

ISDI

Titulación: MDA 11

Modelización avanzada de perfiles de usuarios en Airbnb mediante clusterización de reseñas: optimización para anfitriones, huéspedes y plataforma

Elena Santamaría Izquierdo
Lucía Millán
28/06/2025

Resumen

Este proyecto se enmarca en el campo de la inteligencia artificial aplicada al análisis de datos en plataformas digitales, específicamente en el sector del turismo colaborativo. El objetivo principal es mejorar la experiencia de usuario en Airbnb mediante la identificación de perfiles de huéspedes a través del análisis masivo de reseñas.

El problema abordado radica en la falta de personalización en la interacción entre anfitriones y huéspedes, lo que impacta negativamente en la satisfacción del cliente, reducción de tasa de reservas del anfitrión y en la eficiencia operativa de la plataforma. Para resolverlo, se propone un modelo basado en técnicas de procesamiento del lenguaje natural (NLP) y aprendizaje no supervisado.

La investigación se desarrolló bajo la metodología ágil SCRUM, con sprints quincenales con el Product Owner y reuniones semanales del equipo, lo que permitió una mejora continua del desarrollo. Se analizaron más de 1,4 millones de reseñas de Madrid utilizando primero modelo de TF-IDF y posteriormente el all-MiniLM-L6-v2, que fue finalmente el seleccionado. Para obtener un modelo óptimo se aplicó la reducción de dimensionalidad y visualización en 2D con UMAP y la segmentación mediante KMeans, identificando seis perfiles distintos de usuario.

El MVP fue implementado en arquitectura escalable sobre AWS, permitiendo una futura expansión. Los resultados obtenidos evidencian un impacto positivo en KPIs clave: incremento del ROI en 1,52 anual, aumento en la tasa de reservas con un incremento de ingresos el primer año de 252.306€ y mejora en la satisfacción del cliente. Además, se ha usado Figma para crear una representación de la interfaz de usuario mejorada con la solución de inteligencia artificial

Se concluye que la personalización basada en perfiles derivados del análisis de reseñas mejora significativamente la experiencia del usuario, huésped y plataforma.

Índice de contenidos

| | |
|---|----|
| Resumen | 2 |
| Índice de contenidos | 3 |
| 1. Introducción | 4 |
| 1.1 Resumen ejecutivo del trabajo | 4 |
| 1.2 Metodología del trabajo | 4 |
| 1.3 Recursos y medios empleados | 4 |
| 2. Análisis de contexto | 5 |
| 2.1 Análisis interno (empresa) | 5 |
| 2.2 Análisis externo (sector) | 6 |
| 2.3 Tendencias del sector | 7 |
| 3. Solución propuesta | 8 |
| 3.1 Propósito de la solución | 8 |
| 3.2 Objetivos de la solución | 8 |
| 3.3 Descripción de la solución | 9 |
| 3.4 Tecnología utilizada | 12 |
| 4. Impacto en el negocio | 13 |
| 4.1 Modelo de negocio y flujos de monetización | 13 |
| 4.2 Análisis financiero | 13 |
| 5. Conclusiones | 14 |
| 5.1 Contribuciones realizadas | 14 |
| 5.2 Problemas encontrados | 14 |
| 5.3 Opiniones personales | 14 |
| 6. Apéndice | 15 |
| Anexo I: Diagrama de la Arquitectura de datos | 15 |
| Anexo II: Desglose Financiero Estimado del Proyecto SAA (2025–2027) | 15 |
| 7. Bibliografía | 16 |

1. Introducción

1.1 Resumen ejecutivo del trabajo

Este proyecto propone un modelo de clusterización de reseñas en Airbnb para identificar perfiles de usuario y mejorar la experiencia de anfitriones, huéspedes y la plataforma. Se analizan más de un millón de reseñas de la ciudad de Madrid procesadas mediante técnicas de NLP y el modelo de all-MiniLM-L6-v2 combinados con UMAP y KMeans. El análisis reveló seis perfiles distintos de usuarios, y se desarrolló un MVP con arquitectura en AWS. La implementación del sistema mostró un impacto positivo en los indicadores clave como el ROI, tasa de reservas y satisfacción del cliente. Se concluye que la estrategia de personalización basada en perfiles mejora significativamente la experiencia del usuario y optimiza los recursos de la plataforma.

