

Resultados

airbnb

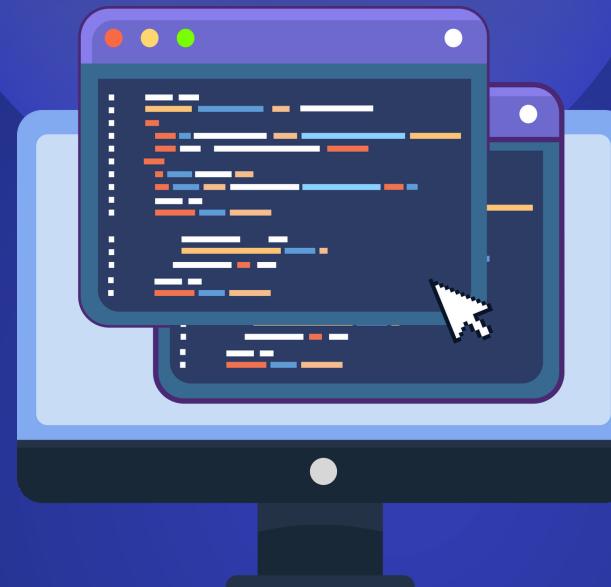
Mizuki Aranzazu Uscanga Pineda A01737787
Osvaldo Terrazas Sánchez A01276389
Jose Manuel Morales Escalante A01737201
Ismael Durán Martínez A01731393



METODOLOGÍA

01

Acciones de
preprocesamiento



02

Selección de las
variables
seleccionadas



03

Análisis
Univariado



04

Análisis
Comparativo



RIO DE JANERIO



CIUDADES

34664 rows ×
75 columns

(26924 rows,
75 columns)

MADRID

PREPROCESAMIENTO

ELIMINACIÓN NULOS: FUNCIÓN ISNULL()



MADRID

description	966
neighborhood_overview	13618
picture_url	2
host_name	4
host_since	4
host_location	8416
host_about	13945
host_response_time	4915

38 COLUMNAS



RIO DE JANEIRO

description	709
neighborhood_overview	14403
host_name	1
host_location	5323
host_about	13624
host_response_time	4418
host_response_rate	4418
host_acceptance_rate	2889
host_is_superhost	614
host_neighbourhood	2482

53 COLUMNAS

```
3 require File.expand_path('../..', __FILE__)
4 # Prevent database truncation if the migration
5 # abort("The Rails environment is running in production mode!")
6 require 'spec_helper'
7 require 'rspec/rails'
8
9 require 'capybara/rspec'
10 require 'capybara/rails'
11
12 Capybara.javascript_driver = :webkit
13 Category.delete_all; Category.create!
14 Shoulda::Matchers.configure do |config|
15   config.integrate do |with|
16     with.test_framework :rspec
17     with.library :rails
18   end
19 end
20
21 # Add additional requires below this line
22
23 # Requires supporting files within the same directory as
24 # spec/support/ and its subdirectories
25 # run as spec files by default. This means you can
26 # in _spec.rb will both be required and
27 # run twice. It is recommended that you
28 # end with _spec.rb. You can
29 # specify the :type option to
30 # action on the second run.
31
32 # Action on the second run
33
34 # No results found for 'mongoid'
35
36 Mongoid
37
38 # buffer
```

```
3   require File.expand_path("../../config/environment", __FILE__)
4   # Prevent database truncation if the test fails
5   abort("The Rails environment is running in production mode")
6   require 'spec_helper'
7   require 'rspec/rails'
8
9   require 'capybara/rspec'
10  require 'capybara/rails'
11
12  Capybara.javascript_driver = :webkit
13  Category.delete_all; Category.create!(name: "Test Category")
14  Shoulda::Matchers.configure do |config|
15    config.integrate do |with|
16      with.test_framework :rspec
17      with.library :rails
18    end
19  end
20
21  # Add additional requires below this line
22
23  # Requires supporting ruby files with
24  # spec/support/ and its subdirectories
25  # run as spec files by default. You can
26  # in _spec.rb will both be required.
27  # run twice. It is recommended that
28  # end with _spec.rb. You can switch
29  # option on the command line to
30
31  # No results found for 'mongoid'
```

PREPROCESAMIENTO

ELIMINACIÓN NULOS: FUNCIÓN DTYPE

```
TIPOS_COLUMNAS = RIO_DE_JANEIRO[COLUMNAS_CON_NULOS.INDEX].DTYPES
COLUMNAS_OBJECT = TIPOS_COLUMNAS[TIPOS_COLUMNAS == 'OBJECT']
COLUMNAS_OBJECT
```



24/34 COLUMN

38/56 COLUMNAS

PREPROCESAMIENTO

ELIMINACIÓN NULOS: FUNCIÓN DTYPES

```
TIPOS_COLUMNAS =  
RIO_DE_JANEIRO[COLUMNAS_CON_NULOS.INDEX].DTYPES  
COLUMNAS_OBJECT = TIPOS_COLUMNAS[TIPOS_COLUMNAS == 'FLOAT']  
COLUMNAS_OBJECT
```

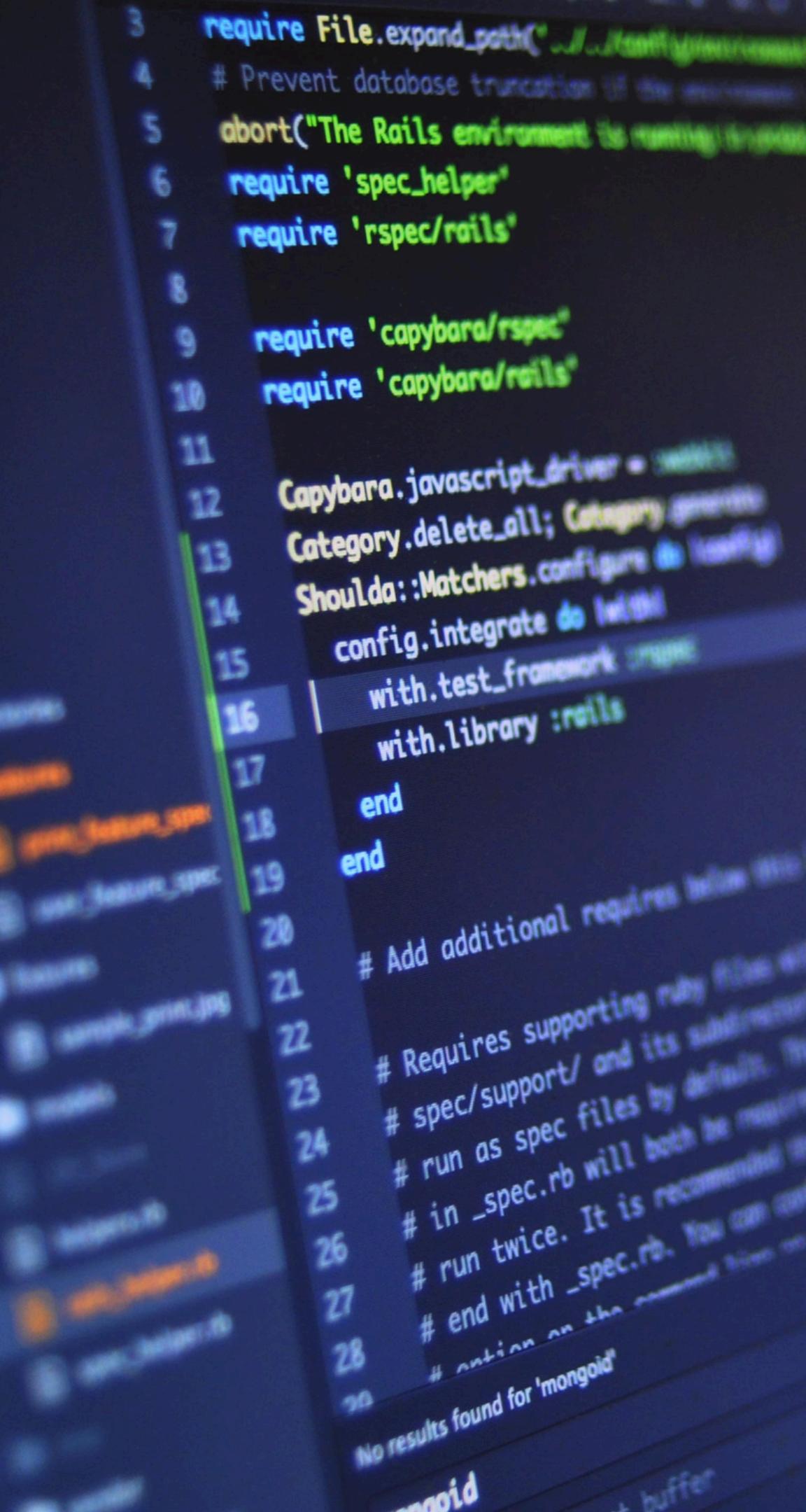


MADRID	
host_listings_count	float64
host_total_listings_count	float64
bathrooms	float64
bedrooms	float64
beds	float64
calendar_updated	float64

RIO DE JANEIRO	
beds	float64
minimum_nights	float64
maximum_maximum_nights	float64
maximum_nights_avg_ntm	float64
number_of_reviews_l30d	float64
review_scores_accuracy	float64

14/34 COLUMNAS

15/53 COLUMNAS



```
3 require File.expand_path('../..', __FILE__)  
4 # Prevent database truncation if the environment is test  
5 abort("The Rails environment is running in production mode!  
6 require 'spec_helper'  
7 require 'rspec/rails'  
8  
9 require 'capybara/rspec'  
10 require 'capybara/rails'  
11  
12 Capybara.javascript_driver = :webkit  
13 Category.delete_all; Category.create  
14 Shoulda::Matchers.configure do |config|  
15 config.integrate do |with|  
16   with.test_framework :rspec  
17   with.library :rails  
18 end  
19  
20 # Add additional requires below this line  
21  
22 # Requires supporting files within the same directory as the spec file  
23 # or specify a full path to the support file.  
24 # run as spec/support/ and its subdirectories  
25 # in _spec.rb will both be required.  
26 # run twice. It is recommended to add  
27 # end with _spec.rb. You can run  
28 # action on the command line to see what  
29 # no results found for 'mongoid'  
30  
# mongoid
```

MÉTODO DE SUSTITUCIÓN



- DESCRIPCIÓN FALTANTE
- SIN INFORMACIÓN
- ANONIMO
- VALIDACIÓN PENDIENTE
- INFORMACIÓN FALTANTE
- NO DEFINIDO
- INFORMACIÓN NO DISPONIBLE
- SIN RESEÑA



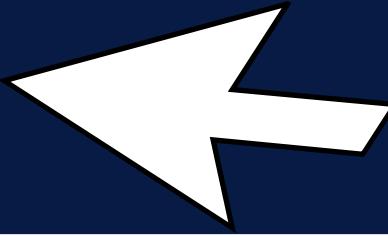
MINIMUM_NIGHTS
MAXIMUM_MAXIMUM_NIGHTS
MAXIMUM_NIGHTS_AVG_NT
NUMBER_OF_REVIEWS_L30D
REVIEW_SCORES_ACCURACY
REVIEW_SCORES_CHECKIN
REVIEW_SCORES_COMMUNICATION
REVIEW_SCORES_VALUE
CALCULATED_HOST_LISTINGS_COUNT
CALCULATED_HOST_LISTINGS_COUNT_ENTIRE_HOMES
CALCULATED_HOST_LISTINGS_COUNT_PRIVATE_ROOMS
CALCULATED_HOST_LISTINGS_COUNT_SHARED_ROOMS
REVIEWS_PER_MONTH



-

0

SELECCION DE VARIABLES



- LISTING_URL
- LAST_SCRAPED
- SOURCE
- NAME
- HOST_URL
- HOST_NAME
- HOST_SINCE
- HOST_LOCATION
- HOST_RESPONSE_TIME
- HOST_RESPONSE_RATE
- HOST_ACCEPTANCE_RATE
- HOST_IS_SUPERHOST
- HOST_NEIGHBOURHOOD
- HOST_VERIFICATIONS
- HOST_HAS_PROFILE_PIC
- HOST_IDENTITY_VERIFIED
- NEIGHBOURHOOD_CLEANSED
- PROPERTY_TYPE
- ROOM_TYPE
- ACCOMMODATES
- BATHROOMS_TEXT
- BEDROOMS
- BEDS
- AMENITIES
- PRICE
- HAS_AVAILABILITY
- NUMBER_OF_REVIEWS
- REVIEW_SCORES_RATING
- INSTANT_BOOKABLE
- CALCULATED_HOST_LISTINGS_CO
UNT
- REVIEWS_PER_MONTH

