



Tecnológico de Monterrey

Actividad 1.1

Extracción de Características

Lucero Jannete López García A01736938

**Analítica de datos y herramientas de inteligencia artificial
I (Gpo 101)**

Fecha

23/03/2025

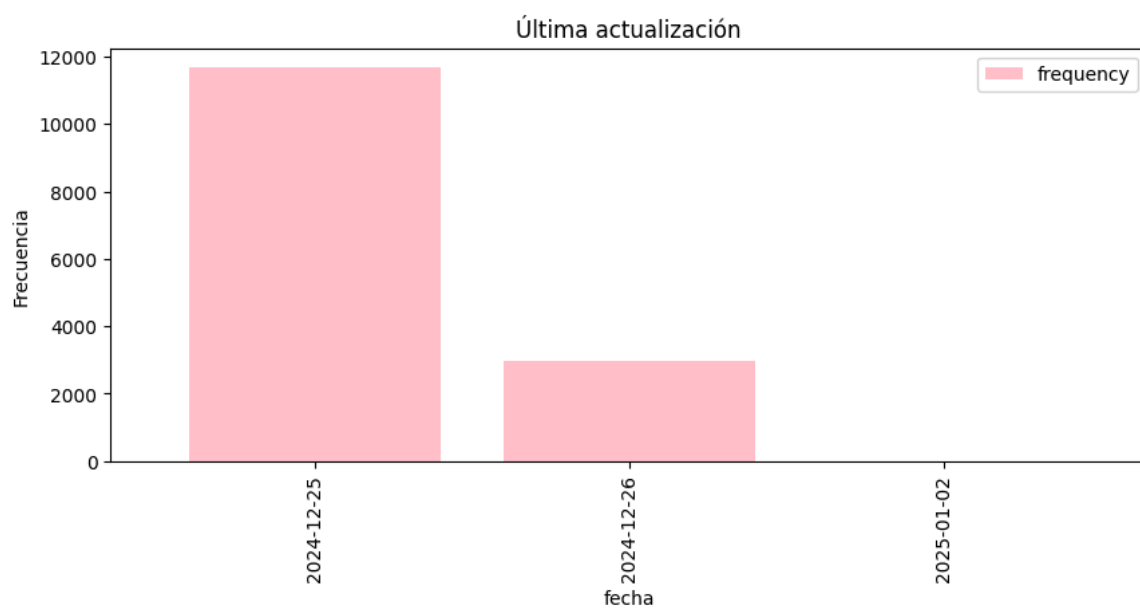
Para realizar los siguientes análisis cuento con un datagrama de 50 columnas y 14642 filas, las cuáles ya fueron previamente limpiadas con respecto a valores nulos y outlets en el bloque 1 de análisis de datos, de igual, el dataframe cuenta con 36 variables de tipo float64, 3 variable de tipo int64, y 12 variables de tipo object.

Cabe destacar que solo seleccionaremos 10 categorías y 10 numéricas para la creación de sus gráficas e interpretaciones. Para generar las gráficas se usaron 3 formatos de tipo barra, de áreas y de tipo pastel, sin embargo solo se seleccionará una por cada variable dependiendo de qué tan factible es el uso.

Variables categóricas:

1. Last scrape/Última fecha de actualización:

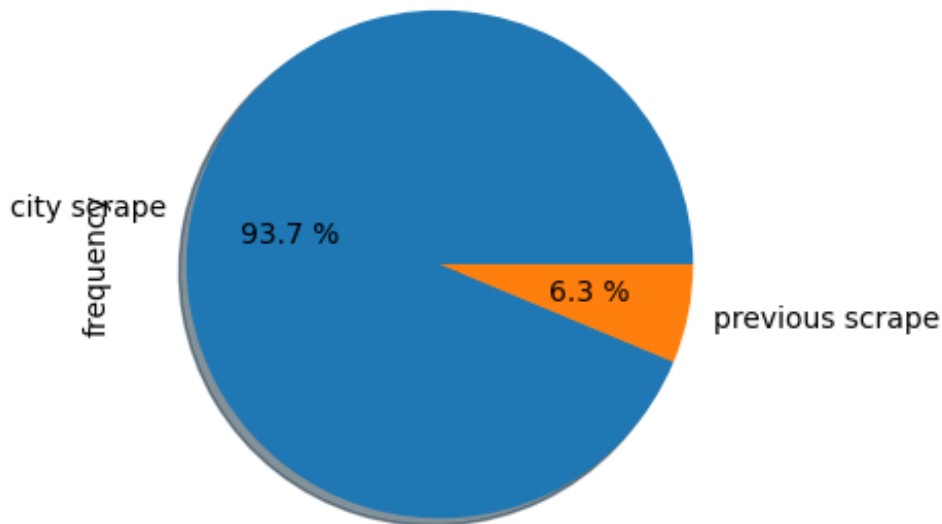
El análisis de la frecuencia de "last_scraped" en Airbnb muestra un pico significativo el 25 de diciembre de 2024, con 11,668 registros, seguido de una caída drástica a 2,973 registros el 26 de diciembre, y una casi total ausencia de datos el 2 de enero de 2025. Esta tendencia sugiere que el scraping de datos fue intenso en un solo día, posiblemente debido a una extracción masiva o una actualización del sistema. La disminución posterior podría indicar restricciones en la recopilación de datos, un cambio en la disponibilidad de anuncios, o ajustes en el proceso de scraping. La ausencia casi total de registros en enero refuerza la posibilidad de que el acceso a la información haya sido limitado o interrumpido.



2. Source/fuente:

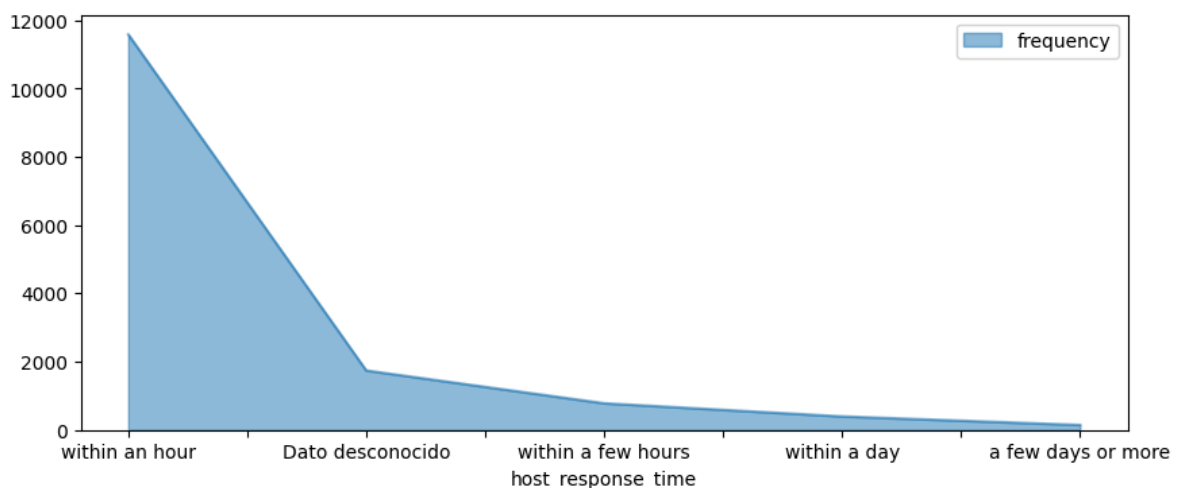
La mayoría de los datos en Airbnb provienen del "city scrape" (13,722 registros), mientras que el "previous scrape" solo tiene 920 registros, lo que muestra una diferencia bastante marcada. Esto me hace pensar que el "city scrape" fue una extracción más reciente y completa, probablemente enfocada en actualizar los listados de una ciudad específica. En cambio, el "previous scrape" parece ser un conjunto de datos más antiguo o menos exhaustivo. Esta variación podría deberse a cambios en la

disponibilidad de anuncios, en la forma en que se recopilan los datos o incluso en restricciones impuestas por la plataforma.



3. Host response time/ Tiempo de respuesta del host:

Podemos observar que conforme va transcurriendo el tiempo, la frecuencia va disminuyendo, lo que significa que la mayoría de los anfitriones responden al momento porque lo que buscan los usuarios es rapidez en el tiempo de atención.

