



Tecnológico de Monterrey
Escuela de Negocios

Instituto Tecnológico y de Estudios Superiores de Monterrey

Escuela de Negocios

Campus Puebla

Plataformas de analítica de negocios

5. Extracción de características

Isaac Miguel Barrón Portillo

A01737199

Docente: PhD Alfredo García Suárez

Fecha: 5 de septiembre de 2025

AIRBNB MADRID, ESPAÑA – Extracción de características }

El presente reporte forma parte de una fase exploratoria del análisis de datos aplicado a un dataset de propiedades y anfitriones de Airbnb en Madrid, España. Esta etapa se centró en el estudio univariado de variables categóricas con el objetivo de comprender la composición, distribución y frecuencia de las principales características que definen tanto a los anfitriones como a los alojamientos ofertados.

Dicho análisis permite identificar patrones iniciales, outliers, homogeneidad o diversidad en las categorías y posibles puntos clave que pueden influir en etapas posteriores del modelado o toma de decisiones.

2. Documentación del Proceso

El trabajo siguió un **workflow estructurado** que incluyó:

- **Limpieza de datos** (tratamiento de outliers en variables cuantitativas previamente).
- **Segmentación de variables categóricas**, diferenciando entre aquellas que requerían recorte de categorías (top 10 más frecuentes) y aquellas que conservaron su distribución completa.
- **Construcción de tablas de frecuencia** usando `.value_counts()` de pandas, almacenando cada tabla en una variable nombrada para facilitar su uso posterior.
- **Visualización personalizada:**
 - Barras verticales para frecuencias absolutas.
 - Gráficos de pastel (pie chart) para variables binarias o de pocas categorías.
 - Estilización grafica por variable para facilitar su interpretación.

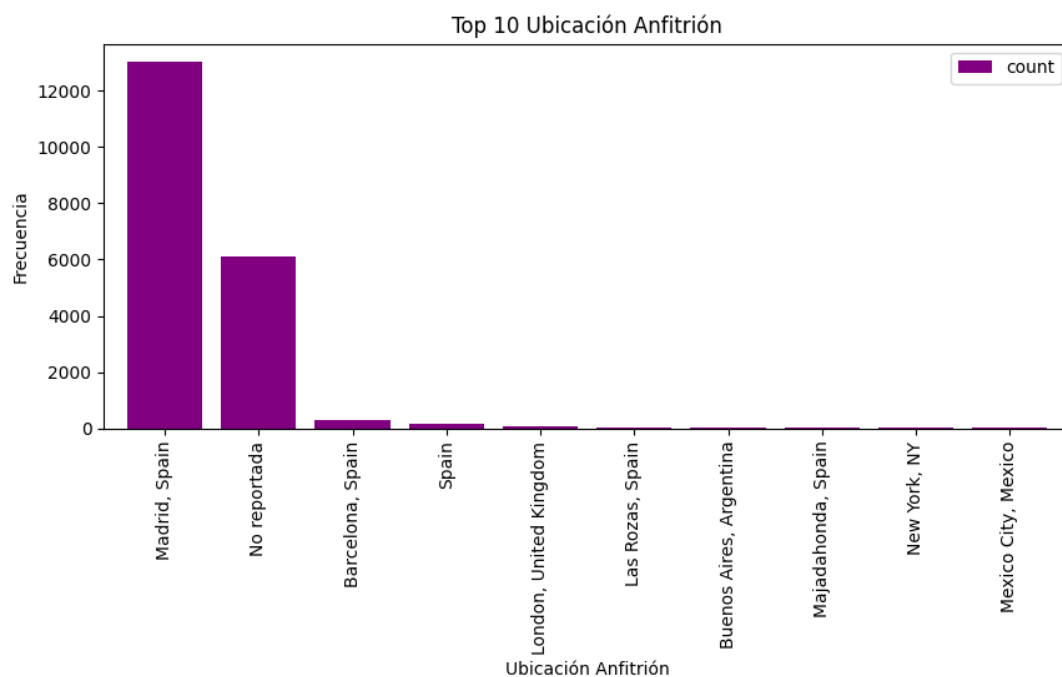
Variables analizadas:

- `ubicacion_anfitrion` (*top 10*)
- `barrio` (*top 10*)

- tipo_propiedad (*top 10*)
- tiempo_respuesta_anfitrion
- es_superanfitrion
- verificaciones_anfitrion
- anfitrion_tiene_foto
- distrito
- tipo_habitacion
- identidad_anfitrion_verificada

