

```
PS C:\Users\Santiago\Desktop\etl_airbnb> python -m venv venv
```

```
PS C:\Users\Santiago\Desktop\etl_airbnb> .\venv\Scripts\Activate.ps1
```

```
(venv) PS C:\Users\Santiago\Desktop\etl_airbnb> python -m pip install --upgrade pip
Requirement already satisfied: pip in c:\users\santiago\desktop\etl_airbnb\venv\lib\site-packages (25.2)
Collecting pip
  Downloading pip-25.3-py3-none-any.whl.metadata (4.7 kB)
  Downloading pip-25.3-py3-none-any.whl (1.8 MB)
    ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 1.8/1.8 MB 5.5 MB/s 0:00:00
Installing collected packages: pip
  Attempting uninstall: pip
    Found existing installation: pip 25.2
    Uninstalling pip-25.2:
      Successfully uninstalled pip-25.2
  Successfully installed pip-25.3
```

```
(venv) PS C:\Users\Santiago\Desktop\etl_airbnb> pip install -r requirements.txt
```

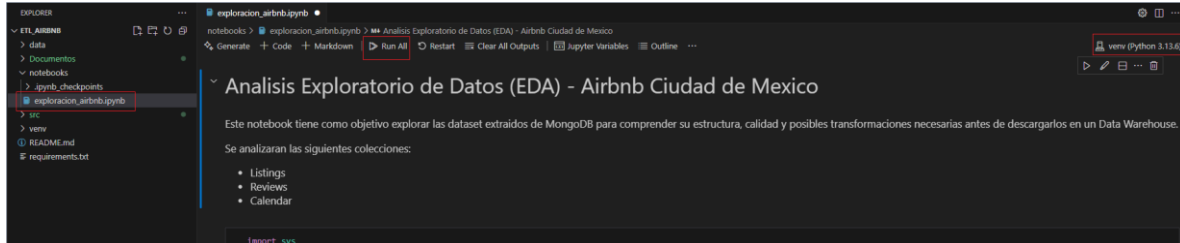
```
(venv) PS C:\Users\Santiago\Desktop\etl_airbnb> Get-Command python | Select-Object Source
```

Source

-----

C:\Users\Santiago\Desktop\etl\_airbnb\venv\Scripts\python.exe

```
(venv) PS C:\Users\Santiago\Desktop\etl_airbnb> 
```



```
[1] ✓ 25.3s Python
...
Conexion exitosa: MX DB
Extraccion 'MX_listings' (459 registros)
Extraccion 'MX_reviews' (218359 registros)
Extraccion 'MX_calendar' (869428 registros)
Listings: (459, 77)
Reviews: (218359, 7)
Calendar: (869428, 6)
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 459 entries, 0 to 458
Data columns (total 77 columns):
 #   Column              Non-Null Count  Dtype
---  -
 0   id                   459 non-null    object
 1   id                   459 non-null    int64
 2   listing_url          459 non-null    object
 3   scrape_id           459 non-null    int64
 4   last_scraped        459 non-null    datetime64[ns]
 5   source              459 non-null    object
 6   name                459 non-null    object
 7   description          449 non-null    object
 8   picture_url         459 non-null    object
 9   host_id             459 non-null    int64
10   host_url            459 non-null    object
11   host_name           459 non-null    object
12   host_since          459 non-null    datetime64[ns]
...
Columnas con listas o diccionarios en df_listings: ['amenities']
Listings duplicados: 0
Reviews duplicados: 0
Calendar duplicados: 0
Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings...
```

Output is truncated. View as a [scrollable element](#) or open in a [text editor](#). Adjust cell output [settings](#).

