001_load_airbnb

May 16, 2025

```
[1]: # Celda 1: Definición de Librerías
     import pandas as pd
     from sqlalchemy import create_engine
     import os
     from dotenv import load_dotenv
     import logging
```

```
[2]: logging.basicConfig(
         level=logging.INFO,
         format='%(asctime)s - %(levelname)s - %(message)s',
         handlers=[
             logging.FileHandler("001_load_airbnb.log"),
             logging.StreamHandler()
         ]
     )

     logging.info("Inicio del notebook de la carga 001_load_airbnb")
```

```
2025-05-16 22:04:18,379 - INFO - Inicio del notebook de la carga 001_load_airbnb
```

```
[3]: # Celda 3: Definición de Variables Constantes y Carga de .env
     CSV_FILE_PATH = '/home/nicolas/Escritorio/proyecto ETL/develop/data/raw/
      ↪Airbnb_Open_Data.csv'
     ENV_FILE_PATH = '/home/nicolas/Escritorio/proyecto ETL/develop/env/.env'
     TABLE_NAME = 'raw_airbnb' # Nombre de la tabla en PostgreSQL

     logging.info(f"Ruta del archivo CSV: {CSV_FILE_PATH}")
     logging.info(f"Ruta del archivo .env: {ENV_FILE_PATH}")
     logging.info(f"Nombre de la tabla en PostgreSQL: {TABLE_NAME}")

     # Cargar variables de entorno
     load_dotenv(ENV_FILE_PATH)

     POSTGRES_USER = os.getenv('POSTGRES_USER')
     POSTGRES_PASSWORD = os.getenv('POSTGRES_PASSWORD')
     POSTGRES_HOST = os.getenv('POSTGRES_HOST')
     POSTGRES_PORT = os.getenv('POSTGRES_PORT')
```

```
POSTGRES_DATABASE = os.getenv('POSTGRES_DATABASE')

if not all([POSTGRES_USER, POSTGRES_PASSWORD, POSTGRES_HOST, POSTGRES_PORT,␣
↪POSTGRES_DATABASE]):
    logging.error("Una o más variables de entorno de PostgreSQL no están␣
↪definidas. Revisa tu archivo .env.")
    # Puedes decidir si detener el script aquí o continuar con valores por␣
↪defecto (no recomendado para producción)
    raise ValueError("Faltan credenciales de PostgreSQL en el archivo .env")
else:
    logging.info("Variables de entorno para PostgreSQL cargadas correctamente.")
```

```
2025-05-16 22:04:18,386 - INFO - Ruta del archivo CSV:
/home/nicolas/Escritorio/proyecto ETL/develop/data/raw/Airbnb_Open_Data.csv
2025-05-16 22:04:18,387 - INFO - Ruta del archivo .env:
/home/nicolas/Escritorio/proyecto ETL/develop/env/.env
2025-05-16 22:04:18,388 - INFO - Nombre de la tabla en PostgreSQL: raw_airbnb
2025-05-16 22:04:18,389 - INFO - Variables de entorno para PostgreSQL cargadas
correctamente.
2025-05-16 22:04:18,387 - INFO - Ruta del archivo .env:
/home/nicolas/Escritorio/proyecto ETL/develop/env/.env
2025-05-16 22:04:18,388 - INFO - Nombre de la tabla en PostgreSQL: raw_airbnb
2025-05-16 22:04:18,389 - INFO - Variables de entorno para PostgreSQL cargadas
correctamente.
```

[4]:
```
# Celda 4: Pre-definición del DataFrame
df_airbnb = pd.DataFrame()
logging.info("DataFrame 'df_airbnb' predefinido como un DataFrame vacío.")
```

```
2025-05-16 22:04:18,396 - INFO - DataFrame 'df_airbnb' predefinido como un
DataFrame vacío.
```

[5]:
```
# Celda 5: Carga del DataFrame desde el archivo CSV
try:
    logging.info(f"Intentando cargar el archivo CSV: {CSV_FILE_PATH}")
    df_airbnb = pd.read_csv(CSV_FILE_PATH, low_memory=False)
    logging.info(f"Archivo CSV '{CSV_FILE_PATH}' cargado exitosamente.")
    logging.info(f"El DataFrame tiene {df_airbnb.shape[0]} filas y {df_airbnb.
↪shape[1]} columnas.")
except FileNotFoundError:
    logging.error(f"Error: Archivo CSV no encontrado en '{CSV_FILE_PATH}'")
    raise
except Exception as e:
    logging.error(f"Ocurrió un error al cargar el CSV '{CSV_FILE_PATH}': {e}")
    raise
```

```
2025-05-16 22:04:18,402 - INFO - Intentando cargar el archivo CSV:
```