1.2 Metodología del trabajo

Para el desarrollo del proyecto se utilizó una metodología ágil basada en Scrum, organizada en sprints de dos semanas, lo que permitió trabajar de forma iterativa, colaborativa y adaptable. Cada sprint tenía objetivos concretos y entregables parciales, facilitando la validación continua y la mejora progresiva. Se utilizó Trello como herramienta de gestión visual para organizar las tareas en columnas tipo Backlog, To Do, In Progress y Done, lo que permitió una clara visibilidad del avance y la asignación de responsabilidades. Además, se realizó una reunión semanal los domingos de seguimiento antes de cada sprint para revisar el progreso, resolver bloqueos y planificar mejoras, asegurando así una evolución constante del proyecto.

1.3 Recursos y medios empleados

Se utilizó el lenguaje de programación Python para desarrollar la solución basada en inteligencia artificial, aplicando técnicas de clustering no supervisado como UMAP y KMeans, junto con modelos de procesamiento de lenguaje natural como all-MiniLM-L6-v2 y TF-IDF.

Además, se generaron mapas de calor para analizar la correlación entre el precio mediano de los alojamientos y las valoraciones promedio de cada clúster de huéspedes.

Para la fase de diseño de experiencia de usuario, se empleó Figma en la creación del prototipo de interfaz del anfitrión, donde se visualizan:

- Las solicitudes de reserva con perfiles de los huéspedes.
- Su historial y preferencias, extraídas automáticamente de sus reseñas.

- Una función de chatbot integrado, capaz de enviar mensajes personalizados en función del perfil y las expectativas del huésped, optimizando la comunicación desde el primer contacto.

2. Análisis de contexto

2.1 Análisis interno (empresa)

Airbnb, fundada en 2008 en San Francisco, aplicando un modelo basado en la economía colaborativa que transformó la industria del alojamiento al popularizar el concepto de alquileres vacacionales entre particulares, consiguiendo su expansión a nivel global. Su plataforma digital que conecta anfitriones con viajeros, permite a los usuarios alquilar alojamientos de forma flexible, eliminando la intermediación tradicional hotelera.

Airbnb cuenta con más de:

- 5 millones de anfitriones
- 2.000 millones de llegadas
- 220 países con anuncios
- 8 millones de anuncios activos

Airbnb opera actualmente en cuatro principales segmentos de alojamiento: **alquileres a corto plazo, viajes de negocios (Airbnb for Work), alquileres de lujo (Airbnb Luxe) y alquiler de experiencias (Airbnb Experiences)**. Esta diversificación permite a la plataforma adaptarse a distintos perfiles de usuarios y posicionarse competitivamente en el sector del turismo global, abarcando desde opciones económicas hasta experiencias premium. Sin embargo, destaca especialmente en el segmento de gama media, donde presenta una gran heterogeneidad tanto en precios como en calidad del servicio. Esta variabilidad permite adaptar la oferta a las preferencias, necesidades y presupuesto de cada usuario, lo que refuerza su atractivo frente a opciones más estandarizadas del mercado hotelero tradicional.

Para el análisis interno, se ha llevado a cabo un estudio detallado empleando dos herramientas estratégicas: **la Matriz de las Cinco Fuerzas de Porter y el Análisis DAFO.**

Dentro de la Matriz de Porter, se identificaron los siguientes factores clave:

- Poder de **negociación con los proveedores es bajo**: Con una amplia oferta y disponibilidad de anfitriones, dependencia de la plataforma para acceder al mercado y facilidad en el reemplazo y sustitución de anfitriones

- Poder de **negociación de compradores es alto**: Diversidad en la oferta de alojamientos disponibles, coste mínimo de migración entre plataformas de hospedaje, impacto de las reseñas y valoraciones en el proceso de decisión de compra y alta sensibilidad al precio
- **Entrada de nuevos competidores es moderado**: Bajas barreras de entrada al mercado, pero alta relevancia en la confianza y reputación del consumidor.
- **Amenaza de sustitutivos es alta**: Existen numerosas alternativas de alojamiento, desde hoteles hasta otras plataformas de economía colaborativa y alojamientos tradicionales.
- **Rivalidad entre competidores es muy alta**: Constante innovación y creación de valor añadido, diferenciación de servicios y experiencias y estrategia de precios agresiva.

