**ESCUELA POLITÉCNICA NACIONAL**

ESCUELA DE FORMACIÓN DE TECNÓLOGOS

**ANÁLISIS DE DATOS**

|  |  |
| --- | --- |
|  | |
| ASIGNATURA: | Análisis de Datos |
| PROFESOR:  FECHA: | Ing. Lorena Chulde / Ing. Juan Pablo Zaldumbide  25 – 11 - 2024 |
| PERÍODO ACADÉMICO: | 2024-B |
|  | |

**CORRECCION**

**EXAMEN DEL PRIMER BIMESTRE**

|  |
| --- |
|  |
|  |
|  |
|  |

**Nombre:**

Erick Nuñez

**PROPÓSITO**

Evaluar los conocimientos adquiridos en el primer trimestre

**INDICACIONES**

Se presentan algunos datasets: archivo1.csv, archivo2.json, archivo3.db y archivo4.csv. Que contienen información sobre vehículos usados, como sus características, precios y opciones de negociación.

Descripción general de las columnas más relevantes:

* + Año: Año del modelo del vehículo.
  + Kilometraje: Distancia recorrida por el vehículo.
  + Precio: Precio del vehículo (en miles).
  + Lugar: Ubicación de venta (e.g., Quito, Guayaquil, Loja).
  + Negociación: Condición de negociación del precio (e.g., "Negociable").
  + Categoría: Condición del vehículo (e.g., "USED" para usado).
  + Marca y Modelo: Marca y modelo del vehículo.
  + Subtipo: Tipo específico del vehículo (e.g., "Todoterreno").

