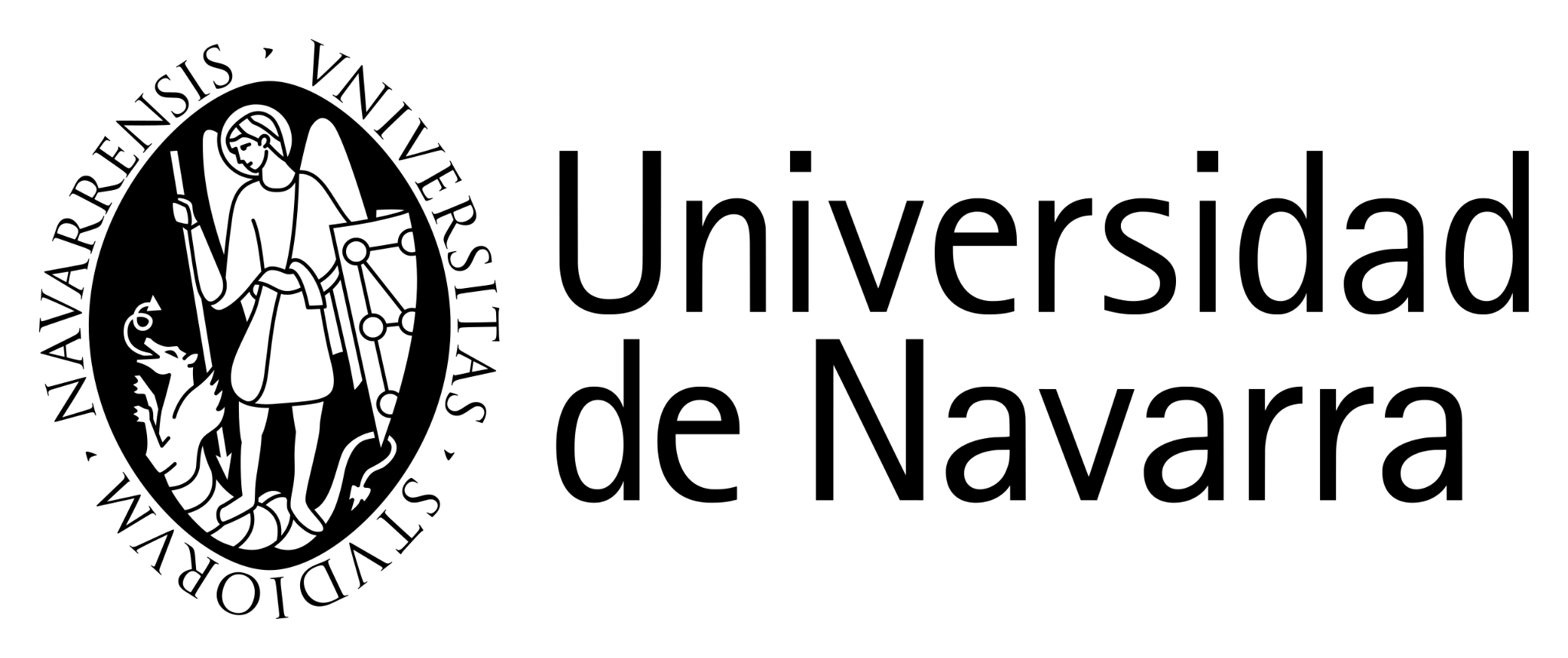
TFM

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**TRABAJO FIN DE GRADO**

Máster en Big Data Science

**Análisis Predictivo de la Satisfacción de Pasajero en Aerolíneas**

Año académico 2025 - 2026

Autores: Diego Rafael de Lemos, Leyre de la Calzada, Carmen Casado, Pablo Benavides, Dick Rogeir Lizana

Fecha: Abril de 2025

**Contenido**

[Resumen Ejecutivo 3](#_gjdgxs)

[Agradecimientos 4](#_30j0zll)

[1. Introducción 5](#_1fob9te)

[1.1. Motivación 6](#_3znysh7)

[1.2. Objetivos 7](#_2et92p0)

[2. Estado del Arte 8](#_3dy6vkm)

[2.1. Introducción al Estado del Arte 8](#_w7wze3o4urit)

[2.2. Conceptualización de la Satisfacción del Pasajero 8](#_grti6383e7a7)

[2.3. Relación entre Satisfacción del Pasajero y Lealtad a la Aerolínea 9](#_dmnqgi11k5io)

[2.4. Evolución del Concepto de Satisfacción: Factores Psicológicos y Contextuales 10](#_m32s6xy36pz5)

[2.5. Métodos de Medición de la Satisfacción del Pasajero 11](#_a1vqsvwpsgkp)

[2.6. Factores de Contexto: Impacto de la Pandemia en la Satisfacción del Pasajero 12](#_kqbod7k1cd5x)

[2.7. Modelos y Técnicas existentes 13](#_c1tb13bj8uu2)

[3. Materiales y Métodos 15](#_u8y14nmhks51)

[3.1. Estructura de los datos 15](#_a5zo7f78q0a6)

[3.2. Descripción de las variables obtenidas 15](#_qmjeadspp1dd)

[3.3. Análisis Exploratorio de Variables 17](#_awglqu7kv299)

[3.4. Selección de variables representativas y reducción de dimensionalidad 30](#_tanmlt2hgg44)

[3.5. Clustering 32](#_wvh0mubzspof)

[3.6. Desarrollo del Modelo 48](#_3cyvw0l3s3pk)

[3.7. Evaluación del Modelo 55](#_18ekl33tx327)

[**4. ROI de los Casos de Uso 60**](#_quaacslcbycg)

[4.2. Desglose de inversiones 60](#_b6r30f5jt69n)

[4.3. Estimación de Beneficios 64](#_3ckjbgficwx1)

[4.4. Cálculo de ROI y VAN 66](#_3we37hoi5ppb)

[4.5. Análisis de sensibilidad 68](#_bdm43bjx496i)

[4.6. Roadmap de implementación 69](#_5bmy3xvbypkd)

[4.7. Conclusiones del ROI ajustado 70](#_2sym3zo0exk2)

[5.1. Principales resultados/conclusiones e implicaciones 73](#_4d34og8)

[5.2. Limitaciones y recomendaciones para futuros trabajos 74](#_2s8eyo1)

[6. Bibliografía 76](#_17dp8vu)

[6.1. Publicaciones 76](#_3rdcrjn)

[6.2. Páginas web 77](#_26in1rg)

[7. Anexo: Código 80](#_5zngffgi2akn)

# Resumen Ejecutivo

El presente trabajo tiene como objetivo principal centrarse en la exploración y modelado de los factores clave que influyen en la percepción de satisfacción de los pasajeros en el sector aeronáutico. Para ello, se emplea un conjunto de datos compuesto por 129,880 observaciones y 25 variables, las cuales reflejan diversas dimensiones de las experiencias de los pasajeros durante su viaje. A través de la implementación de técnicas avanzadas de análisis de datos y la construcción de modelos predictivos robustos, se busca estimar con precisión los niveles de satisfacción de los pasajeros.

Los resultados que se generan permiten una mayor comprensión de las expectativas y necesidades de los pasajeros, proporcionando además una base sólida para el diseño de estrategias de mejora continua en un entorno altamente competitivo como el del transporte aéreo. Este estudio combina un enfoque técnico y práctico, abarcando tanto el análisis de datos como su aplicación directa en la industria de la aviación comercial. Mediante la exploración de variables como la comodidad del pasajero, la calidad del servicio, la puntualidad de los vuelos y la experiencia global del pasajero durante su viaje, se identifican patrones y tendencias clave que posibilitan una comprensión profunda de las expectativas y necesidades de los usuarios.

Además de aportar soluciones a un problema de gran relevancia en el sector aeronáutico, este trabajo demuestra el poder de los datos como herramienta estratégica para transformar la experiencia del pasajero, optimizar procesos operativos y fortalecer la fidelización. Los resultados que se obtienen constituyen una base cuantitativa sólida que respalda la toma de decisiones.

# Agradecimientos

El desarrollo de este trabajo se hace posible gracias a la colaboración, el apoyo y el entusiasmo de quienes han participado en este proceso.

En primer lugar, se agradece a la Universidad de Navarra por proporcionar las herramientas, el conocimiento y el entorno necesario para llevar a cabo esta investigación. Su compromiso con la excelencia académica y la innovación constituye una fuente constante de inspiración.

Asimismo, se expresa gratitud hacia el equipo de mentores, profesores y colegas, cuyas orientaciones y aportaciones enriquecen significativamente este trabajo. Sus perspectivas y críticas constructivas permiten refinar las ideas y superar los desafíos que surgen a lo largo del proceso.

A todos ellos, se extiende el más sincero agradecimiento por ser parte de este logro.

# Introducción

En los últimos años, el sector aeronáutico adquiere una relevancia estratégica tanto en términos económicos como sociales. Más allá de facilitar la movilidad global, las aerolíneas se consolidan como actores clave en el desarrollo del turismo, el comercio y la cohesión territorial. Sin embargo, esta creciente importancia se acompaña de nuevas exigencias por parte de los usuarios, quienes requieren experiencias de viaje más seguras, eficientes, personalizadas y, sobre todo, humanas.

La digitalización del sector y el volumen masivo de datos que generan los pasajeros en cada etapa del viaje posibilitan el uso de tecnologías avanzadas como la inteligencia artificial y, en particular, las redes neuronales. Estas herramientas permiten no solo identificar patrones de comportamiento y necesidades, sino también predecir niveles de satisfacción y anticipar posibles fallos en el servicio. Este potencial tecnológico se convierte en una ventaja competitiva en un contexto donde la fidelidad del pasajero resulta cada vez más volátil y las experiencias negativas se difunden rápidamente a través de redes sociales y medios digitales.

El 26 de marzo de 2025, un incidente ocurrido en el Aeropuerto de Palma de Mallorca pone en evidencia las carencias aún existentes en la atención al pasajero. Un grupo de 189 jubilados gallegos, participantes en un viaje del programa Imserso, queda retenido durante más de seis horas debido a una avería técnica en un vuelo de la aerolínea AlbaStar. Según lo informado por Cadena SER (2025), “*a la falta de información clara se sumaron vales de comida insuficientes, nula asistencia personal y dificultades para presentar reclamaciones digitales*”, lo que agrava la situación, especialmente para personas mayores o con necesidades especiales .

Este tipo de situaciones no solo afectan el bienestar de los pasajeros, sino que también evidencian oportunidades de mejora en la gestión de crisis, la comunicación con el pasajero y la personalización del servicio. En un sector altamente competitivo como el aéreo, la capacidad de respuesta ante imprevistos y el trato al usuario se establecen como factores determinantes para la reputación de marca y la sostenibilidad del modelo de negocio.

## Motivación

Este trabajo surge de la necesidad de encontrar soluciones innovadoras a problemas recurrentes en el sector aeronáutico, tomando como punto de partida una base de datos real que incluye variables relevantes como edad, tipo de pasajero, número de vuelos al año, clase del viaje, retrasos y nivel de satisfacción. A través del uso de redes neuronales, se plantea desarrollar una herramienta predictiva que posibilite anticipar el nivel de satisfacción de un pasajero incluso antes de abordar el vuelo.

La propuesta no se limita a mejorar los indicadores de satisfacción, sino que persigue traducir los datos en acciones concretas: ofrecer asistencia anticipada a pasajeros vulnerables, establecer protocolos de comunicación adaptados a distintos perfiles, optimizar la entrega de recursos durante retrasos y facilitar canales de reclamación accesibles para todos los segmentos de usuarios. Todo ello se traduce en una mejora de la experiencia del pasajero y, en consecuencia, en un incremento del valor percibido del servicio.

Desde una visión empresarial, la implementación de este tipo de soluciones representa un paso hacia la automatización inteligente de la gestión operativa y hacia una relación más empática entre la aerolínea y sus pasajeros. Además, se transforma lo que tradicionalmente constituye un coste, como la atención al pasajero en situaciones críticas, en una inversión estratégica que refuerza la fidelización, reduce quejas y fortalece la imagen de marca.

Aprovechar el potencial de la inteligencia artificial no debe concebirse sólo como una mejora técnica, sino como un cambio de enfoque: se sitúa al pasajero en el centro del servicio, utilizando la tecnología como puente para entenderlo, anticiparlo y cuidarlo. Este trabajo propone, en definitiva, una idea de negocio viable y escalable que posibilita la modernización de las operaciones de las aerolíneas, avanzando hacia un modelo más humano y eficiente de atención al pasajero.

## Objetivos

a) Desarrollar un modelo predictivo para estimar el nivel de satisfacción de los pasajeros en aerolíneas, utilizando técnicas avanzadas de ciencia de datos y machine learning.

b) Impulsar la optimización de la experiencia del pasajero en el sector aeronáutico mediante el análisis y comprensión de variables clave que influyen en la satisfacción de los pasajeros, garantizando resultados que impulsen la fidelización y mejoren la competitividad de las aerolíneas.

# Estado del Arte

## Introducción al Estado del Arte

La satisfacción del pasajero en el ámbito de la aviación se analiza ampliamente en la comunidad académica debido a su implicación directa en la sostenibilidad a largo plazo de las aerolíneas. La comprensión de los factores que determinan esta satisfacción resulta esencial para el desarrollo de estrategias que favorecen la lealtad del pasajero y optimizan la gestión de servicios aéreos.

No obstante, la literatura presenta una notable divergencia en los hallazgos, lo que evidencia la necesidad de un enfoque sistemático y exhaustivo que posibilite la identificación de los factores críticos de éxito en diferentes contextos geográficos y económicos. Asimismo, resulta fundamental considerar cómo los cambios recientes en el entorno global, como la pandemia de COVID-19, modifican las percepciones de calidad y lealtad en el sector aéreo.

El presente apartado aborda el desarrollo teórico y empírico de la satisfacción del pasajero, organizando la información relevante para un análisis profundo y contextualizado. Se incluyen la conceptualización de la satisfacción del pasajero, su relación con la lealtad hacia la aerolínea, la evolución del concepto desde una perspectiva psicológica y contextual, los métodos de medición utilizados y el impacto de situaciones recientes, como la crisis sanitaria global, en la percepción de la calidad del servicio.

De esta manera, se busca ofrecer una visión integral y crítica de los principales estudios realizados, identificando áreas de consenso y puntos de debate que sustentan el enfoque metodológico del presente trabajo.

## Conceptualización de la Satisfacción del Pasajero

La satisfacción del pasajero constituye un constructo complejo y multidimensional que despierta un notable interés en el ámbito de la aviación comercial. Como señala Singh (2021), “*se define como la percepción del pasajero sobre la calidad general del servicio recibido en comparación con sus expectativas previas*”.

El concepto de calidad del servicio percibida (*Perceived Service Quality*, PSQ) se define como “*la evaluación global que realiza el pasajero acerca de la excelencia de los servicios proporcionados por la aerolínea*” (Eshaghi et al., 2024). En este sentido, el valor percibido (*Perceived Value*, PV) se describe como “*la relación entre los costos asumidos (tiempo, dinero y esfuerzo) y los beneficios obtenidos*” (Tam et al., 2022). Estos conceptos resultan esenciales para comprender cómo los pasajeros valoran los servicios ofrecidos.

En el contexto de las aerolíneas de bajo coste (*low-cost*), estudios recientes evidencian variaciones significativas en la percepción de calidad en comparación con las aerolíneas tradicionales. Como señalan Shen y Yahya (2021), “*la percepción de calidad en aerolíneas low-cost del Sudeste Asiático se ve influenciada principalmente por el precio y la eficiencia operativa*”, mientras que en aerolíneas tradicionales, “*factores como la comodidad y el trato personalizado adquieren mayor relevancia*” (Hassan & Salem, 2022).

Por otra parte, la literatura examina el papel de la lealtad del pasajero como resultado directo de la satisfacción. Como señalan Akamavi et al. (2015), “*la satisfacción del pasajero mantiene una relación directa con la fidelidad hacia la aerolínea*”, aunque Shen y Yahya (2021) destacan que “*los resultados presentan variaciones según el contexto geográfico y cultural*”. Este fenómeno enfatiza la necesidad de contextualizar los hallazgos para garantizar conclusiones válidas en diferentes regiones.

## Relación entre Satisfacción del Pasajero y Lealtad a la Aerolínea

La relación entre la satisfacción del pasajero y la lealtad hacia la aerolínea se analiza ampliamente en la literatura académica, dado que representa un factor clave en la gestión estratégica de las compañías aéreas. Como afirman Tam, Barroso y Cruz-Jesus (2022), “*la intención de continuar utilizando servicios de aerolíneas de bajo coste se ve fuertemente influenciada por la percepción de valor y la satisfacción previa*”.

Los estudios indican que la lealtad del pasajero no siempre refleja directamente la satisfacción experimentada. Como señalan Shen y Yahya (2021), “*investigaciones realizadas en el contexto de aerolíneas low-cost en el Sudeste Asiático identifican que la percepción de calidad del servicio y el precio ejercen un impacto significativo en la lealtad*” (p.XX).

Sin embargo, otros estudios realizados en mercados como Arabia Saudí revelan correlaciones más bajas entre satisfacción y fidelidad, lo que indica que factores contextuales pueden incidir en esta relación (Hassan & Salem, 2022).

Asimismo, estudios centrados en mercados europeos, como el de Akamavi et al. (2015), señalan que la lealtad hacia aerolíneas de bajo coste puede estar mediada por expectativas específicas relacionadas con el servicio básico y la eficiencia operativa. Esta variabilidad en los resultados pone de manifiesto la importancia de considerar factores culturales y económicos al examinar la relación entre satisfacción y lealtad.

La pandemia de COVID-19 introduce nuevas dinámicas en esta relación. Según Rita, Moro y Cavalcanti (2022), el contexto de crisis sanitaria modifica las expectativas de los pasajeros, incrementando el valor percibido de medidas de seguridad e higiene, lo que a su vez incide en la percepción de calidad y, por ende, en la lealtad del pasajero.

