** Número total de habitaciones que tiene el hotel.
- **Habitaciones disponibles:** Número de habitaciones que estuvieron sobre las que efectivamente se puede ofrecer el servicio a los clientes del hotel. La fórmula sería la siguiente:

$$\textit{Habitaciones disponibles} = \textit{Habitaciones en el hotel} - \textit{Uso Casa}$$

- **Habitaciones Ocupadas:** Número de habitaciones que fueron efectivamente ocupadas durante el mes.
- **Uso Casa:** Habitaciones utilizadas por el personal del hotel o para otros usos internos no relacionados con la venta a clientes.
- **Hab. Compl.:** Habitaciones de cortesía ofrecidas sin costo, generalmente para invitados especiales, promociones, etc.
- **No. De Personas Alojadas:** Número total de personas que se hospedaron en el hotel durante el mes.
- **Índice de Pax Alojadas:** Promedio de personas alojadas por habitación ocupada.

La fórmula sería la siguiente:

$$\text{índice de Pax Alojadas} = \frac{\text{Habitaciones_Ocupadas}}{\text{No_De_Personas_Alojadas}}$$

No_De_Personas_Alojadas

- **% de Ocup. Con 250 Hab.:** Porcentaje de ocupación calculado sobre una base fija de 250 habitaciones, independientemente del número real de habitaciones disponibles. La fórmula sería la siguiente:

$$\% \text{ de Ocup. Con 250 Hab} = \frac{\text{Habitaciones Ocupadas} \times 100}{250}$$

- **% de Ocup:** Porcentaje de ocupación calculado sobre el total de habitaciones disponibles. La fórmula sería la siguiente:

$$\% \text{ de Ocup} = \frac{\text{Habitaciones ocupadas}}{\text{Habitaciones disponibles}}$$

- **Rev_PAR:** Revenue Per Available Room o Ingreso por Habitación Disponible.

Indica el ingreso generado por cada habitación disponible. La fórmula sería la siguiente:

$$\text{Rev_PAR} = \text{Tarifa Promedio} * (\% \text{ de Ocup Con 250 Hab})$$

- **Tarifa Promedio:** Tarifa diaria promedio por habitación ocupada. La fórmula sería la siguiente:

$$\text{Tarifa Promedio} = \frac{\text{Total Ingresos}}{\text{Habitaciones Ocupadas}}$$

- **Habitaciones:** Ingresos generados específicamente por la venta de habitaciones.

- **Total Ingresos:** Suma total de todos los ingresos generados por el hotel en el mes.

La fórmula sería la siguiente:

$$\text{Total ingresos} = \Sigma \text{ Habitaciones, Room_Service, Restaurante, Bar, Eventos_Ofi_Suites, Telefonos_Internet, Lavanderia, Otros}$$

- **Occ:** Ocupación- Porcentaje de habitaciones ocupadas en el hotel. La fórmula sería la siguiente:

$$\text{Occ:} = \frac{\text{Habitaciones Ocupadas}}{\text{Habitaciones Disponibles}} * 100$$

- **ADR:** Average Daily Rate: Tarifa diaria promedio por habitación. KPI de gestión hotelera más representativo.

$$\text{ADR} = \frac{\text{Ingreso total de habitaciones}}{\text{Número habitaciones vendidas}}$$

Número habitaciones vendidas

- **GOP:** Gross Operating Profit: Beneficio Operativo Bruto, diferencia entre ingresos totales y costos operativos directos.

$$\text{GOP} = \text{Total ingresos} - \text{Total Costos}$$

- **Cump - Ppto:** Cumplimiento del Presupuesto: Comparación de los resultados reales con el presupuesto previsto.

$$\text{Cump Ppto} = \text{Resultado real} - \text{Presupuesto}$$

- **RevPAR:** Revenue Per Available Room: Ingreso por Habitación Disponible.

Número de habitaciones que se reservan en un hotel y los ingresos que generan dichas reservas.

$$\text{RevPAR} = (\text{ADR})(\text{Tasa de ocupación})$$

8. Entendimiento De Los Datos

8.1. Análisis Componentes De Calidad de Datos

Para el diagnóstico y la futura generación de modelos de predicción de costos de la operación de la cadena hotelera en Colombia, la compañía nos ha suministrado tres tipos de bases de datos: i) Estados Financieros (PyG), ii) Estadísticas de monitoreo y evolución de las principales métricas (KPIs) referentes a la operación hotelera y iii) desagregación de las diferentes fuentes generadoras de los costos tanto fijos como variables. De lo anterior, es importante destacar que las bases de datos antes mencionadas fueron compartidas para el periodo comprendido entre 2022 y 2024 en marcos de información mensual y anual.

8.1.1. Tabulación De Variables Presentes

Se tuvieron en cuenta cada una de las variables o rubros a nivel de estadísticas y EEFF presentes en las bases de datos suministradas. Para el caso de los rubros de los EEFF se realizó el mapeo de 19 variables mientras que para el caso de las estadísticas con referencia a diferentes métricas de desempeño de la cadena se identificaron 20 variables.

8.1.2. Cálculo y Revisión De Calidad De Datos

Para cada una de las variables antes mencionadas, se analizaron los siguientes componentes: Completitud, Validez, Puntualidad, Unicidad, Precisión y Consistencia. De lo anterior, es importante tener en cuenta que tal y como se puede observar en los anexos de este documento la totalidad de las variables presentó un resultado del 100% exceptuando

la variable puntualidad, para la cual determinamos un estado de No Aplica (N/A) partiendo de la base que por tratarse de información financiera que debe presentarse a los entes reguladores locales a nivel de Superintendencia, la cadena hotelera tiene unos tiempos límites de presentación que en caso de no generarse podría acarrear sanciones para la compañía.

Ahora bien, si bien es cierto que por el tipo de información suministrada el nivel de calidad de los datos podría calificarse como alto al corresponder a valores de 100%, decidimos pasar a analizar cuáles serían los potenciales factores de riesgo latente que podrían generar que hacia adelante fallara o se disminuyera la alta calidad de los datos suministrados por la cadena.

8.2. Análisis De Factores De Riesgo Latentes

a) **Completitud:** a nivel de EEFF entregados por la compañía no se identificaron riesgos latentes entendiendo que tal y como lo mencionamos anteriormente, esta fuente de información corresponde al mismo PyG presentado por la compañía ante la Superintendencia. No obstante, desde el frente de las estadísticas de KPIs que maneja la cadena existe el riesgo de información incompleta en términos de habitaciones disponibles, habitaciones ocupadas, número de personas alojadas; destacando que la falta de alguno de estos datos automáticamente generaría errores en indicadores como el porcentaje de ocupación, el ADR y el REV_PAR. Así mismo, dentro de esta misma base, la totalidad de los ingresos viene generada por líneas independientes dentro de las cuales el no reporte de líneas como el Bar o Teléfonos-Internet podría generar una disminución en el indicador de completitud.

b) **Validez:** tanto a nivel de EEFF como a nivel de estadísticas el riesgo que podría existir en el envío de información estaría explicado por un cambio en el formato de los datos miles y decimales en algunas variables tabuladas entendiendo que las bases suministradas previamente tienen la definición de (.) para el caso de las unidades en miles y (,) para las unidades decimales. Cualquier cambio en el formato que pueda trocar estas definiciones, causaría de inmediato errores en las diferentes métricas y cálculos que dependen de estos.

c) **Puntualidad:** tal y como lo mencionamos anteriormente, por tratarse de información financiera sensible de cara a los entes reguladores. NO identificamos factores de riesgo latentes en la puntualidad de la presentación de esta información, además el operador cuenta con la presencia constante de una auditora de la fiduciaria, que audita estos reportes.

d) **Unicidad:** la información presentada en las bases de datos NO tiene fuentes potenciales de duplicados de datos en la medida que la información no se asocia a rubros o variables tales como IDs de clientes. Cada rubro presentando tanto a nivel de estadísticas como a nivel de EEFF es un rubro independiente y las bases no presentan uniones entre sí que puedan hacer perder la unicidad.

e) **Precisión:** en términos de precisión encontramos como factor de riesgo el posible incumplimiento de algunas reglas que se detallan a continuación que causarían la disminución de este indicador. Así:

- Las habitaciones ocupadas no pueden ser superiores a las habitaciones disponibles, el porcentaje de ocupación no puede ser superior al 100%, el índice de Pax alojadas debe ser proporcional al No. De Personas

Alojadas, el total de ingresos debe ser igual a la sumatoria de las líneas Habitaciones, Room Service, Restaurante, Bar, Eventos, Teléfonos-Internet, Lavandería y Otros.

f) Consistencia: desde el frente de consistencia, NO encontramos factores de riesgo en la medida que sólo tenemos una fuente de información para cada caso, lo cual a su vez significa que los registros siempre van a ser consistentes con los registros totales.

Tabla 2 Evaluación de componentes calidad de datos EEFF Operador Hotelero

Nombre Dato	Compleitud	Validez	Puntualidad	Unicidad	Precisión	Consistencia
Ingreso alojamiento suits	100%	100%	N/A	100%	100%	100%
Total Costos y Gastos	100%	100%	N/A	100%	100%	100%
Utilidad/Pérdida Alojamiento	100%	100%	N/A	100%	100%	100%
Ingresos alimentos	100%	100%	N/A	100%	100%	100%
Ingresos bebidas	100%	100%	N/A	100%	100%	100%
Costo Alimentos	100%	100%	N/A	100%	100%	100%
Costo Bebidas	100%	100%	N/A	100%	100%	100%
Ingreso Comunicaciones	100%	100%	N/A	100%	100%	100%
Costo Comunicaciones	100%	100%	N/A	100%	100%	100%
Ingreso Lavandería	100%	100%	N/A	100%	100%	100%
Costo Lavandería	100%	100%	N/A	100%	100%	100%
Ingresos						

sobreventas	100%	100%	N/A	100%	100%	100%
Costos Sobreventas	100%	100%	N/A	100%	100%	100%
Utilidad/Pérdida						
Bruta	100%	100%	N/A	100%	100%	100%
Gastos Operaciones	100%	100%	N/A	100%	100%	100%
Utilidad						
Operacional	100%	100%	N/A	100%	100%	100%
Otros Gastos	100%	100%	N/A	100%	100%	100%
Ingresos No	100%	100%	N/A	100%	100%	100%
Operacionales						
Utilidad Neta	100%	100%	N/A	100%	100%	100%

Fuente: elaboración propia

Tabla 3 Evaluación de componentes calidad de datos Estadísticas KPI

Nombre Dato	Compleitud	Validez	Puntualidad	Unicidad	Precisión	Consistencia
Habitaciones Disponibles	100%	100%	N/A	100%	100%	100%
Habitaciones Ocupadas	100%	100%	N/A	100%	100%	100%
Uso Casa	100%	100%	N/A	100%	100%	100%
Hab. Compl.	100%	100%	N/A	100%	100%	100%
No. De Personas Alojadas	100%	100%	N/A	100%	100%	100%
Índice de Pax Alojadas	100%	100%	N/A	100%	100%	100%
% de Ocupación	100%	100%	N/A	100%	100%	100%
% de Ocupación	100%	100%	N/A	100%	100%	100%

Rev_Par	100%	100%	N/A	100%	100%	100%
Tarifa Promedio	100%	100%	N/A	100%	100%	100%
Habitaciones	100%	100%	N/A	100%	100%	100%
Minibares	100%	100%	N/A	100%	100%	100%
Room Service	100%	100%	N/A	100%	100%	100%
Restaurante	100%	100%	N/A	100%	100%	100%
Bar	100%	100%	N/A	100%	100%	100%
Eventos / Ofi- Suites	100%	100%	N/A	100%	100%	100%
Telefonos - Internet	100%	100%	N/A	100%	100%	100%
Lavanderia	100%	100%	N/A	100%	100%	100%
Otros	100%	100%	N/A	100%	100%	100%
TOTAL INGRESOS	100%	100%	N/A	100%	100%	100%

Fuente: elaboración propia

8.3. Análisis Descriptivo De Las Fuentes De Información

El operador hotelero ha proporcionado datos de Ingresos Mensuales y ocupación (Occ- ADR) así como los estados financieros de Pérdidas y Ganancias (PyG) para los periodos 2022,2023 y 2024. Para comprender mejor estas fuentes de información hemos separado los datos de ocupación de los estados financieros, y así se verán a continuación.

8.3.1. Datos De Ocupación, Ingresos Mensuales

A partir de los análisis de Ingresos Mensuales Occ-ADR de los hoteles para los periodos 2022, 2023 y 2024, se realizan los primeros pasos de la exploración de datos así:

Consolidación de Ingresos Mensuales Occ-ADR histórico en los periodos previamente mencionados en una sola base y entendimiento de los diferentes indicadores de negocio presentadas en la base.

Desarrollo de un análisis descriptivo exhaustivo que incluyó la aplicación de estadísticos básicos para entender la distribución y tendencias de los ingresos mensuales. Además, se utilizaron gráficos de cajas y bigotes para visualizar la variabilidad y posibles datos atípicos en los ingresos mensuales a lo largo del periodo analizado.

Tabla 4 Estadísticos descriptivos Ingresos Mensuales OCC

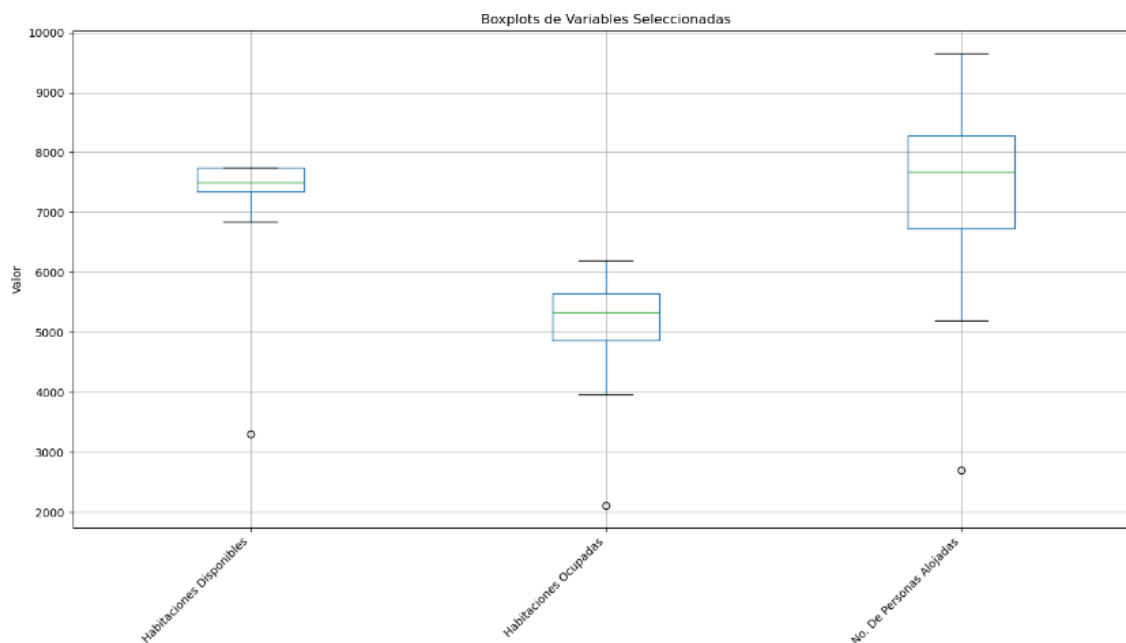
Habits en el Hotel	Habs Disponibles	Habs Ocupadas	Uso Casa	Hab. Com pl.	No. De Pax Alojadas	Índice de Pax Alojadas	% Occ	% Occ 250 Hab	Rev_P ar	Tarifa Promedio	Total, Ingresos
count	31	31	31	31	31	31	27	31	31	31	31
mean	7.581	6.899	4.847	8	13	7.009	135%	66%	66%	153.136	218.992
std	230	2.007	1.519	24	18	2.323	37%	20%	19%	46.377	64.81
min	7	7	3	0	0	6.551	132%	66%	64%	143	202.059
25%	7.5	7.316	4.733	3	6	6.551	132%	66%	64%	143.09	202.059
50%	7.75	7.5	5.331	5	17	7.492	144%	71%	71%	166.419	234.904
75%	7.75	7.75	5.644	5	17	8.242	151%	76%	76%	179.242	251.597

max	7.75	7.75	6.192	132	99	9.654	160%	78%	82%	196.985	282.81
											3

Fuente: Operador hotelero, cálculos propios.