Airbnb se posiciona como una plataforma con gran variedad de alojamientos, destacando los de gama media, en los cuales varían tanto su precio como su calidad en función de las preferencias del cliente

En cuanto al análisis **DAFO** se destaca:

- **Debilidades**: falta de control sobre los estándares de calidad del alojamiento, se servicio y experiencia del huésped y de seguridad y protección
- **Amenazas**: Regulaciones y restricciones gubernamentales que emanan del Ministerio de Interior. (modificación Real Decreto 933/2021) y alta competitividad en el sector de alquileres vacacionales.
- **Fortalezas**: Modelo de negocio basado en la economía colaborativa digital, tecnología avanzada para optimización de precios y personalización de ofertas, diversificación y adaptabilidad de servicios en mercados globales.
- **Oportunidades**: Aumento de digitalización en la búsqueda de alojamientos, nómadas digitales y flexibilidad laboral y tendencia creciente en el uso de reseñas como criterio clave en la selección de alquileres vacacionales¹

2.2 Análisis externo (sector)

El análisis PESTEL (Político, Económico, Social, Tecnológico, Ecológico y Legal) permite examinar los factores del macroentorno que influyen en la estrategia y el desempeño de Airbnb.

- **Político (P)**: Las regulaciones gubernamentales tienen un impacto directo sobre Airbnb, un ejemplo claro es el Real Decreto 933/2021 en España, que obliga a recopilar y reportar los datos personales de los huéspedes, lo que genera retos operativos y de cumplimiento normativo. Además, la

¹ Salesforce. "La importancia de las valoraciones: cómo influyen en la decisión del cliente", Disponible [Internet]: <https://www.salesforce.com/es/blog/importancia-valoraciones/> [28 de junio de 2025]

legislación sobre turismo, fiscalidad local y competencia puede variar significativamente entre países y ciudades, afectando la disponibilidad y rentabilidad de los alojamientos.²

- **Económico (E):** Factores como la inflación, el tipo de cambio y el ciclo económico influyen en la demanda de viajes y en la disposición del consumidor a gastar en alojamiento
- **Social (S):** Cambios en los estilos de vida, como el auge del teletrabajo y los nómadas digitales, han impulsado la demanda de estancias de media y larga duración. También existe una creciente preferencia por experiencias locales y auténticas, lo que encaja perfectamente con el modelo de Airbnb. Las valoraciones sociales y reputación online (reseñas) son fundamentales en las decisiones de reserva.
- **Tecnológico (T):** Grandes avances tecnológicos en el sector como el uso de big data, inteligencia artificial, NLP, y más recientemente, realidad virtual para ofrecer recorridos virtuales. Además, el desarrollo de infraestructuras como el 5G mejora la experiencia de navegación y conectividad de los usuarios.
- **Ecológico (E):** Se ha comenzado a responder a las preocupaciones medioambientales promoviendo alojamientos sostenibles, prácticas de turismo responsable y reduciendo la huella de carbono.
- **Legal (L):** Además del marco regulatorio turístico, los alojamientos de alquiler deben cumplir con normativas de protección de datos (GDPR en Europa, por ejemplo), condiciones contractuales claras, políticas de cancelación y transparencia fiscal. El entorno legal varía por región y puede cambiar rápidamente, afectando tanto a anfitriones como a las plataformas.

2.3 Tendencias del sector

La adopción de **inteligencia artificial (IA)** está marcando una transformación clave en la industria del alojamiento turístico, la integración de tecnologías inteligentes se ha vuelto esencial para mejorar la experiencia del usuario, optimizar la operación y aumentar la personalización del servicio.