PREPROCESAMIENTO

ELIMINACIÓN OUTLIERS: CONOCER LOS TIPOS DE DATOS CON LA FUNCIÓN DTYPES()

listing_url	object
last_scraped	object
source	object
name	object
host_url	object
host_name	object
host_since	object
host_location	object
host_response_time	object
host_response_rate	object
host_acceptance_rate	object
host_is_superhost	object
host_neighbourhood	object
host_verifications	object
host_has_profile_pic	object
host_identity_verified	object
neighbourhood_cleansed	object
property_type	object
room_type	object

accommodates	int64
bathrooms_text	object
bedrooms	object
beds	object
amenities	object
price	object
has_availability	object
number_of_reviews	int64
review_scores_rating	float64
instant_bookable	object
calculated_host_listings_count	int64
reviews_per_month	float64

SEPARAR LOS DATOS EN COLUMNAS
CUANTITATIVAS Y CUALITATIVAS
CON LA FUNCION

SELECT_DTYPES(include == 'OBJECT')

SELECT_DTYPES(include == ('INT64', 'FLOAT64'))

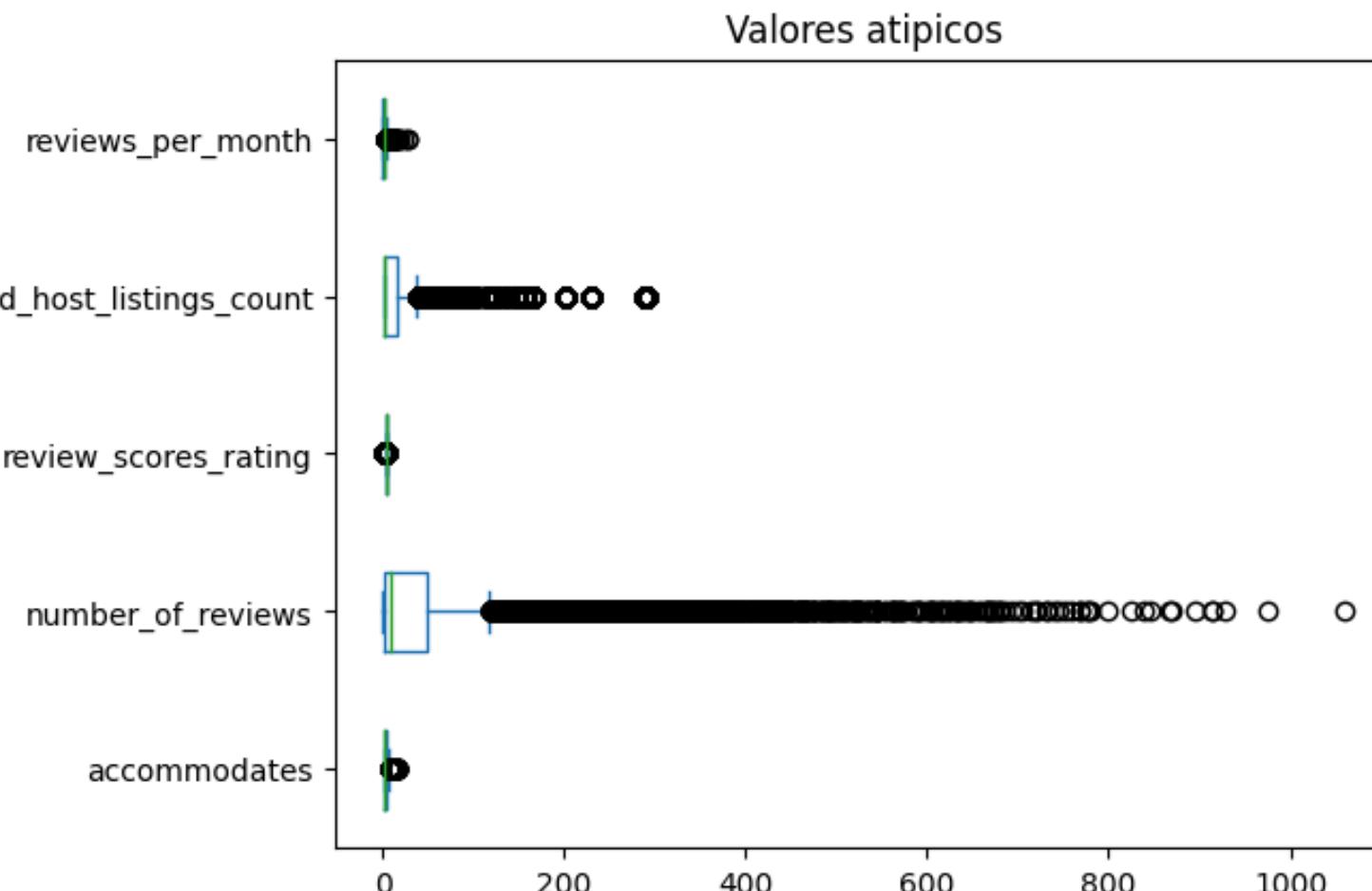
REALIZAR BOXPLOTS



MADRID

```
fig = plt.figure(figsize = (20, 10))
columnas_cuantitativas.plot(kind = 'box', vert = False)
plt.title("Valores atípicos")
plt.show()
```

<Figure size 2000x1000 with 0 Axes>

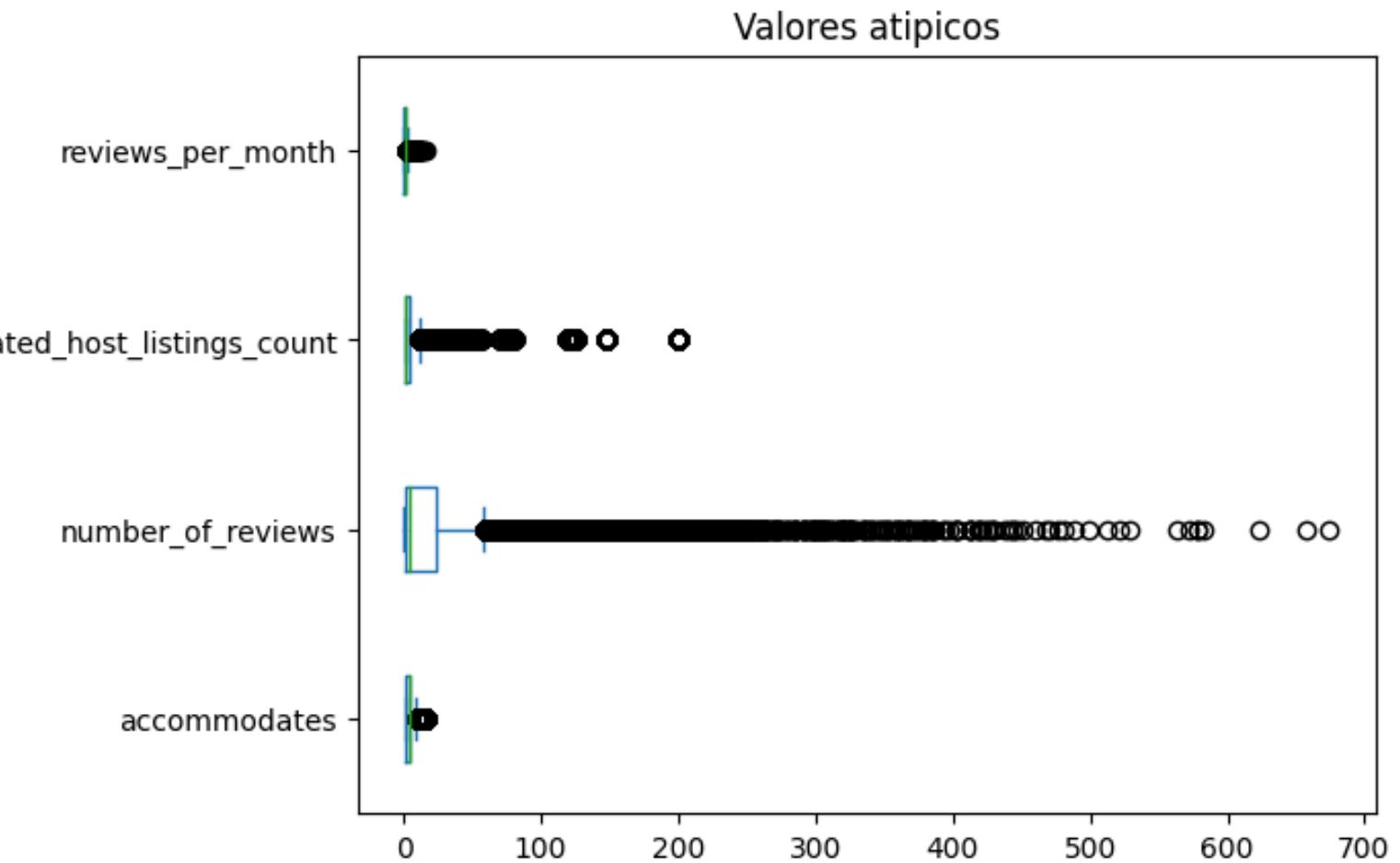


RIO DE JANEIRO



```
fig = plt.figure(figsize = (20, 10))
columnas_cuantitativas.plot(kind = 'box', vert = False)
plt.title("Valores atípicos")
plt.show()
```

<Figure size 2000x1000 with 0 Axes>



OBTENER LIMITES SUPERIOR E INFERIOR



MADRID

```
y = columnas_cuantitativas
Limite_superior = y.mean() + 3*y.std()
Limite_inferior = y.mean() - 3*y.std()
print("Limite superior permitido", Limite_superior)
print("Limite inferior permitido", Limite_inferior)

Limite superior permitido accommodates 8.873245
number_of_reviews      294.833157
review_scores_rating   5.817323
calculated_host_listings_count 214.009952
reviews_per_month      7.090891
dtype: float64
Limite inferior permitido accommodates -2.638808
number_of_reviews      -206.436931
review_scores_rating   3.527078
calculated_host_listings_count -158.844450
reviews_per_month      -3.303818
```



RIO DE JANEIRO

```
y = columnas_cuantitativas
Limite_superior = y.mean() + 3*y.std()
Limite_inferior = y.mean() - 3*y.std()
print("Limite superior permitido", Limite_superior)
print("Limite inferior permitido", Limite_inferior)

Limite superior permitido accommodates 11.026495
number_of_reviews      154.562573
calculated_host_listings_count 81.197680
reviews_per_month      4.208422
dtype: float64
Limite inferior permitido accommodates -2.996839
number_of_reviews      -109.955257
calculated_host_listings_count -63.055861
reviews_per_month      -2.028960
```

FILTRAR VALORES DENTRO DE LOS LIMITES

```
columnas_cuantitativas_sin_outliers = columnas_cuantitativas[(y <= Limite_superior) & (y >= Limite_inferior)]
```

REVISAR VALORES NULOS CON ISNULL().SUM()



MADRID

accommodates	377
number_of_reviews	674
review_scores_rating	488
calculated_host_listings_count	1099
reviews_per_month	486



RIO DE JANEIRO

accommodates	644
number_of_reviews	811
calculated_host_listings_count	716
reviews_per_month	650

HACER UNA COPIA DEL DATAFRAME Y RELLENAR LOS VALORES NULOS CON LA MEDIA

```
columnas_cuantitativas_limpio = columnas_cuantitativas_sin_outliers.copy()  
columnas_cuantitativas_limpio = columnas_cuantitativas_limpio.fillna(round(columnas_cuantitativas_sin_outliers.mean(), 1))
```

JUNTAR LAS COLUMNAS CUANTITATIVAS LIMPIAS CON LAS CUALITATIVAS CON LA FUNCION PD.CONCAT()

```
Madrid_outliers = pd.concat([columnas_cuantitativas_limpio, columnas_cualitativas], axis = 1)  
Madrid_outliers.head()
```

```
Rio_outliers = pd.concat([columnas_cuantitativas_limpio, columnas_cualitativas], axis = 1)  
Rio_outliers.head()
```

CORROBORAR QUE NO HAYA VALORES NULOS

MADRID

RIO DE JANEIRO

accommodates	0
number_of_reviews	0
review_scores_rating	0
calculated_host_listings_count	0
reviews_per_month	0
listing_url	0
last_scraped	0
source	0
name	0
host_url	0
host_name	0
host_since	0
host_location	0
host_response_time	0
host_response_rate	0
host_acceptance_rate	0
host_is_superhost	0
host_neighbourhood	0
host_verifications	0
host_has_profile_pic	0

host_identity_verified	0
neighbourhood_cleansed	0
property_type	0
room_type	0
bathrooms_text	0
bedrooms	0
beds	0
amenities	0
price	0
has_availability	0
instant_bookable	0

