

PROYECTO EDA

Análisis del alojamiento turístico en Madrid



Patricia García

Javier López

María del Pilar Gutiérrez

Índice

1. Introducción

2. Objetivos del proyecto

3. Fuente de datos

4. Metodología y herramientas

5. Exploración, limpieza y preparación del Dataset

6. Análisis Exploratorio de Datos

 1.2 Análisis univariante

 2.2 Análisis bivariante

7. Validación de hipótesis

8. Conclusiones

9. Limitaciones del estudio

10. Líneas futuras

11. Anexos

12. Cierre del Análisis

1. Introducción

En los últimos años, el auge de las plataformas digitales de alquiler turístico ha transformado de manera profunda el mercado inmobiliario y el sector turístico en las grandes ciudades. Plataformas como Airbnb han facilitado el acceso a alojamientos de corta duración, modificando los patrones tradicionales de oferta y demanda y generando nuevos retos económicos, sociales y urbanísticos. Este fenómeno ha suscitado un creciente interés tanto por parte de las administraciones públicas como de investigadores y analistas de datos, debido a su impacto en los precios de la vivienda, la distribución del turismo y la convivencia en los entornos urbanos.

Madrid, como uno de los principales destinos turísticos de Europa, no es ajena a esta transformación. La ciudad ha experimentado un notable incremento en la oferta de alojamientos turísticos, especialmente en determinadas zonas céntricas, donde la presión turística es más intensa. Esta expansión ha dado lugar a un mercado heterogéneo, caracterizado por una amplia variedad de precios, tipos de alojamiento y niveles de uso, lo que hace necesario un análisis detallado que permita comprender su estructura y funcionamiento.

En este contexto, el presente trabajo desarrolla un Análisis Exploratorio de Datos (EDA) con el objetivo de estudiar la estructura, distribución y comportamiento del mercado de alojamientos turísticos en Madrid.

A través de este EDA se pretende identificar patrones relevantes, detectar desigualdades espaciales y analizar los principales factores que influyen en el precio y el uso de los alojamientos turísticos. Para ello, se aplican técnicas de análisis univariante y bivariante apoyadas en visualizaciones gráficas, que permiten extraer conclusiones de forma clara y fundamentada. Asimismo, el estudio busca sentar las bases para análisis más avanzados, como la construcción de modelos predictivos o

el análisis temporal de la demanda turística, aportando una visión integral del mercado de alojamientos turísticos en la ciudad de Madrid.

2. Objetivos del análisis

El objetivo general de este análisis es comprender en profundidad la estructura y el comportamiento del mercado de alojamientos turísticos en la ciudad de Madrid mediante la aplicación de técnicas de Análisis Exploratorio de Datos (EDA). A partir de datos abiertos procedentes de la plataforma Inside Airbnb, se busca identificar patrones, relaciones y características relevantes que permitan explicar las diferencias existentes entre los distintos alojamientos turísticos.

De manera específica, los objetivos del análisis son los siguientes:

- **Analizar la distribución de precios de los alojamientos turísticos**, con el fin de comprender su comportamiento general, identificar posibles asimetrías y evaluar la presencia de valores extremos que puedan afectar a la interpretación del mercado. Este objetivo permite determinar la medida de tendencia central más adecuada para describir el precio y justificar el uso de métricas robustas como la mediana.
- **Estudiar la relación entre el precio y el tipo de alojamiento**, diferenciando entre pisos completos, habitaciones privadas y habitaciones compartidas, con el propósito de evaluar hasta qué punto esta variable contribuye a explicar las diferencias de precio observadas en el mercado turístico de Madrid.
- **Identificar los barrios y distritos con mayor concentración de alojamientos turísticos**, analizando la distribución espacial de la oferta. Este objetivo permite detectar zonas con alta presión turística, así como áreas con menor densidad de alojamientos, que podrían considerarse zonas potencialmente desaprovechadas.

- **Evaluar la disponibilidad anual de los alojamientos como un indicador indirecto del nivel de ocupación**, entendiendo que una menor disponibilidad puede asociarse a una mayor demanda. Este análisis permite aproximarse al comportamiento del uso de los alojamientos en ausencia de datos directos de ocupación o ingresos.
- **Detectar patrones relevantes y posibles valores atípicos** en variables clave como el precio, la disponibilidad o el número de reseñas, con el fin de comprender mejor la heterogeneidad del mercado y tomar decisiones metodológicas adecuadas para los análisis posteriores.

