Analítica de Datos y
Herramientas de Inteligencia Artificial

Reporte de actividad 1.1

Profesor**:** Alfredo García Suárez

Bernardo Quintana López | A01658064

Campus Puebla

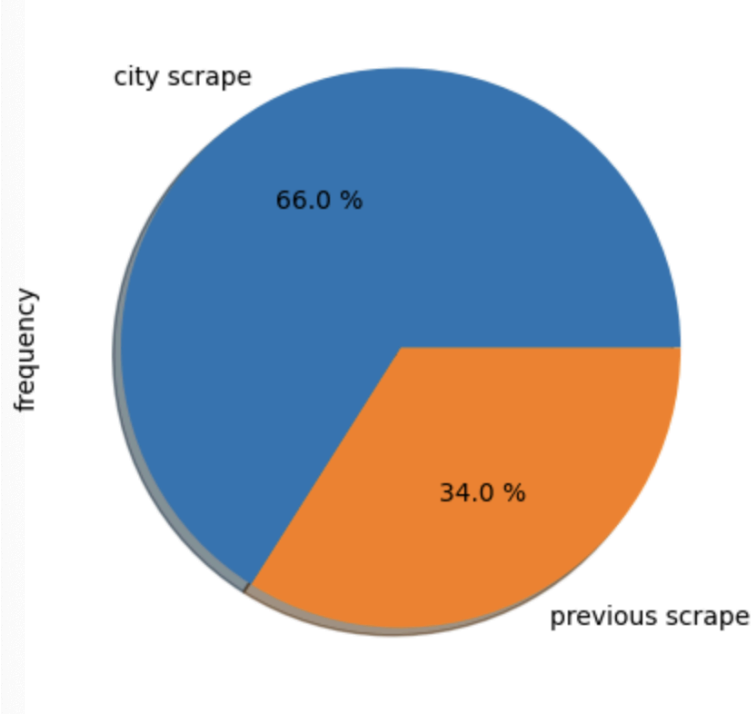
30 de marzo de 2025

Para comenzar, primero se instalaron las librerías necesarias para seguir con el código. Después, se cargó el dataframe de los registros de AirBnB de la ciudad de Londres, Inglaterra, el cual contiene 75 variables diferentes. Posteriormente, se procedió a evaluar la posible existencia de valores nulos. Al identificar que sí existían los mismos, se trataron los datos faltantes con diferentes técnicas como “strings” y el promedio de las variables, segmentando por variables cualitativas y cuantitativas.

De igual manera, se analizaron los valores atípicos mediante el uso de una BoxPlot para identificar las variables que presentaban outliers y así tratar los datos con el método de los percentiles intercuartílicos.

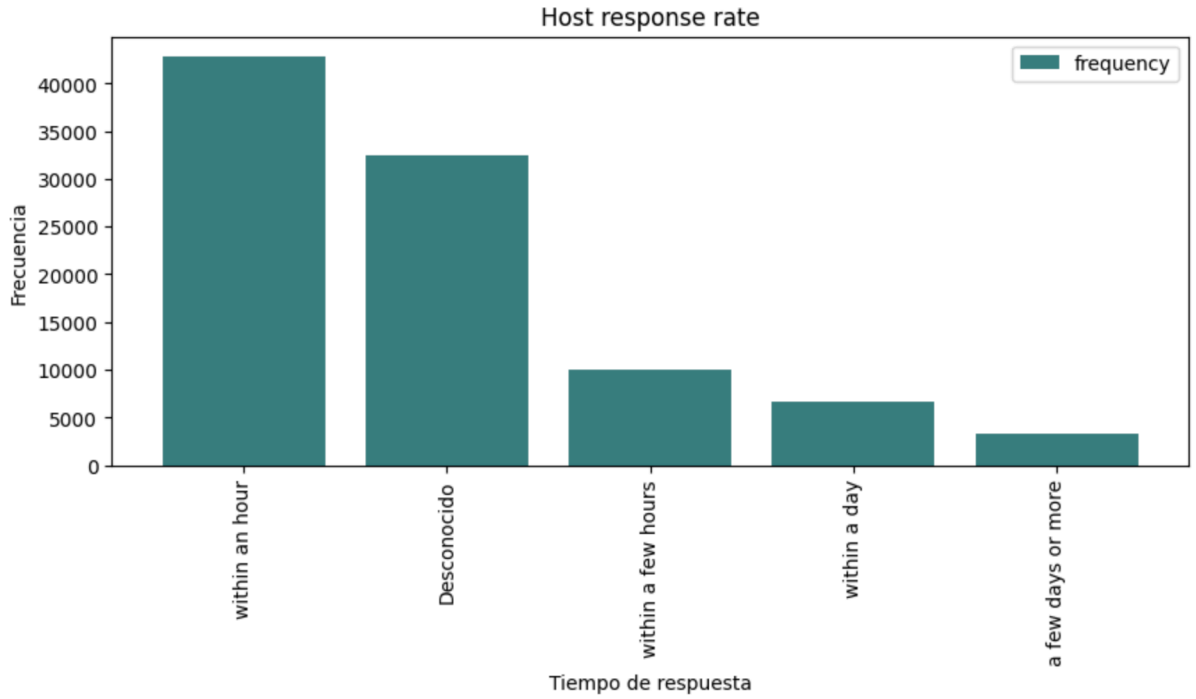
Después de haber realizado la limpia del dataframe final, se utilizó el mismo para realizar un análisis univariado de 15 variables categóricas con la ayuda del código freq_tbl(data). Se analizaron mediante diferentes gráficas, las siguientes variables: source, host_response_time, host_is_superhost, host_has_profile_pic, host_identity_verified, neighbourhood, room_type, has_availability, instant_bookable, bathrooms_text, host_location, host_name, property_type, neighbourhood_cleansed y host_neighbourhood.

Source



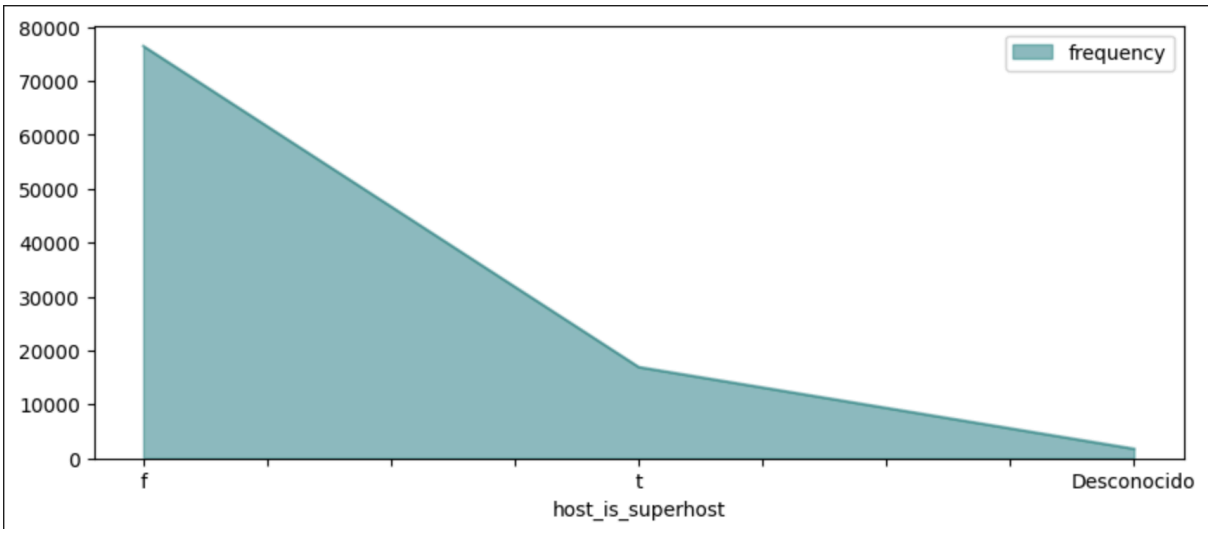
66% de los datos provienen de la última extracción (scrape), mientras que el 34% corresponde a una extracción anterior, lo que sugiere una actualización reciente de la mayoría de los datos