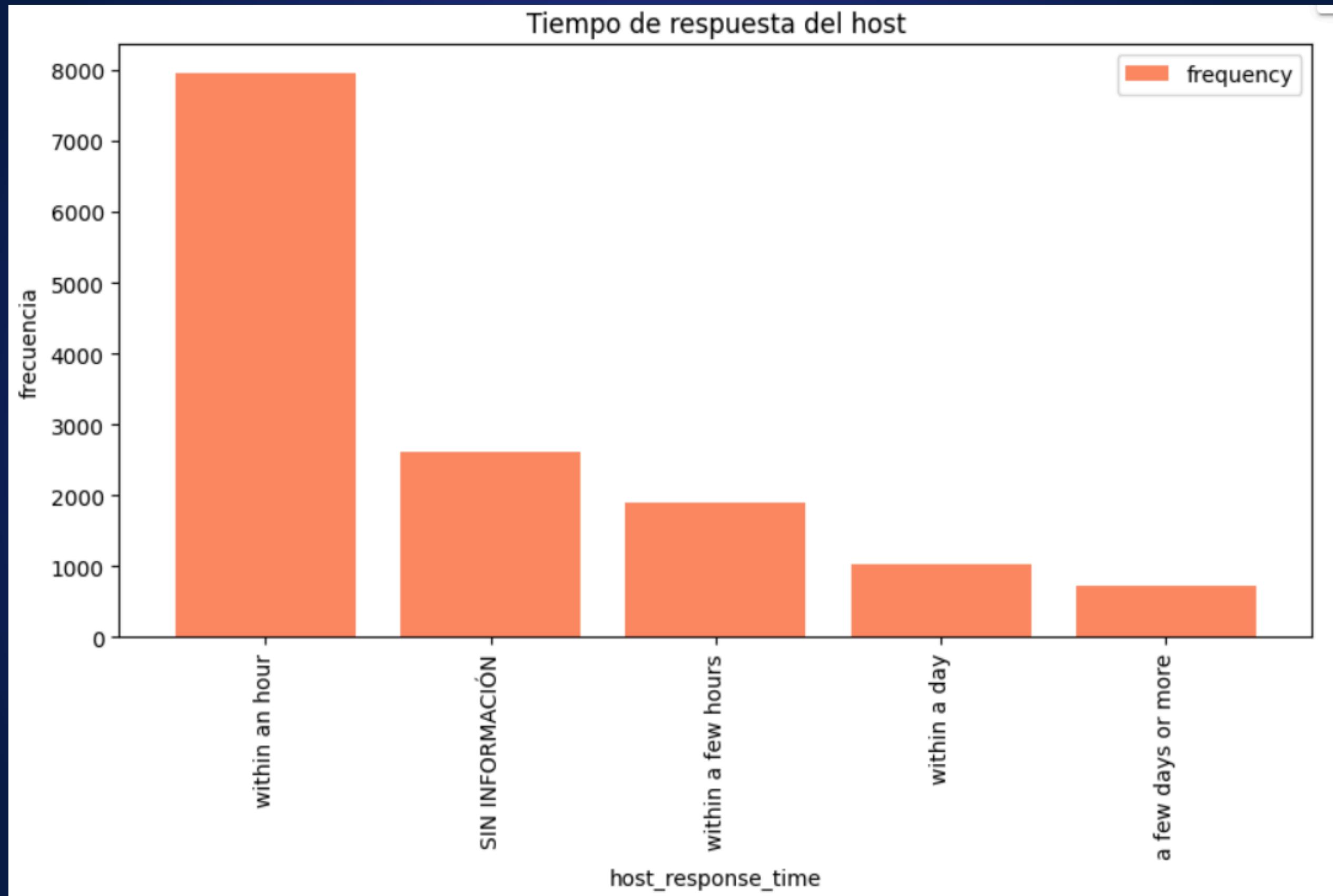


accommodates	0
number_of_reviews	0
calculated_host_listings_count	0
reviews_per_month	0
listing_url	0
last_scraped	0
source	0
name	0
host_url	0
host_name	0
host_since	0
host_location	0
host_response_time	0
host_response_rate	0
host_acceptance_rate	0
host_is_superhost	0
host_neighbourhood	0
host_verifications	0
host_has_profile_pic	0

host_identity_verified	0
neighbourhood_cleansed	0
property_type	0
room_type	0
bathrooms_text	0
bedrooms	0
beds	0
amenities	0
price	0
has_availability	0
review_scores_rating	0
instant_bookable	0

