Universidad de los Andes
Ingeniería de Sistemas y Computación
MINE 4101 – CIENCIA DE DATOS APLICADA - TALLER 1
Caso: Inversiones inmobiliarias para alquiler vacacional.

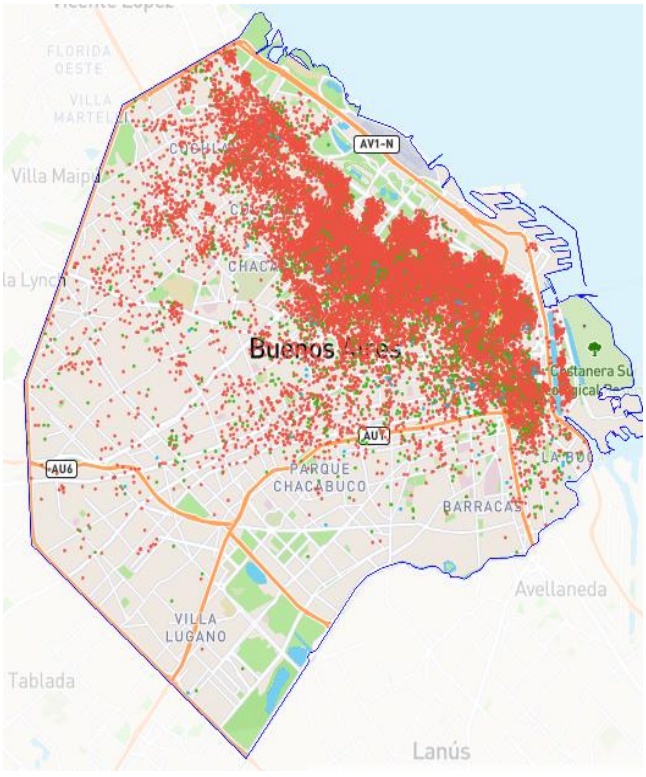
Profesor
Fabián Camilo Peña Lozano

GABRIEL HERNANDEZ REYES
Código 201728589

Contexto del negocio plataforma AIRBNB

Con la creciente popularidad de este modelo de negocio, es cada vez más difícil hacer buenas elecciones de sectores de la ciudad y tipos de propiedades, entre otro tipo de factores, para lograr una rentabilidad adecuada a lo largo del tiempo la cual puede estar dada en términos de una alta tasa de ocupación y un precio por noche que los usuarios estarían dispuestos a pagar.

1. Para el caso se escoge el **dataset** de la ciudad de **Buenos Aires Argentina** con la siguiente información:



2. ENTENDIMIENTO INICIAL DE DATOS:

Generar un breve reporte de entendimiento inicial de datos en donde destaque las dimensiones del dataset, los tipos de datos que contiene y el top 5 de los atributos que considera más importantes para el análisis. Por cada atributo incluya algunos elementos básicos de su comportamiento o distribución (análisis univariado).

- **Reporte de Entendimiento Inicial**
  1. **<u>Carga de datos:</u>**

jupyter TALLER1_AIRBNB_BUENOS_AIRES Last Checkpoint: hace 4 minutos (autosaved)

File Edit View Insert Cell Kernel Widgets Help

Run Code

In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

In [2]:
```
calendar = pd.read_csv('calendar.csv')
vecindarios = pd.read_csv('vecindarios.csv')
listings = pd.read_csv('listings.csv')
reviews = pd.read_csv('reviews.csv')
listings1 = pd.read_csv('listings1.csv')
reviews1 = pd.read_csv('reviews1.csv')
```

In [3]:
```
calendar.head()
```

Out[3]:

	listing_id	date	available	price	adjusted_price	minimum_nights	maximum_nights
0	11508	2023-06-29	f	$16,608.00	$16,608.00	3.0	1125.0
1	11508	2023-06-30	f	$19,163.00	$19,163.00	3.0	1125.0
2	11508	2023-07-01	f	$19,163.00	$19,163.00	3.0	1125.0
3	11508	2023-07-02	f	$16,608.00	$16,608.00	3.0	1125.0
4	11508	2023-07-03	f	$16,608.00	$16,608.00	3.0	1125.0

In [4]:
```
vecindarios.head()
```

Out[4]:

	neighbourhood_group	neighbourhood
0	NaN	Agronomia
1	NaN	Almagro
2	NaN	Balvanera
3	NaN	Barracas
4	NaN	Belgrano

In [5]:
```
reviews.head()
```

Out[5]:

	listing_id	date
0	11508	2012-07-02
1	11508	2012-12-26
2	11508	2013-01-05
3	11508	2013-07-28
4	11508	2013-12-22

GABRIEL HERNANDEZ REYES
Código 201728589

In [6]:
```
listings1.head()
```

Out[6]:

	id	listing_url	scrape_id	last_scraped	source	name	description	neighborhood_overview	pictur
0	11508	https://www.airbnb.com/rooms/11508	20230628025456	2023-06-29	city scrape	Condo in Buenos Aires · ★4.81 · 1 bedroom · 1 ...	LUXURIOUS 1 BDRM APT-POOL/ GYM/ SPA/ 24-HR SE...	AREA: PALERMO SOHO Minutes walking ...	https://a0.muscache/pictures/1935/b1
1	14222	https://www.airbnb.com/rooms/14222	20230628025456	2023-06-28	city scrape	Rental unit in Palermo/Buenos Aires · ★4.78 · ...	Beautiful cozy apartment in excellent location...	Palermo is such a perfect place to explore the...	https://a0.muscache/pictures/469/bb
2	206738	https://www.airbnb.com/rooms/206738	20230628025456	2023-06-29	city scrape	Rental unit in Buenos Aires · ★4.71 · 3 bedroo...	Recently renovated 3 bedroom apartment in Pale...	Palermo is fortunate to have great transport l...	https://a0.muscache/pictures/8b3b1/fe
3	210040	https://www.airbnb.com/rooms/210040	20230628025456	2023-06-29	city scrape	Condo in Buenos Aires · ★4.92 · 1 bedroom · 1 ...	Modern studio apartment on the 13th floor with...	If you're an expat or vacationer spending time...	https://a0.muscache/pictures/2796/cb
4	92228	https://www.airbnb.com/rooms/92228	20230628025456	2023-06-28	city scrape	Rental unit in Buenos Aires · 2 bedrooms · 3 b...	The space Departamento de 3 ambien...	El barrio es muy residencial, con dos supermer...	https://a0.muscache/pictures/b44a/a