3. Análisis y Hallazgos

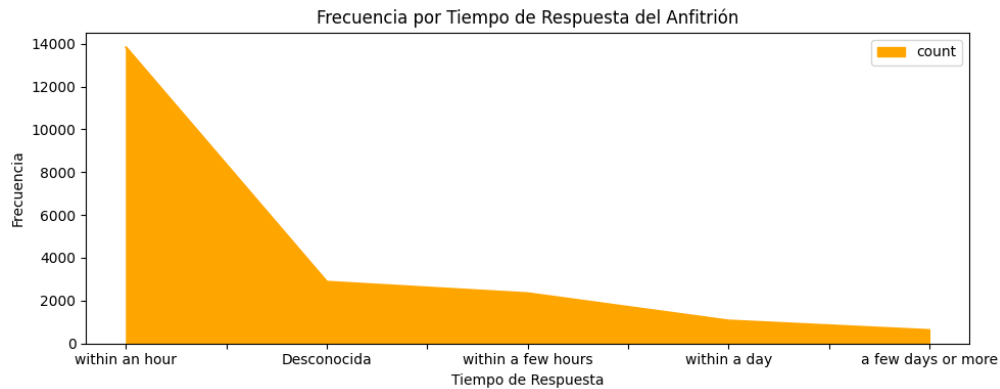
Ubicación del anfitrión



- **Madrid, Spain** lidera ampliamente como la ciudad de residencia de los anfitriones.

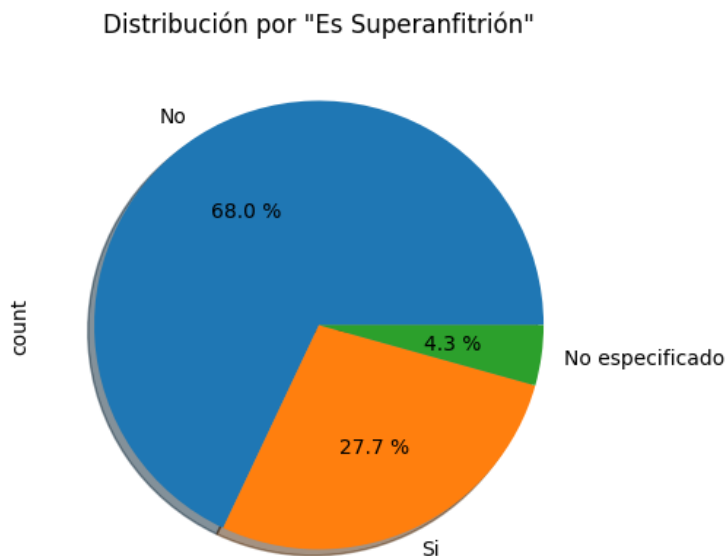
- Le siguen **Barcelona** y valores faltantes (*No reportado*), indicando potenciales vacíos en la calidad de los datos.
- El dominio de ubicaciones en **España** es evidente.

Tiempo de respuesta del anfitrión



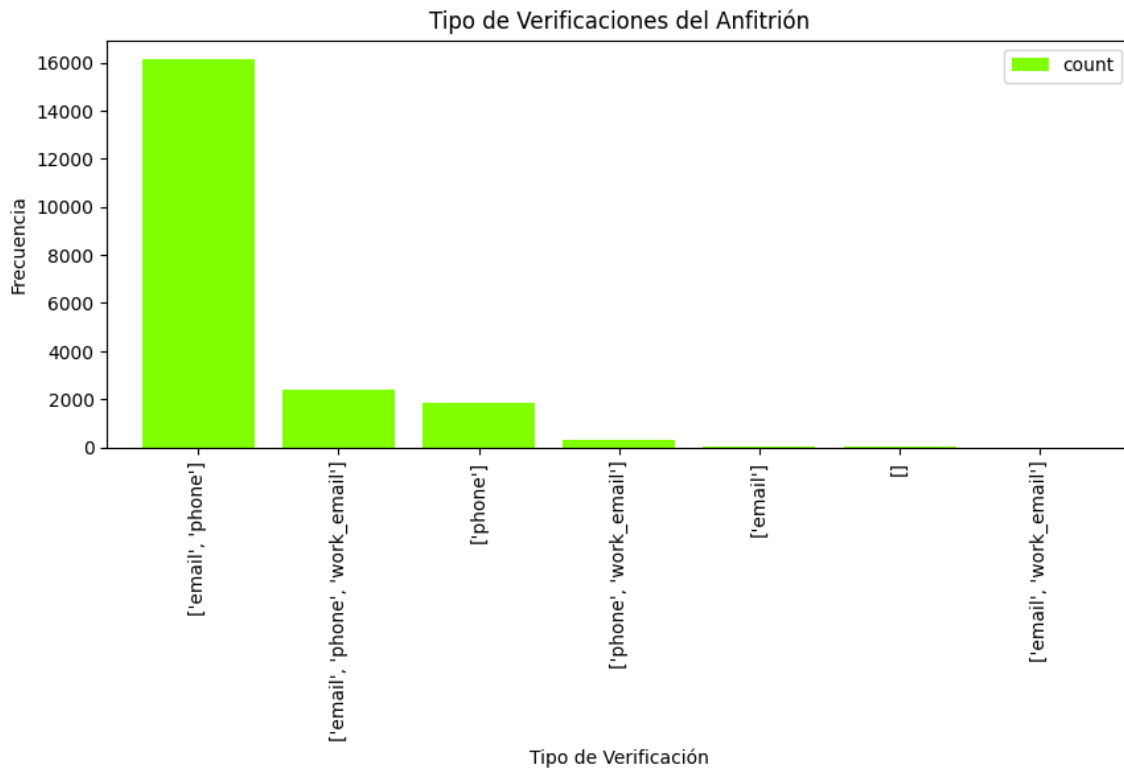
- La mayoría de anfitriones responden "**within an hour**", lo que refleja **alta reactividad** y compromiso.
- Un número considerable aparece como "**Desconocido**", lo que puede indicar registros incompletos o desactualizados.