<https://www.kaggle.com/datasets/joshuastay/ecuador-used-cars-dataset>

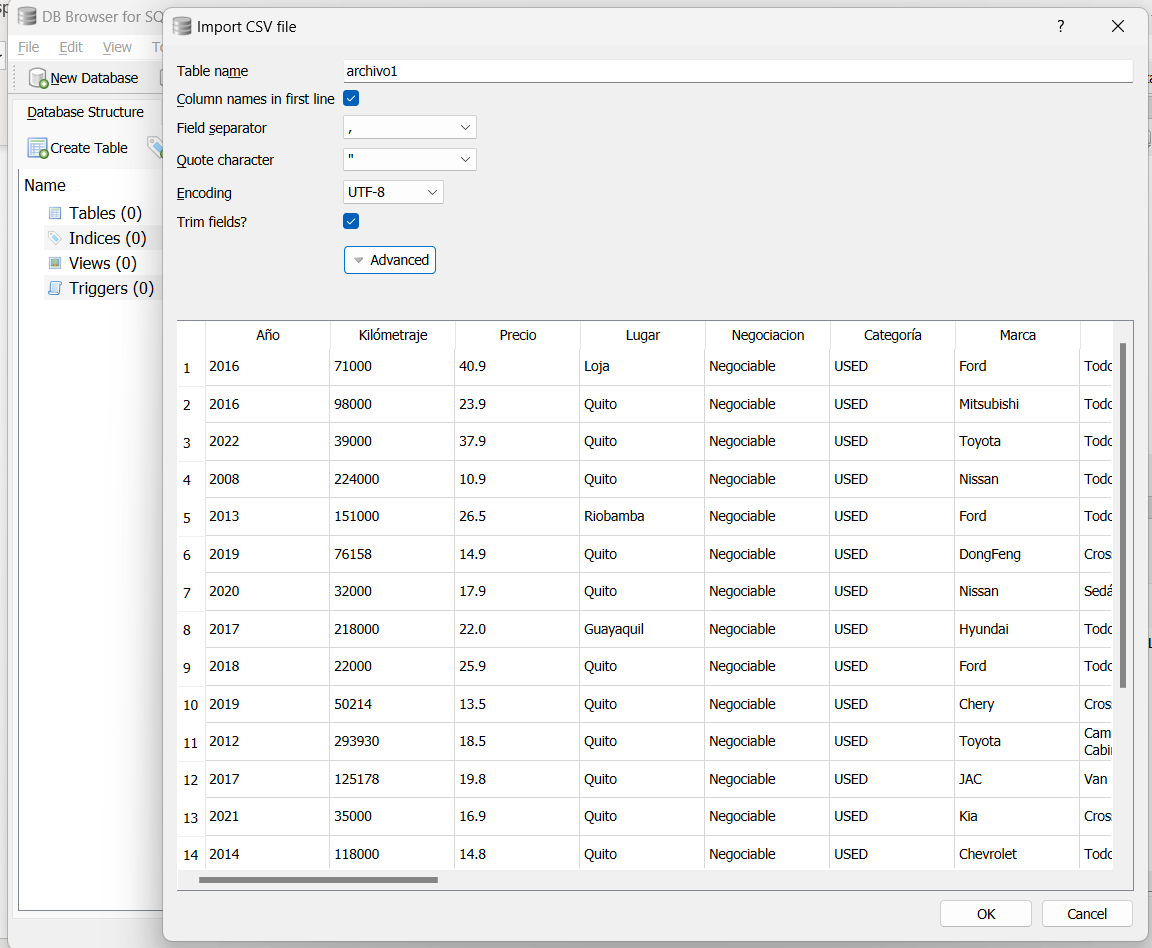
**DESARROLLO**

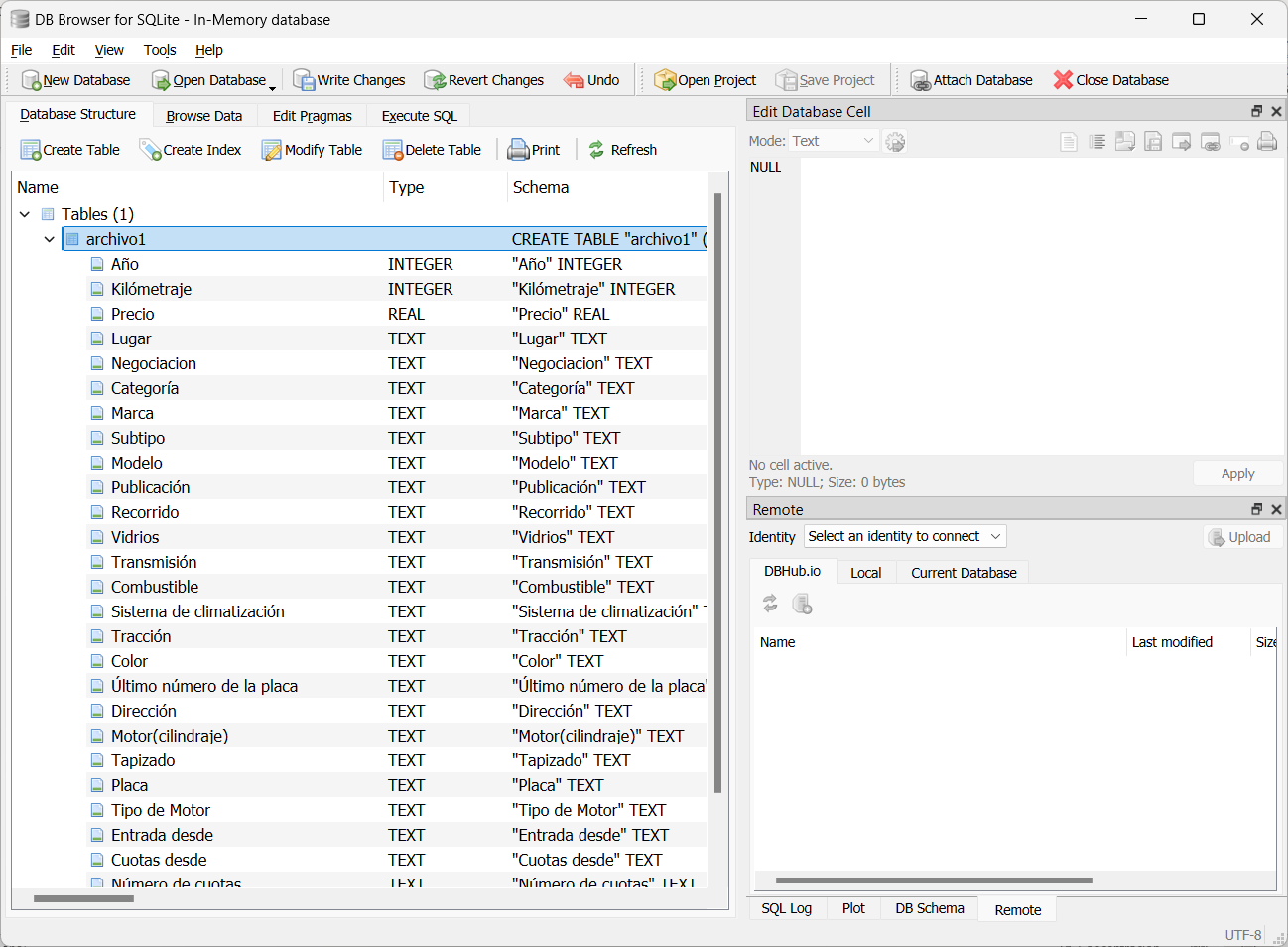
**CARGA DE DATOS**

**(10 puntos)**

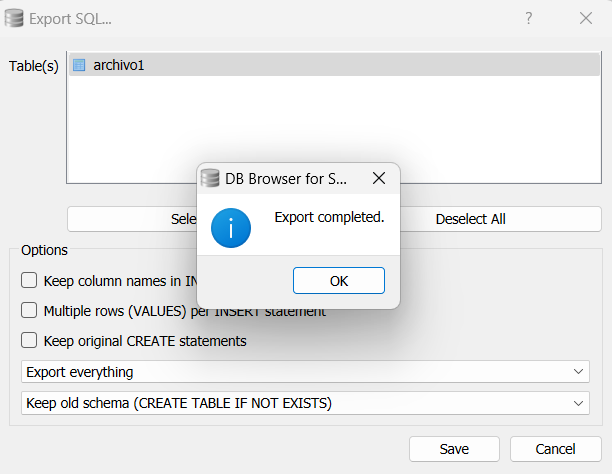
Implementar el proceso de carga de datos, cada item corresponde al número de la ilustración.