5 rows × 75 columns

In [7]:
```
reviews1.head()
```

Out[7]:

	listing_id	id	date	reviewer_id	reviewer_name	comments
0	11508	1615861	2012-07-02	877808	Charlie	Amazing place!\r \r Location: short wa...
1	11508	3157005	2012-12-26	656077	Shaily	Really enjoyed Candela's recommendations and q...
2	11508	3281011	2013-01-05	2835998	Michiel	Candela and her colleague were very attentive ...
3	11508	6050019	2013-07-28	4600436	Tara	The apartment was in a beautiful, modern build...
4	11508	9328455	2013-12-22	3130017	Simon	My stay at Candela's apartment was very enjoya...

In [7]:
```
reviews1.head()
```

Out[7]:

	listing_id	id	date	reviewer_id	reviewer_name	comments
0	11508	1615861	2012-07-02	877808	Charlie	Amazing place!\r \r Location: short wa...
1	11508	3157005	2012-12-26	656077	Shaily	Really enjoyed Candela's recommendations and q...
2	11508	3281011	2013-01-05	2835998	Michiel	Candela and her colleague were very attentive ...
3	11508	6050019	2013-07-28	4600436	Tara	The apartment was in a beautiful, modern build...
4	11508	9328455	2013-12-22	3130017	Simon	My stay at Candela's apartment was very enjoya...

In [8]:
```
listings.head()
```

Out[8]:

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude	room_type	price	minimum_nights	n
0	11508	Condo in Buenos Aires · ★4.81 · 1 bedroom · 1 ...	42762	Candela	NaN	Palermo	-34.581840	-58.424150	Entire home/apt	17339	3	
1	14222	Rental unit in Palermo/Buenos Aires · ★4.78 · ...	87710233	María	NaN	Palermo	-34.586170	-58.410360	Entire home/apt	7807	7	
2	15074	Rental unit in Buenos Aires · 1 bedroom · 1 be...	59338	Monica	NaN	Nuñez	-34.538920	-58.465990	Private room	7665	29	
3	16695	Loft in Buenos Aires · ★4.28 · 1 bedroom · 1 b...	64880	Elbio Mariano	NaN	Monserrat	-34.614390	-58.376110	Entire home/apt	12738	2	
4	20062	Rental unit in Buenos Aires · ★4.93 · 2 bedroo...	75891	Sergio Damian	NaN	Palermo	-34.581848	-58.441605	Entire home/apt	25660	2	

2. Exploración de los datos: **por Columnas**

```
In [14]: calendar.columns, vecindarios.columns, listings.columns, reviews.columns, listings1.columns, reviews1.columns

Out[14]: (Index(['listing_id', 'date', 'available', 'price', 'adjusted_price',
                 'minimum_nights', 'maximum_nights'],
                dtype='object'),
          Index(['neighbourhood_group', 'neighbourhood'], dtype='object'),
          Index(['id', 'name', 'host_id', 'host_name', 'neighbourhood_group',
                 'neighbourhood', 'latitude', 'longitude', 'room_type', 'price',
                 'minimum_nights', 'number_of_reviews', 'last_review',
                 'reviews_per_month', 'calculated_host_listings_count',
                 'availability_365', 'number_of_reviews_ltm', 'license'],
                dtype='object'),
          Index(['listing_id', 'date'], dtype='object'),
          Index(['id', 'listing_url', 'scrape_id', 'last_scraped', 'source', 'name',
                 'description', 'neighborhood_overview', 'picture_url', 'host_id',
                 'host_url', 'host_name', 'host_since', 'host_location', 'host_about',
                 'host_response_time', 'host_response_rate', 'host_acceptance_rate',
                 'host_is_superhost', 'host_thumbnail_url', 'host_picture_url',
                 'host_neighbourhood', 'host_listings_count',
                 'host_total_listings_count', 'host_verifications',
                 'host_has_profile_pic', 'host_identity_verified', 'neighbourhood',
                 'neighbourhood_cleansed', 'neighbourhood_group_cleansed', 'latitude',
                 'longitude', 'property_type', 'room_type', 'accommodates', 'bathrooms',
                 'bathrooms_text', 'bedrooms', 'beds', 'amenities', 'price',
                 'minimum_nights', 'maximum_nights', 'minimum_minimum_nights',
                 'maximum_minimum_nights', 'minimum_maximum_nights',
                 'maximum_maximum_nights', 'minimum_nights_avg_ntm',
                 'maximum_nights_avg_ntm', 'calendar_updated', 'has_availability',
                 'availability_30', 'availability_60', 'availability_90',
                 'availability_365', 'calendar_last_scraped', 'number_of_reviews',
                 'number_of_reviews_ltm', 'number_of_reviews_l30d', 'first_review',
                 'last_review', 'review_scores_rating', 'review_scores_accuracy',
                 'review_scores_cleanliness', 'review_scores_checkin',
                 'review_scores_communication', 'review_scores_location',
                 'review_scores_value', 'license', 'instant_bookable',
                 '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'],
                dtype='object'),
          Index(['listing_id', 'id', 'date', 'reviewer_id', 'reviewer_name', 'comments'], dtype='object'))
```