- **Realidad Virtual (RV) y Realidad Aumentada (RA):** Los usuarios pueden hacer recorridos virtuales de los alojamientos de forma interactiva antes de reservar, con una toma de decisiones más precisa
- **Internet of Things (IoT):** promueve un entorno seguro a través del uso de cámaras inteligentes, cerraduras automáticas y monitoreo en tiempo real, accesibles para huéspedes y anfitriones.

² El País. “Consumo ordena a Airbnb el bloqueo de más de 65.000 anuncios ilegales de pisos turísticos”, Disponible [Internet]: <https://elpais.com/economia/2023-08-07/consumo-ordena-a-airbnb-el-bloqueo-de-mas-de-65000-anuncios-ilegales-de-pisos-turisticos.html> [28 de junio de 2025]

- Big Data y Análisis predictivo: Permite realizar ajustes en función de factores como la demanda o la estacionalidad maximizando los ingresos y la ocupación de los alojamientos.
- Chatbots y asistentes virtuales: Ofrecen respuestas rápidas, gestionan reservas, proporcionan información del alojamiento, etc. sin necesidad de intervención humana.

3. Solución propuesta

3.1 Propósito de la solución

La solución de analítica avanzada (SAA) que se propone en este estudio es transformar el ecosistema Airbnb mejorando la experiencia de los tres stakeholders: usuario (anfitrión), huésped y plataforma, aplicando un modelo de Machine Learning. En un entorno donde el 97% de los usuarios consulta reseñas antes de reservar (*Salesforce, 2023*), la SAA permite aprovechar ese comportamiento para crear agrupaciones de perfiles huésped basados en reseñas. Estos perfiles son la base para personalizar la experiencia, optimizar las recomendaciones y ajustar las expectativas de ambos lados de la transacción.

La solución transforma así un proceso estático y genérico en una experiencia personalizada y predictiva, generando confianza, fidelización y valor para todos los actores involucrados.

Con esta solución, Airbnb no solo mejora un proceso: redefine su propuesta de valor. La SAA transforma datos en conocimiento, conocimiento en acción, y acción en vínculos duraderos. Este enfoque transforma la plataforma en un ecosistema inteligente, inclusivo y centrado en el usuario, y posiciona a Airbnb como líder no solo en oferta, sino en experiencia.

3.2 Objetivos de la solución

El impacto no es unilateral. El anfitrión se convierte en un agente activo con información valiosa que le permite adaptar su comunicación, sus servicios y su estilo de hospitalidad a distintos tipos de huéspedes identificados mediante modelos de clustering. El resultado es un círculo virtuoso de calidad, confianza y reputación que impacta directamente en la tasa de reservas y la valoración promedio de los alojamientos. Los resultados esperados tras dos años muestran un ROI de 1.5 y un incremento de ingresos estimado en 3.6 millones de euros (incluyendo Madrid, Barcelona y top 20 ciudades). Este impacto económico sostenido demuestra que el propósito de la solución va más allá de lo técnico o funcional: es una apuesta estratégica por el valor sostenible, basado en datos y centrado en las personas.

La solución SAA persigue un impacto medible y estratégico en la operativa de Airbnb. Para ello, se definen los siguientes objetivos clave, cada uno acompañado de indicadores (KPIs) que permiten evaluar su cumplimiento de forma objetiva y continua.

1. Incremento promedio de los ingresos por anfitrión(€/año)
2. Tasa de conversión de visitas en reservas (%)
3. Aumento la ocupación media de los alojamientos (%)
4. Reducción el volumen de reseñas negativas (%)
5. Aumento de la satisfacción de los anfitriones con la herramienta (NPS)
6. Optimizar la segmentación de perfiles de huéspedes con
 - Silhouette Score (>0.5): Mide la coherencia y separación entre los clusters generados.
 - Separación de clusters (DBSCAN): Valida que los grupos identificados tienen sentido desde el punto de vista del negocio.
7. Reducción del tiempo medio de navegación del huésped (minutos)
8. Anfitriones activos con SAA (%)
9. ROI

3.3 Descripción de la solución

La solución propuesta consiste en el desarrollo e implementación de Inteligencia Artificial basada en técnicas de Procesamiento de Lenguaje Natural (NLP), cuyo objetivo principal es extraer valor de las reseñas de usuarios en Airbnb para generar perfiles de huéspedes que permitan una personalización de la experiencia en la plataforma.