## Evolución del Concepto de Satisfacción: Factores Psicológicos y Contextuales

La satisfacción del pasajero se desarrolla desde una perspectiva puramente cognitiva hacia una comprensión más integral que incorpora componentes afectivos y conativos. En un primer momento, se entiende la satisfacción como el resultado de una comparación entre el servicio recibido y las expectativas previas (Singh, 2021). No obstante, estudios recientes incluyen elementos emocionales y motivacionales, reconociendo que la satisfacción implica tanto respuestas racionales como emocionales (Tahanisaz, 2020).

Las teorías que sustentan esta evolución se pueden agrupar en dos perspectivas filosóficas. Por un lado, el constructivismo establece que la satisfacción se construye socialmente y surge de la interacción entre el pasajero y su entorno (Ahmadpour et al., 2014). Por otro lado, el empirismo lógico sostiene que la satisfacción debe medirse a partir de fenómenos observables y cuantificables (Tadajewski, 2010).

El uso de modelos formativos y reflexivos en la investigación representa esta dualidad teórica. Los modelos formativos consideran que el constructo es el resultado de la combinación de múltiples indicadores, mientras que los modelos reflexivos suponen que los indicadores reflejan una variable subyacente. Ambos enfoques se aceptan científicamente y se integran en meta-análisis para evaluar la satisfacción del pasajero (Afshardoost et al., 2021).

## Métodos de Medición de la Satisfacción del Pasajero

La medición de la satisfacción del pasajero en el ámbito de la aviación presenta múltiples desafíos debido a su naturaleza subjetiva y multidimensional. A lo largo del tiempo, la literatura desarrolla diversos enfoques metodológicos que posibilitan una evaluación más precisa y representativa de esta variable.

* **Métodos Cuantitativos** Los métodos cuantitativos prevalecen en la evaluación de la satisfacción del pasajero, destacándose especialmente el uso de encuestas estructuradas y cuestionarios. Según Rita, Moro y Cavalcanti (2022), los estudios que implementan encuestas tienden a centrarse en variables clave como la percepción del servicio, el confort a bordo y la relación calidad-precio. Estas encuestas se aplican inmediatamente después del vuelo para capturar percepciones frescas y específicas (Ekinci et al., 2008).  
   Entre las técnicas más utilizadas se incluyen el Análisis de Componentes Principales (PCA) y los Modelos de Ecuaciones Estructurales (SEM), los cuales permiten identificar relaciones entre variables latentes como calidad del servicio y lealtad. Además, el empleo de técnicas estadísticas avanzadas, como el análisis de correlación y regresión, contribuye a evaluar el impacto de cada factor en la satisfacción global (Tam et al., 2022).
* **Métodos Cualitativos** Los métodos cualitativos también resultan fundamentales para comprender la experiencia del pasajero. Las entrevistas semiestructuradas y los grupos focales permiten explorar percepciones subjetivas y emociones asociadas al servicio recibido. Aunque estos métodos proporcionan una visión más detallada y contextualizada, su aplicabilidad a gran escala se encuentra limitada (Kondo, 2001).
* **Análisis de Datos en Redes Sociales** El análisis de datos provenientes de redes sociales adquiere relevancia en la evaluación de la satisfacción del pasajero. El procesamiento de opiniones en plataformas como TripAdvisor y Google Reviews permite analizar la satisfacción de manera continua y en tiempo real (Rita et al., 2022). Los métodos de análisis de sentimientos aplicados a comentarios de usuarios ofrecen una perspectiva actualizada sobre la percepción del servicio, resaltando elementos emocionales que pueden no ser captados por los métodos tradicionales.
* **Consideraciones Críticas** A pesar de la diversidad de técnicas disponibles, no se establece un consenso sobre el método óptimo para medir la satisfacción del pasajero. Mientras que los enfoques cuantitativos proporcionan rigor estadístico, los cualitativos permiten captar matices emocionales y contextuales. Además, la integración de datos no estructurados provenientes de redes sociales agrega una capa adicional de complejidad, ya que se requiere filtrar información irrelevante y gestionar el sesgo del usuario (Rita et al., 2022).

En definitiva, la combinación de métodos cuantitativos, cualitativos y análisis de *big data* se posiciona como el enfoque más robusto para abordar la complejidad inherente a la satisfacción del pasajero en el contexto actual.

## Factores de Contexto: Impacto de la Pandemia en la Satisfacción del Pasajero

La pandemia de COVID-19 impacta de manera significativa en la industria de la aviación, transformando tanto las operaciones de las aerolíneas como las percepciones de los pasajeros respecto a la calidad del servicio. Las medidas sanitarias implementadas durante la pandemia, como el uso de mascarillas, el distanciamiento social y la desinfección frecuente, influyen de manera positiva en la percepción de calidad en ciertos contextos, mientras que en otros generan incomodidad (Rita et al., 2022; Shen & Yahya, 2021).

El impacto en la lealtad del pasajero también resulta considerable, especialmente en el contexto de aerolíneas *low-cost*, donde la limitada capacidad de adaptación a las nuevas normativas afecta la satisfacción del pasajero (Shen & Yahya, 2021). Por otro lado, las aerolíneas tradicionales que implementan medidas robustas logran mantener niveles de lealtad relativamente estables (Akamavi et al., 2015).

## Modelos y Técnicas existentes

El análisis de la satisfacción del pasajero en el ámbito de la aviación emplea una amplia variedad de modelos predictivos y técnicas analíticas para comprender los factores que influyen en la percepción de calidad y lealtad hacia las aerolíneas.

* **Modelos Predictivos más Comunes:** Los modelos de regresión, tanto lineal como logística, se utilizan ampliamente para predecir la satisfacción del pasajero a partir de variables como la percepción de calidad del servicio y el valor percibido. Estos modelos posibilitan la identificación de correlaciones significativas y el establecimiento de relaciones causales en contextos donde las variables independientes ejercen un impacto directo en la variable dependiente (Tam et al., 2022).  
   Otro enfoque común es el uso de árboles de decisión y algoritmos de clasificación como *Random Forest*, que permiten segmentar a los pasajeros según su nivel de satisfacción. Estas técnicas resultan especialmente útiles cuando se manejan múltiples variables categóricas o continuas (Shen & Yahya, 2021).
* **Análisis de Sentimientos y Procesamiento del Lenguaje Natural (PLN):** El análisis de sentimientos adquiere relevancia en el estudio de la satisfacción del pasajero, especialmente mediante comentarios en redes sociales y plataformas de reseñas. Técnicas como el análisis de sentimientos basado en lexicones o el empleo de algoritmos de aprendizaje profundo (como *LSTM* y *BERT*) demuestran ser efectivos para capturar emociones positivas y negativas en los comentarios de los usuarios (Rita et al., 2022).
* **Modelos de Ecuaciones Estructurales (SEM):** Los Modelos de Ecuaciones Estructurales (*SEM*) se posicionan como una herramienta clave en la investigación sobre satisfacción del pasajero debido a su capacidad para manejar múltiples variables exógenas y endógenas, así como variables latentes no observables directamente (Jenatabadi & Ismail, 2014). Estos modelos posibilitan una estimación robusta de relaciones complejas, lo que resulta especialmente útil en el contexto de la aviación, donde factores como la percepción de calidad, el valor percibido y la lealtad pueden encontrarse interrelacionados.  
   Según Jenatabadi e Ismail (2014), el *SEM* se aplica para analizar el rendimiento financiero y no financiero de las aerolíneas, permitiendo la integración de variables económicas y operativas en un modelo unificado. Esto evidencia su versatilidad para evaluar tanto la percepción de los pasajeros como el desempeño empresarial en condiciones económicas variables.
* **Técnicas de Clustering:** El empleo de algoritmos de agrupamiento como *K-means* permite la identificación de grupos homogéneos de pasajeros en función de su nivel de satisfacción. Esta técnica resulta particularmente útil para la segmentación de pasajeros en estudios de mercado, lo que posibilita el diseño de estrategias de fidelización específicas por parte de las aerolíneas (Akamavi et al., 2015).
* **Redes Neuronales y Modelos de Aprendizaje Profundo:** La utilización de redes neuronales para predecir la lealtad del pasajero se incrementa en estudios recientes, especialmente debido a su capacidad para modelar relaciones no lineales y patrones complejos. Singh (2021) establece que los modelos basados en redes neuronales pueden capturar dinámicas complejas entre variables de calidad percibida y satisfacción, lo que permite generar predicciones más precisas en contextos de alta variabilidad.

# Materiales y Métodos

## Estructura de los datos

El conjunto de datos utilizado se compone de 129,880 observaciones, cada una definida por 25 variables que caracterizan tanto aspectos personales de los pasajeros como dimensiones operativas del vuelo y valoraciones subjetivas sobre los distintos servicios ofrecidos por la aerolínea. Esta diversidad posibilita el abordaje del problema desde una perspectiva multidimensional, integrando factores demográficos, hábitos de viaje, rendimiento del servicio y nivel de satisfacción percibido.

Las variables incluyen datos numéricos y categóricos. Entre las variables numéricas se destacan la edad, la distancia del vuelo, los minutos de retraso y diversas puntuaciones de satisfacción expresadas en escalas del 1 al 5. Por otro lado, las variables categóricas agrupan información relacionada con el tipo de pasajero, el motivo del viaje, la clase del billete y la evaluación general de satisfacción.

Este tipo de estructura resulta especialmente útil para la implementación de modelos predictivos, ya que posibilita el análisis de las relaciones entre múltiples factores y la construcción de algoritmos capaces de anticipar la experiencia del pasajero en función de su perfil y las condiciones del vuelo

## Descripción de las variables obtenidas

A continuación, se muestra una tabla que sintetiza las variables incluidas en el análisis, acompañadas de su descripción y tipo de dato.

| **Variable** | **Descripción** | **Tipo de dato** |
| --- | --- | --- |
| Gender | Gender of the passengers (Female, Male) | object |
| Customer Type | The customer type (Loyal customer, disloyal customer) | object |
| Age | The actual age of the passengers | int64 |
| Type of Travel | Purpose of the flight of the passengers (Personal Travel, Business Travel) | object |
| Class | Travel class in the plane of the passengers (Business, Eco, Eco Plus) | object |
| Flights per year | Number of flights per year | float64 |
| Flight Distance | The flight distance of this journey | int64 |
| Inflight wifi service | Satisfaction level of the inflight wifi service (0: Not Applicable; 1-5) | float64 |
| Departure/Arrival time convenient | Satisfaction level of Departure/Arrival time convenience | int64 |
| Ease of Online booking | Satisfaction level of online booking | int64 |
| Gate location | Satisfaction level of Gate location | int64 |
| Food and drink | Satisfaction level of Food and drink | int64 |
| Online boarding | Satisfaction level of online boarding | int64 |
| Seat comfort | Satisfaction level of Seat comfort | int64 |
| Inflight entertainment | Satisfaction level of inflight entertainment | int64 |
| On-board service | Satisfaction level of On-board service | int64 |
| Leg room service | Satisfaction level of Leg room service | int64 |
| Baggage handling | Satisfaction level of baggage handling | int64 |
| Check-in service | Satisfaction level of Check-in service | int64 |
| Inflight service | Satisfaction level of inflight service | int64 |
| Cleanliness | Satisfaction level of Cleanliness | int64 |
| Departure Delay in Minutes | Minutes delayed when departure | int64 |
| Arrival delay in minutes | Minutes delayed when arrival | float64 |
| Satisfaction | Airline satisfaction level (Satisfaction, neutral, or dissatisfaction) | object |

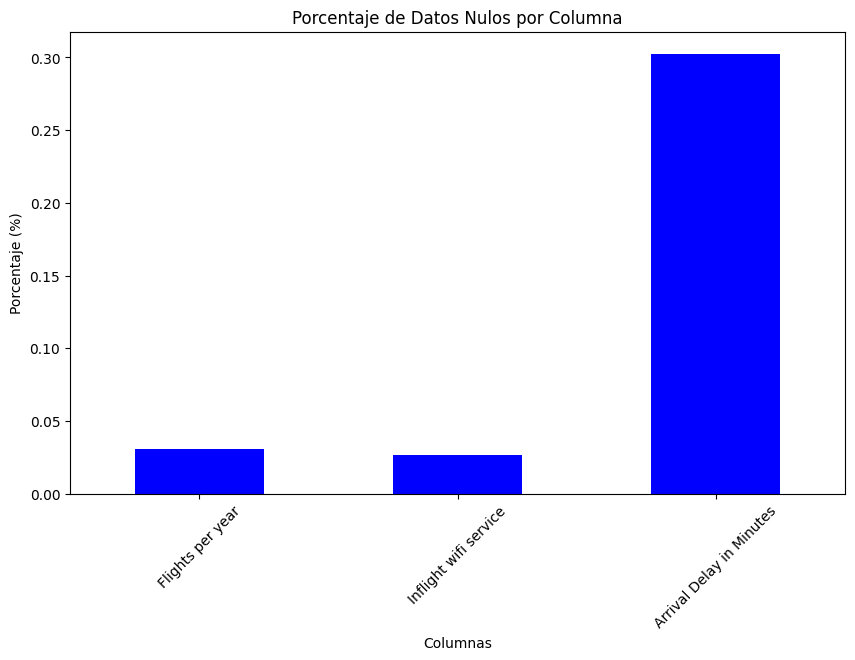
###### Tabla 1: Resumen de variables del dataset con tipo y descripción

## Análisis Exploratorio de Variables

Durante la lectura inicial del conjunto de datos se identifica una columna denominada “Unnamed: 0”, generada por la exportación del índice original. Al no aportar información relevante, se procede a eliminarla mediante el comando: df.drop(['Unnamed: 0'], axis=1, inplace=True).

Se verifica la existencia de valores nulos y registros duplicados usando los métodos df.isna().sum() y df.duplicated().sum() con el objetivo de identificar posibles problemas de calidad en los datos. A continuación, se representa gráficamente el porcentaje de valores nulos por columna, lo que posibilita visualizar de forma clara tres variables afectadas:

* Flights per year
* Inflight wifi service
* Arrival Delay in Minutes



###### Figura 1: Porcentaje de valores nulos por variable

Las dos primeras muestran un porcentaje muy bajo de valores faltantes (alrededor del 0,03%), mientras que la tercera presenta un valor ligeramente superior (algo más del 0,3%). Aunque la proporción es reducida, se decide imputar estos valores para evitar errores en las etapas posteriores del análisis.

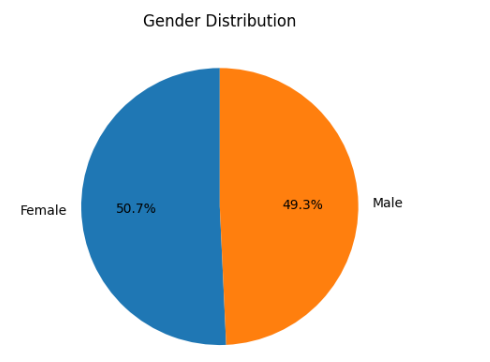
Se utiliza la media para imputar las variables Flights per year e Inflight wifi service, dado que ambas presentan una distribución aproximadamente normal y no muestran valores atípicos significativos. En cambio, Arrival Delay in Minutes es imputada utilizando la mediana, dado que presentan valores extremos que podrían distorsionar la media. Esta estrategia garantiza mayor robustez en la representación central de los datos.

Tras completar las imputaciones, se aplica .astype(int) para convertir estas columnas a tipo entero. Esta conversión se realiza tanto para asegurar la consistencia de los datos discretos como para evitar posibles conflictos con modelos que diferencian entre tipos de datos. Con el conjunto ya limpio, se da comienzo al análisis exploratorio de las principales variables. El objetivo es entender mejor la composición de la muestra y anticipar posibles sesgos o desequilibrios que puedan afectar el modelo.

Con el fin de entender la composición del conjunto de datos y detectar posibles desbalances que puedan afectar al modelo, se analizan las siguientes variables:

**Gender**

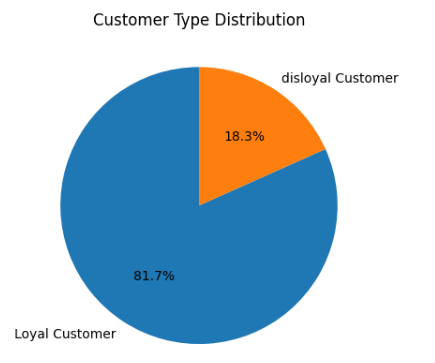
Se observa una distribución equilibrada entre géneros: un 50,7% de mujeres y un 49,3% de hombres. Esta paridad es positiva, ya que minimiza el riesgo de sesgo en los modelos predictivos. La variable es representada gráficamente mediante un diagrama de sectores para facilitar su interpretación.



###### Figura 2: Distribución de género de los pasajeros

**Customer Type**

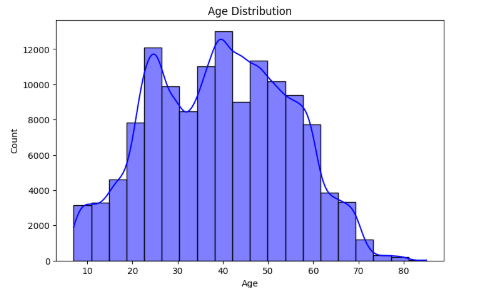
Esta variable distingue entre pasajeros leales y no leales. El análisis revela un claro predominio de pasajeros leales (81,7%), frente a un 18,3% de pasajeros no leales. Este desequilibrio puede influir en la forma en que se percibe el servicio, ya que los pasajeros frecuentes podrían tener expectativas distintas o ser más exigentes.



###### Figura 3: Distribución de tipo de pasajero (leal vs no leal)

**Age**

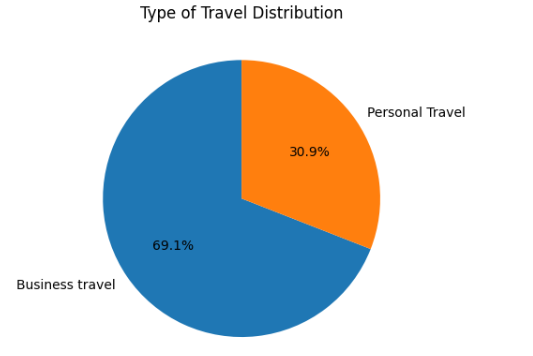
La variable edad se representa mediante un histograma con línea de densidad (KDE), donde se observa una distribución cercana a la normal. La mayor concentración de pasajeros se sitúa entre los 30 y 50 años, aunque hay también un número significativo de usuarios mayores de 60. Esta información resulta útil para adaptar servicios a diferentes perfiles etarios.