2.1 Exploración datos: Tamaños

```
In [17]: calendar.shape
Out[17]: (9564447, 7)

In [18]: vecindarios.shape
Out[18]: (49, 2)

In [19]: listings.shape
Out[19]: (26204, 18)

In [20]: reviews.shape
Out[20]: (571860, 2)

In [21]: listings1.shape
Out[21]: (26204, 75)

In [22]: reviews1.shape
Out[22]: (571860, 6)
```

GABRIEL HERNANDEZ REYES
Código 201728589

2.2 Exploración datos: **Datos <u>NULOS</u>**

```
listings.info(), reviews.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 26204 entries, 0 to 26203
Data columns (total 18 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   id                              26204 non-null  int64
 1   name                            26204 non-null  object
 2   host_id                         26204 non-null  int64
 3   host_name                       26204 non-null  object
 4   neighbourhood_group             0 non-null      float64
 5   neighbourhood                   26204 non-null  object
 6   latitude                        26204 non-null  float64
 7   longitude                       26204 non-null  float64
 8   room_type                       26204 non-null  object
 9   price                           26204 non-null  int64
 10  minimum_nights                  26204 non-null  int64
 11  number_of_reviews               26204 non-null  int64
 12  last_review                     21260 non-null  object
 13  reviews_per_month               21260 non-null  float64
 14  calculated_host_listings_count  26204 non-null  int64
 15  availability_365                26204 non-null  int64
 16  number_of_reviews_ltm           26204 non-null  int64
 17  license                         411 non-null    object
dtypes: float64(4), int64(8), object(6)
memory usage: 3.6+ MB
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 571860 entries, 0 to 571859
```

```
calendar.info(), vecindarios.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9564447 entries, 0 to 9564446
Data columns (total 7 columns):
 #   Column          Dtype
---  ------          -----
 0   listing_id      int64
 1   date            object
 2   available       object
 3   price           object
 4   adjusted_price  object
 5   minimum_nights  float64
 6   maximum_nights  float64
dtypes: float64(2), int64(1), object(4)
memory usage: 510.8+ MB
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 49 entries, 0 to 48
Data columns (total 2 columns):
 #   Column               Non-Null Count  Dtype
---  ------               --------------  -----
 0   neighbourhood_group  0 non-null      float64
 1   neighbourhood        49 non-null     object
dtypes: float64(1), object(1)
memory usage: 916.0+ bytes

(None, None)
```

```
listings1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 26204 entries, 0 to 26203
Data columns (total 75 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   id                              26204 non-null  int64
 1   listing_url                     26204 non-null  object
 2   scrape_id                       26204 non-null  int64
 3   last_scraped                    26204 non-null  object
 4   source                          26204 non-null  object
 5   name                            26204 non-null  object
 6   description                     25635 non-null  object
 7   neighborhood_overview           14590 non-null  object
 8   picture_url                     26204 non-null  object
 9   host_id                         26204 non-null  int64
 10  host_url                        26204 non-null  object
 11  host_name                       26204 non-null  object
 12  host_since                      26204 non-null  object
 13  host_location                   20464 non-null  object
 14  host_about                      14886 non-null  object
 15  host_response_time              22233 non-null  object
 16  host_response_rate              22233 non-null  object
 17  host_acceptance_rate            23334 non-null  object
 18  host_is_superhost               24144 non-null  object
 19  host_thumbnail_url              26204 non-null  object
 20  host_picture_url                26204 non-null  object
 21  host_neighbourhood              17866 non-null  object
 22  host_listings_count             26204 non-null  int64
 23  host_total_listings_count       26204 non-null  int64
 24  host_verifications              26204 non-null  object
 25  host_has_profile_pic            26204 non-null  object
 26  host_identity_verified          26204 non-null  object
 27  neighbourhood                   14590 non-null  object
 28  neighbourhood_cleansed          26204 non-null  object
 29  neighbourhood_group_cleansed    0 non-null      float64
 30  latitude                        26204 non-null  float64
 31  longitude                       26204 non-null  float64
 32  property_type                   26204 non-null  object
 33  room_type                       26204 non-null  object
 34  accommodates                    26204 non-null  int64
 35  bathrooms                       0 non-null      float64
 36  bathrooms_text                  26181 non-null  object
 37  bedrooms                        20659 non-null  float64
 38  beds                            25958 non-null  float64
 39  amenities                       26204 non-null  object
 40  price                           26204 non-null  object
 41  minimum_nights                  26204 non-null  int64
 42  maximum_nights                  26204 non-null  int64
 43  minimum_minimum_nights          26204 non-null  int64
 44  maximum_minimum_nights          26204 non-null  int64
 45  minimum_maximum_nights          26204 non-null  int64
 46  maximum_maximum_nights          26204 non-null  int64
 47  minimum_nights_avg_ntm          26204 non-null  float64
 48  maximum_nights_avg_ntm          26204 non-null  float64
 49  calendar_updated                0 non-null      float64
 50  has_availability                26204 non-null  object
 51  availability_30                 26204 non-null  int64
 52  availability_60                 26204 non-null  int64
 53  availability_90                 26204 non-null  int64
 54  availability_365                26204 non-null  int64
 55  calendar_last_scraped           26204 non-null  object
 56  number_of_reviews               26204 non-null  int64
 57  number_of_reviews_ltm           26204 non-null  int64
 58  number_of_reviews_l30d          26204 non-null  int64
 59  first_review                    21260 non-null  object
 60  last_review                     21260 non-null  object
 61  review_scores_rating            21260 non-null  float64
 62  review_scores_accuracy          21176 non-null  float64
 63  review_scores_cleanliness       21176 non-null  float64
 64  review_scores_checkin           21176 non-null  float64
 65  review_scores_communication     21177 non-null  float64
 66  review_scores_location          21177 non-null  float64
 67  review_scores_value             21177 non-null  float64
 68  license                         411 non-null    object
 69  instant_bookable                26204 non-null  object
 70  calculated_host_listings_count  26204 non-null  int64
 71  calculated_host_listings_count_entire_homes  26204 non-null  int64
 72  calculated_host_listings_count_private_rooms  26204 non-null  int64
 73  calculated_host_listings_count_shared_rooms  26204 non-null  int64
 74  reviews_per_month               21260 non-null  float64
dtypes: float64(17), int64(23), object(35)
memory usage: 15.0+ MB
```

```
In [30]: reviews1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 571860 entries, 0 to 571859
Data columns (total 6 columns):
 #   Column         Non-Null Count   Dtype
---  ------         --------------   -----
 0   listing_id     571860 non-null  int64
 1   id             571860 non-null  int64
 2   date           571860 non-null  object
 3   reviewer_id    571860 non-null  int64
 4   reviewer_name  571860 non-null  object
 5   comments       571832 non-null  object
dtypes: int64(3), object(3)
memory usage: 26.2+ MB
```