El proceso de implementación de la solución analítica se compone de las siguientes etapas:

1. Recolección y procesamiento de datos:

Se parte de un gran volumen de datos públicos importados en dos bases de datos: reseñas y listados extraídos con webscrapping y disponibles en la web de InsideAirbnb³. Se decidió realizar el análisis sobre Madrid ya que dentro de España es de las ciudades con mayor volumen de datos y traducir las reseñas al inglés es más sencillo que ciudades con más idiomas como Barcelona.

Se continúa con una limpieza o pre procesamiento de ambas bases de datos.

- La base de Listados contiene en un inicio alrededor de 24.000 filas y 23 columnas: se eliminan las filas que contienen más de ocho celdas con valores nulos, y en caso de ser inferior; sustituimos los nulos

³ Inside Airbnb. "Get the data", Plataforma abierta de datos de alojamientos, Disponible [Internet]: <http://insideairbnb.com/get-the-data/> [28 de junio de 2025]

por la mediana en campos numéricos y aplicamos la moda sobre los campos categóricos. Finalmente nos cercioramos que la eliminación de nulos no supera el 10% de los registros

- La base de reseñas contiene en un inicio más de 1 millón de filas y 6 columnas: se eliminan los registros vacíos o nulos, los comentarios con menos de 50 caracteres y los comentarios con menos de 2 frases. Para la aplicación del modelo de clustering que realizamos a esta base de datos se ha realizado un paso adicional de traducción de reseñas al inglés y se filtra por una muestra aleatoria con el 1% de los datos (alrededor de 6.000 registros) para acortar el tiempo de procesamiento del modelo

2. Generación de perfiles mediante clustering:

Se han probado con dos modelos de NLP: TF-IDF y all-MiniLM-L6-v2

Primero se probó con **TF-IDF** que es un modelo más clásico que se basa en contar la frecuencia de palabras en cada comentario, sin entender el contexto. Los pasos a seguir son:

1. Preprocesamiento de texto

- Tokeniza: divide el texto en palabras.
- Lematiza o aplica stemming: reduce palabras a su raíz.
- Elimina palabras que no aportan significado ("el", "la", "de")

2. Cálculo de TF-IDF

- TF (frecuencia): cuántas veces aparece una palabra.
- IDF (inversa): mide cuán exclusiva es esa palabra.
- $TF \times IDF$: genera un peso para cada término.

3. Vectorización y clustering

- Cada documento (reseña) se transforma en un vector.
- Se agrupan con KMeans.
- Silhouette Score aceptable si es > 0.2 (el nuestro fue inferior).

Finalmente se descartó este modelo ya que; **no entiende el significado** del texto, solo frecuencia y **peor** desempeño en **Silhouette Score** y menor coherencia entre grupos.

Posteriormente se probó con el modelo **all-MiniLM-L6-v2 + UMAP + KMeans**, que fue el elegido y los pasos que se siguieron fueron:

1. Conversión del texto en vectores

No es necesario un pre-procesamiento ya que el propio modelo tokeniza, limpia el texto y unifica el formato, convierte cada reseña en un vector numérico que refleja **su significado completo**, no solo las palabras y se pueden comparar reseñas midiendo la **similitud de coseno** entre sus vectores: a menor distancia, más parecidas.

2. Reducción de dimensionalidad con UMAP

Ayuda a visualizar los datos y conservar la información relevante y se aplicaron los parámetros:

- `n_neighbors=40`: número de vecinos para preservar la estructura local.
- `min_dist=0.0`: permite que los puntos similares se agrupen más compactamente.
- `n_components=32`: reduce la dimensión para hacer cálculos más eficientes.

3. Agrupamiento con KMeans

Crea clústeres o grupos de huéspedes con comportamientos y preferencias similares.

4. Evaluación del modelo

Usamos **Silhouette Score** para **medir calidad** de los grupos (ideal > 0.5 , en nuestro caso es aceptable con ≈ 0.36).