###### Figura 4: Distribución de edad de los pasajeros

**Type of Travel**

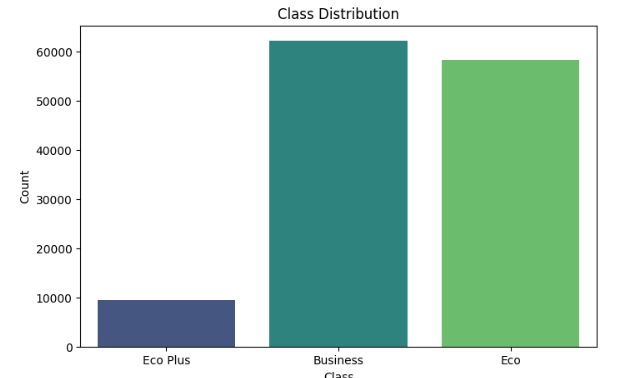
Se identifican dos tipos de viajes: por motivos personales y por trabajo. El 69,1% de los pasajeros viaja por negocios, lo que sugiere un perfil predominantemente corporativo. Esta variable se representa con un gráfico de tarta permitiendo la visualización del desequilibrio. Entender esta diferencia es clave para interpretar cómo valoran los distintos servicios.

****

###### Figura 5: Distribución del tipo de viaje (personal vs negocios)

**Class**

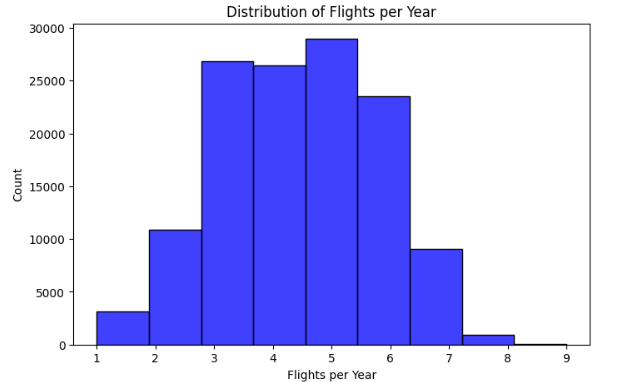
Esta variable muestra en qué clase viajaron los pasajeros. Business es la categoría más frecuente (62.160 registros), seguida por Eco (58.309) y Eco Plus (9.411). Se visualiza con un gráfico de barras que muestra claramente la distribución desigual. Este dato puede condicionar la satisfacción, ya que cada clase conlleva distintos niveles de servicio.



###### Figura 6: Distribución de clase del billete

**Flights per year**

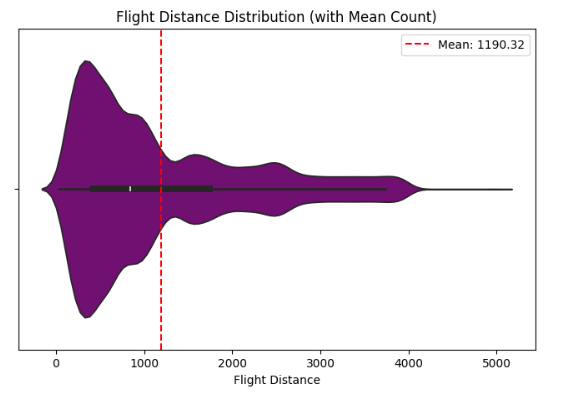
Se examina la frecuencia anual de vuelos. La mayoría de los pasajeros realizan entre 3 y 6 vuelos al año, con un pico en 5. A partir del séptimo vuelo, la frecuencia disminuye de forma pronunciada. Se utiliza un histograma para su representación mostrando la relación entre la fidelidad y el grado de exigencia del pasajero.



###### Figura 7: Distribución de vuelos por año

**Flight Distance**

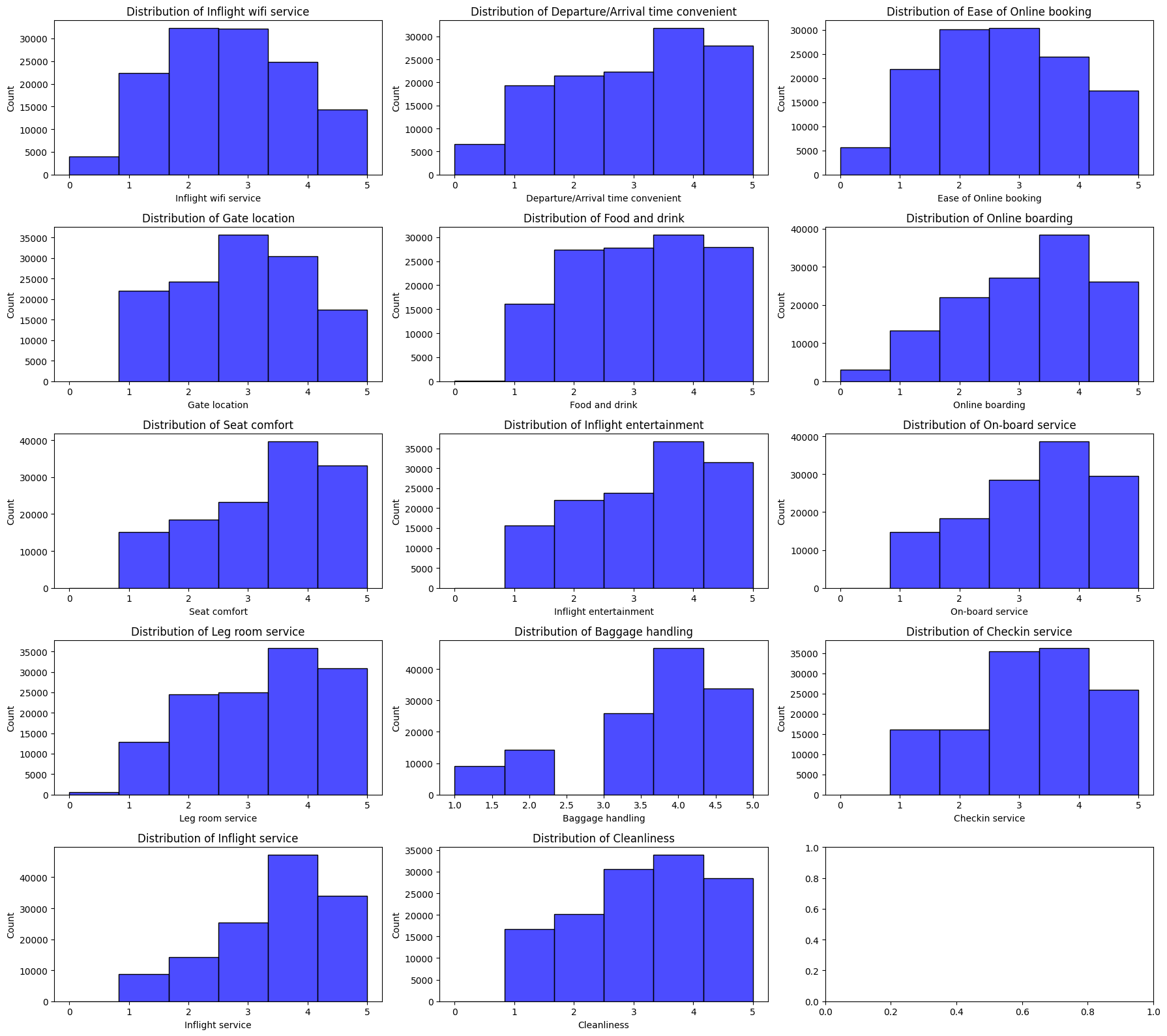
Esta variable presenta una gran dispersión. Se utilizan gráficos de violín para representar su distribución, mostrando una concentración de vuelos por debajo de los 1.000 km, aunque también se detectan trayectos superiores a 4.000 km. La media se sitúa en torno a los 1.190 km. Esta variabilidad permite distinguir entre vuelos de corto y largo recorrido, con posibles implicaciones en la experiencia del usuario.



###### Figura 8: Distribución de distancia de vuelo (gráfico de violín)

### **Análisis de variables relacionadas con la satisfacción**

Dado que el objetivo principal del proyecto es construir un modelo capaz de predecir la satisfacción del pasajero, se dedica una parte específica del análisis exploratorio a examinar las variables que recogen valoraciones directas sobre distintos aspectos del servicio prestado por la aerolínea. Estas variables reflejan de forma cuantitativa la percepción subjetiva del pasajero en distintas etapas del proceso de viaje, desde la reserva hasta el desembarque.

En concreto, se trabaja con un conjunto de 14 variables ordinales que representan distintos componentes de la experiencia del pasajero, tales como la comodidad del asiento, el entretenimiento a bordo, la calidad del servicio durante el vuelo, la facilidad del embarque, la limpieza, el manejo del equipaje o la puntuación del wifi. Todas estas valoraciones se registran en una escala de 0 a 5, siendo 5 el nivel más alto de satisfacción.

###### Figura 9: Distribución de valoraciones de servicios (histogramas)

Para cada una de estas variables se generan histogramas individuales con el fin de visualizar su distribución y detectar patrones, sesgos o irregularidades. Este análisis visual posibilita la identificación de tendencias comunes y diferencias significativas entre ellas.

En general, la mayoría de las variables presentan distribuciones sesgadas hacia la derecha, lo que indica que una parte importante de los pasajeros tiende a valorar positivamente su experiencia. Esto sugiere una percepción global relativamente favorable del servicio ofrecido por la aerolínea, aunque no exenta de puntos débiles.

Entre las variables con mejores valoraciones destacan on-board service, inflight service y cleanliness, que concentran una gran cantidad de puntuaciones de 4 y 5. Esto podría reflejar un esfuerzo consciente por parte de la compañía en garantizar una buena atención a bordo y un entorno cuidado durante el viaje.

Por el contrario, variables como inflight wifi service o gate location muestran distribuciones mucho más dispersas, con presencia de valoraciones bajas (1 y 2) en proporciones relevantes. Esto puede deberse a una experiencia de calidad menos homogénea en estos servicios o a expectativas más exigentes por parte de los usuarios. La variabilidad detectada en estas dimensiones sugiere que hay margen de mejora, especialmente si se quiere reforzar la satisfacción general.

Este análisis es fundamental para orientar la construcción del modelo predictivo. Si bien todas las variables aportan información, no todas tienen el mismo peso en la percepción de calidad. Algunas resultan ser factores clave, mientras que otras podrían influir de forma marginal o incluso estar condicionadas por elementos externos.

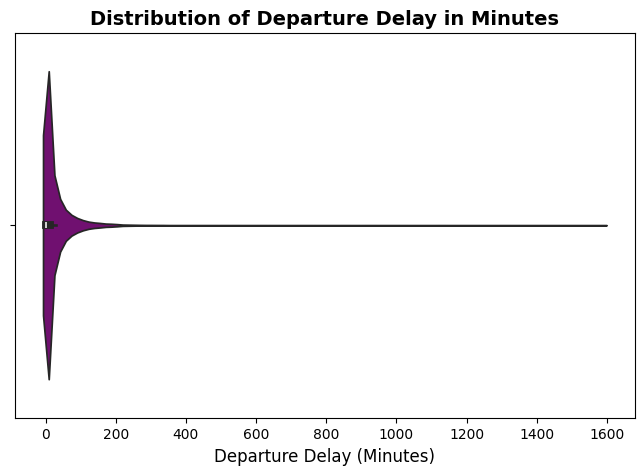
Desde una perspectiva de negocio, conocer el impacto relativo de estas variables sobre la satisfacción permite a la aerolínea tomar decisiones más informadas. Por ejemplo:

* Detectar pasajeros en riesgo de insatisfacción a partir de sus valoraciones en aspectos críticos como el confort del asiento o el entretenimiento.
* Ofrecer servicios personalizados o medidas compensatorias (como mejoras de asiento o embarque preferente) a pasajeros con valoraciones bajas en áreas sensibles.
* Optimizar la inversión en mejoras operativas, priorizando aquellas dimensiones que se relacionan directamente con la experiencia global del pasajero.

En resumen, el análisis detallado de estas variables no solo facilita la preparación de los datos para el modelado, sino que refuerza la lógica empresarial del proyecto, al evidenciar de forma cuantitativa qué aspectos del servicio deben ser objeto de especial atención si se desea aumentar la satisfacción y la fidelización del pasajero.

**Departure Delay in Minutes**

Más de 73.000 vuelos no presentan retraso en el despegue. No obstante, hay registros que alcanzan demoras superiores a 12 horas. Esta variable se representa con un gráfico de violín, que evidencia la concentración en torno a cero y la presencia de outliers. Aunque son casos puntuales, pueden tener un gran impacto en la percepción del servicio.

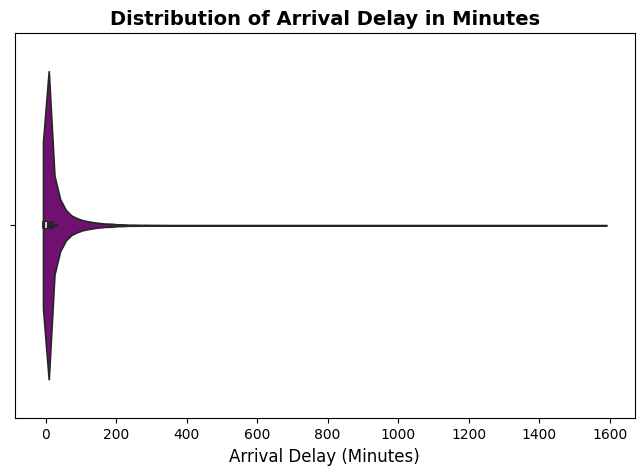
.

###### Figura 10: Distribución del retraso en despegue (gráfico de violín)

**Arrival Delay in Minutes**

Se analiza de forma paralela el retraso en la llegada. Los patrones observados fueron prácticamente idénticos: la mayoría de valores en 0 y una cola larga de valores extremos que alcanza los 1.584 minutos (casi 26 horas). Este tipo de datos requiere atención especial, ya que incluso una pequeña proporción de vuelos con retrasos graves puede generar un fuerte impacto negativo en la satisfacción general y la imagen de la aerolínea.

Los gráficos de violín ayudan a detectar visualmente estos valores extremos, que podrían necesitar ser tratados con técnicas robustas (como truncado o winsorización) antes de ser introducidos en el modelo de machine learning.

****

###### Figura 11: Distribución del retraso en llegada (gráfico de violín)

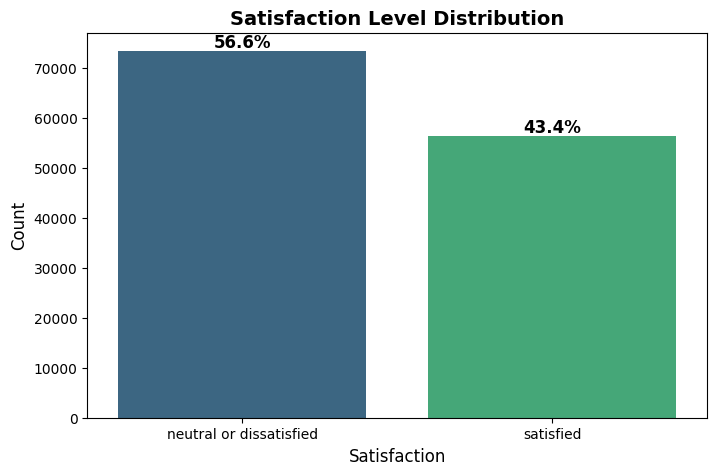
### **Distribución de la variable objetivo: satisfacción del pasajero**

La variable satisfaction fue recodificada en dos clases:

* 0: neutral or dissatisfied
* 1: satisfied

Al contar los valores, se identifica que el 56% de los pasajeros están satisfechos, mientras que el 44% no lo están o son neutrales. Esta distribución relativamente equilibrada justifica el enfoque de clasificación binaria en el modelo predictivo, y sugiere que hay margen de mejora en la experiencia de vuelo para un porcentaje importante de pasajeros.

Esta observación es especialmente relevante porque refuerza la motivación del proyecto: si el modelo puede identificar qué combinación de factores reduce la satisfacción, la aerolínea podrá intervenir de forma preventiva y segmentada.



###### Figura 12: Distribución binaria de satisfacción del pasajero

**Codificación de variables categóricas**

Antes de entrenar el modelo predictivo, es necesario transformar las variables categóricas a formato numérico. Esto se debe a que los algoritmos de machine learning no pueden trabajar directamente con valores tipo texto. A continuación se detallan las transformaciones aplicadas:

* Gender. La variable Gender es codificada como binaria, asignando el valor 1 a los pasajeros de género masculino y 0 a los de género femenino. Para facilitar su interpretación, la variable es renombrada como Gender\_male.
* Customer Type. Se asigna el valor 1 a los pasajeros leales y 0 a los no leales. Esta variable también es renombrada como is\_loyal, lo que facilita su lectura y comprensión dentro del modelo.
* Type of Travel. Esta variable distingue entre vuelos de negocios y personales. Se asigna el valor 1 a los vuelos por motivos de trabajo y 0 a los personales, renombrando la variable a Business Travel.
* Class. Como la clase del billete tiene un orden lógico (Eco, Eco Plus y Business Class), se aplica label encoding para representarla como una variable ordinal, asignando 0 a clase económica, 1 a Eco Plus y 2 a Business Class. Esta codificación permite al modelo aprovechar la jerarquía implícita de la variable.
* Satisfaction: Esta variable es convertida valor binario, donde 1 representa que el pasajero está satisfecho y 0 que está insatisfecho o neutral, lo que facilita tanto el entrenamiento como la evaluación del modelo.