2.3 Exploración datos: **Distribución de las Variables Numéricas**

```
calendar.describe()
```

	listing_id	minimum_nights	maximum_nights
count	9.564447e+06	9.564415e+06	9.564415e+06
mean	4.208753e+17	6.253008e+00	5.743989e+05
std	3.952350e+17	2.251984e+01	3.509441e+07
min	1.150800e+04	1.000000e+00	1.000000e+00
25%	3.307263e+07	2.000000e+00	3.650000e+02
50%	6.043267e+17	3.000000e+00	1.125000e+03
75%	8.078057e+17	4.000000e+00	1.125000e+03
max	9.232677e+17	1.000000e+03	2.147484e+09

```
vecindarios.describe()
```

	neighbourhood_group
count	0.0
mean	NaN
std	NaN
min	NaN
25%	NaN
50%	NaN
75%	NaN
max	NaN

```
reviews.describe()
```

	listing_id
count	5.718600e+05
mean	1.523683e+17
std	2.933727e+17
min	1.150800e+04
25%	1.502689e+07
50%	3.126255e+07
75%	5.261362e+07
max	9.219057e+17

```
listings.describe()
```

	id	host_id	neighbourhood_group	latitude	longitude	price	minimum_nights	number_of_reviews	r
count	2.620400e+04	2.620400e+04	0.0	26204.000000	26204.000000	2.620400e+04	26204.000000	26204.000000	
mean	4.208766e+17	1.633505e+08	NaN	-34.590773	-58.418323	1.752933e+04	6.300145	21.823386	
std	3.952425e+17	1.686749e+08	NaN	0.018311	0.030168	1.754018e+05	24.286468	38.715892	
min	1.150800e+04	1.342600e+04	NaN	-34.693700	-58.530890	1.750000e+02	1.000000	0.000000	
25%	3.307579e+07	1.796440e+07	NaN	-34.601158	-58.438060	7.406500e+03	2.000000	1.000000	
50%	6.043269e+17	9.768438e+07	NaN	-34.590055	-58.420240	1.019000e+04	3.000000	7.000000	
75%	8.078068e+17	2.641341e+08	NaN	-34.580430	-58.393920	1.528600e+04	4.000000	25.000000	
max	9.232677e+17	5.222350e+08	NaN	-34.534980	-58.355410	2.529509e+07	1000.000000	657.000000	

```
reviews.describe()
```

reviews_per_month	calculated_host_listings_count	availability_365	number_of_reviews_ltm
21260.000000	26204.000000	26204.000000	26204.000000
1.395734	16.062815	213.903335	9.561327
1.384079	33.051039	127.418208	13.972815
0.010000	1.000000	0.000000	0.000000
0.380000	1.000000	89.000000	0.000000
1.000000	2.000000	227.000000	4.000000
2.000000	11.000000	342.000000	14.000000
20.470000	188.000000	365.000000	377.000000

```
listings1.describe()
```

	id	scrape_id	host_id	host_listings_count	host_total_listings_count	neighbourhood_group_cleansed	latitude	longitude
count	2.620400e+04	2.620400e+04	2.620400e+04	26204.000000	26204.000000	0.0	26204.000000	26204.000000
mean	4.208766e+17	2.023063e+13	1.633505e+08	19.949626	27.689399	NaN	-34.590773	-58.418323
std	3.952425e+17	1.164866e+01	1.686749e+08	53.366923	87.616200	NaN	0.018311	0.030168
min	1.150800e+04	2.023063e+13	1.342600e+04	1.000000	1.000000	NaN	-34.693700	-58.530890
25%	3.307579e+07	2.023063e+13	1.796440e+07	1.000000	1.000000	NaN	-34.601158	-58.438060
50%	6.043269e+17	2.023063e+13	9.768438e+07	3.000000	4.000000	NaN	-34.590055	-58.420240
75%	8.078068e+17	2.023063e+13	2.641341e+08	15.000000	18.000000	NaN	-34.580430	-58.393920
max	9.232677e+17	2.023063e+13	5.222350e+08	1837.000000	3334.000000	NaN	-34.534980	-58.355410

accommodates	bathrooms	...	review_scores_cleanliness	review_scores_checkin	review_scores_communication	review_scores_location
26204.000000	0.0	...	21176.000000	21176.000000	21177.000000	21177.000000
2.875248	NaN	...	4.725821	4.861277	4.856843	4.867302
1.480234	NaN	...	0.402382	0.320230	0.330554	0.279053
1.000000	NaN	...	1.000000	1.000000	1.000000	1.000000
2.000000	NaN	...	4.640000	4.850000	4.840000	4.840000
2.000000	NaN	...	4.830000	4.960000	4.970000	4.950000
4.000000	NaN	...	5.000000	5.000000	5.000000	5.000000
16.000000	NaN	...	5.000000	5.000000	5.000000	5.000000

```
listings1.describe()
```

review_scores_value	calculated_host_listings_count	calculated_host_listings_count_entire_homes	calculated_host_listings_count_private_rooms	calculat
21177.000000	26204.000000	26204.000000	26204.000000	
4.717414	16.062815	15.213746	0.622996	
0.400193	33.051039	32.455952	2.693113	
1.000000	1.000000	0.000000	0.000000	
4.640000	1.000000	1.000000	0.000000	
4.820000	2.000000	2.000000	0.000000	
5.000000	11.000000	10.000000	0.000000	
5.000000	188.000000	188.000000	29.000000	

