



ESCUELA POLITÉCNICA NACIONAL

ESCUELA DE FORMACIÓN DE TECNÓLOGOS



ANÁLISIS DE DATOS

ASIGNATURA:

Análisis de Datos

PROFESOR:

Ing. Lorena Chulde / Ing. Juan Pablo Zaldumbide

PERÍODO ACADÉMICO:

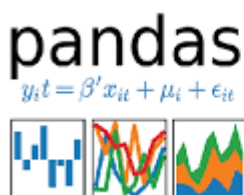
2023-B

TALLER DE REFUERZO

ESTUDIANTES

Marcelo Pinzón

Silvia Chaluisa



	SanName	MaxWeightMg	WeightMg
0	Castañeda	450	450
1	Blue	510	450
2	Green	580	530
3	Red	670	640
4	Newborn	880	850
5	ShortHorned_1	1850	1570
6	ShortHorned_2	2260	2110
7	Gray	1300	1360

2023 - B

Objetivo:

Consultar y manipular varios data set mediante bases de datos SQL y NoSQL locales y en la nube para la descripción y análisis de la información de los clientes de todas las bases de la empresa inmobiliaria “El Búho”.

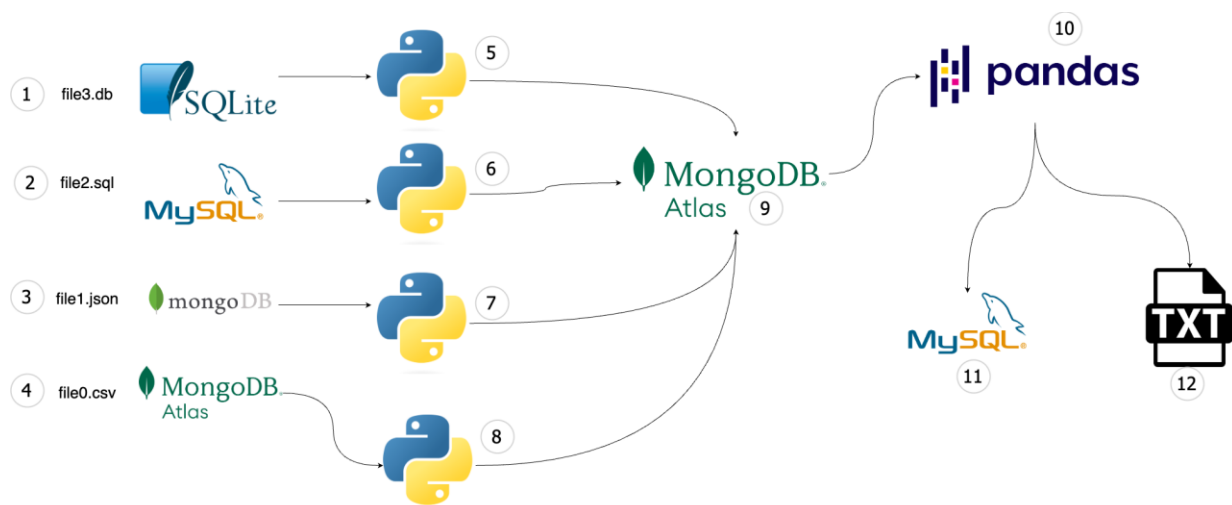
Objetivos específicos:

- Importar y exportar archivos desde varias Bases de Datos
- Integrar los data set de las cuatro sucursales mediante MongoDB Atlas para la accesibilidad, manipulación y análisis de los datos.

Planteamiento del problema:

La empresa inmobiliaria “El Búho” tiene como objetivo averiguar la información de sus clientes, lastimosamente sus datos están divididos en las bases de datos de sus cuatro sucursales.

Dada la siguiente arquitectura:



Indicaciones:

1. SQLite

- Importar file3.db desde la base de datos SQLite.
- Exportar con la extensión que crea conveniente

Importar las bibliotecas necesarias, crear la conexión con la BDD, extraer la información, convertirla a DataFrame y guardarla en un archivo con formato CSV

```
Click here to ask Blackbox to help you code faster |
# Importar las bibliotecas necesarias
import os
import pandas as pd
import sqlite3
# Utilizar la ruta de trabajo actual
ruta=os.getcwd()
rutaarch=ruta+"/Archivos"
# Verificar la existencia de la ruta en la que se guardaran los DataFrames exportados como archivos CSV o JSON
if not os.path.exists(rutaarch):
    os.makedirs(rutaarch)

a. Importar file3.db desde la base de datos SQLite y exportarlo como archivo CSV.

Click here to ask Blackbox to help you code faster |
# Conexion con la base de datos
conexion=sqlite3.connect(ruta+"/file3.db")
consulta="SELECT * FROM TABLE_NAME"
datos=pd.read_sql_query(consulta, conexion)
conexion.commit()
conexion.close()
# Guardar el DataFrame como archivo CSV
datos.to_csv(rutaarch+"/Datos file3.csv", index=False)
```

o > Documentos > GitHub > Analisis-de-datos > Repaso_1B

Nombre	Estado	Fecha de modificación	Tipo	Tamaño
.ipynb_checkpoints	✓	21/12/2023 8:24 PM	Carpeta de archivos	
Archivos	✓	24/12/2023 6:53 PM	Carpeta de archivos	
Act_repaso.ipynb	✓	24/12/2023 6:55 PM	Archivo de origen ...	17 KB
file0.csv	✓	20/12/2023 7:13 PM	Archivo de valores...	51 KB
file1.json	✓	20/12/2023 7:13 PM	Archivo de origen ...	91 KB
file2.sql	✓	20/12/2023 7:13 PM	SQL Text File	85 KB
file3.db	✓	20/12/2023 7:13 PM	Data Base File	280 KB

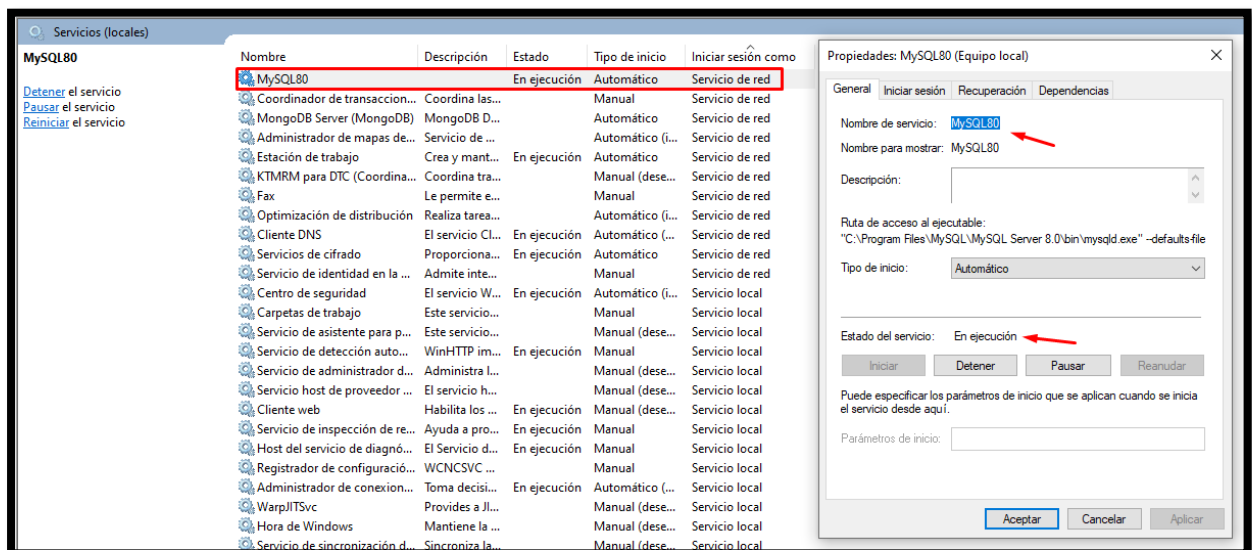
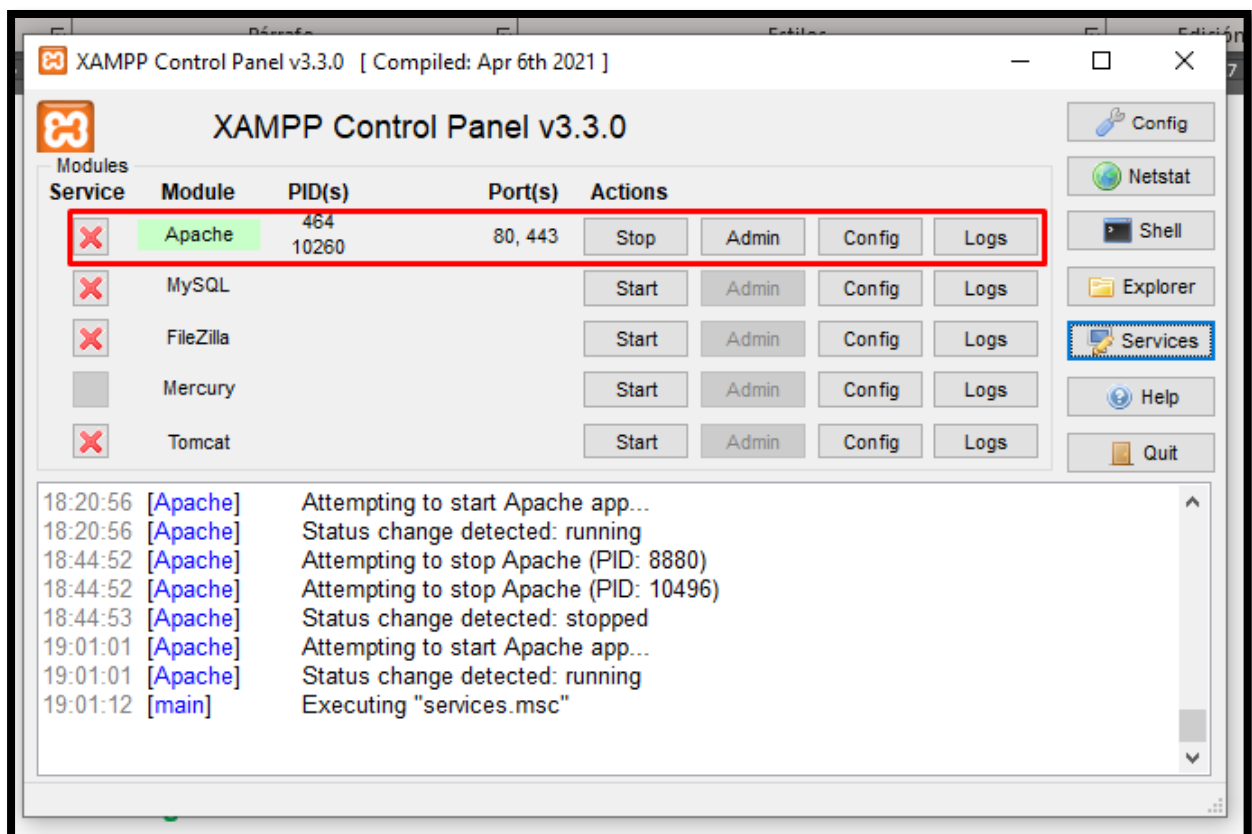
o > Documentos > GitHub > Analisis-de-datos > Repaso_1B > Archivos

Nombre	Estado	Fecha de modificación	Tipo	Tamaño
Datos file3.csv	✓	24/12/2023 6:53 PM	Archivo de valores...	171 KB

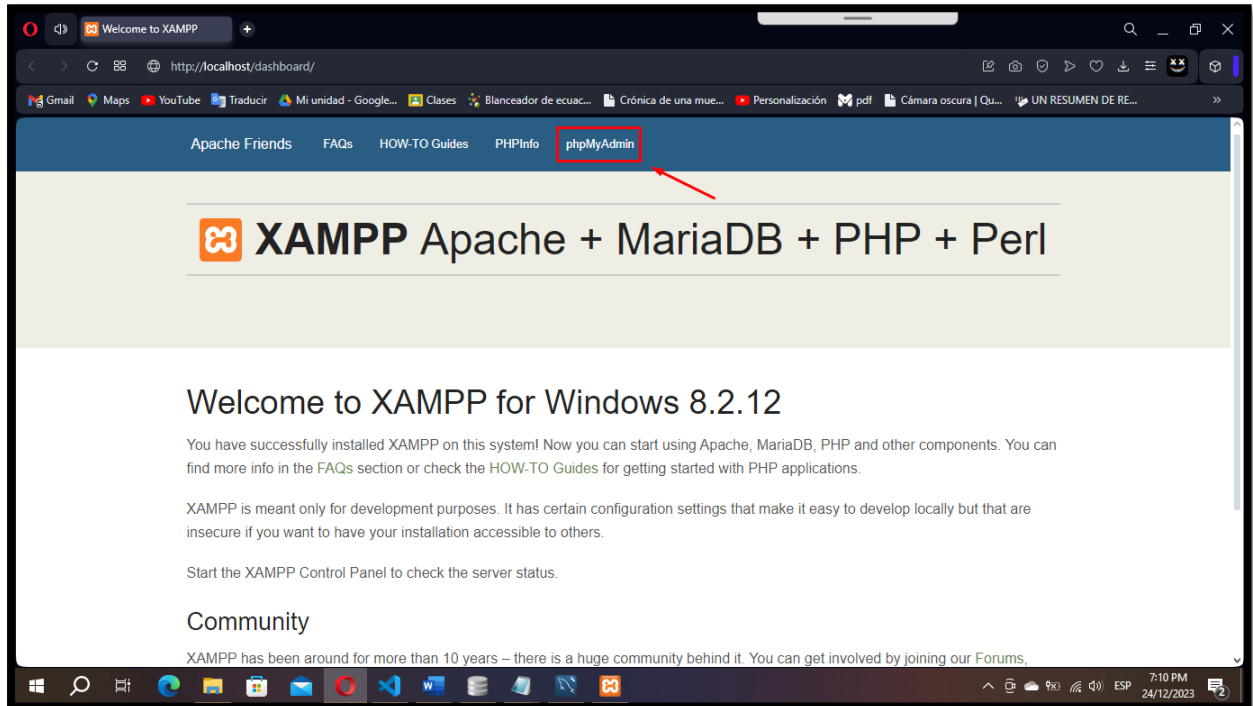
2. MySQL

- Importar file2.sql desde la base de datos MySQL. (Corregir el script en el caso que tenga errores)
- Exportar con la extensión que crea conveniente.