1. El archivo llamado archivo1.csv debe ser importado a SQLite.

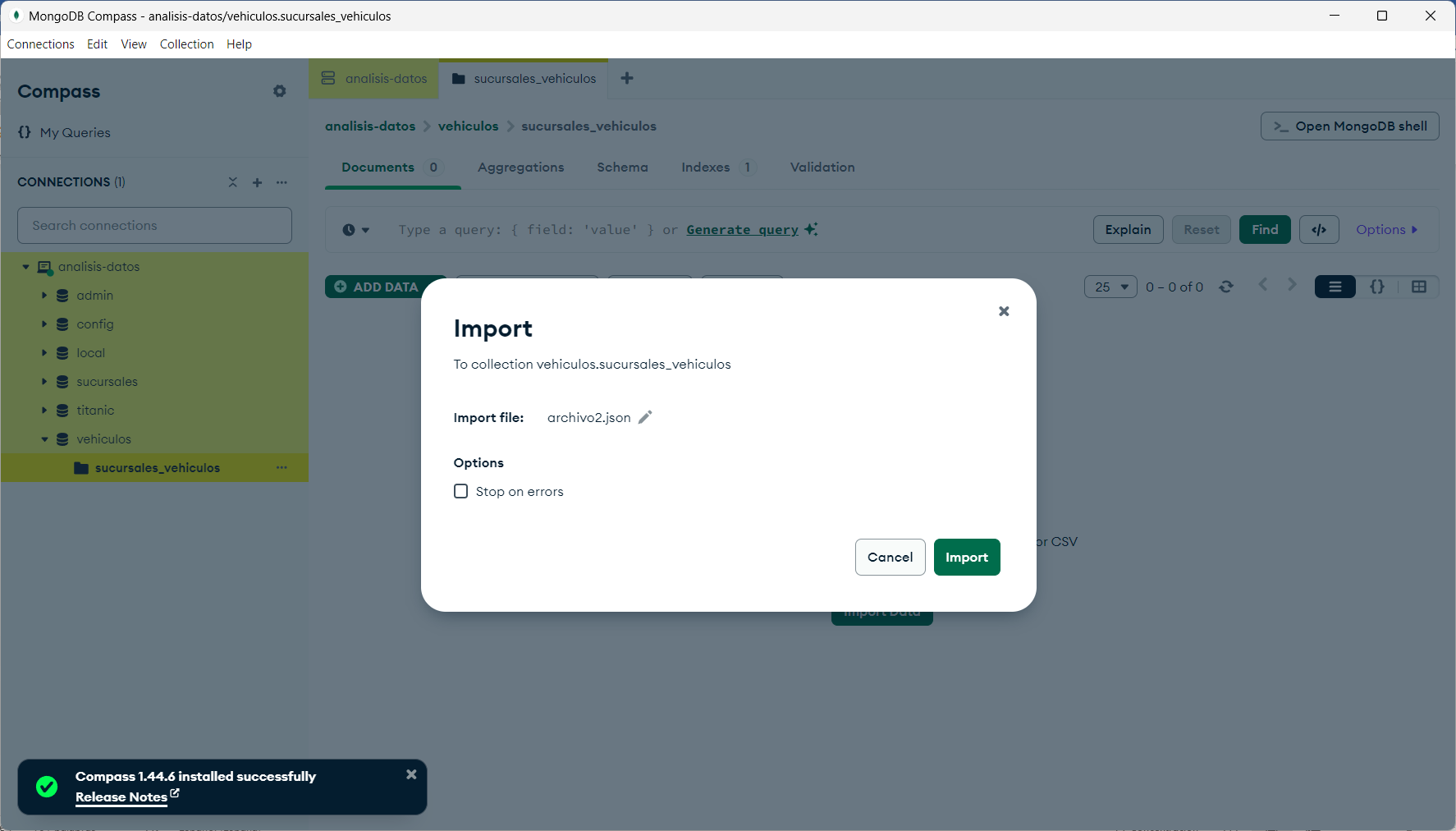
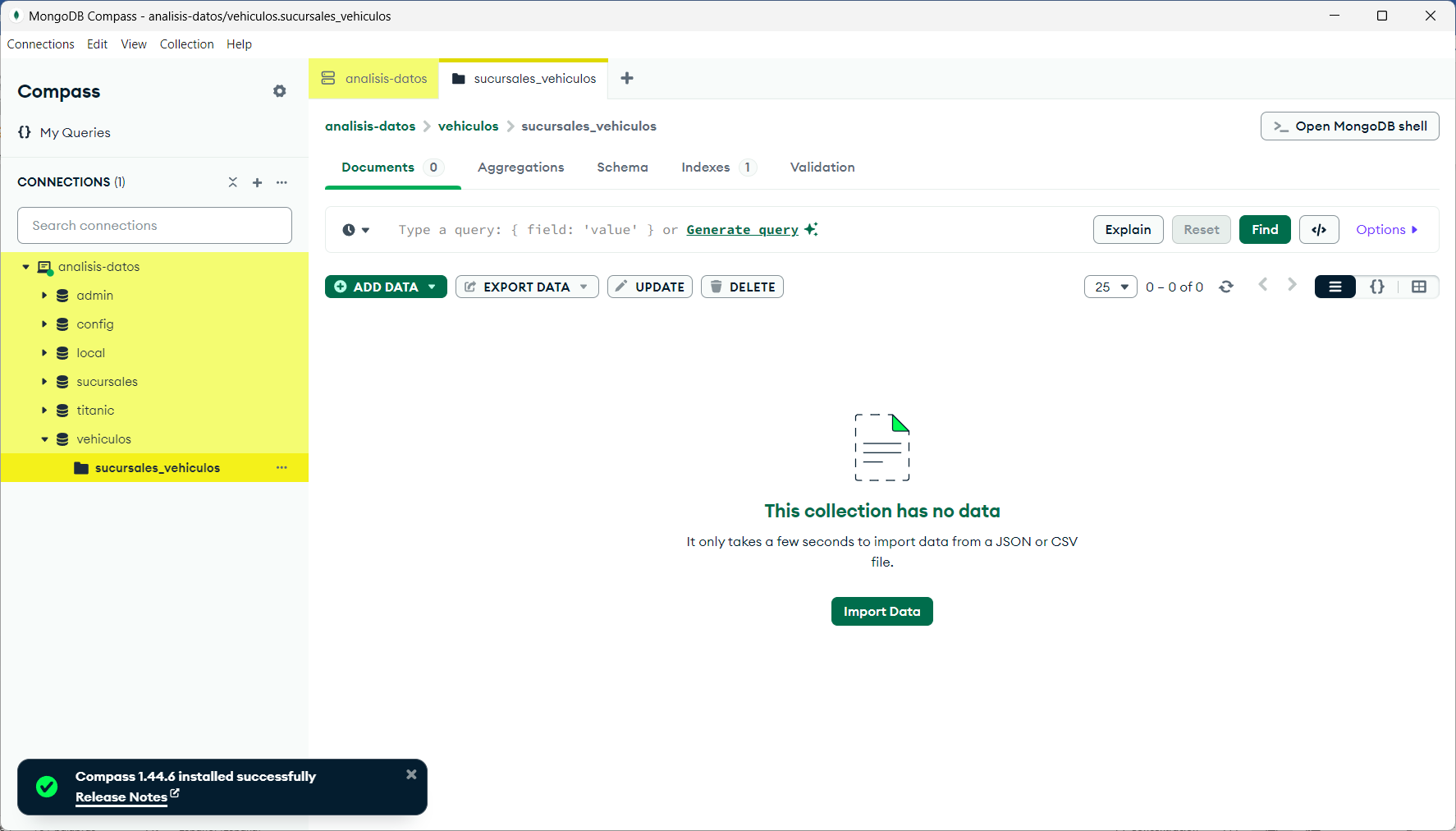