Host response rate



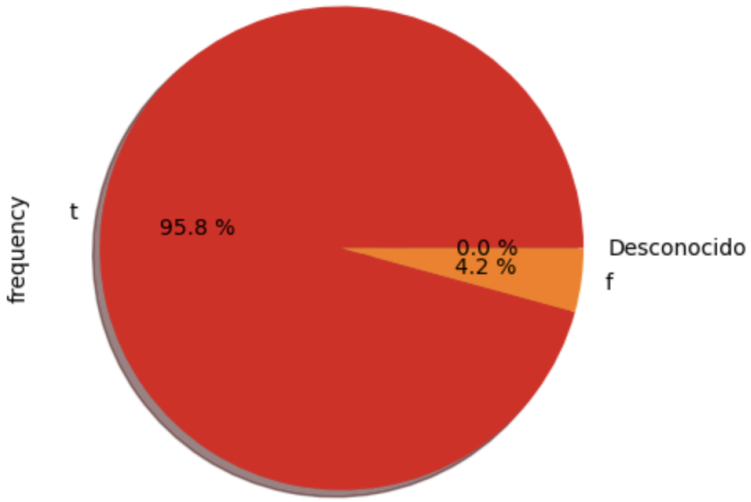
La mayoría de hosts responden dentro de una hora, algunas horas o un día, lo cual indica una buena comunicación propietario cliente y la atención brindada

Host is superhost



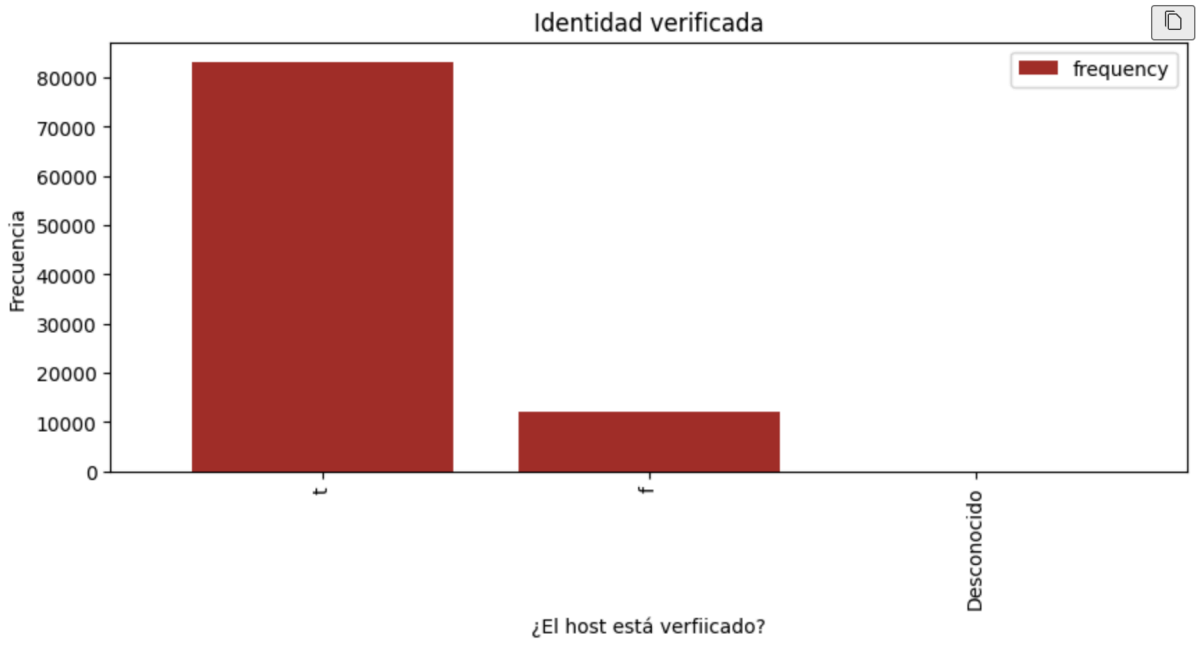
La mayoría de hosts no son considerados super hosts

Host has profile pic



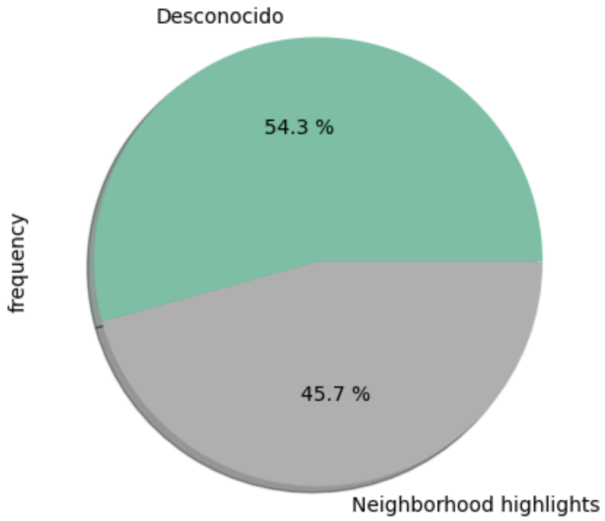
El 95.8% de los hosts tiene foto de perfil, mientras que un 4.2% carece de ella o no se ha registrado, mostrando una alta adopción de este elemento

Identidad verificada



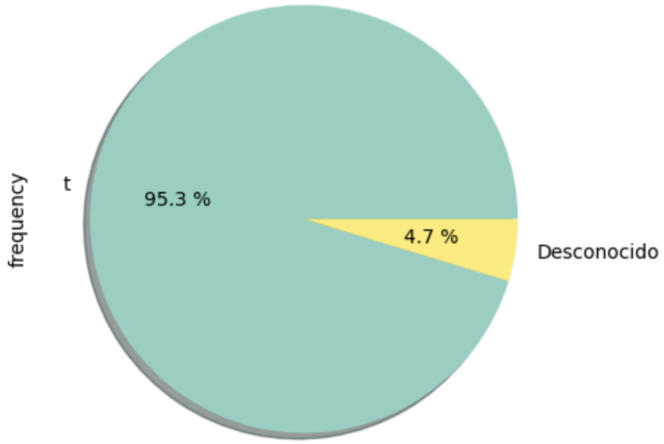
La mayoría de hosts están verificados, lo cual puede indicar una mayor confianza de parte del cliente interesado en rentar.

Neighbourhood



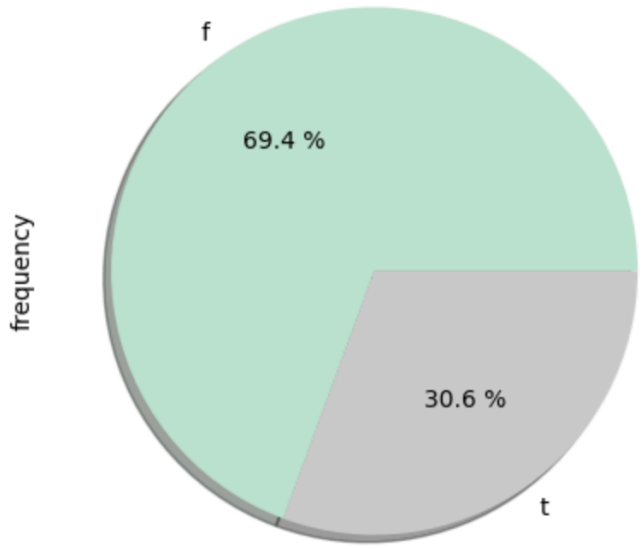
El 54.3% de los datos de barrios son desconocidos, mientras que el 45.7% está registrado, revelando una brecha significativa en la información geográfica

Has availability



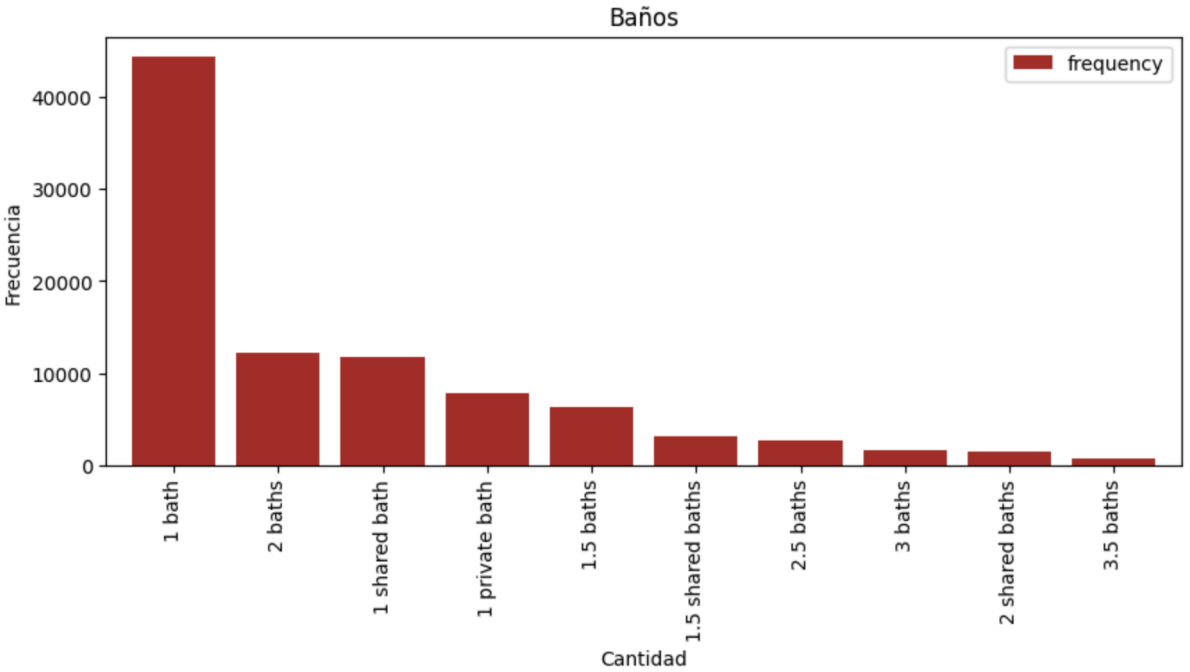
El 95.3% de las propiedades están disponibles, y solo un 4.7% no tiene datos, lo que sugiere una alta tasa de disponibilidad