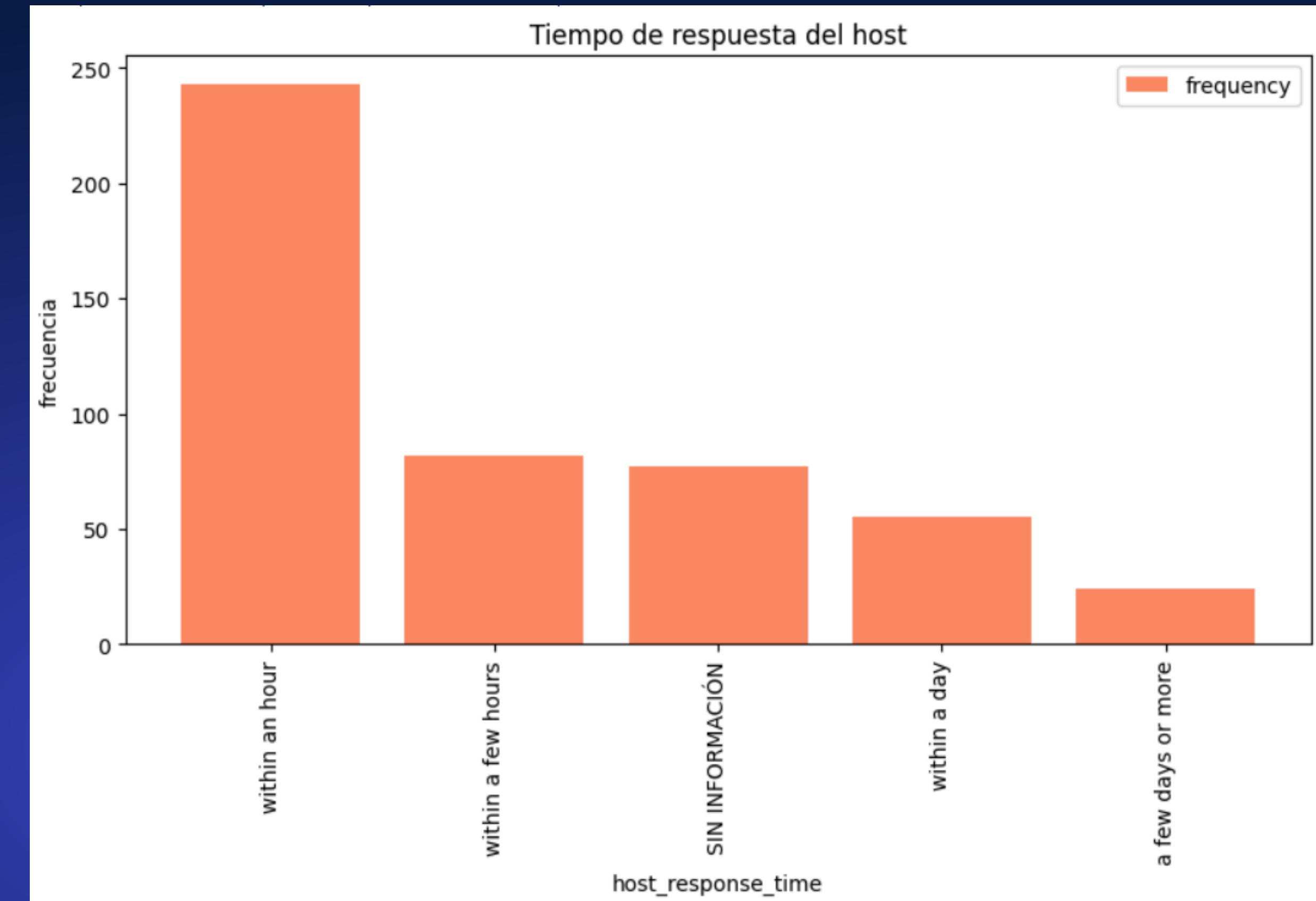


GRÁFICAS



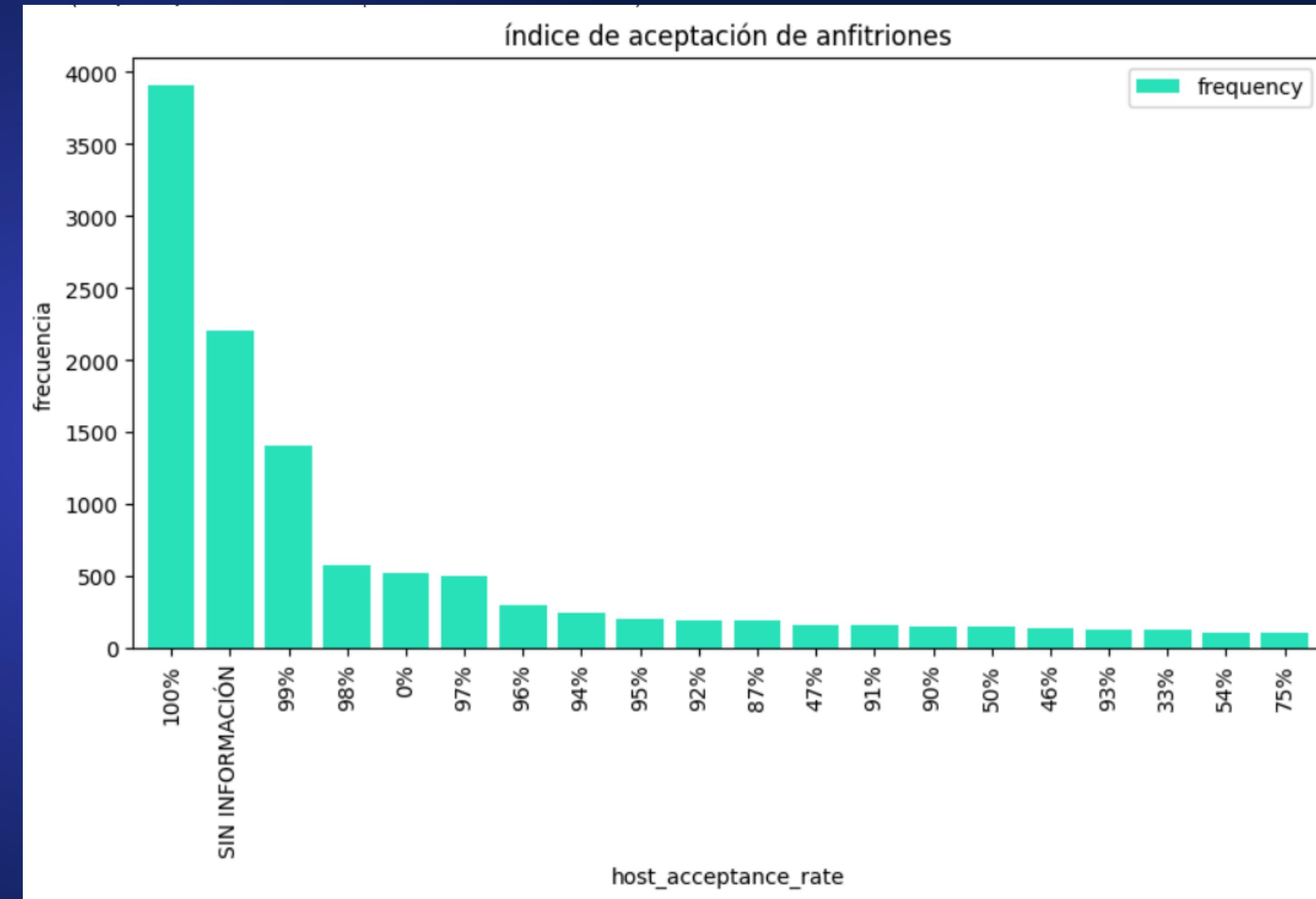
GRÁFICAS

RIO DE JANEIRO



GRÁFICAS

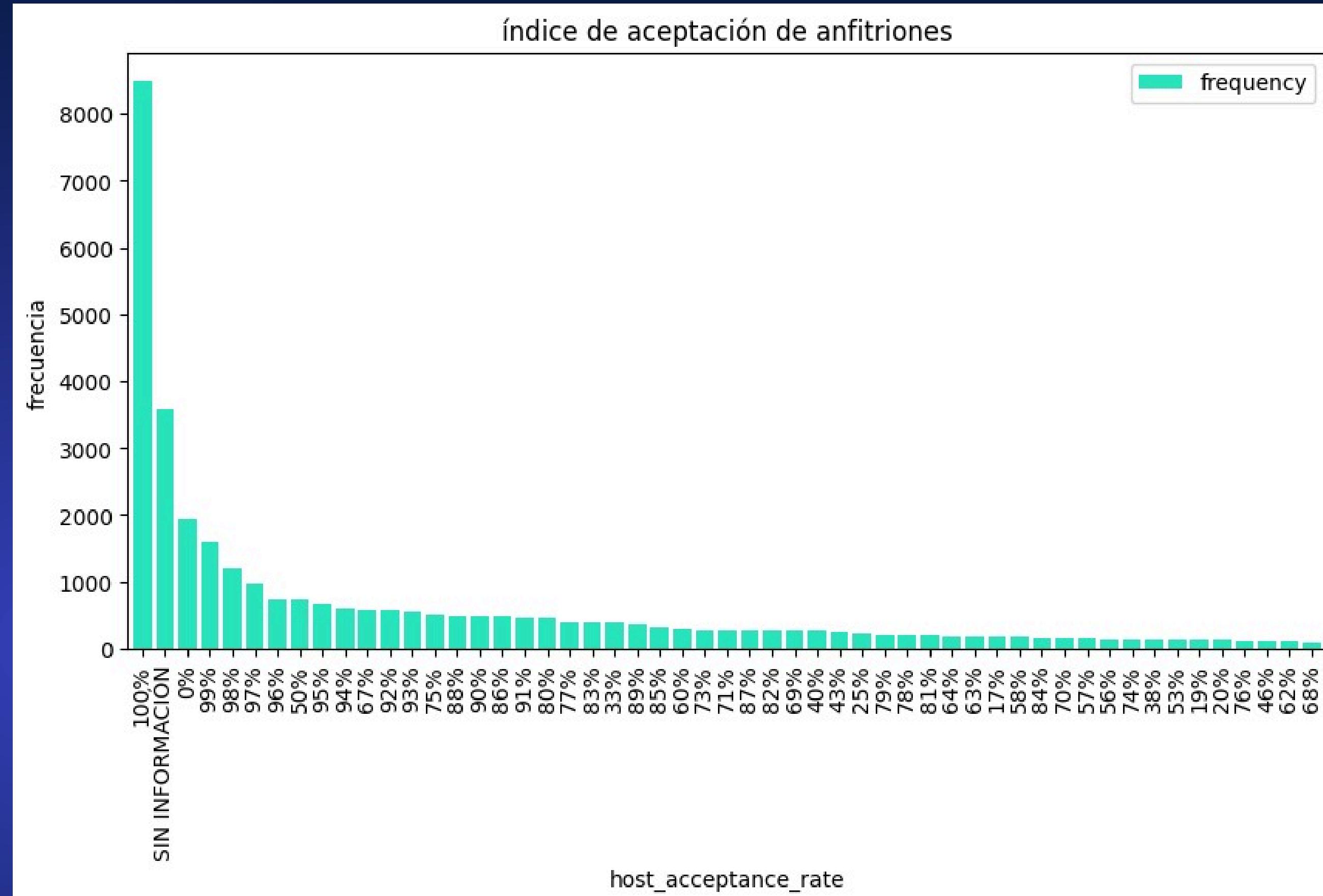
MADRID



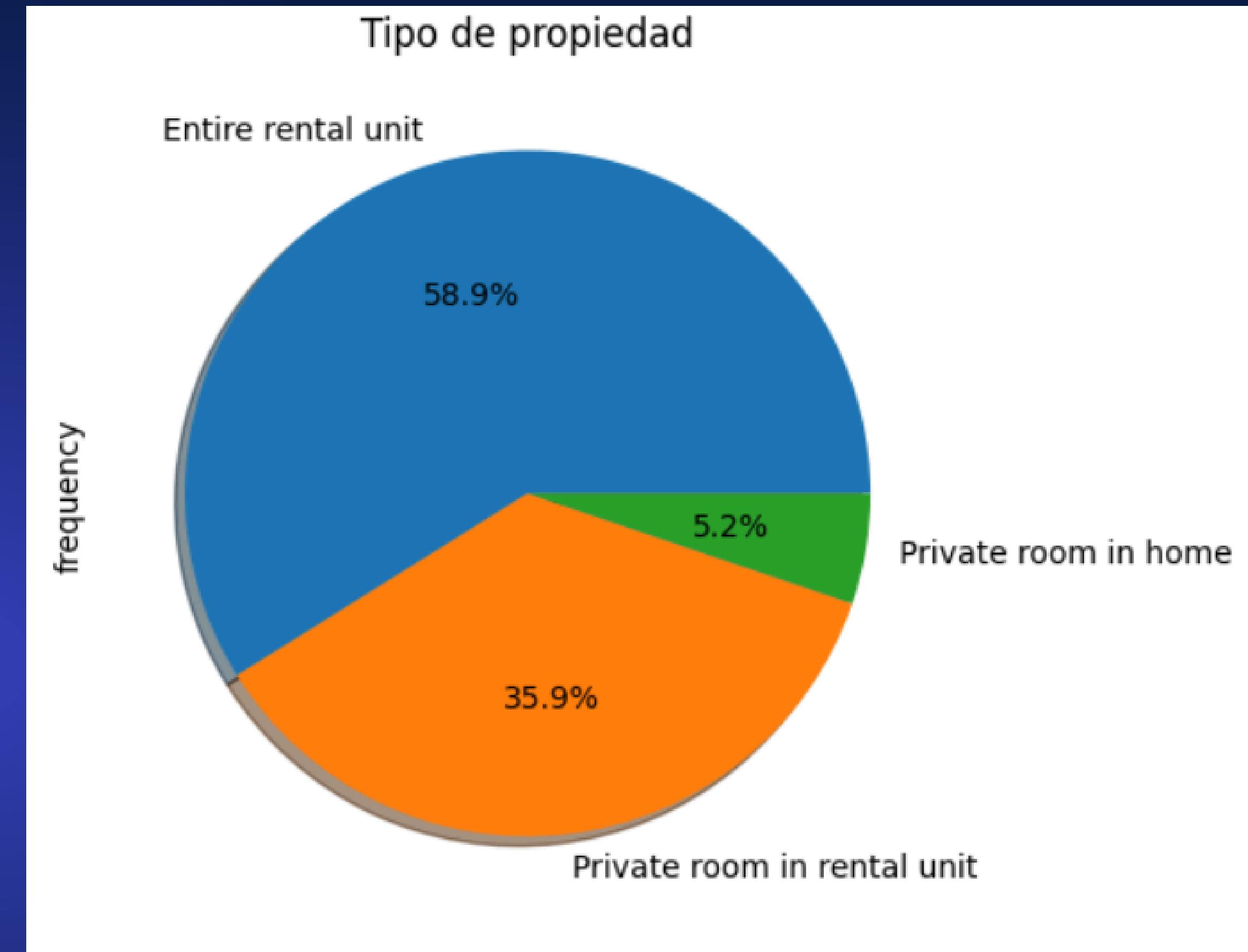