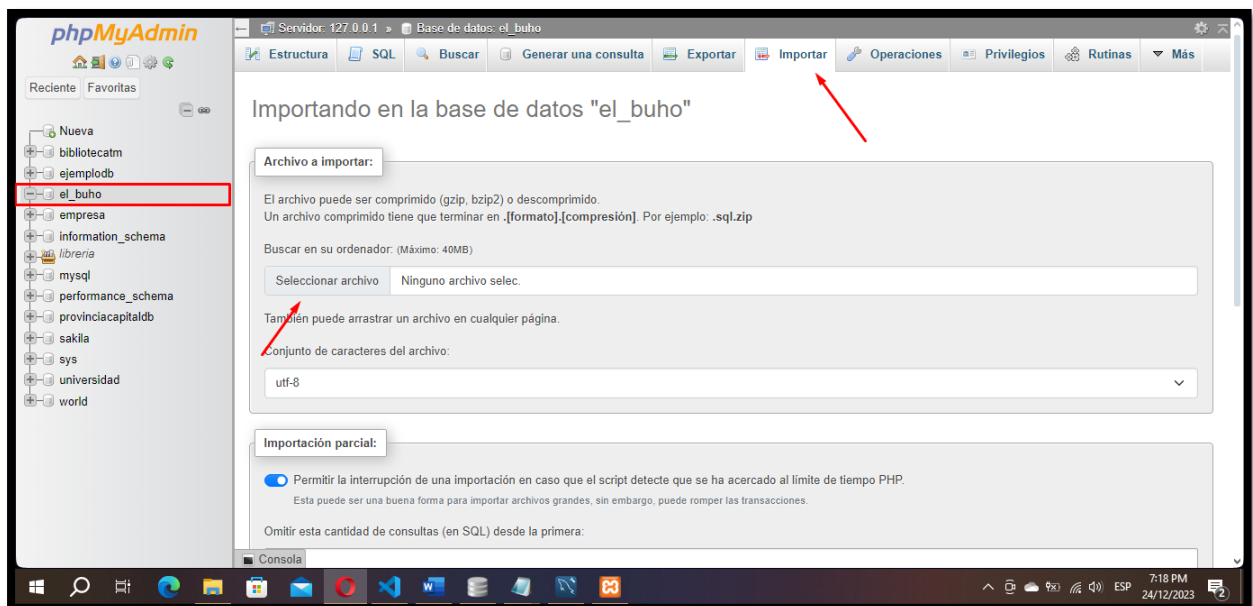
En mi caso solo inicio el servicio “Apache” debido a que ya tengo un servicio para MySQL ejecutándose

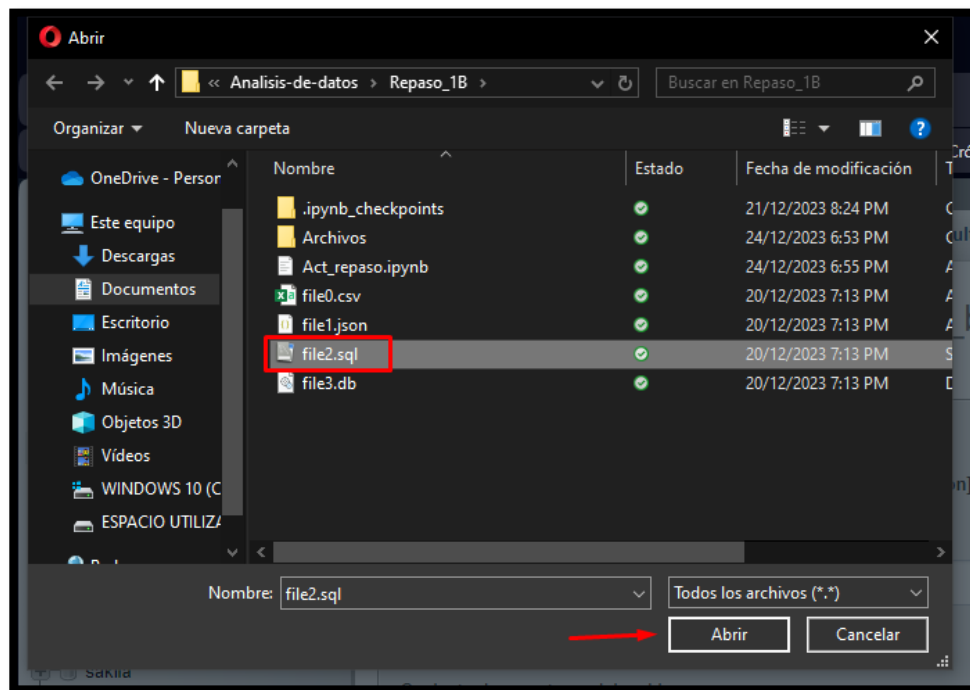


Establecer la conexión con el servidor



Importar “file2.sql” en una base de datos creada con el nombre de “el_buho”

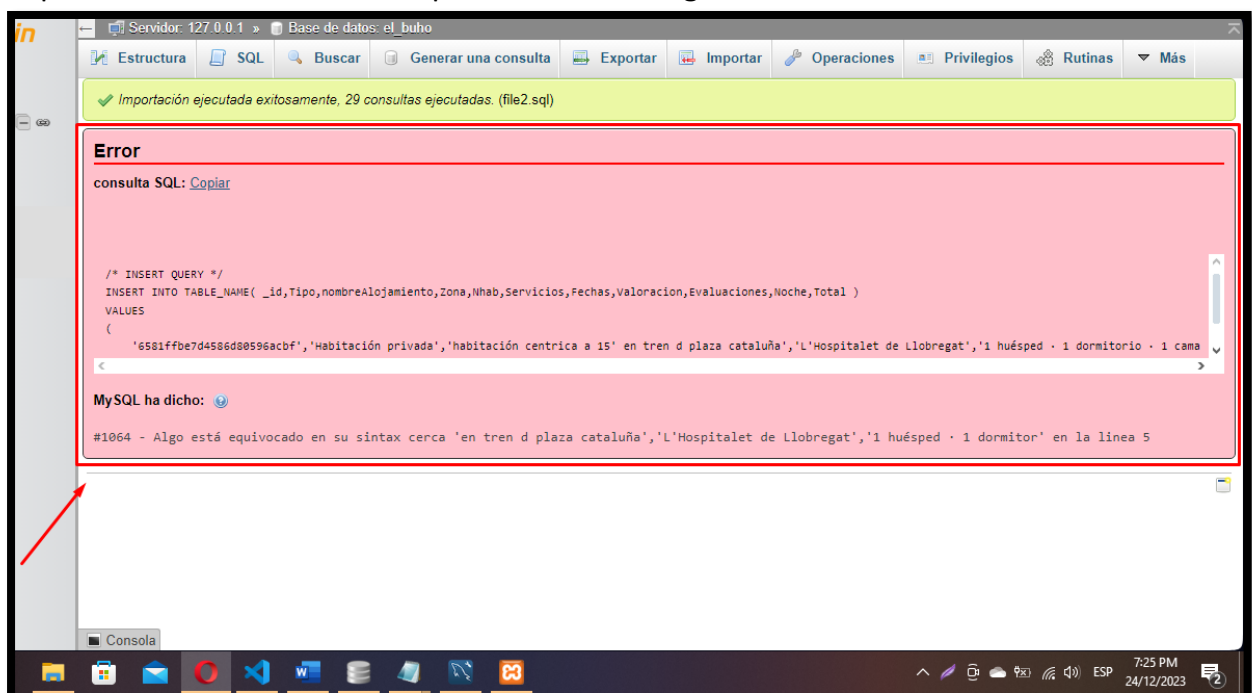




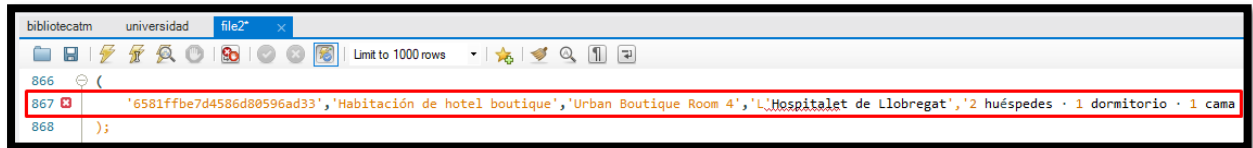
Importar el archivo a la base de datos



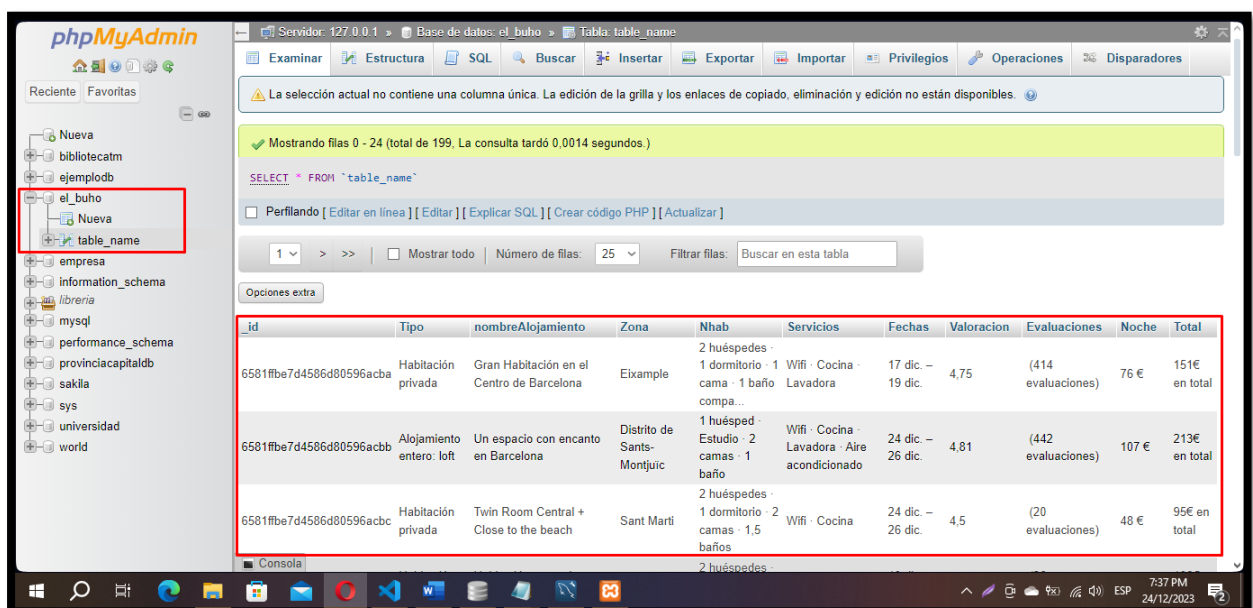
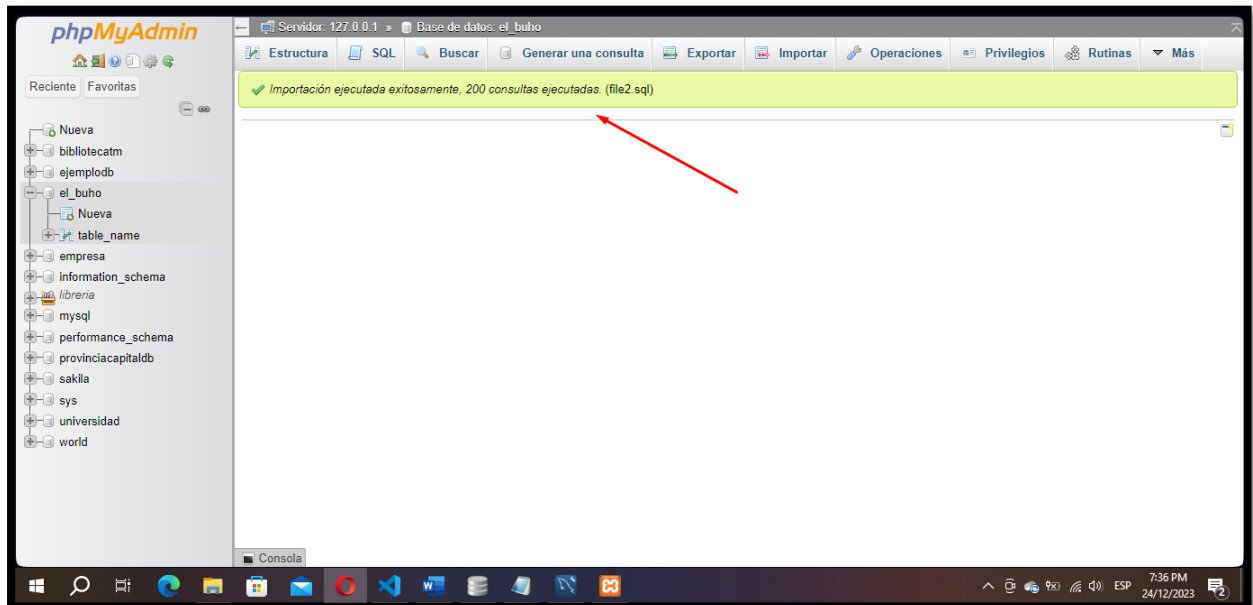
Se procederá a modificar el script debido al error generado