/home/nicolas/Escritorio/proyecto ETL/develop/data/raw/Airbnb_Open_Data.csv
2025-05-16 22:04:18,835 - INFO - Archivo CSV '/home/nicolas/Escritorio/proyecto
ETL/develop/data/raw/Airbnb_Open_Data.csv' cargado exitosamente.
2025-05-16 22:04:18,835 - INFO - El DataFrame tiene 102599 filas y 26 columnas.

[6]:
```python
# Celda 6: Verificación de la carga del DataFrame
if not df_airbnb.empty:
    logging.info("Mostrando las primeras 5 filas del DataFrame df_airbnb
↪(formato markdown):")
    logging.info(f"\n{df_airbnb.head().to_markdown(index=False)}")
else:
    logging.warning("El DataFrame df_airbnb está vacío. No se puede mostrar el
↪head.")
```

2025-05-16 22:04:18,840 - INFO - Mostrando las primeras 5 filas del DataFrame
df_airbnb (formato markdown):
2025-05-16 22:04:18,847 - INFO -

id	NAME	host id	host_identity_verified	host name	neighbourhood group	neighbourhood	lat	long	country	country code	instant_bookable	cancellation_policy	room type	Construction year	price	service fee	minimum nights	number of reviews	last review	reviews per month	review rate number	calculated host listings count	availability 365	house_rules	license
1001254	Clean & quiet apt home by the park	80014485718	unconfirmed	Madaline	Brooklyn	Kensington	40.6475	-73.9724	United States	US	False	strict	Private room	2020	$966	$193	10	9	10/19/2021	0.21	4	6	286	Clean up and treat the home the way you'd like your home to be treated. No smoking.	nan
1002102	Skylit Midtown Castle	52335172823	verified	Jenna	Manhattan	Midtown	40.7536	-73.9838	United States	US	False

3

moderate | Entire home/apt | 2007 | $142 | $28
| 30 | 45 | 5/21/2022 | 0.38 |
4 | 2 | 228 | Pet friendly but
please confirm with me if the pet you are planning on bringing with you is OK. I
have a cute and quiet mixed chihuahua. I could accept more guests (for an extra
fee) but this also needs to be confirmed beforehand. Also friends traveling
together could sleep in separate beds for an extra fee (the second bed is either
a sofa bed or inflatable bed). Smoking is only allowed on the porch. | nan
|
| 1002403 | THE VILLAGE OF HARLEM…NEW YORK ! | 78829239556 | nan
| Elise | Manhattan | Harlem | 40.809 | -73.9419 |
United States | US | True | flexible |
Private room | 2005 | $620 | $124 |
3 | 0 | nan | nan |
5 | 1 | 352 | I encourage you to
use my kitchen, cooking and laundry facilities. There is no additional charge to
use the washer/dryer in the basement. No smoking, inside or outside. Come home
as late as you want. If you come home stumbling drunk, it's OK the first time.
If you do it again, and you wake up me or the neighbors downstairs, we will be
annoyed. (Just so you know . . .) | nan |
| 1002755 | nan | 85098326012 |
unconfirmed | Garry | Brooklyn | Clinton Hill
| 40.6851 | -73.9598 | United States | US | True |
moderate | Entire home/apt | 2005 | $368 | $74
| 30 | 270 | 7/5/2019 | 4.64 |
4 | 1 | 322 | nan
| nan |
| 1003689 | Entire Apt: Spacious Studio/Loft by central park | 92037596077 |
verified | Lyndon | Manhattan | East Harlem
| 40.7985 | -73.944 | United States | US | False |
moderate | Entire home/apt | 2009 | $204 | $41
| 10 | 9 | 11/19/2018 | 0.1 |
3 | 1 | 289 | Please no smoking in
the house, porch or on the property (you can go to the nearby corner).
Reasonable quiet after 10:30 pm. Please remove shoes in the house.
| nan |
2025-05-16 22:04:18,847 - INFO -
| id | NAME | host id |
host_identity_verified | host name | neighbourhood group | neighbourhood
| lat | long | country | country code | instant_bookable |
cancellation_policy | room type | Construction year | price |
service fee | minimum nights | number of reviews | last review |
reviews per month | review rate number | calculated host listings count |
availability 365 | house_rules
| license |
|--------:|:---|------------:|:---
---------------------|:-----------|:---------------------|:--------------|-
-------:|---------:|:-------------|:--------------|:------------------|:-----