Se observó que la ocupación promedio fue del 66%, con picos del 78%, indicando una utilización variable de la capacidad hotelera. El RevPAR promedio fue de 153.136 unidades monetarias, reflejando la eficiencia en la generación de ingresos por habitación disponible.

Además, se identificó una media de 7.009 personas alojadas, con un índice de ocupación del 135%, señalando una gestión activa de la capacidad. Los datos mostraron una variabilidad considerable, evidenciada por desviaciones estándar significativas en los ingresos y ocupación, lo que sugiere la presencia de estacionalidades o eventos específicos afectando



los resultados financieros del hotel.

Figura 5 Gráfico de caja y bigotes para ocupación de habitaciones

Fuente: Operador hotelero, cálculos propios.

La Figura 5 muestra la distribución de Habitaciones Disponibles, Habitaciones Ocupadas y No. De Personas Alojadas. La mediana de habitaciones disponibles es de aproximadamente 7500, mientras que la de habitaciones ocupadas está justo por encima de 5000 y la de personas alojadas cerca de 7000. Las tres variables presentan valores atípicos significativamente bajos. La relación entre la disponibilidad de habitaciones y el número de personas alojadas es alta, pero la menor ocupación efectiva indica una baja tasa de ocupación relativa, sugiriendo posibles ineficiencias en la gestión de ocupación.

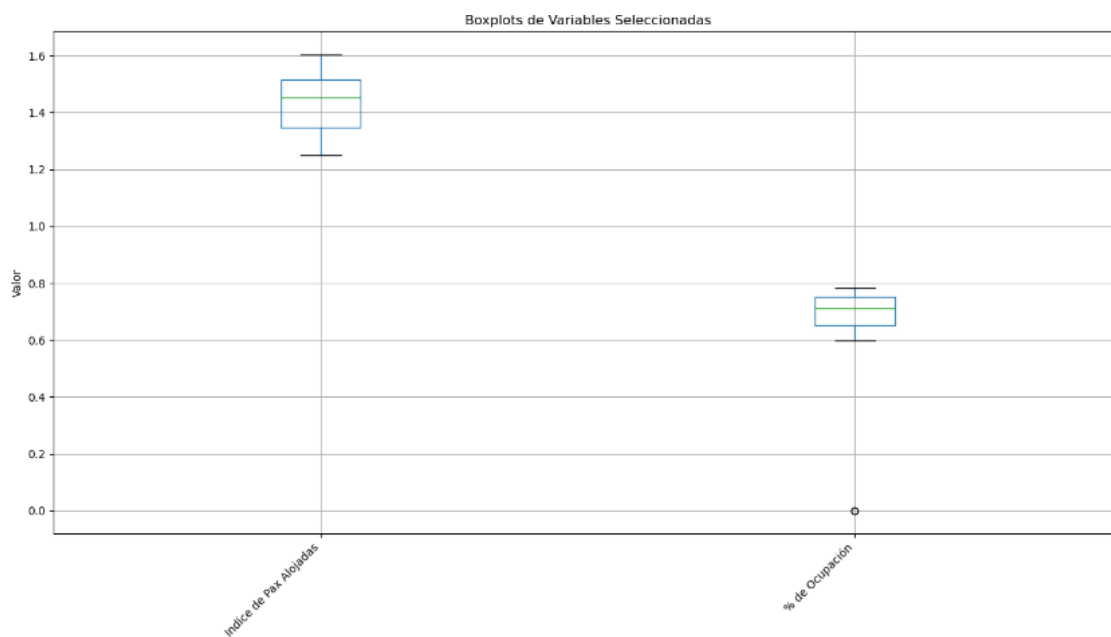


Figura 6 Índice de personas alojadas

Fuente: operador hotelero, cálculos propios.

De acuerdo con la Figura 6, la mediana del Índice de Pax Alojadas está cerca de 1.4, con un rango intercuartílico entre aproximadamente 1.3 y 1.5, sin valores atípicos. En

hoteles deben enfocarse en estrategias para aumentar su tasa de ocupación, ya que pueden mejorar significativamente su rendimiento financiero y la rentabilidad.

2. ¿Hay una relación directa entre el número de personas alojadas y la ocupación?

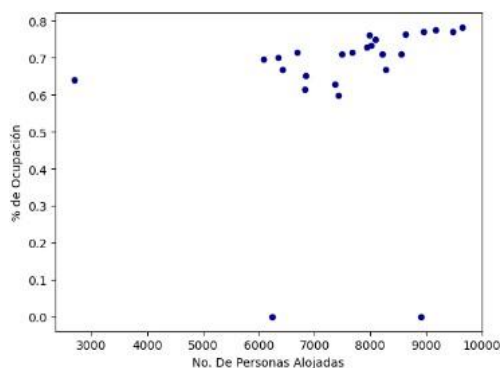


Figura 8 Relación #pax alojadas vs. ocupación

Fuente: Elaboración propia.

Sí, como se observa en la figura 8, la relación positiva entre el porcentaje de ocupación y el número de personas alojadas sugiere que el número de personas alojadas es un factor importante que determina el porcentaje de ocupación. Esto se debe a que un mayor número de personas alojadas significa que el hotel está vendiendo más habitaciones, lo que lleva a un mayor porcentaje de ocupación.

8.4. Análisis de Estados Financieros, PyG

A partir de la información proporcionada se realizan los primeros pasos de la exploración de dichos datos así:

Consolidación PyG histórico en los periodos previamente mencionados en una sola base y entendimiento de las diferentes cuentas contables presentadas en el Estado

Financiero.

Tabulación y formulación de los EEFF, en este caso el PyG. Para la formulación fue necesario entender cada dato de la estructura especialmente en materia de costos y gastos teniendo en cuenta que las bases de datos suministradas por el hotel se encontraban pegadas como valores. Las bases consolidadas se presentan en marcos mensuales con el fin de analizar datos atípicos y estacionalidades más adelante.

Desarrollo de análisis vertical del Estado de Resultados. Este análisis se realiza con base en los ingresos totales de la compañía con el fin de identificar nivel de representatividad de ítems como costos, gastos operacionales y no operacionales con respecto a los ingresos y al mismo tiempo, identificar tendencias a partir de la evolución mensual. En línea con lo anterior, se logra identificar que el mayor factor de disminución en los márgenes netos frente a márgenes brutos y operacionales está dado por los Gastos Operacionales (promedios entre 35% y 40% sobre los ingresos totales) y no por los costos. Más aún, se logra identificar que los gastos operacionales más relevantes corresponden a Gastos de administración, Mantenimiento, Servicios Públicos y Ventas; de los cuales se presentará el respecto análisis más adelante.

Ahora bien, en materia de Gastos Fijos correspondientes a la categoría No Operacional, se observa un solo rubro representativo (Retribución Operador % GOP), el cual en los últimos dos años explica cerca del 80% de los Gastos No operacionales.

Por otra parte, también se logra identificar que de las cinco líneas de ingreso con las que cuenta el hotel (Ingreso Alojamiento Suites, Ingresos Alimentos y Bebidas, Ingresos Comunicaciones, Ingresos Lavandería y Otros Ingresos Operacionales), el Ingreso por alojamiento representa históricamente cerca del 84% de los ingresos totales.

Desarrollo de gráficos de reconocimiento inicial. Se realizan dos gráficos de diagnóstico rápido. Por un lado, se grafica la evolución mensual del margen neto para identificar el estado de rentabilidad en los últimos dos años. Por otro lado, se grafica la distribución de ingresos por línea de ingresos observando una clara supremacía de los ingresos por alojamiento antes mencionado. A continuación, se presentan las gráficas:

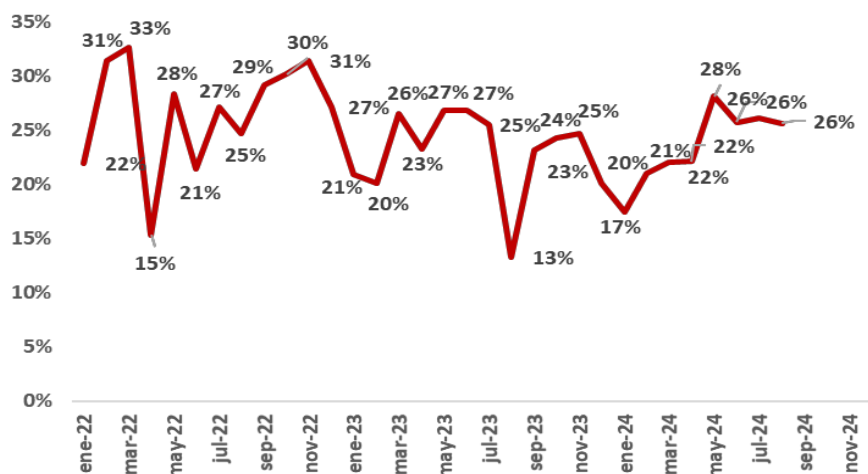


Figura 9 Evolución mensual del margen neto

Fuente: Operador hotelero. Cálculos propios

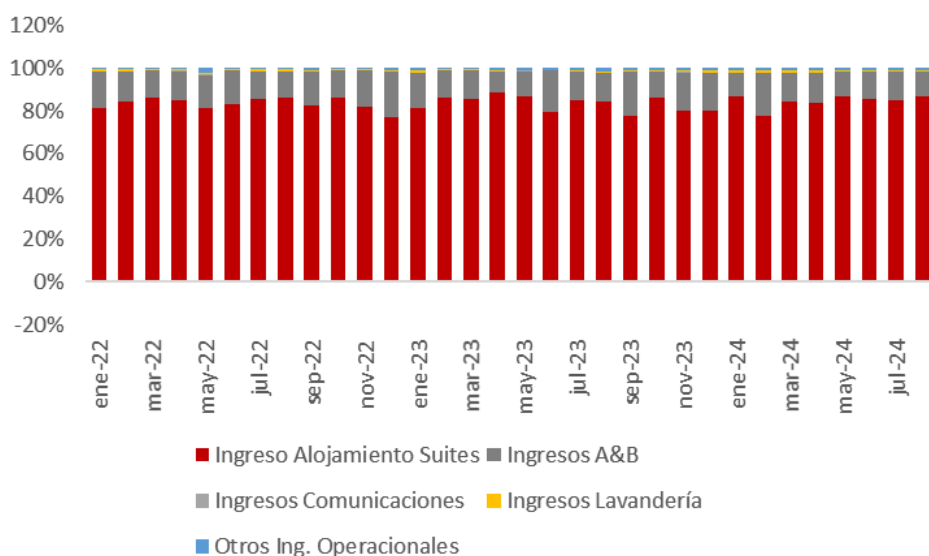


Figura 10 Distribución de ingresos por línea

Fuente: Operador hotelero. Cálculos propios

Análisis de evolución de ítems de ingreso total y de gastos operacionales. Para cada uno de los ítems antes mencionados, se calcula la variación anual, algunos datos estadísticos básicos para describir el comportamiento de las cifras (Promedio mensual, Mediana mensual, Desviación estándar mensual, Máximo, Mínimo), se realiza agrupación trimestral con el fin de encontrar cuál es la participación de cada trimestre y así poder identificar posibles estacionalidades. A continuación, se presentan los resultados obtenidos para cada rubro representativo en el PyG y comentarios de análisis que servirán de base para una mayor profundización de los datos en futuras reuniones con el equipo del hotel.

Tabla 5 Ingresos Totales

PERIODO	INGRESOS TOTALES	VARIACIÓN INGRESOS Y/Y (BASE ENE-ABRIL)
---------	------------------	---

TOTAL 2022	16,441,151,587	
TOTAL 2023	18,535,324,631	
YTD 2024	13,932,919,000	14.27%
ENE-AGOSTO 2023	12,192,862,292	19.01%
ENE-AGOSTO 2022	10,245,366,831	

Fuente: elaboración propia.

Tabla 6 Estadísticos descriptivos total ingresos

Descripción	2022	2023	2024
Promedio Mensual	1,370,095,966	1,544,610,386	1,741,614,875
Mediana Mensual	1,399,058,083	1,568,238,910	1,755,656,000
Desviación Mensual	193,858,324	158,011,099	147,561,210
Máximo	1,641,882,022	1,743,229,618	1,915,076,000
Mínimo	1,029,236,487	1,298,468,365	1,476,866,000

Fuente: elaboración propia.

Tabla 7 Total ingresos por trimestre

Descripción	2022	% Part.	2023	% Part.	2024
1T	3,516,637,116	21%	4,738,349,061	26%	4,750,921,000
2T	3,722,385,171	23%	4,615,813,225	25%	5,458,374,000
3T	4,517,430,043	27%	4,541,464,674	25%	
4T	4,684,699,257	28%	4,639,697,672	25%	

Fuente: elaboración propia.

Se observa un 2023 con crecimiento notorio en el rubro de ingresos ante un mal 2022, mientras que en 2024, las cifras son bastante estables frente a las observadas en 2023 con un crecimiento de 3,8% tal y como también lo sustentan datos estadísticos básicos. Las

variaciones 2024 vs 2023 en meses como marzo y abril se explican por las bases estadísticas de la Semana Santa.

Por otro lado, no se identifican estacionalidades claras. Ningún trimestre tiene una participación representativa sobre el total de los ingresos del año, al menos con los datos comprendidos en el periodo 2022-2024.

Tabla 8 Gastos de administración

Periodo	G. Administración	Variación Y/Y	Correlación vs Ingresos
Total 2022	755,145,215		69.03%
Total 2023	906,790,965		
YTD 2024	715,289,000	19.85%	
Ene-Agosto 2023	596,844,719	19.78%	
Ene-Agosto2022	498,297,149		

Fuente: elaboración propia.

Se observa una tendencia creciente en métricas anuales en los Gastos de Administración. Llama la atención el notorio crecimiento de los gastos en 2024 vs 2023 (23,31%) en los primeros cuatro meses del año comparado con la estabilidad de los ingresos totales los cuáles registran un crecimiento para el mismo periodo de 3.8%, lo cual se traduce en un factor de disminución clara en los márgenes de rentabilidad de la compañía.

Tabla 9 Estadísticos Gastos de administración

Descripción	2022	2023	2024
Promedio Mensual	62,928,768	75,565,914	89,411,125
Mediana Mensual	63,253,282	76,683,572	90,205,000
Desviación Mensual	5,217,629	6,488,870	2,294,544

Máximo	70,924,091	81,572,477	92,235,000
Mínimo	51,615,635	55,309,595	84,617,000

Fuente: elaboración propia.

Lo anterior, también se puede observar en un contexto en el cual la estabilidad de los ingresos se expresa a través de una disminución constante en los últimos dos años en la desviación mensual de los ingresos totales frente a un incremento sostenido en la desviación mensual de los gastos de administración.

Tabla 10 Variación gastos de administración

Descripción	2022	% Part.	2023	% Part.	2024
1T	177,599,407	24%	204,872,881	23%	264,952,000
2T	194,350,820	26%	232,696,148	26%	271,293,000
3T	197,271,012	26%	235,313,274	26%	
4T	185,923,975	25%	233,908,662	26%	

Fuente: elaboración propia.