hared_rooms	reviews_per_month
26204.000000	21260.000000
0.082392	1.395734
0.914151	1.384079
0.000000	0.010000
0.000000	0.380000
0.000000	1.000000
0.000000	2.000000
18.000000	20.470000

```
reviews1.describe()
```

	listing_id	id	reviewer_id
count	5.718600e+05	5.718600e+05	5.718600e+05
mean	1.523683e+17	4.313380e+17	1.680009e+08
std	2.933727e+17	3.810591e+17	1.507216e+08
min	1.150800e+04	4.403400e+04	1.000000e+00
25%	1.502689e+07	4.392309e+08	4.006871e+07
50%	3.126255e+07	5.883773e+17	1.220752e+08
75%	5.261362e+07	7.891023e+17	2.564149e+08
max	9.219057e+17	9.239680e+17	5.221917e+08

- **Top 5 de los Atributos más Importantes:**

  i. Vecindarios (bathrooms)
  ii. Tipo de propiedad o habitavion (room_type)
  iii. Disponibilidad (availability – 30 – 60 – 90 – 365 dias)
  iv. Comentarios (reviews)
  v. Precio (Price – price_adjusted)

3. ESTRATEGIA DE ANÁLISIS:

A partir de ciertos hitos y momentos donde de alquiler en alojamientos de larga y mediana estadía, se valida si se condice con el crecimiento de alquileres de propiedades temporales.

- Se realizar un análisis al **comportamiento de las variables categóricas**

In [41]:
```
listings.describe(include=['O'])
```

Out[41]:

	name	host_name	neighbourhood	room_type	last_review	license
count	26204	26204	26204	26204	21260	411
unique	13038	3404	48	4	1495	395
top	Rental unit in Buenos Aires · ★New · 1 bedroom...	Pablo	Palermo	Entire home/apt	2023-06-19	123456
freq	880	382	8800	23556	1018	3

In [42]:
```
listings1.describe(include=['O'])
```

Out[42]:

	listing_url	last_scraped	source	name	description	neighborhood_overview	picture_url	host_url	host_name
count	26204	26204	26204	26204	25635	14590	26204	26204	26204
unique	26204	2	2	13038	23781	12036	25613	13612	3404
top	https://www.airbnb.com/rooms/11508	2023-06-28	city scrape	Rental unit in Buenos Aires · ★New · 1 bedroom...	Disfrutá de la sencillez de este alojamiento t...	Monserrat and San Telmo are the oldest neighbo...	https://a0.muscache.com/pictures/6ca7baf3-0cd5...	https://www.airbnb.com/users/show/1021694	Pablo
freq	1	14706	24079	880	103	47	15	188	382

In [42]:
```
listings1.describe(include=['O'])
```

Out[42]:

host_since	...	room_type	bathrooms_text	amenities	price	has_availability	calendar_last_scraped	first_review	last_review	license	instant_bookal
26204	...	26204	26181	26204	26204	26204	26204	21260	21260	411	262
3976	...	4	44	24615	4274	2	2	3263	1495	395	
2011-08-25	...	Entire home/apt	1 bath	["Wifi", "TV", "Air conditioning", "Kitchen"]	$8,917.00	t	2023-06-28	2023-04-09	2023-06-19	1149457243	
188	...	23556	18045	147	462	24728	14706	153	1018	3	184

```
In [43]: reviews.describe(include=['O'])
```

Out[43]:

	date
count	571860
unique	4394
top	2023-04-09
freq	2398

```
reviews1.describe(include=['O'])
```

	date	reviewer_name	comments
count	571860	571860	571832
unique	4394	54499	543104
top	2023-04-09	Pablo	Excelente
freq	2398	4917	1328

- **Técnicas a Utilizar**

  Se realizar el estudio y análisis de la variable TARGUET donde se agrupan dos categorías principales de los dataset y a la vez se realiza un conteo respecto a otra variable, con el proposito de identificar patrones como vecindarios con últimos y mejores comentaros, así como también la cantidad de vecindario con licenciamiento exigido

```
In [10]: listings.groupby(['neighbourhood','room_type']).count()['last_review']
```

```
Out[10]: neighbourhood     room_type
         Agronomia         Entire home/apt     27
                           Private room         5
         Almagro           Entire home/apt    617
                           Hotel room           0
                           Private room       102
                                              ...
         Villa Santa Rita  Private room         1
         Villa Soldati     Private room         0
         Villa Urquiza     Entire home/apt    222
                           Private room        15
                           Shared room          1
         Name: last_review, Length: 134, dtype: int64
```

```
In [9]: listings.groupby(['neighbourhood','room_type']).count()['license']
```