Instant bookable



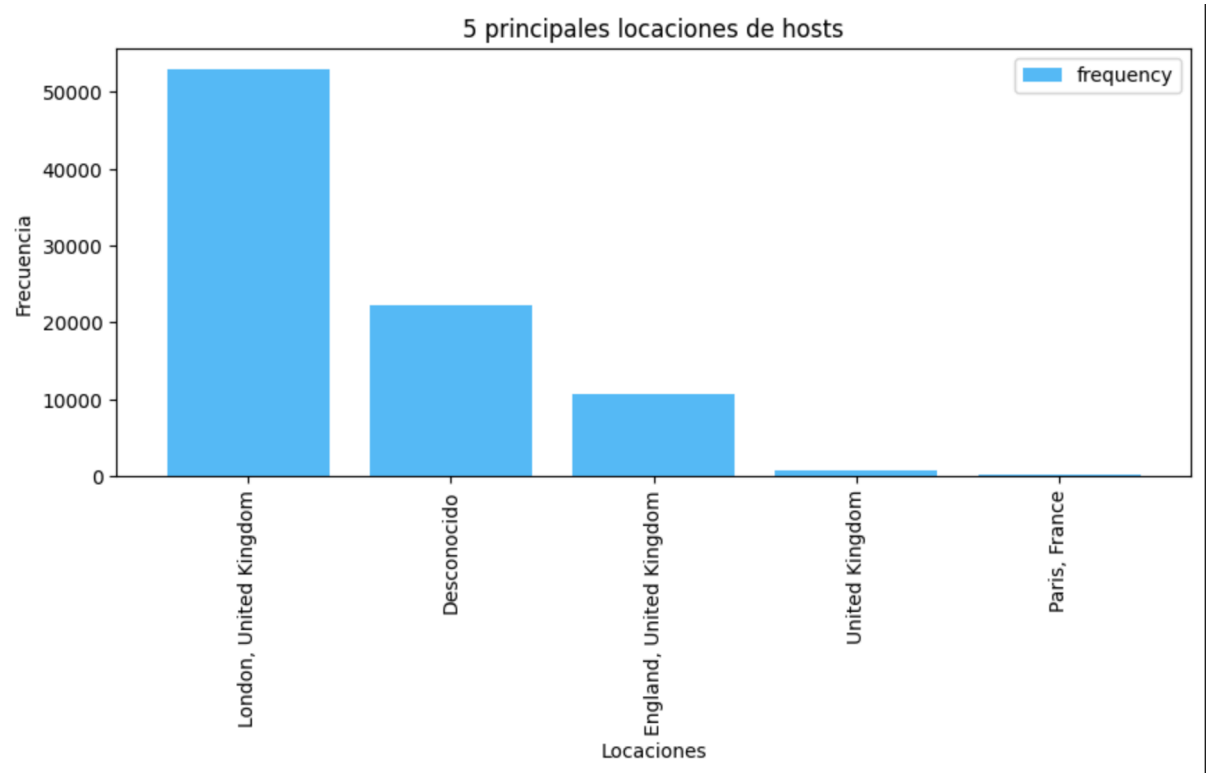
El 30.6% de las propiedades permiten reserva instantánea, frente a un 69.4% que no

Baños



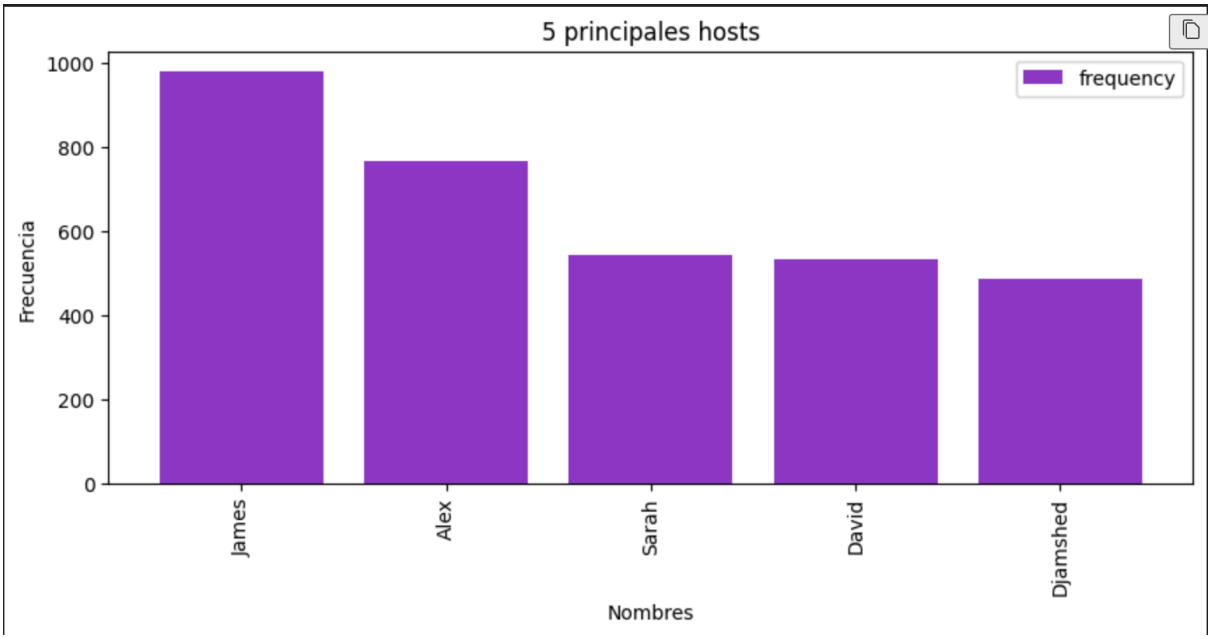
La mayoría de propiedades ofertadas cuentan con 1 baño, 2 baños o 1 baño compartido.

Locación de hosts



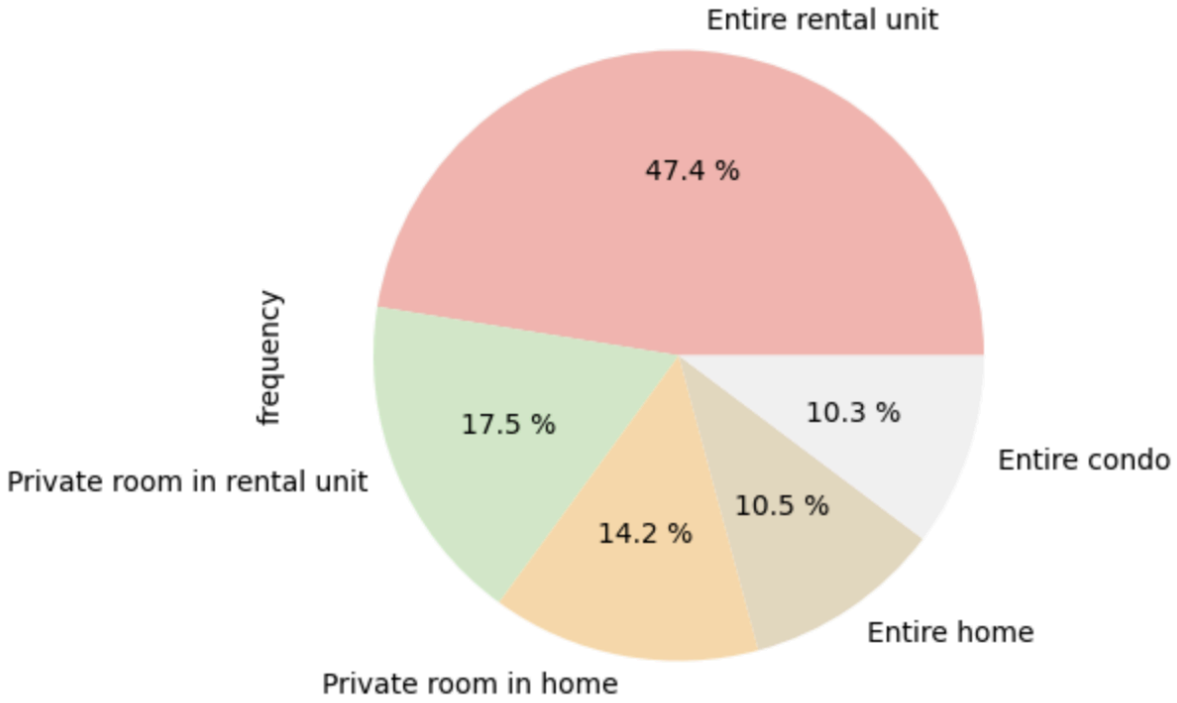
La mayoría de hosts se encuentran en el Reino Unido, mayormente en Londres.