Al igual que en el caso de los ingresos totales, no se observan estacionalidades claras y prácticamente, los gastos de administración se distribuyen de manera uniforme a lo largo de los cuatro trimestres del año.

Tabla 11 Gastos de Ventas

Periodo	Gastos de Ventas	Variación Gastos Ventas Y/Y	Correlación Gasto de Ventas vs
---------	------------------	-----------------------------	--------------------------------

		Ingresos Totales
Total 2022	1,982,456,623	81.78%
Total 2023	2,207,155,421	
YTD 2024	2,021,116,000	39.24%
Ene-Agosto 2023	1,451,569,547	16.64%
Ene-Agosto 2022	1,244,492,597	

Fuente: elaboración propia.

Tabla 12 Estadísticos Gastos de Ventas

Descripción	2022	2023	2024
Promedio Mensual	165,204,719	183,929,618	252,639,500
Mediana Mensual	158,403,852	193,253,559	256,241,500
Desviación Mensual	31,498,076	46,077,841	18,084,259
Máximo	220,150,627	260,165,455	273,769,000
Mínimo	123,142,064	70,592,395	218,210,000

Fuente: elaboración propia.

En línea con lo observado en el rubro de gastos de administración, existe una tendencia creciente en la desviación estándar mensual impactada el bajo nivel de gasto de ventas para enero 2024, el cual termina también impactando de manera directa el promedio de este rubro.

Tabla 13 Variación Gastos de Ventas

Descripción	2022	% Part.	2023	% Part.	2024
1T	410,809,813	21%	588,937,625	27%	702,963,000
2T	439,087,429	22%	544,618,764	25%	782,497,000
3T	556,358,823	28%	578,178,613	26%	
4T	576,200,558	29%	495,420,419	22%	

Fuente: elaboración propia.

A diferencia de los ingresos totales y los gastos de administración, para los años 2022 y 2023 se observan tendencias crecientes y decrecientes, respectivamente, en materia de distribución por trimestre. Ahora bien, es importante destacar que este tipo de ítem tiende a tener una correlación positiva alta con los ingresos totales. Dicho esto, hacia adelante es clave entender el porqué del bajo gasto de ventas en el mes de enero de 2024 el cual distorsionó completamente el resultado en lo corrido de 2024.

Tabla 14 Gastos Servicios Públicos

Descripción	Servicios Públicos	Var. Servicios Públicos Y/Y	Correlación S. Públicos vs Ingresos
Total 2022	895,640,048		72.28%
Total 2023	971,164,759		
YTD 2024	721,010,000	9.47%	
Ene-Agosto 2023	658,616,414	15.83%	
Ene-Agosto 2022	568,611,387		

Fuente: elaboración propia.

Tabla 15 Estadísticos Gastos Servicios Públicos

Descripción	2022	2023	2024
Promedio Mensual	74,636,671	80,930,397	90,126,250
Mediana Mensual	73,324,938	82,464,655	91,001,000
Desviación Mensual	8,552,125	8,819,188	6,073,505
Máximo	91,474,178	92,197,140	98,070,000
Mínimo	60,873,986	60,424,410	79,525,000

Fuente: elaboración propia.

Tabla 16 Variación Gastos Servicios Públicos

Descripción	2022	% Part.	2023	% Part.	2024
1T	193,136,499	22%	263,113,584	27%	261,527,000
2T	214,756,037	24%	254,742,430	26%	278,897,000
3T	245,110,197	27%	214,735,920	22%	
4T	242,637,315	27%	238,572,825	25%	

Fuente: elaboración propia.

Como era de esperarse, al tener una correlación alta con el rubro de ingresos totales, no se observan estacionalidades significativas en la distribución del gasto por servicios públicos el cual es casi uniforme a lo largo de los cuatro trimestres del año.

Tabla 17 Gastos Mantenimiento

Descripción	Mantenimiento	Var. Mantenimiento Y/Y	Correlación Mantenimiento vs Ingresos
Total 2022	499,599,454		62.26%
Total 2023	708,511,809		
YTD 2024	464,651,000	-0.62%	
Ene-Agosto 2023	467,559,518	50.96%	
Ene-Agosto 2022	309,727,589		

Fuente: elaboración propia.

Tabla 18 Estadísticos Gastos Mantenimiento

Descripción	2022	2023	2024
Promedio Mensual	41,633,288	59,042,651	58,081,375
Mediana Mensual	41,783,971	58,811,469	56,953,500
Desviación Mensual	5,925,068	5,496,823	5,222,136
Máximo	52,661,472	71,388,451	67,783,000
Mínimo	32,009,301	50,257,697	50,813,000

Fuente: elaboración propia.

Al igual que en los gastos de administración, se han presentado en los últimos dos años incrementos significativos en este rubro incluso mayores a los observados en los ingresos totales, factor que sin lugar a duda explica parte del por qué la disminución de márgenes de la compañía. En conjunto con el ítem de Servicios Públicos se observa un nivel de correlación alto con el rubro de ingresos totales.

Tabla 19 Variación Gastos de Mantenimiento

Descripción	2022	% Part.	2023	% Part.	2024
1T	119,871,944	24%	167,880,634	24%	171,598,000
2T	108,908,701	22%	183,873,145	26%	175,342,000
3T	123,522,645	25%	176,613,714	25%	
4T	147,296,164	29%	180,144,315	25%	

Fuente: elaboración propia.

8.4.1. Metodología Basada En Análisis Vertical

En un primer momento el operador hotelero consideraba que su problemática se enfocaba en poder de pronosticar de una manera acertada los costos y gastos de su negocio. No obstante, en la fase de inicial de entendimiento de los datos, se logró identificar que la relación tanto de los costos como de los gastos del hotel frente a los ingresos de la operación

se mantenía bastante estable en el tiempo. Lo anterior, se pudo determinar a partir de la realización de un análisis vertical del Estado de Pérdidas y Ganancias con base en la información inicialmente suministrada.

De acuerdo con lo anterior, se modificó el planteamiento inicial hacia un enfoque en el cual la principal problemática a resolver fuera desarrollar modelos estadísticos robustos que permitieran pronosticar los ingresos, para con base en estos pronósticos y entendiendo la relación estable antes mencionada de los costos y los gastos, poder pronosticar de una manera más adecuada estos últimos entendiendo que el principal indicador financiera a cumplir por parte del operador hotelero siempre ha sido mantener un Margen Neto de Rentabilidad superior al 30%.

A continuación, se presentan los hallazgos realizados en este análisis vertical para uno de los hoteles del operador hotelero.

Tabla 20 Análisis Vertical

Fecha	Ingresos Totales	% Costos/Ingresos	% Gastos/Ingresos
1/01/2024	\$ 1.805.145.000	4,4%	71,1%
1/02/2024	\$ 2.190.962.000	4,7%	67,8%
1/03/2024	\$ 2.175.397.000	5,3%	66,9%
1/04/2024	\$ 2.338.858.000	5,0%	65,9%
1/05/2024	\$ 2.527.788.000	5,1%	63,5%
1/06/2024	\$ 2.529.882.000	5,7%	63,0%
1/07/2024	\$ 2.338.907.000	4,5%	66,2%
1/08/2024	\$ 2.378.208.000	4,8%	65,8%
1/09/2024	\$ 2.597.945.000	5,7%	62,5%
1/10/2024	\$ 2.689.967.000	5,3%	62,2%
1/11/2024	\$ 2.678.787.000	5,4%	62,1%
1/12/2024	\$ 1.938.414.000	5,4%	71,8%
Promedio	\$ 2.349.188.333	5,1%	65,7%

Fuente: elaboración propia.

Tal y como se puede observar en la tabla 20, mes a mes, los costos representan una proporción sobre los ingresos totales en un rango entre 4,5% y 5,5%, mientras que los gastos como un todo lo hacen en un rango entre 63% y 67%. Por esta razón, para este trabajo de grado se generarán modelos de predicción de ingresos para posteriormente, con base en la metodología de análisis vertical poder pronosticar tanto los costos y gastos para el periodo apoyándose en la estabilidad en la relación entre estos rubros.

9. Preparación De Base De Datos

9.1. Datos Internos Del Hotel

Como se mencionó con anterioridad a partir del análisis vertical se determinó que a partir de los indicadores del hotel se busca predecir los ingresos, para con estas predicciones controlar costos y gastos. Así, se creó una base de datos inicial, con 92 columnas, y 48 filas. Las 92 columnas que corresponden a todos los indicadores del hotel, así como los valores de las cuentas de los estados financieros (ingresos, costos, gastos, utilidad, etc.). Y las 48 filas son los 48 meses de la ventana de tiempo que se tomó desde enero de 2021 a diciembre de 2024.

9.2. Definición De Variables Exógenas

En la construcción de un modelo de pronóstico de ingresos para el sector hotelero, resulta fundamental incorporar variables macroeconómicas debido a su impacto directo e indirecto sobre la dinámica del consumo, la demanda turística y la capacidad de gasto de

los clientes (en este caso de las empresas entendiendo que el operador hotelero en el cual se enfoca este trabajo de grado tiene una alta dependencia de sus ingresos en el segmento corporativo). Estas variables actúan como factores exógenos que influyen en la toma de decisiones tanto de viajeros nacionales como internacionales, así como en la operación misma del establecimiento.

La inflación, por ejemplo, afecta el poder adquisitivo de los consumidores, reduciendo la disponibilidad de recursos para actividades como el turismo y el alojamiento. Una inflación elevada puede restringir el consumo de servicios no esenciales y generar ajustes en los precios de los productos y servicios que ofrece el hotel, afectando directamente sus ingresos.

La tasa de interés del Banco de la República incide en el costo del dinero y en las condiciones de financiación tanto para los hogares como para las empresas. Tasas de interés elevadas encarecen los créditos y desincentivan el gasto, lo cual repercute negativamente en la ocupación hotelera, especialmente en segmentos sensibles a los ciclos económicos.

La tasa de cambio también juega un papel clave. Un peso colombiano depreciado frente al dólar puede favorecer el turismo extranjero al hacer más atractivos los destinos locales desde el punto de vista del costo. Sin embargo, también puede encarecer insumos importados, elevando los costos operativos del hotel y afectando su rentabilidad.

El Indicador de Seguimiento a la Economía (ISE) permite observar la evolución de la actividad económica en el corto plazo. Su comportamiento refleja el nivel de dinamismo productivo del país, lo cual guarda relación con el flujo de viajeros por motivos laborales, comerciales o de recreación. Un entorno económico activo favorece

la movilidad y el gasto de los hogares, condiciones favorables para el sector hotelero.

Finalmente, el crecimiento económico es una medida agregada de la salud de la economía. En periodos de expansión económica, aumenta el empleo, mejora la confianza del consumidor y crecen los ingresos disponibles, lo cual se traduce en una mayor disposición para viajar, hospedarse y consumir servicios turísticos. Por estas razones, integrar estas variables al modelo de pronóstico permite capturar los efectos del entorno económico general sobre los ingresos del hotel, mejorando la precisión del modelo y permitiendo una planificación estratégica más robusta.

Así las cosas, a partir de un criterio de experto, se definieron como variables exógenas para incluir en los modelos a realizar: la inflación mensual y anual total en Colombia y a su vez, el subcomponente de servicios; el crecimiento económico generalizado de la economía Colombiana y el indicador de seguimiento económico para la industria publicado de manera mensual por el DANE; la tasa de cambio (Peso colombiano frente al Dólar estadounidense) y la evolución de la tasa de referencia de política monetaria determinada por el Banco de La República de manera mensual.

Con esta información, se complementó la base de datos inicial, agregando 16 columnas de indicadores económicos, dentro de la misma ventana de tiempo determinada. A lo largo de las pruebas con los distintos modelos, esta base se fue modificando según las variables que se iban a tener en cuenta.

10. Exploración De Modelos Para Predicción De Ingresos

10.1. Regresión Lineal

El análisis de regresión lineal es una herramienta estadística que permite modelar la relación entre una variable dependiente y un conjunto de variables independientes, con el fin de explicar y predecir el comportamiento de la primera. En este caso, se utilizó la regresión lineal para explicar los ingresos totales mensuales del operador hotelero, a partir de variables seleccionadas por su relevancia económica y su relación estadística con la variable objetivo.

Para este modelo, se cargó la base de datos histórica del hotel, y se calculó la matriz de correlación entre las variables, lo que permitió identificar las variables con mayor asociación lineal con los ingresos totales. En la figura 14 se muestra el mapa de calor generado a partir de dicha matriz, donde se destacan correlaciones fuertes (superiores a 0.7 o inferiores a -0.7) entre los ingresos y variables como la tarifa promedio, número de habitaciones disponibles y número de eventos realizados. Esta visualización permitió definir un subconjunto óptimo de predictores para el modelo.

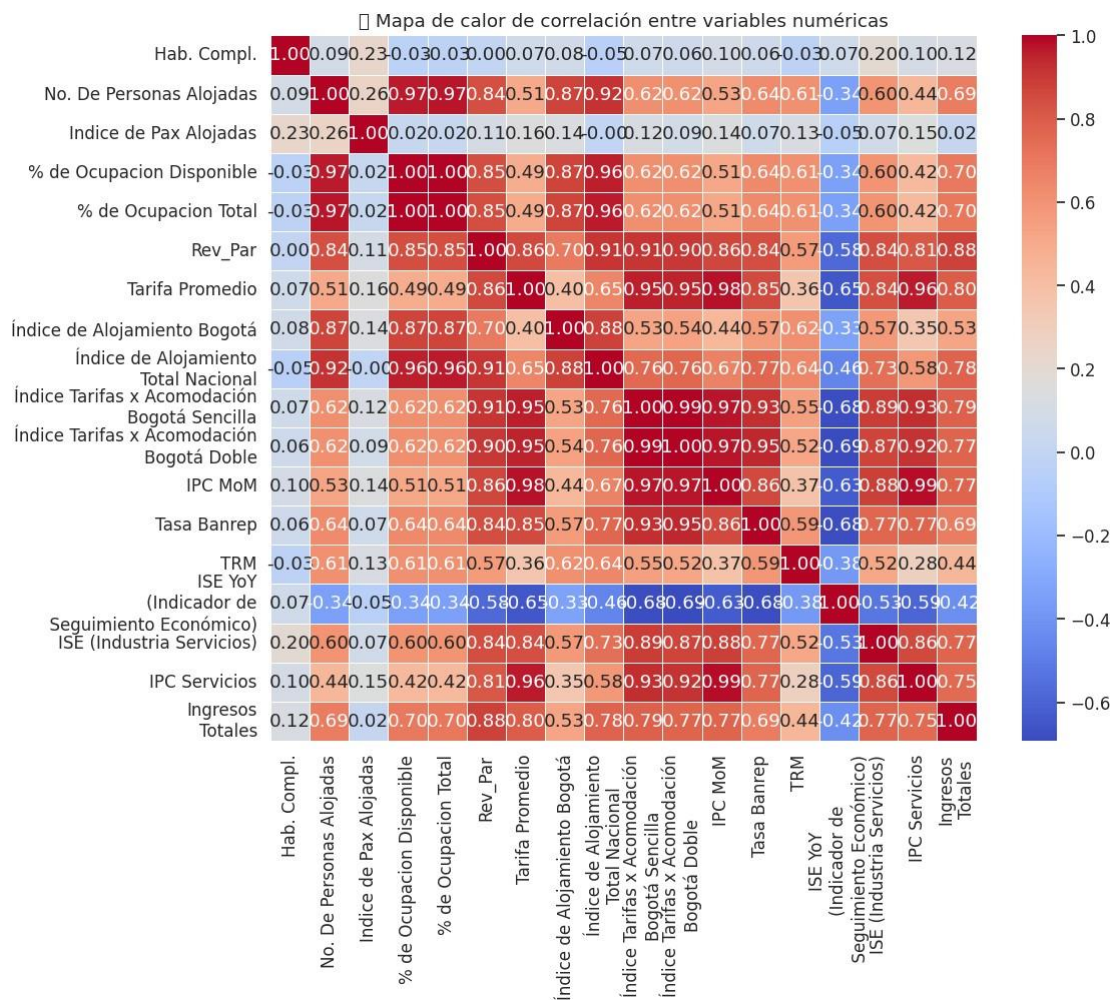


Figura 11 Mapa de Calor Correlaciones entre variables - Regresión Lineal

Fuente: elaboración propia.