**Matriz correlación**  
Una vez aplicadas todas las transformaciones, se verifica que todas las variables estuvieran correctamente tipadas como numéricas y sin valores nulos. El conjunto de datos resultante incluye 25 columnas preparadas para ser utilizadas en el modelo de predicción. El uso de nombres claros como is\_loyal, Gender\_male o Business Travel facilita la posterior interpretación de los resultados y su aplicación práctica en la toma de decisiones de negocio.

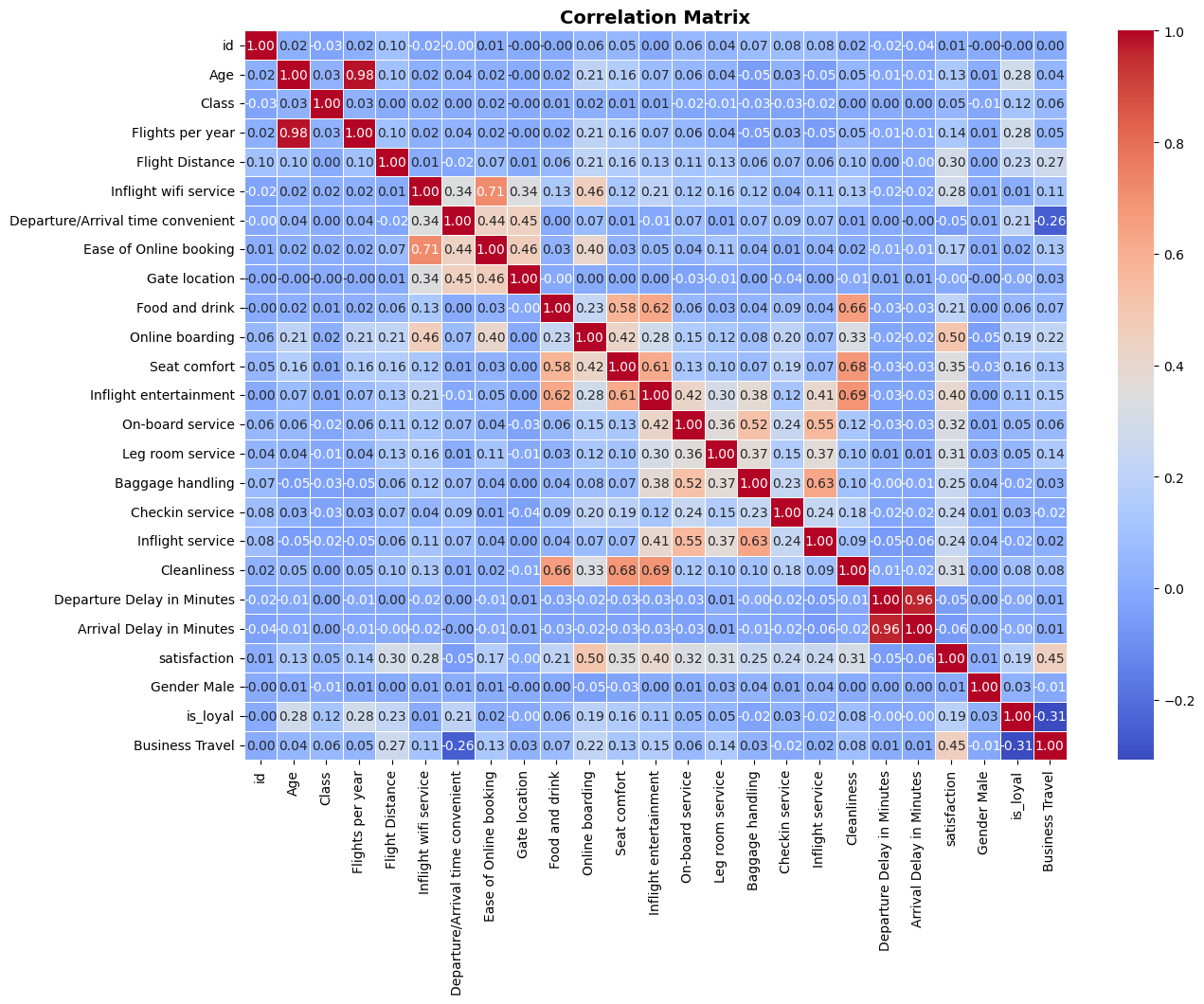
Como paso final del análisis exploratorio, se construye una matriz de correlación entre las variables numéricas del conjunto de datos. El objetivo es identificar relaciones lineales significativas entre las distintas variables, especialmente aquellas que pueden influir directamente en la variable objetivo: la satisfacción del pasajero.

En esta matriz se utilizan colores para visualizar la intensidad y el signo de las correlaciones: los tonos rojos indican correlaciones positivas, mientras que los azules representan relaciones negativas. Las correlaciones cercanas a cero sugieren ausencia de relación lineal.

Entre los hallazgos más relevantes destacan los siguientes:

* La variable satisfaction muestra correlaciones positivas moderadas con variables como on-board service (0.42), inflight service (0.41), seat comfort (0.31) y cleanliness (0.31). Esto sugiere que estos factores influyen de forma significativa en la percepción de satisfacción general del pasajero.
* También se observa correlaciones negativas entre satisfaction y los retrasos, especialmente arrival delay in minutes (-0.20), lo que indica que los retrasos prolongados afectan negativamente la experiencia del usuario.
* Algunas variables muestran una alta relación entre sí, como inflight service y on-board service (0.69), o leg room service y seat comfort (0.62), lo que puede interpretarse como dimensiones del servicio que los pasajeros tienden a valorar conjuntamente.

Este análisis permite identificar las variables más relevantes para el modelo predictivo posterior y anticipar la posible multicolinealidad entre variables similares. Además, justifica de forma empírica que mejorar ciertos aspectos del servicio (como la atención a bordo o la comodidad del asiento) podría traducirse en un incremento de la satisfacción general del pasajero.

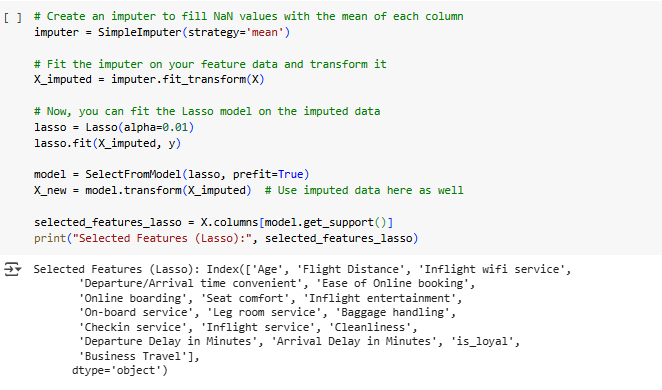
Se incluye a continuación la matriz de correlación correspondiente con valores anotados para su interpretación detallada.

###### Figura 13: Matriz de correlación entre variables

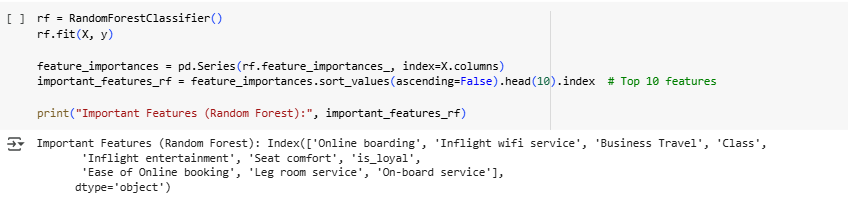
## Selección de variables representativas y reducción de dimensionalidad

Con el objetivo de construir un modelo más eficiente y centrado en las variables más influyentes, se aplican distintas técnicas de selección de características. Estas técnicas permiten reducir la complejidad del modelo y mejorar su rendimiento, al eliminar atributos que aportan poca o ninguna información predictiva.

* **Lasso:** Se aplica una regresión Lasso con regularización (α = 0.01) para identificar las variables con mayor impacto en la predicción de la satisfacción. Lasso tiende a reducir a cero los coeficientes de aquellas variables menos relevantes, facilitando la selección automática. Entre las variables seleccionadas se encuentran inflight service, seat comfort, online boarding, inflight entertainment, leg room service, departure delay y cleanliness, entre otras.

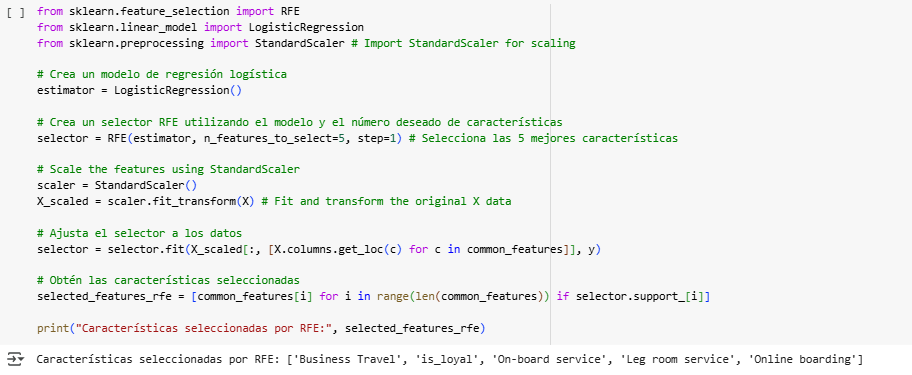


* **Random Forest:** Se emplea este modelo para estimar la importancia relativa de cada variable. Esta técnica es capaz de capturar relaciones no lineales entre las variables. Las más destacadas son online boarding, inflight wifi service, business travel, class, inflight entertainment e is\_loyal.

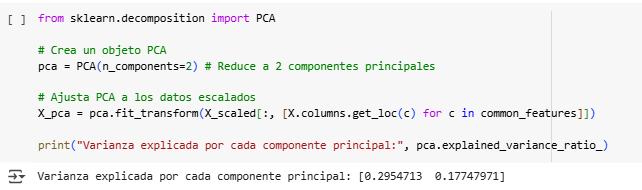
Tras comparar ambos métodos, se identifican variables comunes, como: inflight wifi service, business travel, inflight entertainment, is\_loyal, on-board service, leg room service, ease of online booking in online boarding. Estas características son seleccionadas como base para el modelo final, ya que su relevancia es confirmada por ambos enfoques.



* **RFE (Recursive Feature Elimination):** A partir del conjunto de variables comunes, se aplica el método de eliminación recursiva de características (RFE) utilizando un modelo de regresión logística. Este método selecciona las cinco variables más representativas del conjunto: business travel, is\_loyal, on-board service, leg room service y online boarding. Estas variables se utilizan como input principal para el modelo final de clasificación.



* **PCA (Análisis de Componentes Principales):** Se realiza un análisis de componentes principales (PCA) sobre el subconjunto reducido, con el fin de evaluar la proporción de varianza explicada. Los dos primeros componentes principales explican conjuntamente un 47% de la varianza, lo que respalda la efectividad de la reducción de dimensionalidad en la simplificación del modelo sin pérdida significativa de información.



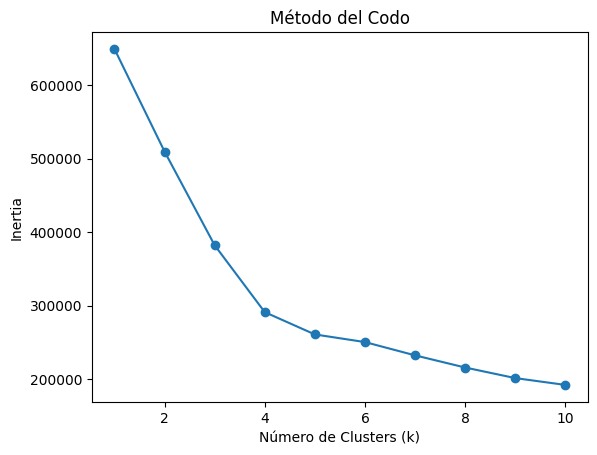
Estas técnicas permiten seleccionar un conjunto compacto y eficaz de variables predictoras, optimizando el rendimiento y la interpretabilidad del modelo.

## Clustering

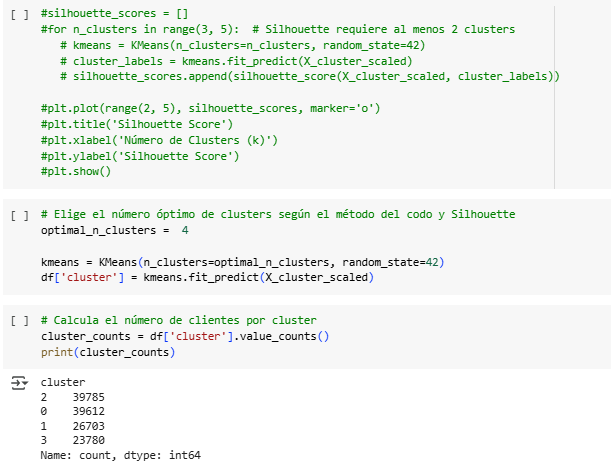
Con el fin de identificar perfiles de pasajeros que compartan características similares, se aplica un análisis de clustering utilizando las variables seleccionadas previamente mediante el método de eliminación recursiva de características (RFE). Esta técnica permite descubrir grupos naturales dentro de los datos que pueden ser utilizados para la personalización de servicios, mejoras operativas y toma de decisiones estratégicas por parte de la aerolínea.

Para preparar los datos, se estandarizan las variables con un escalado z-score, asegurando que todas tuvieran el mismo peso dentro del algoritmo. A continuación, se utiliza el algoritmo KMeans para agrupar a los pasajeros en distintos clusters.

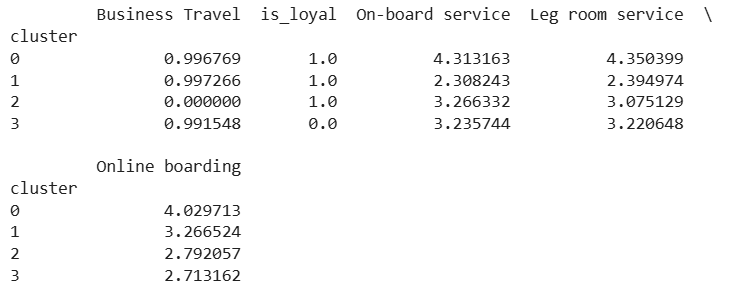
El número óptimo de grupos se determina combinando dos enfoques complementarios: el método del codo, que identifica el punto de inflexión en la curva de inercia, y el índice de Silhouette, que evalúa la calidad de los agrupamientos mediante la cohesión interna y separación externa de los clusters. Ambos métodos coincidieron en que el valor más adecuado era k = 4, lo cual garantiza una segmentación equilibrada y representativa de los distintos perfiles de usuarios.



###### Figura 14: Representación gráfica de los clusters tras aplicación de K-Means



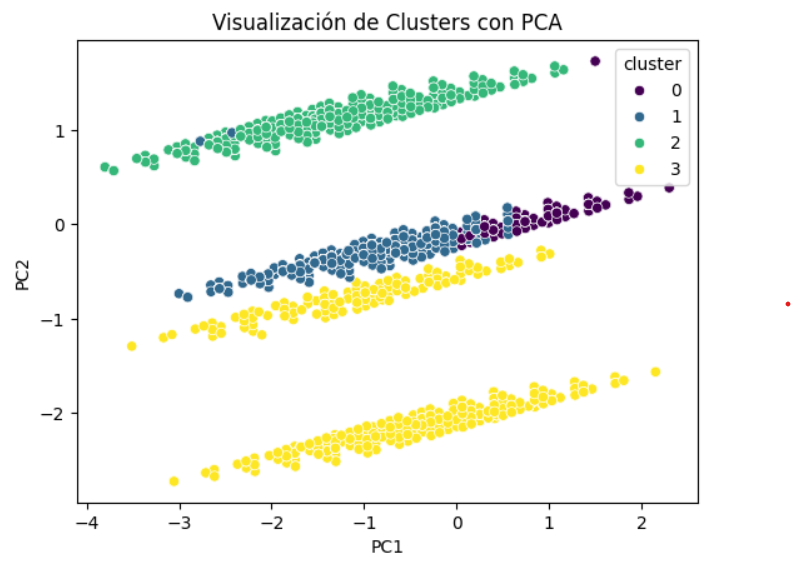
Tras asignar una etiqueta de cluster a cada pasajero, se calculan las medias de las variables incluidas en el análisis para cada grupo, permitiendo así una interpretación clara de las diferencias entre segmentos. A modo de resumen, se obtienen las siguientes conclusiones principales:



* Cluster 0: Agrupa principalmente a pasajeros de viajes de negocios, con una fidelidad elevada (is\_loyal) y puntuaciones altas en servicios como embarque online, atención a bordo y espacio para las piernas. Representa un perfil premium que valora la eficiencia y el confort.
* Cluster 1: También está compuesto por pasajeros leales, aunque con puntuaciones más bajas en variables como “on-board service” o “leg room service”. Puede interpretarse como un grupo con potencial de mejora mediante acciones de refuerzo del valor percibido.
* Cluster 2: Se distingue por estar compuesto exclusivamente por pasajeros que no viajan por trabajo. A pesar de su lealtad a la aerolínea, presentan puntuaciones más bajas en todos los aspectos del servicio. Este grupo puede representar un segmento con expectativas no satisfechas o necesidades específicas no cubiertas.
* Cluster 3: Tiene una baja proporción de pasajeros leales y puntuaciones medias en las variables consideradas. Puede considerarse un grupo más indiferente, con mayor riesgo de abandono si no se implementan acciones personalizadas.

Este análisis de clustering aporta una segmentación clara y operativa de los pasajeros. Desde el punto de vista del negocio, proporciona una base sólida para diseñar estrategias diferenciadas por perfil, como por ejemplo ofrecer mejoras específicas al cluster 2 para aumentar su satisfacción, o mantener programas exclusivos para fidelizar aún más a los pasajeros del cluster 0. Además, esta información puede alimentar futuros modelos predictivos o sistemas de recomendación en tiempo real.

El siguiente gráfico de dispersión muestra la visualización de la distribución y separación de los clusters en un espacio reducido de dos dimensiones, obtenido a través de Análisis de Componentes Principales (PCA) permitiendo interpretar patrones y diferencias entre los grupos de pasajeros de manera más clara. Los cluster 0 y 1 tienen solapamiento indicando que tienen características similares.