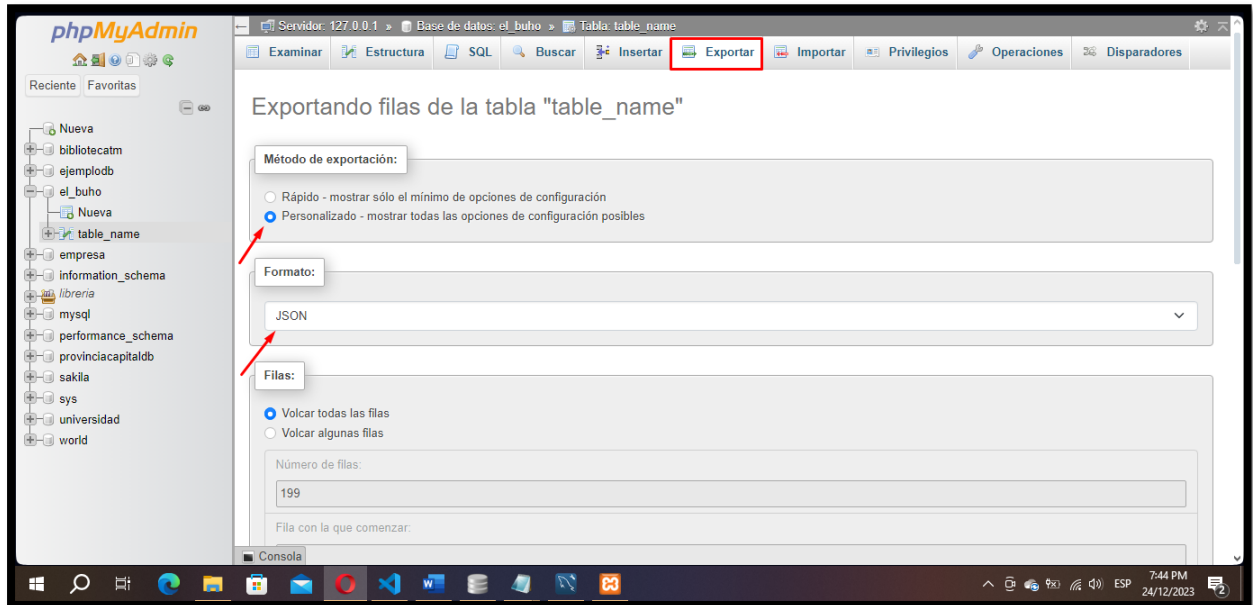
Editar todas las líneas con errores y guardar el script editado



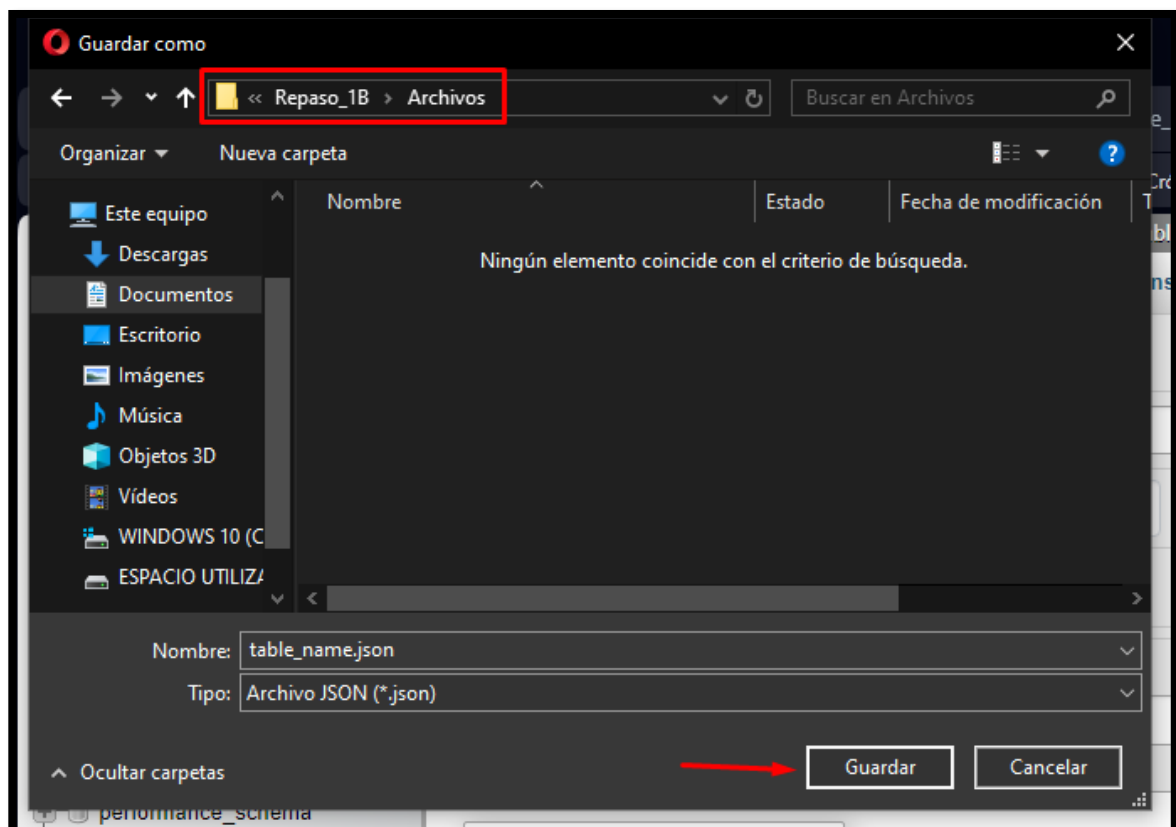
Intentar nuevamente la importación (No debe dar ningún error)

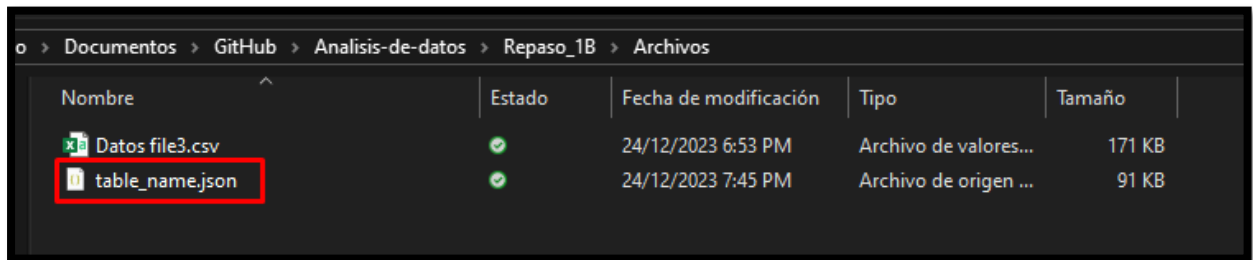


Exportar el archivo como JSON



Guardarlo en la ruta "Archivos"



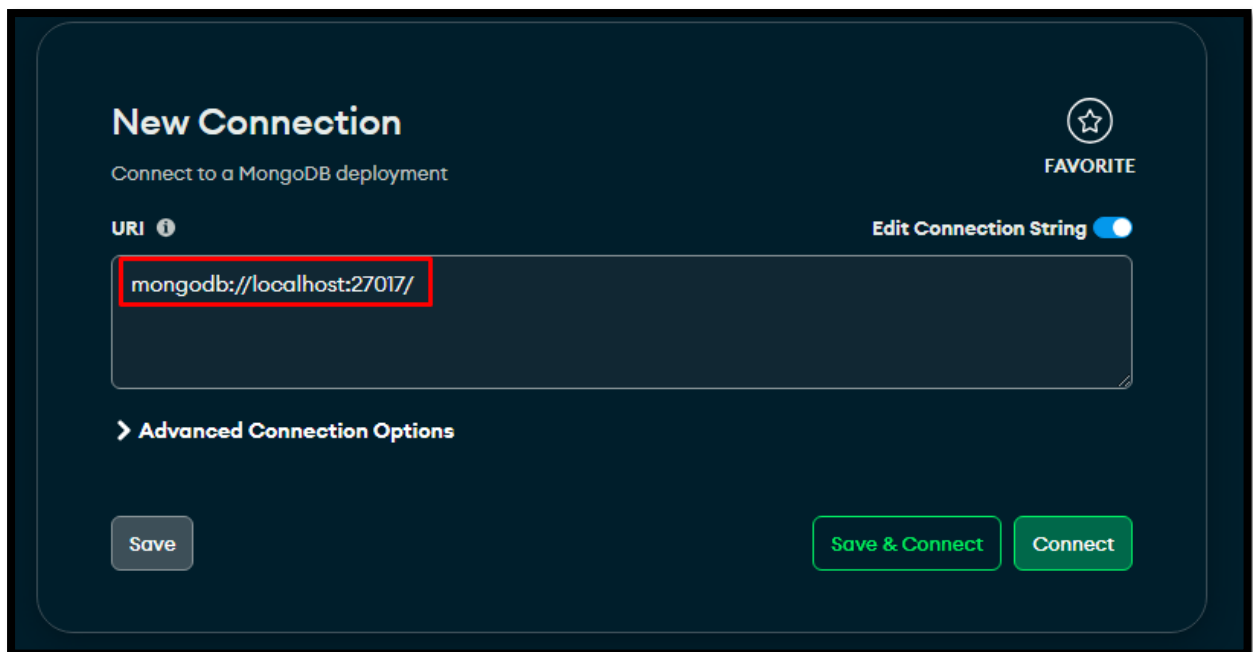


Nombre	Estado	Fecha de modificación	Tipo	Tamaño
Datos file3.csv	✓	24/12/2023 6:53 PM	Archivo de valores...	171 KB
table_name.json	✓	24/12/2023 7:45 PM	Archivo de origen ...	91 KB

3. MongoDB


- Importar file1.json desde la base de datos MongoDB.
- Exportar con la extensión que crea conveniente.

Conexión a la base de datos



New Connection

Connect to a MongoDB deployment

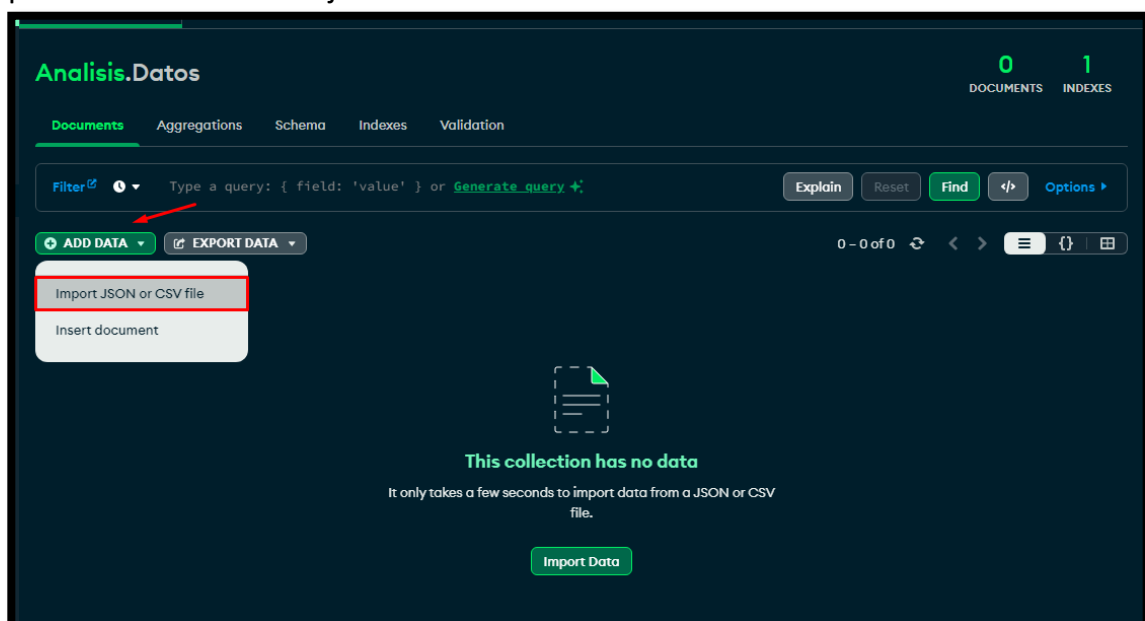
URI  Edit Connection String ☒

`mongodb://localhost:27017/`

[Advanced Connection Options](#)

Save Save & Connect Connect




Importar el archivo “file1.json”





Analisis.Datos

0 DOCUMENTS 1 INDEXES







Documents Aggregations Schema Indexes Validation

Filter  Type a query: { field: 'value' } or [Generate query](#)  Explain Reset Find  Options