Es superanfitrión



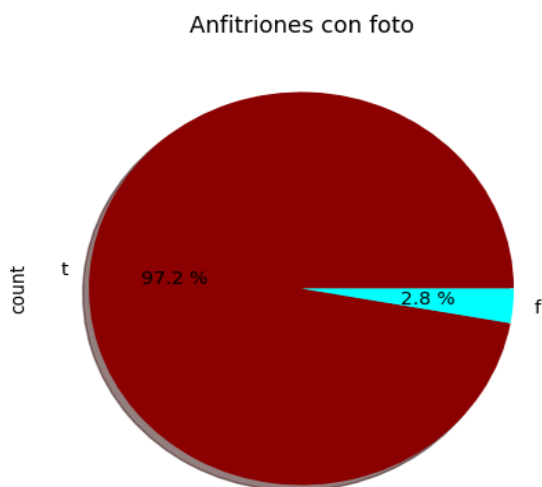
- Alrededor del **68% no son superanfitriones**, mientras que el **27.7% sí lo son**.
- La proporción sugiere que convertirse en superanfitrión sigue siendo una **distinción limitada** a un grupo reducido.

Verificaciones del anfitrión



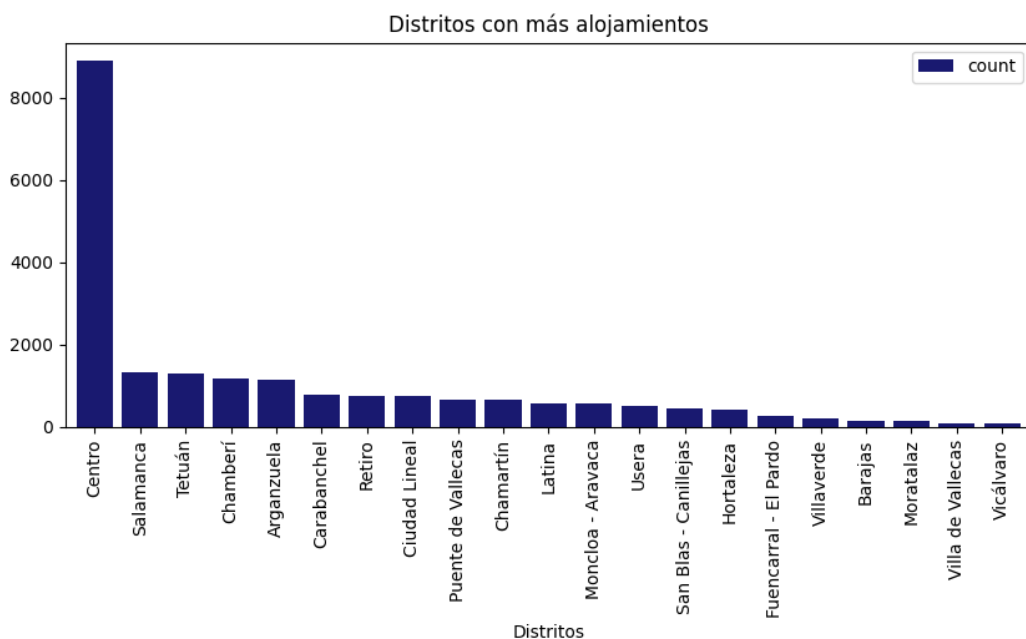
- Las verificaciones más comunes son combinaciones entre "email" y "phone".
- Aparecen múltiples combinaciones, pero también categorías **vacías o mínimas** como [], que deben considerarse para imputación o limpieza.

¿El anfitrión tiene foto?



- Mayoría (**97.2%**) **sí tienen foto**, lo cual contribuye a la **confianza del huésped**.
- Solo un pequeño porcentaje no cuenta con imagen, aunque esto puede afectar la conversión en la plataforma.