id	NAME	host id	host_identity_verified	host name	neighbourhood group	neighbourhood	lat	long	country	country code	instant_bookable	cancellation_policy	room type	Construction year	price	service fee	minimum nights	number of reviews	last review	reviews per month	review rate number	calculated host listings count	availability 365	house_rules	license
1001254	Clean & quiet apt home by the park	80014485718	unconfirmed	Madaline	Brooklyn	Kensington	40.6475	-73.9724	United States	US	False	strict	Private room	2020	$966	$193	10	9	10/19/2021	0.21	4	6	286	Clean up and treat the home the way you'd like your home to be treated. No smoking.	nan
1002102	Skylit Midtown Castle	52335172823	verified	Jenna	Manhattan	Midtown	40.7536	-73.9838	United States	US	False	moderate	Entire home/apt	2007	$142	$28	30	45	5/21/2022	0.38	4	2	228	Pet friendly but please confirm with me if the pet you are planning on bringing with you is OK. I have a cute and quiet mixed chihuahua. I could accept more guests (for an extra fee) but this also needs to be confirmed beforehand. Also friends traveling together could sleep in separate beds for an extra fee (the second bed is either a sofa bed or inflatable bed). Smoking is only allowed on the porch.	nan
1002403	THE VILLAGE OF HARLEM…NEW YORK !	78829239556	nan	Elise	Manhattan	Harlem	40.809	-73.9419	United States	US	True	flexible	Private room	2005	$620	$124	3	0	nan	nan	5	1	352	I encourage you to use my kitchen, cooking and laundry facilities. There is no additional charge to use the washer/dryer in the basement. No smoking, inside or outside. Come home as late as you want. If you come home stumbling drunk, it's OK the first time. If you do it again, and you wake up me or the neighbors downstairs, we will be annoyed. (Just so you know . . .)	nan
1002755	nan	85098326012	unconfirmed	Garry	Brooklyn	Clinton Hill	40.6851	-73.9598	United States	US	True	moderate	Entire home/apt	2005	$368	$74	30	270	7/5/2019	4.64	4	1	322	nan	nan
1003689	Entire Apt: Spacious Studio/Loft by central park	92037596077																							

```
verified                    | Lyndon    | Manhattan              | East Harlem
| 40.7985 | -73.944  | United States | US                 | False            |
moderate                    | Entire home/apt |                  2009 | $204    | $41
|              10 |                      9 | 11/19/2018        |                  0.1  |
3 |                                      1 |                    289 | Please no smoking in
the house, porch or on the property (you can go to the nearby corner).
Reasonable quiet after 10:30 pm.  Please remove shoes in the house.
|        nan |
```