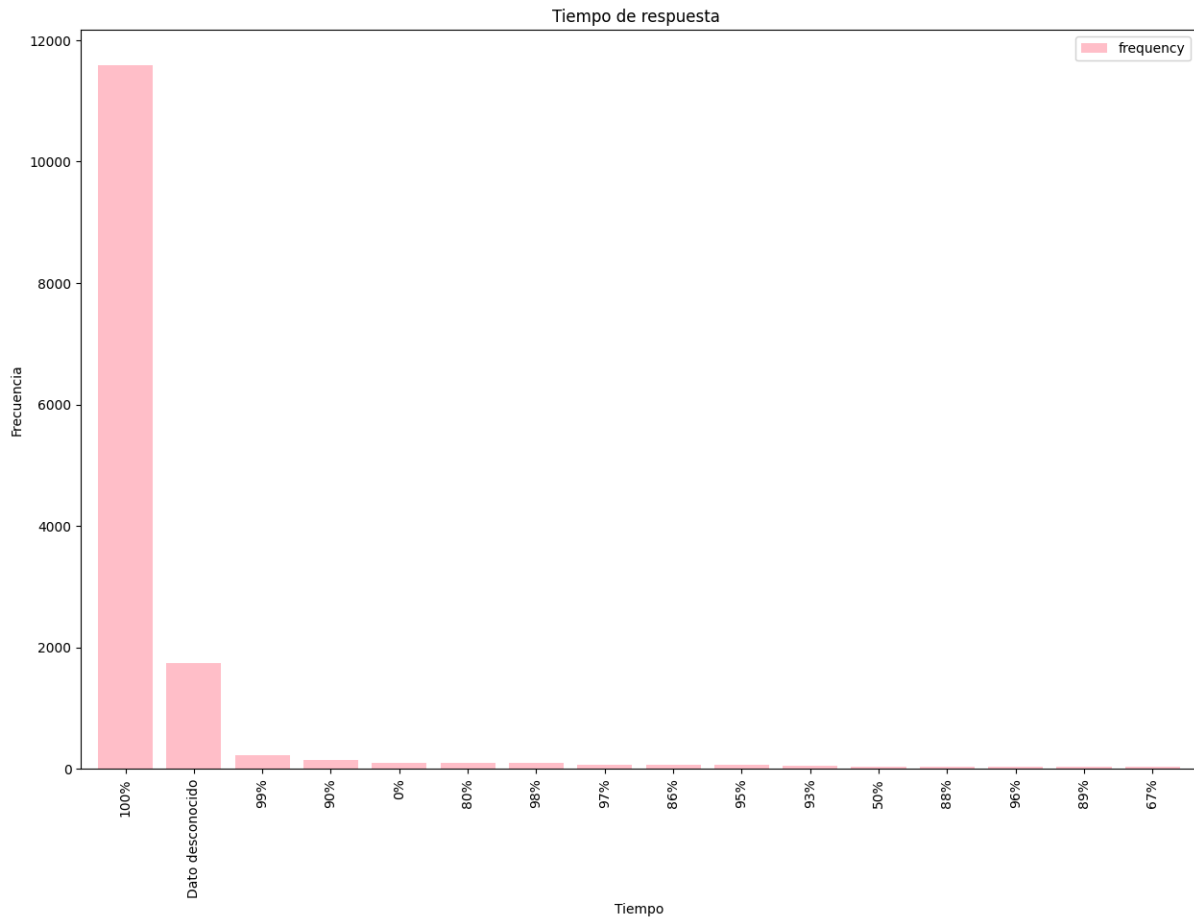


4. Host response rate/ Tasa de respuesta:

Para visualizar estos datos, la mejor opción sería una gráfica de barras, ya que permite comparar fácilmente la frecuencia de cada nivel de `host_response_rate`.

En cuanto al significado de los datos, la mayoría de los anfitriones tienen una tasa de respuesta del 100% (11,590 casos), lo que indica que son muy activos respondiendo a los huéspedes. Sin embargo, hay 1,739 registros con datos desconocidos, lo que podría deberse a la falta de interacción o a que Airbnb no ha registrado respuestas en

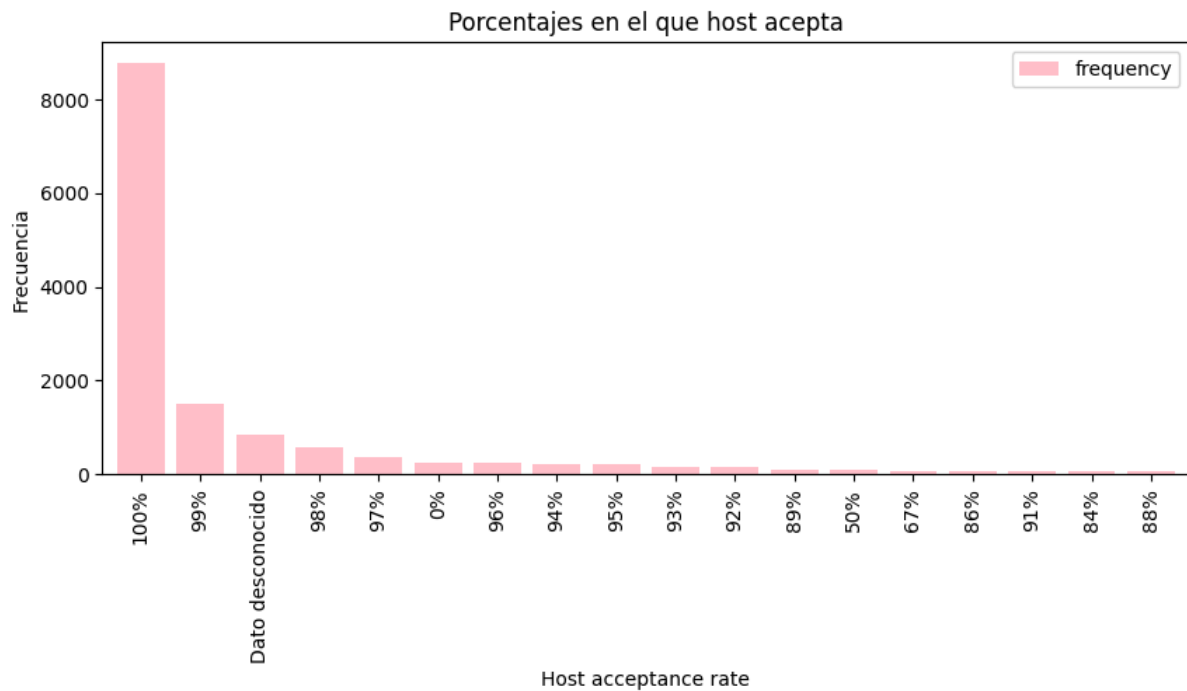
esos casos. A medida que la tasa de respuesta disminuye, también lo hace el número de anfitriones, lo que sugiere que la mayoría mantiene una comunicación constante con los huéspedes. Un pequeño grupo tiene tasas de respuesta muy bajas, como 0% o 50%, lo que podría significar que son anfitriones inactivos o que no revisan sus mensajes con frecuencia.



5. Host acceptance rate /Tasa de aceptación:

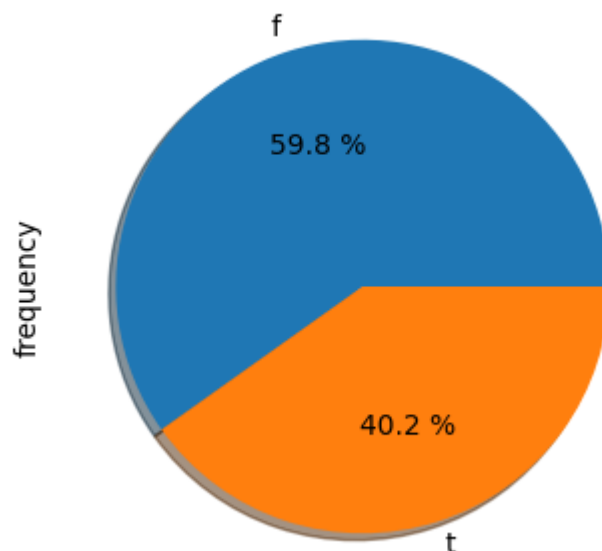
La mayoría de los anfitriones tienen una tasa de aceptación del 100% (8,781 registros), lo que indica que aceptan casi todas las solicitudes de reserva. Le siguen los anfitriones con una tasa del 99% (1,507 casos), lo que refuerza la idea de que la mayoría tiene una alta disposición a aceptar huéspedes. Sin embargo, hay 831 registros con datos desconocidos, lo que podría indicar hosts que no han recibido suficientes solicitudes o cuyos datos no han sido registrados por la plataforma.