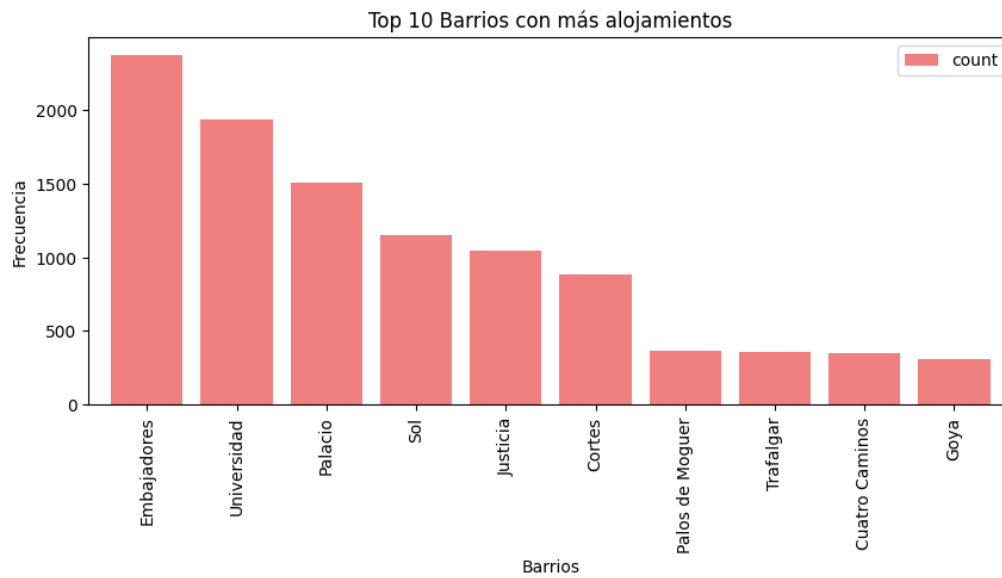
Distrito



- **Centro** y **Salamanca** dominan como zonas con mayor concentración de alojamientos.

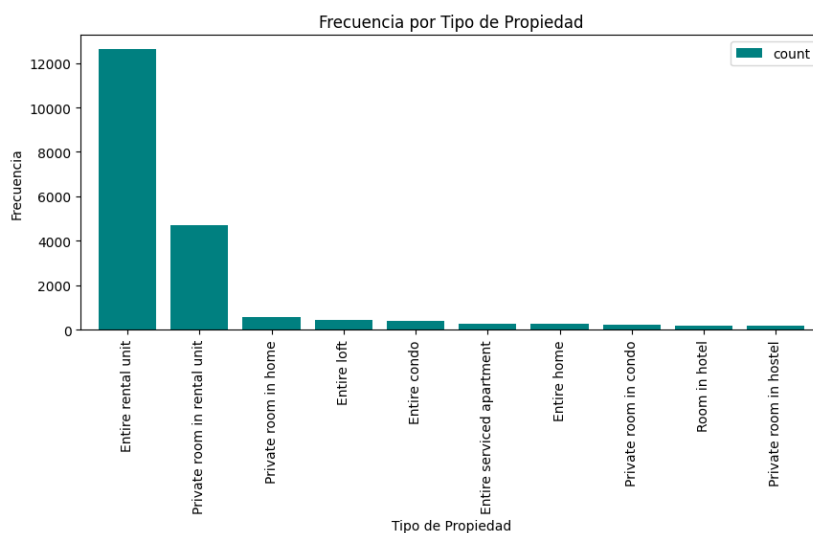
- Muestra una **concentración urbana** fuerte, con caídas progresivas hacia otros distritos.

Barrio



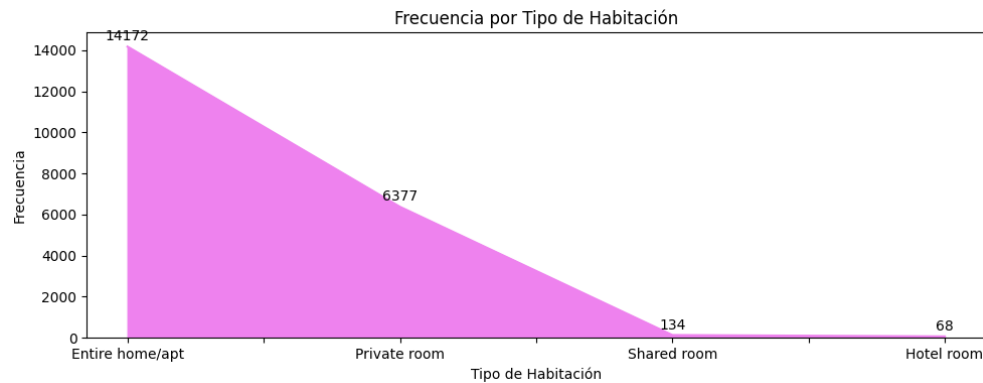
- Los **barrios más populares** están encabezados por **Embajadores, Universidad y Palacio**.
- Es evidente la **distribución desigual**, con la mayoría de los alojamientos en pocas zonas.

Tipo de propiedad



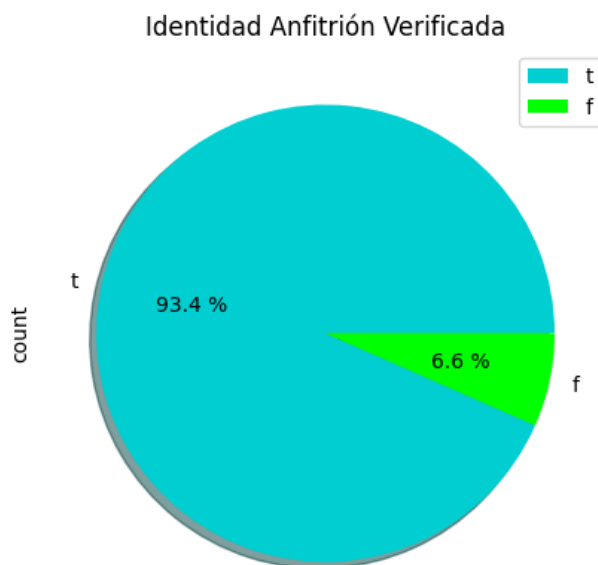
- Predomina la oferta de "**Entire rental unit**", seguido por habitaciones privadas.
- Muestra una **preferencia por alojamientos independientes**, aunque también hay diversidad.