Con base en el análisis anterior, se seleccionaron las siguientes variables como predictoras: Tarifa Promedio, Habitaciones, Eventos, Índice de Alojamiento Bogotá, Variación Anual Tarifas Hoteleras Bogotá (Sencilla), IPC MoM, IPP MoM e ISE, ya que son las 8 variables con mayor correlación la variable de ingresos totales, como se observa en la figura 15. Además, para cumplir con los supuestos del modelo y estabilizar la

varianza de la variable objetivo, se aplicó una transformación logarítmica a los ingresos totales.

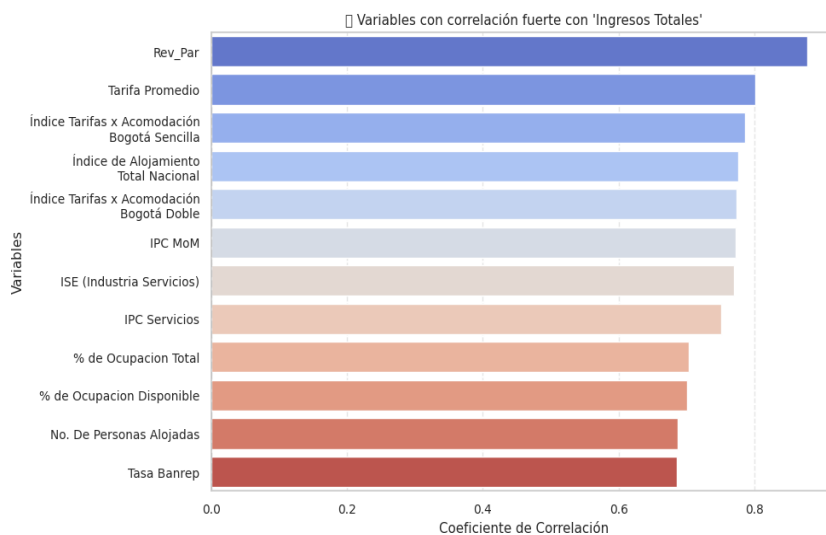


Figura 12 Variables con alta correlación con "Ingresos Totales" – Regresión Lineal

Fuente: elaboración propia.

Posteriormente, se entrenó el modelo de regresión lineal utilizando la biblioteca statsmodels, lo que permitió obtener no solo las predicciones, sino también los coeficientes de cada variable y su nivel de significancia estadística (valor p). En la figura 13 se pueden observar los resultados del modelo.

Tabla 21 Resultados Regresión lineal

Variable	B	EE	t	p	IC 95%
Constante	20.380	0.888	22.961	<.001	[18.580, 22.180]
Rev_Par	-0.000001	0.000003	-0.257	.799	[-0.000006, 0.000004]
Tarifa Promedio	0.000006	0.000002	2.530	.016	[0.000001, 0.000011]
Índice Tarifas Acomod. Bogotá (Sencilla)	0.0029	0.011	0.257	.799	[-0.020, 0.025]
Índice Alojamiento Nacional Total	-0.0015	0.010	-0.156	.877	[-0.021, 0.018]
Índice Tarifas Acomod. Bogotá (Doble)	-0.0151	0.015	-1.039	.306	[-0.045, 0.014]
IPC MoM	-0.0166	0.044	-0.375	.710	[-0.106, 0.073]

ISE (Industria - Servicios)	0.0022	0.001	1.530	.135	[-0.001, 0.005]
IPC Servicios	0.0154	0.049	0.316	.754	[-0.084, 0.114]
% Ocupación Total	37.882	15.707	2.412	.021	[6.027, 69.737]
% Ocupación Disponible	-35.899	15.560	-2.307	.027	[-67.457, -4.341]
No. de Personas Alojadas	-0.00007	0.00005	-1.586	.121	[-0.0002, 0.00002]
Tasa Banrep	1.1339	3.702	0.306	.761	[-6.375, 8.643]

Fuente: elaboración propia.

El modelo de regresión lineal múltiple mostró una alta capacidad explicativa, con un coeficiente de determinación $R^2 = 0.835$, lo que indica que aproximadamente el 83.5% de la variabilidad en los ingresos totales del hotel puede ser explicada por las variables incluidas en el modelo. Al observar los coeficientes, se identifican como predictores estadísticamente significativos la tarifa promedio ($B=0.000006$, $p=.016$), el porcentaje de ocupación total ($B=37.88$, $p=.021$) y el porcentaje de ocupación disponible ($B=-35.90$, $p=.027$). Estos resultados sugieren que, en el contexto del operador hotelero, tanto el comportamiento tarifario como la dinámica de ocupación tienen un impacto directo en la generación de ingresos. En contraste, variables como el IPC, ISE, o la tasa del Banco de la Republica no resultaron significativas en este modelo. Es importante destacar que el valor del número de condición ($3.73e+08$) indica la posible presencia de multicolinealidad entre algunas variables, lo cual podría estar afectando la estabilidad de los coeficientes estimados.

Aunque se realizó un análisis de VIF para reducir la multicolinealidad, se determinó que este modelo no era adecuado para lo que buscábamos con el operador hotelero, ya que al hacer el análisis de VIF se eliminaban variables altamente importantes, y además si se quisiera implementar este modelo en el hotel se requeriría hacer una previa proyección de cada variable para poder luego aplicar el modelo de regresión lineal. Este modelo fue descartado, pero muy útil para la comprensión de los datos, y el proceso de aprendizaje.

10.2. Random Forest

Se utilizó la base creada inicialmente, con variables históricas entre enero de 2021 y

diciembre de 2024. Se cargó la base de datos y se realizó una limpieza y transformación previa, que incluyó convertir la columna Fecha a formato de fecha (datetime), y la creación de variables categóricas como año, mes y día, derivadas de dicha columna. Estas variables se crearon para que el modelo pudiera captar variaciones estacionales y tendencias cronológicas de manera más eficiente.

Asimismo, se seleccionaron como predictoras varias variables que reflejan el comportamiento económico y operativo del hotel, incluyendo Tarifa Promedio, Eventos, Habitaciones, IPC, IPP, ISE y algunos indicadores externos del sector hotelero. Se aplicó también la transformación logarítmica a la variable objetivo (Ingresos Totales), para estabilizar su varianza y facilitar el ajuste del modelo.

Para entrenar el modelo, se dividió la base de datos en entrenamiento (80%) y prueba (20%) utilizando `train_test_split` de `sklearn`, para que el modelo se entrenara con una muestra representativa y se evaluara con datos no vistos. A continuación, se entrenó el modelo `RandomForestRegressor`, utilizando validación cruzada para garantizar robustez y evitar sobreajuste. En la figura 13, se pueden observar algunos de los árboles creados, dentro del modelo.

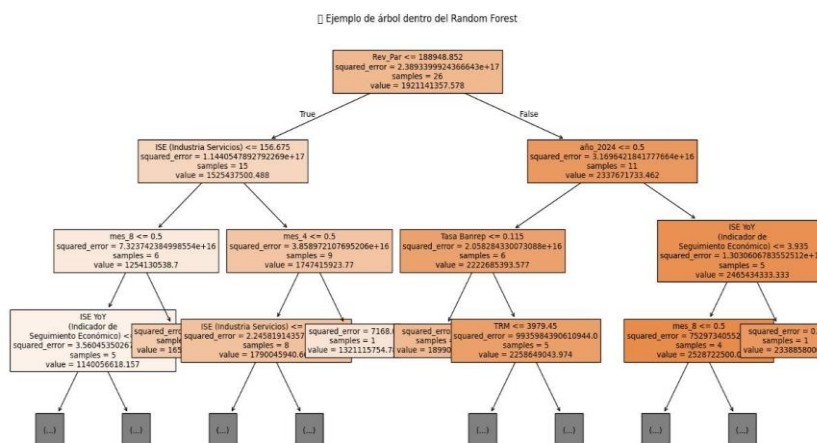


Figura 13 Ejemplo de árboles - Random Forest

Fuente: elaboración propia.

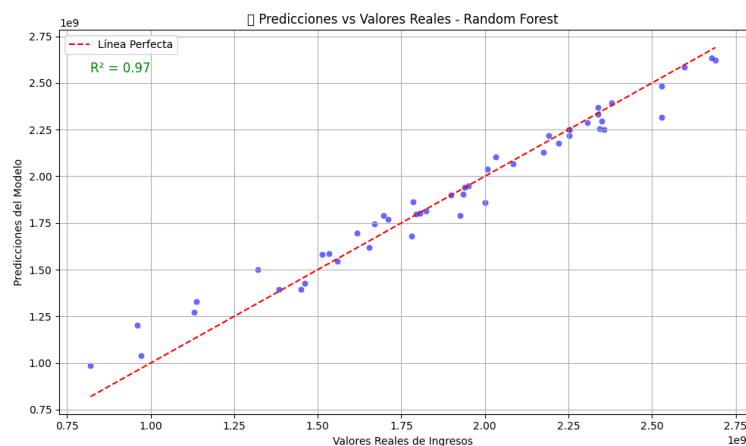


Figura 14 Predicciones vs Valores Reales - Random Forest

Fuente: elaboración propia.

En la figura 14 se puede ver el resultado de la predicción del modelo vs los valores reales, se observa una tendencia creciente en los ingresos y un buen ajuste del modelo. Sin embargo, el modelo fue evaluado sobre el conjunto de prueba utilizando métricas estándar: MAE (Error Absoluto Medio), MAPE (Error Porcentual Absoluto Medio) y R^2 (Coeficiente de Determinación). El modelo presentó un buen desempeño, con un R^2 cercano a 0.92, lo que indica que logra capturar gran parte de la variabilidad de los ingresos, aunque advierte también un riesgo de sobreajuste. En cuanto al MAE, esta por encima de los 200 millones de pesos, un valor alto para el error. Y el MAPE, está en 12%, que, si bien no es un error significativamente alto, otros modelos mostraron mejor desempeño. Además, al igual que para la regresión lineal este modelo requería para poder ser implementado proyectar primero las variables predictoras, lo que aumentaría el riesgo de error del modelo.

10.3. Modelos De Series De Tiempo

10.3.1. Prophet

El análisis de series de tiempo permite modelar fenómenos que evolucionan temporalmente, y así detectar patrones y poder generar pronósticos útiles para la toma de decisiones. (Jiménez, 2012). Para aplicar la serie de tiempo a la base de datos se aplicó el modelo Prophet, un modelo de series de tiempo, desarrollado por Facebook (Taylor & Letham, 2018), es un modelo basado en descomposición aditiva, que captura tendencias no lineales y estacionalidades anuales o semanales, lo cual lo hace adecuado para sectores como el hotelero, donde existen picos de demanda estacionales predecibles. Cabe aclarar que para el caso del operador hotelero, las estacionalidades que se espera encontrar son contrarias a las de un hotel que se dedica al turismo, pues el negocio de este hotel, es principalmente hospedaje para viajes de negocios y evento empresariales.

Se utilizó una serie mensual de ingresos del hotel, con los ingresos totales registrados desde enero de 2021 a diciembre de 2024. Se dividió esta base en base de entrenamiento y base de evaluación, la base de entrenamiento fue de enero de 2021 a junio 2024 y los meses restantes, fueron base de evaluación. Para cumplir con los requisitos de Prophet, se renombraron las columnas a `ds` (Fecha) y `y` (Ingresos Totales), y se aplica logaritmo natural a `y` para estabilizar la varianza. Se aplicó el código `df_prophet['y'] = np.log(df_prophet['y'])` haciendo uso de Google Colab.

Para el entrenamiento del modelo se configuró Prophet con estacionalidad anual y modo aditivo, desactivando estacionalidades semanales o diarias. Luego se ajustó sobre la

serie transformada y se generaron predicciones a 12 meses.

A continuación, se presentan gráficamente los resultados del modelo. En la figura 15 se presenta la serie observada, observando valles en los ingresos en los meses de enero, y según el año abril o marzo, es posible decir que estos valles pueden estar asociados a las temporadas de vacaciones y semana santa.



Figura 15 Serie Observada Ingresos Totales - Modelo Prophet

Fuente: Elaboración propia

Una vez analizada la serie observada se genera entonces la predicción de Ingresos Totales con Prophet, en escala original. En la figura 16 se muestra en la línea azul la serie original de ingresos (en millones), sin transformaciones. Ahora, en la línea naranja punteada se observa la predicción que hace Prophet a futuro, transformada de nuevo a la escala original (en millones), y finalmente en el área sombreada naranja se representa el intervalo de confianza al 95% de la predicción, indicando el rango donde es probable que caigan los valores futuros.

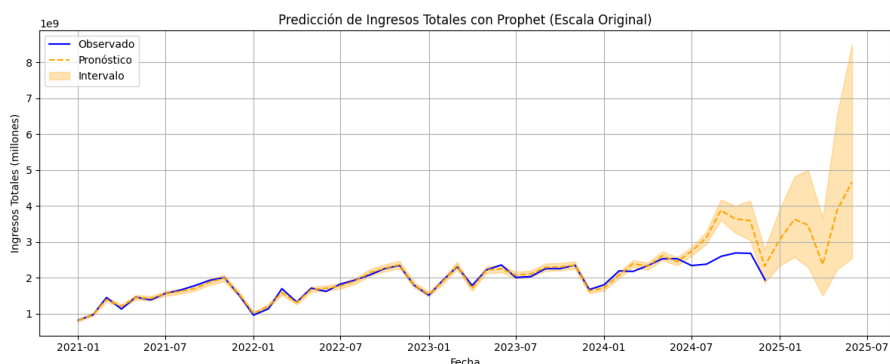


Figura 16 Predicción de Ingresos Totales con Prophet

Fuente: Elaboración propia

A partir de esta grafica presentada en la figura 16 podemos interpretar que Prophet captura bien la tendencia general creciente de los ingresos. Así como la estacionalidad, que parece estar bien reflejada, pues se ven subidas y bajadas en los mismos periodos del año. El intervalo se ensancha en el futuro, lo que es normal al aplicar estos modelos, pues hay mayor incertidumbre conforme avanza el horizonte de predicción. Sin embargo, al observar el periodo de evaluación se ve que los valores reales no se encuentran en el intervalo de confianza, y por el contrario se encuentran significativamente por debajo de lo pronosticado, esto para el hotel sería negativo, pues se vería afectado su margen.

Como parte del modelo Prophet, se realizó el análisis de residuos, que se observa en la figura 17, estos residuos muestran la diferencia entre los valores reales y los valores pronosticados. Al interpretar estos resultados observamos que los residuos oscilan alrededor de la línea cero, lo que sugiere que el modelo no tiene sesgo y no está sobreestimando ni subestimando los ingresos. Además, no se observa una tendencia clara en los residuos, lo que indica que el modelo captura adecuadamente la estructura temporal principal de los datos. Sin embargo, y reforzando la alerta observada en la base de evaluación, se evidencia que en algunos puntos (por ejemplo, inicios de 2022 y 2024), se observan residuos relativamente grandes. Esto puede deberse a eventos atípicos o factores externos que el modelo no logró anticipar correctamente, por esta razón este modelo fue descartado, ya que al ser una serie de tiempo sencilla, solo toma en cuenta la estacionalidad de la serie observada y no el efecto de variables exógenas, que pueden aportar a predecir

los ingresos.