Por otro lado, un grupo más pequeño tiene tasas de aceptación bajas, con 258 anfitriones en 0%, lo que sugiere que rechazan todas las solicitudes o han dejado de gestionar su alojamiento activamente. También se observan algunos valores intermedios, como 50% o 67%, lo que podría representar hosts más selectivos con sus huéspedes. En general, la distribución muestra que la mayoría de los anfitriones en Airbnb tienden a aceptar la mayoría de las reservas, mientras que un grupo minoritario mantiene un nivel de selectividad o inactividad en la plataforma.



6. Host is superhost:

Hay 8,762 anfitriones que no son Superhosts y 5,880 que sí lo son. Esto indica que, aunque una parte significativa de los anfitriones ha alcanzado el estatus de Superhost, la mayoría aún no lo ha hecho. Esto podría deberse a que no cumplen con ciertos criterios como alta tasa de respuesta, buenas calificaciones o un número mínimo de reservas en el período requerido.

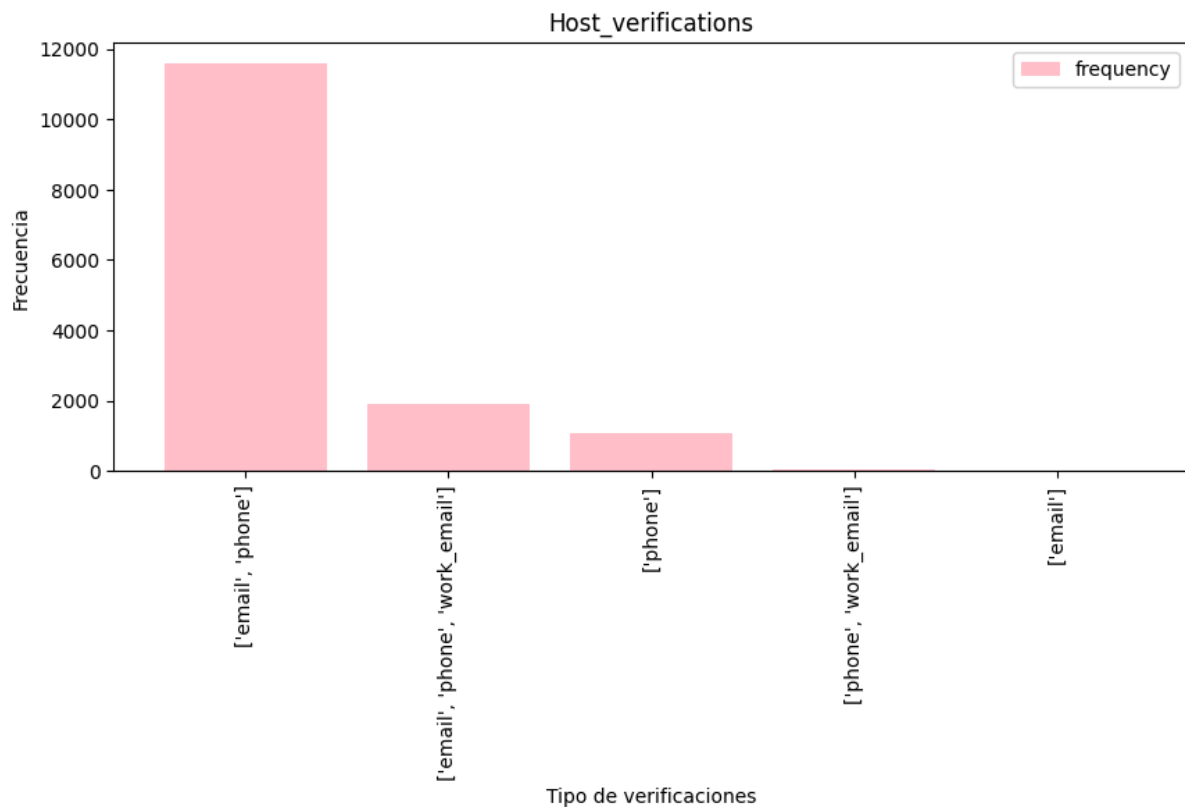


7. Host verifications/verificaciones:

La mayoría de los anfitriones (11,604) han verificado su email y teléfono, lo que indica que estos son los métodos más comunes y probablemente los mínimos

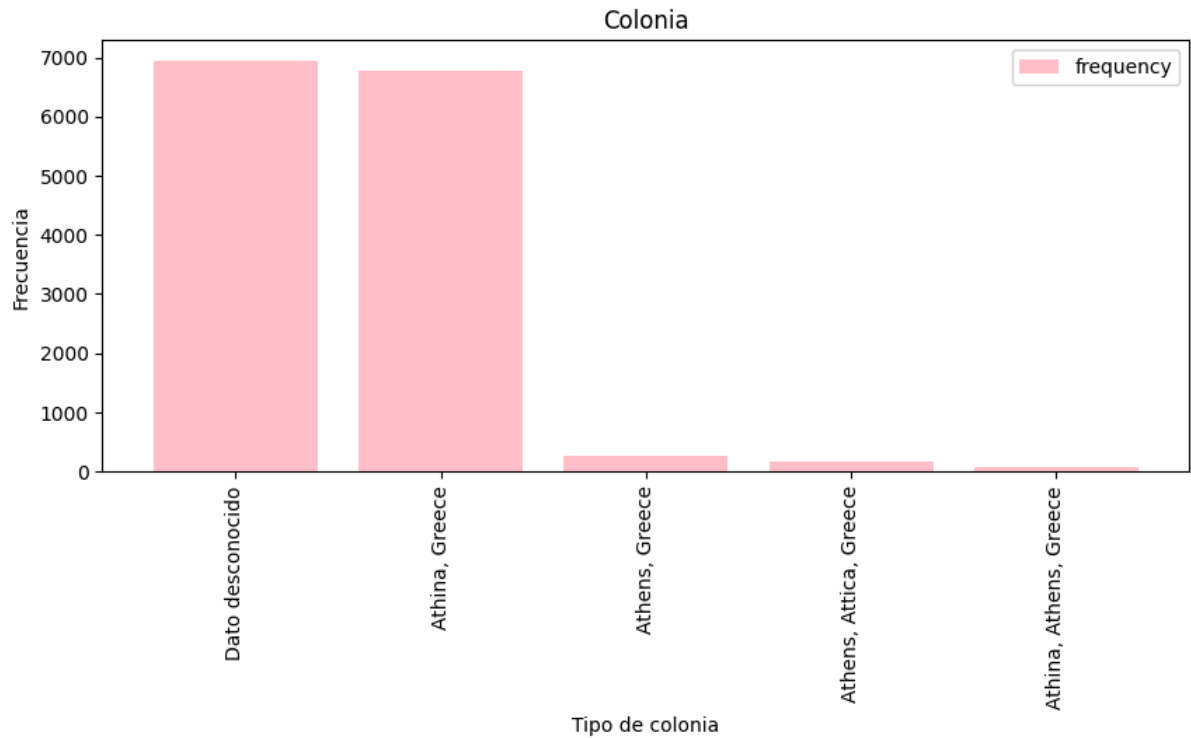
requeridos por Airbnb. Un grupo menor (1,885) también ha verificado un correo de trabajo, posiblemente por motivos profesionales.

Por otro lado, 1,072 solo han verificado su teléfono, y un pequeño grupo tiene verificaciones menos convencionales, como solo email (18 anfitriones) o solo teléfono y correo de trabajo (62 anfitriones). Esto sugiere que la mayoría cumple con verificaciones básicas, mientras que pocos agregan credenciales adicionales, posiblemente por confianza en la plataforma o requisitos mínimos de Airbnb.