****

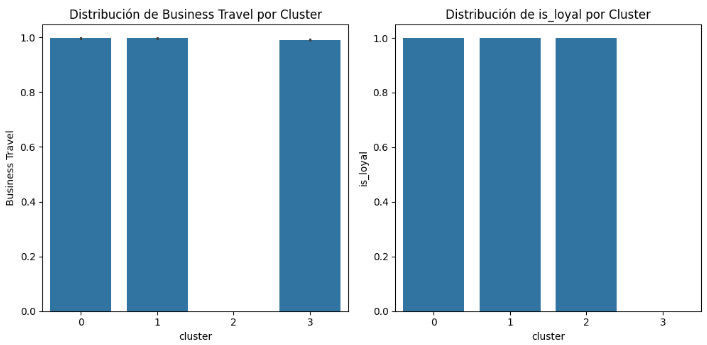
###### Figura 15: Comparativa de variables por cluster: Business Travel, is\_loyal, on-board service, legroom y online boarding

Los siguientes 5 gráficos de barras representan la distribución de cada variable significativa seleccionada a través de RFE.

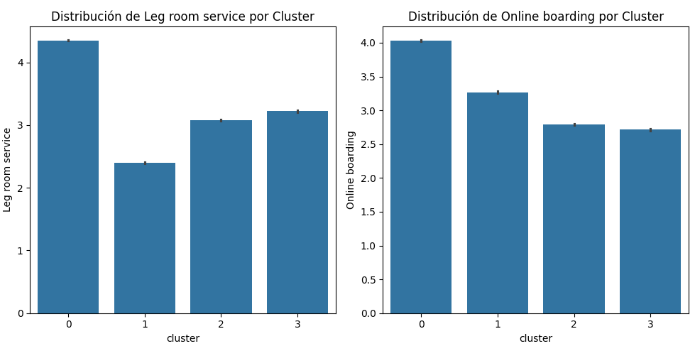
* Para Business Travel, los cluster 0, 1 y 3 agrupan principalmente pasajeros que viajan por motivos de trabajo, mientras que el cluster 2 representa viajeros por viajes personales.
* Para is\_loyal, el cluster 3 representa pasajeros no leales, mientras que los otros reúnen pasajeros frecuentes.
* Para on-board service, mientras el cluster 0 destaca por su alta satisfacción con el servicio a bordo, el cluster 1 agrupa pasajeros menos satisfechos.
* Para leg room service, el cluster 1 destaca por su insatisfacción con el espacio para las piernas, mientras que el cluster 0 tiene la mejor evaluación.
* Para Online boarding, el cluster 0 reúne pasajeros que valoran el proceso de embarque digital, mientras que los clusters 2 y 3 tienen una experiencia menos positiva.

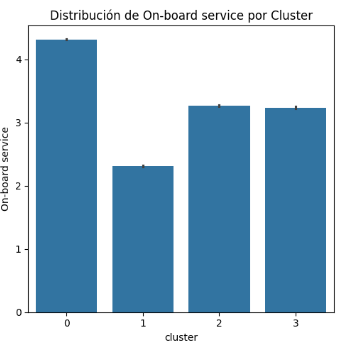
Por lo tanto, se puede afirmar que:

* Cluster 0: Clientes leales y satisfechos, especialmente en servicios a bordo, espacio y embarque en línea.
* Cluster 1: Clientes leales pero insatisfechos, con puntuaciones bajas en servicio y comodidad.
* Cluster 2: Clientes no comerciales (viajes personales), con satisfacción moderada en todos los aspectos.
* Cluster 3: Clientes no leales, con evaluaciones intermedias en todos los servicios.

****

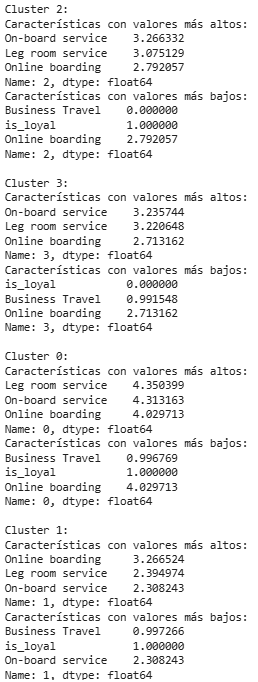
###### Figura 16: Distribución de satisfacción por cluster en variables clave

****

****

###### Figura 17: Visualización PCA de clusters generados por K-Means

Los siguientes resultados muestran como las características clave varían en cada cluster identificado en el análisis.

****

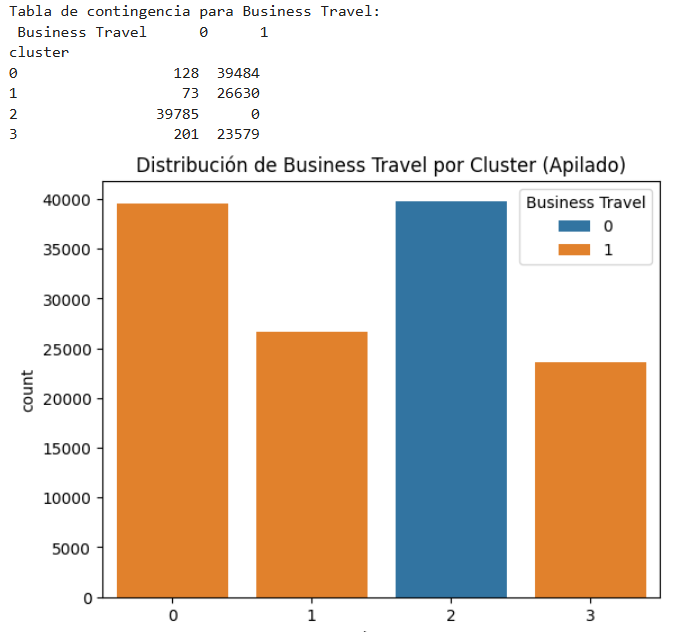
| Cluster | Perfil | Características altas | Características bajas |
| --- | --- | --- | --- |
| 0 | Viajeros de negocios leales y satisfechos | Leg room service = 4.35  On board service = 4.31  Online boarding = 4.03 | Business travel = 0.99  is\_loyal = 1.0 |
| 1 | Viajeros de negocios leales pero insatisfechos con comodidad y servicio | online boarding = 3.26 | On-board service = 2.30  Leg room service = 2.39  Business travel = 0.99  is\_loyal = 1 |
| 2 | Viajeros por motivos personales, pasajeros leales con satisfacción moderada | on-board service = 3.26  Leg room service = 3.07 | Business travel = 0  is\_loyal = 1.0  Online\_boarding = 2.79 |
| 3 | Pasajeros no leales con satisfacción intermedia en los servicios | On-board service = 3.23  Leg service = 3.22 | is\_loyal = 0  Business travel = 0.99  Online boarding = 2.71 |

###### Tabla 2: Características promedio por cluster (clustering interpretado)

Se analiza la distribución de las variables significativas en los diferentes clusters mediante una tabla de contingencia y un gráfico de barras

Business travel:

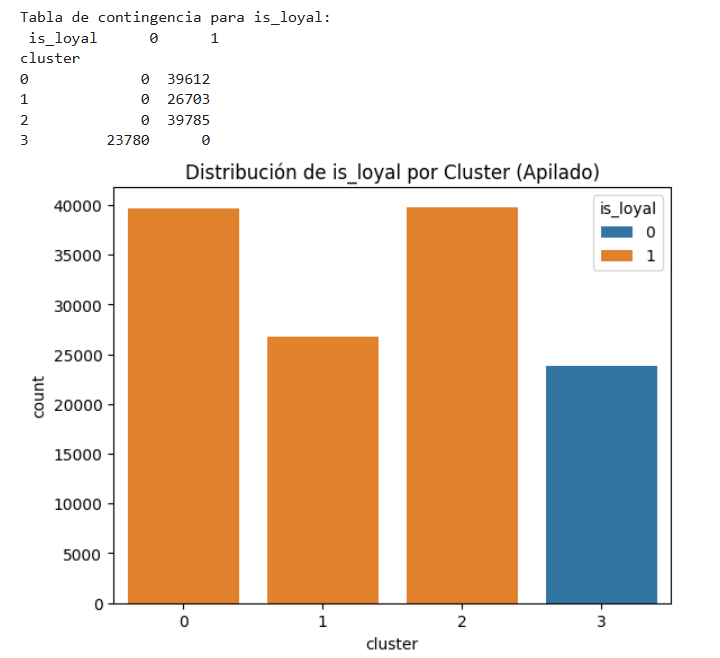
* Cluster 0, 1 y 3, la mayoría de los pasajeros en estos grupos viajan por negocios.
* Cluster 2, ninguno (valor = 0).



###### Figura 18: Distribución por tipo de viaje (Business Travel) según cluster

is\_loyal:

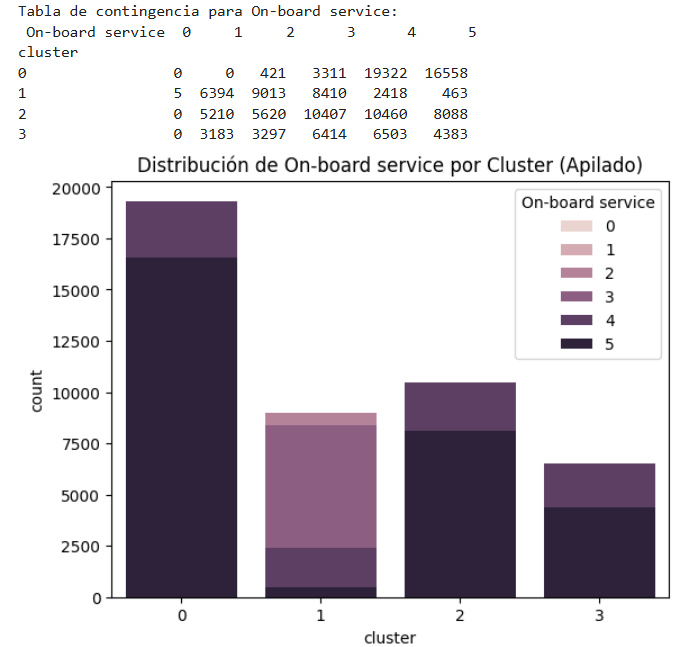
* Clusters 0, 1 y 2: Todos los pasajeros en estos grupos son leales
* Cluster 3: contiene únicamente pasajeros no leanes (valor = 0)



###### Figura 19: Distribución de lealtad (is\_loyal) según cluster

On-borad service:

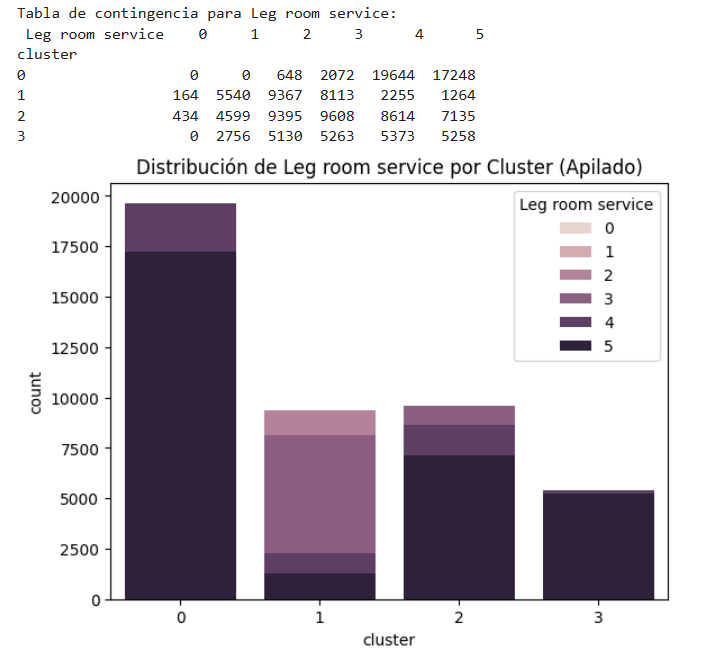
* Cluster 0 tiene una alta proporción de pasajeros con niveles 4 y 5, lo que indica alta satisfacción con el servicio a bordo.
* Cluster 1 tiene más pasajeros con niveles bajos 1 y 2, lo que sugiere insatisfacción con el servicio a bordo.
* Cluster 2 y 3 muestran una distribución más equilibrada, con una proporción notable de niveles 3 y 4.



###### Figura 20: Distribución de satisfacción con on-board service por cluster

Leg Room Service

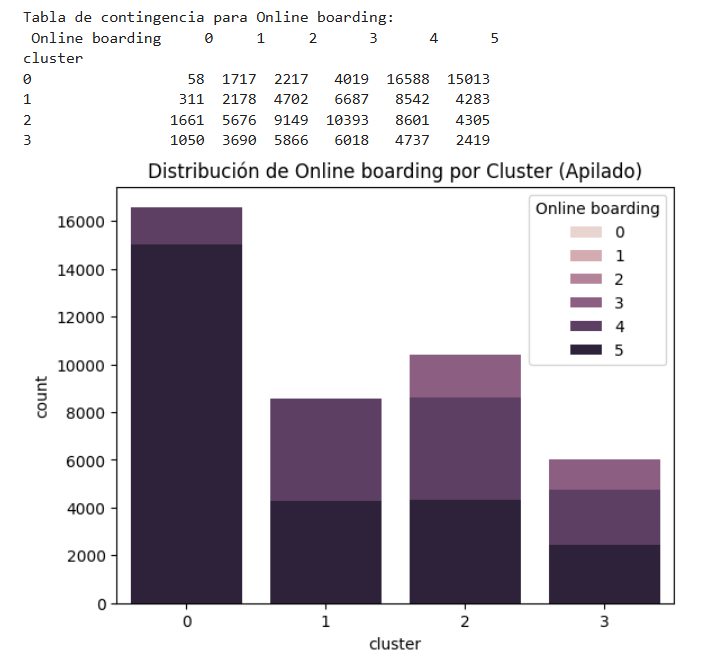
* Cluster 0 tiene una gran cantidad de pasajeros con niveles altos de satisfacción 4 y 5 , indicando que perciben mayor comodidad en el espacio para las piernas.
* Cluster 1 predomina la insatisfacción, con una alta cantidad de pasajeros en los niveles 1 y 2.
* Cluster 2 muestra una distribución más equilibrada, con muchos pasajeros en los niveles 3 y 4.
* Cluster 3 también tiene una distribución uniforme, con pasajeros en todos los niveles.



###### Figura 21: Distribución de satisfacción con leg room service por cluster

Online boarding

* Cluster 0 tiene valores altos de satisfacción, con muchos pasajeros en los niveles 4 y 5.
* Cluster 1 distribución equilibrada, aunque con una mayor presencia de pasajeros con niveles 2 y 3.
* Cluster 2 contiene la mayor cantidad de pasajeros con niveles 2 y 3
* Cluster 3 mayor concentración en niveles 3 y 4 con menos pasajeros en el nivel 5.



###### Figura 22: Distribución de satisfacción con online boarding por cluster

Conclusiones del Análisis de Clustering

**1. Número de Clusters y Tamaño**

El análisis de clustering realizado mediante el método de K-Means y el criterio del codo, determina que la estructura óptima para este conjunto de datos es de cuatro clusters. La distribución de pasajeros en cada grupo es la siguiente:

* **Cluster 0:** 39,612 pasajeros
* **Cluster 1:** 26,703 pasajeros
* **Cluster 2:** 39,785 pasajeros
* **Cluster 3:** 23,780 pasajeros

Estos grupos reflejan diferencias significativas en la percepción del servicio, hábitos de viaje y nivel de fidelización, proporcionando una base valiosa para la toma de decisiones estratégicas.

**2. Caracterización de los Clusters**

Cada cluster presenta perfiles distintos en función de los aspectos analizados:

Cluster 0: Viajeros frecuentes altamente satisfechos: Suelen viajar en Business o Eco Plus, tanto por negocios como por placer, y representan un segmento premium para la aerolínea. Se recomienda fortalecer programas de fidelización y ofrecer servicios exclusivos para mantener su satisfacción.

Este grupo reúne pasajeros leales que priorizan la comodidad y el entretenimiento en sus vuelos. Destacan por su alta valoración en:

* Comodidad del asiento
* Entretenimiento a bordo
* Servicio de wifi

Cluster 1: Pasajeros con menor satisfacción: Suelen viajar en clase económica y tienden a ser menos leales a la aerolínea. Para mejorar su experiencia, se pueden optimizar procesos, como el embarque y la oferta de entretenimiento, reduciendo puntos de fricción que afectan su satisfacción.

Este grupo muestra niveles de satisfacción más bajos, reflejando una experiencia menos positiva en:

* Comodidad del asiento
* Entretenimiento a bordo
* Frecuencia de retrasos

Cluster 2: Pasajeros orientados a la eficiencia: Incluye viajeros de negocios y de placer, con una fuerte tendencia a la lealtad. Se recomienda reforzar beneficios relacionados con la eficiencia, como prioridad en el embarque o acceso a salas VIP, para alinearse con sus expectativas.

Los pasajeros de este grupo valoran la agilidad y facilidad de los procesos, mostrando alta satisfacción en:

* Reserva online
* Embarque digital
* Ubicación de la puerta de embarque

Cluster 3: Pasajeros mixtos con satisfacción moderada: Predominan los viajeros corporativos, que buscan una buena relación calidad-precio. Es crucial comprender mejor sus preferencias para personalizar la oferta, mejorando aquellos aspectos que podrían fortalecer su satisfacción y fidelización.

Este grupo se distingue por una combinación de características, con énfasis en:

* Servicio a bordo
* Espacio para las piernas
* Alta frecuencia de vuelos al año

**3. Aplicaciones y Utilidad del Análisis**

Este análisis de clustering permite a la aerolínea implementar estrategias más precisas en distintos ámbitos:

* Segmentación de pasajeros: Identificar grupos con necesidades similares para ofrecer productos y servicios más adaptados.
* Estrategias de marketing: Diseñar campañas personalizadas, resaltando los atributos que cada cluster valora más.
* Mejora del servicio: Detectar áreas críticas que impactan la satisfacción, optimizando los aspectos clave en cada segmento.
* Programas de fidelización: Crear incentivos específicos para cada cluster, asegurando que los beneficios sean realmente relevantes para cada tipo de pasajero.