Host name



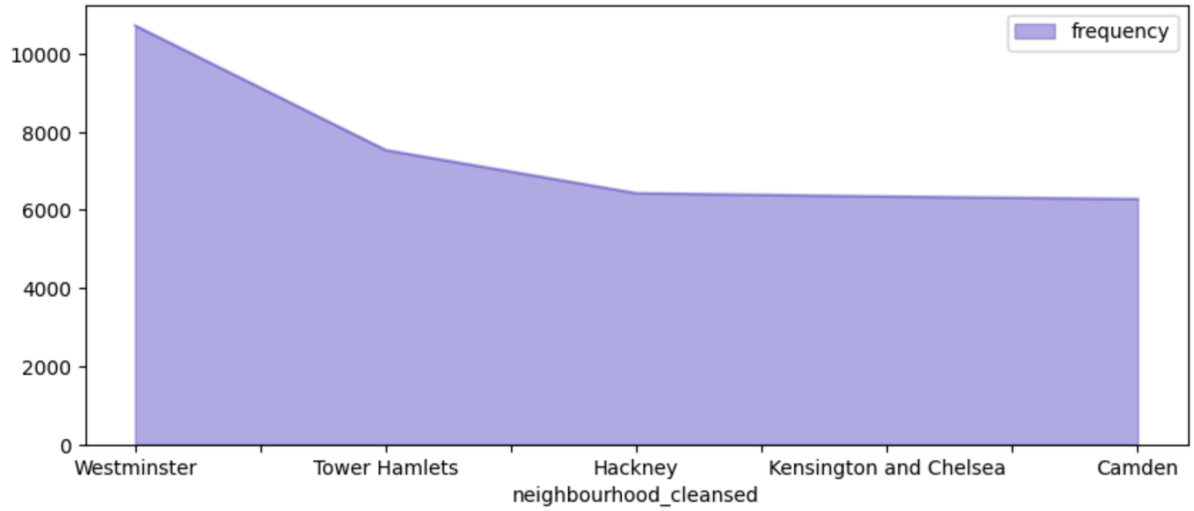
Entre los hosts que más ofrecen sus propiedades están James, Alex y Sarah

Property type



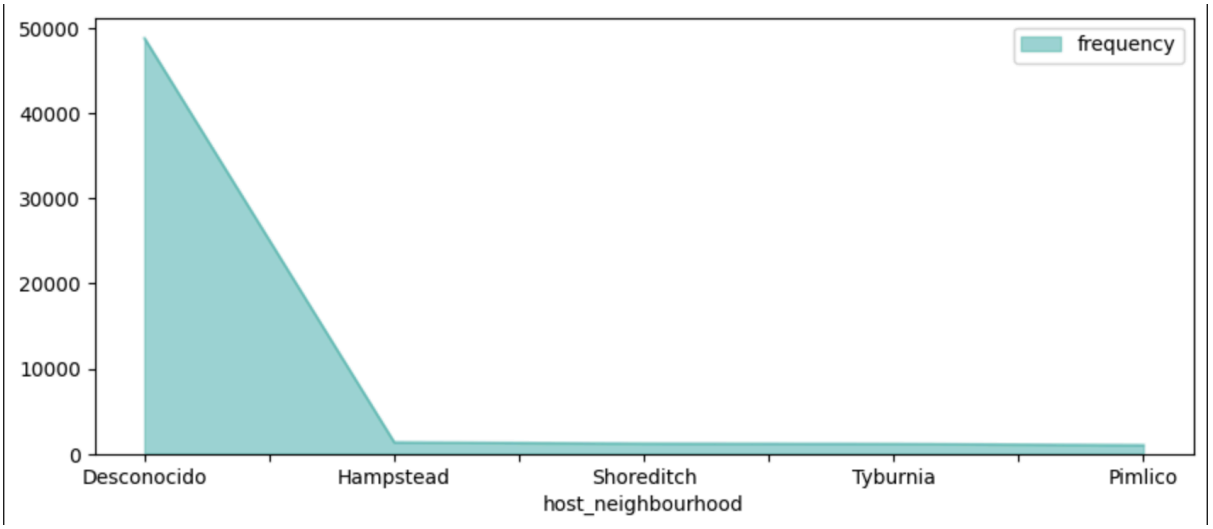
"Entire rental unit" es el tipo más común (47.4%), seguido de "Private room in rental unit" (17.5%), reflejando una tendencia hacia alojamientos completos

Neighbourhood cleansed



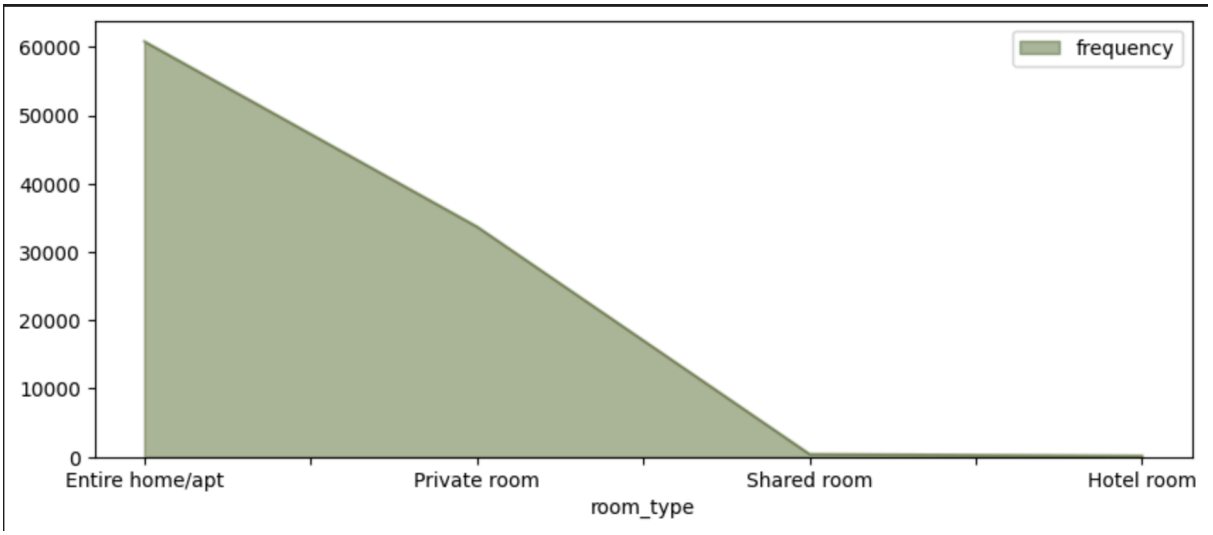
La mayoría de propiedades tienen registros en Westminster, Tower Hamlets y Hackney

Host Neighbourhood



La mayoría de hosts se encuentran en los barrios de Hampstead, Shoreditch y Tyburnia

Room type



Las casas y apartamentos completos son el tipo de propiedad más ofrecido en los registros, seguido de las recámaras personales y compartidas. Las habitaciones de hotel son las menos ofrecidas

Estas gráficas se realizaron mediante el filtrado de las frecuencias de cada variable con la ayuda de un código.

```
table3 = freq_tbl(data['host_response_time'])
table3
```