Se decantó por la elección de este modelo ya que:

- Captura el significado del lenguaje, más allá de palabras específicas.
- Agrupa perfiles, aunque los usuarios se expresen de manera diferente.
- Es eficiente y ligero (versión reducida de BERT).
- Ideal para detectar patrones emocionales y contextuales en las reseñas.
- Funciona muy bien con UMAP y KMeans para segmentar perfiles reales.

3. Identificación de seis tipos principales de huéspedes:

A partir del análisis del contenido de cada grupo de reseñas, se describieron los siguientes perfiles:

Cluster 1: Viajeros afectuosos: Valoran hospitalidad cálida, detalles personalizados y atención cercana del anfitrión.

Cluster 2: Viajeros prácticos: Buscan funcionalidad, limpieza y buena ubicación, especialmente en zonas céntricas.

Cluster 3: Turistas exploradores y comunicativos: Priorizan la ubicación estratégica y valoran respuestas rápidas ante inconvenientes.

Cluster 4: Huéspedes empáticos: Aprecian el trato humano, la comprensión del anfitrión y una experiencia acogedora, incluso ante imprevistos.

Cluster 5: Viajeros exigentes: Enfocados en la eficiencia, la tranquilidad y la claridad en normas y condiciones.

Cluster 6: Huéspedes cálidos y familiares: Buscan comodidad, buena conexión y una atención amable que los haga sentir como en casa

4. Análisis de correlación lineal entre el precio de los alojamientos y las valoraciones

Posteriormente al clustering se realizó un análisis para entender si el precio influía en la valoración de los alojamientos según el barrio de Madrid, esto se ha representado con un mapa de calor

El ejemplo más claro de relación directa entre precio y valoraciones es el clúster 3. Este perfil valora especialmente la **comunicación rápida y clara** del anfitrión, así como la **precisión en la información** y la **facilidad de check-in**.

Existe una **fuerte relación entre precio y valoración**: el **81 % de la valoración en comunicación se explica por el precio**, lo que indica que a mayor coste, mayores expectativas. Si no se cumple con este nivel de atención, las valoraciones generales pueden verse afectadas. Se concentra en barrios como Trafalgar, Imperial, Legazpi, Lista y Prosperidad.

La estrategia recomendada al anfitrión para mejorar su servicio es: comunicación fluida y transparente, información precisa y recomendaciones turísticas personalizadas

3.4 Tecnología utilizada

Para llevar a cabo la solución analítica propuesta, se ha utilizado un conjunto de tecnologías y herramientas que permiten procesar lenguaje natural, segmentar perfiles de usuario y visualizar resultados de forma efectiva, todo este análisis se ha realizado en Python.

- **TF-IDF:** Modelo NLP permite identificar las palabras más relevantes dentro de una reseña, en función de su frecuencia en el documento y en el conjunto total de textos. Fue útil para una primera aproximación a los patrones de lenguaje
- **All-MiniLM-L6-v2:** Este modelo de **embeddings semánticos** convierte las reseñas en vectores numéricos que reflejan el significado del texto, más allá de las palabras concretas. Es una versión ligera y rápida de modelos como BERT o RoBERTa, ideal para tareas de clustering y análisis de sentimientos.
- **UMAP:** Técnica para **reducir la dimensionalidad de los vectores** generados por MiniLM, lo que permite **visualizar** la estructura interna de los datos en un espacio de 2D o 3D.

Esta técnica conserva la relación de cercanía entre puntos, lo que facilita detectar agrupaciones naturales de huéspedes según sus opiniones.

- **KMeans:** Algoritmo de clustering no supervisado que **agrupa puntos de datos en K clústeres basándose en su cercanía** (normalmente usando distancia euclidiana). El usuario debe definir previamente el número de clústeres (K).
- **Método del codo:** técnica para ayudar a elegir el **valor óptimo de K** para aplicar en KMeans
- **Arquitectura en AWS⁴:** creación de una arquitectura óptima que cumpliera con los requisitos necesarios de: escalabilidad, eficiencia, económica, segura, fácil de conectar, confiable, para todo el equipo y que apoya los objetivos propuestos⁵