ADD DATA  **EXPORT DATA** 

Import JSON or CSV file

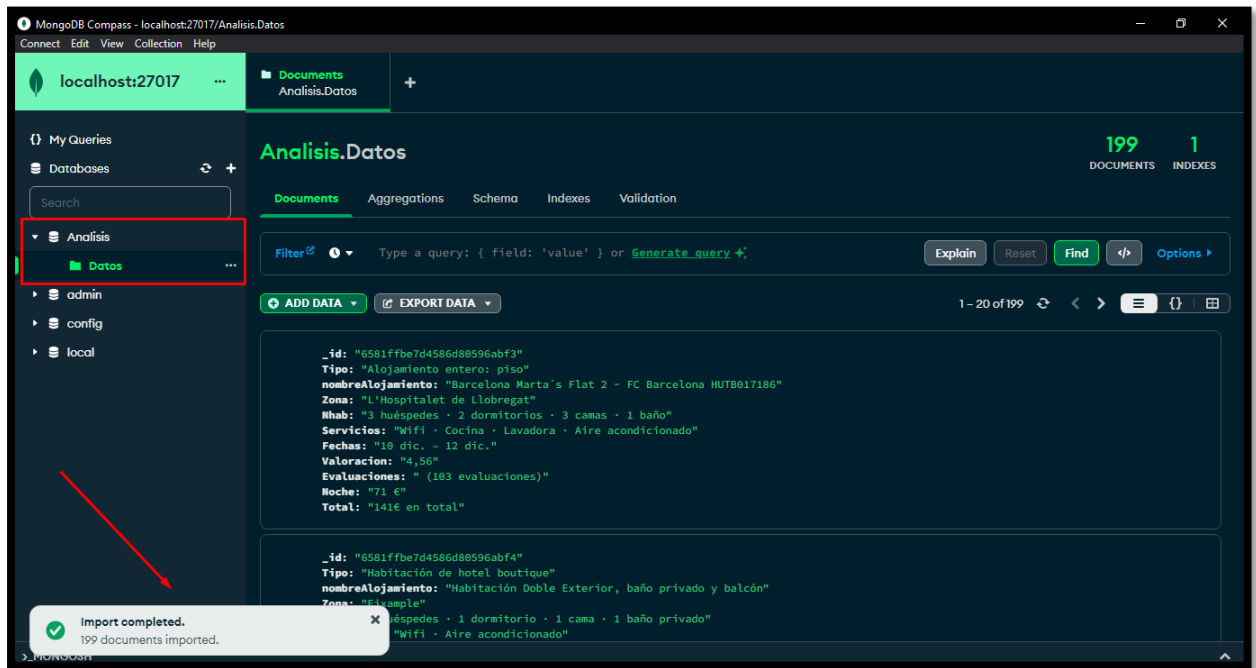
Insert document

0 - 0 of 0      

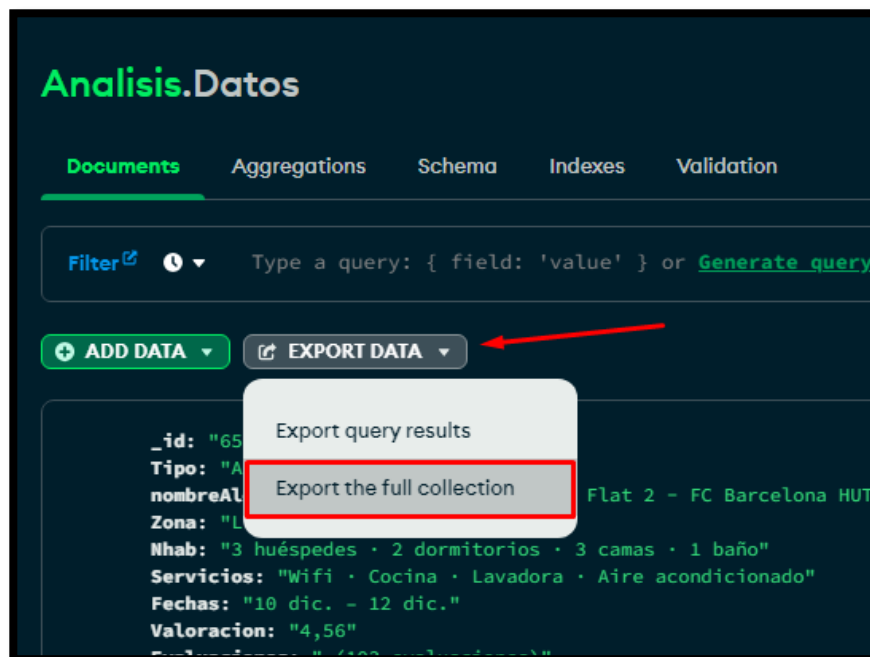
This collection has no data

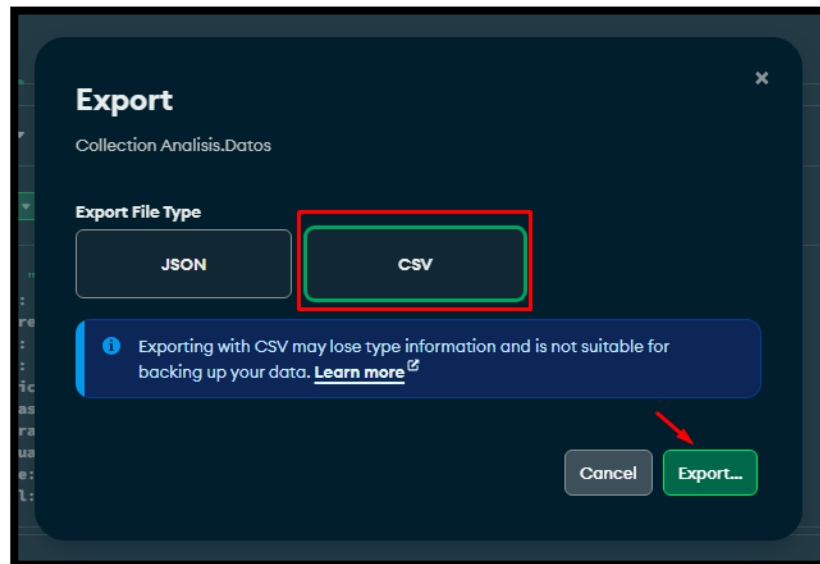
It only takes a few seconds to import data from a JSON or CSV file.

Import Data

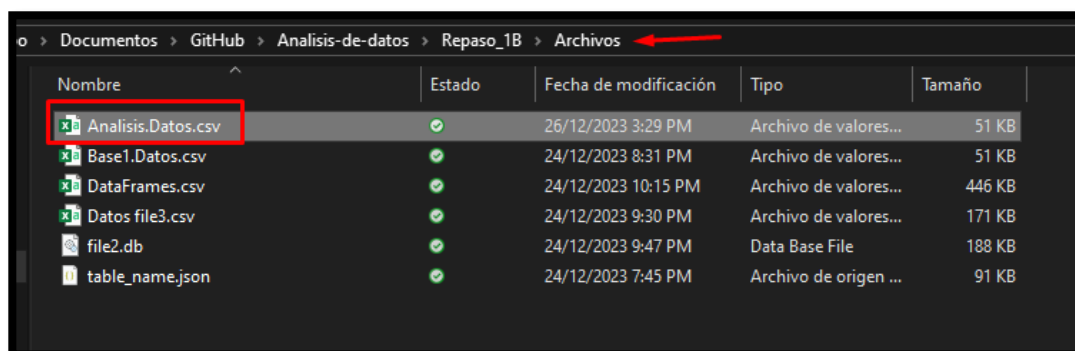
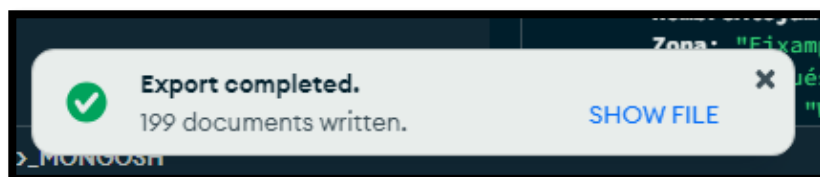
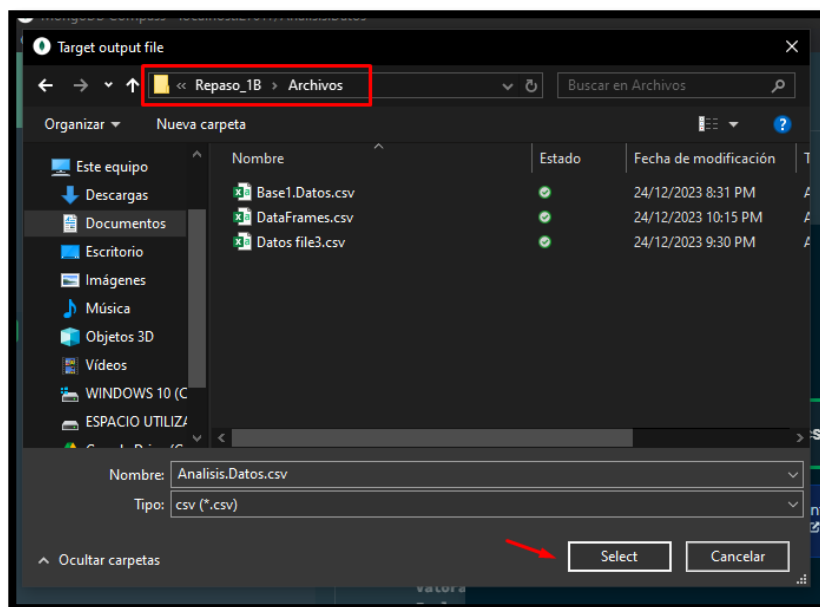


Exportar el archivo con extensión CSV





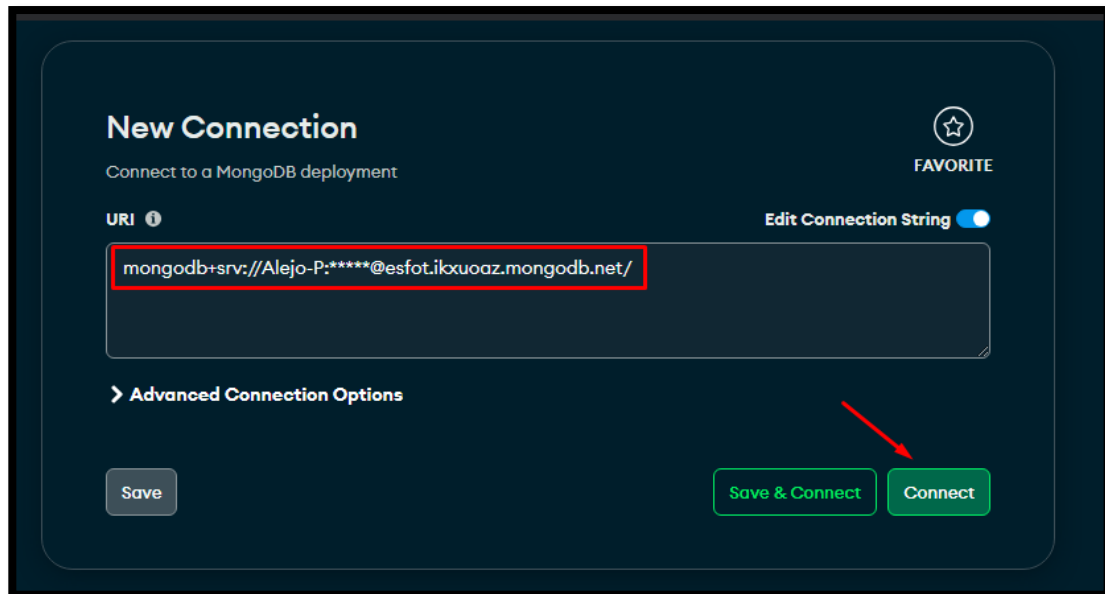
Guardar el archivo en la ruta “Archivos”



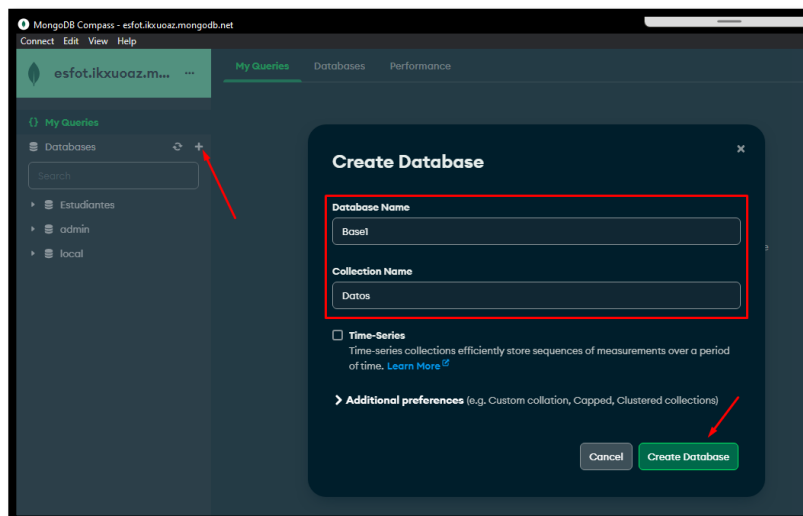
4. MongoDB Atlas

- Importar file0.csv desde la base de datos MongoDB Atlas.
- Exportar con la extensión que crea conveniente.