[19] ✓ 0.0s

	host_response_time	frequency	percentage	cumulative_perc
0	within an hour	42781	0.449645	0.449645
1	Desconocido	32435	0.340904	0.790549
2	within a few hours	9980	0.104894	0.895443
3	within a day	6635	0.069736	0.965179
4	a few days or more	3313	0.034821	1.000000

```
table4 = table3.drop(['percentage', 'cumulative_perc'], axis=1)
table4
```

[20] ✓ 0.0s

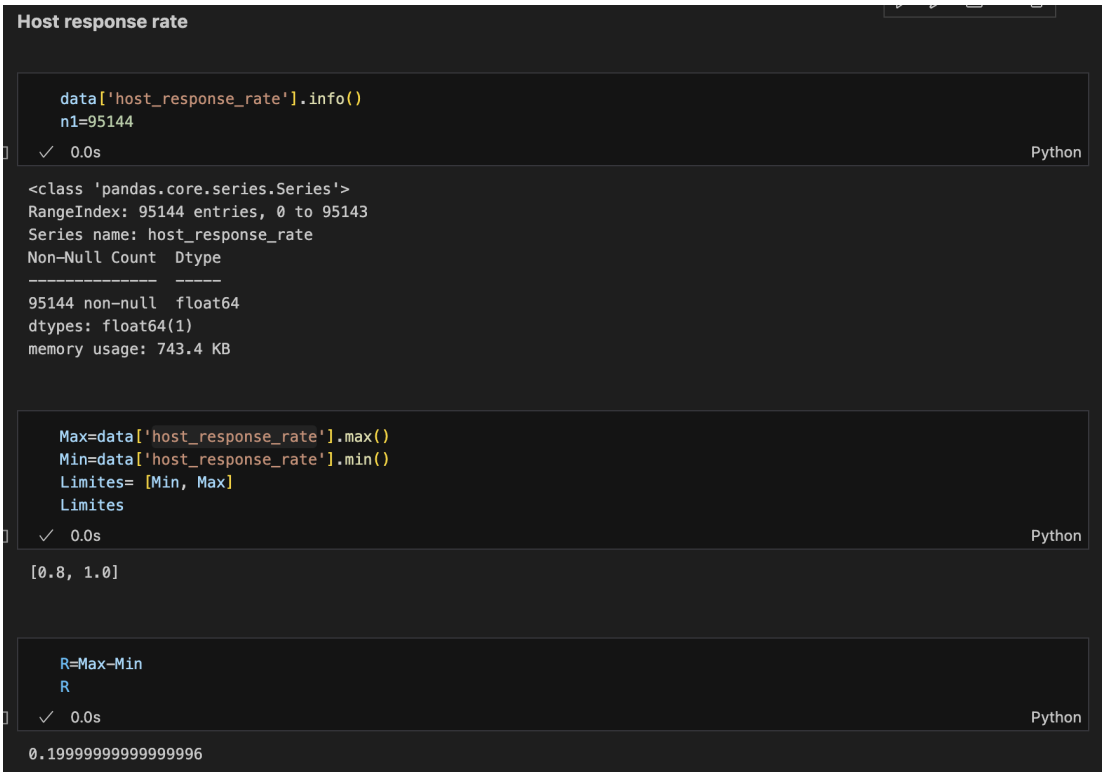
	host_response_time	frequency
0	within an hour	42781
1	Desconocido	32435
2	within a few hours	9980
3	within a day	6635
4	a few days or more	3313

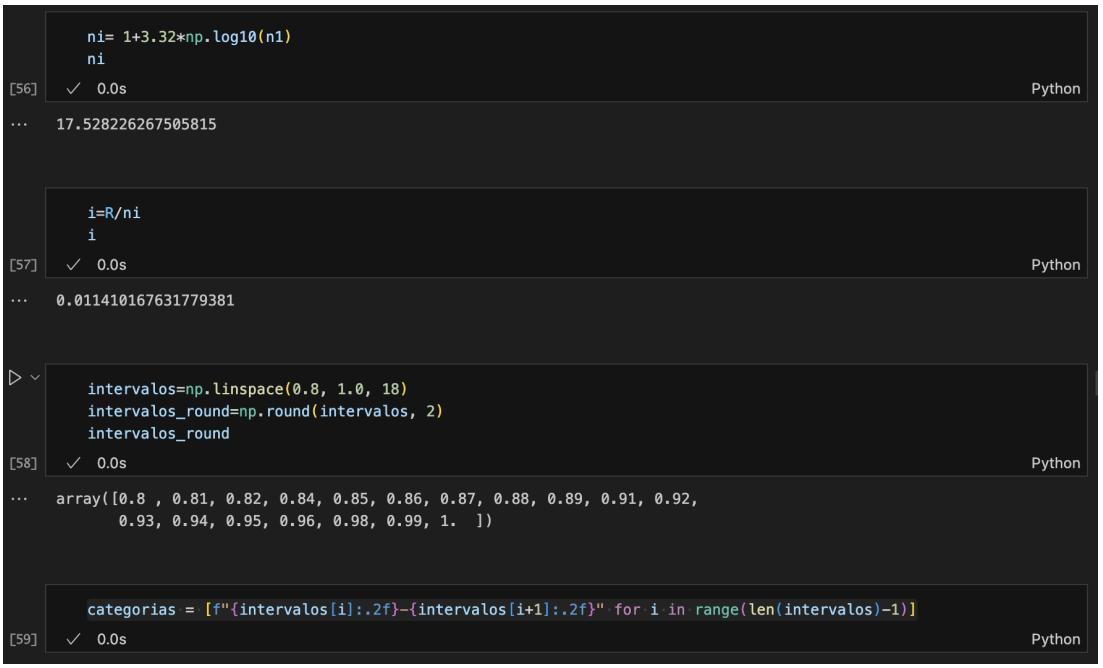
```
Filtro_index1 = table4.set_index('host_response_time')
Filtro_index1
```

[21] ✓ 0.0s

host_response_time	frequency
within an hour	42781
Desconocido	32435
within a few hours	9980
within a day	6635
a few days or more	3313

De la misma manera, se hizo la categorización de las variables requeridas: ” host_response_rate”, “host_acceptance_rate”, “host_total_listings_count”, “accommodates”, “bathrooms_text”, “beds”, “price”, “maximum_nights_avg_ntm”, “availability_365”, “number_of_reviews”, “review_scores_value”, “reviews_per_month”, mediante la regla de Sturges. Se muestra un ejemplo con la variable host acceptance rate a continuación:

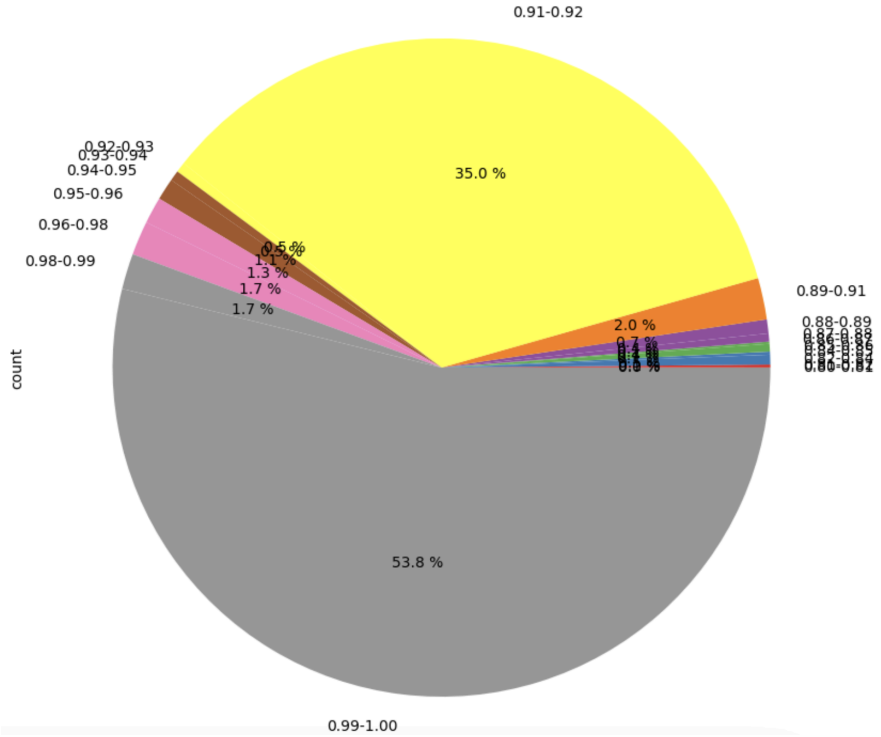




```
[60]  data['host_response_rate']=pd.cut(x= data['host_response_rate'], bins=intervalos_round, labels=categorias)
      data['host_response_rate']
      ✓ 0.0s                                                                                    Python

...   0        0.98-0.99
      1        0.99-1.00
      2        0.98-0.99
      3        0.98-0.99
      4        0.91-0.92
               ...
      95139    0.91-0.92
      95140    0.91-0.92
      95141    0.99-1.00
      95142    0.91-0.92
      95143    0.91-0.92
      Name: host_response_rate, Length: 95144, dtype: category
      Categories (17, object): ['0.80-0.81' < '0.81-0.82' < '0.82-0.84' < '0.84-0.85' ... '0.95-0.96' < '0.96-0.98' < '0.9

[61]  conteo_categorias = data['host_response_rate'].value_counts().sort_index()
      ✓ 0.0s                                                                                    Python

[62]  conteo_categorias.plot(kind='pie', figsize=(15,10), shadow=False, autopct='%0.1f %%', colormap='Set1')
      ✓ 0.0s                                                                                    Python

...   <Axes: ylabel='count'>
```