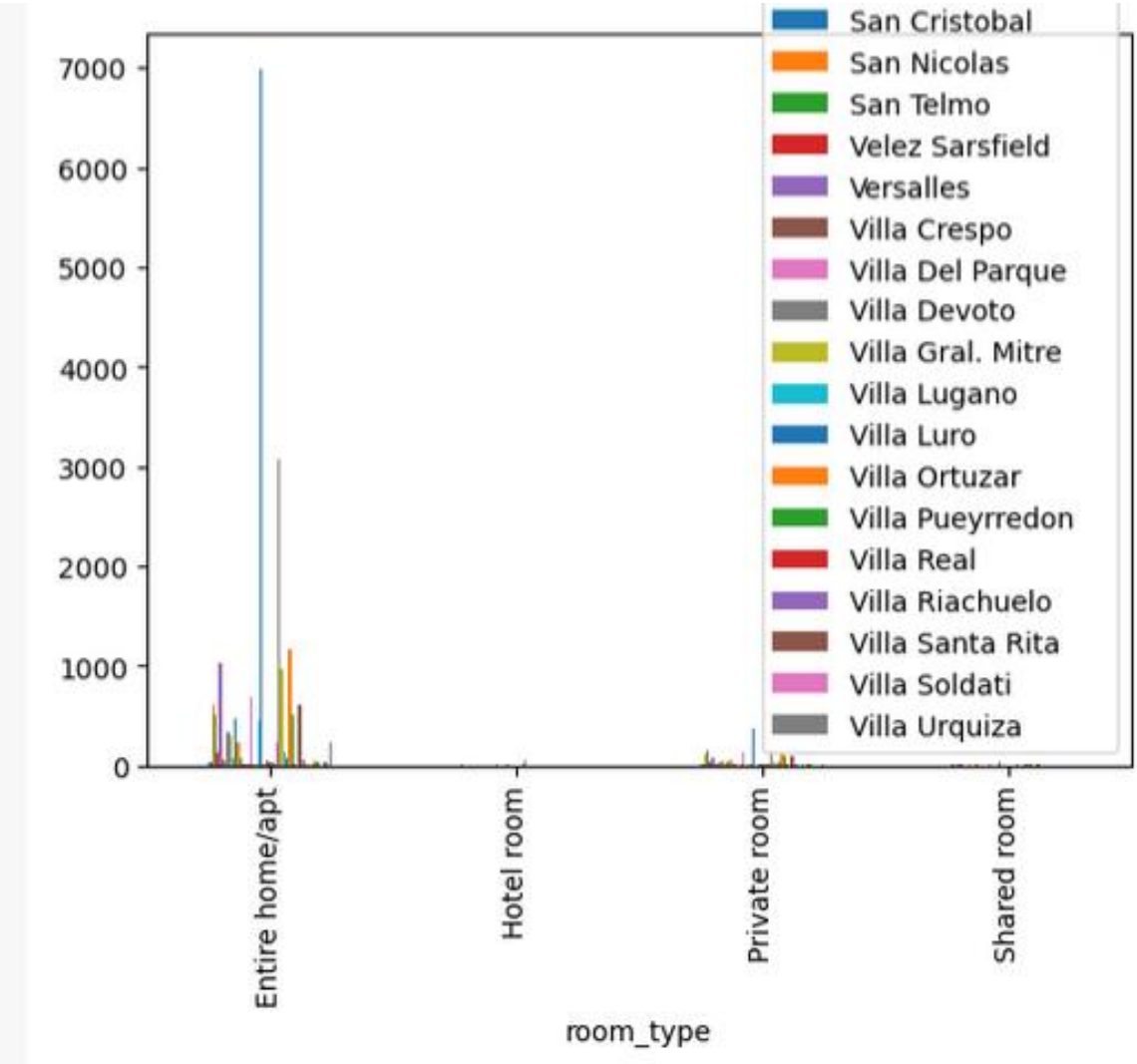
```
Out[9]: neighbourhood     room_type
        Agronomia         Entire home/apt     2
                          Private room        0
        Almagro           Entire home/apt    21
                          Hotel room          0
                          Private room        2
                                             ..
        Villa Santa Rita  Private room        0
        Villa Soldati     Private room        0
        Villa Urquiza     Entire home/apt     7
                          Private room        0
                          Shared room         0
        Name: license, Length: 134, dtype: int64
```

- Seguidamente se crea una **variable de agrupación** empleando UNSTACK de acuerdo a la al análisis previo de la variable TARGUET.

```
In [11]: grouped_bourhood = listings.groupby(['neighbourhood','room_type']).count()['last_review']
         print(grouped_bourhood)
         (grouped_bourhood.unstack(level=0).plot.bar())
         plt.show()
```

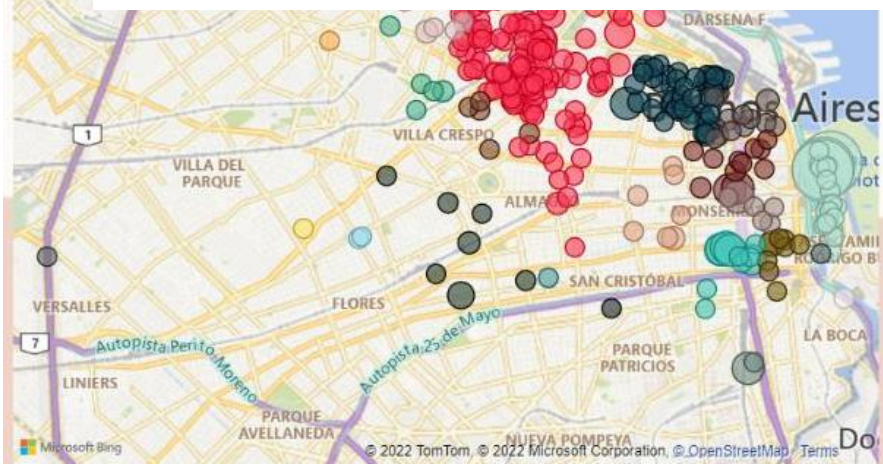
```
neighbourhood     room_type
Agronomia         Entire home/apt     27
                  Private room         5
Almagro           Entire home/apt    617
                  Hotel room           0
                  Private room       102
                                     ...
Villa Santa Rita  Private room         1
Villa Soldati     Private room         0
Villa Urquiza     Entire home/apt    222
                  Private room        15
                  Shared room          1
Name: last_review, Length: 134, dtype: int64
```

Tipo de vivienda / vecindario / mejores comentarios



```
grouped_bourhood = listings.groupby(['neighbourhood','room_type']).count()['license']
print(grouped_bourhood)
(grouped_bourhood.unstack(level=0).plot.bar())
plt.show()
```