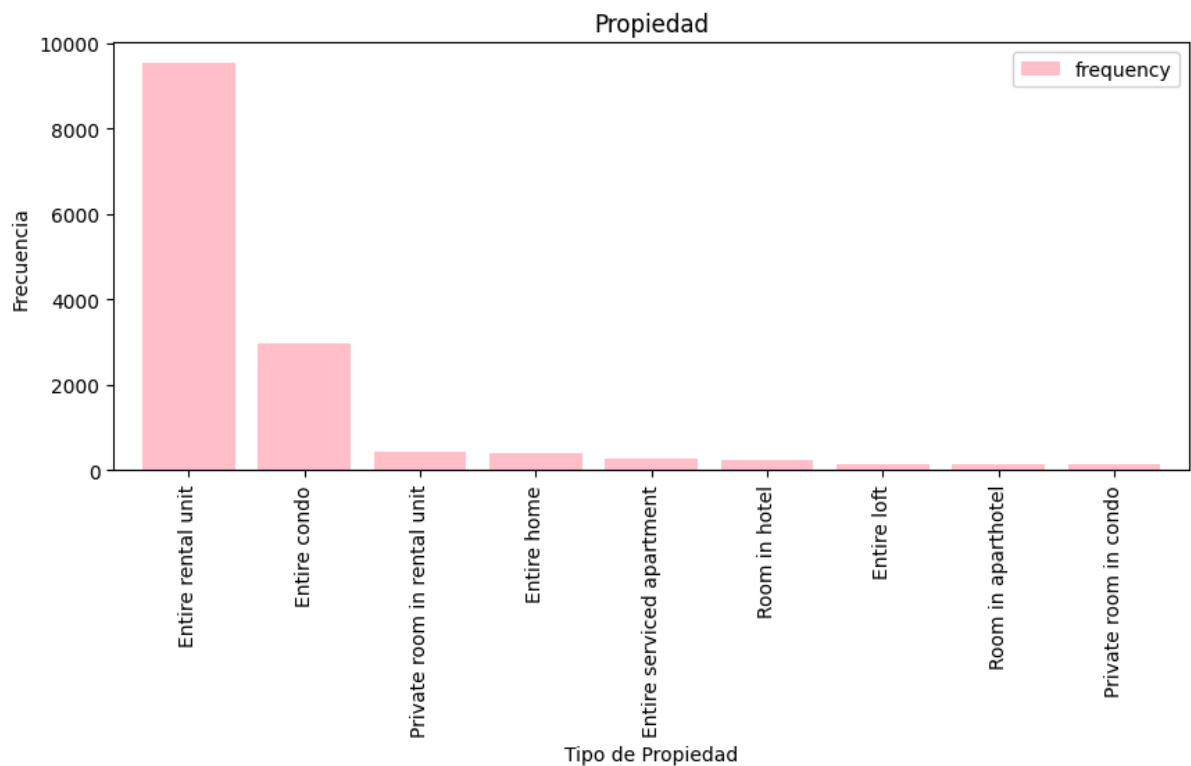
8. neighborhood /Vecindario

Se puede observar que el mayor registro es el de datos desconocidos (6944), esto se debe a que no airbnb no registró los datos o no se verificó en dónde se están hospedando, luego el registro que mayor registro tiene es Athina, Grecia. Este lugar puede ser llamativo para los usuarios debido a que es una zona turística, tiene mejores reseñas, el vecindario es tranquilo etc.



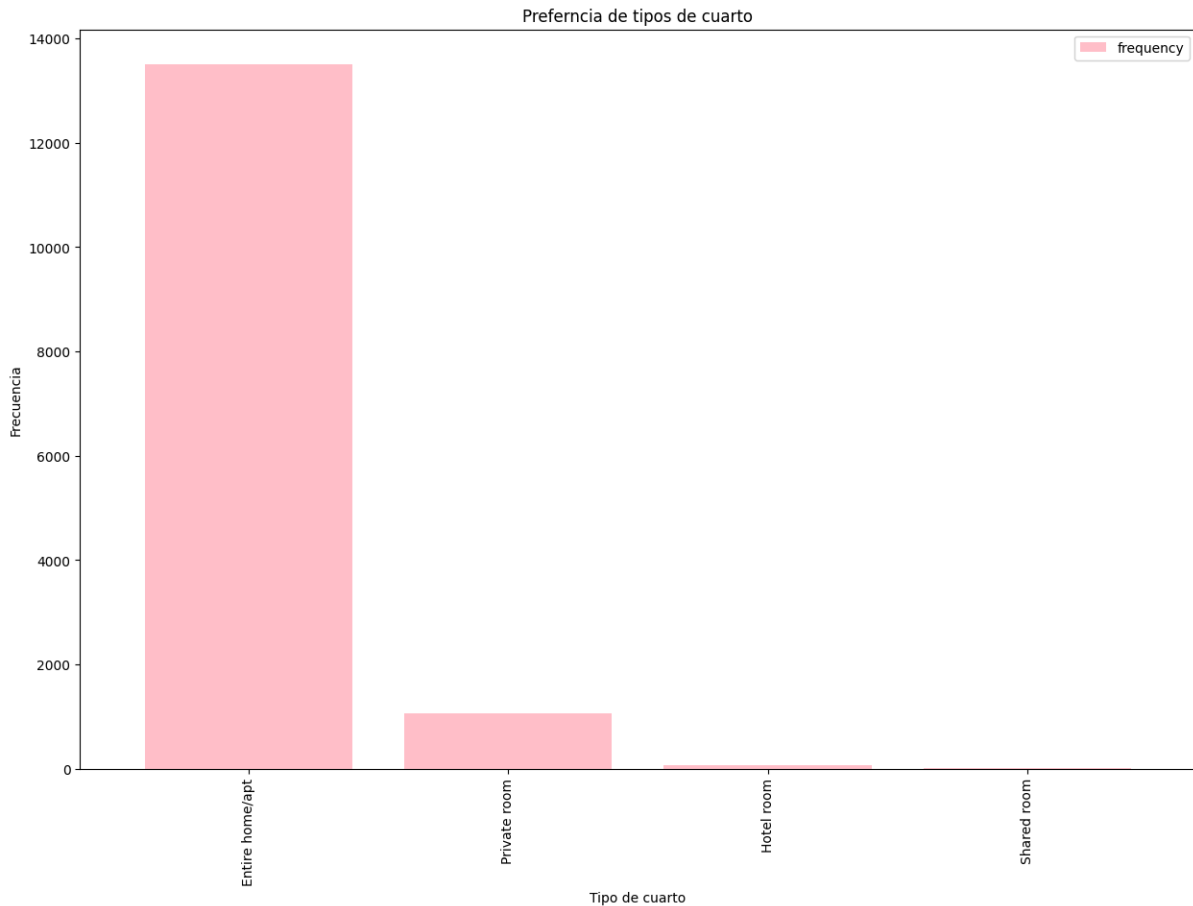
9. Tipo de propiedad:

Para está gráfica se tomaron en cuenta todos los tipos de propiedad y su preferencia, es por ello, que la gráfica de barras puede ayudarnos a visualizar la comparativa. Por lo que podemos observar que el tipo de propiedad en el que mayor se hospedan los usuarios son unidades entrega en renta (9530) debido a que una familia entera viaja o la persona que se hospeda le gusta tener un lugar amplio pero sí mimo y poder disfrutar.



10. Tipo de cuarto:

Ligado a l tipo de propiedad, también existen el tipo de cuarto que más resort los usuario, por ejemplo el patron que se sigue repitiendo es que el usuario renta lugare/apartamento enteros, debido a la comodidad, accesibilidad o que son muchas personas las que viajan juntas y requieren de un lugar amplio.