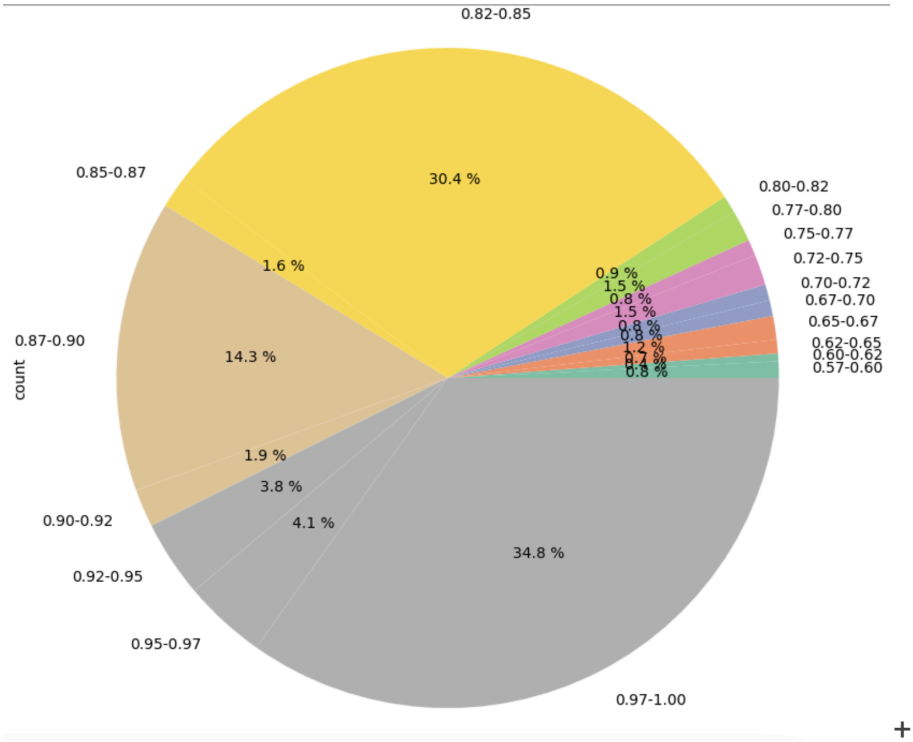
Después de aplicar este código a todas las variables mencionadas, se obtuvieron las siguientes gráficas

Host response rate



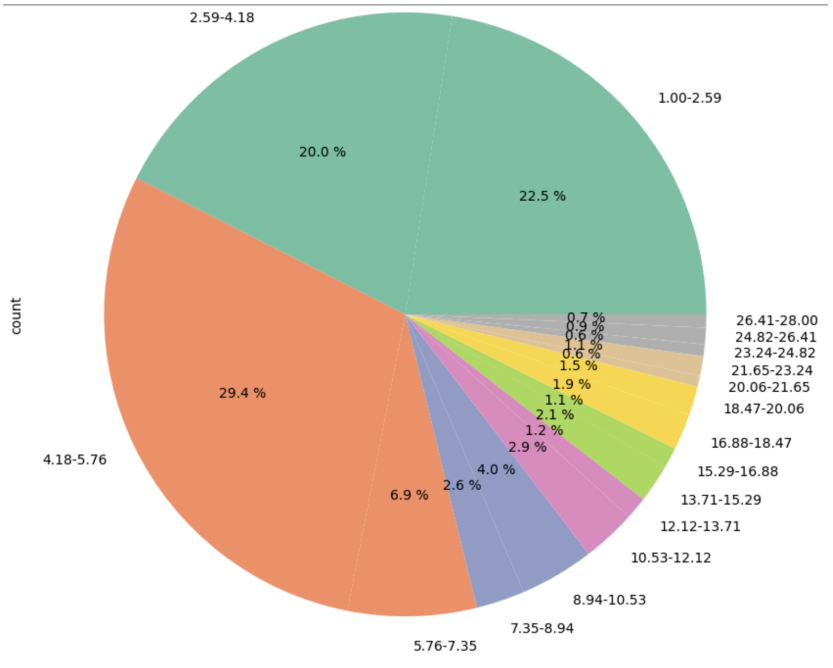
La mayoría de registros están en el rango de 0.99 a 1.

Host acceptance rate



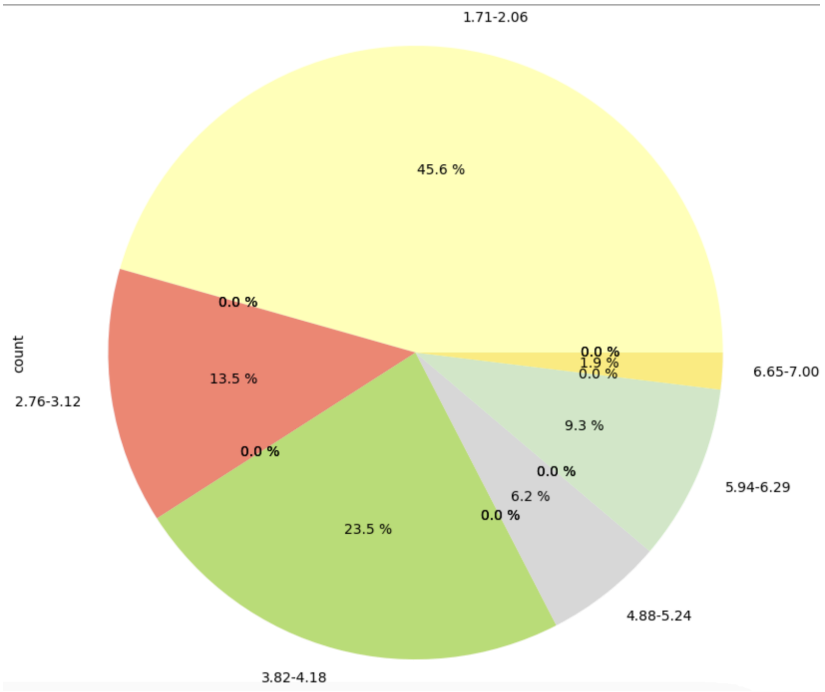
La mayoría de registros se encuentran en el rango 0.97-1.00

Host total listings count



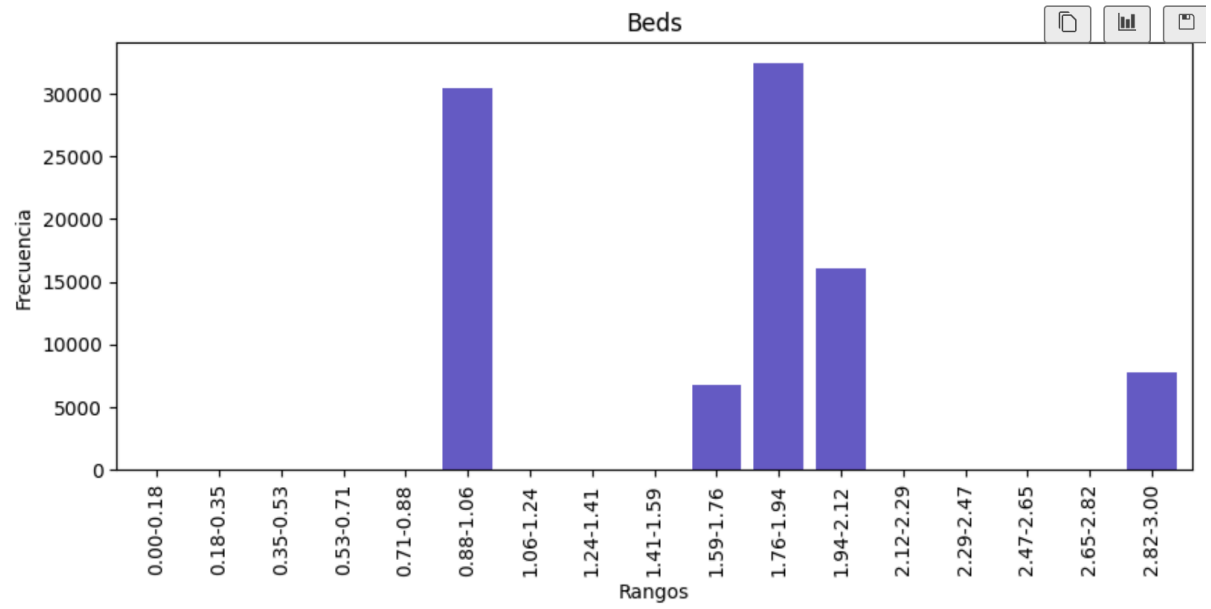
La mayoría de registros se encuentran en el rango 4.18-5.76