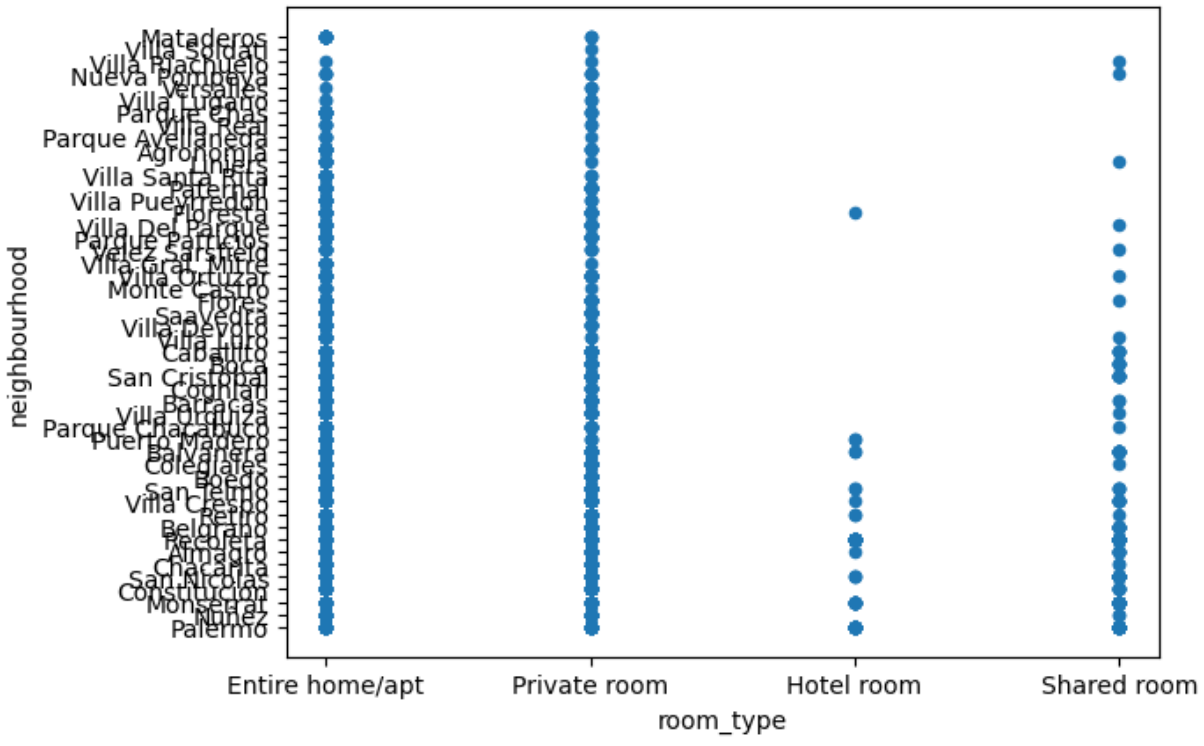
```
neighbourhood     room_type
Agronomia         Entire home/apt     2
                  Private room        0
Almagro           Entire home/apt    21
                  Hotel room          0
                  Private room        2
                                     ..
Villa Santa Rita  Private room        0
Villa Soldati     Private room        0
Villa Urquiza     Entire home/apt     7
                  Private room        0
                  Shared room         0
Name: license, Length: 134, dtype: int64
```



Propiedades más baratas

Barrio	Precio
Nueva	
33344062	$1.230,00
34576896	$1.230,00
14756491	$1.232,00
Parque Avellaneda	
16513075	$1.191,00
23680793	$1.350,00
11396410	$1.356,00
Monte Castro	
37125234	$1.065,00
28952101	$1.479,00
990271	$1.602,00

```
listings.plot.scatter(x="room_type", y="neighbourhood")
plt.show()
```



4. DESARROLLO DE LA ESTRATEGIA:

1. Se realizaron los siguientes análisis:

- **Predictivo**

  Validando la concentración por barrios o vecindarios de la ciudad, se pudo detectar un patrón de zonas de turistas y una tasa de crecimiento promedio de acuerdo a las fechas y comentarios de las personas.

- **Prescriptivo**

  La intención de generar posibles acciones de la ciudad para brindar mayor confiabilidad y comodidad en dichos barrios o vecindarios, así como activar un comercio mayor.

2. Se estudiar las diferentes columnas de los datos, identificando, agrupando y contando las variables principales, teniendo en cuenta los valores usentes o nulos de los dataset,

tanto crear como modificar columnas y generar diversos modelos con los que realizar predicciones.

Se realiza comparación de vecindarios respecto a comentarios y licencias :

```
In [5]: listings.groupby(['neighbourhood']).count()['last_review']
```

```
Out[5]: neighbourhood
Agronomia              32
Almagro               722
Balvanera             669
Barracas              161
Belgrano             1102
Boca                   59
Boedo                  63
Caballito             363
Chacarita             343
Coghlan                68
Colegiales            492
Constitucion          277
Flores                 81
Floresta               21
Liniers                13
Mataderos              14
Monserrat             853
Monte Castro            6
Nueva Pompeya           7
Nuñez                 466
Palermo              7410
Parque Avellaneda       4
Parque Chacabuco       50
Parque Chas            42
Parque Patricios       47
Paternal               24
Puerto Madero         233
Recoleta             3268
Retiro               1004
Saavedra              137
San Cristobal          98
San Nicolas          1293
San Telmo             603
Velez Sarsfield         8
Versalles               1
Villa Crespo          704
Villa Del Parque       60
Villa Devoto           55
Villa Gral. Mitre      10
Villa Lugano            3
Villa Luro             18
Villa Ortuzar          70
Villa Pueyrredon       42
Villa Real              2
Villa Riachuelo         0
Villa Santa Rita       24
Villa Soldati           0
Villa Urquiza         238
Name: last_review, dtype: int64
```

```
In [6]: listings.groupby(['neighbourhood']).count()['license']
```