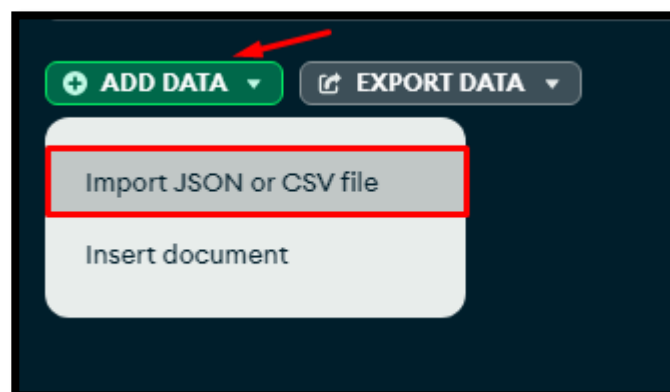
Conectarse a la base de datos

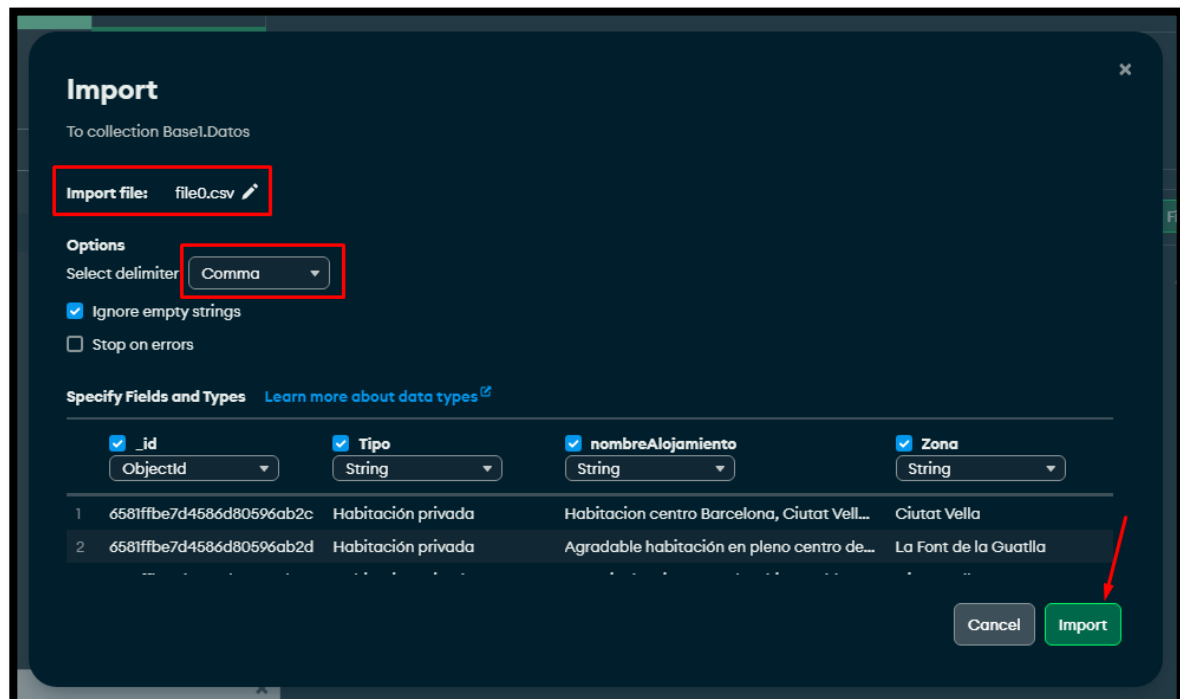
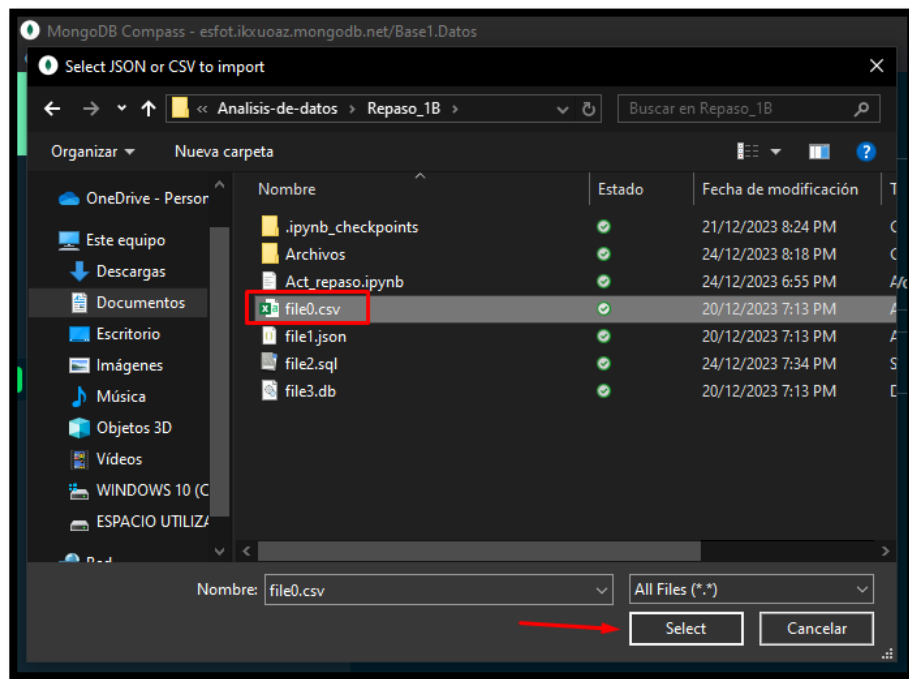


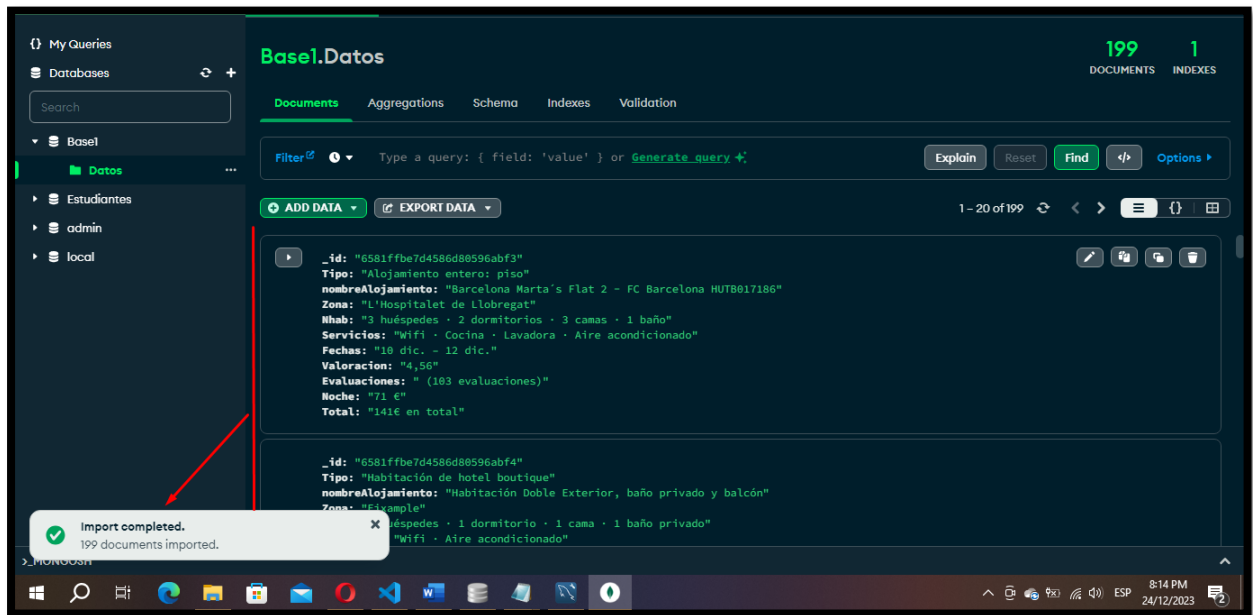
Crear la base de datos



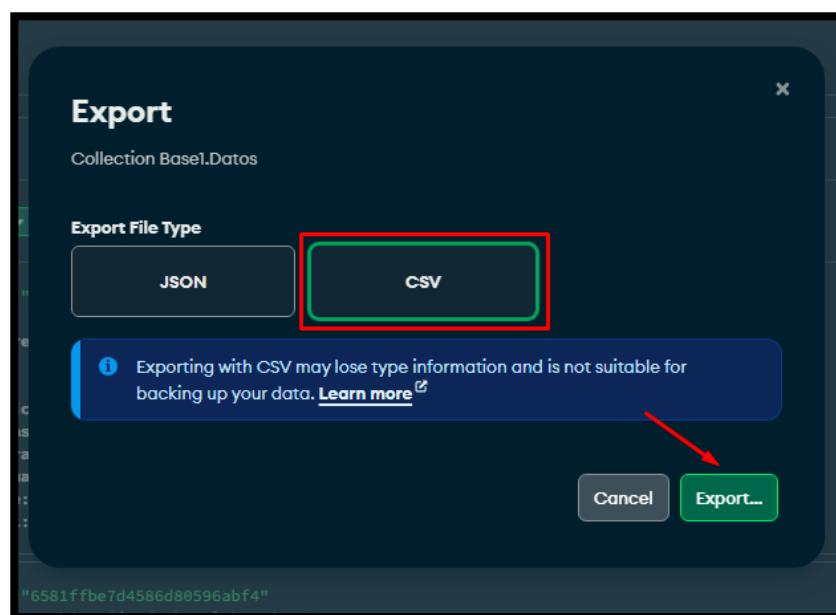
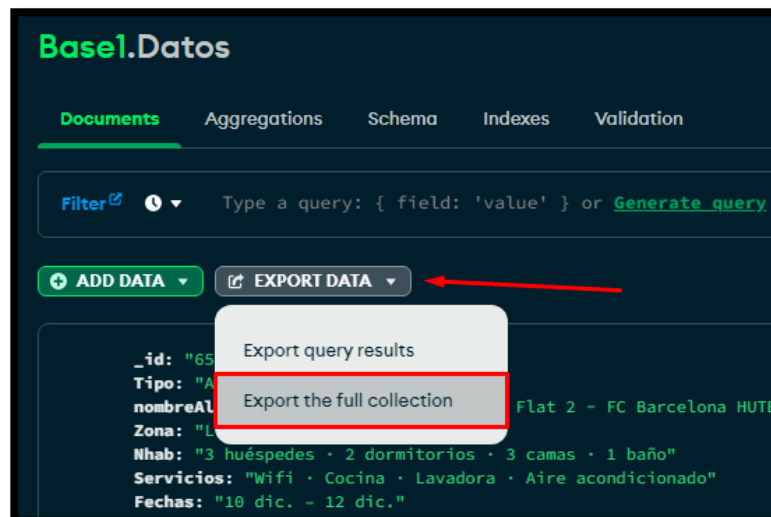
Importar el archivo "file0.JSON" a la base de datos



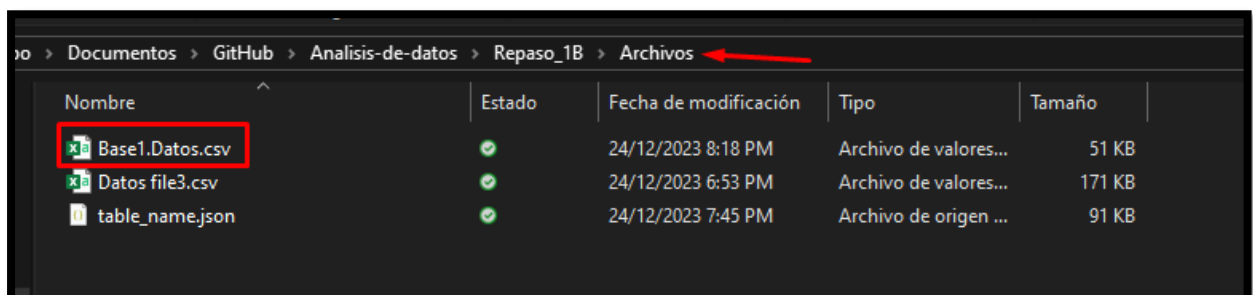
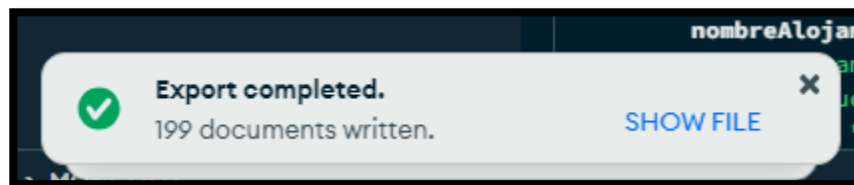
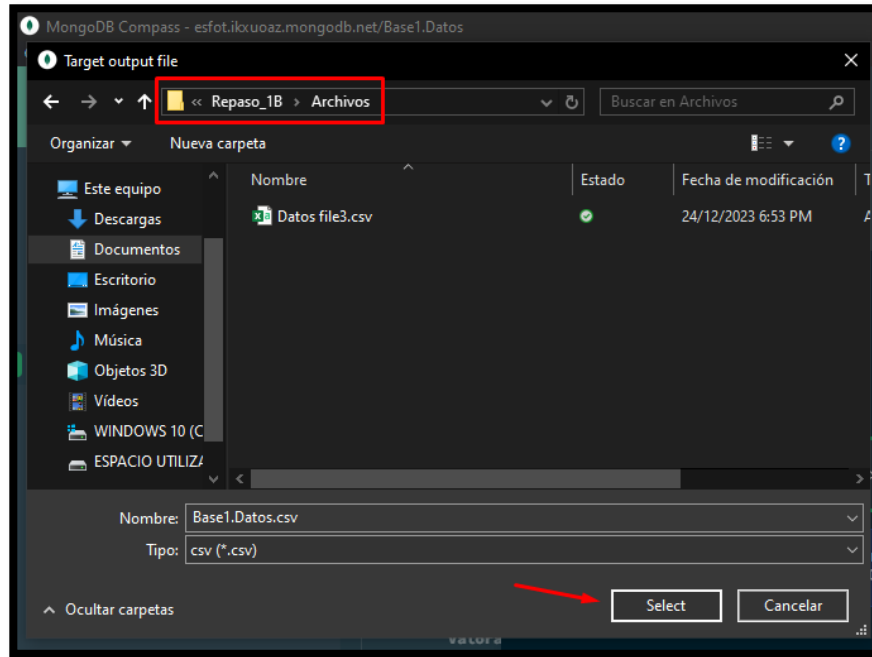




Exportar toda la información como archivo CSV



Guardar el archivo en la ruta “Archivos”



5. Leer file3 desde Jupyter con Python

```

Click here to ask Blackbox to help you code faster |
# Conexión con la base de datos
conexion=sqlite3.connect(ruta+"/file3.db")
consulta="SELECT * FROM TABLE_NAME"
datos=pd.read_sql_query(consulta, conexion)
conexion.commit()
conexion.close()
# Guardar el DataFrame como archivo CSV
datos.to_csv(rutaarch+"/Datos_file3.csv", index=False)
# Mostrar la información (10 primeras filas)
datos.head(10)

```

[14] ✓ 0.0s Python

	_id	Tipo	nombreAlojamiento	Zona	Nhab	Servicios	Fechas	Valoracion	Evaluaciones	Noche	Total
0	6581ffbe7d4586d80596ad81	Habitación privada	Beautiful and sunny apartment at Eixample!	El Fort Pienc	2 huéspedes · 1 dormitorio · 1 cama · 1 baño	Wifi · Cocina · Lavadora	17 dic. – 19 dic.	NA	No hay evaluaciones	58 €	116€ en total
1	6581ffbe7d4586d80596ad82	Habitación privada	Room near the center	Distrito de Sants-Montjuic	2 huéspedes · 1 dormitorio · 1 cama · 1 baño	Wifi	17 dic. – 19 dic.	4,21	(185 evaluaciones)	36 €	71€ en total
2	6581ffbe7d4586d80596ad81	Habitación privada	Beautiful and sunny apartment at Eixample!	El Fort Pienc	2 huéspedes · 1 dormitorio · 1 cama · 1 baño	Wifi · Cocina · Lavadora	17 dic. – 19 dic.	NA	No hay evaluaciones	58 €	116€ en total
3	6581ffbe7d4586d80596ad82	Habitación privada	Room near the center	Distrito de Sants-Montjuic	2 huéspedes · 1 dormitorio · 1 cama · 1 baño	Wifi	17 dic. – 19 dic.	4,21	(185 evaluaciones)	36 €	71€ en total
4	6581ffbe7d4586d80596ad81	Habitación privada	Beautiful and sunny apartment at Eixample!	El Fort Pienc	2 huéspedes · 1 dormitorio · 1 cama · 1 baño	Wifi · Cocina · Lavadora	17 dic. – 19 dic.	NA	No hay evaluaciones	58 €	116€ en total