Accommodates



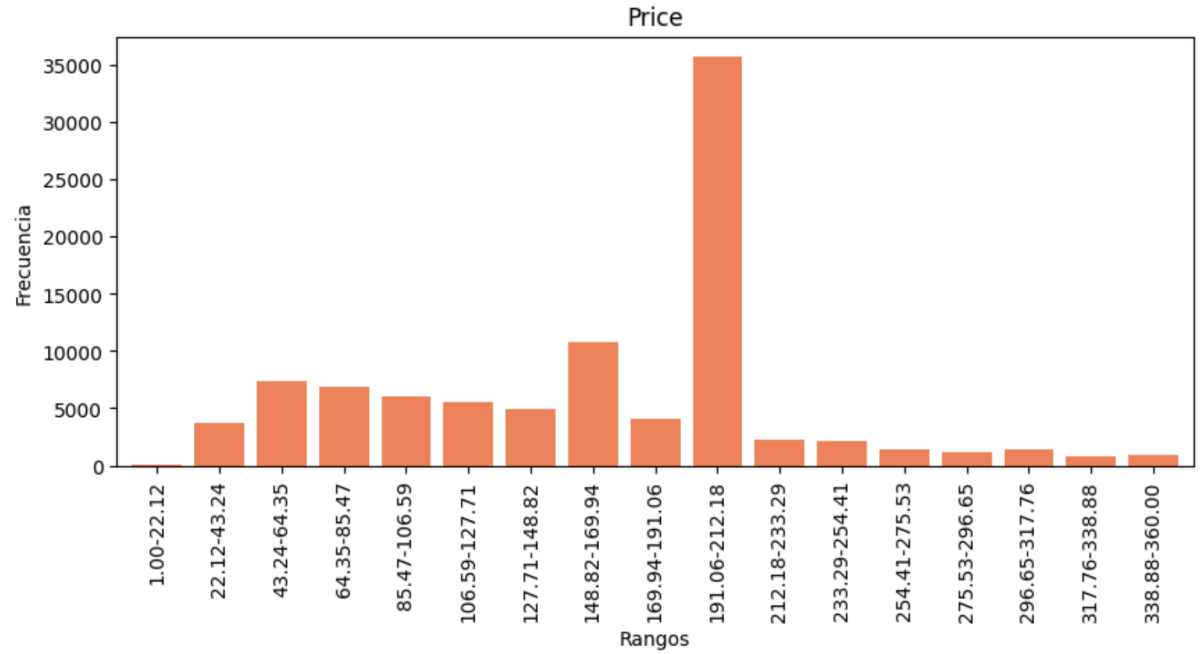
La mayoría de registros se encuentran en el rango 1.71-2.06

Beds



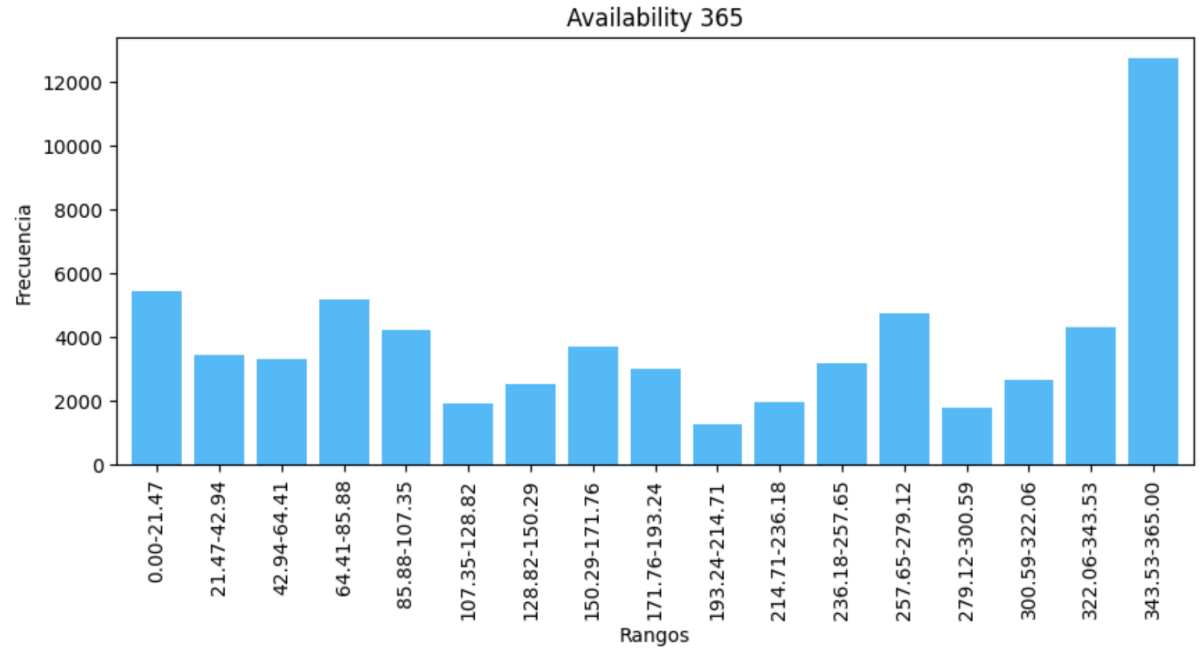
La mayoría de propiedades ofrecen dos camas aproximadamente

Price



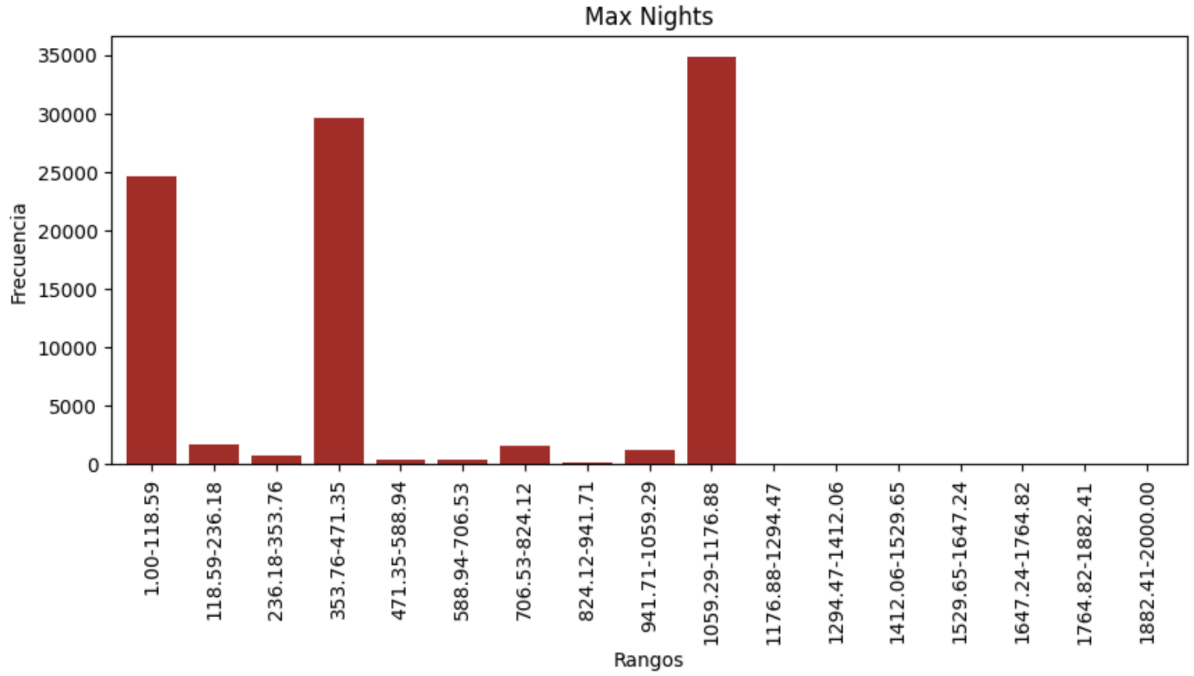
La mayoría de propiedades oscilan entre un precio de 191.06 y 212.18 dólares la noche