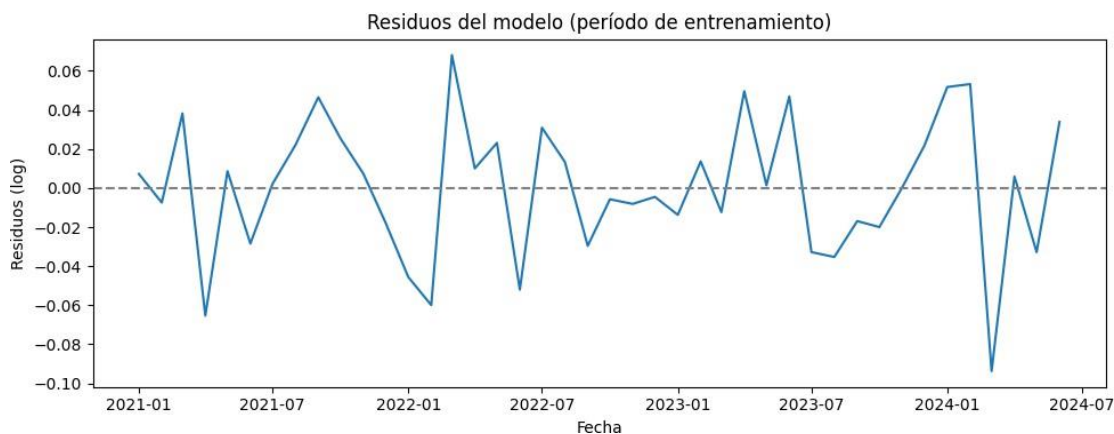


Figura 17 Residuos Modelo Prophet

Fuente: elaboración propia.

10.3.2. ARIMA

El modelo ARIMA (Autoregressive Integrated Moving Average) es un modelo estadístico versátil que captura comportamientos a corto plazo y tendencias a largo plazo, adecuado para series de tiempo estacionarias, proporcionando coeficientes interpretables que ayudan en la comprensión de las dinámicas subyacentes (Sharma & Choudhary, 2022, p. 477). Para el caso del operador hotelero objeto del trabajo de grado y según la naturaleza temporal de sus datos lo hace candidato para métodos de series de tiempo, como el ARIMA, adicional la serie del hotel demuestra una clara tendencia, pero no es estacionalidad definida por lo que el ARIMA nos ayuda a capturar la tendencia mediante los parámetros de diferenciación.

Según Noble (2024), dentro del modelo ARIMA existen tres parámetros importantes

para su implementación:

- p : significa el orden de la parte autorregresiva del modelo
- d : el grado de diferenciación involucrado en la serie de tiempo
- q : el orden de la parte del promedio móvil

A continuación, se detalla el paso a paso de la creación del modelo:

Se inició con la carga y limpieza de la base de datos, estandarizando los nombres de las columnas, eliminando valores nulos y ajustando la estructura temporal de la serie. Se definió la fecha como índice y se aseguró el orden cronológico ascendente.

Para poder contemplar mejor la serie histórica del operador hotelero se lleva a cabo mediante la representación gráfica de la serie, como se observa en la figura 18, en esta se observó una tendencia creciente sostenida en los ingresos desde 2021 hasta 2024. Asimismo, se identificaron caídas recurrentes en el mes de diciembre de cada año, lo que podría sugerir una estacionalidad puntual. Sin embargo, este comportamiento no se presenta con suficiente regularidad como para considerar la serie como estacional estructurada, motivo por el cual se optó inicialmente por un modelo ARIMA sin componente estacional.

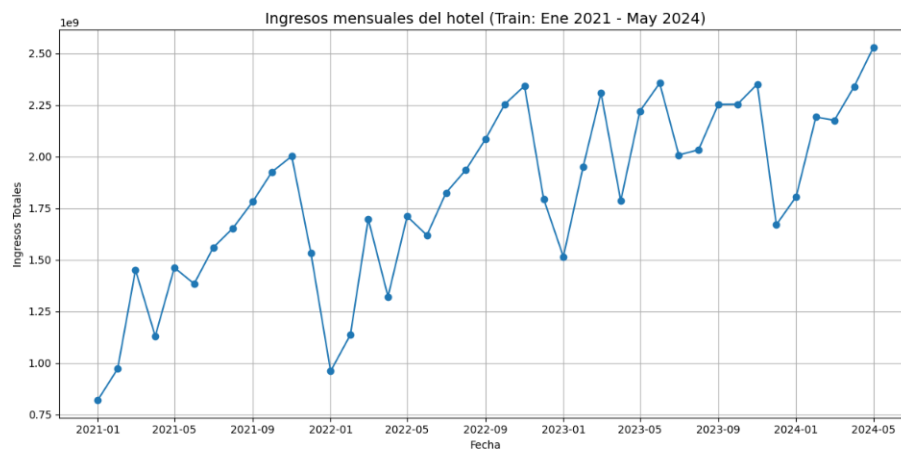


Figura 18 Ingresos base de entrenamiento – ARIMA

Fuente: elaboración propia.

Con el fin de identificar si la serie de tiempo presentada es estacionaria se aplicó a la serie la prueba de Dickey-Fuller Aumentada (ADF) dando como resultado un p value aproximado de 0,056, lo cual indica que no se puede rechazar la hipótesis nula de que la serie no es estacionaria. Teniendo en cuenta el resultado anterior se debe hacer la serie estacionaria probando con una diferencia de primer orden dando como resultado un nuevo p value aproximado de 0.00000037 indicando que la serie diferenciada es estacionaria lo que indica que para el modelo ARIMA podemos utilizar el valor $d = 1$.

Para seguir identificando la estructura de la serie y apoyar la definición de los parámetros para el modelo ARIMA, se realiza la prueba de autocorrelación (ACF) y autocorrelación parcial (PACF) sobre la serie diferenciada.

Con la prueba de autocorrelación (ACF) se midió el grado de correlación que existe entre los valores actuales de la serie diferenciada y sus valores pasados, el resultado, como se ve en la figura 19, muestra un pico significativo en el primer regazo y una caída posterior dentro de los intervalos de confianza, por lo cual se sugiere un componente de media móvil de orden 1 ($q = 1$).

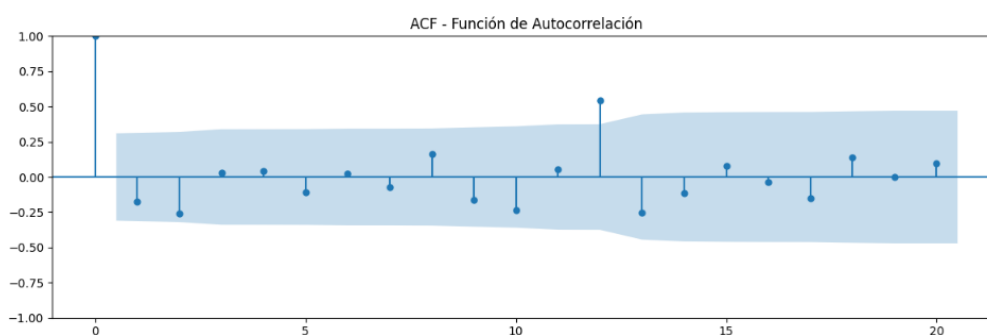


Figura 19 Autocorrelación – ARIMA

Fuente: elaboración propia.

Por otro lado, con la prueba de autocorrelación parcial ideal para identificar la posible presencia de un componente autorregresivo, mostrando un valor significativo en el primer rezago como se observa en la figura 20 y cortando a cero posteriormente, indicando la presencia de un componente autorregresivo de orden 1 ($p = 1$).

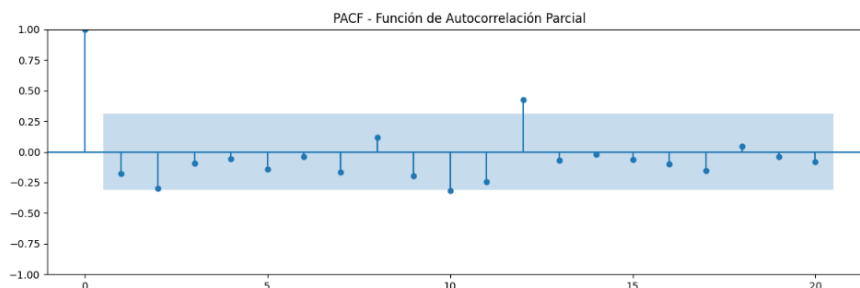


Figura 20 Prueba autocorrelación parcial – ARIMA

Fuente: elaboración propia.

Para el entrenamiento del modelo, se divide la serie en entrenamiento (enero 2021 a mayo 2024) y en prueba (junio 2024 a diciembre 2024), con esto de igual manera se ajusta el modelo con la configuración propuesta ARIMA (1,1,1) generando las predicciones correspondientes para el periodo de prueba. Con los resultados se grafica la serie de entrenamiento, predicciones y los intervalos de confianza, que se observa en la

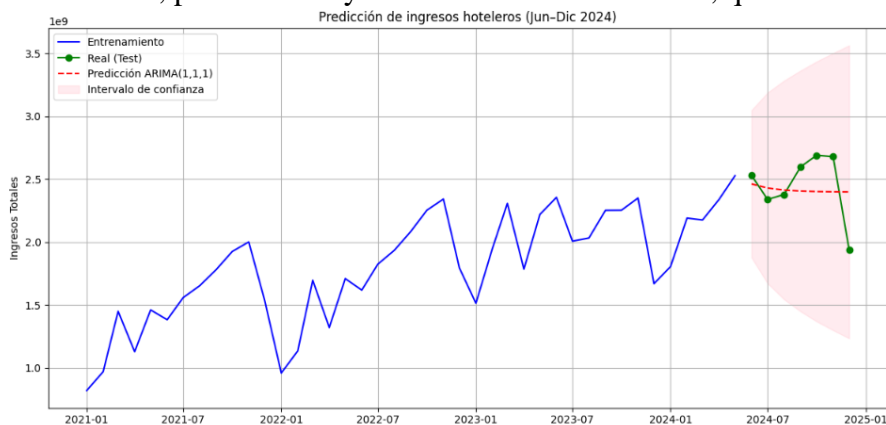


figura 21.

Figura 21 Predicción ingresos – ARIMA

Fuente: elaboración propia.

Con la finalidad de medir la efectividad del modelo se calcularon métricas de desempeño, dando los siguientes resultados:

- MAE: \$192,470,228
- RMSE: \$248,030,135
- MAPE: 8.54%
- R^2 : -0.03

Los resultados muestran que el modelo logra un error porcentual absoluto medio (MAPE) aceptable, lo que indica una precisión razonable en la predicción relativa. Sin embargo, el valor negativo del R^2 evidencia que el modelo no captura adecuadamente la variabilidad total de la serie, lo cual puede atribuirse a la alta volatilidad observada.

Adicionalmente, se aplicó la prueba de Ljung-Box sobre los residuos del modelo, obteniendo un p value de 0,4667, lo que indica que no se evidencian correlaciones significativas remanentes en los residuos, es decir, estos se comportan como ruido blanco.

En consecuencia, si bien el modelo ARIMA (1,1,1) es un punto de partida sólido para contemplar la dinámica de la serie, presenta una capacidad razonable para capturar la tendencia de la serie, se recomienda considerar modelos más robustos que integren

variables exógenas y permitan modelar mejor la irregularidad y la volatilidad observada en los ingresos hoteleros.

10.3.3. ARIMAX

El modelo ARIMAX (Autoregressive Integrated Moving Average with Exogenous Variables) se implementó como paso previo al modelo SARIMAX, con el objetivo de evaluar la capacidad de predicción de los ingresos totales del hotel al incorporar variables externas relevantes. A diferencia del ARIMA tradicional, ARIMAX permite integrar información exógena que influye directamente sobre la serie de tiempo, lo que resulta útil cuando existen factores externos, como indicadores económicos, que afectan los ingresos del hotel (Ortiz Martín, 2022).

Tras el análisis de estacionariedad (ADF), se determinó que la serie original no era estacionaria ($p > 0.05$), por lo que se aplicó una diferenciación, definiéndose el parámetro $d = 1$. Posteriormente, mediante el análisis de autocorrelación (ACF) y autocorrelación parcial (PACF), se sugirió un modelo con parámetros $p = 1$ y $q = 1$, resultando en un modelo ARIMAX(1,1,1).

Para el análisis y selección de variables exógenas se hacen distintas pruebas que se explican a continuación:

- Correlación: En la **Figura 22** se muestra las variables que tienen mayor correlación con la variable a predecir, ingresos totales.

```

 Correlación de variables con IngresosTotales:
Rev_Par                                0.877499
Tarifa_Promedio                         0.800912
Índice_Tarifas_x_AcomodaciónBogotá_Sencilla  0.786438
Índice_de_AlojamientoTotal_Nacional       0.775740
Índice_Tarifas_x_AcomodaciónBogotá_Doble  0.774064
IPC_MoM                                  0.772169
ISE_(Industria_Servicios)                0.769431
IPC_Servicios                            0.751176
%_de_Ocupacion_Total                     0.701883
%_de_Ocupacion_Disponible                 0.698506
No._De_Personas_Alojadas                  0.686695
Tasa_Banrep                               0.681873
Índice_de_Alojamiento_Bogotá             0.529961
TRM                                       0.438569
Hab._Compl.                              0.121054
Indice_de_Pax_Alojadas                    0.003909
ISE_YOY(Indicador_de_Seguimiento_Económico) -0.424218
dtype: float64

```

Figura 22 Correlación de variables con los ingresos totales

Fuente: elaboración propia.

- Causalidad de granger: Con las variables exógenas se aplica la prueba de granger y se grafican en la **Figura 23** solo aquellas que muestran una causalidad significativa (p value menor a 0.05) con ingresos totales.

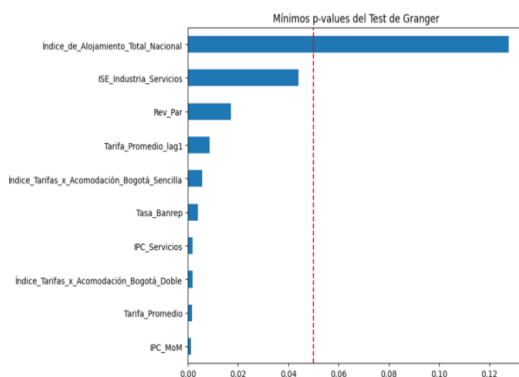


Figura 23 Causalidad de granger

Fuente: elaboración propia.

Con los resultados de las anteriores pruebas, se procede a entrenar el modelo ARIMAX (1,1,1). Primero dividiendo la serie en entrenamiento (enero 2021 a mayo 2024) y en prueba (junio 2024 a diciembre 2024), generando las predicciones correspondientes para el periodo de prueba. Con los resultados se grafica **Figura 24** la serie de entrenamiento, predicciones y los intervalos de confianza.

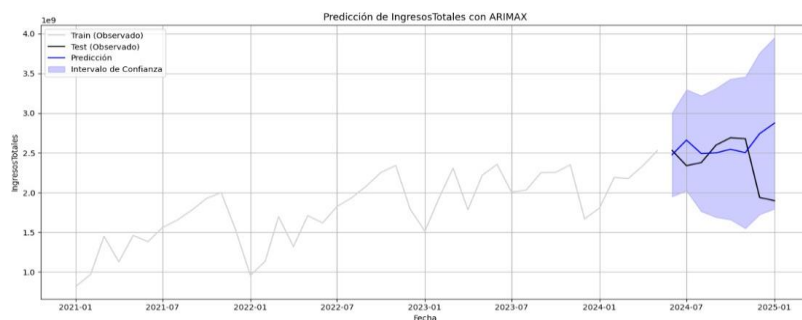


Figura 24 Predicción de ingresos – ARIMAX

Fuente: elaboración propia.

