

Tarea BBDD noSQL

Maria Fernanda Lizano Chaves

Descripción y características de la base de datos:

Para realizar este proyecto, se trabajó con una base de datos que contiene los alojamientos disponibles en Airbnb dentro la ciudad de New York, donde se incluye información general y a la vez importante sobre cada uno de los alojamientos, como lo es el barrio y la zona donde se ubican, el precio, mínimo de noches e incluso la valoración promedio que tiene cada alojamiento según las reseñas de sus clientes.

Esta base de datos, se importó a MongoDB – NoSQLBooster, donde se realizaron diversos análisis enfocados a entender el comportamiento general de algunas características, además de diferencias entre una zona y otra o entre un tipo de alojamiento y otro; se realizaron entre otras cosas análisis de los costos de hospedaje y de las valoraciones promedio de los mismos, esto para entender si hay grandes diferencias o si las características de los alojamientos son similares entre sí. Finalmente se busca identificar algunas opciones de alojamiento que cumplan con ciertas condiciones consideradas “ideales” y el costo total del hospedaje para cada opción.

Nota: debido a la gran cantidad de campos que posee esta base de datos, el análisis se limitó a analizar solo algunos de estos campos puesto que se consideran de mucha relevancia al momento de buscar o analizar alojamientos desde el punto de vista turístico como lo son: las reseñas de los clientes, el costo, mínimo de noches requerido, longevidad del inmueble o la ubicación del mismo dentro de la ciudad.

Características y estructura de la base de datos:

- Fuente: <https://www.kaggle.com/datasets>
- Formato: csv
- Cantidad de documentos: 102599
- Cantidad de campos: 25
- Hay missing values?: Si (como parte del análisis, se muestran 2 formas de trabajar con ellos: filtrar esos documentos para evitar trabajar con ellos y rellenar los espacios en blanco con un valor/comentario)
- Diccionario:

Cada alojamiento tiene los siguientes campos:

Id	Identificador único del alojamiento en Airbnb
NAME	Nombre del alojamiento
host id	Identificador del anfitrión del alojamiento en Airbnb
host_identity_verified	Indica si la identidad del anfitrión está verificada o no
host name	Nombre del anfitrión
neighbourhood group	Zona donde está ubicado el alojamiento
neighbourhood	Barrio donde se encuentra
lat	Latitud de la ubicación
long	Longitud de la ubicación
country	País
country code	Código de país
instant_bookable	Indica si el alojamiento tiene opción de reserva inmediata o no
cancellation_policy	Política de cancelación
room type	Tipo de alojamiento
Construction year	Año de construcción
price	Precio por noche del alojamiento
service fee	Cargo por servicio. Este se cobra una sola vez para cada estadía
minimum nights	Mínimo de noches que acepta el anfitrión
number of reviews	Cantidad de reseñas documentadas para el alojamiento
last review	Ultima reseña documentada
reviews per month	Promedio de reseñas mensuales que recibe el alojamiento
review rate number	Valoración promedio del alojamiento
calculated host listings count	Cantidad de alojamientos que posee registrados cada anfitrión
availability 365	Disponibilidad en días durante los próximos 365 días
house_rules	Reglas de la casa

Carga del fichero:

```
localhost master
1 const contents = [
2   {
3     content: "C:\\Users\\ferli\\OneDrive\\Desktop\\BD No SQL\\Data sets\\Airbnb_Open_Data.csv",
4     collection: "Airbnb_Open_Data",
5     idPolicy: "overwrite_with_same_id", //overwrite_with_same_id|always_insert_with_new_id|insert_with_new_id_if_id_exists|skip_documents_with_existing_
6     //Use the transformer to customize the import result
7     //transformer: (doc)=>{ //async (doc)=>{
8       // doc["importDate"] = new Date()
9       // return doc; //return null skips this doc
10    }
11  }
12 ];
13
14 mb.importContent({
15   connection: "localhost",
16   database: "master",
17   fromType: "file",
18   batchSize: 2000,
19   contents
20 })
```