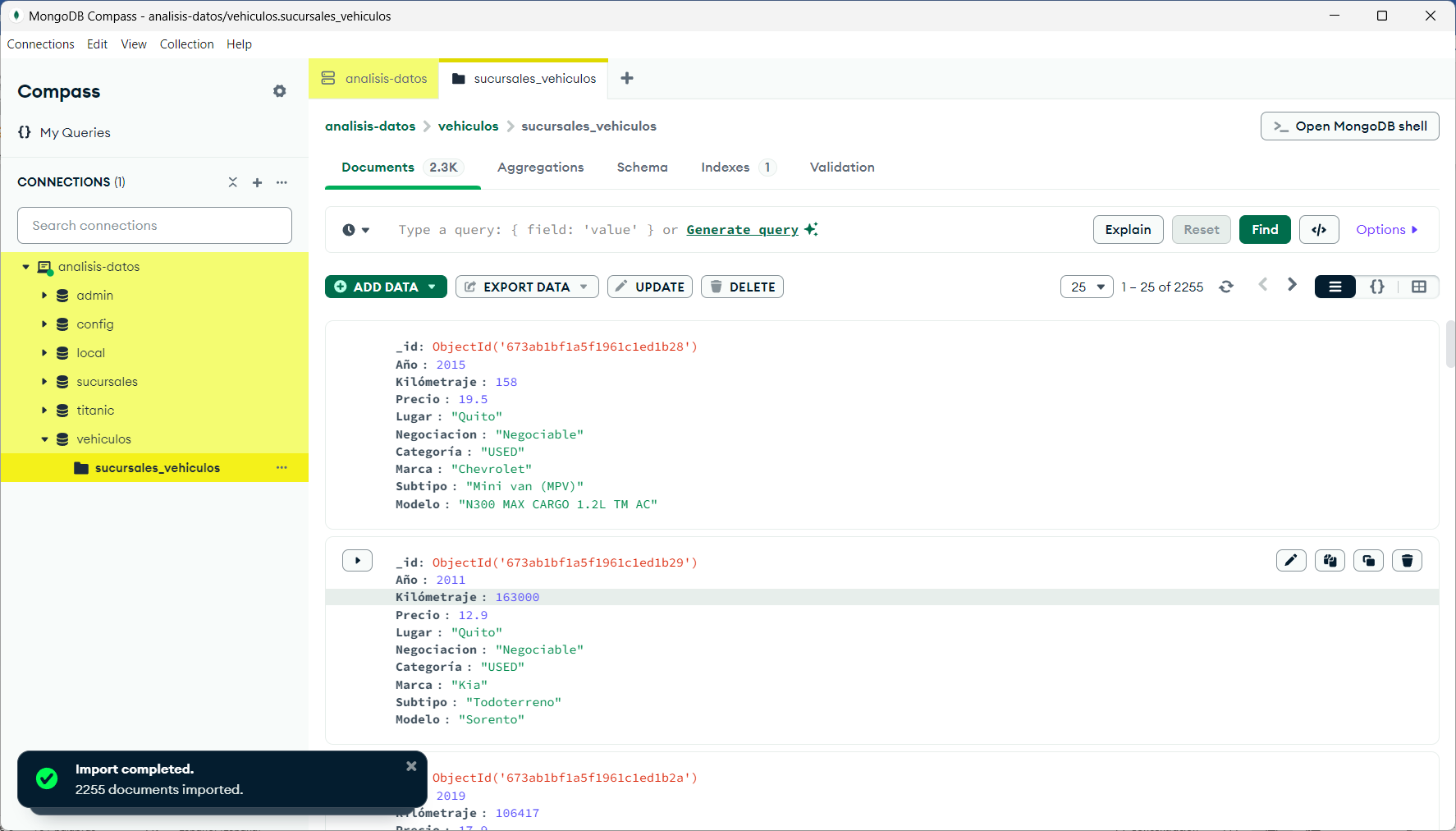


**Exportamos la Base de Datos en formato (.csv)**

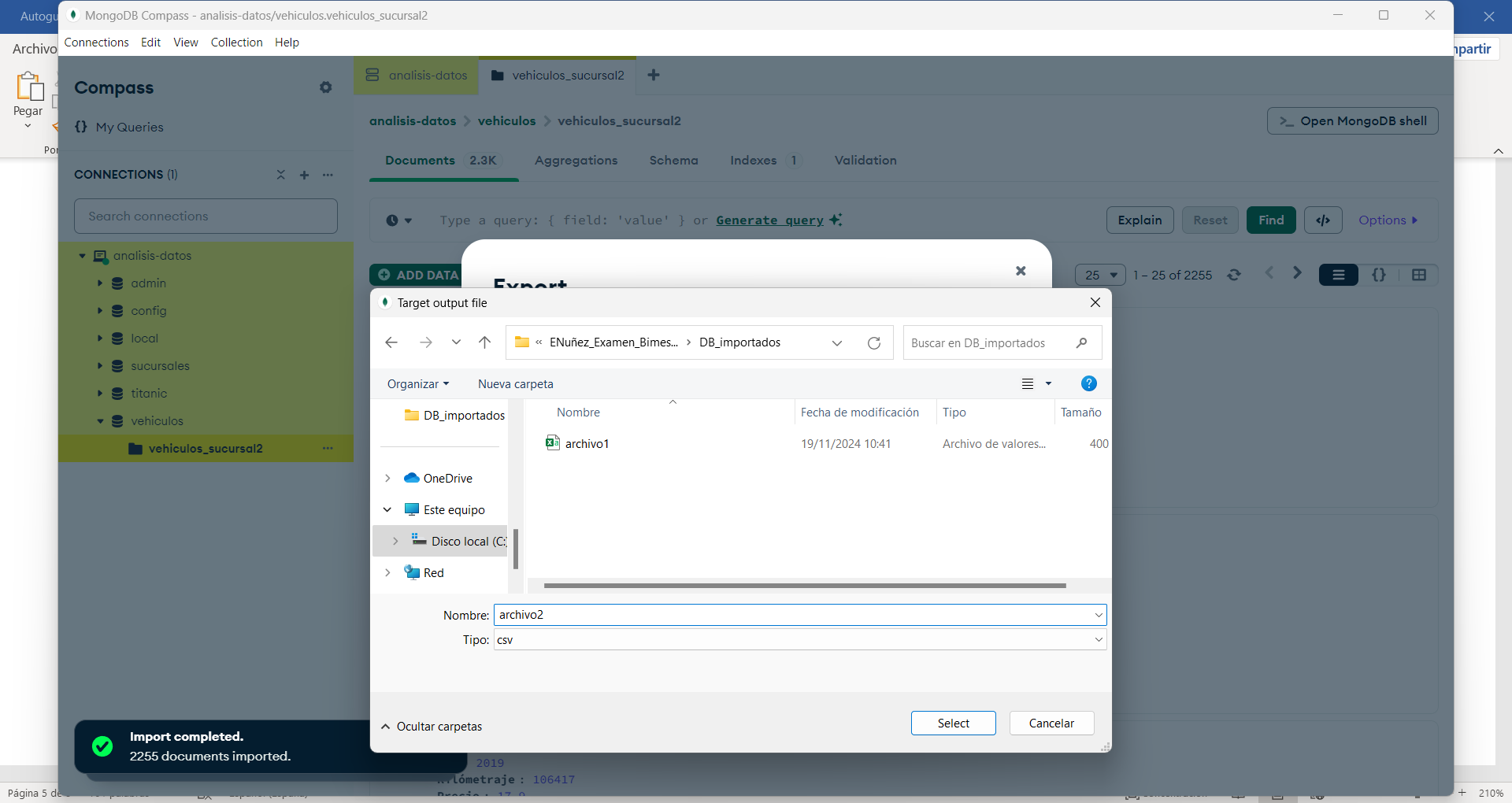