En conjunto, estos objetivos pretenden ofrecer una visión global y fundamentada del mercado de alojamientos turísticos en Madrid, sentando las bases para análisis más avanzados y proporcionando información útil para la toma de decisiones estratégicas en el ámbito del turismo y la planificación urbana.

3. Fuente de datos

Los datos utilizados en el presente análisis proceden de **Inside Airbnb**, una iniciativa independiente que recopila y publica datos abiertos con el objetivo de facilitar el estudio del impacto social, económico y urbano de la plataforma Airbnb en diferentes ciudades del mundo. Este proyecto proporciona información detallada y accesible que permite analizar de forma transparente el funcionamiento del mercado de alojamientos turísticos, siendo una fuente ampliamente utilizada en estudios académicos y trabajos de análisis de datos.

El dataset correspondiente a la ciudad de Madrid recoge información a nivel de alojamiento individual, lo que permite realizar un análisis granular del mercado. Entre las principales variables incluidas se encuentran el **precio del alojamiento**, el **tipo de habitación** (piso completo, habitación privada o compartida), la **localización geográfica** mediante coordenadas de latitud y longitud, el **número de reseñas** y la **disponibilidad anual**, entre otras características relevantes.

Estas variables permiten abordar diferentes dimensiones del mercado turístico, como la estructura de precios, la distribución espacial de la oferta, los patrones de uso y la popularidad de los alojamientos. No obstante, es importante señalar que el dataset presenta algunas limitaciones inherentes a su naturaleza. En particular, no incluye información directa sobre ingresos reales ni sobre fechas concretas de

reserva, por lo que ciertas métricas, como el nivel de ocupación o la demanda, deben aproximarse mediante variables proxy, como la disponibilidad anual o el número de reseñas recientes.

A pesar de estas limitaciones, los datos de Inside Airbnb constituyen una fuente fiable y adecuada para la realización de un Análisis Exploratorio de Datos. Su nivel de detalle y cobertura permiten extraer conclusiones relevantes sobre el mercado de alojamientos turísticos en Madrid y sirven como base para futuros análisis más avanzados, como estudios temporales o modelos predictivos de precios.

4. Metodología y herramientas

El análisis exploratorio de datos se ha desarrollado utilizando el lenguaje de programación Python, ampliamente empleado en el ámbito del análisis de datos por su versatilidad y su amplio ecosistema de librerías especializadas. Para la manipulación, limpieza y transformación de los datos se han utilizado principalmente las librerías Pandas y NumPy, que permiten trabajar de forma eficiente con grandes volúmenes de información y realizar operaciones estadísticas básicas.

En cuanto a la visualización de los datos, se han empleado las librerías Matplotlib y Seaborn, que facilitan la creación de gráficos descriptivos y comparativos. Estas herramientas han sido fundamentales para identificar patrones, tendencias y posibles valores atípicos, así como para comunicar de manera clara los resultados obtenidos durante el análisis.

La metodología seguida se basa en las fases clásicas de un Análisis Exploratorio de Datos (EDA). En primer lugar, se realizó una exploración inicial del conjunto de datos para comprender su estructura y las variables disponibles. Posteriormente, se llevó a cabo un proceso de limpieza y preparación de los datos, que incluyó el tratamiento de valores nulos, la corrección de formatos y la gestión de valores extremos. A continuación, se desarrolló un análisis univariante para estudiar el comportamiento individual de las principales variables y un análisis bivariante para explorar las relaciones existentes entre ellas. Finalmente, se procedió a la extracción de conclusiones, sintetizando los hallazgos más relevantes y relacionándolos con los objetivos planteados en el estudio.

5.Exploración, limpieza y preparación del Dataset

Durante la fase de exploración inicial se realizó un análisis preliminar del conjunto de datos con el objetivo de comprender su estructura general, el volumen de registros disponibles y el tipo de variables incluidas. Este primer acercamiento permitió identificar posibles inconsistencias y establecer las bases para las etapas posteriores del análisis.

En esta fase se examinó el tipo de cada variable, diferenciando entre variables numéricas y variables categóricas. Entre las variables numéricas destacan el precio del alojamiento, el número de reseñas y la disponibilidad anual, mientras que entre las variables categóricas se identificaron el tipo de alojamiento y el barrio o distrito en el que se localiza cada inmueble. Asimismo, se analizó de forma general la distribución de estas variables, lo que permitió detectar asimetrías, concentraciones de valores y la posible presencia de valores extremos, especialmente en la variable precio.