**Conclusión:** La información obtenida proporciona una base sólida para maximizar la satisfacción del pasajero, aumentar la retención y mejorar la rentabilidad mediante estrategias de diferenciación ajustadas a cada segmento de viajeros.

## Desarrollo del Modelo

### **3.6.1.** **Selección del Modelo**

El presente modelo se basa en una red neuronal profunda (DNN - Deep Neural Network) con el objetivo de predecir la satisfacción del pasajero en una aerolínea a partir de diversas características que capturan múltiples dimensiones en su experiencia. La satisfacción del pasajero es un elemento fundamental en la industria de la aviación, ya que influye en la fidelización, reputación de la compañía y toma de decisiones estratégicas para mejorar el servicio.

La satisfacción del pasajero está influenciada por una combinación de todas las variables que se describen en los apartados anteriores.

Por lo tanto, se ha optado por una red neuronal profunda, la cual presenta múltiples ventajas:

* Capacidad de capturar relaciones complejas: El modelo puede aprender de datos en los que la satisfacción del pasajero no sigue una relación lineal con las variables analizadas, optimizando la capacidad del modelo para encontrar patrones en los datos.
* Adaptabilidad a la segmentación de pasajeros (Clusters): El conjunto de datos utilizado ha sido previamente segmentado en 4 clusters, lo que indica que existen diferentes grupos de pasajeros con distintos niveles de expectativas. Una red neuronal se ajusta bien a estos grupos porque aprende automáticamente las diferencias entre cada cluster, lo que permite hacer predicciones más personalizadas y mejorar la toma de decisiones estratégicas.
* Manejo de alta dimensionalidad y volumen de datos: El conjunto de datos incluye 129,880 registros y 24 características predictoras, lo que genera un espacio de información amplio. Las redes neuronales profundas son especialmente adecuadas para manejar grandes volúmenes de datos, ya que pueden procesar múltiples variables simultáneamente sin que la precisión se vea afectada.

Este modelo de red neuronal profunda (DNN) se ha seleccionado debido a su capacidad para captar relaciones complejas, adaptarse a la segmentación de pasajeros y manejar datos de alta dimensión, todo mientras proporciona un desempeño significativamente mejor en comparación con métodos tradicionales. Gracias a esta arquitectura, el modelo no solo predice la satisfacción del pasajero, sino que también permite generar información valiosa para la mejora de los servicios de la aerolínea, facilitando la optimización de la atención al pasajero y la personalización de la experiencia de viaje.

**3.6.2. Preparación de Datos**

Para la implementación del modelo de red neuronal, se ha exportado el conjunto de datos final a un archivo de texto denominado “aerolinea\_final.csv”.

El conjunto de datos utilizado contiene 129.880 registros con 25 características predictoras. Además, se han implementado 4 clústers de pasajeros, lo que permite segmentar el análisis y mejorar la interpretación de resultados.

A continuación, se presenta una tabla que resume las variables consideradas en el análisis, junto con su descripción y tipo de dato:

| **Variable** | **Descripción** | **Tipo de dato** |
| --- | --- | --- |
| Age | The actual age of the passengers | int64 |
| Class | Travel class in the plane of the passengers | int64 |
| Flights per year | Number of flights per year | int64 |
| Flight Distance | The flight distance of this journey | int64 |
| Inflight wifi service | Satisfaction level of the inflight wifi service | int64 |
| Departure/Arrival time convenient | Satisfaction level of Departure/Arrival time convenience | int64 |
| Ease of Online booking | Satisfaction level of online booking | int64 |
| Gate location | Satisfaction level of Gate location | int64 |
| Food and drink | Satisfaction level of Food and drink | int64 |
| Online boarding | Satisfaction level of online boarding | int64 |
| Seat comfort | Satisfaction level of Seat comfort | int64 |
| Inflight entertainment | Satisfaction level of inflight entertainment | int64 |
| On-board service | Satisfaction level of On-board service | int64 |
| Leg room service | Satisfaction level of Leg room service | int64 |
| Baggage handling | Satisfaction level of baggage handling | int64 |
| Check-in service | Satisfaction level of Check-in service | int64 |
| Inflight service | Satisfaction level of inflight service | int64 |
| Cleanliness | Satisfaction level of Cleanliness | int64 |
| Departure Delay in Minutes | Minutes delayed when departure | int64 |
| Arrival delay in minutes | Minutes delayed when arrival | int64 |
| Satisfaction | Airline satisfaction level | int64 |
| Gender Male | Gender of the passengers | int64 |
| Is\_loyal | The customer type | int64 |
| Business Travel | Purpose of the flight of the passengers | int64 |
| Cluster | Clúster 0, 1 , 2 y 3 | Int64 |

###### Tabla 3: Variables incluidas en el modelo de red neuronal profunda (DNN) con tipo y descripción

Durante la lectura del conjunto de datos se detecta una columna denominada “id”, al no aportar información relevante, es eliminada mediante el comando: df = df.drop(columns=['id'])

Para el entrenamiento de modelos de aprendizaje, se considera la división del conjunto de datos de la siguiente manera.

* 75% para entrenamiento
* 25% para pruebas

**3.6.3. Arquitectura del Modelo**

El modelo propuesto sigue una estructura de Red Neuronal Profunda (DNN) con capas completamente conectadas:

model = Sequential([

Dense(64, input\_shape=(X\_train.shape[1],), activation='relu'),

Dense(32, activation='relu'),

Dense(1, activation='sigmoid')

])

Donde:

* La entrada tiene las 24 variables predictoras.
* En la primera capa tiene 64 neuronas con activación ReLU
* En la segunda capa tiene 32 neuronas con activación ReLU
* En la capa de salida tiene 1 neurona con activación sigmoid, lo que permite clasificar la satisfacción del pasajero.

Para la optimización:

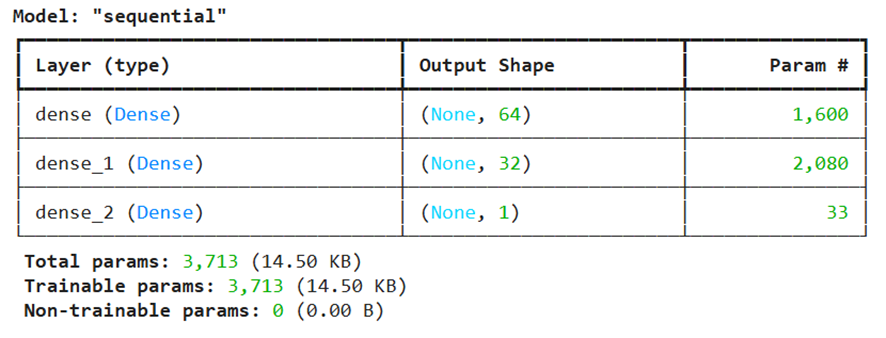
model.compile(optimizer='adam',

loss='binary\_crossentropy',

metrics=['accuracy'])

Donde:

* Función de pérdida es binary cross entropy qué es la adecuada para clasificación binaria.
* Optimizador es Adam, que ajusta los pesos de manera eficiente.
* Métricas de evaluación son accuracy, precisión, recall y F1-score.



### **3.6.4. Entrenamiento del Modelo**

El modelo se ha entrenado con los siguientes hiperparámetros:

* Número de épocas: 20
* Tamaño del batch: 32
* Validación del modelo: validation\_split=0.2

Se obtienen los siguientes resultados:

Para precisión:

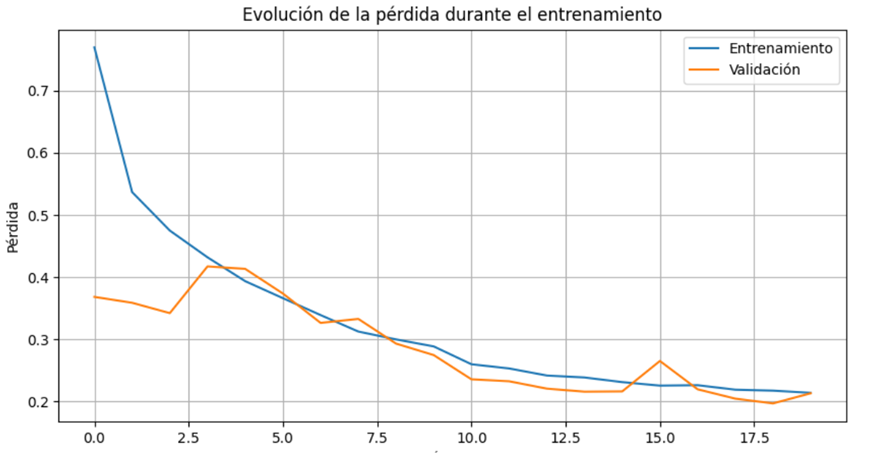
* Inicio: En la primera época, la precisión del entrenamiento es 67.07%, mientras que la precisión en validación alcanza 85.94%
* Final: En la última época, la precisión del entrenamiento aumenta a 91.17% y la precisión en validación alcanza 92.03% en su mejor punto.
* Esto indica que el modelo aprende correctamente con cada iteración, mejorando su capacidad de clasificación.

### 

###### Figura 23: Evolución de la precisión del modelo durante el entrenamiento

Para la pérdida:

* Inicio: La pérdida en entrenamiento comienza en 1.3703, lo que indica un alto error en las predicciones iniciales.
* Final: La pérdida se reduce a 0.2157, lo que muestra que el modelo ha minimizado los errores en la clasificación
* En validación, la pérdida tiene un comportamiento similar y alcanza 0.1971 en la mejor época.

****

###### Figura 24: Reducción de la pérdida (loss) durante el entrenamiento del modelo

## Evaluación del Modelo

Después del entrenamiento, el modelo es probado con datos no vistos (X\_test, y\_test). Las métricas obtenidas en la evaluación son:

* Precisión en test: 90.81%, lo que indica que el modelo clasifica correctamente a los pasajeros en satisfechos e insatisfechos.
* Pérdida final en test: 0.2192, lo que confirma que el error en la clasificación es bajo.

1015/1015 ━━━━━━━━━━━━━━━━━━━━ 2s 1ms/step

- accuracy: 0.9076 - loss: 0.2192

Test Accuracy: 0.9081

La matriz de confusión muestra el desempeño del modelo de red neuronal al identificar correctamente a los pasajeros satisfechos e insatisfechos con una precisión elevada.

Para Precisión:

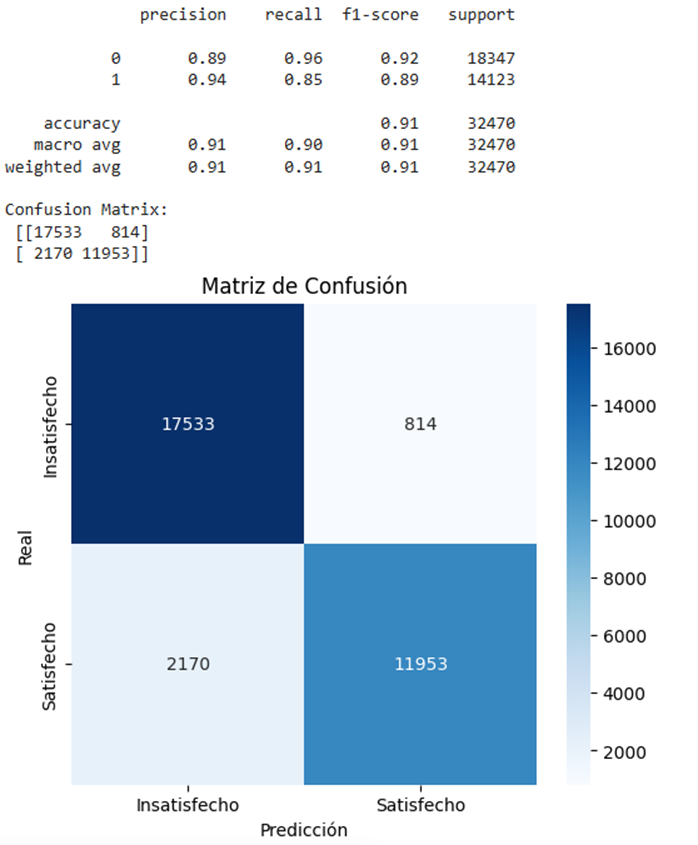
* Para la clase 0 (insatisfecho): 89% cuando el modelo predice a un pasajero como insatisfecho.
* Para la clase 1 (satisfecho): 94% cuando el modelo predice satisfacción.

Para Sensibilidad, capacidad de detección:

* Para la clase 0 (insatisfecho): 96% el modelo identifica correctamente a los pasajeros insatisfechos.
* Para la clase 1 (satisfecho): 85% el modelo identifica correctamente a los pasajeros satisfechos, lo que indica que pierde un pequeño porcentaje de pasajeros satisfechos que clasifica erróneamente como insatisfechos.

Para F1-Score, balance entre precisión y recall:

* Para la clase 0 (insatisfecho): 0.92% clasificación muy buena en insatisfacción.
* Para la clase 1 (satisfecho): 0.89 buen equilibrio, aunque ligeramente menos que en clase 0.



###### Figura 25: Matriz de confusión del modelo de red neuronal

Interpretaciones de valores:

* 17533 pasajeros insatisfechos están correctamente clasificados.
* 814 pasajeros insatisfechos están mal clasificados como satisfechos (falsos positivos)
* 11953 pasajeros satisfechos están correctamente clasificados.
* 2170 pasajeros satisfechos están mal clasificados como insatisfechos (falsos positivos)

El modelo funciona mejor en la detección de pasajeros insatisfechos (96% de acierto), pero pierde algunos pasajeros satisfechos al clasificarlos erróneamente como insatisfechos.

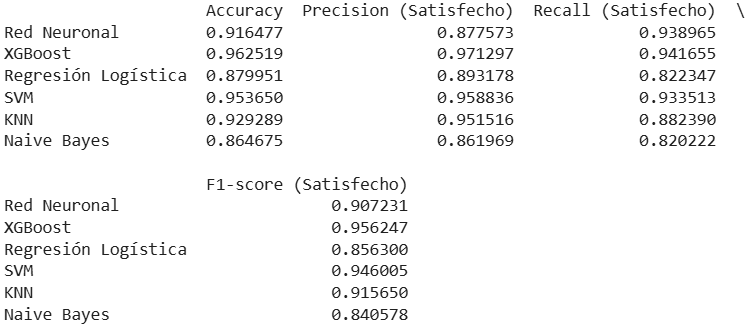
Conclusión y posibles mejoras:

El modelo de red neuronal tiene una precisión general del 91%, lo que indica que es sólido para predecir la satisfacción del pasajero. Sin embargo, se puede mejorar algunos aspectos:

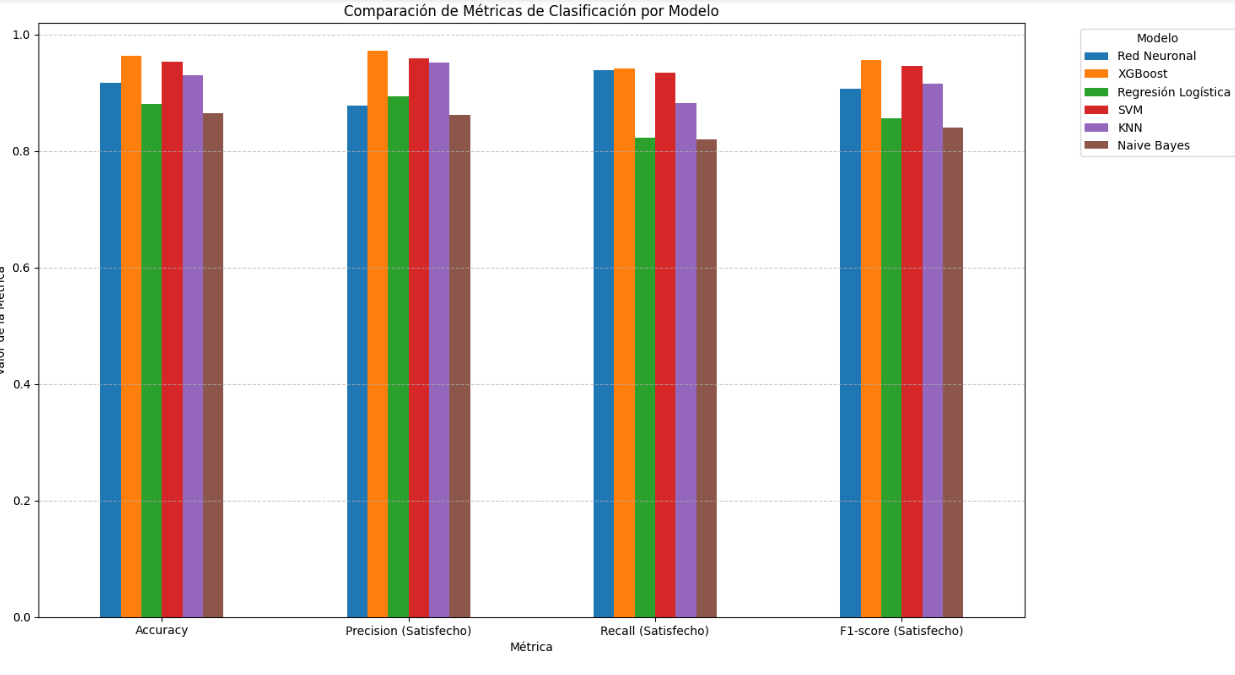
* Ajustar el umbral de clasificación (sigmoid > 0.5) para mejorar la detección de pasajeros satisfechos.
* Implementar Dropout en las capas ocultas para reducir posibles fluctuaciones en validación.
* Probar más neuronas en la capa oculta, modificando la arquitectura de la red.
* Variar el número de épocas y usar Early Stopping para evitar sobreajuste.
  1. **Propuesta de otros Modelos Alternativos**

Para la evaluación de distintos algoritmos de aprendizaje automático es fundamental seleccionar la mejor opción según precisión y rendimiento. En este caso, se compararon seis modelos como la Red Neuronal, XGBoost, Regresión Logística, SVM, KNN y Naive Bayes en términos de accuracy (precisión global), precisión (satisfecho), recall (satisfecho) y F1-score (satisfecho). Teniendo los siguientes resultados:

* Accuracy (Precisión Global): Indica el porcentaje de predicciones correctas. XGBoost destaca con 96.25%, seguido por SVM (95.37%) y Red Neuronal (91.65%).
* Precision (Satisfecho): Representa la proporción de predicciones positivas que fueron correctas. XGBoost obtiene el mayor valor (97.13%), seguido de KNN (95.15%) y SVM (95.88%).
* Recall (Satisfecho): Mide la capacidad del modelo para identificar correctamente todas las instancias positivas. Red Neuronal tiene el mejor desempeño (93.89%), seguido de XGBoost (94.16%) y SVM (93.35%).
* F1-score (Satisfecho): Es un balance entre precisión y recall. XGBoost lidera con 95.62%, seguido de SVM (94.60%) y Red Neuronal (90.72%).