Tipo de habitación



- Se confirma lo anterior: **más del 50% de las propiedades ofrecen la totalidad del espacio**, seguidas por habitaciones privadas.

Identidad verificada



- Más del **93% de los anfitriones tienen su identidad verificada**, lo que mejora la **confiabilidad de la plataforma**.

Conclusiones

- El análisis revela **fuertes concentraciones geográficas** (Madrid y distritos céntricos), y un **alto grado de profesionalización** por parte de los anfitriones (foto, identidad verificada, rápida respuesta).
- La presencia de valores como "Desconocido", "[]", "No reportado" plantea desafíos de **data quality** que deberán atenderse.
- Estas variables y sus distribuciones serán clave para análisis bivariados y para predecir el desempeño de un alojamiento.

Categorización y Preparación de Variables Cualitativas

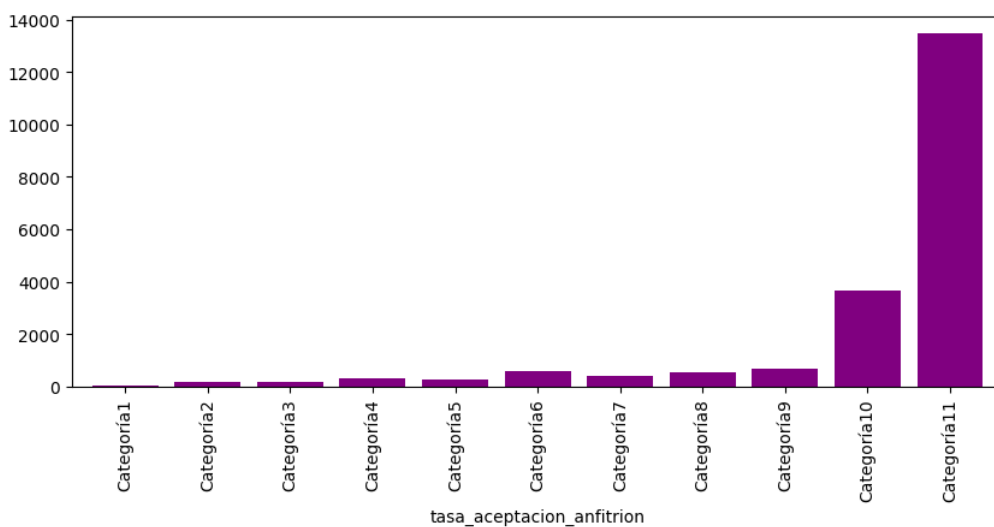
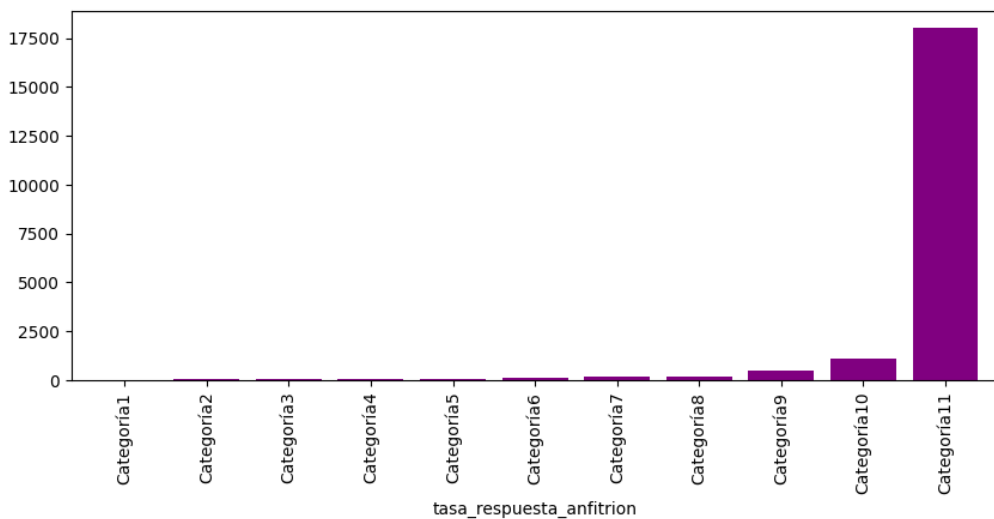
En esta fase se realizó el análisis univariado de variables categóricas:

- Se generaron **tablas de frecuencia** para cada variable categórica de interés, con el objetivo de conocer su distribución y preparar visualizaciones adecuadas.
- Para evitar dispersión en gráficos y análisis, se limitaron a las **10 categorías con mayor frecuencia** las siguientes variables:
 - ubicacion_anfitrión
 - barrio
 - tipo_propiedad
- Para el resto de las variables categóricas como distrito, tipo_habitacion, verificaciones_anfitrión, etc., se mantuvieron todas sus categorías, pero se les estableció un índice categórico para facilitar el acceso y visualización en análisis posteriores.