6. Leer file2 desde Jupyter con Python

Leer el archivo file2.sql en Python

```

Click here to ask Blackbox to help you code faster |
# Conecta con la base de datos (crea el archivo si no existe)
conexion = sqlite3.connect(rutaarch+"/file2.db")

# Crea un objeto cursor para ejecutar comandos SQL
cursor = conexion.cursor()

# Lee el contenido del archivo .sql con diferentes encodings
with open(ruta + "/file2.sql", 'r', encoding="latin-1") as archivo_sql:
    comandos_sql = archivo_sql.read()

# Ejecuta los comandos SQL en la base de datos
cursor.executescript(comandos_sql)

# Guarda los cambios y cierra la conexión
conexion.commit()
Dataset2=pd.read_sql_query("SELECT * FROM table_name", conexion)
conexion.close()

```

[10] ✓ 1.9s Python

7. Leer file1 desde Jupyter con Python

7. Leer file1.json desde Jupyter con Python

```

Click here to ask Blackbox to help you code faster |
dataset3=pd.read_json(ruta+"/file1.json")
# Mostrar las primeras 3 filas
dataset3.head(3)

```

[16] ✓ 0.0s Python

	_id	Tipo	nombreAlojamiento	Zona	Nhab	Servicios	Fechas	Valoracion	Evaluaciones	Noche	Total
0	6581ffbe7d4586d80596abf3	Alojamiento entero: piso	Barcelona Marta's Flat 2 - FC Barcelona HUTB01...	L'Hospitalet de Llobregat	3 huéspedes · 2 dormitorios · 3 camas · 1 baño	Wifi · Cocina · Lavadora · Aire acondicionado	10 dic. – 12 dic.	4,56	(103 evaluaciones)	71 €	141€ en total
1	6581ffbe7d4586d80596abf4	Habitación de hotel boutique	Habitación Doble Exterior, baño privado y balcón	Eixample	2 huéspedes · 1 dormitorio · 1 cama · 1 baño p...	Wifi · Aire acondicionado	17 dic. – 19 dic.	4,56	(9 evaluaciones)	87 €	174€ en total
2	6581ffbe7d4586d80596abf5	Habitación privada	Luxury room with private bathroom 2	Sant Pere-Santa Caterina i la Ribera	3 huéspedes · 1 dormitorio · 1 baño privado	Wifi · Cocina · Lavadora	10 dic. – 12 dic.	4,53	(15 evaluaciones)	76 €	151€ en total

8. Leer file0 desde Jupyter con Python

```

dataset4=pd.read_csv(ruta+"file0.csv", delimiter=";", index_col="_id")
# Mostrar las ultimas 5 filas
dataset4.tail(5)

```

_id	Tipo	nombreAlojamiento	Zona	Nhab	Servicios	Fechas	Valoracion	Evaluaciones	Noche	Total
6581ffbe7d4586d80596abee	Alojamiento entero: piso	Precioso Estudio en Barcelona *Bcn* A01	Eixample	3 huéspedes · 2 camas · 1 baño	Tu anfitrión es una empresa · Wifi · Cocina · ...	17 dic. – 19 dic.	4,76	(46 evaluaciones)	103 €	206€ en total
6581ffbe7d4586d80596abef	Habitación privada	Habitación doble (tranquila) cerca de Las Ram...	Ciutat Vella	2 huéspedes · 1 dormitorio · 2 camas · 2 baños...	Wifi · Lavadora	10 dic. – 12 dic.	4,9	(10 evaluaciones)	50 €	100€ en total
6581ffbe7d4586d80596abf0	Habitación de hotel boutique	Urban Boutique Room 3	L'Hospitalet de Llobregat	2 huéspedes · 1 dormitorio · 1 cama · 1 baño p...	Wifi · Aire acondicionado	17 dic. – 19 dic.	4,4	(20 evaluaciones)	51 €	102€ en total
6581ffbe7d4586d80596abf1	Habitación de hotel	Estilo atemporal en habitación doble con balcón	Eixample	2 huéspedes · 1 dormitorio · 1 baño privado	Wifi · Aire acondicionado	10 dic. – 12 dic.	4,52	(462 evaluaciones)	84 €	168€ en total
6581ffbe7d4586d80596abf2	Habitación privada	Luxurious Room with En Suite Bathroom	Ciutat Vella	2 huéspedes · 1 dormitorio · 1 baño privado	Wifi · Cocina · Aire acondicionado	24 dic. – 26 dic.	4,88	(42 evaluaciones)	95 €	189€ en total

9. Integrar los cuatro data set y enviar a MongoDB Atlas (puede usar concat)

```

# Concatenar los datasets
combinacion=pd.concat([dataset1, dataset2, dataset3, dataset4])

# Ubicar la unificación en un archivo CSV
combinacion.to_csv(rutaarch+"/DataFrames.csv", index=False)

```

Nombre	Estado	Fecha de modificación	Tipo	Tamaño
Base1.Datos.csv	✓	24/12/2023 8:31 PM	Archivo de valores...	51 KB
DataFrames.csv	✓	24/12/2023 9:47 PM	Archivo de valores...	436 KB
Datos file3.csv	✓	24/12/2023 9:30 PM	Archivo de valores...	171 KB
file2.db	🔄	24/12/2023 9:47 PM	Data Base File	188 KB
table_name.json	✓	24/12/2023 7:45 PM	Archivo de origen ...	91 KB

Base1.combinacion 700 DOCUMENTS 1 INDEXES

Documents Aggregations Schema Indexes Validation

Filter Generate query Explain Reset Find Options

ADD DATA EXPORT DATA 1 - 19 of 700

```

{
  "_id": {},
  "Tipo": "Habitación privada",
  "nombreAlojamiento": "Room near the center",
  "Zona": "Distrito de Sants-Montjuïc",
  "Nhab": "2 huéspedes · 1 dormitorio · 1 cama · 1 baño",
  "Servicios": "Wifi",
  "Fechas": "17 dic. - 19 dic.",
  "Valoracion": "4,21",
  "Evaluaciones": "(185 evaluaciones)",
  "Noche": "36 €",
  "Total": "716 en total"
}

```

```

{
  "_id": {},
  "Tipo": "Habitación privada",
  "nombreAlojamiento": "Habitación en Gracia cerca del Park Güell",
  "Zona": "El Poble Sec",
  "Nhab": "2 huéspedes · 1 dormitorio · 1 cama · 1 baño",

```

10. Importar el archivo desde MongoDB Atlas hasta Jupyter haciendo uso de las librerías NumPy y pandas y realizar lo siguiente:

- a. Verificar el tiempo que se demora en cargar el data set.

```
a. Verificar el tiempo que se demora en cargar el data set.
```

```
Click here to ask Blackbox to help you code faster |
# Importar las bibliotecas necesarias (Pandas ya se importo antes)
import numpy as np
import time
t_inicial=time.time()
Dataframe=pd.read_csv(rutaarch+"/DataFrames.csv", sep=",", index_col="_id")
t_final=time.time()
print("El tiempo demorado en cargar el DataSet fue: {} segundos".format(round((t_final-t_inicial), 2)))
```

[11] ✓ 0.1s Python

... El tiempo demorado en cargar el DataSet fue: 0.13 segundos

- b. Verificar el número de registros.

```
Click here to ask Blackbox to help you code faster |
regs=len(Dataframe)
print(f"El total de registros en el DataFrame es: {regs} registros")
```

[19] ✓ 0.0s Python

... El total de registros en el DataFrame es: 1861 registros

- c. Verificar si hay registros duplicados, en el caso de haberlos, deberá eliminar los duplicados.

```
Click here to ask Blackbox to help you code faster |
#Verificar valores duplicados en cada registro
duplicados=Dataframe.duplicated()
Dataframe=Dataframe.drop_duplicates() # Eliminar filas duplicadas
no_duplicados=Dataframe.duplicated() # Volver a comprobar valores duplicados
print(duplicados,"\n",no_duplicados)
```

[38] ✓ 0.0s Python

... 0 False
1 False
2 True
3 True
4 True
...
1856 False
1857 False
1858 False
1859 False
1860 False
Length: 1861, dtype: bool

→ Valores duplicados

0 False
1 False
8 False
9 False
10 False
...
1856 False
1857 False
1858 False
1859 False
1860 False
Length: 700, dtype: bool

→ Valores no duplicados

d. Verificar si hay celdas nulas

```
Click here to ask Blackbox to help you code faster |
#Sumar la cantidad de celdas vacías de cada columna en el DataFrame
Dataframe.isnull().sum()

[39] ✓ 0.0s Python
```

...	_id	198
	Tipo	0
	nombreAlojamiento	0
	Zona	0
	Nhab	0
	Servicios	0
	Fechas	0
	Valoracion	30
	Evaluaciones	1
	Noche	1
	Total	1
	dtype:	int64

e. En caso de haber celdas nulas, se deberá sacar el promedio y completar la información con la media.

```
Click here to ask Blackbox to help you code faster | Comment Code |
def Generar_ID():
    PreID="6581ffbe7d4586d80596" # Obtener parte del ID predefinido
    # Usar la funcion time() para generar numeros en serie, convertirlos a hexadecimal y extraer los 4 ultimos digitos
    contador=hex(int(time.time()))[-4:]
    return PreID+contador # Retornar la ID generada

Valores=Dataframe["_id"]
#Iterar entre cada valor del diccionario Valores
for ind, valor in Valores.items():
    if pd.isna(valor): # Comprobar que el valor no sea nulo
        Dataframe.at[ind, "_id"]=Generar_ID()

[171] ✓ 0.0s Python
```

```
Click here to ask Blackbox to help you code faster |
Dataframe.isnull().sum()

[172] ✓ 0.3s Python
```

...	_id	0
	Tipo	0
	nombreAlojamiento	0
	Zona	0
	Nhab	0
	Servicios	0
	Fechas	0
	Valoracion	30
	Evaluaciones	1
	Noche	1
	Total	1
	dtype:	int64

```

def Decimal(numero):
    """Convierte un numero en decimal"""
    numstr=""
    punto=0
    for digito in str(numero): #Iterar sobre cada caracter de la cadena
        if (digito=='.' or digito==',') and punto<1: # Comprobar si el digito de la cadena es '.' o ','
            punto+=1
            numstr+="."
        elif digito.isdigit(): # Comprobar si la cadena es un digito
            numstr+=str(digito)
    return float(numstr) # Retornar la cadena convertida a decimal

# Extraer los valores no nulos de la fila "Valoracion"
Valoracion=Dataframe["Valoracion"]
Valor_no_nulos=Dataframe["Valoracion"].dropna()
suma, iteracion=0, 0 # Definir variables para calcular el promedio
for _, valor in Valor_no_nulos.items(): # Iterar entre cada valor del diccionario "Valor_no_nulos"

    #Debido a que los valores que se extraen "3,9" no son numeros validos, se usa una funcion creada para convertir
    #la cadena a numeros validos "3.9" para calcular el promedio
    suma+=Decimal(valor) #conversion de cadena a numero, remplazar la ',' por '.'
    iteracion+=1 # Almacenarla cantidad de veces que se repite el proceso

# Calcular el promedio
promedio=round(suma/iteracion,2)
# Volver a ubicar la ',' en lugar del '.'
prom=str(promedio).replace('.', ',')
#Iterar entre cada elemento de "Valoracion" y remplazar los valores nulos
for clave, valor in Valoracion.items():
    if pd.isna(valor):
        Dataframe.at[clave,"Valoracion"]=prom # Reemplazar los valores nulos con el promedio
Dataframe.isnull().sum() # Ver si hay valores nulos

```

```

_id          0
Tipo         0
nombreAlojamiento  0
Zona         0
Nhab         0
Servicios    0
Fechas       0
Valoracion   0
Evaluaciones 1
Noche        1
Total        1
dtype: int64

```

```

Click here to ask Blackbox to help you code faster |
import random # Importar la biblioteca para generar numeros aleatorios
# Extraer valores de la columna "Evaluaciones"
Evaluaciones=Dataframe["Evaluaciones"]
# Iterar por cada valor, buscar y remplazar valores nulos
for clave, valor in Evaluaciones.items():
    if pd.isna(valor):
        Dataframe.at[clave, "Evaluaciones"]=f"({random.randint(22, 750)} evaluaciones)"
Dataframe.isnull().sum() # Ver si hay valores nulos

```

[247] ✓ 0.0s Python

```

... _id          0
    Tipo        0
    nombreAlojamiento  0
    Zona       0
    Nhab        0
    Servicios   0
    Fechas      0
    Valoracion  0
    Evaluaciones 0
    Noche       1
    Total      1
    dtype: int64

```

```

Click here to ask Blackbox to help you code faster |
# Extraer valores completos y no nulos de la columna "Noches"
Noche=Dataframe["Noche"].dropna()
Noches=Dataframe["Noche"]
Valores=[] # Lista donde se guardaran los valores numericos
for clave in Noche.keys(): # Iterar entre cada clave del diccionario "Noche"
    # Agregar a la lista el valor en el indice 0 de la lista separada del valor del indice del diccionario
    Valores.append(int(Noche[clave].split(" ")[0]))
# Calcular el promedio
Costo_noche=round(sum(Valores)/len(Valores))

for clave, valor in Noches.items(): # Iterar entre cada clave del diccionario "Noches"
    if pd.isnull(valor): # Comprobar valores nulos y remplazarlos
        Noches[clave]=f"{Costo_noche} €"

Dataframe.isnull().sum() # Ver si hay valores nulos

```

[271] ✓ 0.0s Python

```

... _id          0
    Tipo        0
    nombreAlojamiento  0
    Zona       0
    Nhab        0
    Servicios   0
    Fechas      0
    Valoracion  0
    Evaluaciones 0
    Noche       0
    Total      1
    dtype: int64

```

```

Click here to ask Blackbox to help you code faster |
# Extraer valores completos y no nulos de la columna "Total"
Totales=Dataframe["Total"]

for clave, valor in Totales.items(): # Iterar entre cada clave del diccionario "Totales"
    if pd.isnull(valor): # Comprobar valores nulos y remplazarlos
        Totales[clave]=f"{random.randint(40, 260)}€ en total" #Rellenar el valor nulo con un numero aleatorio

Dataframe.isnull().sum() # Ver si hay valores nulos

```

[322] ✓ 0.3s Python

```

... _id          0
    Tipo        0
    nombreAlojamiento  0
    Zona       0
    Nhab        0
    Servicios   0
    Fechas      0
    Valoracion  0
    Evaluaciones 0
    Noche       0
    Total      0
    dtype: int64

```

f. Describir el data set

Click here to ask Blackbox to help you code faster | `Dataframe.describe()`

[324] ✓ 1.4s Python

	_id	Tipo	nombreAlojamiento	Zona	Nhab	Servicios	Fechas	Valoracion	Evaluaciones
count	700	700	700	700	700	700	700	700	700
unique	700	14	377	41	154	23	17	80	168
top	6581ffbe7d4586d80596ad81	Habitación privada	Habitación doble en Sagrada Familia	Ciutat Vella	2 huéspedes · 1 dormitorio · 1 cama · 1 baño c...	Wifi · Cocina · Lavadora · Aire acondicionado	17 dic. – 19 dic.	4,64	No hay evaluaciones
freq	1	324	4	152	111	130	174	36	29

g. Describir usando campos definidos, mostrando 10 filas.