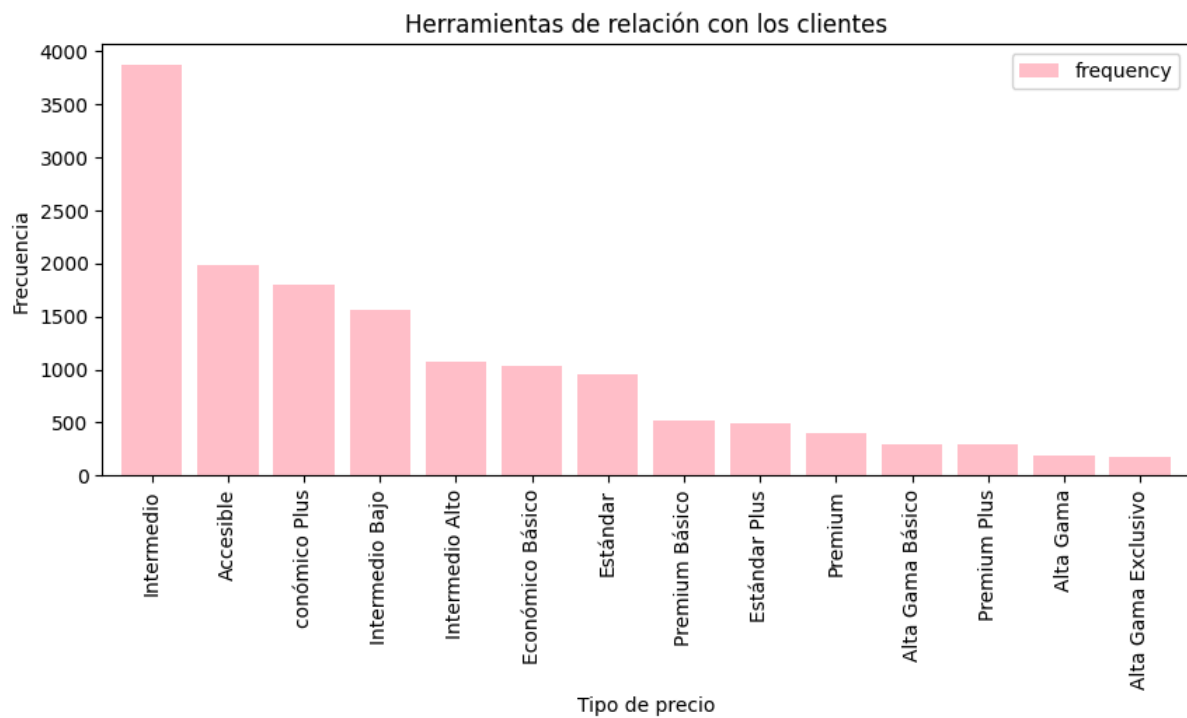
Variable numéricas pasadas a categorías para un mejor análisis, en esta sección la mayoría de los datos fueron rango que se categorizaron en 13 variables, sin embargo, solo se tomaron en cuenta las frecuencias que más destacan para conocer lo que prefiere más el usuario de AIRBNB. De igual forma, solo se tomará una sola gráfica dependiendo de cuál convenga más para el análisis de cada variable.

11. Price /precio:

Para esta variable se crearon 13 categorías que fueron: Intermedio, Accesible, económico Plus Intermedio Bajo, Intermedio Alto, Económico Básico, Estándar, Premium Basic, Estándar Plus, Premium, Alta Gama Básic, Premium Plus, Alta Gama, Alta Gama Exclusivo.

Así mismo, todas las categorías contaban con registros, por lo que, no se eliminó ninguna.

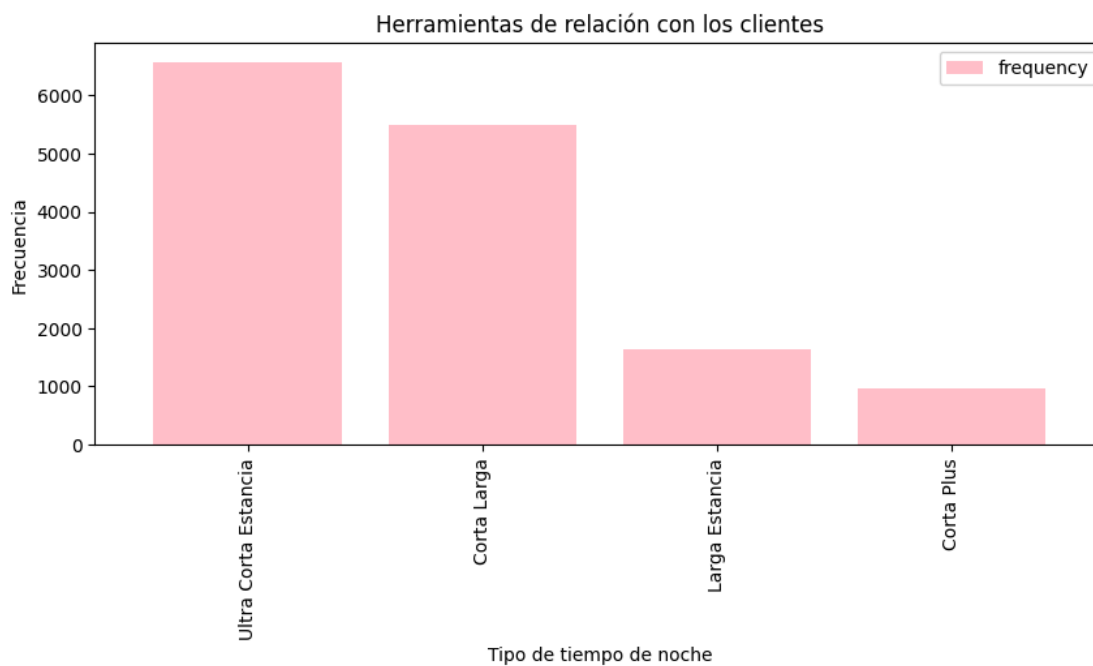
Podemos observar que la que tiene mayores registro es el precio de tipo intermedio (3874) y las que escogieron precios más elevados contaron con menor frecuencia (176)



12. Minimum nights/Noches mínimas:

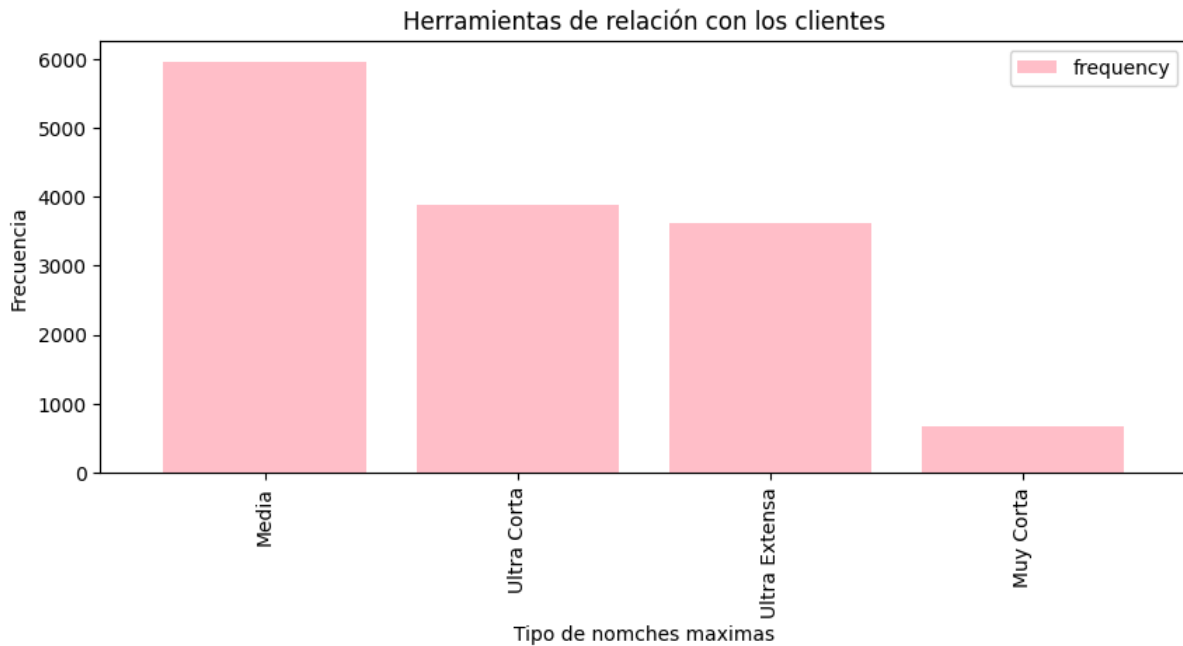
Sun 13 categorías fueron: Ultracorta Estancia, Estancia Express, Estancia Muy Corta, Estancia Corta, Corta Plus, Corta Extendida, Corta Larga, Estándar, Estándar Plus, Media Estancia, Media Plus, Media Extendida, Larga Estancia Básica, Larga Estancia, sin mango no todas contaban con registros solamente las que se muestra en la tabla de abajo.