GRÁFICAS



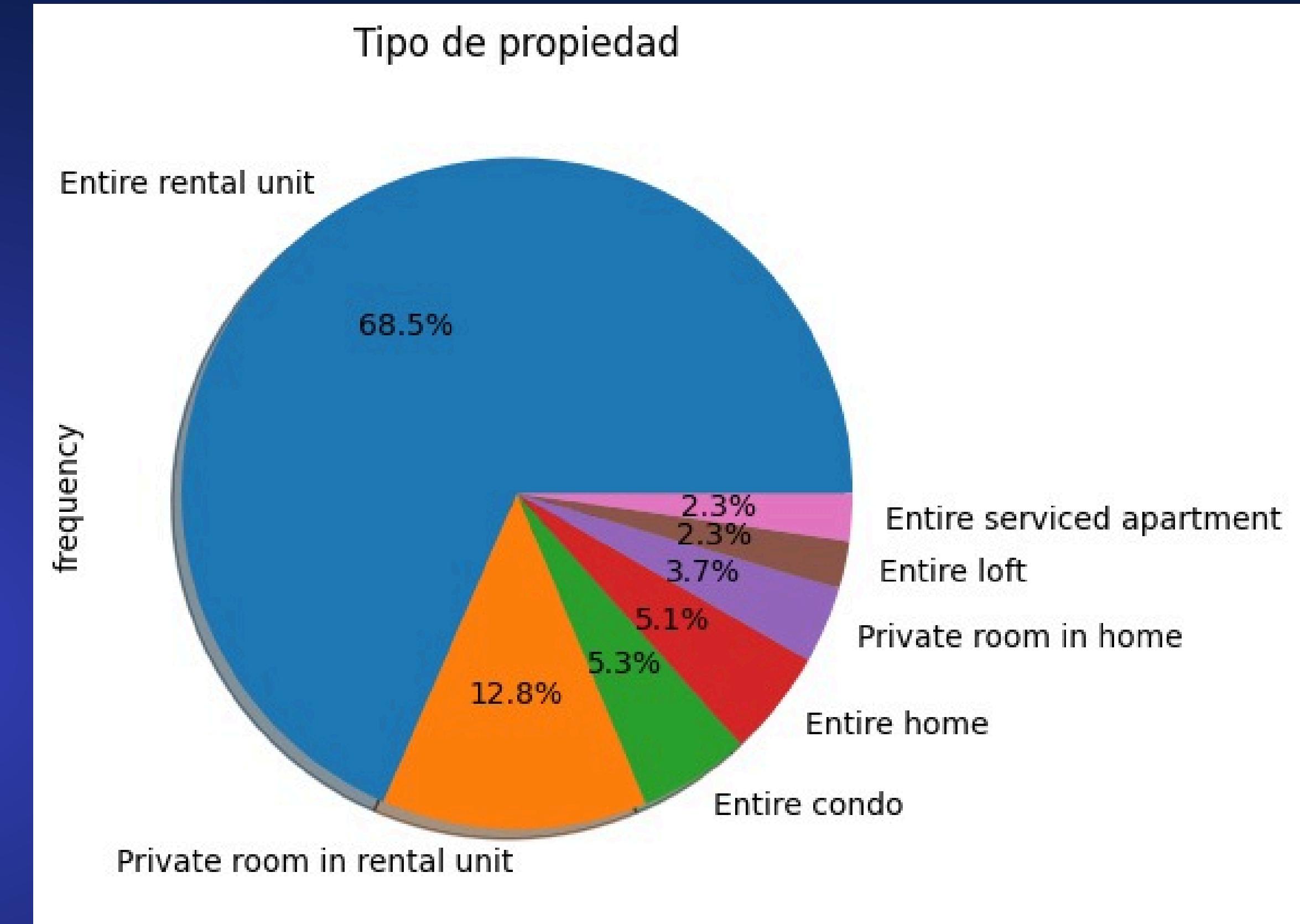
índice de aceptación de anfitriones



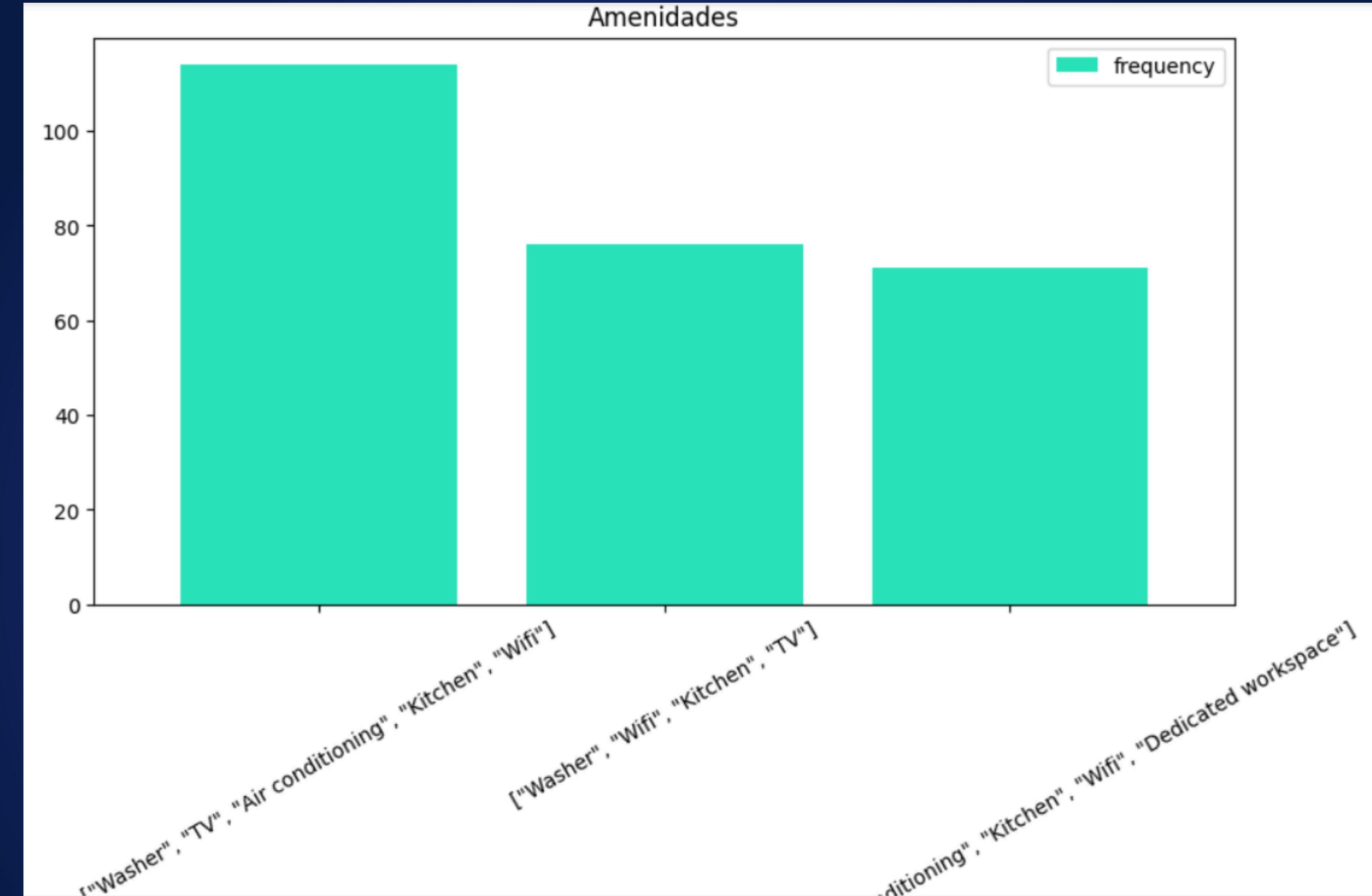
GRÁFICAS



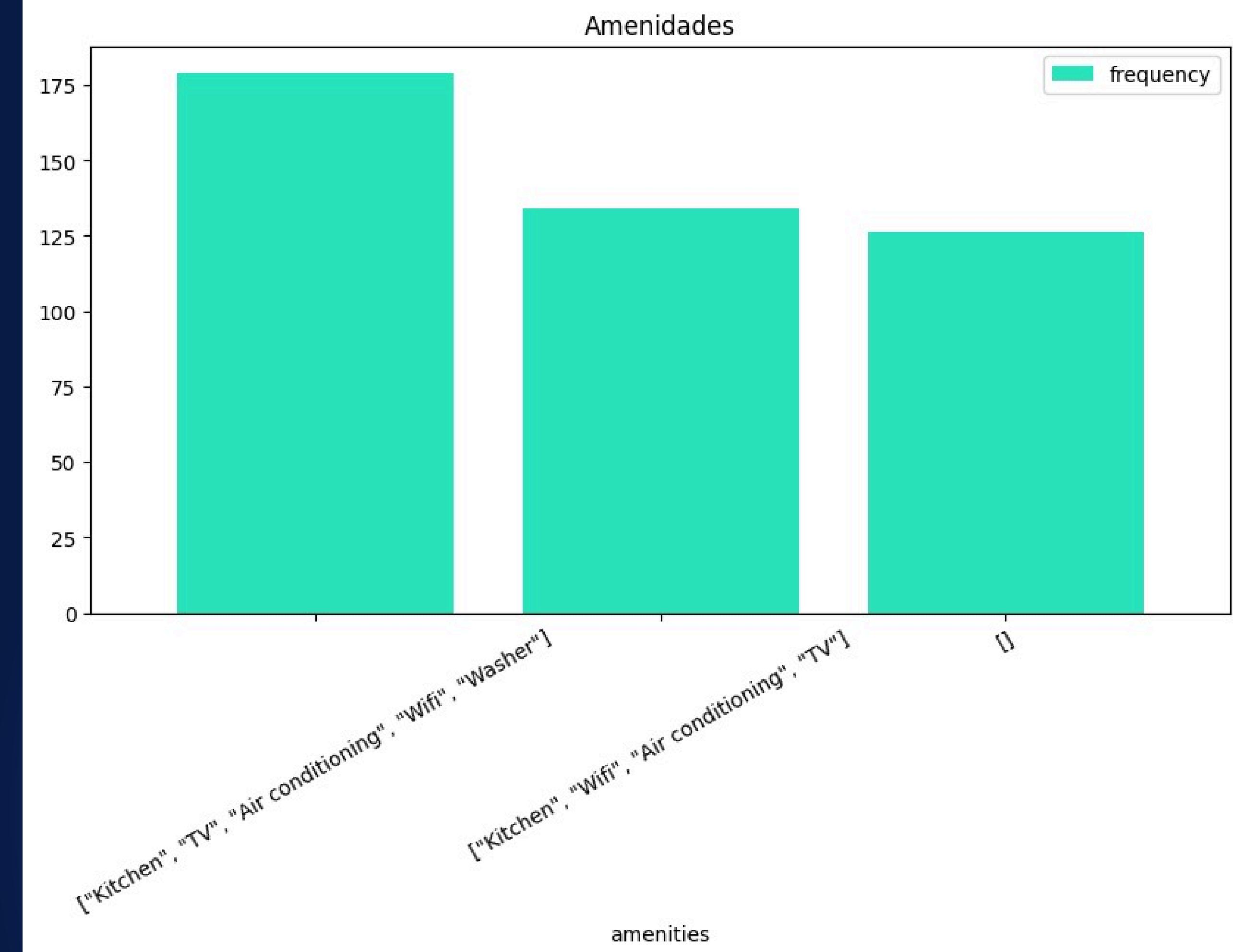