Consultas realizadas para analizar la base de datos

1. Para empezar vamos a agregar una línea con un nuevo lugar en Manhattan, primero se ordena para no repetir ID, y se inserta un nuevo documento con un ID menor:

```
db.Airbnb.find().sort({id: 1})
```

```
db.Airbnb.insertOne({
  "id": 1001056,
  "NAME": "Best place in the City",
  "host id": 1487835973,
  "host_identity_verified": "verified",
  "host name": "Fernanda",
  "neighbourhood group": "Manhattan",
  "neighbourhood": "East Village",
  "lat": 40.7224,
  "long": -73.983,
  "country": "United States",
  "country code": "US",
  "instant_bookable": true,
  "cancellation_policy": "moderate",
  "room type": "Entire home/apt",
  "Construction year": 2007,
  "price": "$725 ",
  "service fee": "$193 ",
  "minimum nights": 2,
  "number of reviews": 1,
  "last review": "11/07/2022",
  "reviews per month": 0.01,
  "review rate number": 5,
  "calculated host listings count": 1,
  "availability 365": 286,
  "house_rules": "No smoking. Parties until 23:00. Pets allowed"
})
```

0.010 s			Tree	
Key	Value	Type		
(1)	{ acknowledged : true, insertedId : ObjectId("636d29898756df2ff2746e19") }	Object		
acknowledged	true	Bool		
insertedId	636d29898756df2ff2746e19	ObjectId		

Airbnb		0.060 s	102,600 Docs		50		p. 1		1 - 50		Table
_id	id	NAME	host id	host_identity_verified	host name	neighbourhood group	neighbourhood	lat	long		
1	636d29898756df2ff2746e19	Best place in the City	1487835973	verified	Fernanda	Manhattan	East Village	40.7224	-73.983		
2	636d29528756df2ff272c1	Clean & quiet apt home by	-15898925	unconfirmed	Madaline	Brooklyn	Kensington	40.6475	-73.9724		
3	636d29528756df2ff272c1	Skyliit Midtown Castle	79556527	verified	Jenna	Manhattan	Midtown	40.7536	-73.9838		

2. Hay algunos lugares que tienen en blanco el campo de reglas de la casa, por lo que vamos a actualizar esos campos para indicar que no hay reglas definidas para dicho lugar:

```
db.Airbnb.updateMany(
  {"house_rules": {$exists: false}},
  {$set: {"house_rules": "No Rules"}}
)
```

Key

Value

Type

(1)

acknowledged

matchedCount

modifiedCount

{ acknowledged : true, matchedCount : 52131, modifiedCount : 52131 }

true

52,131 (52.1K)

52,131 (52.1K)

Object

Bool

Int32

Int32

Airbnb

0.120 s

102,600 Docs

50

p. 1

1 - 50

Table

_id	Constructi	price	service fee	minimum r	number of	last review	reviews per month	review rate	calculated	availability	house_rules	
12	636b80d68756df2ff26e	2004	\$319	\$64	1	160	6/9/2019	1.33	3	4	1	No Rules
13	636b80d68756df2ff26e	2008	\$606	\$121	5	53	6/22/2019	0.43	4	1	163	My ideal guests would be w
14	636b80d68756df2ff26e	2008	\$714	\$143	2	188	6/23/2019	1.5	4	1	258	- One of the bedroom close
15	636b80d68756df2ff26e	2010	\$580	\$116	4	167	6/24/2019	1.34	4	3	47	No Rules
16	636b80d68756df2ff26e	2019	\$149	\$30	2	113	7/5/2019	0.91	3	1	68	No Rules
17	636b80d68756df2ff26e	2018	\$578		90	27	10/31/2018	0.22	3	1	100	Arrival time can be no later
18	636b80d68756df2ff26e	2009	\$778		2	148	6/29/2019	1.2	3	1	197	Absolutely no smoking in th
19	636b80d68756df2ff26e	2006	\$656		2	198	6/28/2019	1.72	5	1	96	- Please be mindful of the n
20	636b80d68756df2ff26e	2008	\$460		1	260	7/1/2019	2.12	3	1	325	No Rules

3. Manhattan es un lugar muy turístico, por lo que nos interesa saber cuántos lugares de hospedaje hay en Manhattan:

```
db.Airbnb.find({"neighbourhood group": "Manhattan"}).count()
```

35	db.Airbnb.find({"neighbourhood group": "Manhattan"}).count()
36	
37	
0.164 s	
1	43793

4. De todos los lugares en Manhattan, cuántos y cuáles alojamientos enteros tienen reserva inmediata, con review rate de 5 (mínimo 100 reviews para aumentar la confianza en el rate) y que permitan reservar una sola noche. Mostrando solamente los campos importantes:

```
db.Airbnb.find(
  {"neighbourhood group": "Manhattan",
   "instant_bookable": true,
   "room type": "Entire home/apt",
   "minimum nights": {$eq:1},
   "review rate number": {$eq:5},
   "number of reviews": {$gt:100}
 },
 {
   "_id":0, "id":1, "NAME":1, "neighbourhood group":1,
   "instant_bookable":1, "room type":1, "price":1,
   "number of reviews":1, "review rate number":1
 })
```