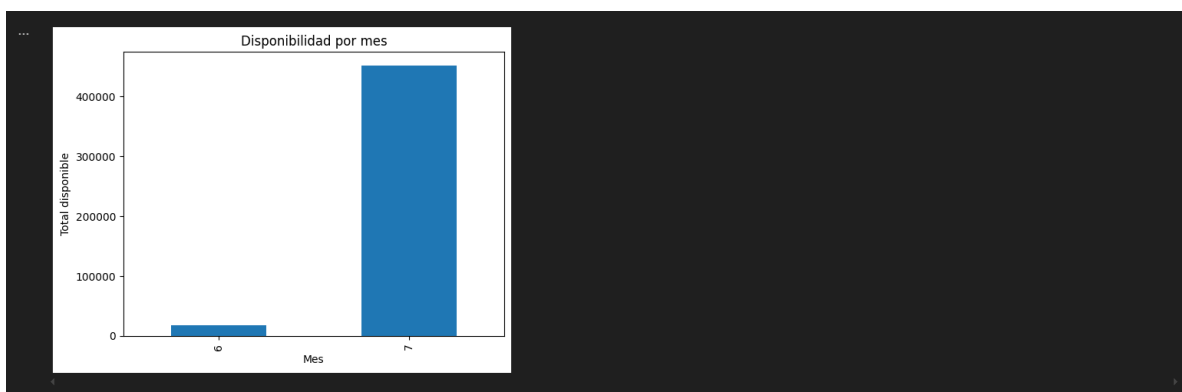
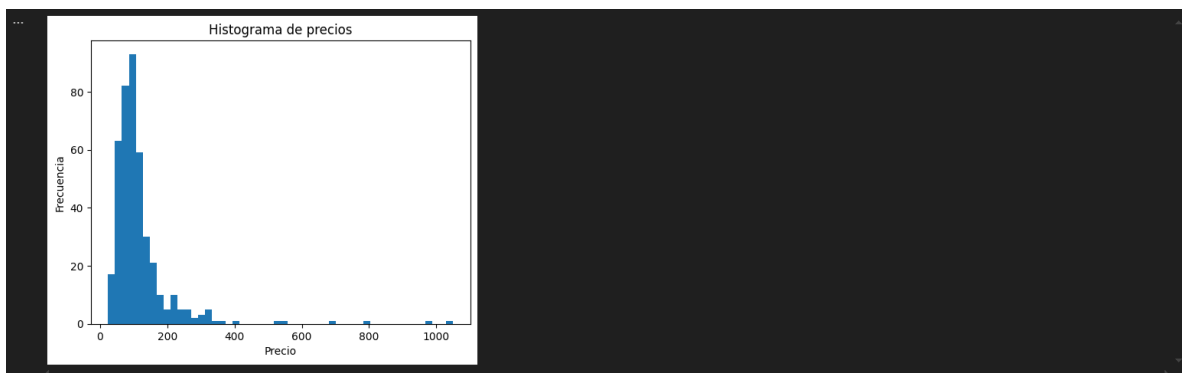
## Conclusiones del Analisis I

- Conexión: Exitosa a la base de datos airbnb.
- Tamaño de los datasets:
  - listings: 26.401 registros y 77 columnas.
  - reviews: 1.388.226 registros y 7 columnas.
  - calendar: 9.636.365 registros y 6 columnas.
- Duplicados: No se detectaron duplicados en ninguna colección.
- Valores nulos:

En listings, varias columnas de host e información descriptiva tienen nulos (host\_about, neighborhood\_overview, review\_scores\_\*), lo que indica posibles campos opcionales.

En reviews y calendar, no hay nulos significativos.

- Tipos de datos: Las fechas están correctamente como datetime64, y los precios aún deben limpiarse (actualmente son object).
- Columnas con estructuras anidadas: host\_verifications y amenities son listas → podrían requerir desanidado o conteo de elementos más adelante.



## Conclusiones del Analisis II

- La mayoría de los precios se concentran en valores bajos, con algunos outliers que podrían corresponder a alojamientos de lujo o errores de registro.
- Los meses con menor disponibilidad (ej. junio-julio) pueden indicar alta demanda en temporada vacacional.
- El dataset contiene columnas con listas (`amenities`, `host_verifications`) que deben ser transformadas si se desea un análisis más profundo.
- Tras la limpieza y validación, los datos están listos para un análisis avanzado o modelado predictivo.