[7]:
```python
# Celda 7: Crear motor de conexión a PostgreSQL
if not df_airbnb.empty:
    try:
        DATABASE_URL = f"postgresql+psycopg2://{POSTGRES_USER}:
 ↪{POSTGRES_PASSWORD}@{POSTGRES_HOST}:{POSTGRES_PORT}/{POSTGRES_DATABASE}"
        engine = create_engine(DATABASE_URL)
        logging.info("Motor de SQLAlchemy para PostgreSQL creado.")

        with engine.connect() as connection:
            logging.info("Conexión a la base de datos PostgreSQL establecida y
 ↪verificada exitosamente.")
    except Exception as e:
        logging.error(f"Error al crear el motor de SQLAlchemy o al conectar a
 ↪PostgreSQL: {e}")
        engine = None
else:
    logging.warning("El DataFrame está vacío, se omite la creación del motor de
 ↪DB.")
```

```
2025-05-16 22:04:18,871 - INFO - Motor de SQLAlchemy para PostgreSQL creado.
WARNING:  la base de datos «airbnb» tiene una discordancia de versión de
ordenamiento ("collation")
DETAIL:  La base de datos fue creada usando la versión de ordenamiento 2.31,
pero el sistema operativo provee la versión 2.35.
HINT:  Reconstruya todos los objetos en esta base de datos que usen el
ordenamiento por omisión y ejecute ALTER DATABASE airbnb REFRESH COLLATION
VERSION, o construya PostgreSQL con la versión correcta de la biblioteca.
2025-05-16 22:04:18,881 - INFO - Conexión a la base de datos PostgreSQL
establecida y verificada exitosamente.
WARNING:  la base de datos «airbnb» tiene una discordancia de versión de
ordenamiento ("collation")
DETAIL:  La base de datos fue creada usando la versión de ordenamiento 2.31,
pero el sistema operativo provee la versión 2.35.
HINT:  Reconstruya todos los objetos en esta base de datos que usen el
ordenamiento por omisión y ejecute ALTER DATABASE airbnb REFRESH COLLATION
VERSION, o construya PostgreSQL con la versión correcta de la biblioteca.
2025-05-16 22:04:18,881 - INFO - Conexión a la base de datos PostgreSQL
establecida y verificada exitosamente.
```

```python
[8]: # Celda 8: Crear tabla vacía en PostgreSQL (si no existe o reemplazarla)
if not df_airbnb.empty and 'engine' in locals() and engine is not None:
    try:
        logging.info(f"Intentando crear/reemplazar la tabla vacía␣
   ↪'{TABLE_NAME}' en PostgreSQL.")
        df_airbnb.head(0).to_sql(TABLE_NAME, engine, if_exists='replace',␣
   ↪index=False)
        logging.info(f"Tabla '{TABLE_NAME}' creada/reemplazada exitosamente␣
   ↪(estructura definida, sin datos).")
    except Exception as e:
        logging.error(f"Error al crear la tabla '{TABLE_NAME}' en PostgreSQL:␣
   ↪{e}")
else:
    if df_airbnb.empty:
        logging.warning("DataFrame está vacío. No se intentará crear la tabla␣
   ↪en la base de datos.")
    if 'engine' not in locals() or engine is None:
        logging.warning("Motor de base de datos no inicializado. No se␣
   ↪intentará crear la tabla.")
```

```
2025-05-16 22:04:18,889 - INFO - Intentando crear/reemplazar la tabla vacía
'raw_airbnb' en PostgreSQL.
2025-05-16 22:04:18,918 - INFO - Tabla 'raw_airbnb' creada/reemplazada
exitosamente (estructura definida, sin datos).
```

```python
[9]: # Celda 9: Insertar el DataFrame en la tabla de PostgreSQL
if not df_airbnb.empty and 'engine' in locals() and engine is not None:
    try:
        logging.info(f"Intentando insertar datos del DataFrame en la tabla␣
   ↪'{TABLE_NAME}'.")
        df_airbnb.to_sql(TABLE_NAME, engine, if_exists='append', index=False)
        logging.info(f"Datos insertados exitosamente en la tabla '{TABLE_NAME}'.
   ↪")
    except Exception as e:
        logging.error(f"Error al insertar datos en la tabla '{TABLE_NAME}':␣
   ↪{e}")
else:
    if df_airbnb.empty:
        logging.warning("DataFrame está vacío. No hay datos para insertar en la␣
   ↪base de datos.")
    if 'engine' not in locals() or engine is None:
        logging.warning("Motor de base de datos no inicializado. No se pueden␣
   ↪insertar datos.")
```

```
2025-05-16 22:04:18,924 - INFO - Intentando insertar datos del DataFrame en la
tabla 'raw_airbnb'.
2025-05-16 22:04:24,674 - INFO - Datos insertados exitosamente en la tabla
```