En la tabla de barras se puede observar que sus estancias mínimas fueron bastante cortas (6571)



13. Maximum nights:

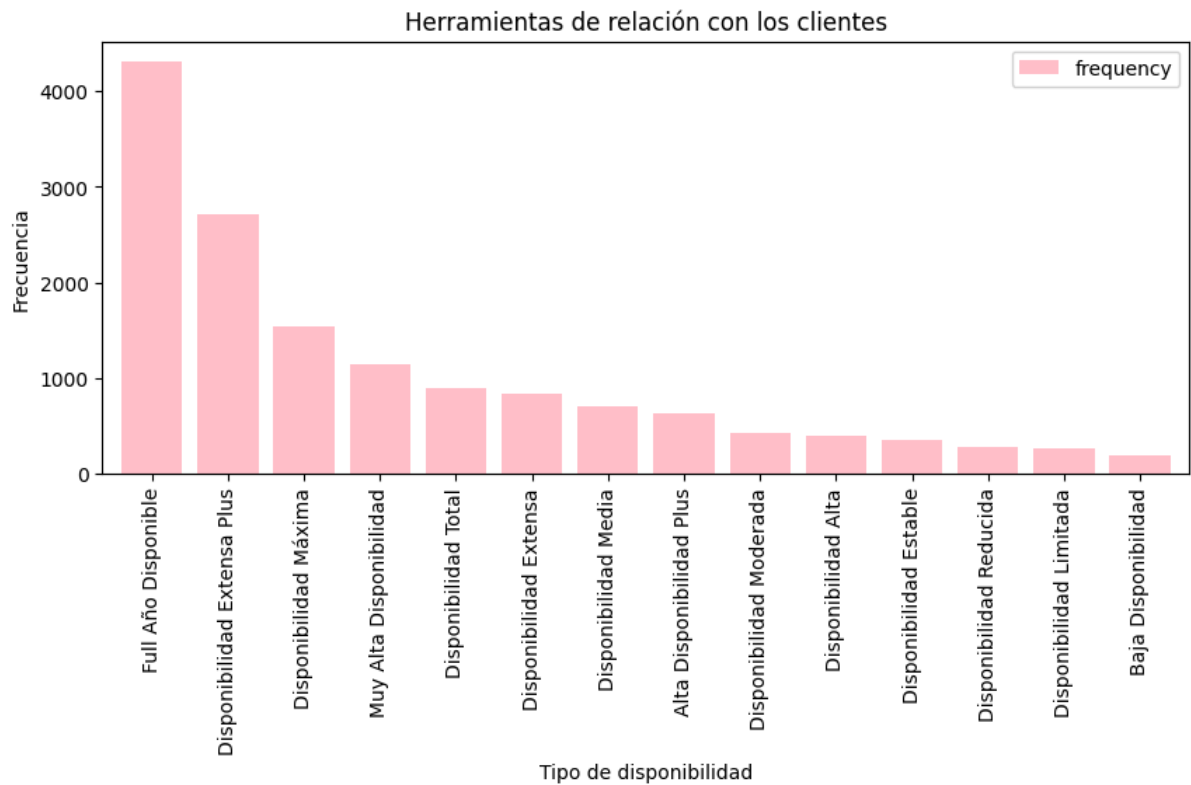
Al tener 13 categorías, varias contaban con frecuencias muy bajas, lo que no nos daría insight significativos, por lo que, se filtró la variable y solo se tomaron las categorías que tuvieran mayor de 500 en frecuencia porque nos interesa saber los valores máximos de cuánta es su duración en un lugar dándonos así las 4 categorías más importantes: Media, ultracorta, ultra extensa, muy corta. Analizando la gráfica podemos observar que los usuarios tienen estancias medias en los lugares en los que se hospedan. (5690)



14. Disponibilidad 365:

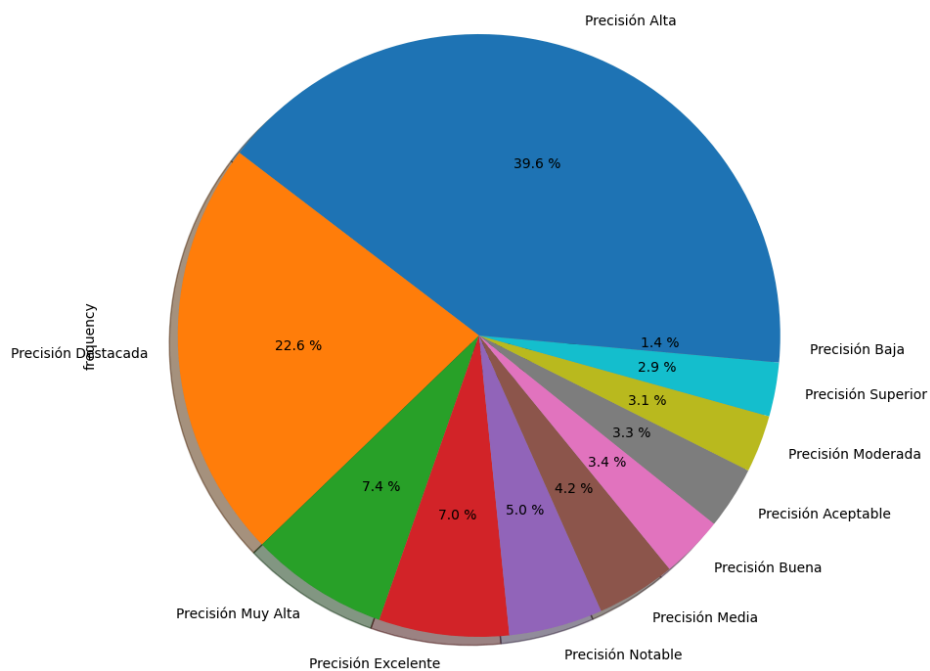
Aquí se tomaron en cuenta las 13 categorías debido a que todas contaban con registros con altas frecuencias e importante para conocer qué prefiere el usuario de airbnb. La mayoría de los anfitriones tienen una disponibilidad extensa o completa. 4,308 anfitriones están totalmente disponibles todo el año, mientras que 2,711 tienen una disponibilidad extensa plus, lo que sugiere que ofrecen una gran cantidad de días disponibles para reservas. A medida que disminuye la disponibilidad, el número de anfitriones también disminuye, con 275 anfitriones con disponibilidad reducida y solo 190 con baja disponibilidad.

Esto refleja que la gran mayoría de los anfitriones están dispuestos a recibir huéspedes en muchas fechas, lo que facilita la reserva y la flexibilidad para los viajeros. Sin embargo, también existe un grupo que tiene una disponibilidad más limitada, lo que podría estar relacionado con su ubicación, tipo de propiedad o preferencias personales sobre la cantidad de huéspedes que aceptan.