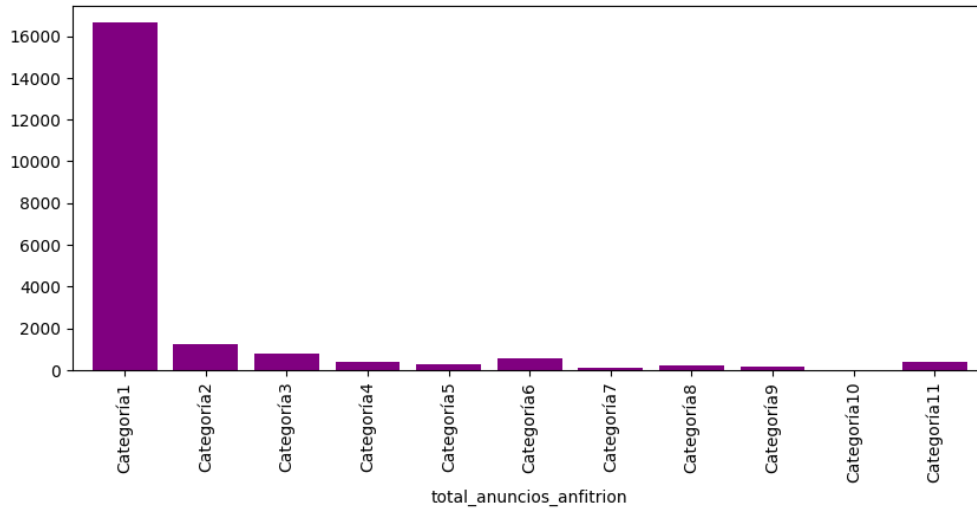
Esto permitió controlar la complejidad y preparar visualizaciones más legibles, mejorando así la interpretación de los resultados.

3. Análisis de Gráficas – Variables Cuantitativas Categorizadas

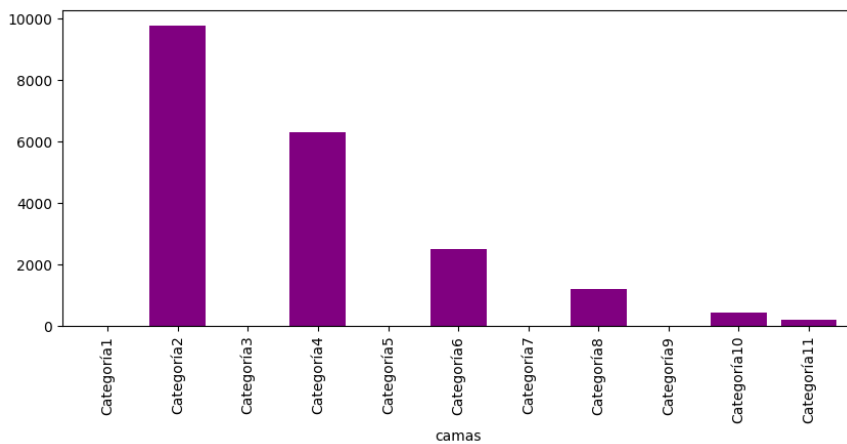
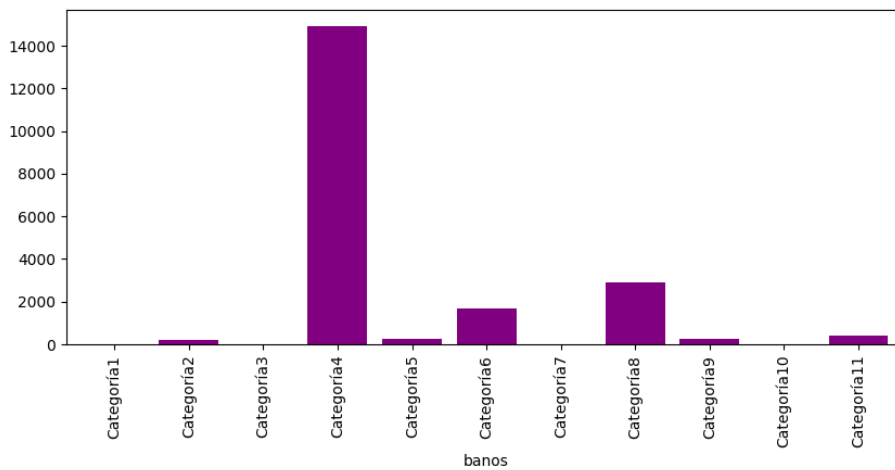
A continuación, se presenta un resumen de hallazgos con base en las visualizaciones generadas:



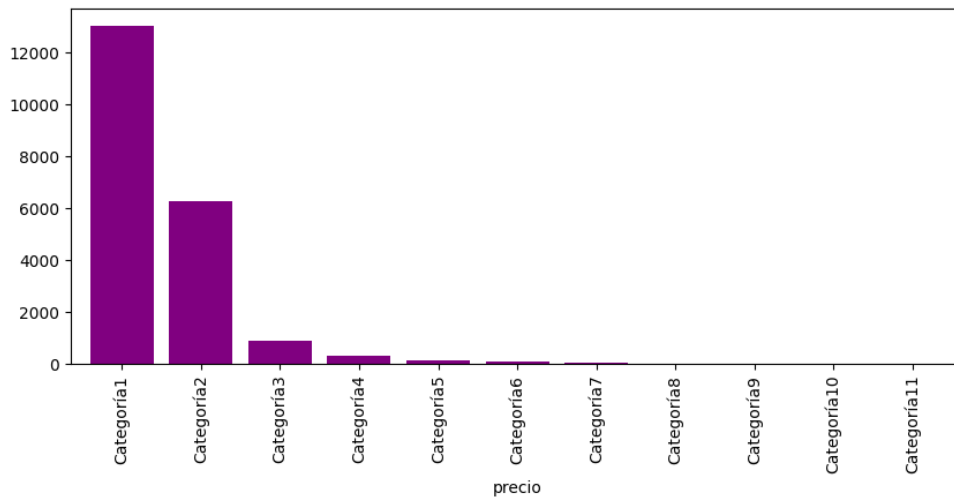
- **Tasa de respuesta y aceptación del anfitrión:** Se observó una gran concentración en las categorías más altas, indicando que la mayoría de los anfitriones tienen una buena interacción con los usuarios.



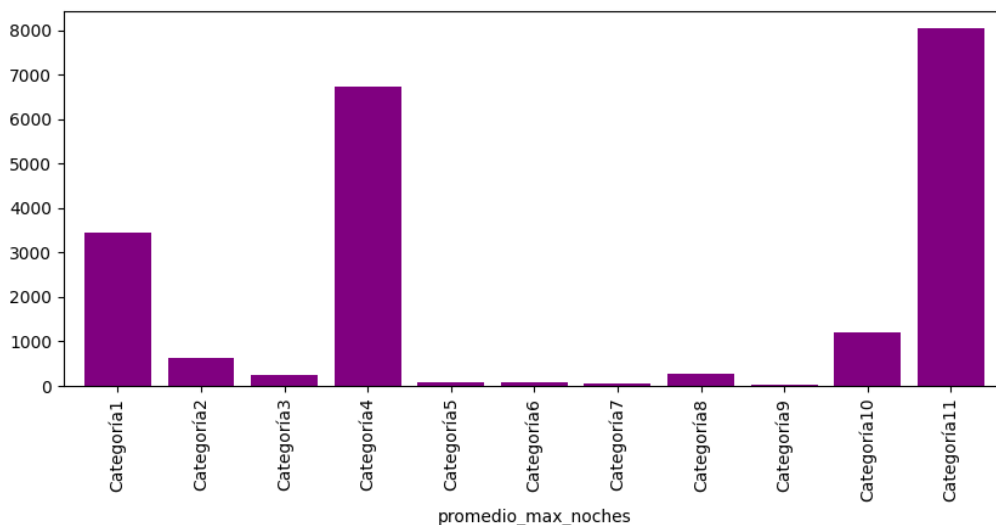
- **Total de anuncios del anfitrión:** Aunque la mayoría de los anfitriones tienen pocos anuncios, existen algunos con muchísimos, lo que sugiere la presencia de operadores profesionales o empresas.



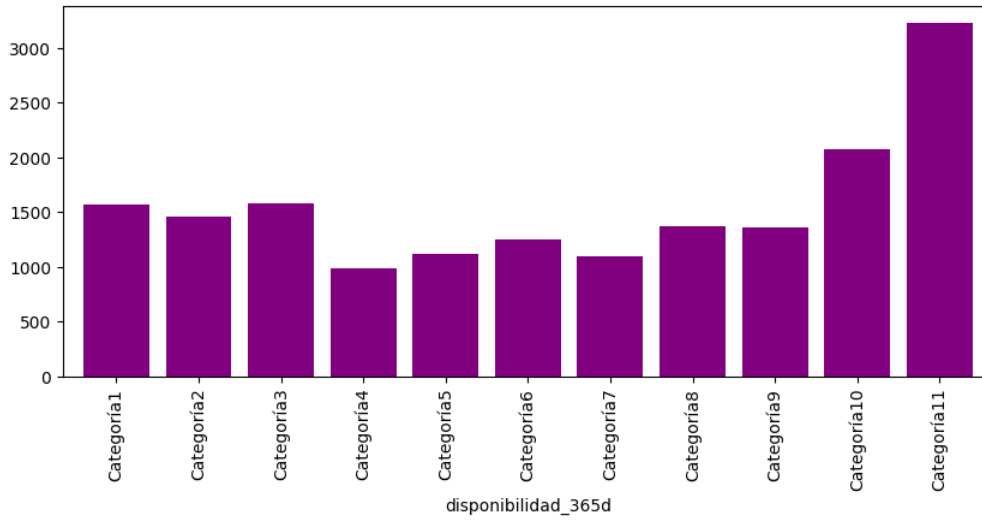
- **Baños y camas :** Estas variables muestran patrones similares, donde lo más común es tener configuraciones básicas (1 a 2 baños, 2 a 4 camas, 1 a 2 habitaciones), aunque existen propiedades de gran tamaño.