Para finalizar la ejecución de modelo ARIMAX (1,1,1) se sacan las métricas de evaluación dando los siguientes resultados

- MAE : 335355076.22
- RMSE : 471099685.97
- R^2 : -1.5982
- MAPE : 16.13%
- Ljung-Box p-value: 0.5881

Los indicadores anteriores muestran que, en promedio el modelo se equivoca por unos \$335 millones al mes y hay algunos periodos con errores muy grandes (RMSE). Así mismo el R^2 que el resultado predictivo del modelo no se ajusta a la realidad del operador hotelero y a sus predicciones financieras. Por último indicando que no hay autocorrelación en los residuos, evitnado dejar información importante por fuera del modelo.

Los resultados anteriores indican que el modelo ARIMAX (1,1,1) no captura correctamente la dinámica de ingresos del operador hotelero, surgiendo que los ingresos presentan un patrón más complejo, dado que este presenta estacionalidades claras, sería conveniente utilizar un modelo SARIMAX, el cual contempla la estacionalidad de forma explícita

10.3.4. SARIMA

El modelo SARIMA (Seasonal ARIMA) está diseñado para manejar series de tiempo

con tendencia y patrones estacionales dominantes en una variable. Este algoritmo logra capturar las variaciones periódicas que se presentan intervalos regulares de tiempo y los combina con los componentes no estacionales de un ARIMA tradicional (Molero, 2022). SARIMA es un modelo popular en la industria hotelera debido a que la estacionalidad es un componente inherente de la variable a proyectar, además que las predicciones a nivel mensual o trimestral son muy útiles a la hora de generar estrategias en este sector. En resumen, cuando una serie de tiempo financiera presenta componentes estacionales significativos, el modelo SARIMA suele ser una opción popular y bien fundamentada para generar proyecciones estables. (Brida & Risso, 2011)

La estructura SARIMA agrega el componente no estacional del ARIMA con componentes adicionales para establecer la estacionalidad. La notación del SARIMA es $(p,d,q)(P,D,Q)_s$ donde los parámetros en mayúscula corresponden a los parámetros estacionales:

- P – orden autorregresivo estacional.
- D – orden de diferenciación.
- Q – orden de medias móviles estacional.
- s – longitud del período estacional, es decir, el número de unidades de tiempo que abarca un ciclo completo de la temporada. (Molero, 2022)

Para la predicción de los ingresos para el operador hotelero se implementó un modelo SARIMA $(p,d,q) = (2,1,2)$ y $(P,D,Q,s) = (1,1,1,12)$ lo cual refleja tanto la dinámica de la serie a corto plazo como los ciclos anuales presentes en la serie. El entrenamiento se realizó con una base de datos mensual con una ventana de tiempo comprendida desde enero de 2021 y diciembre de 2024. El año 2024 se guardó como conjunto de validación

de la capacidad predictiva del modelo.

El proceso de Ajuste del modelo SARIMA incluyó los siguientes pasos:

1. Frecuencia temporal: se transformó la variable ingresos en una serie temporal mensual para asegurar la consistencia estacional en el modelo.
2. Validación Cruzada: Los datos fueron divididos en dos subconjuntos: entrenamiento (enero 2021 – diciembre 2023) y prueba (enero – diciembre 2024). Esta separación permite evaluar el desempeño real del modelo.
3. Evaluación: El modelo SARIMA presenta los siguientes estadísticos de evaluación.
 - R2: 0.38
 - MAPE: 6.85%

El modelo presenta un buen desempeño en términos de error porcentual, sin embargo, la variabilidad del ingreso es explicada un 38%, lo que indica la necesidad del uso de modelo SARIMAX donde se incluyan variables exógenas que ayuden a absorber esa volatilidad de la serie. La siguiente grafica muestra la calidad de la predicción del modelo SARIMA sobre los ingresos mensuales del operador hotelero.

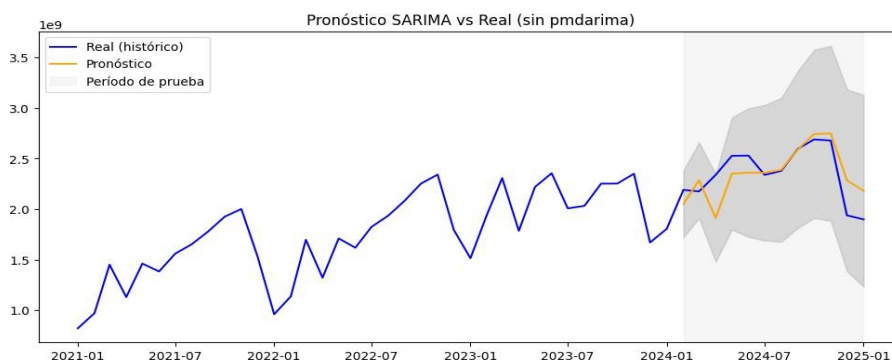


Figura 25 Predicción ingresos – SARIMA

Fuente: elaboración propia.

A pesar de que el modelo SARIMA logró capturar parcialmente la estacionalidad presente en los ingresos de operador hotelero, los resultados de la evolución del modelo muestran que sigue siendo una solución limitada. Por esta razón se descarta el modelo SARIMA y se opta por un modelo SARIMAX que incorpora variables exógenas relevantes.

10.4. Comparación De Modelos

La Tabla 22 presenta los principales resultados de desempeño obtenidos para cada uno de los modelos evaluados. El modelo SARIMAX se posiciona como el más sólido, alcanzando un R^2 de 0.8841 y un MAPE de 2.97%. Además, su p-valor en la prueba de Ljung-Box (0.077) indicando así que los residuos no presentan autocorrelación significativa, lo cual valida su ajuste desde un enfoque estadístico clásico (Box, Jenkins, Reinsel & Ljung, 2015). Aunque el modelo Prophet mostró un R^2 aún más alto (0.9508), su MAPE fue mayor (4.55%) y no se cuenta con una validación formal de autocorrelación de residuos, lo cual limita su análisis en términos clásicos. El modelo SARIMA tuvo un desempeño inferior ($R^2 = 0.3811$; MAPE = 6.85%) y además falló la prueba de Ljung-Box ($p = 0.0087$), lo que sugiere que deja residuos autocorrelacionados. El modelo Random Forest, aunque útil como enfoque no paramétrico, no cumplió con los criterios de validación estadística y tuvo un MAPE del 12%. Finalmente, los modelos ARIMA y ARIMAX presentaron un R^2 negativo y errores significativamente más altos, lo cual evidencia que no logran capturar adecuadamente la

dinámica de ingresos del operador hotelero, ya que no contemplan el componente de estacionalidad. En conclusión, el modelo SARIMAX representa la mejor alternativa metodológica, al combinar una alta precisión predictiva con cumplimiento riguroso de supuestos estadísticos fundamentales.

Tabla 222 Comparativo de desempeño de modelos predictivos de ingresos

Modelo	R²	MAPE (%)	Ljung-Box p-value	Ranking
SARIMAX (1,1,1)(0,1,0,12)	0.8841	2.97	0.077	1er lugar
Prophet	0.9508	4.55	—	2do lugar
SARIMA (2,1,2)(1,1,1,12)	0.3811	6.85	0.0087	3er lugar
Random Forest	0.6953	12.00	—	No cumple validación estadística
ARIMA (1,1,1)	-0.0156	8.54	0.4667	R ² negativo
ARIMAX (2,1,2)	-1.5982	16.13	0.5881	R ² negativo + MAPE alto

11. Modelo Definitivo De Predicción De Ingresos: SARIMAX

El modelo SARIMAX (Seasonal AutoRegressive Integrated Moving Average with eXogenous variables) es una extensión del modelo ARIMA que incorpora tanto comportamientos estacionales como variables exógenas. Este enfoque permite capturar de manera más precisa las dinámicas reales de series temporales económicas y financieras que están influenciadas por múltiples factores internos y externos (Hyndman & Athanasopoulos, 2018)

El modelo SARIMAX se compone de los siguientes elementos:

- **AR (AutoRegresivo):** captura la relación entre el valor actual de la serie y sus valores pasados.
- **I (Integrado):** permite diferenciar la serie para hacerla estacionaria.
- **MA (Media Móvil):** modela la dependencia entre el valor actual y los errores pasados.
- **S (Estacionalidad):** incorpora patrones que se repiten a intervalos regulares, como los ciclos mensuales o anuales.
- **X (Variables exógenas):** permite la inclusión de variables independientes que ayudan a explicar el comportamiento de la serie dependiente (Box, Jenkins, Reinsel & Ljung, 2015)

Para el caso del operador hotelero objeto de este trabajo de grado, el modelo Sarimax es pieza fundamental para el desarrollo de pronósticos futuros en las cifras financieras del operador considerando un importante complemento en su metodología actual que si bien es adecuada para el funcionamiento actual puede robustecerse con esta propuesta a la hora de estimar las proyecciones financieras para cada año. Es importante recordar que a lo largo de su historia, el acercamiento al pronóstico de este tipo de variables (Ingresos, Costos y Gastos) ha sido con formas lineales en las cuales simplemente se han utilizado las evoluciones de las mismas métricas internas del hotel, es decir, en ningún caso, el operador hotelero se ha apoyado en variables exógenas concretas tales como el crecimiento económico, las expectativas en términos de inflación, tasa de referencia del Banco de la República y tasa de cambio entre otras. Es aquí en donde se encontró tal vez el principal valor agregado que tiene el modelo y es que, a diferencia de otros modelos estadísticos autorregresivos mencionados con anterioridad, este, permite incorporar variables explicativas externas. De esta manera, el modelo al final permite alcanzar pronósticos con la mejor información de los dos mundos al contemplar tanto las variables o KPIs internos como las diferentes expectativas que se tienen sobre variables macroeconómicas como tal.

Más aún, también resulta clave mencionar bondades adicionales que tiene el modelo tales como: **i)** una captura bastante adecuada y una facilidad clara para adaptarse a comportamientos estacionales especialmente bajo periodicidades mensuales, que resultan propios de la actividad hotelera como tal, **ii)** buena capacidad de pronóstico de variables

objetivo en un contexto de bajo volumen de información que termina siendo fundamental considerando que la información provista por el operador hotelero en materia de Estados Financieros sólo contemplaba la historia en el periodo comprendido entre enero de 2021 y diciembre de 2024, es decir, el modelo tuvo que llegar a un excelente resultado con tan solo 48 datos, **iii)** flexibilidad en escenarios económicos complejos teniendo en cuenta que el modelo se adapta bien a sectores sensibles a factores externos como el turismo, donde los ingresos pueden verse afectados por choques económicos, regulatorios o sanitarios (Song, Li & Witt, 2012).

Ahora bien, para poder detallar el paso a paso de la creación y ajuste del modelo y a su vez, presentar los resultados del mismo, es importante resaltar los diferentes supuestos estadísticos con los que cuenta este modelo, así:

1. **Estacionariedad (en media varianza):** El modelo asume que la serie de tiempo es estacionaria o puede transformarse en estacionaria mediante diferenciación. Esto implica que las propiedades estadísticas de la serie, como la media y la varianza, son constantes a lo largo del tiempo. También se requiere estacionariedad estacional para las componentes estacionales. Box et al. (2015); Hyndman & Athanasopoulos (2018)
2. **Linealidad:** SARIMAX es un modelo lineal, por lo tanto, se asume que la relación entre los valores pasados, los errores pasados y las variables exógenas con la variable dependiente es lineal. Hamilton (1994); Stock & Watson (2015)
3. **Residuos (errores) con media cero, homocedásticos e incorrelacionados:** los residuos del modelo deben cumplir con tres condiciones: Tener **media cero**, ser

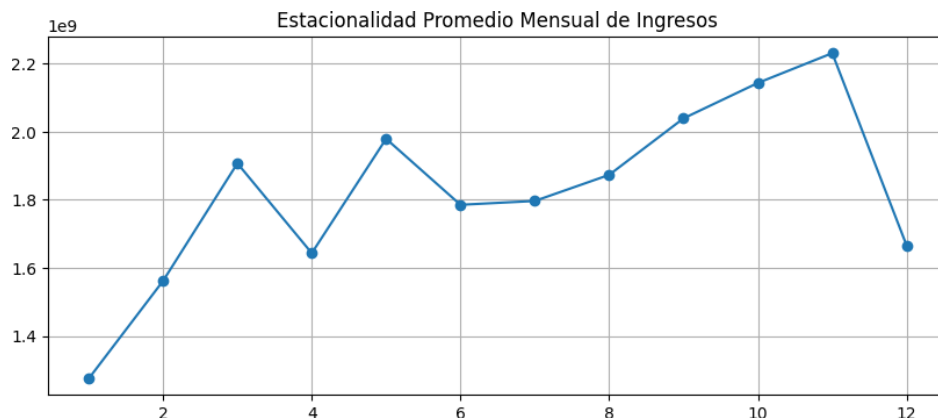
homocedásticos (varianza constante) y ser **no autocorrelacionados** (ruido blanco). Enders (2014)

4. **Estructura de rezagos correctamente especificada:** La elección de los parámetros (p , d , q) y sus equivalentes estacionales (P , D , Q , s) debe capturar adecuadamente la dinámica temporal. Una especificación incorrecta puede generar errores correlacionados, sesgo o sobreajuste. Box et al. (2015)

A continuación, se detalla el paso a paso de la creación del modelo:

11.1. Análisis Exploratorio Inicial

Dentro del análisis exploratorio inicial de la serie de tiempo, se graficó la serie original dando como resultado la figura 26, donde se detalla visualmente una tendencia creciente y algunas fluctuaciones periódicas (estacionales). En la imagen, se observa el promedio mensual de ingresos que se tiene en cada uno de los meses, los cuales están representados por los números 1 a 12 en el eje horizontal. De esta, cabe destacar como para el caso del operador hotelero meses tales como enero, junio, julio y diciembre son



de ingresos muy bajos lo cual es consecuente con un *core* de negocio enfocado en clientes corporativos y no en turismo.

Figura 26 Estacionalidad promedio Mensual Ingresos

Fuente: elaboración propia.

A su vez, dentro del análisis exploratorio se procedió a construir una matriz de correlaciones con el objetivo de determinar la relación entre las variables tanto endógenas como exógenas y de esta manera, empezar a tener bases que permitieran más adelante escoger qué variables se iban a considerar para el modelo. En la figura 27, se presenta la matriz de correlaciones obtenida.

```

 Correlación de variables con IngresosTotales:
Rev_Par 0.877499
Tarifa_Promedio 0.800912
Índice_Tarifas_x_AcomodaciónBogotá_Sencilla 0.786438
Índice_de_AlojamientoTotal_Nacional 0.775740
Índice_Tarifas_x_AcomodaciónBogotá_Doble 0.774064
IPC_MoM 0.772169
ISE_(Industria_Servicios) 0.769431
IPC_Servicios 0.751176
%_de_Ocupacion_Total 0.701883
%_de_Ocupacion_Disponible 0.698506
No._De_Personas_Alojadas 0.686695
Tasa_Banrep 0.681873
Índice_de_Alojamiento_Bogotá 0.529961
TRM 0.438569
Hab._Compl. 0.121054
Indice_de_Pax_Alojadas 0.003909
ISE_YoY(Indicador_de_Seguimiento_Económico) -0.424218
dtype: float64

```

Figura 27 Correlación de variables frente al ingreso

Fuente: elaboración propia.

A partir de lo anterior, como un punto de partida se tomaron como referencia aquellas variables que tuvieran una correlación positiva superior a 0.7 o una correlación negativa inferior a -0.7, con respecto a la variable a pronosticar que en este caso es “Ingresos Totales”.