Click here to ask Blackbox to help you code faster |

```
#Describir la columna "Fechas" y mostrar las 10 primeras filas
print(Dataframe["Fechas"].describe())
Dataframe["Fechas"].head(10)
```

[327] ✓ 0.0s Python

count 700
unique 17
top 17 dic. – 19 dic.
freq 174
Name: Fechas, dtype: object

0 17 dic. – 19 dic.
1 17 dic. – 19 dic.
8 10 dic. – 12 dic.
9 17 dic. – 19 dic.
10 17 dic. – 19 dic.
11 10 dic. – 12 dic.
12 17 dic. – 19 dic.
13 17 dic. – 19 dic.
14 26 nov. – 28 nov.
15 10 dic. – 12 dic.
Name: Fechas, dtype: object

Click here to ask Blackbox to help you code faster |

```
#Describir la columna "Fechas" y mostrar las 10 primeras filas
print(Dataframe["Total"].describe())
Dataframe["Total"].head(10)
```

[328] ✓ 0.0s Python

count 700
unique 271
top 83€ en total
freq 24
Name: Total, dtype: object

0 116€ en total
1 71€ en total
8 57€ en total
9 83€ en total
10 45€ en total
11 139€ en total
12 59€ en total
13 88€ en total
14 63€ en total
15 66€ en total
Name: Total, dtype: object

```

Click here to ask Blackbox to help you code faster |
#Describir la columna "Evaluaciones" y mostras las 10 primeras filas
print(Dataframe["Evaluaciones"].describe())
Dataframe["Evaluaciones"].head(10)
[329] ✓ 0.0s Python
...
count          700
unique          168
top      No hay evaluaciones
freq           29
Name: Evaluaciones, dtype: object

...
0      No hay evaluaciones
1      (185 evaluaciones)
8      (201 evaluaciones)
9      (22 evaluaciones)
10     (232 evaluaciones)
11     (65 evaluaciones)
12     (276 evaluaciones)
13     (15 evaluaciones)
14     (146 evaluaciones)
15     (314 evaluaciones)
Name: Evaluaciones, dtype: object

```

h. Indique el alojamiento más costoso por noche.

```

Click here to ask Blackbox to help you code faster |
# Mostrar el alojamiento mas costoso por noche
# Algoritmo de conversion de valor y busqueda (mayor valor)
Alojamiento_costoso=Dataframe["Noche"]
valor=int(str(Alojamiento_costoso[0]).split()[0])
for clave in Alojamiento_costoso.keys():
    num=int(str(Alojamiento_costoso[clave]).split()[0])
    if num>valor:
        valor=int(str(Alojamiento_costoso[clave]).split()[0])
#Completar y convertir el indice de busqueda
Ind=str(f"{valor} €")
# Filtar y mostrar las filas que cumplan con la condicion
Dataframe[Dataframe["Noche"]==Ind]
[350] ✓ 0.3s Python
...

```

_id	Tipo	nombreAlojamiento	Zona	Nhab	Servicios	Fechas	Valoracion	Evaluaciones	Noche	Total
6581ffb7d4586d80596ac37	Alojamiento entero: piso	My Address in Plaza Real views	Ciutat Vella	4 huéspedes · 2 dormitorios · 4 camas · 1 baño	Tu anfitrión es una empresa · Wifi · Aire acon...	10 dic. – 12 dic.	4,42	(38 evaluaciones)	175 €	350€ en total

i. Indique el total cancelado más costoso por un cliente.

```

Click here to ask Blackbox to help you code faster |
# Total cancelado mas costoso
Comment Code
def Extraer_numero(Cadena):
    #Retornar solo el numero de la cadena enviada a la funcion
    return int(''.join([i for i in str(Cadena) if str(i).isdigit()]))

Total=Dataframe["Total"]
Precios_totales=[]
for _, valor in Total.items():
    Precios_totales.append(Extraer_numero(valor)) # Guardar el valor como numero en la lista
# Ordenar la lista de forma descendente para obtener el valor mas alto primero
Precios_totales.sort(reverse=True)
Valor_alto=str(f"{Precios_totales[0]}€ en total")
# Filtrar y mostrar las filas que cumplan con la condicion
Dataframe[Dataframe["Total"]==Valor_alto]

```

[368] ✓ 0.0s Python

...	_id	Tipo	nombreAlojamiento	Zona	Nhab	Servicios	Fechas	Valoracion	Evaluaciones	Noche	Total
6581ffbe7d4586d80596ac37	Alojamiento entero: piso	My Address in Plaza Real views	Ciutat Vella	4 huéspedes - 2 dormitorios - 4 camas - 1 baño	Tu anfitrión es una empresa · Wifi · Aire acon...	10 dic. - 12 dic.	4,42	(38 evaluaciones)	175 €	350€ en total	

j. Indique el tipo de alojamiento que hay para 6 personas.

```

Click here to ask Blackbox to help you code faster |
# tipo de alojamiento que hay para 6 personas
# Extraer los valores de la columna "Nhab" como diccionario
Num_hab=Dataframe["Nhab"]
# Crear una lista por comprension de los valores de ese diccionario
Elementos=[valor for _,valor in Num_hab.items()]
tipo=[] # Crear una lista donde e almacenaran los tipos de alojamiento y la cantidad de huespedes

# Iterar entre cada valor extraido
for i in range(len(Elementos)):
    if str(Elementos[i]).startswith("6"): # Comprobar si el valor extraido empieza en '6'
        # Añadir a la lista en tipo de alojamiento y la cantidad de huespedes en forma de lista de listas
        tipo.append([Dataframe["Tipo"][Dataframe["Nhab"]==Elementos[i]].to_list(), [Elementos[i]]])
# Iterar entre cada lista de la lista creada
for Alojamiento, Personas in tipo:
    # Convertir los valores a cadena de caracteres
    tipos=', '.join(Alojamiento)
    persona=', '.join(Personas)
    # Mostrar los resultados
    print(f"tipo: {tipos}\nPara: {persona.split(',')[0][::-1]}\n")

```

9] ✓ 0.0s Python

```

tipo: Habitación privada, Alojamiento entero: piso
Para: 6 huéspedes

tipo: Alojamiento entero: piso
Para: 6 huéspedes

tipo: Alojamiento entero: piso
Para: 6 huéspedes

```

k. Verifique cuál es la valoración más alta y cuál es la valoración más baja de los tipos d alojamientos.

Click here to ask Blackbox to help you code faster |

```
#Cuál es la valoración más alta y cuál es la valoración más baja de los tipos de alojamientos
valoracion=Dataframe["Valoracion"]
# Obtener la valoracion alta
valor_alto=valoracion.max()
# Obtener la valoracion baja
valor_bajo=valoracion.min()
# Filtrar y mostrar las filas que cumplen con la condicion (primeras 5 filas)
Dataframe[Dataframe["Valoracion"]==valor_alto].head()
```

✓ 0.2s Python

_id	Tipo	nombreAlojamiento	Zona	Nhab	Servicios	Fechas	Valoracion	Evaluaciones	Noche	Total
86d80596ad9e	Alojamiento entero: piso	Single Room in the heart of Barcelona	El Prat de Llobregat	1 huésped · 1 dormitorio · 1 cama · 1 baño pri...	Wifi · Aire acondicionado	17 dic. – 19 dic.	5	(22 evaluaciones)	57 €	114€ en total
86d80596adb5	Alojamiento entero: piso	Spacious Cosy room & Private balcony and bathroom	Eixample	5 huéspedes · 1 dormitorio · 1 cama · 1 baño	Wifi · Aire acondicionado	24 dic. – 26 dic.	5	(3 evaluaciones)	53 €	105€ en total
86d80596add1	Habitación privada	Nice Flat at Barcelona	L Hospitalet de Llobregat	4 huéspedes · 2 dormitorios · 3 camas · 1 baño	Wifi · Cocina · Lavadora · Aire acondicionado	17 dic. – 19 dic.	5	(9 evaluaciones)	130 €	259€ en total
86d80596aceb	Habitación privada	Sagrada familia, céntrico	La Sagrada	2 huéspedes · 1 dormitorio · 1 cama · 1 baño	Wifi · Cocina · Lavadora · Aire acondicionado	17 dic. – 19 dic.	5	(4 evaluaciones)	45 €	90€ en total

Click here to ask Blackbox to help you code faster |

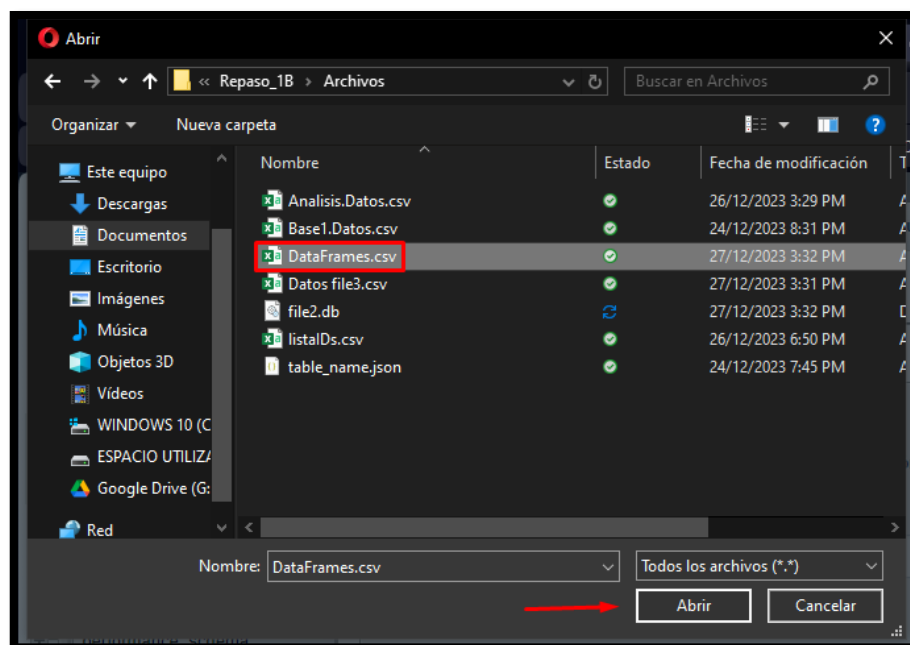
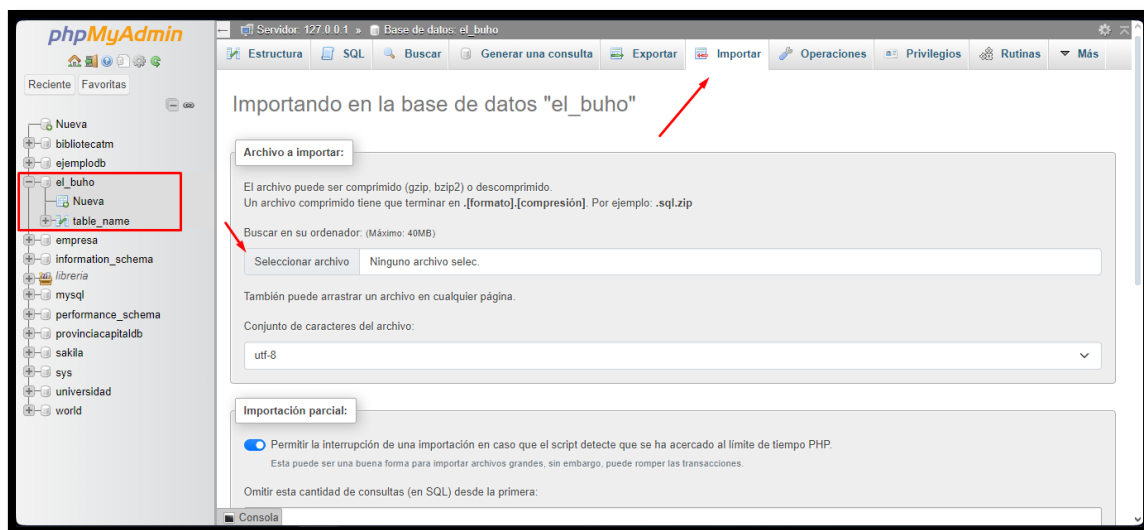
```
# Filtrar y mostrar las filas que cumplen con la condicion (ultimas 5 filas)
Dataframe[Dataframe["Valoracion"]==valor_alto].tail()
```