Esta exploración inicial resultó clave para orientar las decisiones metodológicas adoptadas posteriormente, como la selección de métricas robustas y el diseño de las visualizaciones más adecuadas para el análisis.

La fase de limpieza y preparación de los datos tuvo como objetivo garantizar la calidad y coherencia del conjunto de datos antes de proceder al análisis exploratorio propiamente dicho. En primer lugar, se realizó un tratamiento de los valores nulos, prestando especial atención a su impacto en las variables relevantes para el estudio. En el caso de variables no críticas, los valores ausentes fueron gestionados de forma que no afectaran a los análisis posteriores.

La variable precio requirió un tratamiento específico, ya que inicialmente se encontraba almacenada como texto con símbolos monetarios. Para su correcta utilización, se eliminaron dichos símbolos y se convirtió la variable a formato numérico. Asimismo, se eliminaron aquellos registros con precios nulos o iguales a cero, al considerarse no representativos del mercado real de alojamientos turísticos.

Adicionalmente, se llevaron a cabo comprobaciones de consistencia y se revisaron posibles valores atípicos, especialmente en las variables relacionadas con el precio y la disponibilidad. Estas tareas permitieron obtener un dataset limpio y preparado

para el desarrollo de los análisis univariantes y bivariantes, asegurando la fiabilidad de los resultados obtenidos.

6. Análisis Exploratorio de Datos

El Análisis Exploratorio de Datos constituye el núcleo central de este trabajo, ya que permite comprender en profundidad el comportamiento de las variables incluidas en el dataset y detectar patrones relevantes en el mercado de alojamientos turísticos de Madrid. Esta fase del análisis se ha estructurado en dos bloques principales: un análisis univariante, orientado al estudio individual de cada variable, y un análisis bivariante, centrado en la exploración de las relaciones existentes entre las variables clave.

A través de estas dos aproximaciones se busca no solo describir el comportamiento de los datos, sino también identificar tendencias, desigualdades espaciales y posibles factores explicativos que ayuden a interpretar la dinámica del mercado turístico.

6.1 Análisis univariante

El análisis univariante permitió examinar la distribución individual de las principales variables del estudio, proporcionando una primera visión de la estructura del mercado. En el caso de la variable precio, se observó una distribución claramente asimétrica, caracterizada por una alta concentración de alojamientos en rangos de precios bajos y medios, junto con la presencia de valores extremos en los tramos más elevados. Este comportamiento justificó el uso de la mediana como medida de tendencia central más representativa frente a la media.

De forma similar, el número de reseñas mostró una distribución desigual, con un elevado número de alojamientos que cuentan con pocas o ninguna reseña, frente a un grupo reducido que concentra una parte significativa de la actividad. Este patrón refleja la heterogeneidad en el uso de los alojamientos turísticos y pone de manifiesto la existencia de distintos niveles de demanda.

Asimismo, el análisis de la distribución geográfica evidenció una alta concentración de alojamientos en determinados distritos, especialmente en zonas céntricas, lo que sugiere una oferta turística desigual desde el punto de vista espacial. Estas observaciones iniciales resultaron fundamentales para orientar los análisis posteriores y seleccionar las visualizaciones más adecuadas.

El análisis bivariante se centró en estudiar las relaciones existentes entre el precio y otras variables relevantes, con el objetivo de identificar posibles factores explicativos de las diferencias observadas. En primer lugar, se analizó la relación entre precio y ubicación, observándose que los alojamientos situados en distritos céntricos presentan precios significativamente más elevados que aquellos ubicados en zonas periféricas. Además, estas zonas muestran una mayor dispersión de precios, lo que indica un mercado más heterogéneo.

En segundo lugar, se examinó la relación entre precio y tipo de alojamiento, constatando que los pisos completos tienden a presentar precios más altos en comparación con las habitaciones privadas o compartidas. Este resultado pone de manifiesto la importancia del tipo de alojamiento como factor determinante en la fijación del precio.

De forma complementaria, se exploraron relaciones entre el precio y variables asociadas al uso del alojamiento, como el número de reseñas o la disponibilidad anual. Los resultados sugieren que un precio más elevado no implica necesariamente una mayor demanda, lo que apunta a la existencia de rangos de precio óptimos y a la influencia de otros factores adicionales en el éxito de los alojamientos turísticos.