GRÁFICAS



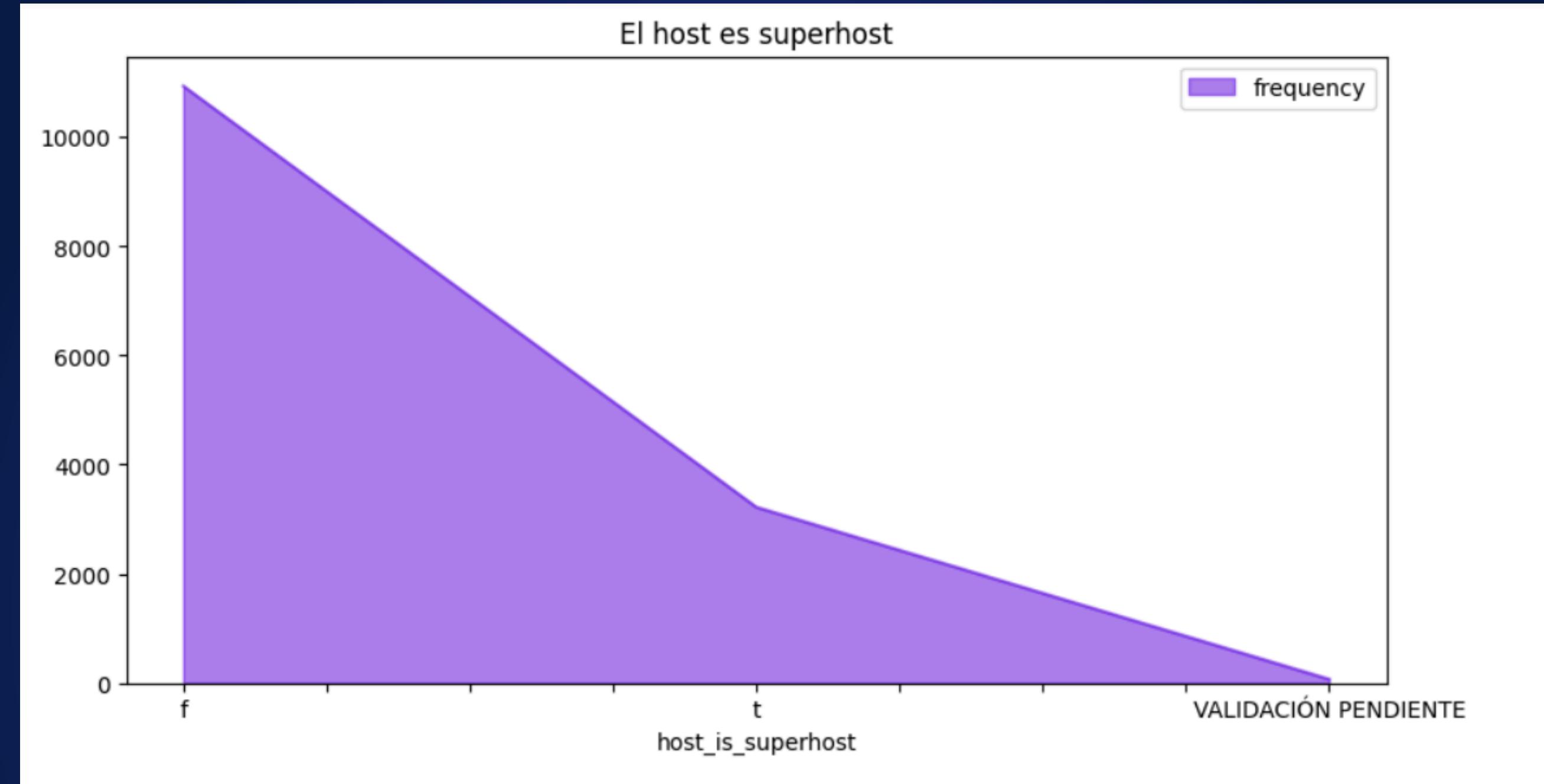
GRÁFICAS



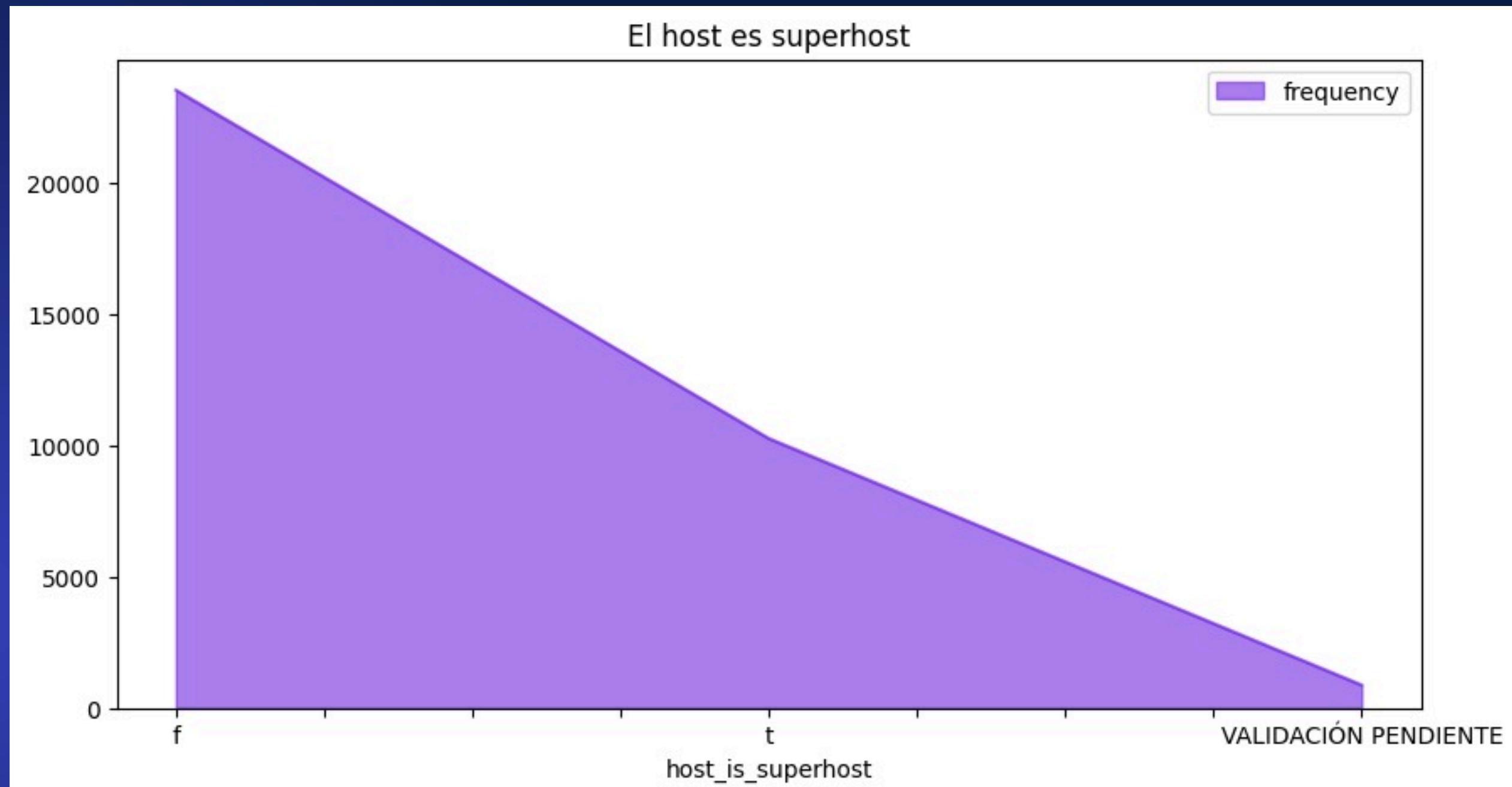
GRÁFICAS



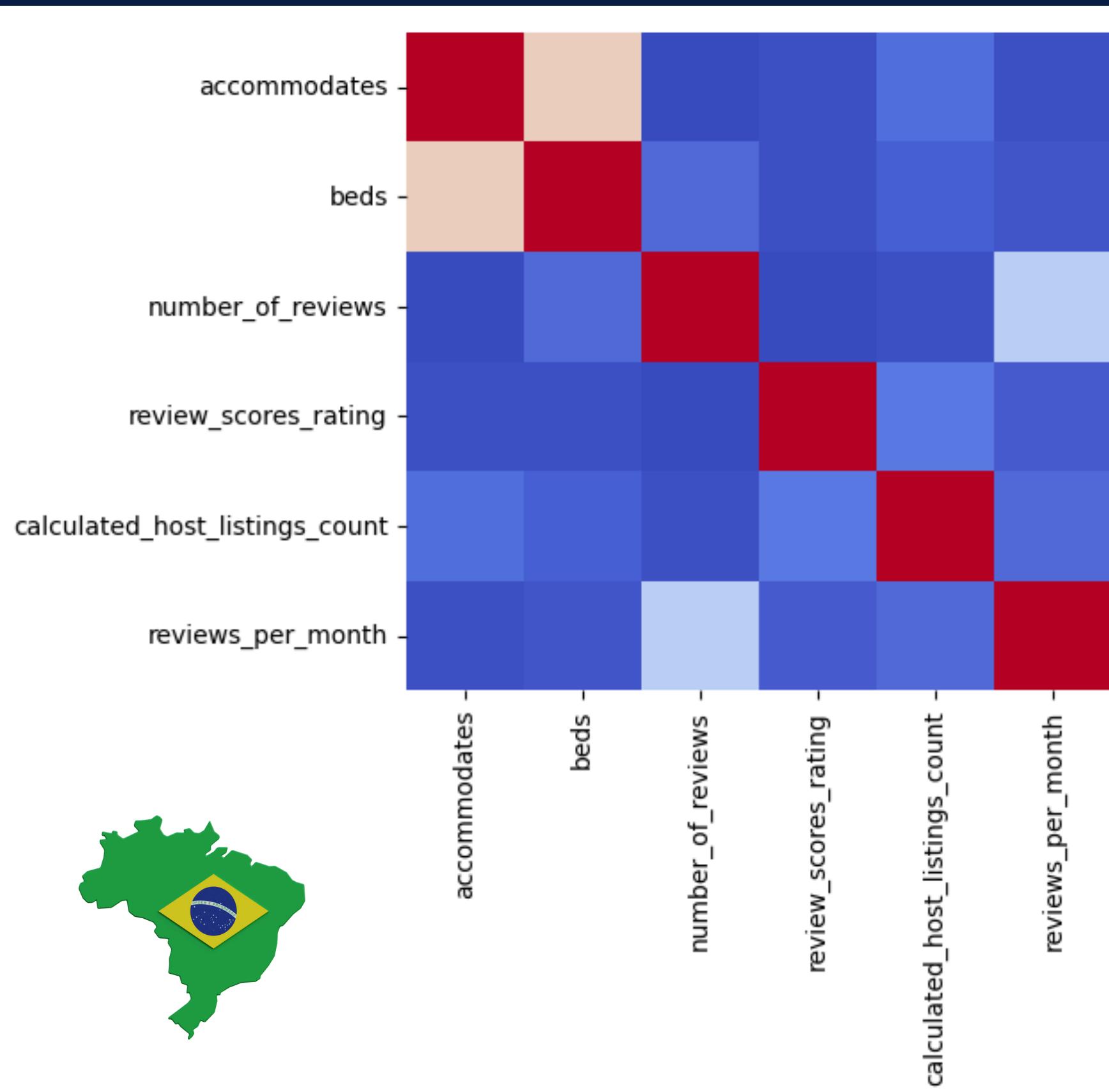
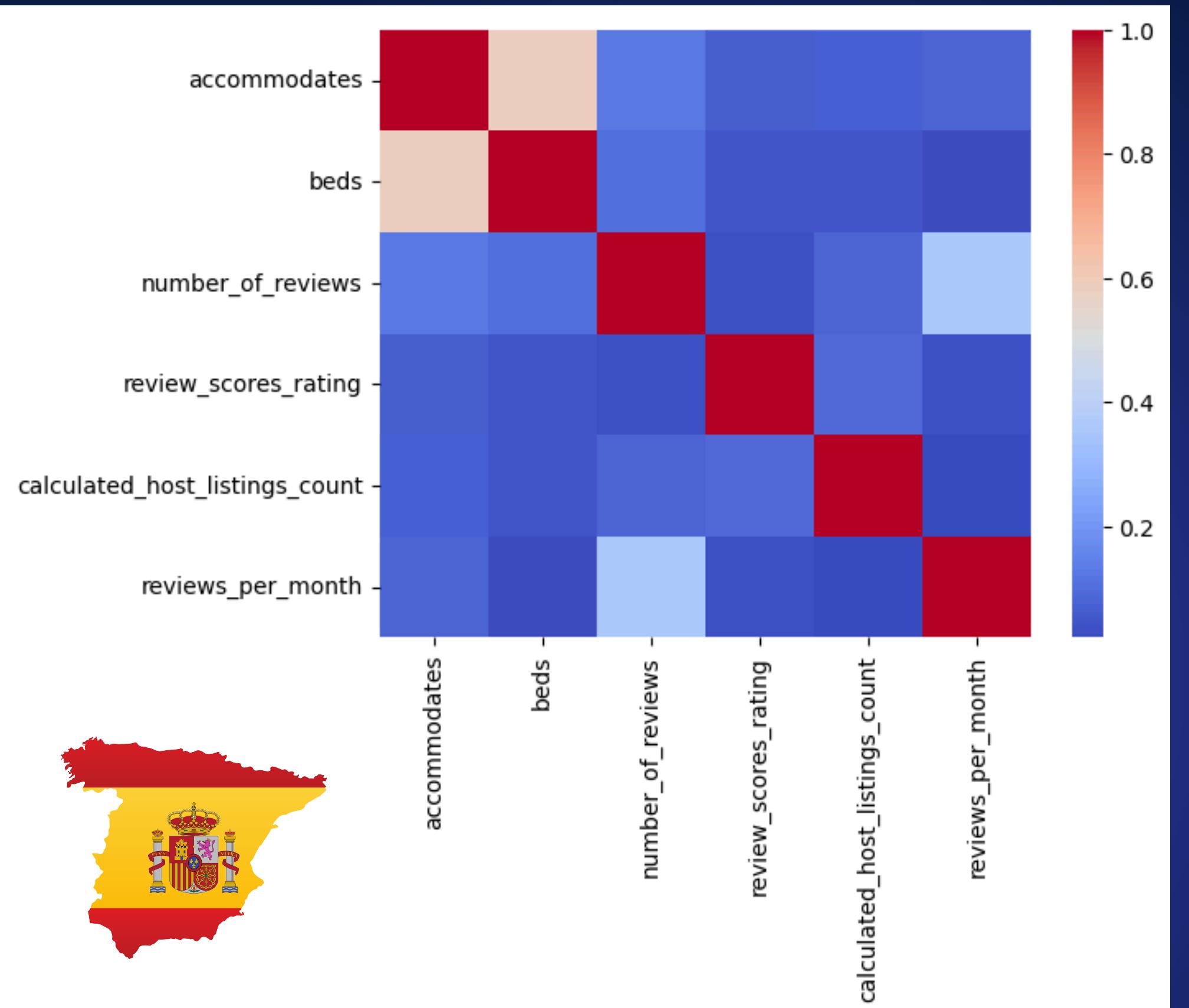
GRÁFICAS