✓ 0.0s Python

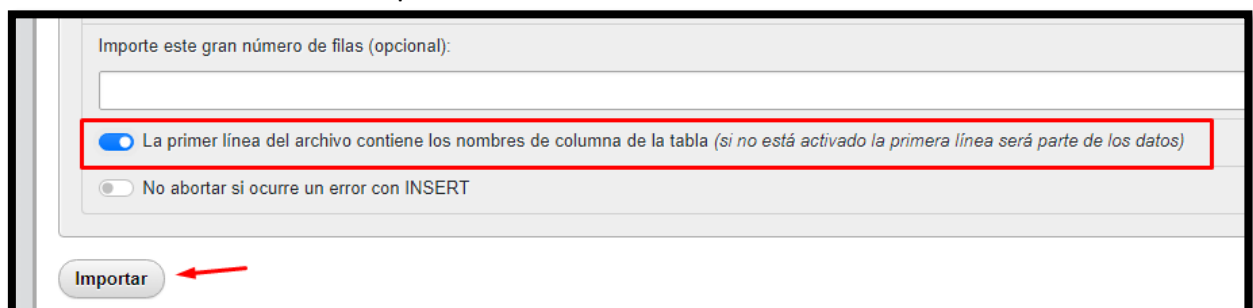
_id	Tipo	nombreAlojamiento	Zona	Nhab	Servicios	Fechas	Valoracion	Evaluaciones	Noche	Total
7d4586d805966de2	Habitación privada	Beach modern spacious room Queen bed in Poblenou	Sant Marti	2 huéspedes · 1 dormitorio · 1 cama · 1 baño c...	Wifi · Cocina · Lavadora · Aire acondicionado	17 dic. – 19 dic.	5	(12 evaluaciones)	53 €	106€ en total
7d4586d805966de2	Habitación privada	Room to rent in central Barcelona near la rambla	Ciutat Vella	2 huéspedes · 1 dormitorio · 1 baño privado	Wifi · Cocina · Lavadora	17 dic. – 19 dic.	5	(3 evaluaciones)	64 €	127€ en total
7d4586d805966de2	Habitación privada	Sagrada familia, céntrico exterior.Piso partic...	La Sagrada Familia	2 huéspedes · 1 dormitorio · 1 cama · 1,5 baño...	Wifi · Cocina · Lavadora · Aire acondicionado	17 dic. – 19 dic.	5	(4 evaluaciones)	45 €	90€ en total
7d4586d805966de2	Habitación de hostel	Habitación Doble con baño privado - 2 camas	Ciutat Vella	2 huéspedes · 1 dormitorio	Wifi · Aire acondicionado	24 dic. – 26 dic.	5	(7 evaluaciones)	69 €	138€ en total

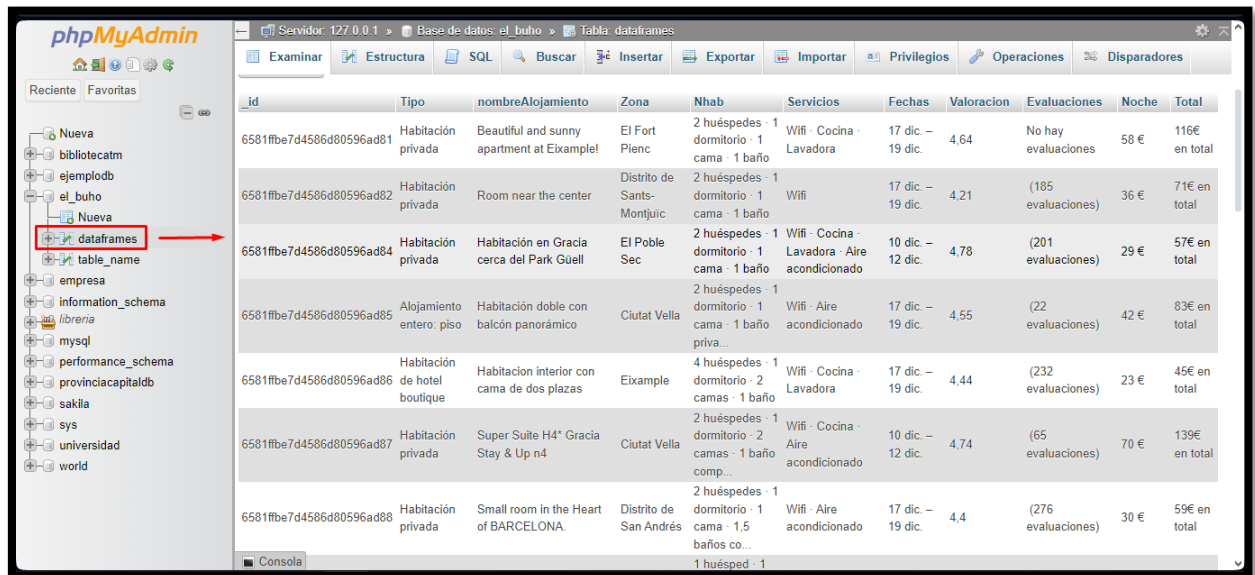
11. Exportar a MySQL el archivo unificado.

Crear una tabla en la base de datos creada “el_buho” e importar el archivo “DataFrames.csv” que contiene los 4 datasets unificados



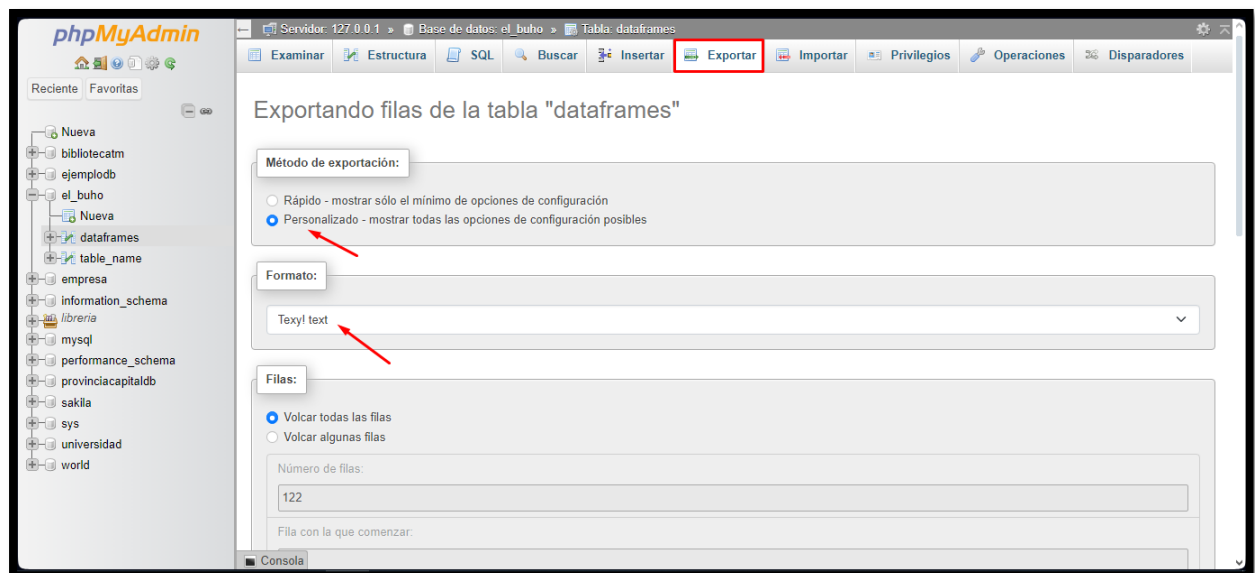
Marcar la siguiente opción para indicar que la fila 1 de los dato del archivo corresponde a los atributos de la tabla e importar los datos





id	Tipo	nombreAlojamiento	Zona	Nhab	Servicios	Fechas	Valoracion	Evaluaciones	Noche	Total
6581f1be7d4586d80596ad81	Habitación privada	Beautiful and sunny apartment at Eixample!	El Fort Pienc	2 huéspedes - 1 dormitorio - 1 cama - 1 baño	Wifi - Cocina - Lavadora	17 dic. - 19 dic.	4,64	No hay evaluaciones	58 €	116€ en total
6581f1be7d4586d80596ad82	Habitación privada	Room near the center	District de Sants-Montjuic	2 huéspedes - 1 dormitorio - 1 cama - 1 baño	Wifi	17 dic. - 19 dic.	4,21	(185 evaluaciones)	36 €	71€ en total
6581f1be7d4586d80596ad84	Habitación privada	Habitación en Gracia cerca del Park Güell	El Poble Sec	2 huéspedes - 1 dormitorio - 1 cama - 1 baño	Wifi - Cocina - Lavadora - Aire acondicionado	10 dic. - 12 dic.	4,78	(201 evaluaciones)	29 €	57€ en total
6581f1be7d4586d80596ad85	Alojamiento entero: piso	Habitación doble con balcón panorámico	Ciutat Vella	2 huéspedes - 1 dormitorio - 1 cama - 1 baño privada...	Wifi - Aire acondicionado	17 dic. - 19 dic.	4,55	(22 evaluaciones)	42 €	83€ en total
6581f1be7d4586d80596ad86	Habitación de hotel boutique	Habitacion interior con cama de dos plazas	Eixample	4 huéspedes - 1 dormitorio - 2 camas - 1 baño	Wifi - Cocina - Lavadora	17 dic. - 19 dic.	4,44	(232 evaluaciones)	23 €	45€ en total
6581f1be7d4586d80596ad87	Habitación privada	Super Suite H4* Gracia Stay & Up n4	Ciutat Vella	2 huéspedes - 1 dormitorio - 2 camas - 1 baño comp...	Wifi - Cocina - Aire acondicionado	10 dic. - 12 dic.	4,74	(65 evaluaciones)	70 €	139€ en total
6581f1be7d4586d80596ad88	Habitación privada	Small room in the Heart of BARCELONA.	District de San Andrés	2 huéspedes - 1 dormitorio - 1 cama - 1.5 baños co...	Wifi - Aire acondicionado	17 dic. - 19 dic.	4,4	(276 evaluaciones)	30 €	59€ en total

12. Además, exporte el archivo unificado con extensión .txt.



Exportando filas de la tabla "dataframes"

Método de exportación:

☐ Rápido - mostrar sólo el mínimo de opciones de configuración

☒ Personalizado - mostrar todas las opciones de configuración posibles

Formato:

Texty text

Filas:

☒ Volcar todas las filas

☐ Volcar algunas filas

Número de filas:

122

Fila con la que comenzar:

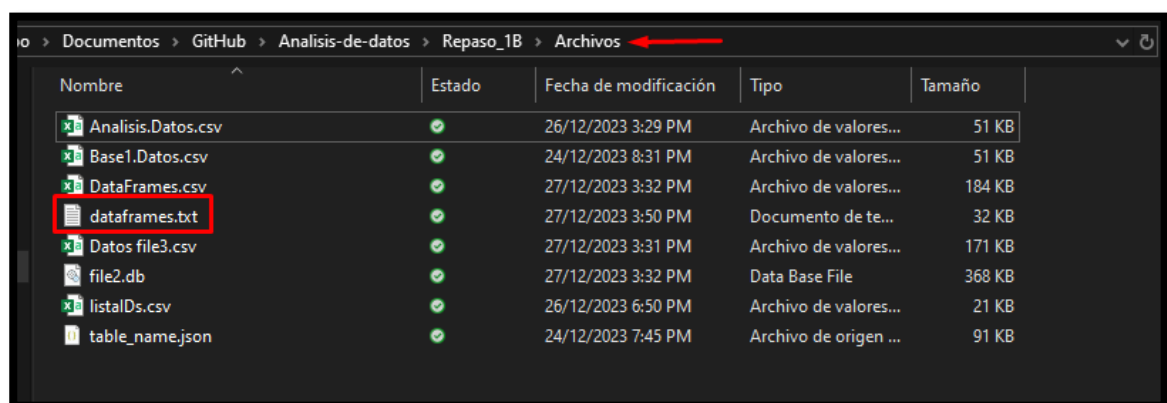
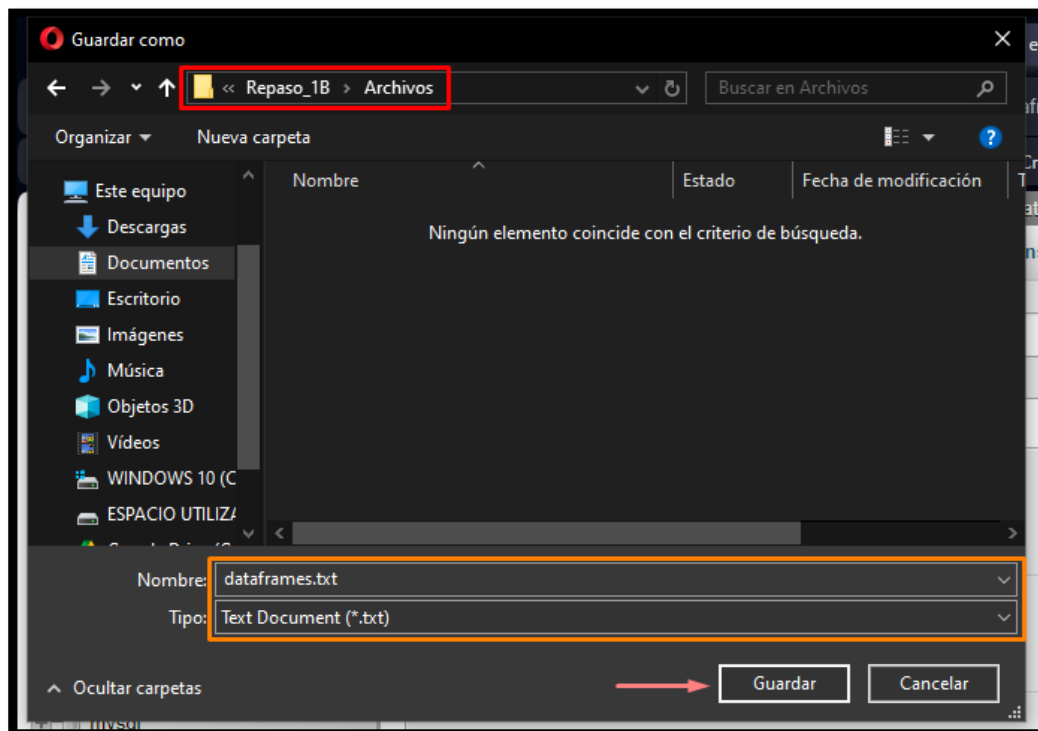


Reemplazar NULL con:

NULL

Exportar

Guardar el archivo en la ruta "Archivos"



PRESENTACIÓN

Al finalizar tu práctica deberás subir:

- 1.- Código fuente <apellido.py> o <apellido.ipynb>
- 2.- pdf con capturas de pantalla de la ejecución de cada ejercicio.
- 3.- Archivos .txt generados

Poner todo en un archivo comprimido con el nombre: (Taller-Refuerzo-AnalisisDatos-
NApellido) y entregarlo en la tarea indicada.

RECURSOS NECESARIOS

- Material de clase
- Internet
- MySQL

- MongoDB
- MongoDB Atlas
- SQLite
- Json, CSV
- Jupyter
- Pandas, NumPy