11.2. Análisis De Causalidad De Granger

De manera complementaria a ese primer momento exploratorio en el cual se calcularon las diferentes correlaciones entre las variables tanto endógenas como exógenas, se procedió a hacer un análisis de Causalidad de Granger, que se observa en la figura 28, con el objetivo de más allá de entender si existía o no correlación entre las variables independientes y la variable dependiente, poder determinar si el comportamiento de las series temporales de las variables independientes verdaderamente servían para poder predecir el comportamiento futuro de la variable dependiente (Ingresos Totales).

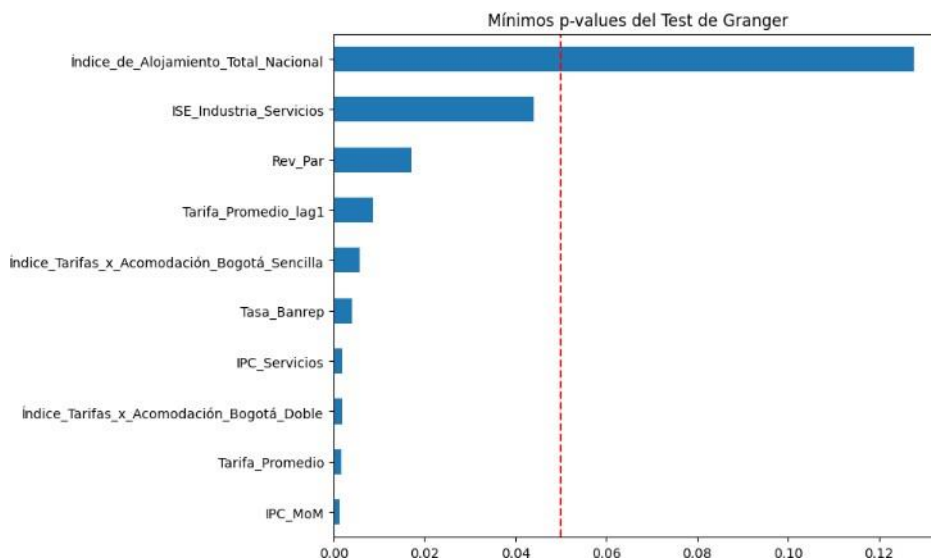


Figura 28 Causalidad de granger

Fuente: elaboración propia.

Al observar dichos resultados, se tomaron como referencia para la estructuración del modelo, las variables con un p-value inferior a 0.05, las cuales se encuentran antes de la línea roja punteada en la imagen.

11.3. Prueba De Estacionariedad

En línea con ir evaluando el cumplimiento de los supuestos estadísticos del modelo SARIMAX, se procedió a realizar una prueba de Dickey Fuller aumentada. Con respecto a esta prueba, se obtuvo un ADF estadístico de -2.8141 y un p-value de 0.0563, lo cual indicaba en primera instancia que la serie no era estacionaria. No obstante, al realizar la misma prueba sobre la primera diferencia de la base se obtuvo un ADF estadístico de -5.8448 y un p-value de 0.01; indicando así estacionariedad y la posibilidad de trabajar con un modelo de parámetros $d=1$ y $D=1$.

11.4. Autocorrelación (ACF) y Autocorrelación Parcial (PACF)

Al ejecutar la prueba ACF identificamos que la serie diferenciada está aprendiendo inmediatamente de los resultados del pasado para ajustar los valores siguientes, dando como sugerencia utilizar orden 1 ($q = 1$).

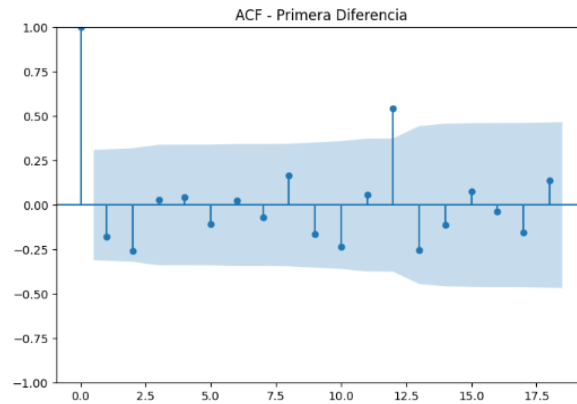


Figura 29 Prueba ACF

Fuente: elaboración propia.

Seguido con la prueba PACF se identifica la presencia de un componente autorregresivo, indicando la presencia de un componente autorregresivo de orden 1 ($p = 1$).

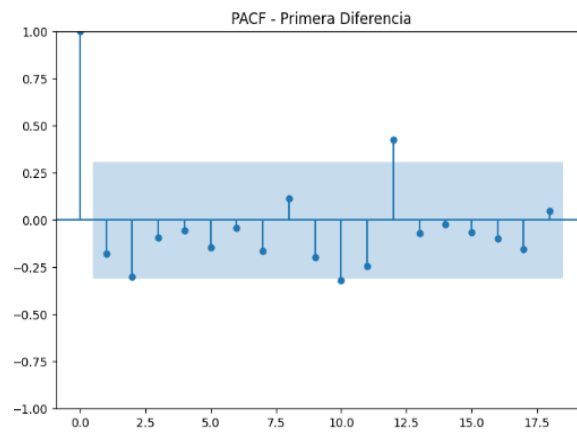


Figura 30 Prueba PACF

Fuente: elaboración propia.

11.5. Entrenamiento Del Modelo, Ajuste del Modelo y Pruebas De Validación

Con los resultados de las anteriores pruebas, se procede a entrenar el modelo SARIMAX (1,1,1) (0,1,0,12). Primero dividiendo la serie en entrenamiento (enero 2021 a mayo 2024) y en prueba (junio 2024 a diciembre 2024), generando las predicciones correspondientes para el periodo de prueba. Con los resultados se gráfica, como se muestra en la Figura 31, la serie de entrenamiento, predicciones y los intervalos de confianza.

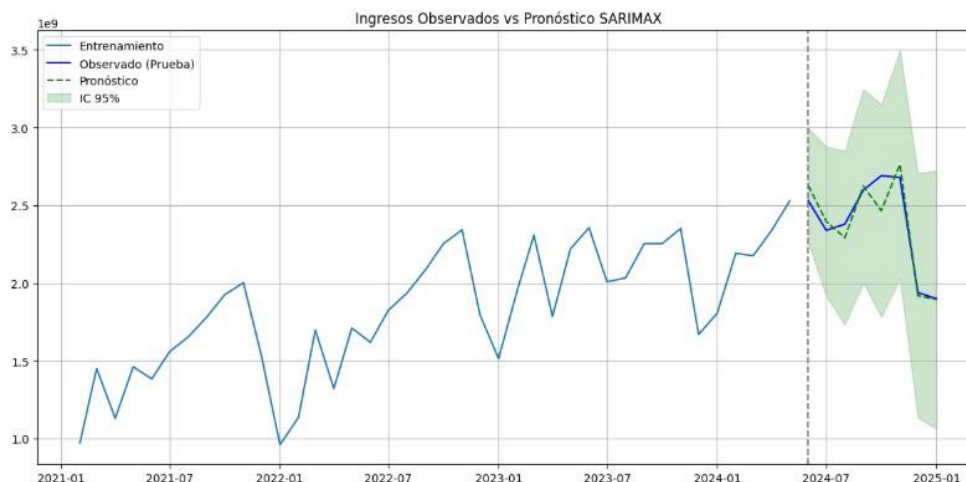


Figura 31 Pronóstico modelo SARIMAX

Fuente: elaboración propia.

Tal y como se puede observar en la imagen presentada anteriormente, los valores pronosticados para un periodo de seis meses muestran una similitud importante con los datos de prueba observados, razón por la cual, al ver tan ajustados estas predicciones se procedió a realizar varias pruebas que permitieran evaluar el riesgo de sobre ajuste del

modelo. Los resultados obtenidos en esta validación fueron los siguientes: MAPE en Test: 0.0297, R2 en Test: 0.8841, encontrando así que el modelo no estaba sobre ajustado y por el contrario generaliza correctamente a datos futuros.

Más aún, en línea con lo anterior, se procedió a realizar diversas pruebas estadísticas con el fin de determinar la validez del modelo dentro de las cuales se destacaron: Shapiro-Wilk, Jarque-Bera, D'agostino, Ljung-Box y Breusch-Pagan. A continuación, se presentan los resultados.

```

Resultados de las pruebas estadísticas sobre los residuos:
      Prueba  p-value  ¿Pasa? (p>0.05)
0  Shapiro-Wilk  0.463821      True
1  Jarque-Bera  0.378746      True
2  D'Agostino  0.239679      True
3  Ljung-Box  0.877829      True
4  Breusch-Pagan  0.148688      True
5  Breusch-Godfrey  0.884376      False

```

Figura 32 Resultado estadísticos de evaluación modelo SARIMAX

Fuente: elaboración propia.

Los resultados arrojan un modelo sólido que pasa todas las pruebas estadísticas.

11.6. Puesta En Producción Del Modelo

Una vez obtenidos los pronósticos del modelo SARIMAX junto con sus respectivos intervalos de confianza, se procedió a la fase de puesta en producción del modelo, con el fin de facilitar su reutilización y aplicación futura. Para ello, el modelo ajustado fue serializado y guardado en formato. pkl (Pickle), lo cual permite almacenar su estructura, parámetros y estado interno de forma eficiente. Posteriormente, se realizó una predicción específica para el mes de enero de 2025, utilizando las proyecciones más recientes de las

variables exógenas correspondientes a dicho periodo. Esta predicción final refleja la capacidad del modelo para generalizar y ofrecer estimaciones actualizadas bajo nuevas condiciones económicas.

11.7. Conclusiones Del Modelo SARIMAX

Durante el desarrollo del modelo de pronóstico de ingresos del operador hotelero, se comprobó que SARIMAX ofrecía ventajas relevantes:

- El análisis de estacionalidad evidenció ciclos mensuales consistentes.
- Las pruebas de causalidad de Granger confirmaron la relación significativa entre los ingresos y variables externas.
- La inclusión de rezagos mejoró la calidad del ajuste.
- Las métricas de validación (MAE, RMSE, MAPE) mostraron mejoras frente a modelos que no consideran estacionalidad ni exógenas.

Por estas razones, el modelo SARIMAX resultó ser una herramienta metodológicamente sólida y empíricamente eficaz para el pronóstico objeto de este trabajo.

12. Modelo De Anomalías Para Costos y Gastos

En la medida que el objetivo general de este trabajo de grado es el de lograr a través del uso de herramientas analíticas hacer predicciones de ingresos, costos y gastos de un operador hotelero, resulta fundamental observar cómo interactúan estos tres rubros en el desarrollo de este negocio. A lo largo de su existencia, el operador más allá de sólo centrarse en la evolución de sus ingresos dentro de las diferentes líneas de negocio, se ha enfocado claramente en un solo indicador: el Margen Neto.

De lo anterior, es importante mencionar que parte del por qué ajustar o buscar soluciones óptimas para predecir de una mejor manera estos rubros se ha vuelto meta del operador hotelero en los últimos años, pues dado un contexto de fuertes choques en las cadenas de suministro, de inflación al alza y tasas del Banco de la República al alza a niveles no vistos desde finales de los años 90, el margen neto se ha visto deteriorado de una manera importante (la meta del operador hotelero en términos de margen neto es del 30% al año).

Así las cosas, para poder generar un verdadero valor agregado al control financiero y

a la toma de decisiones de inversión, no sólo se debe tener un enfoque en la predicción de ingresos, costos y gastos sino también se debe complementar este esfuerzo con la presentación de herramientas que permitan de una u otra forma, anticipar sorpresas negativas en rubros como los costos y los gastos, de tal manera que el operador hotelero pueda tener planes de contingencia que les permitan sortear estos eventos y así mismo no desviarse en el porcentaje de cumplimiento de sus metas para el año. Por esta razón, de manera complementaria a la generación de modelos de pronóstico desarrollados tales como el SARIMAX, previamente mencionado, se implementaron modelos de detección de anomalías dentro de la base de datos de costos y gastos del hotel. A continuación, se detallan los pasos realizados.

12.1. Identificación De Modelos De Detección De Anomalías Útiles Para Este Caso de Estudio

Para este caso en particular se tomaron como referencia tres modelos de detección de anomalías: modelo Z-score, modelo Isolation Forest y modelo LOF (Local Outlier Factor). Para objeto de este trabajo de grado, sólo se profundizará en los dos últimos modelos mencionados.

El Modelo Isolation Forest es un modelo de aprendizaje automático no supervisado diseñado para identificar anomalías dentro de un conjunto de datos. Su lógica se basa en que los puntos anómalos son pocos y están ubicados en regiones menos densas, lo que facilita su aislamiento frente al resto de los datos normales.

A diferencia de otros enfoques que miden la densidad o la distancia, este modelo se enfoca en separar los puntos de manera aleatoria. Construye múltiples árboles de decisión, llamados árboles de aislamiento, donde los datos se dividen de forma sucesiva

mediante cortes aleatorios en sus atributos. Los puntos que requieren menos divisiones para quedar completamente aislados son considerados anomalías, ya que se separan fácilmente del resto del conjunto.

La eficiencia del modelo se debe a que no necesita calcular distancias entre puntos ni densidades locales. Además, tiene una buena capacidad para trabajar con grandes volúmenes de datos y detectar anomalías incluso en conjuntos de datos de alta dimensión. Su entrenamiento es rápido y su interpretación es directa a partir de la profundidad promedio con la que un punto es aislado en los árboles.

En resumen, este modelo identifica anomalías de forma eficiente al analizar qué tan rápido puede aislar un punto respecto a los demás, lo que lo convierte en una técnica efectiva, robusta y escalable para detectar comportamientos atípicos.

El modelo LOF (Local Outlier Factor) es una técnica no supervisada para la detección de anomalías que se basa en el análisis de la densidad de los datos. Su objetivo principal es identificar observaciones que se encuentren en zonas significativamente menos densas en comparación con sus vecinos más cercanos.

La idea central del modelo es que un punto puede ser considerado anómalo si su entorno inmediato tiene una densidad de datos notablemente mayor. Para lograr esto, LOF calcula una puntuación comparando la densidad local de un punto con la de sus vecinos más cercanos. Esta densidad se estima utilizando distancias relativas y el número de vecinos especificado previamente.

El resultado de este análisis es un valor numérico, conocido como factor LOF. Un valor cercano a uno indica que el punto tiene un comportamiento similar al de su entorno. En cambio, si el valor es significativamente mayor que uno, se considera que el

punto está aislado respecto a sus vecinos, y por lo tanto puede tratarse de una anomalía.

Esta técnica es especialmente útil en situaciones donde los datos tienen estructuras locales complejas, ya que no depende de un patrón global, sino de relaciones vecinales. LOF es capaz de detectar outliers que no necesariamente presentan valores extremos, pero que se comportan de forma inusual en su contexto más cercano.

En resumen, este modelo identifica anomalías al detectar desviaciones en la densidad local de los datos, ofreciendo una manera eficaz de encontrar valores atípicos contextualizados dentro del conjunto.

12.2. Identificación De Anomalías En Costos y Gastos Totales Del Operador

Hotelero

Una vez escogidos los dos modelos con para analizar la existencia o no de anomalías en los costos y gastos del operador hotelero. Se procedió a correr estos modelos con la base de datos en el periodo correspondiente entre enero de 2021 y diciembre de 2024. A continuación, se presentan de manera gráfica los resultados.



Figura 33 Resultado Modelo de anomalías Costos Totales

Fuente: elaboración propia.



Figura 34 Resultado Modelo de anomalías Gastos Totales

Fuente: elaboración propia.

A partir de los resultados observados de manera gráfica, se puede ver como para el caso de los costos totales tanto el modelo *Isolation Forest* como el modelo *LOF* identificaron la existencia de anomalías para los meses de enero de 2022-2024, agosto 2022 y febrero de 2023; mientras que, en el caso de los gastos totales, se presentaron de manera consistente para los meses de enero y diciembre de 2022-2024.