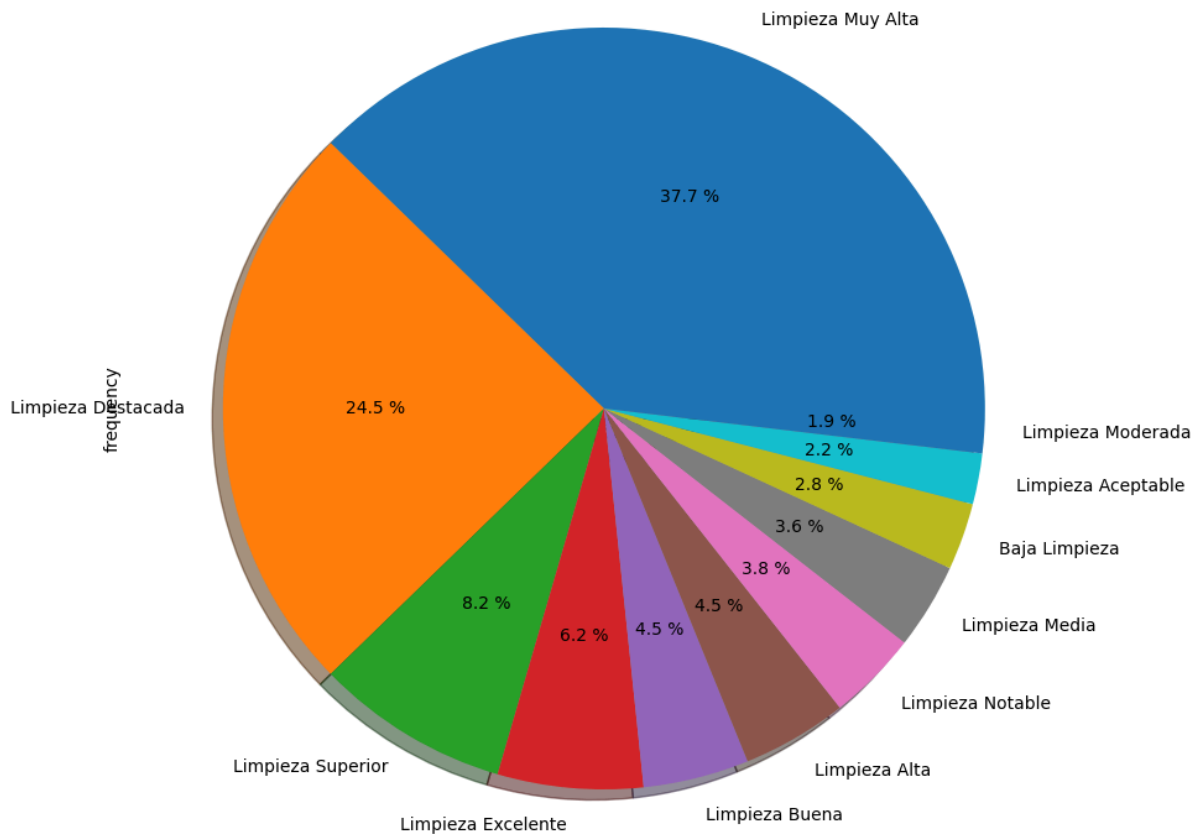
15. Nivel de precisión

Se puede observar que los anfitriones cuentan con un alto nivel de presión con un 36.6%, lo que indica que la mayoría de las propiedades en Airbnb coinciden bien con las descripciones y fotos proporcionadas. A medida que disminuye la precisión, el número de anfitriones también disminuye.



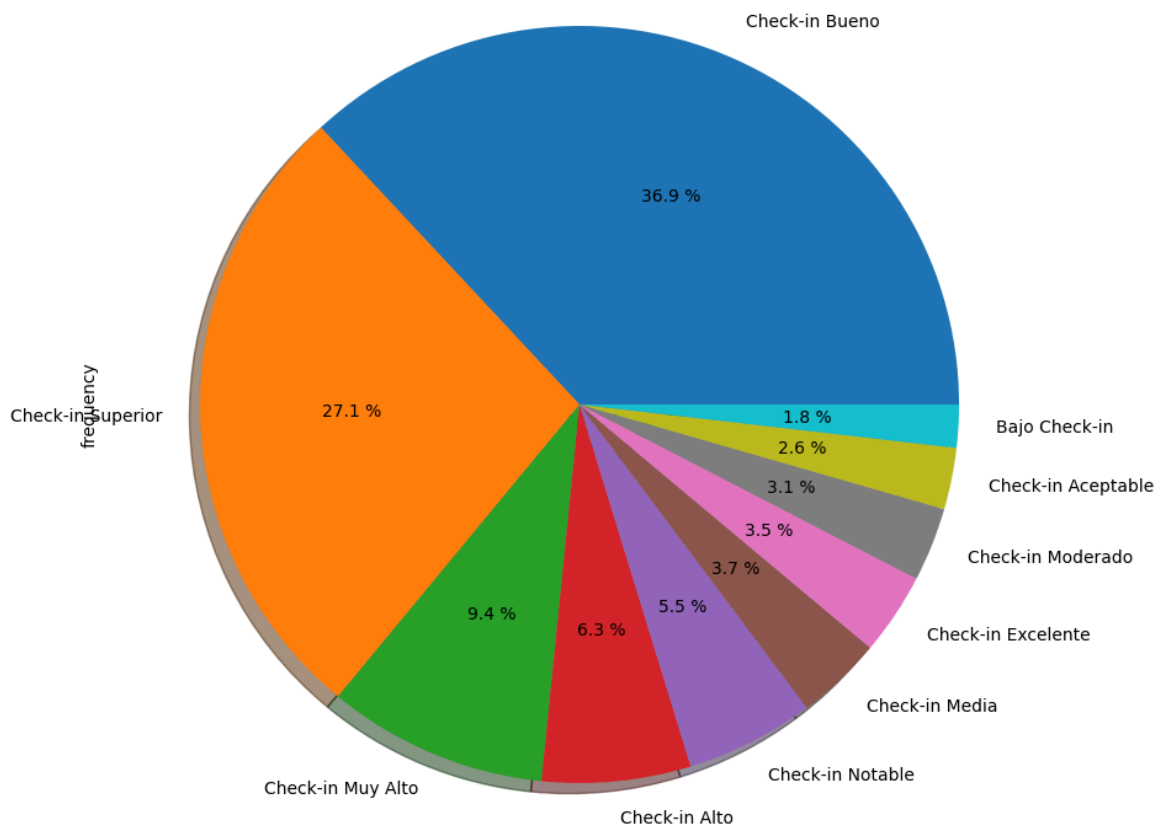
16. Nivel de limpieza:

De igual forma que el nivel de precisión, el 37.7% de los anfitriones mantienen una alta limpieza en los lugares, de ahí un 24.5 % es limpieza destacable, por lo que podríamos coincidir que más del 50% de los anfitriones cuidan y mantienen los lugares de renta en perfecto estado generando que el usuario se encuentre satisfecho.



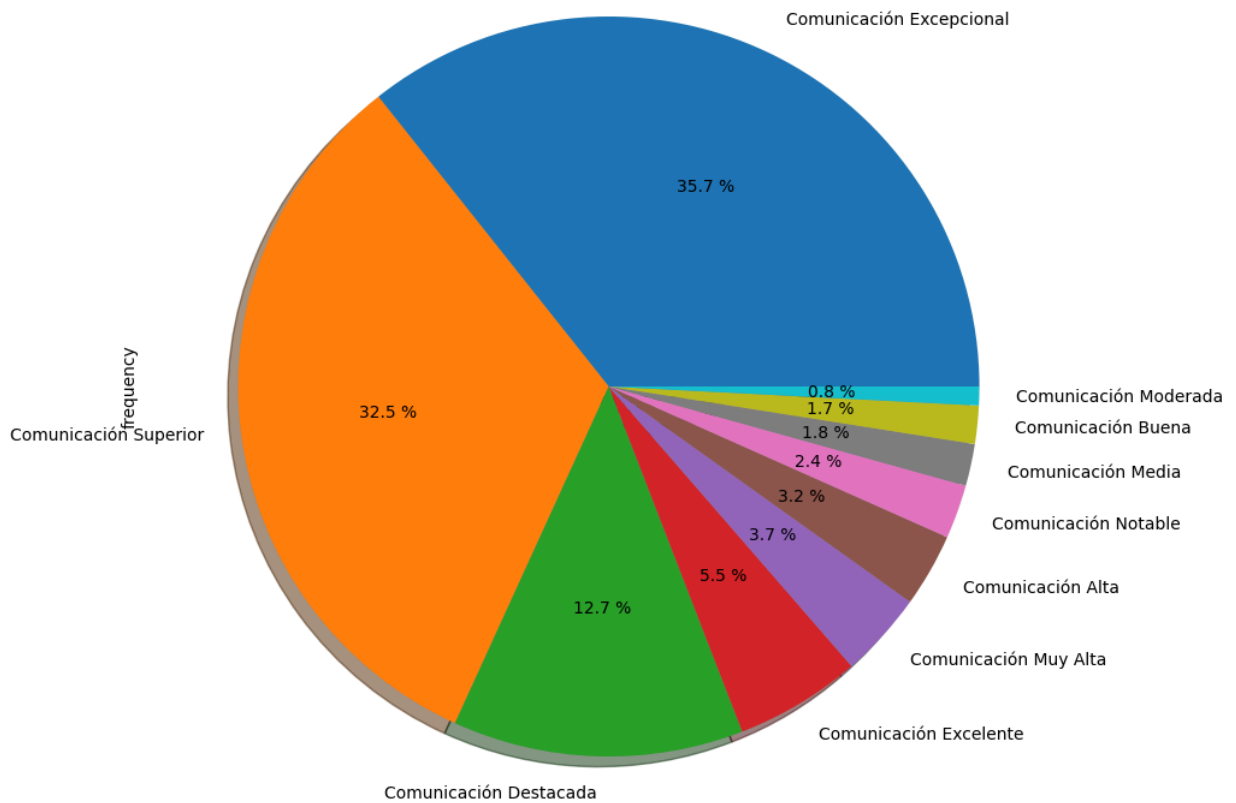
17. Nivel de check in:

La mayoría de los anfitriones de Airbnb reciben calificaciones positivas en el check-in, con más del 64% obteniendo una evaluación de Check-in Bueno o superior. Un 9.38% se destaca con un Check-in Muy Alto. Sin embargo, un pequeño porcentaje (1.83%) tiene un Bajo Check-in, lo que indica que algunos anfitriones enfrentan problemas en este aspecto. En general, el proceso de check-in es mayormente eficiente, pero hay oportunidades de mejora para garantizar una experiencia consistente para todos los huéspedes.