'raw_airbnb'.

```python
# Celda 10: Verificar los primeros 5 datos desde la base de datos PostgreSQL
if 'engine' in locals() and engine is not None:
    try:
        logging.info(f"Consultando las primeras 5 filas de la tabla
'{TABLE_NAME}' desde PostgreSQL.")
        query = f"SELECT * FROM \"{TABLE_NAME}\" LIMIT 5;"
        df_from_db = pd.read_sql_query(query, engine)
        logging.info("Primeras 5 filas obtenidas de PostgreSQL (formato
markdown):")

        if not df_from_db.empty:
            print(df_from_db.to_markdown(index=False))
        else:
            logging.info("La consulta no devolvió filas o la tabla está vacía.")

    except Exception as e:
        logging.error(f"Error al consultar datos desde la tabla '{TABLE_NAME}'
en PostgreSQL: {e}")
else:
    logging.warning("Motor de base de datos no inicializado. No se puede
verificar la tabla.")
```

```
2025-05-16 22:04:24,680 - INFO - Consultando las primeras 5 filas de la tabla
'raw_airbnb' desde PostgreSQL.
2025-05-16 22:04:24,684 - INFO - Primeras 5 filas obtenidas de PostgreSQL
(formato markdown):
```

```
|      id | NAME                                               |      host id |
host_identity_verified   | host name    | neighbourhood group   | neighbourhood
|     lat |     long | country       | country code  | instant_bookable   |
cancellation_policy   | room type      |   Construction year | price   |
service fee   |   minimum nights |   number of reviews | last review   |
reviews per month |   review rate number |   calculated host listings count |
availability 365 | house_rules
| license   |
|--------:|:--------------------------------------------------|------------:|:---
---------------------|:------------|:----------------------|:---------------|-
-------:|---------:|:-------------|:-------------|:------------------|:-----
----------------|:------------|-------------------:|:-------|:-----------
---|----------------:|-------------------:|:-------------|------------------
-:|---------------------:|-----------------------------:|------------------
:|:------------------------------------------------------------------------
------------------------------------------------------------------------------
------------------------------------------------------------------------------
------------------------------------------------------------------------------
------------------------------------------------------------------------------
```

-------|:----------|

id	NAME	host id	host_identity_verified	host name	neighbourhood group	neighbourhood	lat	long	country	country code	instant_bookable	cancellation_policy	room type	Construction year	price	service fee	minimum nights	number of reviews	last review	reviews per month	review rate number	calculated host listings count	availability 365	house_rules
1001254	Clean & quiet apt home by the park	80014485718	unconfirmed	Madaline	Brooklyn	Kensington	40.6475	-73.9724	United States	US	false	strict	Private room	2020	$966	$193	10	9	10/19/2021	0.21	4	6	286	Clean up and treat the home the way you'd like your home to be treated. No smoking.
1002102	Skylit Midtown Castle	52335172823	verified	Jenna	Manhattan	Midtown	40.7536	-73.9838	United States	US	false	moderate	Entire home/apt	2007	$142	$28	30	45	5/21/2022	0.38	4	2	228	Pet friendly but please confirm with me if the pet you are planning on bringing with you is OK. I have a cute and quiet mixed chihuahua. I could accept more guests (for an extra fee) but this also needs to be confirmed beforehand. Also friends traveling together could sleep in separate beds for an extra fee (the second bed is either a sofa bed or inflatable bed). Smoking is only allowed on the porch.
1002403	THE VILLAGE OF HARLEM…NEW YORK !	78829239556		Elise	Manhattan	Harlem	40.809	-73.9419	United States	US	true	flexible	Private room	2005	$620	$124	3	0		nan	5	1	352	I encourage you to use my kitchen, cooking and laundry facilities. There is no additional charge to use the washer/dryer in the basement. No smoking, inside or outside. Come home as late as you want. If you come home stumbling drunk, it's OK the first time. If you do it again, and you wake up me or the neighbors downstairs, we will be annoyed. (Just so you know . . .)
1002755		85098326012	unconfirmed	Garry	Brooklyn	Clinton Hill	40.6851	-73.9598	United States	US	true	moderate	Entire home/apt	2005	$368	$74	30	270	7/5/2019	4.64	4	1	322	
1003689	Entire Apt: Spacious Studio/Loft by central park	92037596077	verified	Lyndon	Manhattan	East Harlem	40.7985	-73.944	United States	US	false	moderate	Entire home/apt	2009	$204	$41	10	9	11/19/2018	0.1	3	1	289	Please no smoking in the house, porch or on the property (you can go to the nearby corner). Reasonable quiet after 10:30 pm. Please remove shoes in the house.

```
[11]: # Celda 11: Finalización (Opcional: cerrar conexiones si es necesario, aunque
      ↪SQLAlchemy suele manejarlas)
      logging.info("Script de carga de datos finalizado.")
      if 'engine' in locals() and engine is not None:
          engine.dispose()
          logging.info("Conexiones del motor de SQLAlchemy dispuestas (cerradas).")
      else:
          logging.info("Celda 11: Script finalizado.")
```

```
2025-05-16 22:04:24,692 - INFO - Script de carga de datos finalizado.
2025-05-16 22:04:24,693 - INFO - Conexiones del motor de SQLAlchemy dispuestas
(cerradas).
2025-05-16 22:04:24,693 - INFO - Conexiones del motor de SQLAlchemy dispuestas
(cerradas).
```