Se obtuvieron 39 resultados:

Airbnb 0.437 s 39 Docs								
	id	NAME	neighbourhood group	instant_bookab	room type	price	number of review	review rate number
1	1697786	Stylish Uptown Westside Apt.	Manhattan	true	Entire home/apt	\$272	102	5
2	2196513	TriBeCa 2500 Sq Ft w/ Priv Elevator	Manhattan	true	Entire home/apt	\$70	447	5
3	6775633	Upscale Apt blocks to Columbia Med./Pr	Manhattan	true	Entire home/apt	\$499	133	5
4	6864553	Luxury, Cozy 1BD, walk to CentrPark	Manhattan	true	Entire home/apt	\$1,188	126	5
5	7975228	Huge 1BD near Central Park Sleeps 5	Manhattan	true	Entire home/apt	\$395	136	5
6	8719729	Beautiful Upper East Side Apartment	Manhattan	true	Entire home/apt	\$527	138	5
7	8779377	East Village 3 Beds, 2 bedrooms. Sleeps	Manhattan	true	Entire home/apt	\$596	101	5
8	8876582	MANHATTAN Studio 10 SECONDS Wal	Manhattan	true	Entire home/apt	\$478	149	5
9	9782354	LIVE HERE!! COZY STUDIO IN CHELS	Manhattan	true	Entire home/apt	\$608	198	5
10	12391419	Sunny Private Studio-Apartment	Manhattan	true	Entire home/apt	\$429	142	5
11	12623385	Nice Private Studio Apartment	Manhattan	true	Entire home/apt	\$707	124	5
12	12647686	Cozy one bedroom apartment	Manhattan	true	Entire home/apt	\$321	145	5
13	13935650	Entire One Bedroom Apt-Great Location	Manhattan	true	Entire home/apt	\$955	120	5
14	18055255	Spacious Stylish 2 br duplex in HK ft for	Manhattan	true	Entire home/apt	\$207	118	5
15	28413090	Cozy Flat in the Heart of the Upper West	Manhattan	true	Entire home/apt	\$932	104	5
16	28641742	Studio in the Heart of Chelsea Neighborh	Manhattan	true	Entire home/apt	\$287	157	5
17	28736737	Midtown great location private Apt	Manhattan	true	Entire home/apt	\$283	346	5
18	28753859	Nice Private Studio Apartment	Manhattan	true	Entire home/apt	\$491	281	5
19	29541439	Charming Studio in a Great location	Manhattan	true	Entire home/apt	\$1,092	219	5
20	30352767	S+ Location Midtown Studio Apt	Manhattan	true	Entire home/apt	\$1,034	418	5
21	30498574	LUXURY APT w PRIVATE GARDEN NY	Manhattan	true	Entire home/apt	\$268	312	5
22	30713971	Cozy Private studio-apartment	Manhattan	true	Entire home/apt	\$766	206	5
23	30774172	Superb Location Midtown Sutudio	Manhattan	true	Entire home/apt	\$482	148	5
24	30829402	Modern Duplex 1 bdrm , 1.5 Baths + terr	Manhattan	true	Entire home/apt	\$571	146	5
25	30879109	Manhattan renovated flat near Central P	Manhattan	true	Entire home/apt	\$648	134	5
26	39351391	Stylish Uptown Westside Apt.	Manhattan	true	Entire home/apt	\$272	102	5
27	39850118	TriBeCa 2500 Sq Ft w/ Priv Elevator	Manhattan	true	Entire home/apt	\$70	447	5
28	41046952	Cute East Village 2 Bedroom Apt.	Manhattan	true	Entire home/apt	\$381	146	5
29	41780959	Spacious Stylish 2 br duplex in HK ft for	Manhattan	true	Entire home/apt	\$207	118	5
30	43963096	Sunny Private Studio-Apartment	Manhattan	true	Entire home/apt	\$429	142	5
31	44195062	Nice Private Studio Apartment	Manhattan	true	Entire home/apt	\$707	124	5
32	44414325	Luxury 2Bed/2Bth Private Loft Best NYC	Manhattan	true	Entire home/apt	\$1,018	106	5
33	47439272	MANHATTAN Studio 10 SECONDS Wal	Manhattan	true	Entire home/apt	\$478	149	5
34	48344492	Great Modern Studio in Chelseal	Manhattan	true	Entire home/apt	\$68	132	5
35	48345044	LIVE HERE!! COZY STUDIO IN CHELS	Manhattan	true	Entire home/apt	\$608	198	5
36	50954110	Sunny Private Studio-Apartment	Manhattan	true	Entire home/apt	\$429	142	5
37	52813151	Spacious Stylish 2 br duplex in HK ft for	Manhattan	true	Entire home/apt	\$207	118	5
38	56855435	Nice Private Studio Apartment	Manhattan	true	Entire home/apt	\$707	124	5
39	56879736	Cozy one bedroom apartment	Manhattan	true	Entire home/apt	\$321	145	5