4. Impacto en el negocio

4.1 Modelo de negocio y flujos de monetización

El proyecto se articula propone una solución tecnológica basada en Machine Learning y Procesamiento de Lenguaje Natural, capaz de analizar reseñas de usuarios para generar **perfiles de huéspedes** y personalizar la experiencia de reserva. Este modelo aporta valor a tres stakeholders principales: Airbnb, anfitriones y huéspedes

Airbnb monetiza el impacto del proyecto a través del incremento en los ingresos generados por los anfitriones, reducción de costes y el impacto en la inversión

En un periodo de 2 años:

- Mejora ADR (Average Daily Rate) del 3%, pasando de 87,4 € a 90 €.
- Incremento anual por anfitrión de 786 € con una adopción parcial (30%) en Madrid: 3.210 anfitriones que se traduce en 2.523.060 €/año
- Comisión Airbnb (10%): 252.306 €/año⁶
- Reducción de reseñas negativas con un ahorro anual estimado de 126.150 € en costes operativos.

4.2 Análisis financiero

Se estima un ROI por el proyecto piloto en Madrid de 1,52. $(252.306 - 100.000) / 100.000 = 1,52x$ (beneficio neto: 152.306 €)

Con un escalado a 2 años⁷:

⁴ Amazon Web Services. "AWS Pricing Calculator", Disponible [Internet]: <https://aws.amazon.com/es/pricing/> [28 de junio de 2025]

⁵ Diagrama de la arquitectura en el anexo I

⁶ Airbnb. "Página oficial de Airbnb España", Disponible [Internet]: <https://www.airbnb.es/> [28 de junio de 2025]

⁷ Desglose de los gastos en el anexo II

- Ingresos proyectados: 3.600.000 €
- Comisión estimada para Airbnb: 360.000 €
- Reducción de costes agregada: 126.150 €/año

5. Conclusiones

5.1 Contribuciones realizadas

En el proyecto he contribuido en mayor medida en la creación del modelo de clustering y en su posterior análisis, así como en la infraestructura de la arquitectura y los costes asociados a esta.

5.2 Problemas encontrados

Se han identificado problemas en:

- Gestión de datos nulos o ceros
- Diversidad lingüística en reseñas
- Ajuste de modelos para encontrar una precisión óptima con un modelo funcional para el caso de uso
- Implementar una idea novedosa que no se esté realizando aún y con potencial
- Organización de tareas y reuniones de equipo regulares por compatibilidad de horarios
- El equipo se ha visto reducido de 6 a 4 personas con lo cual ha tocado más trabajo por persona del esperado

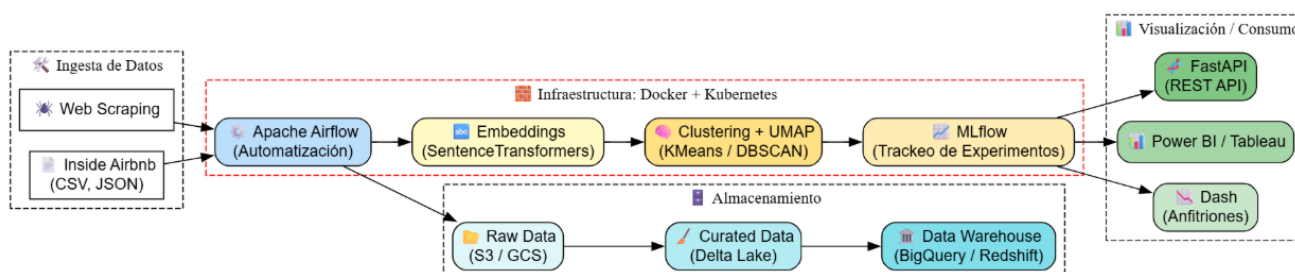
5.3 Opiniones personales

Este proyecto troncal ha sido una oportunidad clave para actualizarme en el uso de herramientas de inteligencia artificial y profundizar en el valor estratégico de la arquitectura del dato. Me ha permitido comprender cómo diseñar e implementar una solución tecnológica de principio a fin dentro de una empresa internacional y líder en su sector, integrando tanto la perspectiva técnica como la de negocio.