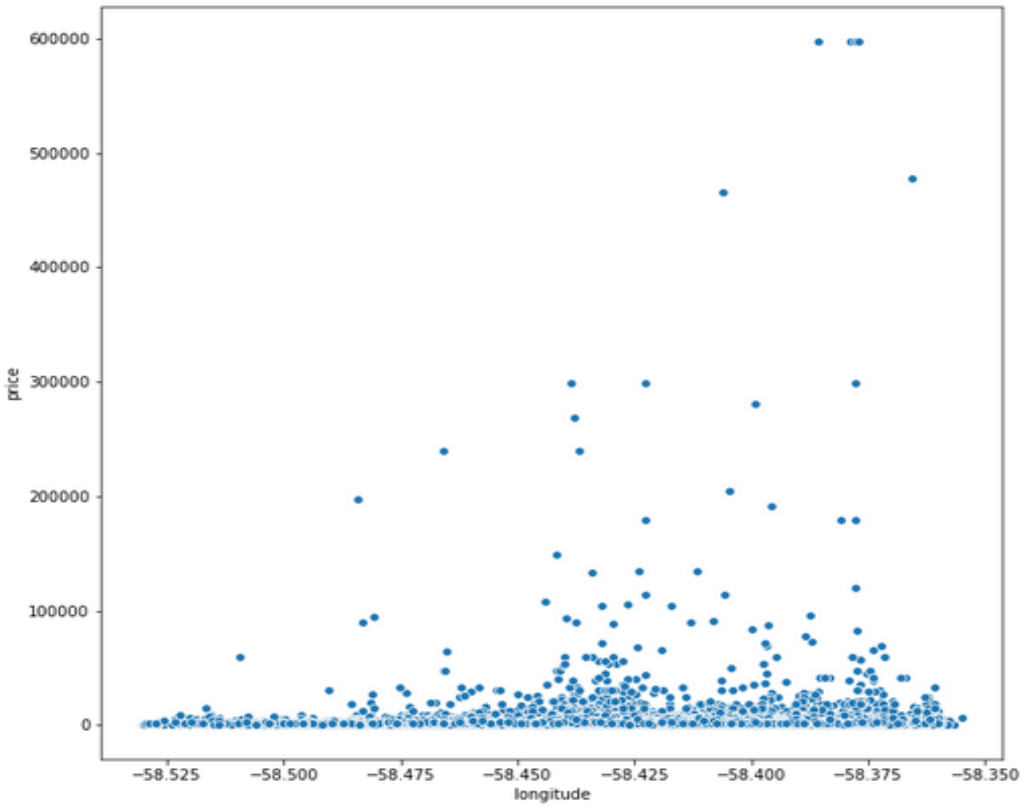
```
Out[6]: neighbourhood
Agronomia              2
Almagro               23
Balvanera             19
Barracas               7
Belgrano              26
Boca                   2
Boedo                  2
Caballito              7
Chacarita              5
Coghlan                1
Colegiales            11
Constitucion           4
Flores                 5
Floresta               0
Liniers                0
Mataderos              0
Monserrat             17
Monte Castro           1
Nueva Pompeya          0
Nuñez                 11
Palermo               95
Parque Avellaneda      0
Parque Chacabuco       2
Parque Chas            0
Parque Patricios       1
Paternal               0
Puerto Madero          3
Recoleta              54
Retiro                23
Saavedra               4
San Cristobal          3
San Nicolas           34
San Telmo             15
Velez Sarsfield        0
Versalles              0
Villa Crespo          18
Villa Del Parque       2
Villa Devoto           3
Villa Gral. Mitre      1
Villa Lugano           0
Villa Luro             0
Villa Ortuzar          2
Villa Pueyrredon       1
Villa Real             0
Villa Riachuelo        0
Villa Santa Rita       0
Villa Soldati          0
Villa Urquiza          7
Name: license, dtype: int64
```

3. Luego se genera una segmentación elegida a los gráficos anteriores por:

   - Tipo de propiedad. VS Barrio.

```
plt.figure(figsize=(10,10))
sns.scatterplot(x='longitude',y='price',data=df)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f19d0ba3d50>
```



Luego en algunos casos especiales se seleccionó filtros como:

- Puntuación de la propiedad.
- Período de tiempo - Año.

4. Transformación de los datos:

   ✓ **Insights Neighborhood:**

   Se eliminó la columna neighborhood_overview ya que no era utilizada.

   ✓ **Insights Review:**

Se modificó el tipo de dato de los siguientes campos: {{"number_of_reviews", Int64.Type},

- "number_of_reviews_ltm", Int64.Type
- "number_of_reviews_l30d", Int64.Type
- "first_review", type date
- "last_review", type date
- "reviews_per_month", type number

✓ **Insights room_type:**

Se modificaron los tipos de campos:

- "id_room", Int64.Type
- "type_room", type text

5. GENERACIÓN DE RESULTADOS:

En Buenos Aires, la ciudad podría observar desde el año 2012 que Airbnb contempla actualmente una gran demanda en Argentina y las cifras siguen aumentando a lo largo de los años.

Como conclusión podemos decir que se han obtenido datos relevantes en el análisis del dashboard tales como:

1. El tipo de propiedad u hospedaje más ofrecido y por ende más solicitado es de un 'Entire home/apt', el cual dispone de un lugar solo para el huésped, sin compartir espacios comunes.
2. Palermo es el barrio con más concentración de hospedajes, por lo que se puede impulsar un crecimiento de negocios en el área.
3. Los precios más altos y como era de esperarse, se ofrecen en Puerto Madero, por lo que podemos delimitar la clase monetaria de huéspedes que puede recibir dicha zona a la hora de plantear nuevos negocios y nevos hospedajes.

Hasta el día de la fecha el año que más registros de anfitriones ha tenido es el 2016, pero lo más drástico a tener en cuenta fue la cantidad de anfitriones por debajo del promedio que se registraron en la pandemia y como esto tuvo un efecto en la oferta de hospedaje. Este dato se encuentra repuntando en el año 2022.Por último se considera que Buenos Aires tiene un crecimiento mucho mayor esperado y con los valores y detalles mencionados podemos proyectar ese crecimiento de la mejor manera posible.