Así mismo tenemos el siguiente diagrama de barras realizando la misma comparativa.



###### Figura 26: Comparación de métricas de rendimiento entre modelos (Accuracy, Precision, Recall, F1-score)

Por lo tanto se afirma que:

* El modelo XGBoost: Presenta el mejor desempeño general, obteniendo altos valores en todas las métricas. Es ideal para tareas de clasificación donde se requiere alto rendimiento y estabilidad
* El modelo SVM: Tiene excelente precisión y recall, lo que lo hace adecuado para problemas con datos bien estructurados y separables.
* Red Neuronal: Aunque muestra un desempeño sólido, su rendimiento es ligeramente inferior al de XGBoost y SVM en algunas métricas

# ROI de los Casos de Uso

**4.1. Propuestas de casos de uso basados en el análisis predictivo de satisfacción.**

* Compensaciones dinámicas de comida a bordo por demoras o cancelaciones: Cuando el modelo predictivo identifica que un vuelo presenta un alto riesgo de demora o cancelación, puede activar un mecanismo automático de emisión de vouchers o compensaciones preaprobadas para comida a bordo o en aeropuerto. Se brinda una oferta de compensación anticipada donde el pasajero recibe un cupón digital para comprar alimentos o bebidas en salas VIP o restaurantes del aeropuerto o a bordo. Esta acción proactiva genera percepción de cuidado y reduce la frustración. La finalidad es la reducción de quejas y reclamaciones, al ofrecer una solución inmediata, se mitiga la insatisfacción.
* Asistencia anticipada a pasajeros vulnerables: Identificar pasajeros con características de vulnerabilidad como edad avanzada, necesidades especiales o historial de reclamaciones para anticipar acciones de apoyo como acompañamiento en aeropuerto, embarque asistido o acceso preferente. Para ello, se detecta a partir de datos de reserva como la edad, historial de servicios solicitados para etiquetarlos en el modelo como “vulnerable”. Se tiene alertas y notificaciones via app o SMS antes y durante el vuelo, se envían recomendaciones personalizadas como guía de aeropuerto o asistencia en tránsito de conexiones. La aerolínea puede asignar un personal de apoyo en puntos clave como en el counter de check in a la espera del pasajero que requiere la atención especial. La finalidad es la reducción de incidencias para pasajeros vulnerables y reclamaciones posteriores. La mejora significativa de marca especialmente en segmentos con alta sensibilidad como pasajeros mayores, familias con niños pequeños y personas con discapacidad.

## 4.2. Desglose de inversiones

### **4.2.1. Infraestructura tecnológica y licencias**

Para el Almacenamiento y procesamiento de datos como servidores Cloud.

* Inversión inicial estimada: 80.000 €: Basada en cálculos de proveedores cloud para un entorno de datos de tamaño medio (ITrexGroup, 2023), la infraestructura puede costar entre $150–$300/mes para CPU básicas, hasta varios miles de dólares/mes si se requieren instancias de mayor capacidad o GPU para tareas de ML intensivas. Para un proyecto piloto con datos de vuelos, se asume un desembolso inicial en aprovisionamiento, configuración y pruebas de aproximadamente 80.000 € (equivalente a ~USD 85.000 en coste real de ingeniería e infraestructura).
* Mantenimiento anual: 20.000 €: Considera los costes de operación continua, almacenamiento incremental, instancias de procesamiento de bajo uso, backups, monitorización y pequeñas ampliaciones a lo largo del año (ITrexGroup, 2023), estimados en 20.000 € anuales para una infraestructura de pilotaje y despliegue moderado

Para la Plataforma ML / herramientas de integración: Incluye licencias o servicios de MLOps (plataformas SaaS o licencias empresariales), API de estado de vuelo en tiempo real y sistemas de notificaciones (SMS/app, servicios de mensajería y gestión de vouchers).

* Inversión inicial: 40.000 €: Plataformas de MLOps comerciales o SaaS como Databricks, API de estado de vuelo en tiempo real como FlightAware (ITrexGroup, 2023). Estos elementos suman un coste cercano a 40.000 € iniciales (configuración, licencias mínimas y pruebas).

Para Mantenimiento anual: 15.000 €: Renovación de licencias de MLOps en piloto, suscripción a APIs de vuelo y servicios de mensajería mínima para el volumen estimado en fase de producción inicial (ITrexGroup, 2023), asumido en unos 15.000 € anuales.

### **4.2.2. Desarrollo e integración**

Para el Equipo Data Science:

* 2 profesionales senior durante 6–9 meses para construir modelos predictivos básicos (demoras/cancelaciones y detección de pasajeros vulnerables).
* Coste aproximado (Glassdoor, 2025): 2 × (70.000 €/año prorrateado a 9 meses) = 105.000 €
* Basado en:
  + Salario bruto medio de Data Scientist senior en España: según Glassdoor, el salario promedio ronda 41.000 USD/año (aprox. 38.000 €) y (Levels.fyi, 2025) sitúa rangos en Barcelona de €46.9K–€80.4K para Data Scientists.
  + El costo total para la empresa (salario bruto + cargas sociales + beneficios) se estima en torno a 1.3× salario bruto, lo que lleva a un coste por FTE senior de 70.000 € al año.
  + Prorrateo a 9 meses: 70.000 € × 0.75 = 52.500 € por profesional; para 2 profesionales = 105.000 €.

Para el equipo de desarrollo/DevOps

* 2 ingenieros para integración de pipelines, notificaciones automáticas de vouchers y alertas de asistencia (Glassdoor, 2025).
* Coste: 1,5 × (60.000 €/año prorrateado a 9 meses) = 67.500 €

o El salario bruto medio de un ingeniero DevOps/Backend senior en España es de 50.000 € – 60.000 €/año, con costo completo es 60.000 €/año.

o Prorrateo a 9 meses: 60.000 € × 0.75 = 45.000 €; con 1,5 FTE (para cubrir picos o complementariedad de perfiles) = 67.500 €.

Para la consultoría externa y auditoría de datos

* Evaluación de calidad de datos de reservas y segmentación de pasajeros vulnerables.
* Coste estimado: 30.000 €

o Basado en tarifas de KodyTechnolab. (2025), en Europa Occidental (aprox. $100–$150/h) y alcance de varias semanas de trabajo de analistas y expertos en datos de aviación.

o Por ejemplo, si un equipo de 2 consultores trabaja 3–4 semanas a jornada parcial o total, se llega a una cifra de decenas de miles de euros.

### **4.2.3. Formación y gestión del cambio**

Para la formación de equipos operativos (atención al cliente, personal de tierra) en nuevos flujos automáticos (emisión de vouchers, alertas de asistencia)

El Coste es de 15.000 €: Basado en programas y workshops de transformación digital en aviación (LPC Training, 2025) como cursos especializados que oscilan en decenas de miles de euros, adaptados al número de participantes y duración como e-learning, sesiones presenciales o híbridas.

En la comunicación interna y roles de Product Owner: Talleres, documentación y coordinación de stakeholders, el coste es de 20.000 €: Incluye horas de reuniones, elaboración de documentación, workshops internos, consultoría ligera en gestión del cambio (LPC Training, 2025), apoyado por prácticas de gestión de proyectos ágiles y transformación digital en aerolíneas, donde se reservan partidas para comunicación y training interno.

### **4.2.4. Operación y mantenimiento**

Para Soporte Data Science/IT tras lanzamiento (Glassdoor, 2025): 1 Data Scientist/ingeniero ML a tiempo completo a 70.000 € al añ**o** para monitorizar modelos, ajustes y nuevos datos.

Para los costes de licencias y APIs (mantenimiento continuo): El costo es de 35.000 € al año: Incluye renovación de licencias MLOps en fase productiva, suscripciones a API de estado de vuelo en tiempo real, servicios de mensajería/SMS/app y operación de infraestructura cloud básica. Basado en estimaciones de renovaciones anuales para volumen medio de peticiones y uso, consistente con costes pilotos de MLOps. Las Referencias de costes cloud y API: ITRex indica que infra básica puede costar $1.460 –$3.600 al año para CPU bajas, y APIs de vuelo suelen presupuestarse aparte según volumen de llamadas.

Para los Ajustes y mejoras menores como optimización de modelos y actualizaciones (Coherent Solutions, 2025), el coste es de 30.000 €/año: El esfuerzo continuo para reentrenar, recalibrar umbrales, incorporar nuevos datos o variables, mejoras en pipelines y pequeños cambios de funcionalidad. Basado en experiencias de proyectos ML que destinan un 20–30% del coste inicial de desarrollo en mantenimiento evolutivo.

### **4.2.5. Totales**

* Total de inversión año 1:

o Infraestructura + licencias: 80.000 € + 40.000 € = 120.000 €

o Desarrollo e integración: 105.000 € + 67.500 € + 30.000 € = 202.500 €

o Formación y gestión cambio: 15.000 € + 20.000 € = 35.000 €

o Total aproximado en año 1: ~357.500 €

* Costes recurrentes años 2 y 3:

o Soporte Data Science/IT: 70.000 €

o Licencias/APIs/infraestructura: 35.000 €

o Ajustes y mejoras menores: 30.000 €

o Total aproximado anual: ~135.000 €

## 4.3. Estimación de Beneficios

Para compensaciones dinámicas de comida a bordo por demoras o cancelaciones: Según la regulación europea, las aerolíneas deben proporcionar comidas o refrescos en relación con el tiempo de espera en retrasos prolongados (Reglamento (CE) No 261/2004). La emisión proactiva de vouchers mejora la percepción del pasajero y puede reducir reclamaciones formales posteriores; prácticas en la industria muestran que ofrecer refreshment temprano alivia frustración.

* Coste medio actual de compensación reactiva (LPC Training, 2025): Incluye reembolso o compensaciones como el hotel, comidas, transporte alternativo, atención al cliente, se asume 50 € por pasajero afectado
* Coste medio con mecanismo proactivo: Emisión de voucher de comida/refreshment anticipado: valor nominal 10 € (solo si el modelo predice riesgo > umbral). Si se confirma demora grave o cancelación, se añade complemento de 15 € adicionales; coste total en eventos confirmados: 25 € promedio. Si no se confirma, el voucher se revoca o se reemplaza por un beneficio menor por ejemplo 5 € en millas, pero para simplificar, asumimos coste neto de 10 € en eventos no confirmados.
* Número de eventos/anuales: En el Supuesto de 50.000 vuelos afectables al año según el histórico de riesgo moderado o alto de demora y 1 pasajero promedio por vuelo que activa el voucher.
* Cálculo de ahorro según Reglamento CE 261/2004:

o Eventos confirmados (40% de 50.000 = 20.000): ahorro por evento 50 € − 25 € = 25 €, por lo tanto, 20.000 × 25 € = 500.000 €.

o Eventos no confirmados (60% de 50.000 = 30.000): coste proactivo 10 € frente a coste cero si no se emite en reactivo ⇒ 30.000 × 10 € = 300.000 €

o Gasto total proactivo: 800.000 €

o Coste total reactivo (sin mecanismo): 20.000 × 50 € = 1.000.000 €

o Ahorro directo anual estimado: 1.000.000 € − 800.000 € = 200.000

o Reducción de reclamaciones formales y carga de atención: Si se reducen reclamaciones en un 20% sobre 20.000 reclamaciones potenciales, con un coste de atención de 30 €/reclamación, por lo tanto, 20.000 × 0.20 × 30 € = 120.000 €.

o Beneficio total asociado: 200.000 € + 120.000 € = 320.000 € anuales

Para la asistencia anticipada a pasajeros vulnerables: La asistencia a pasajeros con movilidad reducida o necesidades especiales es obligatoria y valorada en la satisfacción del pasajero (IATA Global Passenger Survey refleja alta valoración cuando se presta asistencia adecuada). En los procesos de atención planificada pueden reducir costes operativos y tiempos de espera: soluciones de self-service y asignación anticipada de personal muestran ahorros en personal y mejora de throughput (informes de SITA/Valour Consultancy).

* Volumen de pasajeros vulnerables: De 10 millones de pasajeros anuales, el 2% (200.000 pasajeros) se etiquetan como “vulnerables” (edad avanzada, familias con niños, necesidades especiales)
* Coste medio actual de incidencias y reclamos en este segmento: Los reclamos incluyen retrasos en asistencia, por mal manejo, tiempo extra de personal; Se asume 40 € por pasajero vulnerable, entonces el total de coste actual: 200.000 × 40 € = 8.000.000 €.
* Coste de la asistencia anticipada: Las notificaciones, personal asignado proactivamente y coordinación, el coste medio estimado es 15 € por pasajero, entonces el total coste asistencia anticipada: 200.000 × 15 € = 3.000.000 € anuales.
* Reducción de incidencias y reclamos:

o Se supone que la asistencia anticipada reduce incidencias en un 50%

o Las incidencias evitadas: 100.000 eventos × 40 € = 4.000.000 € ahorrados

o El Beneficio neto: 4.000.000 € − 3.000.000 € = 1.000.000 € anuales

· Beneficio adicional de fidelización (IATA, 2024): Si el segmento vulnerable tiene un valor de ciclo de vida promedio de 200 €/año y la mejora incrementa retención en 2% sobre 200.000 pasajeros (4.000 pasajeros retenidos más), beneficio = 4.000 × 200 € = 800.000 €. Por lo que, se puede imputar 50%: 400.000 €.

* Beneficio total asociado:

o Ahorro neto en incidencias/reclamaciones: 1.000.000 €.

o Beneficio por retención (50%): 400.000 €.

o Total anual estimado: 1.400.000 €.

## 4.4. Cálculo de ROI y VAN

### **4.4.1. Flujos de caja estimados**

* Año 1 (piloto e implementación inicial)

o Inversión total año 1: −357.500 €.

o Beneficios año 1 (fase piloto, 30% de beneficio esperado): Compensaciones dinámicas: 0,3 × 320.000 € = 96.000 €, Asistencia vulnerables: 0,3 × 1.400.000 € = 420.000 €, Total beneficio año 1: 516.000 €.

o Flujo neto año 1: 516.000 € − 357.500 € = +158.500 €.

* Año 2 (despliegue completo)

o Costes recurrentes: −135.000 €.

o Beneficios plenos: Compensaciones dinámicas: 320.000 €. Asistencia a vulnerables: 1.400.000 €. Total de beneficio año 2: 1.720.000 €.

o Flujo neto año 2: 1.720.000 € − 135.000 € = +1.585.000 €.

* Año 3 (operación estable y optimización)

o Costes recurrentes: −135.000 €.

o Beneficios año 3 (ligera mejora 5%): Compensaciones dinámicas: 320.000 € × 1,05 = 336.000 €. La asistencia a vulnerables: 1.400.000 € × 1,05 = 1.470.000 €. Total de beneficio año 3: 1.806.000 €.

o Flujo neto año 3: 1.806.000 € − 135.000 € = +1.671.000 €.

### **4.4.2. VAN al 8% de descuento**

* Año 1 (t=1): 158.500 € / 1.08 ≈ 146.759 €.
* Año 2 (t=2): 1.585.000 € / (1.08²) ≈ 1.358.453 €.
* Año 3 (t=3): 1.671.000 € / (1.08³) ≈ 1.324.231 €.
* VAN total ≈ 2.829.443 €.
* Al ser positivo, se considera la inversión rentable bajo estos supuestos.

### **4.4.3. Payback**

En los flujos acumulados:

* Fin año 1: +158.500 € (queda pendiente recuperar 199.000 € de la inversión inicial).
* Fin año 2: +1.743.500 € acumulados; la inversión inicial se recupera en algún punto de ese año 2.

Periodo de recuperación: alrededor de 1,2 años. Esto indica un payback relativamente rápido.

### **4.4.4. Tasa Interna de Retorno (TIR)**

* Serie de flujos: Año 0: −357.500 €; Año 1: +516.000 €; Año 2: +1.585.000 €; Año 3: +1.671.000 €.
* La TIR resultará muy superior al 8% de referencia, reforzando la viabilidad. Se aconseja calcularla con herramientas financieras (Excel, Python, calculadora financiera) para el valor exacto.

## 4.5. Análisis de sensibilidad

Para el escenario conservador (beneficios 50%)

* Compensaciones dinámicas: 160.000 € anuales; Asistencia vulnerables: 700.000 €.
* Año 1: 0,3 × (160.000 + 700.000) = 258.000 €; Flujo neto = 258.000 − 357.500 = −99.500 €.
* Año 2: 860.000 − 135.000 = 725.000 €.
* Año 3: 903.000 − 135.000 = 768.000 €.
* VAN ≈ −92.130 + 622.041 + 608.468 = 1.138.379 € > 0.
* Payback en año 2. Aunque en el piloto da negativo, el proyecto sigue siendo positivo en el horizonte de 3 años.

Para el Escenario optimista (+20% beneficios)

* Compensaciones: 384.000 €; Asistencia: 1.680.000 €.
* Año 1: 0,3 × 2.064.000 = 619.200 €; Flujo neto = 619.200 − 357.500 = 261.700 €.
* Año 2: 2.064.000 − 135.000 = 1.929.000 €.
* Año 3: 2.167.200 − 135.000 = 2.032.200 €.
* VAN ≈ 3.514.136 €.
* Proyecto aún más atractivo.