Además, he fortalecido mis habilidades de trabajo en equipo, contribuyendo activamente desde mi perfil técnico y colaborando estrechamente con mis compañeros para superar los retos del desarrollo. He intentado aportar al máximo mis conocimientos para apoyar en la parte técnica del proyecto, al tiempo que aprendía de otras áreas clave como la analítica de negocio, UX y estrategia de implantación. Este proyecto troncal me ha servido para actualizarme sobre las herramientas de IA, aprender sobre la arquitectura del dato y la importancia de esta y como implementar de inicio a fin una idea de tecnológica en una empresa internacional y líder en el mercado.

6. Apéndice

Anexo I: Diagrama de la Arquitectura de datos



Anexo II: Desglose Financiero Estimado del Proyecto SAA (2025–2027)

Desglose del análisis financiero estimado en 2años:

Costes de Inversión (CAPEX)

| Fase | Concepto | Monto (€) | Año |
|-------------------------|-------------------------------|------------------|-----|
| Fase 1: Madrid | Infraestructura | 15.000 | 1 |
| | Machine Learning | 20.000 | 1 |
| | UX/UI Producto | 10.000 | 1 |
| | Recursos Humanos | 25.000 | 1 |
| | Desarrollo | 30.000 | 1 |
| | Subtotal Madrid | 100.000 | |
| Fase 2: BCN + Mejora | Optimización del modelo | 40.000 | 1 |
| | Formación a anfitriones | 20.000 | 1 |
| | Nuevas funcionalidades UX | 30.000 | 1 |
| | Infraestructura adicional | 20.000 | 1 |
| | Subtotal Barcelona | 110.000 | |
| Fase 3: Escalado Global | Escalado a 20 ciudades | 60.000 | 2 |
| | Formación internacional | 30.000 | 2 |
| | Mantenimiento y soporte | 20.000 | 2 |
| | Data training adicional | 30.000 | 2 |
| | Subtotal Global | 140.000 | |
| Consultoría Total | Servicios externos (2 años) | 90.000 | 1–2 |
| | TOTAL INVERSIÓN 2 AÑOS | 440.000 € | |

Ingresos Proyectados por Comisiones (10% de aumento de ingresos a anfitriones)

| Región / Escenario | Año 1 (€) | Año 2 (€) | Total 2 años (€) |
|-------------------------------|-----------|-----------|--------------------|
| Madrid (30% adopción) | 252.306 | 252.306 | 504.612 |
| Barcelona + mejora Madrid | — | 296.000 | 296.000 |
| 20 ciudades globales | — | 3.000.000 | 3.000.000 |
| Total ingresos por comisiones | 252.306 | 3.548.306 | 3.800.612 € |

Reducción de Costes Operativos (reseñas negativas evitadas)

| Categoría | Año 1 (€) | Año 2 (€) | Total 2 años (€) |
|-----------------------------|----------------|------------------------|------------------|
| Madrid (reseñas evitadas) | 126.150 | 126.150 | 252.300 |
| Escalado (estimado) | – | Incluido en mejora ROI | – |
| Total ahorro directo | 126.150 | 126.150 | 252.300 € |

7. Bibliografía

[1] Salesforce. “La importancia de las valoraciones: cómo influyen en la decisión del cliente”, Disponible [Internet]:

<https://www.salesforce.com/es/blog/importancia-valoraciones/> [28 de junio de 2025]

[2] Inside Airbnb. “Get the data”, Plataforma abierta de datos de alojamientos, Disponible [Internet]:

<http://insideairbnb.com/get-the-data/> [28 de junio de 2025]

[3] El País. “Consumo ordena a Airbnb el bloqueo de más de 65.000 anuncios ilegales de pisos turísticos”, Economía, Disponible [Internet]:

<https://elpais.com/economia/2023-08-07/consumo-ordena-a-airbnb-el-bloqueo-de-mas-de-65000-anuncios-ilegales-de-pisos-turisticos.html> [28 de junio de 2025]

[4] Airbnb. “Página oficial de Airbnb España”, Disponible [Internet]:

<https://www.airbnb.es/> [28 de junio de 2025]

[5] Amazon Web Services. “AWS Pricing Calculator”, Disponible [Internet]:

<https://aws.amazon.com/es/pricing/> [28 de junio de 2025]