7. Validación de la hipótesis

Los resultados obtenidos a lo largo del Análisis Exploratorio de Datos permiten confirmar parcialmente la hipótesis planteada al inicio del estudio. En particular, el análisis muestra que la ubicación del alojamiento y el tipo de alojamiento son los factores que presentan una mayor capacidad explicativa en relación con las diferencias de precio observadas en el mercado de alojamientos turísticos de Madrid.

El estudio de la variable ubicación pone de manifiesto que los alojamientos situados en zonas céntricas presentan precios significativamente más elevados que aquellos ubicados en distritos periféricos. Además, estas zonas muestran una mayor dispersión de precios, lo que indica la coexistencia de alojamientos de diferentes segmentos dentro de un mismo distrito. Este comportamiento confirma el papel central de la localización como determinante del precio.

De igual forma, el análisis por tipo de alojamiento evidencia que los pisos completos tienden a alcanzar precios superiores en comparación con las habitaciones privadas o compartidas. Este resultado refuerza la idea de que el tipo de alojamiento constituye un factor clave en la fijación del precio, independientemente de otros aspectos del anuncio.

Por otro lado, las características del anuncio, como la capacidad, el número de noches mínimas o la disponibilidad anual, muestran una influencia más moderada y actúan de forma complementaria a los factores principales. Si bien contribuyen a explicar parte de la variabilidad del precio, su impacto es menor en comparación con la ubicación y el tipo de alojamiento.

Finalmente, la estacionalidad no pudo analizarse de forma directa debido a las limitaciones del dataset, que no incluye información temporal detallada sobre fechas de reserva o ingresos. No obstante, se utilizaron variables proxy, como el número de reseñas recientes y la disponibilidad anual, para aproximarse de manera indirecta al comportamiento de la demanda. En conjunto, estos resultados permiten validar la

hipótesis de forma parcial y subrayan la importancia de considerar las limitaciones de los datos en la interpretación de los hallazgos.

8. Conclusiones

El análisis exploratorio realizado pone de manifiesto que el mercado de alojamientos turísticos en Madrid es altamente heterogéneo y presenta una fuerte concentración espacial en las zonas céntricas de la ciudad. Estas áreas concentran tanto una mayor densidad de alojamientos como los precios más elevados, reflejando una presión turística superior en comparación con los distritos periféricos.

Los resultados del estudio muestran que la ubicación y el tipo de alojamiento son los factores que ejercen una mayor influencia sobre el precio. Los alojamientos situados en distritos centrales y aquellos que corresponden a pisos completos tienden a presentar precios significativamente más altos, mientras que las habitaciones privadas y compartidas se sitúan en rangos de precios inferiores. Asimismo, se observa una elevada dispersión de precios en las zonas céntricas, lo que evidencia la coexistencia de diferentes segmentos de mercado.

En relación con el uso de los alojamientos, el análisis sugiere que un precio elevado no garantiza necesariamente una mayor demanda, ya que la actividad se concentra principalmente en rangos de precios intermedios. Este resultado pone de relieve la importancia de considerar un equilibrio entre precio y características del alojamiento para maximizar su atractivo.

En conjunto, este Análisis Exploratorio de Datos proporciona una visión clara y fundamentada de los principales patrones del mercado de alojamientos turísticos en Madrid. Los hallazgos obtenidos constituyen una base sólida para futuros estudios más avanzados, como el desarrollo de modelos predictivos de precios o el análisis temporal de la demanda, y pueden resultar de utilidad para la toma de decisiones en el ámbito del turismo y la planificación urbana.

9. Limitaciones del estudio

A pesar de que los datos utilizados permiten realizar un Análisis Exploratorio de Datos sólido y obtener conclusiones relevantes sobre el mercado de alojamientos turísticos en Madrid, el estudio presenta una serie de limitaciones que deben ser tenidas en cuenta a la hora de interpretar los resultados.

La principal limitación del análisis es la ausencia de información temporal detallada, como fechas concretas de reserva o evolución de los precios a lo largo del tiempo. Esta carencia impide analizar de manera directa la estacionalidad del mercado turístico y evaluar cómo varían la demanda y los precios en función de la época del año. Como consecuencia, no es posible identificar patrones temporales precisos ni comparar comportamientos entre distintas temporadas.

Asimismo, el dataset no incluye datos reales de ingresos ni de ocupación, lo que limita la capacidad de evaluar el rendimiento económico de los alojamientos. Para aproximarse a estos aspectos, se recurrió al uso de variables proxy, como la disponibilidad anual o el número de reseñas recientes, que permiten inferir de manera indirecta el nivel de demanda. No obstante, estas aproximaciones pueden introducir cierto grado de incertidumbre y no sustituyen completamente a los datos reales de ingresos.