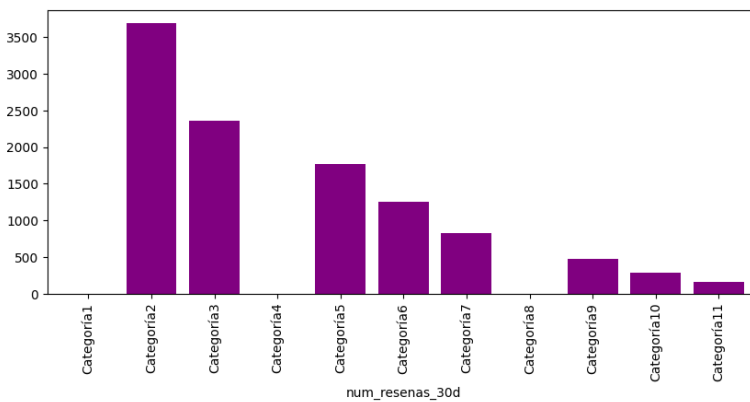
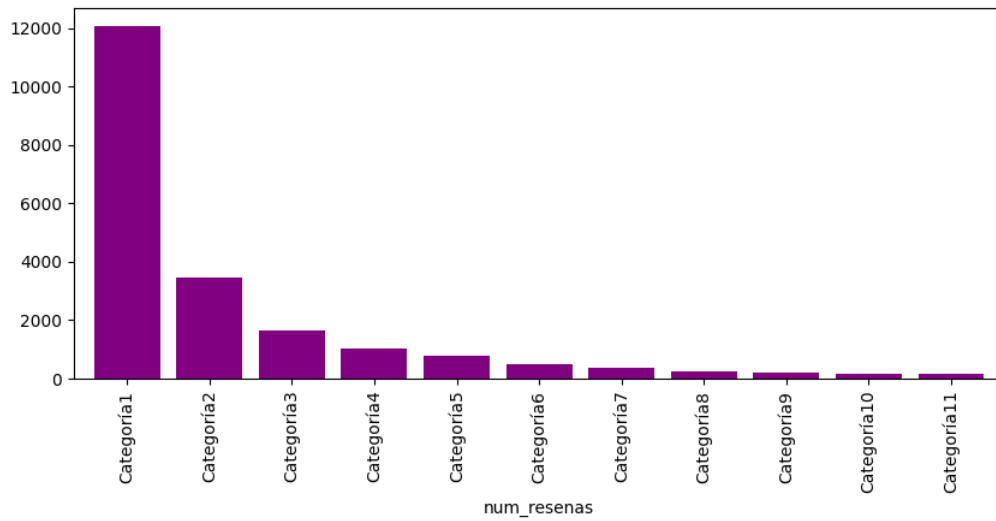
- **Precio:** Predomina una gran cantidad de propiedades con precios bajos a moderados, pero existen propiedades de lujo con precios elevados (outliers).



- **Promedio de noches mínimas y máximas:** Se presentan patrones de agrupamiento en pocos valores frecuentes, indicando que los anfitriones tienden a ofrecer estancias típicas (mínimo 1 o 2 noches, máximo usual entre 7 y 30 noches).



- **Disponibilidad 365 días:** La distribución es relativamente equilibrada, aunque hay una ligera tendencia a mantener la propiedad disponible la mayor parte del año.



- **Número de reseñas:** La mayoría de las propiedades tienen pocas reseñas, lo que puede deberse a nuevas publicaciones o baja demanda, aunque también existen casos con cientos de reseñas.

Conclusión General

El preprocesamiento aplicado a esta base de datos de Airbnb en Madrid ha permitido limpiar, imputar y categorizar variables de forma estratégica y fundamentada. Gracias a ello, ahora contamos con un conjunto de datos confiable y representativo para posteriores análisis exploratorios o segmentación de mercado.

Las visualizaciones han permitido identificar patrones clave del mercado de hospedaje en Madrid, como:

- La profesionalización de ciertos anfitriones.
- La predominancia de estancias básicas.
- La variabilidad en precios y disponibilidad.
- La alta interacción de los anfitriones con los usuarios (respuesta y aceptación).

Este análisis proporciona un punto de partida sólido para estudios más avanzados sobre comportamiento del mercado, optimización de precios, segmentación de propiedades y recomendaciones para nuevos anfitriones.