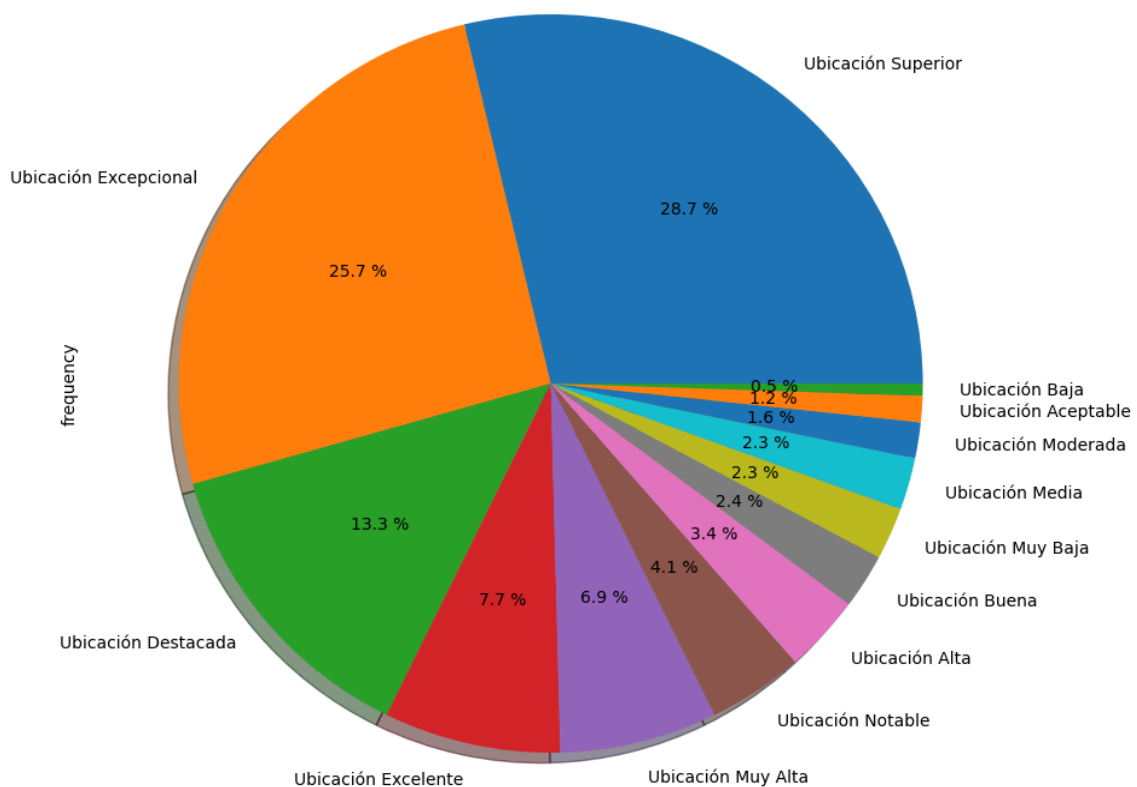
18. Nivel de comunicación:

La mayoría de los anfitriones tienen una comunicación excepcional o superior, con más del 68% de ellos recibiendo estas calificaciones. Un porcentaje más pequeño tiene comunicaciones destacadas o de calidad inferior, lo que sugiere que la comunicación con los huéspedes generalmente es muy buena en Airbnb, pero aún existen áreas para mejorar, especialmente en los casos con calificaciones más bajas.



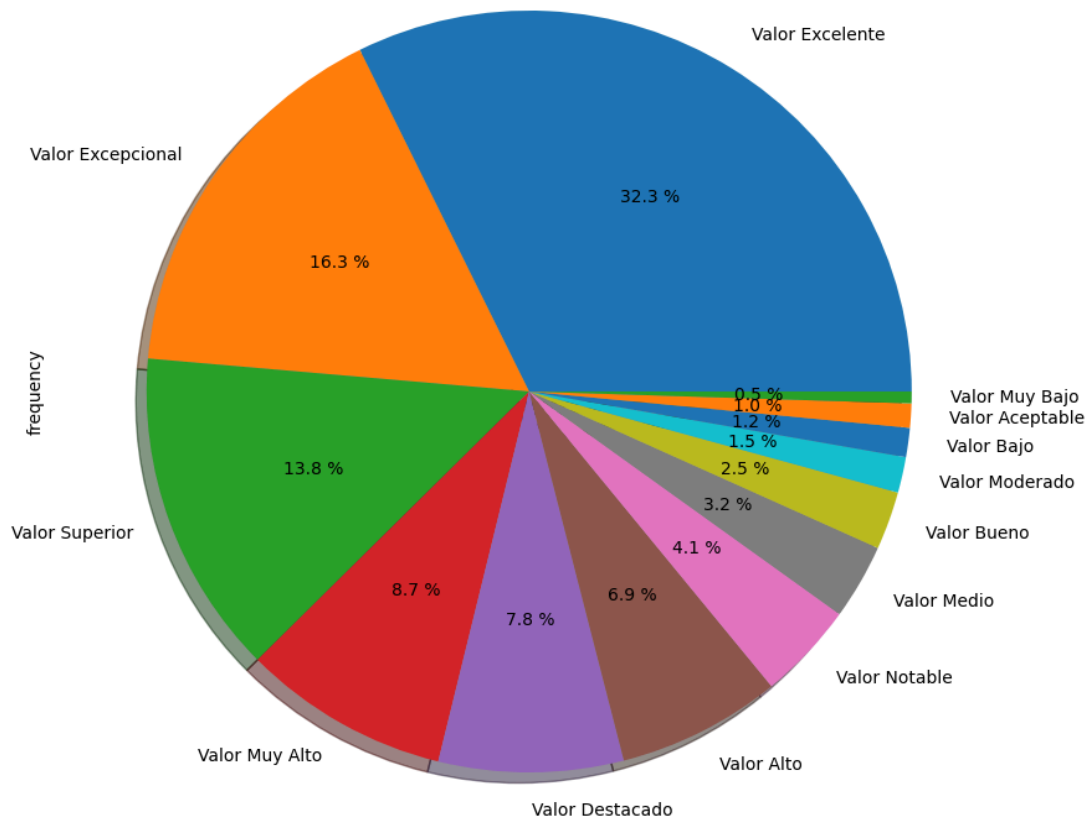
19. Nivel de ubicación:

La mayoría de los anfitriones reciben calificaciones positivas sobre la ubicación, con más del 54% obteniendo calificaciones de Ubicación Superior o Excepcional. Esto indica que muchas propiedades se encuentran en lugares altamente valorados por los huéspedes. Sin embargo, un pequeño porcentaje presenta calificaciones bajas o aceptables, lo que sugiere que aunque la mayoría de las ubicaciones son atractivas, hay algunas áreas que podrían mejorar en términos de ubicación para aumentar la satisfacción de los huéspedes.



20. Nivel de valor:

Finalmente en el nivel de valor que reciben los anfitriones 32.3% reciben una calificación excelente, un 16.3% valor excepcional, 13.8% valor superior y un 8.7% un valor muy alto, lo que podemos concluir que más del 50% de los anfitriones reciben calificaciones altas por parte del usuario respecto a que cumplen con todo lo anterior de limpieza, presión, ubicación etc. generando que la respuesta del usuario hacia el anfitrión sea positiva.



Conclusión:

La mayoría de las gráficas a utilizar fueron de barras y de pastel porque lo que necesitaban era comparar categorías y observar el comportamiento de preferencia de valores, por lo que, solamente una vez se utilizó de áreas porque necesitábamos ver la evolución del tiempo de respuesta conforme transcurrían las horas.

Respecto a los datos, podemos concluir que los anfitriones tienen noción de lo que los usuarios buscan y siempre tratan de brindar el mejor servicio para la satisfacción del cliente.