5. Comparemos: Cuál es el review rate promedio de las habitaciones privadas en Brooklyn y cual es en Manhattan?

```
db.Airbnb.aggregate([
  {$match: {"neighbourhood group":"Brooklyn"}},
  {$group: { _id: "$Promedio", PromBrooklyn: {$avg: "$review rate number"}}
})
db.Airbnb.aggregate([
  {$match: {"neighbourhood group":"Manhattan"}},
  {$group: { _id: "$Promedio", PromManhattan: {$avg: "$review rate number"}}
})
```

Airbnb 0.169 s 1 Doc		
	_id	PromBrooklyn
1	null	3.2591

Airbnb 0.147 s 1 Doc		
	_id	PromManhattan
1	null	3.2766

6. Cuáles son los 5 lugares en Brooklyn con mayor cantidad de reseñas? ¿Cuál es el costo y el review rate de cada uno?

```
db.Airbnb.aggregate([
  {$match: {"neighbourhood group": "Brooklyn"}},
  {$project:
    {"_id":0, "NAME":1, "number of reviews":1, "review rate number":1, "price":1}},
  {$sort:
    {"number of reviews": -1}},
  {$limit: 5}
])
```

Airbnb 0.234 s 5 Docs				
	NAME	price	number of reviews	review rate number
1	Entire New Apartment in Pa	\$847	679	2
2	Williamsburg Penthouse Gt	\$281	648	2
3	Williamsburg Penthouse Hi	\$758	638	4
4	Couples Luxury suite.	\$697	560	2
5	Room with En Suite Bathro	\$769	501	3

7.Cuál es el año de construcción de los alojamientos más antiguos y cual es el de los más nuevos?

```
db.Airbnb.aggregate([
  {$match: {"Construction year": {$exists: true}}},
  {$group: {_id: null, Mas_Viejos: {$min: {$toString: "$Construction year"}}, Mas_Nuevos: {$max:
    {$toString: "$Construction year"}}}},
  {$project: {_id:0}}
])
```

Airbnb 0.716 s 1 Doc		
	Mas_Viejos	Mas_Nuevos
1	2003	2022

*Se hizo una conversión de los resultados a formato string, esto por un tema visual, puesto que la fase group da los resultados en formato número y esto implicaba que se imprimieran los años con separadores de miles (2,003)

Podría decirse que las construcciones de los alojamientos en New York son bastante recientes.

8. Los datos de precio están en formato string, tienen signo \$, además de que tienen un espacio al final del valor y los precios de más de mil tienen una “,” en formato string; por lo que para poder analizarlos, primero es necesario limpiarlos y convertirlos a números. Para esto creo un nuevo campo llamado Precio con los datos de “price” ya optimizados y convertidos a Int.

Para realizar esta tarea, se ha utilizado un update con pipeline de agregación:

```
db.Airbnb.updateMany({},
  [{$set:
    {Precio:
      {$convert: {input:
        {$replaceAll: {input:
          {$trim: {input:
            {$substrBytes: ["$price", 1, 5]}},
          find:",", replacement:""},
          to: "int", onError: null}}}
      ]}}}
  ]})
```

6.381 s

Key	Value	Type
(1)	{ acknowledged : true, matchedCount : 102600, modifiedCount : 102600 }	Object
acknowledged	true	Bool
matchedCount	102,600 (0.10M)	Int32
modifiedCount	102,600 (0.10M)	Int32