GRÁFICAS

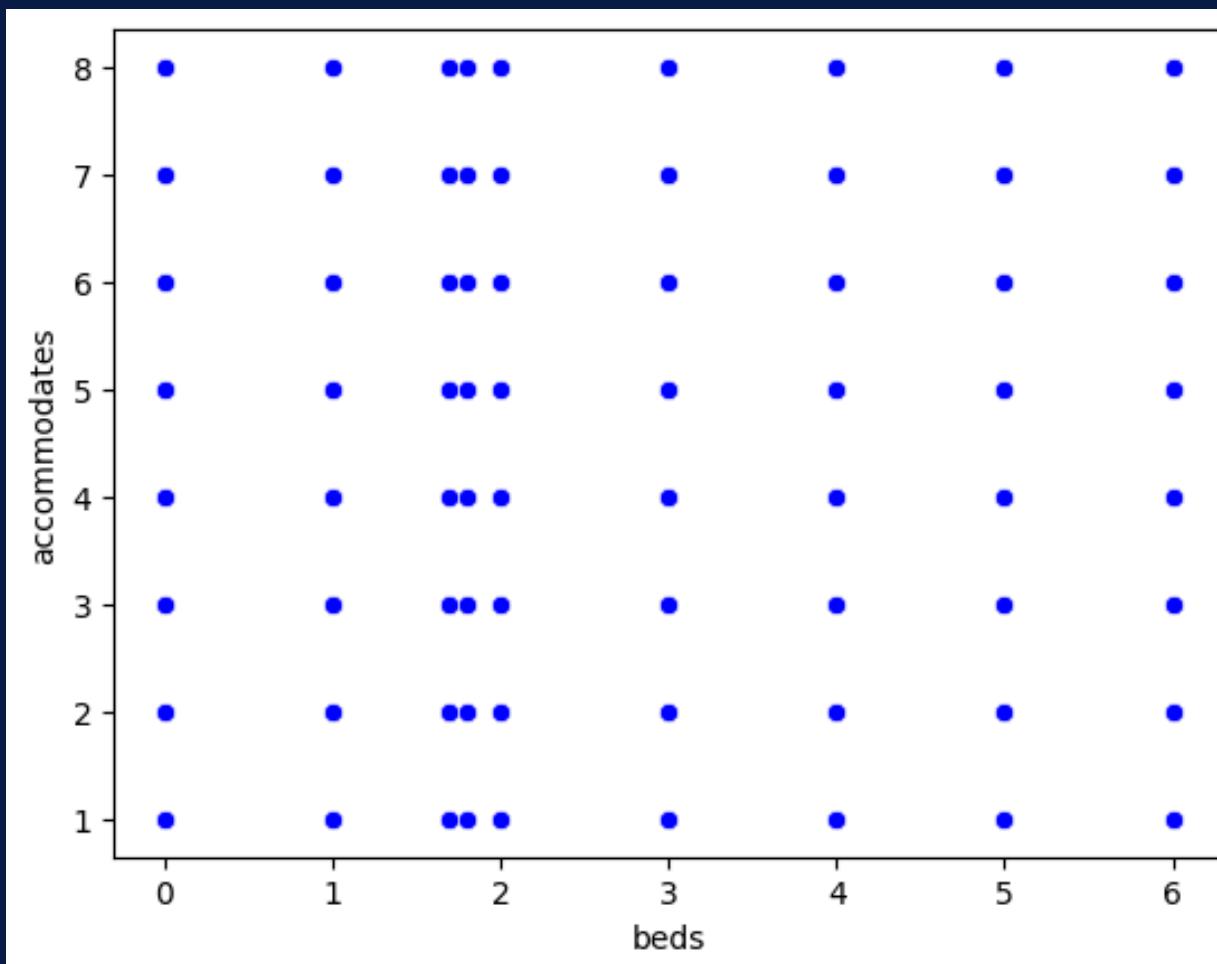


MAPA DE CALOR



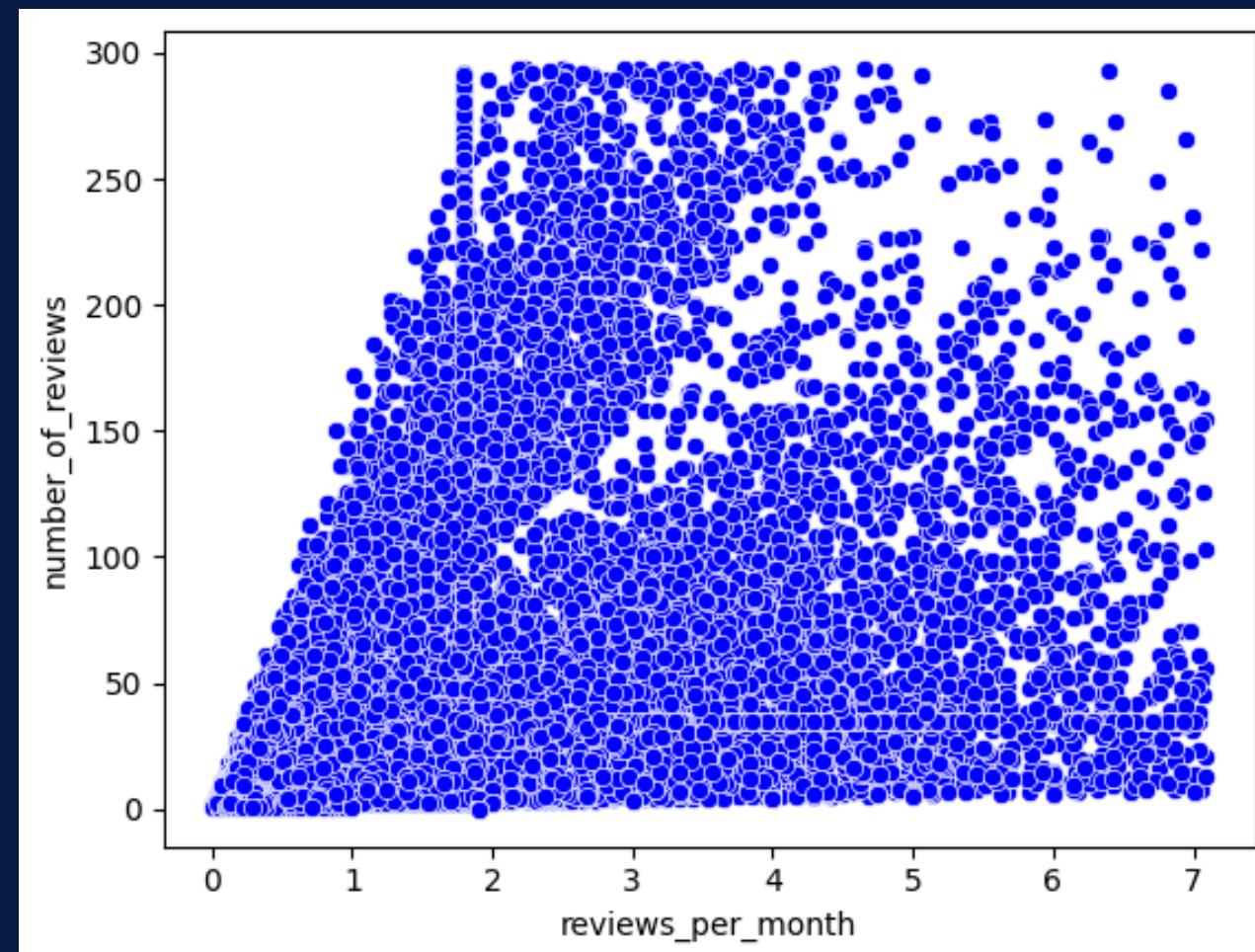


TOP 3 MODELOS



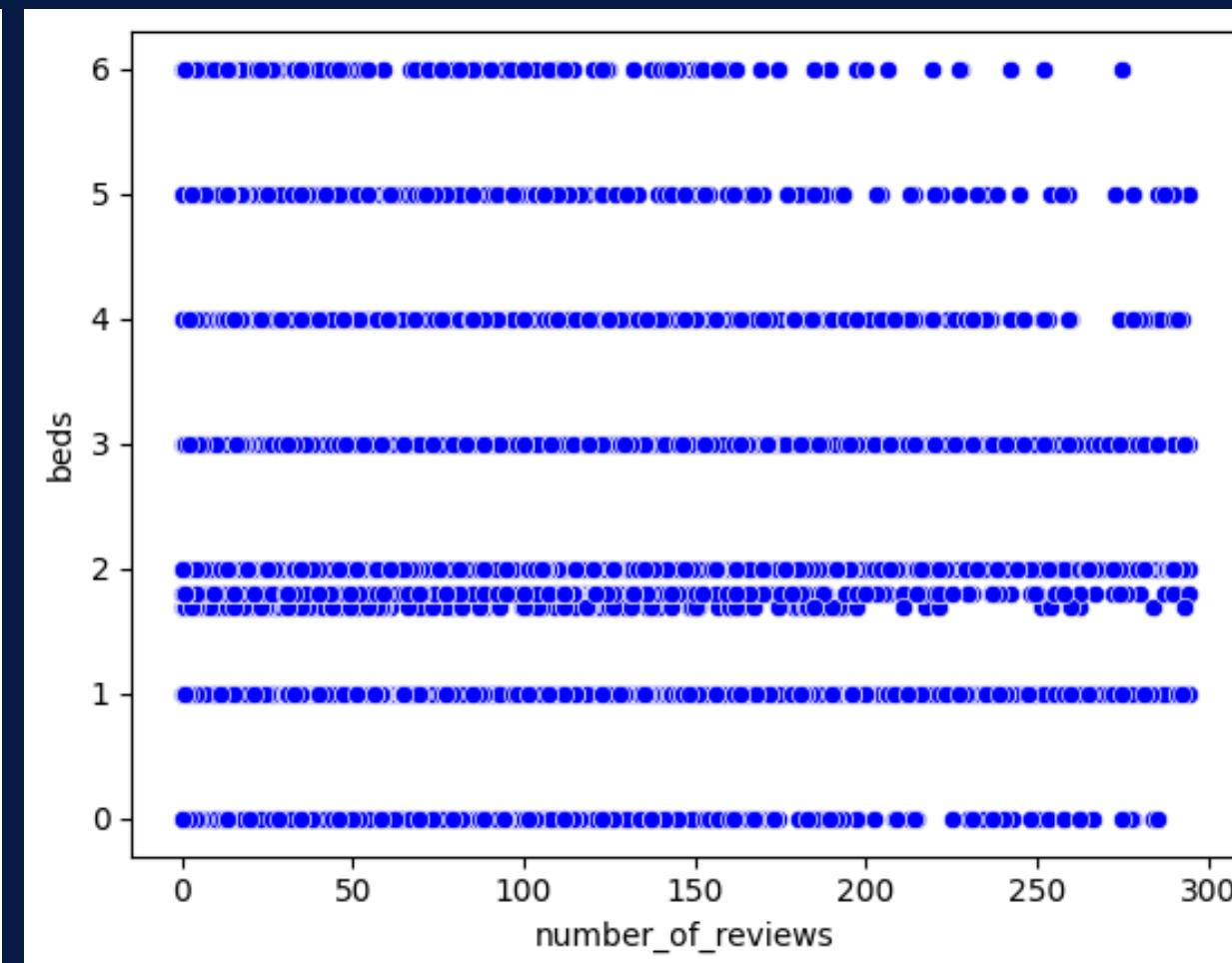
$$y = 1.53 + 0.86X$$

R² = 0.58



$$y = 10.09 + 13.88X$$

R² = 0.35

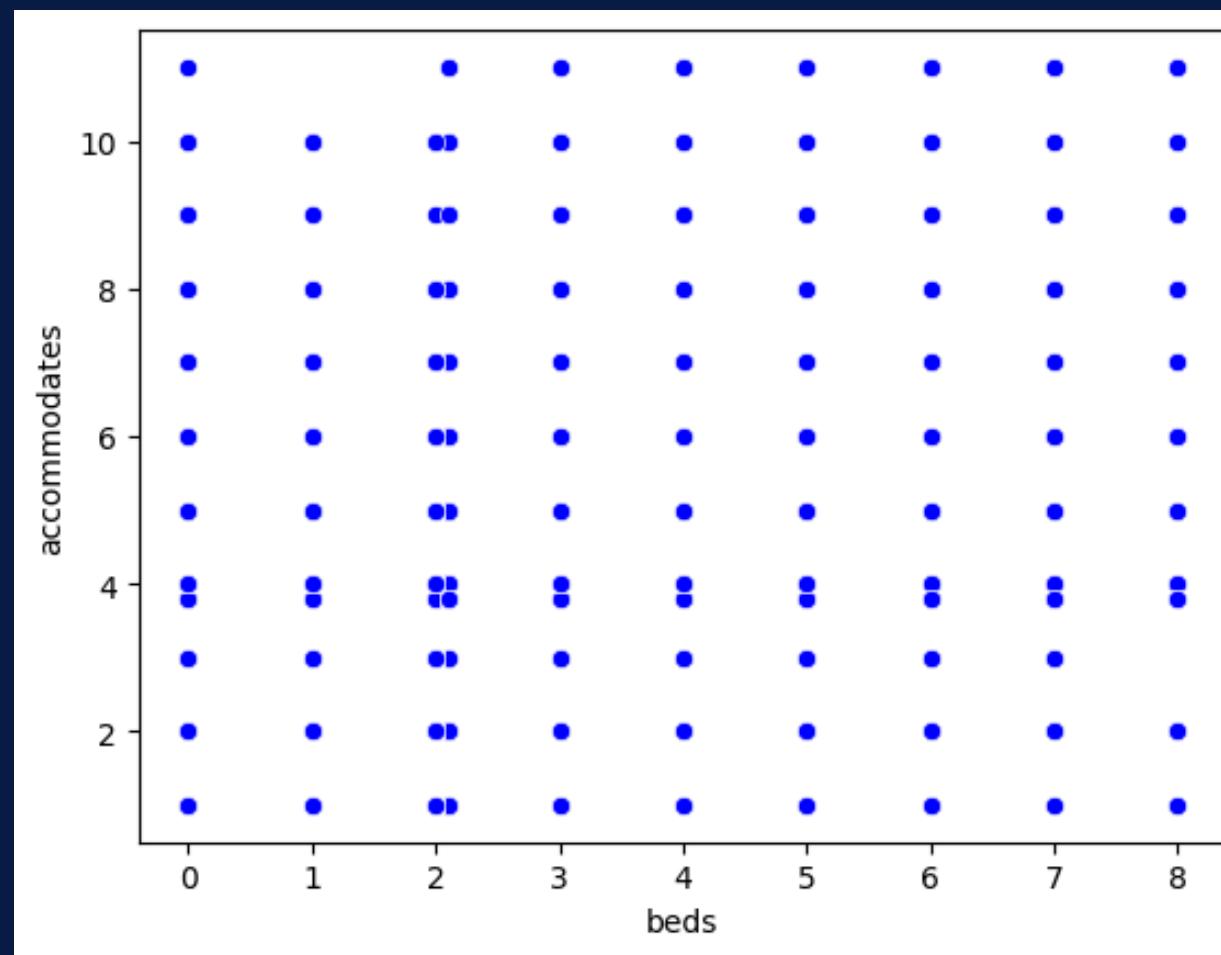


$$y = 1.64 + 0.0021X$$

R² = 0.102

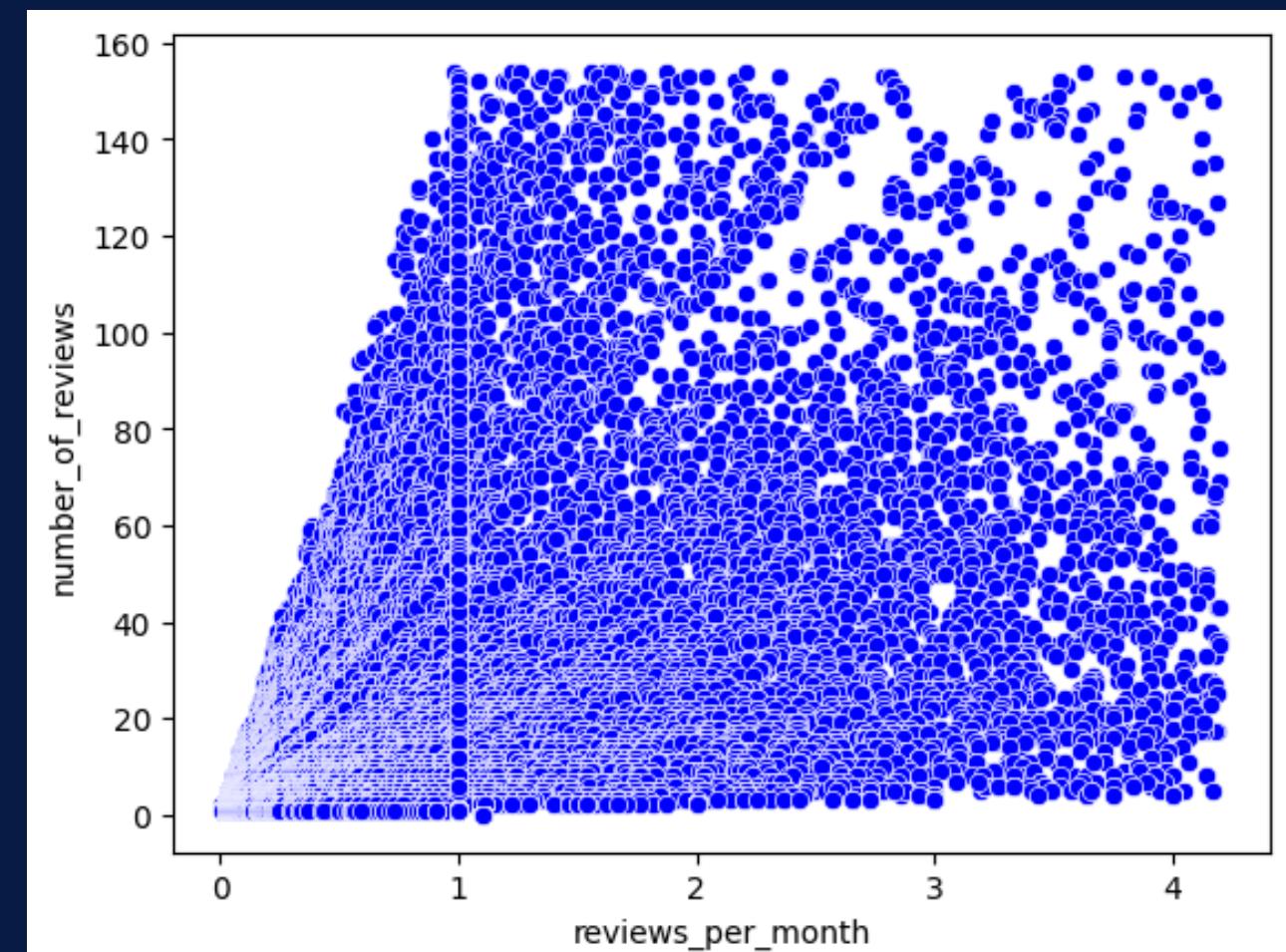


TOP 3 MODELOS



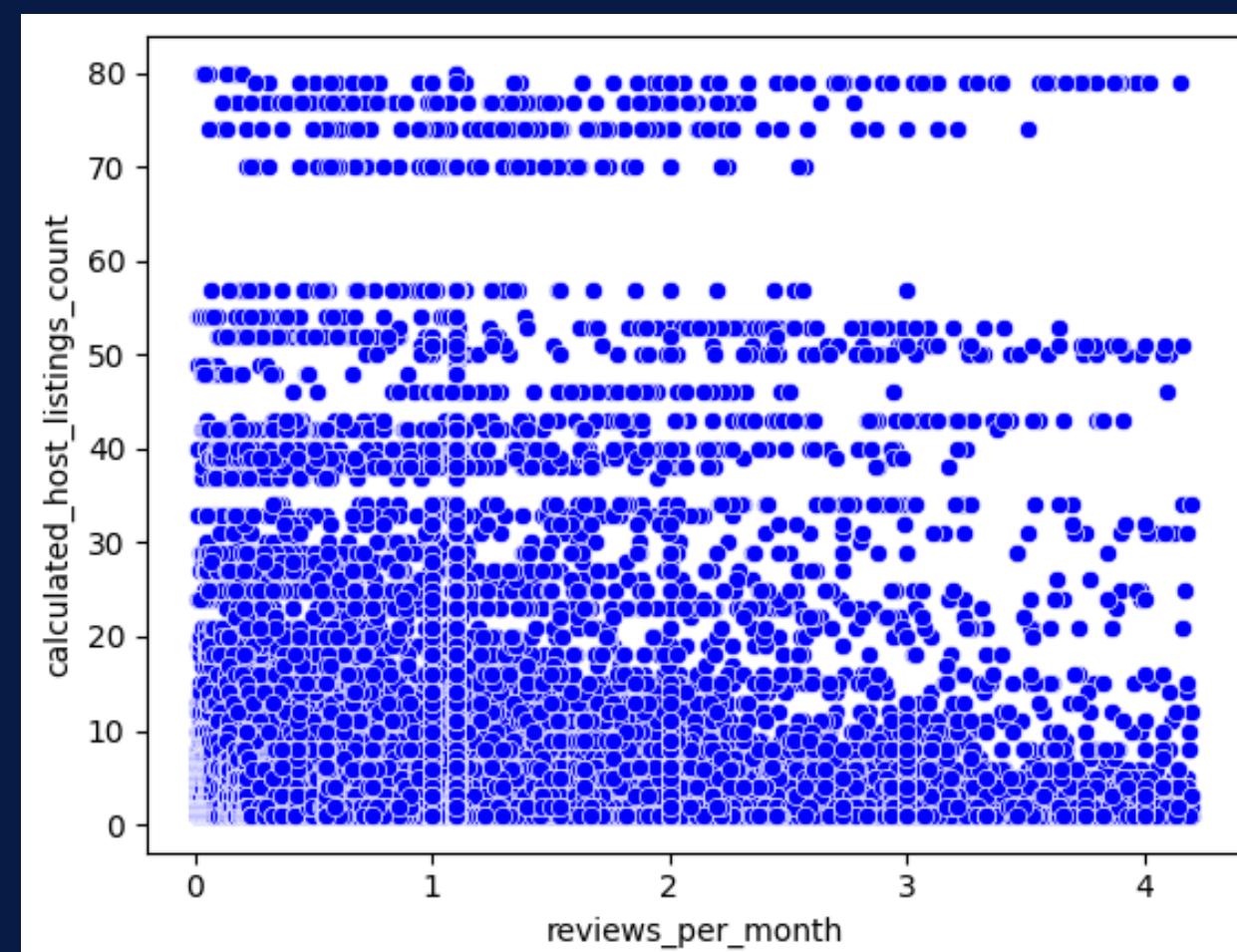
$$y = 2.35 + 0.70X$$

R² = 0.58



$$y = 4.15 + 13.12X$$

R² = 0.38



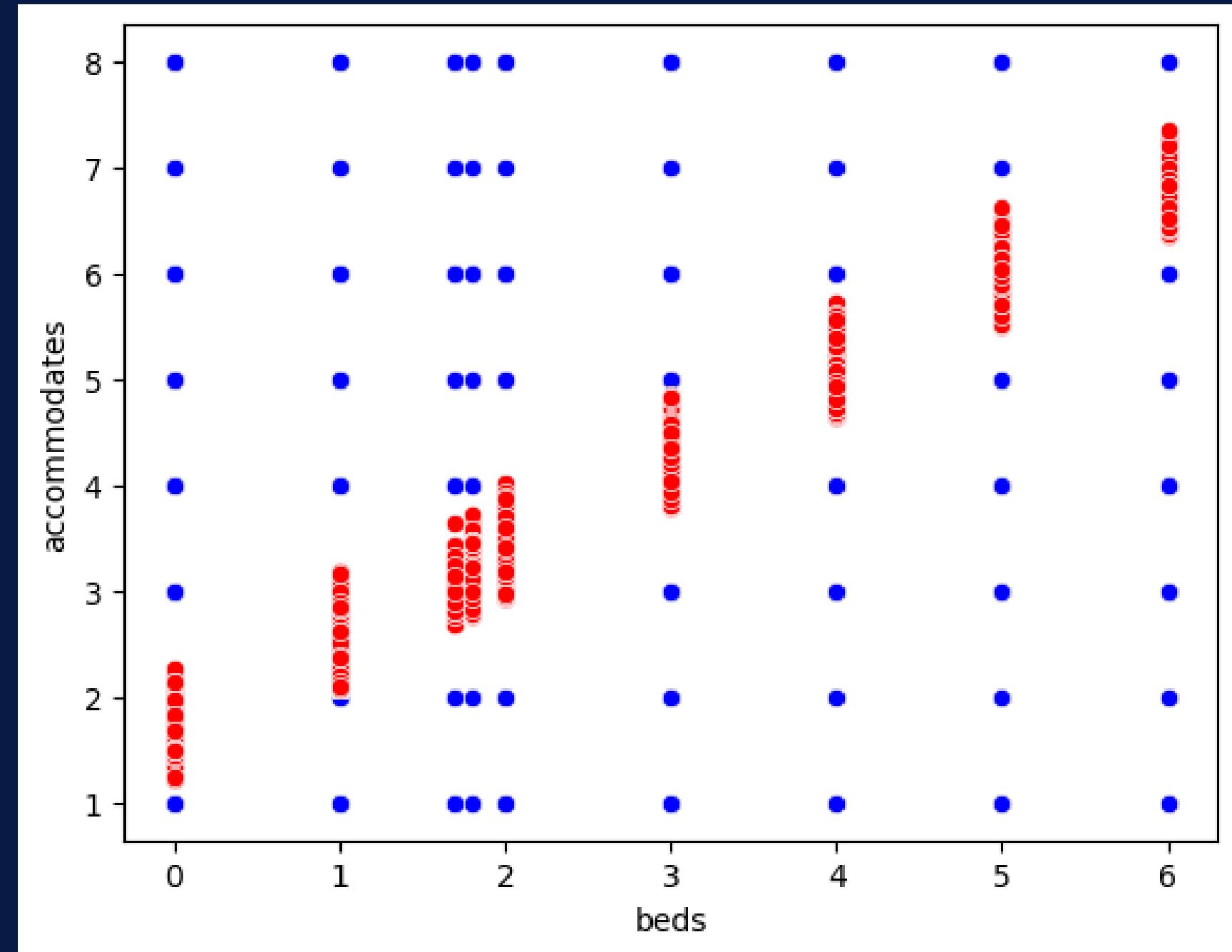
$$y = 39.05 - 6.82X$$

R² = 0.11

REGRESIÓN MULTIPLE



- Variables cuantitativas:
- Beds
 - Number of reviews
 - Review scores rating
 - calculated host listings count
 - Reviews per month



$$y=2.44 +0.85X_1+0.001X_2 -0.24X_3 +0.001X_4 +0.09X_5$$

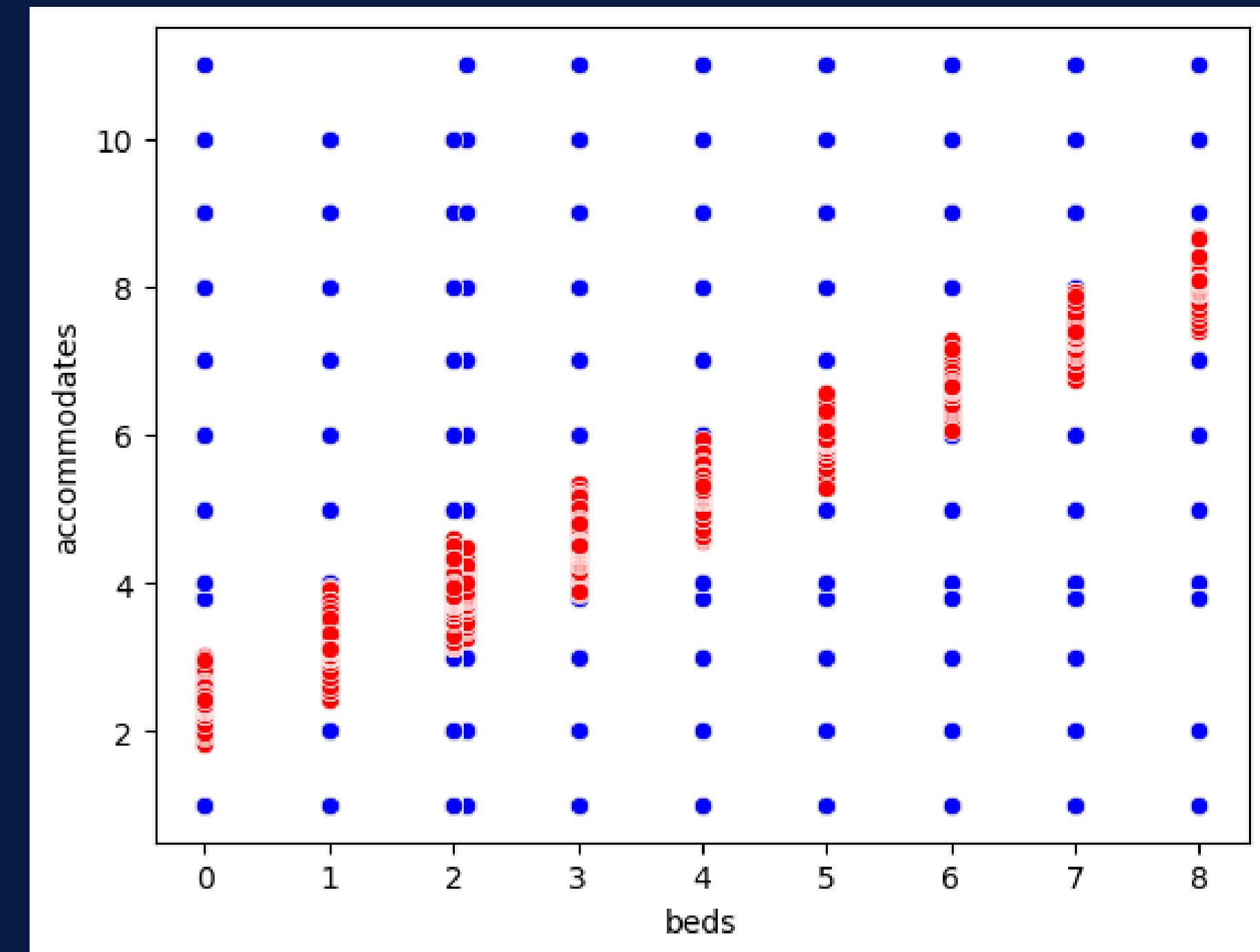
$$R^2= 0.60$$

REGRESIÓN MULTIPLE



Variables cuantitativas:

- Beds
- Number of reviews
- Review scores rating
- calculated host listings count
- Reviews per month



$$y = 2.40 + 0.70X_1 - 0.004X_2 - 0.031X_3 + 0.007X_4 + 0.12X_5$$

$$R^2 = 0.60$$

CONCLUSIONES

- Las variables cuantitativas obtenidas al final del procesamiento de información no muestran una gran correlación.
- Se deberá plantear otro modelo en el cual se consideren variables de tipo int o float omitidas en este proceso, por ejemplo; price.
- Ambas ciudades, a pesar de pertenecer a distintos continentes, muestran un comportamiento bastante similar respecto al análisis de frecuencia. Esto nos puede hablar acerca de otras características que comparten en un trasfondo, como por ejemplo el turismo, condiciones geográficas, entre otras.
- Gracias a la limpieza de valores nulos y outliers, se puede obtener un análisis mucho más acertado y que refleje todavía más información que sea de utilidad.

Muchas Gracias