## 4.6. Roadmap de implementación

· Fase 0: Preparación de datos y gobernanza (2–3 meses)

* Auditoría de calidad de datos de reservas, perfil de pasajeros y datos históricos de demoras (consultoría externa).
* Asegurar consentimiento para uso de datos de perfil.
* Definir KPIs (p. ej., tasa de falsos positivos en demora, reducción de incidencias en vulnerables).

· Fase 1: Piloto compensaciones dinámicas y detección de vulnerables (4–6 meses)

* Desarrollar modelo de riesgo de demora/cancelación con histórico y APIs (meteorología, tráfico aéreo).
* Implementar MVP de emisión de vouchers: integrar mensajería (app/SMS) y sistema de vouchers digitales.
* Definir lógica de etiquetado de pasajeros vulnerables a partir de datos de reserva; integración manual inicial de asignación de personal de apoyo en rutas seleccionadas.
* Medir en subconjunto de rutas: emisión vs. ahorro, feedback de personal y pasajeros.

· Fase 2: Ajuste y escalado parcial (6–8 meses)

* Calibrar umbrales del modelo de demora para minimizar falsos positivos.
* Refinar proceso de asistencia vulnerable: confirmaciones previas y coordinación con ground handling.
* Extender a más rutas/mercados: priorizar donde haya más pasajeros vulnerables y retrasos frecuentes.
* Desarrollar dashboard en tiempo real: número de vouchers emitidos, validaciones, ahorro acumulado; asistencias programadas e incidencias evitadas.

· Fase 3: Operación completa y optimización continua (9–12 meses)

* Modelo en producción con pipelines MLOps para reentrenamiento periódico y monitorización de drift.
* Automatización total de identificación de vulnerables con confirmaciones automáticas y asignación de personal mediante sistema de gestión de ground handling.
* Monitoreo continuo y A/B testing (variaciones de valor de voucher, niveles de servicio de asistencia).
* Feedback loop: datos reales de ahorro y satisfacción para ajustar supuestos financieros.

· Fase 4: Innovación y expansión de alcance (a partir de año 2)

* Vincular incentivos de fidelización cuando se usan vouchers o asistencia; estudiar efectos en retención global.
* Integrar comunicaciones proactivas complementarias (alertas de puerta, estado de embarque) que refuercen estos casos de uso.
* Revisar periódicamente el ROI con datos reales y ajustar presupuestos y prioridades.

## 4.7. Conclusiones del ROI ajustado

* **Rentabilidad**: Con inversión focalizada ~357.500 € en primer año y costes recurrentes ~135.000 €/año, los beneficios directos (ahorro en compensaciones y reducción de incidencias de pasajeros vulnerables más retención) generan un VAN positivo (~2,8 M€) y payback corto (~1,2 años) en escenario base.
* **Importancia del piloto**: En escenario conservador, el piloto podría dar flujo negativo en año 1; por ello conviene diseñar piloto escalonado que valide supuestos y ajuste umbrales antes de desplegar completamente.
* **Impacto cualitativo**: Mejora de percepción de la aerolínea al gestionar demoras y vulnerabilidades de forma proactiva, reforzando marca y fidelidad.
* **Riesgos gestionables**: Con gobernanza de datos, calibración de modelos y buena gestión del cambio, se minimizan riesgos de baja precisión y adopción.

· **Recomendación**:

* Iniciar un piloto sólido con métricas claras y recursos limitados en rutas con alta incidencia de demoras y elevado porcentaje de pasajeros vulnerables.
* Ajustar rápidamente tras resultados iniciales.
* Escalar a toda la operación si los KPIs de ahorro y satisfacción confirman los supuestos.
* Monitorear continuamente y recalcular ROI con datos reales para futuras inversiones y expansión de casos de uso.

1. **Conclusiones**

Este trabajo ha demostrado el potencial de los modelos predictivos basados en machine learning para generar valor real dentro del sector de la aviación comercial. A través del análisis profundo de los datos disponibles de pasajeros y vuelos, hemos podido construir una herramienta que no solo permite entender mejor los factores que influyen en la satisfacción del cliente, sino también anticiparse a experiencias negativas y tomar decisiones proactivas.

Desde un punto de vista empresarial, la propuesta va más allá de un simple modelo técnico: se trata de una solución que puede integrarse fácilmente en los sistemas actuales de una aerolínea para mejorar la relación con el cliente y optimizar la operativa diaria. El hecho de poder predecir con antelación qué pasajeros tienen mayor probabilidad de insatisfacción permite actuar de forma personalizada y eficiente, algo especialmente importante en un sector donde la experiencia del cliente es clave para la fidelización y la reputación. Además, se ha podido comprobar que la mayoría de las variables necesarias para alimentar este modelo ya están disponibles en las bases de datos operativas de las aerolíneas, lo que facilita su implementación, por lo que, la incorporación de este tipo de herramientas puede suponer una ventaja competitiva clara, al permitir ofrecer un servicio más ajustado a las expectativas del pasajero, reducir quejas y generar una percepción de marca mucho más positiva.

El análisis de correlación, la selección de variables representativas y las técnicas de reducción de dimensionalidad aplicadas han servido no solo para optimizar el rendimiento del modelo, sino también para identificar de forma clara cuáles son los factores que más afectan a la satisfacción. Esta información, además de alimentar el modelo, puede guiar decisiones estratégicas en otras áreas: mejoras en el servicio, diseño de nuevas tarifas, campañas de fidelización, o decisiones operativas como la asignación de recursos en vuelos con alta probabilidad de incidencias.

En resumen, se ha desarrollado una solución realista, escalable y orientada al cliente, que utiliza datos ya disponibles para transformar la experiencia del pasajero en una fuente de valor para la compañía aérea.

## Principales resultados/conclusiones e implicaciones

Los principales resultados obtenidos en este trabajo pueden agruparse en dos grandes bloques: la validación técnica del modelo predictivo y el diseño de una propuesta de negocio viable en torno a su aplicación.

Desde el punto de vista técnico, se ha demostrado que es posible predecir el nivel de satisfacción de un pasajero con una precisión considerable utilizando únicamente variables que la aerolínea ya recoge en sus operaciones habituales. El análisis ha revelado que existen patrones claros: los retrasos, la clase del billete, la atención a bordo o la limpieza son aspectos que tienen un impacto significativo en la percepción del cliente. A partir de esta información, se han implementado modelos de clasificación binaria (satisfecho/no satisfecho), con un rendimiento adecuado tanto en métricas de precisión como de recall.

Desde el punto de vista estratégico, las implicaciones son relevantes. La herramienta puede aplicarse en distintos momentos del viaje: antes del embarque para alertar de posibles pasajeros en riesgo de insatisfacción, durante el vuelo para personalizar el servicio, o después del mismo para activar acciones correctivas o preventivas (como encuestas específicas, compensaciones automáticas o mejoras en vuelos futuros).

Además, si se dispusiera de información adicional (como el destino, el motivo del viaje, el alojamiento, la existencia de quejas anteriores, o incluso variables emocionales a partir de feedback), se podría llevar esta propuesta aún más lejos. Por ejemplo, se podrían generar alertas automáticas para que el personal en tierra reciba a un pasajero con un problema o para que se le ofrezca un pequeño obsequio en el hotel si el vuelo tuvo una incidencia importante.

También se abre la puerta a futuros desarrollos como sistemas de recomendación personalizados, mejoras dinámicas en tiempo real según la situación del vuelo, o dashboards operativos que permitan a los responsables de atención al cliente priorizar su tiempo de forma más eficiente.

En definitiva, el modelo desarrollado no solo cumple una función predictiva, sino que permite a la aerolínea evolucionar hacia una gestión más proactiva, empática y centrada en el cliente. Esto representa un avance importante en la digitalización de la experiencia de vuelo, y marca una línea clara de cómo los datos pueden convertirse en una herramienta de transformación empresarial real.

## Limitaciones y recomendaciones para futuros trabajos

A pesar de los resultados obtenidos, el presente trabajo presenta una serie de limitaciones que es importante tener en cuenta:

El modelo se basa en un conjunto de datos que, aunque representativo, puede no reflejar completamente la diversidad de contextos y aerolíneas a nivel global. Algunos atributos clave para entender la satisfacción del pasajero, como factores culturales, expectativas personales o circunstancias específicas del vuelo (clima, retrasos imprevistos, trato del personal), no están representados o presentan valores faltantes.

La satisfacción del pasajero es una variable compleja, influida por factores emocionales y contextuales que pueden cambiar de forma dinámica. Por tanto, su medición a través de escalas estandarizadas o encuestas puede no capturar toda la riqueza de la experiencia real del cliente.

Es posible que existan sesgos en la recopilación de datos, por ejemplo, una sobrerrepresentación de pasajeros insatisfechos (quienes tienen más probabilidad de completar encuestas) o de ciertos segmentos demográficos. Esto puede afectar la generalización del modelo.

El modelo desarrollado es estático, es decir, se entrena una vez con un conjunto de datos fijo. Sin mecanismos de actualización automática o aprendizaje en tiempo real, su rendimiento puede degradarse con el tiempo si cambian las condiciones del mercado o las expectativas de los pasajeros.

Además, con el objetivo de enriquecer la investigación, sería de relevancia dar continuación a este proyecto tratando las siguientes vías:

Incorporar el Net Promoter Score (NPS) como variable objetivo podría ser especialmente útil. El NPS refleja la lealtad del cliente y su probabilidad de recomendar la aerolínea. Predecirlo por segmento ayudaría a detectar grupos con alto riesgo de abandono y a enfocar mejoras donde generen mayor impacto en la reputación y fidelización.

También sería valioso ampliar el análisis con fuentes de feedback no estructurado, como comentarios en redes sociales, encuestas abiertas o reseñas. Técnicas de procesamiento de lenguaje natural (NLP) permitirían capturar aspectos emocionales o específicos que influyen en la satisfacción y que no están presentes en las variables numéricas.

Por último, una mejora clave sería implementar modelos adaptativos o en tiempo real. El enfoque actual es estático y podría perder precisión si cambian los patrones de comportamiento. Usar aprendizaje continuo o modelos en streaming permitiría actualizar predicciones de forma dinámica y tomar decisiones más ágiles en operaciones.

# Bibliografía

## Publicaciones

* Eshaghi, M. S., Afshardoost, M., Lohmann, G., & Moyle, B. D. (2024). Drivers and outcomes of airline passenger satisfaction: A meta-analysis. Journal of the Air Transport Research Society, 3, 100034. <https://doi.org/10.1016/j.jatrs.2024.100034>
* Afshardoost, M., & Eshaghi, M. S. (2020). Destination image and tourist behavioural intentions: A meta-analysis. *Tourism Management*, *81*, 104154. <https://doi.org/10.1016/j.tourman.2020.104154>
* Hassan, T. H., & Salem, A. E. (2022). Impact of service quality of low-cost carriers on airline image and consumers’ satisfaction and loyalty during the COVID-19 outbreak. *International Journal of Environmental Research and Public Health, 19*(1), 83. <https://doi.org/10.3390/ijerph19010083>
* Akamavi, R. K., Mohamed, E., Pellmann, K., & Xu, Y. (2015). Key determinants of passenger loyalty in the low-cost airline business. *Tourism Management, 46*, 528–545. <https://doi.org/10.1016/j.tourman.2014.07.010>
* Shen, C., & Yahya, Y. (2021). The impact of service quality and price on passengers’ loyalty towards low-cost airlines: The Southeast Asia perspective. *Journal of Air Transport Management, 91*, 101966. <https://doi.org/10.1016/j.jairtraman.2020.101966>
* Tam, C., Barroso, M., & Cruz-Jesus, F. (2022). Understanding the determinants of users’ continuance intention to buy low-cost airline flights online. *Journal of Hospitality and Tourism Technology*. <https://doi.org/10.1108/JHTT-08-2021-0211>
* Kondo, Y. (2001). Customer satisfaction: How can I measure it? *Total Quality Management, 12*(7–8), 867–872. <https://doi.org/10.1080/09544120100000044>
* Tahanisaz, S. (2020). Evaluation of passenger satisfaction with service quality: A consecutive method applied to the airline industry. *Journal of Air Transport Management, 83*, Article 101764. <https://doi.org/10.1016/j.jairtraman.2020.101764>
* Singh, B. (2021). Predicting airline passengers’ loyalty using artificial neural network theory. *Journal of Air Transport Management, 94*, Article 102080. <https://doi.org/10.1016/j.jairtraman.2021.102080>
* Rita, P., Moro, S., & Cavalcanti, G. (2022). The impact of COVID-19 on tourism: Analysis of online reviews in the airlines sector. *Journal of Air Transport Management, 104*, Article 102277. <https://doi.org/10.1016/j.jairtraman.2022.102277>
* Jenatabadi, H. S., & Ismail, N. A. (2014). Application of structural equation modelling for estimating airline performance. Journal of Air Transport Management, 40, 25–33. <https://doi.org/10.1016/j.jairtraman.2014.05.005>
* Ahmadpour, N., Lindgaard, G., Robert, J. M., & Pownall, B. (2014). The thematic structure of passenger comfort experience and its relationship to the context features in the aircraft cabin. Ergonomics, 57(6), 801–815. <https://doi.org/10.1080/00140139.2014.899632>
* Tadajewski, M. (2010). Critical marketing studies: logical empiricism, ‘critical performativity’ and marketing practice. Marketing Theory, 10(2), 210-222. (Original work published 2010) <https://doi.org/10.1177/1470593110366671>
* Ekinci, Y., Dawes, P. L., & Massey, G. R. (2008). An extended model of the antecedents and consequences of consumer satisfaction for hospitality services. European Journal of Marketing, 42(1/2), 35–68. <https://doi.org/10.1108/03090560810840907>

## Páginas web

* Asociación de Líneas Aéreas. (2023). *Un nuevo análisis de la industria de las aerolíneas revela la popularidad de las tarjetas de crédito y los programas de fidelización de las aerolíneas y su impacto en las economías locales*.<https://ala.aero/es/2023/09/un-nuevo-analisis-de-la-industria-de-las-aerolineas-revela-la-popularidad-de-las-tarjetas-de-credito-y-los-programas-de-fidelizacion-de-las-aerolineas-y-su-impacto-en-las-economias-locales/>
* Cadena SER. (2025, 26 de marzo). *Indignación entre mayores gallegos tras un vuelo del Imserso: “Nos trataron como si no importáramos”*. Radio Coruña.<https://cadenaser.com/galicia/2025/03/26/indignacion-entre-mayores-gallegos-tras-un-vuelo-del-imserso-nos-trataron-como-si-no-importaramos-radio-coruna/>
* El País. (2025, 28 de marzo). *Trece años de un machine learning de muy altos vuelos*.<https://elpais.com/tecnologia/branded/especial-innovacion/2025-03-28/trece-anos-de-un-machine-learning-de-muy-altos-vuelos.html>
* Glassdoor. (2025). *Data Scientist Salaries in Spain*. Recuperado de https://www.glassdoor.com/Salaries/spain-data-scientist-salary-SRCH\_IL.0,5\_IN219\_KO6,20.htm (consultado junio 2025) [glassdoor.com](https://www.glassdoor.com/Salaries/spain-data-scientist-salary-SRCH_IL.0%2C5_IN219_KO6%2C20.htm?utm_source=chatgpt.com).
* Levels.fyi. (2025). *Data Scientist Salary in Barcelona, Spain*. Recuperado de https://www.levels.fyi/t/data-scientist/locations/barcelona-esp (consultado 13 de junio de 2025) [levels.fyi](https://www.levels.fyi/t/data-scientist/locations/barcelona-esp?utm_source=chatgpt.com).
* ITrexGroup. (2023). *Machine Learning Costs: Price Factors and Real-World Estimates*. Recuperado de https://itrexgroup.com/blog/machine-learning-costs-price-factors-and-estimates/ (consultado junio 2025) [itrexgroup.com](https://itrexgroup.com/blog/machine-learning-costs-price-factors-and-estimates/?utm_source=chatgpt.com).
* KodyTechnolab. (2025). *IT Consulting Hourly Rates in 2025: Western Europe $100–$150/h*. Recuperado de https://kodytechnolab.com/blog/it-consulting-hourly-rates/ (consultado junio 2025) [kodytechnolab.com](https://kodytechnolab.com/blog/it-consulting-hourly-rates/?utm_source=chatgpt.com).
* LPC Training. (2025). *Digital Transformation in Aviation program*. Recuperado de https://www.lpcentre.com/it-and-computer-science/digital-transformation-in-aviation (consultado junio 2025) [lpcentre.com](https://www.lpcentre.com/it-and-computer-science/digital-transformation-in-aviation?utm_source=chatgpt.com).
* Coherent Solutions. (2025). *AI Development Cost Estimation: Pricing Structure, Implementation ...*. Recuperado de https://www.coherentsolutions.com/insights/ai-development-cost-estimation-pricing-structure-roi (consultado junio 2025) [coherentsolutions.com](https://www.coherentsolutions.com/insights/ai-development-cost-estimation-pricing-structure-roi?utm_source=chatgpt.com).
* Reglamento (CE) No 261/2004. *Regulation (EC) No 261/2004 of the European Parliament and of the Council*. Recuperado de EUR-Lex (consultado junio 2025) [coherentsolutions.com](https://www.coherentsolutions.com/insights/ai-development-cost-estimation-pricing-structure-roi?utm_source=chatgpt.com).
* IATA. (2024). *Global Passenger Survey*. [Informe sobre satisfacción y valoración de asistencia a pasajeros con necesidades especiales] [levels.fyi](https://www.levels.fyi/companies/amazon/salaries/data-scientist/locations/spain?utm_source=chatgpt.com).
* SITA / Valour Consultancy. (2023). *Air Transport IT Insights: Passenger Survey; estudios de throughput y atención a pasajeros vulnerables* [glassdoor.com](https://www.glassdoor.com/Salaries/madrid-spain-data-scientist-salary-SRCH_IL.0%2C12_IM1030_KO13%2C27.htm?utm_source=chatgpt.com).

# Anexo: Código