## Hallazgos del EDA

- La columna `price` contiene símbolos que deben limpiarse para convertir a tipo numérico.
- Se detectaron valores extremos en `minimum_nights`, algunos con más de 1000 noches.
- `amenities` es una lista que puede expandirse en columnas binarias para análisis.
- El campo `date` en `calendar` puede usarse para derivar mes, año y trimestre.
- Existen valores nulos en varias columnas que deben tratarse en la fase de transformación.

```
• (venv) PS C:\Users\Santiago\Desktop\etl_airbnb> python src\test_transformacion.py
```

```
=====
```

```
🔧 INICIANDO TRANSFORMACIÓN DE DATOS DESDE MONGODB
```

```
SQLQuery1.sql - DE...8SRF\Santiago (62))* ✕
```

```
CREATE DATABASE Airbnb_DW;
```

```
100 %
```

```
Messages
```

```
Commands completed successfully.
```

```
Completion time: 2025-10-26T18:10:33.6928049-05:00
```

```
src > Logs > logs_extraccion.txt
```

```
1 2025-10-26T18:01:48.278298 | Colección: Conexion | Cantidad: 0 | Estado: Éxito
```

```
2 2025-10-26T18:01:48.349229 | Colección: PK listings | Cantidad: 459 | Estado: Éxito
```

```
(venv) PS C:\Users\Santiago\Desktop\etl_airbnb> python src\test_transformacion.py
```

```
Logs en "logs/test_transformacion_log.txt"
```

```
(venv) PS C:\Users\Santiago\Desktop\etl_airbnb> python src\test_carga.py
```

```
INICIANDO PRUEBA DE CARGA A SQL SERVER
```

```
=====
```

```
Configuración:
```

```
Server: DESKTOP-DE28RPF\SQLEXPRESS
```

```
Base de datos: Airbnb_DW
```

```
Usuario: sam
```

```
=====
```

```
PAZO 1: Cargando datos transformados del test...
```

```
-----
```

```
✓ Listings cargados: 30 registros
```

```
✓ Reviews cargados: 10 registros
```

```
✓ Calendar cargados: 10 registros
```

```
=====
```

```
PAZO 2: Conectando a SQL Server...
```

```
-----
```

```
✓ Conexión exitosa a SQL Server
```

```
=====
```

```
PAZO 3: Cargando datos a SQL Server...
```

```
=====
```

```
▲ Cargando tabla "listings"...
```

```
✓ Datos cargados en tabla "listings" (30 registros)
```

```
✓ tabla "listings" creada con 10 existencias
```

Object Explorer

Connect to: DESKTOP-DE28RPF\SQLEXPRESS (SQL Server 16.0.1000 - DESKTOP)

SQLQuery1.sql - DE...8SRF\Santiago (62))\* ✕

```
--select top 4 * from [dbo].[calendar]
```

```
--select top 4 * from [dbo].[listings]
```

```
--select top 4 * from [dbo].[reviews]
```

id	listing_id	date	available	minimum_nights	maximum_nights	fecha_transformacion
1	68bc31c0051b3a0533a59f	35787	2025-06-26	nan	1	7
2	68bc31c0051b3a0533a59f	35787	2025-06-27	nan	1	7
3	68bc31c0051b3a0533a59f	35787	2025-06-28	nan	1	7
4	68bc31c0051b3a0533a59f	35787	2025-06-29	nan	1	7

id	id	listing_url	source	name	description	picture_url
1	68bc2a73051b3a05339fa0	2952450	https://www.airbnb.com/rooms/2952450	city scrape	Luxury 2 bedroom apartment	The apartment is located in a quiet neighborhood. F...
2	68bc2a73051b3a05339fa0	10762745	https://www.airbnb.com/rooms/10762745	city scrape	Ab hospital area studio both wif. (Pet)	Spacious warm studio in 1940 house close to hospit...
3	68bc2a73051b3a05339fa0	3020211	https://www.airbnb.com/rooms/3020211	city scrape	Forky Urban Gem: Prime Central Location + Parking!	Step into the charming and comfy 1650'1650 square...
4	68bc2a73051b3a05339fa0	5651579	https://www.airbnb.com/rooms/5651579	city scrape	Large studio apt. by Capital Center & ESP@	Spacious studio with hardwood floors, fully equippe...

id	listing_id	id	date	reviewer_id	reviewer_name	comments	comment_length	fecha_transformacion
1	68bc2a73051b3a05339fa0	44816	112485057068038852	2025-01-01	67488775	Wife	Classy charm, great location and lots of space. Good for couples or families wanting peace...	129
2	68bc2a73051b3a05339fa0	56074	136670054384327521	2025-02-27	30291552	Ricardo	Spacious with a great view. Close to metro and an easy five walk to Revolution monument...	132
3	68bc2a73051b3a05339fa0	6044877	1347999418671061277	2025-02-02	2773255	Shayla	I had a wonderful stay. There's a lot of restaurants and bars in the area and it's also very clo...	210
4	68bc2a73051b3a05339fa0	6044877	136033568133302856	2025-02-19	211391901	Yuka	ロケーションもよく部屋も清潔に感じました。ホストは女性で丁寧な対応でとても良かったです。	54