Airbnb 0.119 s 102,600 Docs

	_id	price	service fee	minimum	number of	last review	reviews per month	review rat	calculated	availability	house_rules	Precio
1	636d582a8756df2f277f	\$725	\$193	2	1	11/07/2022	0.01	5	1	286	No smoking. Parties until 2:	725
2	636d57ee8756df2f275f	\$966	\$193	10	9	10/19/2021	0.21	4	6	286	Clean up and treat the hom	966
3	636d57ee8756df2f275f	\$142	\$28	30	45	5/21/2022	0.38	4	2	228	Pet friendly but please conf	142
4	636d57ee8756df2f275f	\$620	\$124	3	0			5	1	352	I encourage you to use my	620
5	636d57ee8756df2f275f	\$368	\$74	30	270	7/5/2019	4.64	4	1	322	No Rules	368
6	636d57ee8756df2f275f	\$204	\$41	10	9	11/19/2018	0.1	3	1	289	Please no smoking in the h	204
7	636d57ee8756df2f275f	\$577	\$115	3	74	6/22/2019	0.59	3	1	374	No smoking, please, and no	577
8	636d57ee8756df2f275f	\$71	\$14	45	49	10/5/2017	0.4	5	1	224	Please no shoes in the hou	71
9	636d57ee8756df2f275f	\$1,060	\$212	45	49	10/5/2017	0.4	5	1	219	House Guidelines for our B	1,060 (1.1

9. Calculamos la cantidad de alojamientos y el promedio del costo de estos según el vecindario, utilizando el stage “Buket” del pipeline de agregación:

```
db.Airbnb.aggregate ([
  {$bucket: {
    groupBy: "$neighbourhood group",
    boundaries: ["Bronx", "Brooklyn", "Manhattan", "Queens", "Staten Island"],
    default: "Z_Other",
    output: {
      "Total": {$sum: 1},
      "Prom_Precio" : {$avg: "$Precio"}
    }
  }
}])
```

	_id ↕	Total ↕	Prom_Precio ↕
1	Bronx	2,712 (2.7)	627.7664
2	Brooklyn	41,842 (41)	626.5622
3	Manhattan	43,793 (43)	622.4426
4	Queens	13,267 (13)	630.2076
5	Z_Other	986	625.2434

10. Ahora hacemos el mismo cálculo, pero únicamente para alojamientos enteros para tener una comparación de alojamientos relativamente “equivalentes”:

```
db.Airbnb.aggregate ([
  {$match: {"room type": "Entire home/apt"}},
  {$bucket: {
    groupBy: "$neighbourhood group",
    boundaries: ["Bronx", "Brooklyn", "Manhattan", "Queens", "Staten Island"],
    default: "Z_Other",
    output: {
      "Total": {$sum: 1},
      "Prom_Precio" : {$avg: "$Precio"}
    }
  }
}])
```

Airbnb 0.865 s 5 Docs			
	_id ↕	Total ↕	Prom_Precio ↕
1	Bronx	1,022 (1.0)	620.2326
2	Brooklyn	20,575 (20)	626.8964
3	Manhattan	26,474 (26)	623.1514
4	Queens	5,146 (5.1)	627.0304
5	Z_Other	485	643.2402

11. Es curioso que el precio promedio de los alojamientos enteros sea muy similar al costo promedio total. Vamos a enfocarnos en Manhattan y calcular el precio promedio pero en este caso por tipo de habitación:

```
db.Airbnb.aggregate ([
  {$match: {"neighbourhood group": "Manhattan"}},
  {$bucket: {
    groupBy: "$room type",
    boundaries: ["Entire home/apt", "Hotel room", "Private room", "Shared room"],
    default: "Z_Other",
    output: {
      "Total": {$sum: 1},
      "Prom_Precio" : {$avg: "$Precio"}
    }
  }
}]
```

Airbnb 0.589 s 4 Docs			
	_id	Total	Prom_Precio
1	Entire home/apt	26,474 (26 623.1514)	
2	Hotel room	100	681.87
3	Private room	16,312 (16 620.3679)	
4	Z_Other	907	632.5221

12. Siendo las habitaciones de hotel las más costosas (en promedio), veremos la lista de los 10 hoteles más costosos y su ubicación:

```
db.Airbnb.aggregate([
  {$match: {"room type": "Hotel room"}},
  {$project: {_id:0, "NAME": 1, "Precio": 1, "neighbourhood group": 1, "neighbourhood": 1}},
  {$sort: {"Precio": -1}},
  {$limit: 10}
])
```