Availability 365



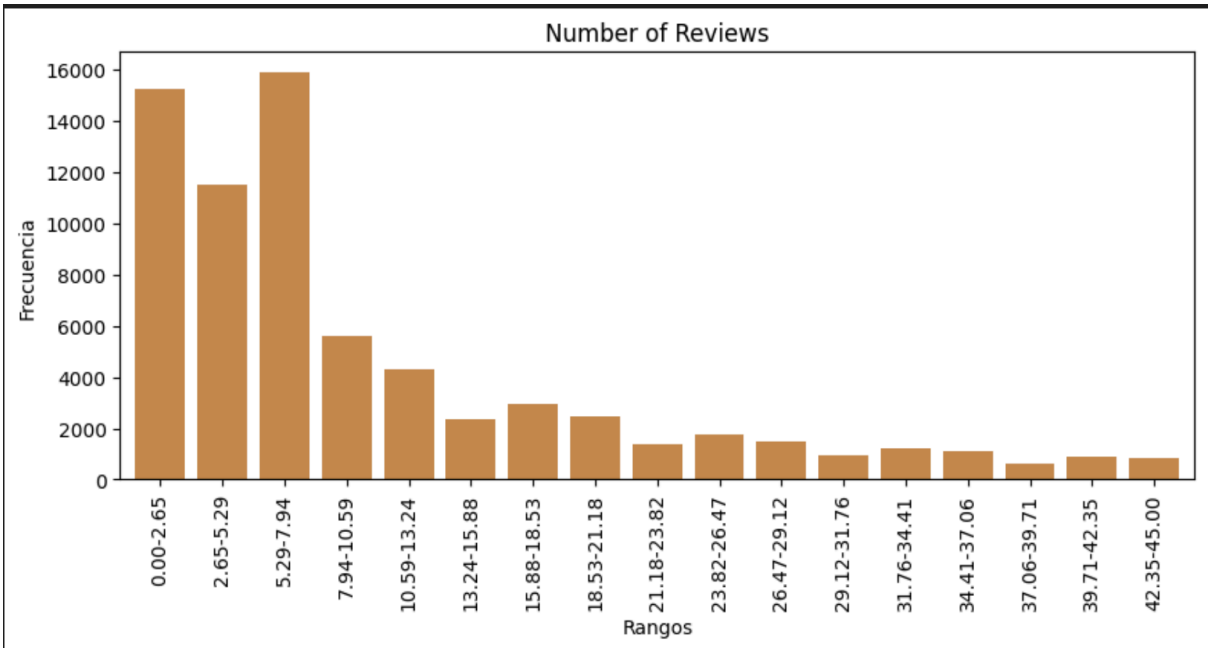
La mayoría de propiedades se encuentran disponibles a lo largo del año, entre 343 y 365 días

Noches máximas



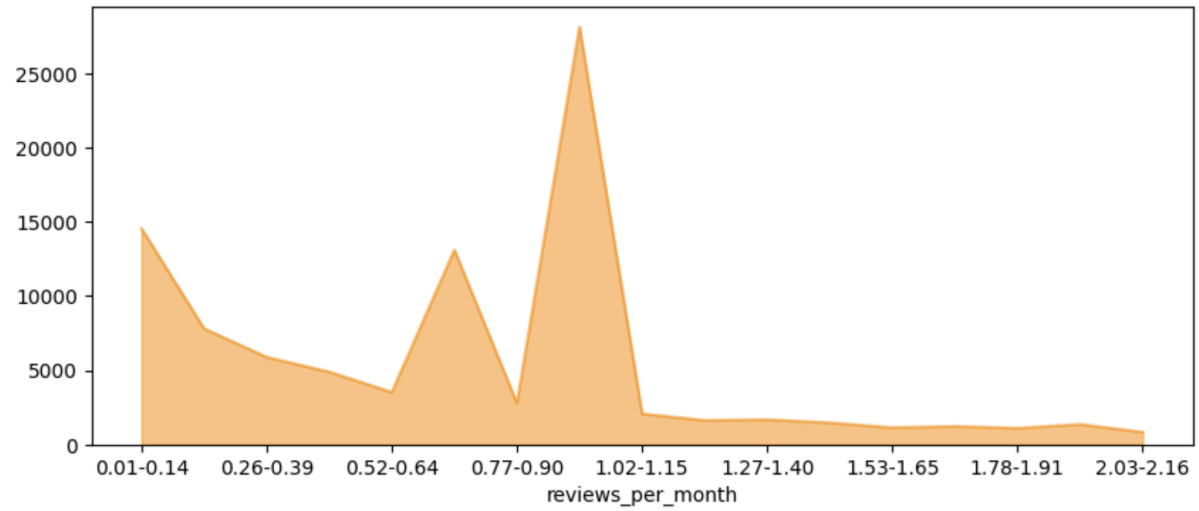
La mayoría de propiedades permiten un máximo de noches entre 1059 y 1176

Número de reseñas



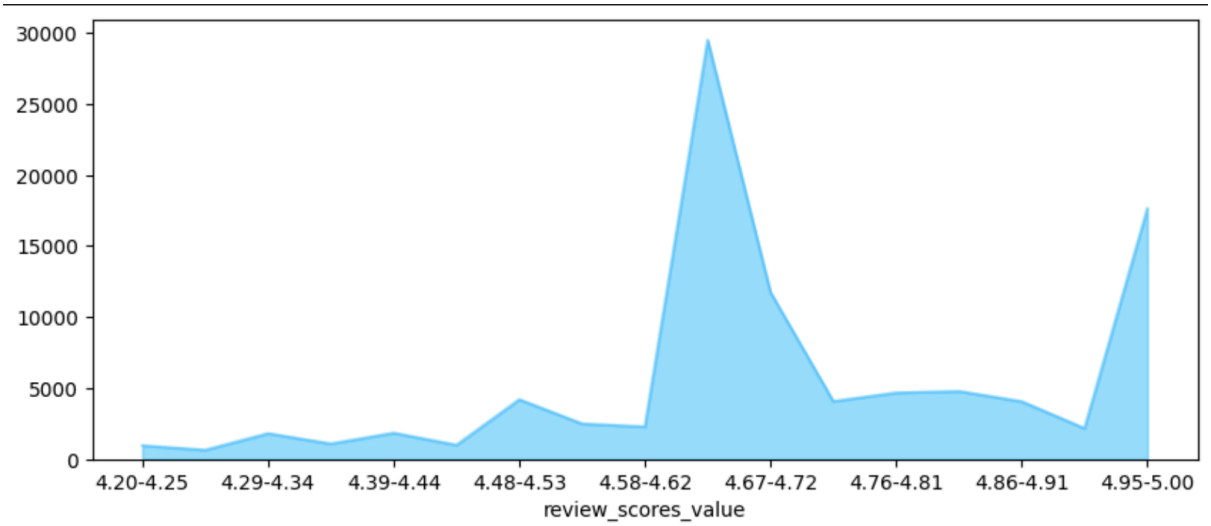
La mayoría de propiedades tienen entre 5 y 8 reseñas

Reviews per month



Las propiedades cuentan con al menos una reseña por mes

Review scores value



El score más común en esta categoría ronda entre 4.62 y 4.67 puntos

El análisis del dataset de Airbnb permitió identificar varios aspectos clave sobre las propiedades en renta. En primer lugar, se detectó la necesidad de manejar valores nulos y outliers para garantizar la calidad de los datos. Las visualizaciones revelaron que la mayoría de las propiedades tienen reseñas positivas y que los precios suelen concentrarse en un rango moderado.

Además, la categorización de variables como el precio y el tipo de habitación facilitó la identificación de tendencias en el mercado. Por ejemplo, la predominancia de

alojamientos completos sugiere que los anfitriones prefieren ofrecer mayor privacidad a los huéspedes.

Como recomendaciones futuras, se podría profundizar en el análisis de las propiedades más valoradas y explorar la relación entre ubicación geográfica y precios. Estos insights podrían ser útiles para anfitriones que buscan optimizar sus listados y para plataformas como Airbnb que desean mejorar la experiencia del usuario.

En resumen, este reporte proporciona una base sólida para entender el comportamiento del mercado de Airbnb y ofrece oportunidades para futuros análisis más detallados.