12.3. Análisis De Causalidad Anomalías Detectadas Por Cada Línea de Negocio

En búsqueda de poder generar valor agregado al operador hotelero y utilizar estos hallazgos en términos de detección de anomalías, se decidió profundizar en este análisis

de tal forma que se pudiera concluir la razones por las cuales se presentaban esas anomalías en estos rubros a lo largo del tiempo. Por esta razón, se implementaron los mismos modelos de detección de anomalías pero en esta ocasión, en cada una de las líneas de negocio que se tienen (Alojamiento, Alimentos y Bebidas, Comunicaciones, Lavandería y Otros), con el fin de encontrar una especie de causalidad de tal manera que en meses determinados, a la hora de realizar la planeación financiera se tengan alertas y planes de contingencias para no permitir que dichas anomalías afecten el margen de rentabilidad del operador tal y como se había mencionado previamente. A continuación, se presentan los hallazgos obtenidos producto de dicho análisis.

Mes y año	Costos Totales (Costo)	Otros Gastos + Costos Alojamiento Suites (Costo)	Costo A&B (Costo)	Costo Comunicaciones (Costo)	Costo Lavandería (Costo)	Costo Otros Ingresos o menores (Costo)	Gastos Totales (Gasto)	Total Gasto Alojamiento Suites (Gasto)	Total Gasto A&B (Gasto)	Total Gasto Comunicaciones (Gasto)	Total Gasto Lavandería (Gasto)	Total Gastos Otros Ingresos (Gasto)	Gastos Sobreventas (Gasto)
1 ene 2022	▶	▶			▶		▶	▶	▶			▶	
1 feb 2022	▶	▶	▶				▶		▶				
1 ago 2022	▶			▶		▶							
1 mar 2023	▶	▶					▶				▶		
1 nov 2023	▶		▶				▶		▶			▶	
1 ene 2024							▶	▶					
1 dic 2024							▶	▶					

Figura 35 Resultado Modelo de anomalías líneas de negocio

Fuente: elaboración propia.

Las etiquetas en verde presentes en la tabla muestran una existencia de causalidad de determinada línea de negocio, dentro de la materialización de la anomalía a nivel de las líneas generales (costos totales y gastos totales).

12.4. Integración Con El Modelo Predictivo De Ingresos

Finalmente, como se mencionó con anterioridad, la idea de la implementación de estos modelos de detección de anomalías es generar valor agregado en la toma de decisiones en un contexto en el cual los datos observados puedan complementar a las predicciones realizadas en términos ingresos. Por esta razón, la información previamente presentada referente a los modelos de detección de anomalías y los hallazgos en términos de causalidad dentro de las diferentes variables generales se incluyen dentro del entregable de tablero de visualización al operador hotelero. De esta forma, el área encargada de la planeación financiera y de control de presupuesto tendrán más información que les permita definir esa guía de ruta a la hora de estructurar planes de contingencia para que estas anomalías no afecten el margen del negocio.

13. Tablero De Control

El tablero de control diseñado y desarrollado para el operador hotelero tiene como objetivo principal centralizar y visualizar estratégicamente los indicadores clave, así como los modelos implementados para mejorar la toma de decisiones a futuro. Su diseño se basó tanto en criterios técnicos de visualización de datos como en las necesidades importantes del negocio, y tiene la siguiente estructura:

1. **Análisis descriptivo:** Esta sección presenta análisis estadístico y de estacionalidad de los indicadores core del negocio y de las variables exógenas. Incluye gráficos de tendencias mensuales y anuales para identificar patrones globales y estacionales.
2. **Análisis financiero:** Esta parte del tablero descompone los costos y gastos, mostrando su participación relativa frente a los ingresos. Se incluyen gráficos tipo Pareto, matrices de comparación y barras apiladas para facilitar la identificación de los rubros con mayor impacto en la operación.
3. **Análisis temporal y de anomalías:** Esta sección del tablero combina el análisis temporal con la detección automatizada de comportamientos atípicos, permitiendo al usuario no solo comprender la evolución de los ingresos, costos y gastos del operador hotelero, sino también identificar desviaciones significativas que podrían afectar la interpretación financiera. Mediante modelos como Isolation Forest y Local Outlier Factor, se identifican anomalías en el comportamiento de ingresos, costos y gastos. De igual manera, se incorporan las predicciones del

modelo SARIMAX por medio de la integración del modelo entrenado en formato pkl.

4. Glosario: Con el objetivo de garantizar la usabilidad y el entendimiento del tablero, se implementa una sección de glosario de términos de indicadores clave de negocio. En estas definiciones clave se encuentra tanto el significado teórico como las fórmulas de cálculo para dar contexto completo de cada variable.

Dentro de la estructura del tablero, se garantiza que sea escalable y de fácil actualización, para que esta herramienta tenga alto impacto en la formulación de estrategias del operador hotelero. Lo anterior implica la automatización de predicciones y la transformación de datos mediante la implementación de medidas DAX y objetos Python que complementan el análisis y garantizan su veracidad.

14. Conclusiones y Recomendaciones

El presente proyecto permitió validar el potencial de la ciencia de datos como herramienta clave para fortalecer la toma de decisiones financieras en el sector hotelero, específicamente para el operador. A partir del análisis de los estados financieros históricos, complementado con indicadores de la industria y variables exógenas, se identificó un modelo financiero de negocio estable, por consecuente se desarrolló una solución integral orientada a mejorar la capacidad de planeación financiera y control del margen de rentabilidad, agregando valor de manera inmediata. Durante el proceso, se evaluaron diversas metodologías de modelado, incluyendo regresión lineal, Random Forest, ARIMA, ARIMAX y SARIMA, concluyendo que el modelo SARIMAX ofrecía el mejor desempeño. Este modelo no solo permitió capturar la estacionalidad característica del sector hotelero, sino también integrar de manera eficiente variables exógenas relevantes, generando proyecciones de ingresos mensuales más robustas, ajustadas a las dinámicas internas y del entorno.

Asimismo, a partir del análisis vertical de los estados financieros, fue posible proyectar los costos y gastos asociados, proporcionando una visión integral de la rentabilidad esperada en el mediano plazo. Esto nos permite concluir que al implementar la analítica en el sector empresarial, para áreas financieras, siempre será necesario primero determinar cuál es el objetivo financiero (aumentar ingresos, disminuir gastos, optimizar recursos, mantener el margen de rentabilidad, etc) ya que esto determinará el camino a seguir en el

proceso analítico.

En una industria altamente expuesta a choques por la cadena de suministro, la implementación del modelo de detección de anomalías demostró ser una herramienta de valor agregado al permitir identificar desviaciones en las principales líneas de costos y gastos, facilitando una gestión proactiva y oportuna sobre el margen financiero del negocio, a partir de la generación de información valiosa para la estructuración de planes de contingencia.

Adicionalmente, se evidenció la importancia de contar con información de calidad, estructurada y periódica que permita alimentar tanto modelos predictivos como de anomalías, destacando la necesidad de fortalecer la cultura de la toma de decisiones basada en datos al interior de la organización para garantizar la efectividad de estas herramientas.

Finalmente, la combinación de los modelos predictivos, el sistema de alertas por anomalías y el tablero de control interactivo permitió consolidar una solución integral enfocada en una implementación sencilla y amigable con el operador hotelero con el objetivo de potenciar las capacidades analíticas y de monitoreo del operador, brindándole una herramienta ágil, intuitiva y basada en la analítica para así, enfocar la planeación estratégica en aumentar los ingresos al mismo tiempo que se controlan los costos y gastos para mantener el margen.

Referencias Bibliográficas

- Andriawan, Z.A., Purnama, S.R., Darmawan, A.S., Ricko, Wibowo, A., Sugiharto, A., & Wijayanto, F. (2020). Prediction of Hotel Booking Cancellation using CRISP-DM. 2020 4th International Conference on Informatics and Computational Sciences (ICICoS), 1-6.
- Banco de la República. (2023). Análisis económico del turismo en Bogotá y Colombia. Recuperado de: <https://www.ccb.org.co/informacion-especializada/observatorio/analisis-economico/turismo>
- Brida, J. G., & Risso, W. A. (2011). Tourism Demand Forecasting with SARIMA Models – the Case of South Tyrol. Universidad de la República de Uruguay.
- Carrillo, E. & Gómez, Y. (2017). Medición de la eficiencia de hoteles: caso de estudio en Colombia Revista Virtual Universidad Católica del Norte, 51,143-155. Recuperado de: <http://revistavirtual.ucn.edu.co/index.php/RevistaUCN/article/view/849/1367>
- Mayenberger, A.S. (2017). Evaluación operacional y financiera: aplicada a la industria hotelera - 4ta. Edición. Universidad Externado. 9789587727395.
- Molero, A. A. (2022). SERIES TEMPORALES APLICADAS. Universidad de Granada.
- Noble, J. (2024, 24 de mayo). ¿Qué son los modelos ARIMA? IBM. <https://www.ibm.com/mx-es/think/topics/arima-model>
- OMT. (2022). Impacto económico de la crisis de la COVID-19 en el turismo. Recuperado de: <https://www.unwto.org/tourism-data/international-tourism-and-covid-19>
- Ortiz Martín, R. (2022). Metodología ARIMAX aplicada al análisis de la evolución de datos de suicidios en España (Trabajo de Fin de Máster). Departamento de Estadística e

Investigación Operativa, Universidad de Barcelona.

Pardo, Navarro. (2020). *Analítica de datos para toma de decisiones en las pymes y los micro establecimientos del sector turístico de Colombia 2015 – 2019*. Fundación Universitaria

Compensar. Recuperado de:

<https://repositoriocrai.ucompensar.edu.co/server/api/core/bitstreams/aff75f6-6990-41a1-9201-7bb5c9c90f9c/content>

Sharma, R., & Choudhary, A. (2022). Forecasting the Consumer Price Index using SARIMAX Modeling. *International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)*, 13(2), 204-210. <https://doi.org/10.48175/IJARSCT-3987>

Declaración de originalidad y autenticidad

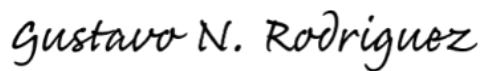
Declaramos bajo la gravedad del juramento, que hemos escrito el presente proyecto aplicado empresarial (PAE), en la propuesta de solución a una problemática en el campo de conocimientos del programa Maestría por nuestra propia cuenta y que, por lo tanto, su contenido es original.

Declaramos que hemos indicado clara y precisamente todas las fuentes directas e indirectas de información y que este PAE no ha sido entregado a ninguna otra institución con fines de calificación o publicación.



Firma Digital

Ana María Mosquera Morales




Firma Digital

Gustavo Nicolas Rodríguez Mejía



Firma Digital

Linda Liceth López López

A handwritten signature in black ink, consisting of several overlapping loops and a vertical line extending downwards.

Firma Digital

Sara Juliana Martínez López

Firmado en Bogotá, D.C el 25 de mayo de 2025

Declaración de exoneración de responsabilidad

Declaramos que la responsabilidad intelectual del presente trabajo es exclusivamente de sus autores. La Universidad del Rosario no se hace responsable de contenidos, opiniones o ideologías expresadas total o parcialmente en él.



Firma Digital

Ana María Mosquera Morales

Gustavo N. Rodríguez
Firma Digital

Gustavo Nicolas Rodríguez Mejía



Firma Digital

Linda Liceth López López

A handwritten signature in black ink, consisting of several overlapping loops and a vertical line extending downwards from the right side.

Firma Digital

Sara Juliana Martínez López

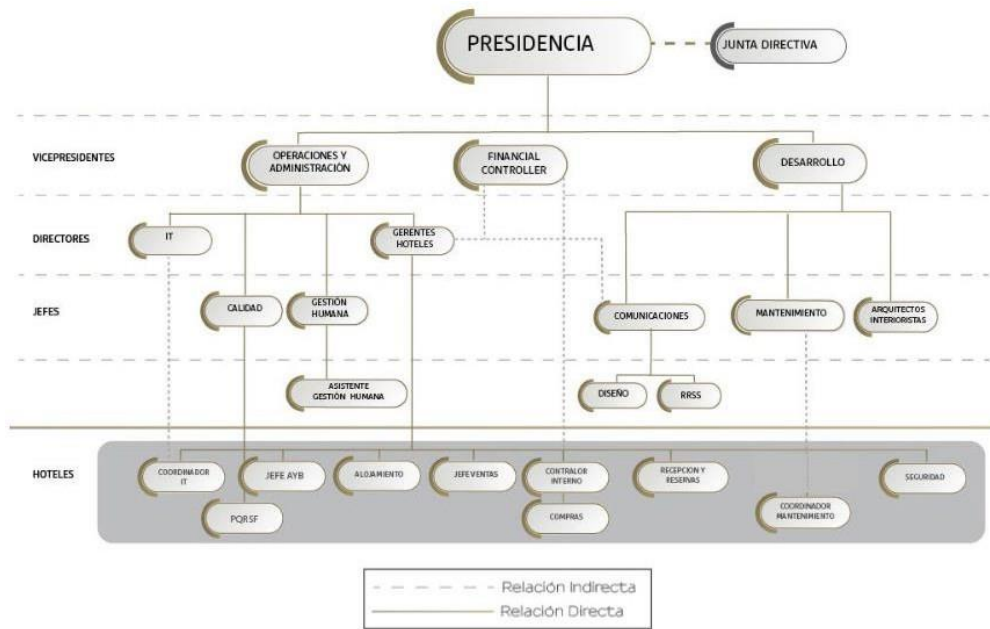
Firmado en Bogotá, D.C el 25 de mayo de 2025

Anexos

Anexo 1. Validación de estrategia de PE a través de LEGO Serious Play



Anexo 2. Organigrama Corporativo Operador Hotelero



Anexo 3. Entrevista semiestructurada: Mapeo de procesos Operador Hotelero

¿Cómo es el flujo de la información, antes de que lleguen a las bases de datos que vamos a utilizar en el modelo?

1. Infraestructura y Almacenamiento

¿Dónde se almacenan actualmente los datos?

¿Qué herramientas o sistemas utilizan para gestionar los datos?

¿Qué tipos de datos manejan? (estructurados, no estructurados, semiestructurados)

¿Tienen un catálogo o inventario de datos?

2. Calidad de los Datos

¿Qué tan actualizados están sus datos?

¿Han identificado problemas comunes como duplicados, datos incompletos o inconsistencias?

¿Existen procesos para validar y limpiar los datos?

¿Qué métricas utilizan para evaluar la calidad de los datos?

3. Acceso y Seguridad

¿Quién tiene acceso a los datos y cómo se controla ese acceso?

¿Utilizan políticas de seguridad o cifrado para proteger los datos sensibles?

¿Cumplen con alguna normativa de protección de datos?

4. Procesos de Gestión

¿Cómo se recopilan los datos actualmente?

¿Tienen documentados los flujos de datos desde su origen hasta su uso final?

¿Existen responsables asignados para la gestión de datos?

5. Integración y Uso

¿Cómo se integran los datos entre los diferentes sistemas o departamentos?

¿Qué tan fácil es acceder a los datos para su análisis o toma de decisiones?

¿Qué herramientas utilizan para el análisis o visualización de datos?

6. Problemas y Desafíos

¿Cuáles son los principales retos que enfrentan con sus datos?

¿Han identificado cuellos de botella en el flujo de datos o en los procesos analíticos?

¿Qué procesos o sistemas consideran obsoletos o ineficientes?

7. Estrategia y Futuro

¿Tienen una estrategia de datos definida?

¿Qué objetivos tienen para el uso de datos en el futuro?

8. Gobernanza de Datos

¿Existe un marco de gobernanza de datos en la empresa?

¿Hay políticas claras sobre el ciclo de vida de los datos (creación, uso, almacenamiento, eliminación)?

¿Cómo se asegura la calidad y la responsabilidad en los datos compartidos externamente?