Finalmente, es importante señalar que el análisis se basa en una fotografía puntual del mercado, correspondiente a un momento concreto en el tiempo. Esto implica que los resultados reflejan la situación del mercado en ese periodo específico y pueden variar en función de cambios normativos, económicos o en el comportamiento de los usuarios. Estas limitaciones no invalidan los resultados obtenidos, pero deben considerarse como un marco contextual para su correcta interpretación.

10. Líneas futuras

A partir de los resultados obtenidos y de las limitaciones identificadas en el presente estudio, se abren diversas líneas de trabajo futuro que permitirían profundizar en el análisis del mercado de alojamientos turísticos en Madrid y ampliar el alcance del estudio.

Una de las principales líneas futuras consiste en el desarrollo de modelos predictivos de precios, utilizando técnicas de aprendizaje automático o modelos estadísticos avanzados. Estos modelos permitirían estimar el precio óptimo de los alojamientos en función de variables como la ubicación, el tipo de alojamiento, las características del anuncio y los patrones de demanda observados, aportando un enfoque más aplicado y orientado a la toma de decisiones.

Asimismo, resulta de especial interés realizar un análisis temporal del uso de los alojamientos turísticos, incorporando información sobre fechas de reserva y evolución de la demanda a lo largo del tiempo. Este enfoque permitiría estudiar la estacionalidad del mercado, identificar períodos de mayor y menor actividad y analizar cómo varían los precios y la ocupación en función de la época del año.

Por último, una línea futura relevante sería la incorporación de nuevas fuentes de datos, como información socioeconómica por distrito, datos de movilidad turística o registros administrativos. La integración de estos datos permitiría enriquecer el análisis, mejorar la explicación de los patrones observados y ofrecer una visión más completa del impacto del alojamiento turístico en el entorno urbano.

11. Anexos

En los anexos se recoge información complementaria que amplía y respalda el análisis presentado en el cuerpo principal del documento. Este apartado tiene como objetivo proporcionar mayor transparencia metodológica y facilitar la reproducibilidad

del estudio, sin sobrecargar las secciones principales con detalles técnicos adicionales.

Concretamente, los anexos incluyen gráficos adicionales que permiten profundizar en determinados aspectos del análisis, así como tablas resumen con estadísticas descriptivas de las principales variables. Asimismo, se incorporan fragmentos de código y referencias a los notebooks utilizados durante las distintas fases del proceso, como la limpieza de datos, el análisis univariante y bivariante y la generación de visualizaciones.

Este material complementario resulta de utilidad para comprender en detalle las decisiones metodológicas adoptadas y para servir de base a posibles extensiones del análisis en trabajos futuros. Aunque los anexos no forman parte del núcleo principal del estudio, constituyen un soporte fundamental para garantizar la solidez y coherencia del Análisis Exploratorio de Datos realizado.

12. Cierre del Análisis Exploratorio de Datos

El Análisis Exploratorio de Datos desarrollado en este trabajo ha permitido obtener una visión global y estructurada del mercado de alojamientos turísticos en la ciudad de Madrid. A través del estudio de la distribución de precios, la localización geográfica, el tipo de alojamiento y los indicadores de uso, se han identificado patrones claros que ayudan a comprender la dinámica y heterogeneidad de este mercado.

Los resultados obtenidos ponen de manifiesto la importancia de la ubicación y del tipo de alojamiento como principales factores explicativos del precio, así como la existencia de una fuerte concentración de la oferta en determinadas zonas de la ciudad. Al mismo tiempo, el análisis evidencia que el éxito de los alojamientos turísticos no depende exclusivamente de un precio elevado, sino de un equilibrio entre precio, características del alojamiento y demanda.

Si bien el estudio presenta limitaciones derivadas de la naturaleza de los datos disponibles, el enfoque exploratorio adoptado ha resultado adecuado para detectar relaciones relevantes y formular hipótesis fundamentadas. Este EDA sienta así las bases para análisis más avanzados, como la modelización predictiva de precios o el estudio temporal de la demanda, y aporta una base sólida para la toma de decisiones en el ámbito del turismo urbano y la planificación territorial.

En definitiva, este trabajo demuestra el valor del Análisis Exploratorio de Datos como herramienta fundamental para comprender fenómenos complejos a partir de datos reales, ofreciendo una aproximación rigurosa y replicable al estudio del mercado de alojamientos turísticos en Madrid.