Airbnb 0.158 s 10 Docs				
	NAME	neighbourhood group	neighbourhood	Precio
1	Floor 35th in the heart of New York	Manhattan	Theater District	1,193 (1.2K)
2	Dakota Room at Incentra Village Guest House	Manhattan	West Village	1,172 (1.2K)
3	Roomy 1 bedroom suite in midtown Manhattan	Manhattan	Murray Hill	1,172 (1.2K)
4	Brodway- Double Double Room	Manhattan	Hell's Kitchen	1,164 (1.2K)
5	Empire City - King Lux King Room	Manhattan	Hell's Kitchen	1,151 (1.2K)
6	Spacious King Room near Times Square	Manhattan	Theater District	1,150 (1.1K)
7	Hudson Yard - Lux King Room	Manhattan	Hell's Kitchen	1,149 (1.1K)
8	Comfortable studio suite in midtown Manhattan	Manhattan	Murray Hill	1,141 (1.1K)
9	Homey studio in Midtown Manhattan	Manhattan	Murray Hill	1,136 (1.1K)
10	Impressive 1 bedroom near Sloan Kettering	Manhattan	Murray Hill	1,128 (1.1K)

13. Finalmente haremos un ejercicio para estimar el costo aproximado de hospedarnos 4 noches en Manhattan (teniendo en cuenta solo los alojamientos con un mínimo permitido de 4 o menos noches), específicamente en la zona de SoHo y bajo condiciones óptimas:

En alojamientos con un precio menor al promedio, pero con review rate de 5 (en más de 30 reviews), que tengan hosts con identidad verificada y políticas de cancelación flexible. Veamos qué opciones tenemos que cumplan estas características:

```
db.Airbnb.aggregate([
  {$match:
    {"neighbourhood": "SoHo", "minimum nights": {$lte: 4}, "review rate number": {$eq: 5},
    "number of reviews": {$gt: 30}, "host_identity_verified": "verified", "cancellation_policy":
    "flexible", "Precio": {$lt: 622}
  }},
  {$project: {_id: 0, "NAME": 1, "neighbourhood": 1, "review rate number": 1, "price": 1,
    "room type": 1,
    "Costo_Total":
      {$sum: [
        {$toInt:
          {$trim: {input:
            {$replaceAll: {input: "$service fee", find: {$literal: "$"}, replacement: ""}}}},
          {$multiply: ["$Precio", 4]}
        ]}
      }
    }
  })
])
```

Airbnb 0.115 s 4 Docs						
	NAME	neighbourhood	room type	price	review rate number	Costo_Total
1	Right in the middle of it all!	SoHo	Private room	\$468	5	1,966 (2.0K)
2	SoHo! Private Room in Comfy Apt!	SoHo	Private room	\$470	5	1,974 (2.0K)
3	City Queen	SoHo	Hotel room	\$451	5	1,894 (1.9K)

Conclusiones y recomendaciones:

- Hay gran cantidad de alojamientos en New York y no es coincidencia que se escuche con frecuencia decir que esta ciudad es muy cara, los precios promedio de los alojamientos en todas las zonas es bastante elevado!
- A pesar del costo promedio elevado, fue posible visualizar varios alojamientos con precios menores a \$100 dentro de la lista de alojamientos con valoración de 5.
- A pesar de que Manhattan es la zona más exclusiva de New York, la valoración promedio de sus alojamientos es bastante similar a la de Brooklyn, y el precio promedio por noche también lo es! De hecho la diferencia promedio entre zonas es casi insignificante.
- A pesar del punto anterior, podemos ver que los hoteles mas lujosos (o al menos los más costosos, todos están en Manhattan.
- La cantidad de hoteles publicados en Airbnb es bastante baja comparada con la cantidad de habitaciones privadas (en casa o apartamento particular) y alojamientos enteros; lo que nos confirma que este tipo de hospedajes siguen siendo el marco principal de la plataforma.
- SoHo es uno de los barrios más exclusivos de Manhattan, e incluso ahí pudimos encontrar alojamientos cuyo precio estuviera por debajo del promedio y que cumplieran con una serie de exigencias que los hace alojamientos casi ideales.
- Se debe revisar los valores numéricos de las bases de datos: en el caso de la base de datos utilizada, los precios venían con símbolo de dólar, un espacio al final del número y con el signo de coma (,) lo que lo convertía en datos string (y bastante sucios) que debían trabajarse con diversas fórmulas para convertirse a integer.
- La limpieza de datos en general también es importante, en este caso nos encontramos con muchos valores faltantes en diferentes campos.