1. El archivo llamado archivo2.json debe ser importado a mongoDB.

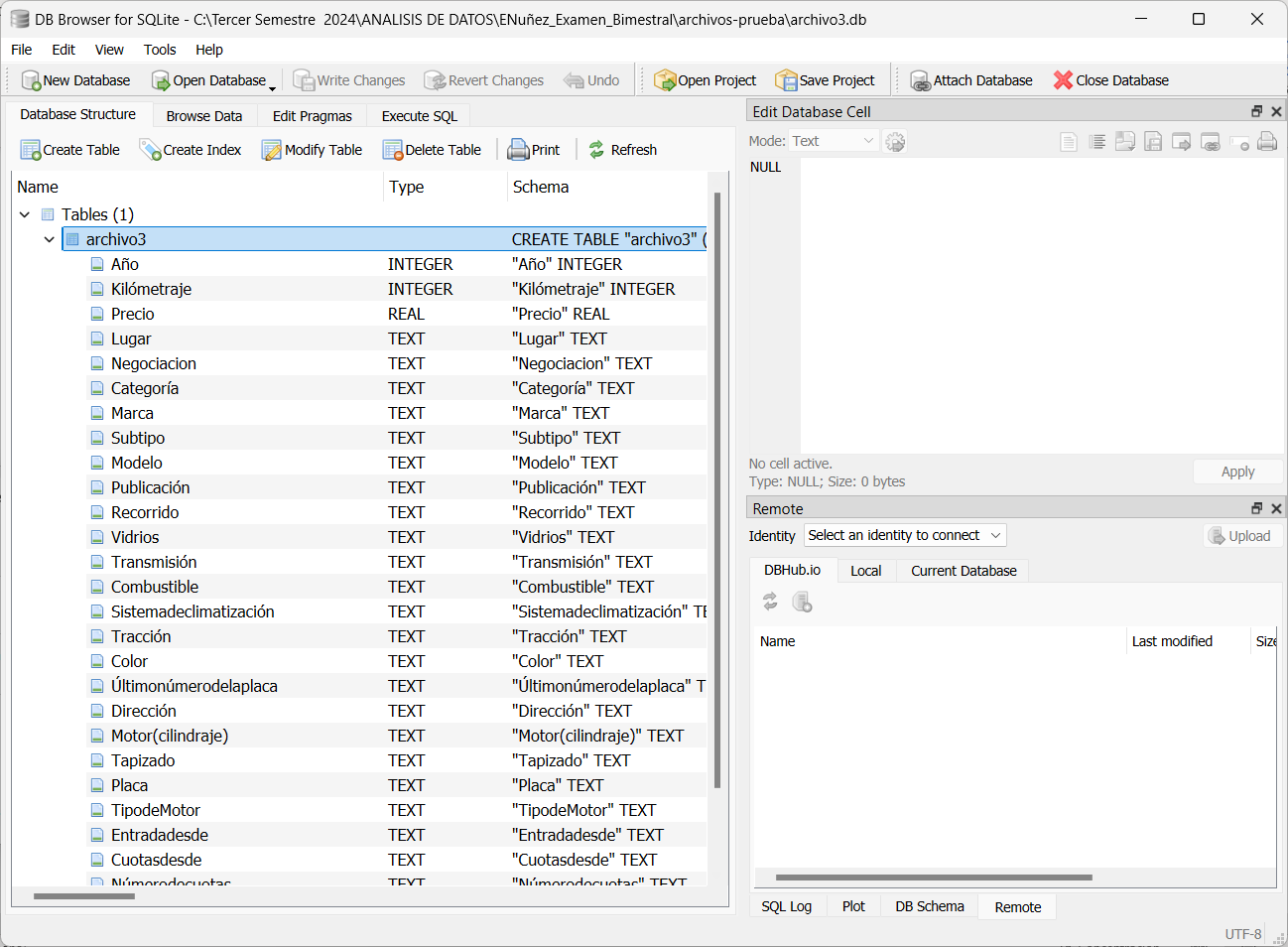




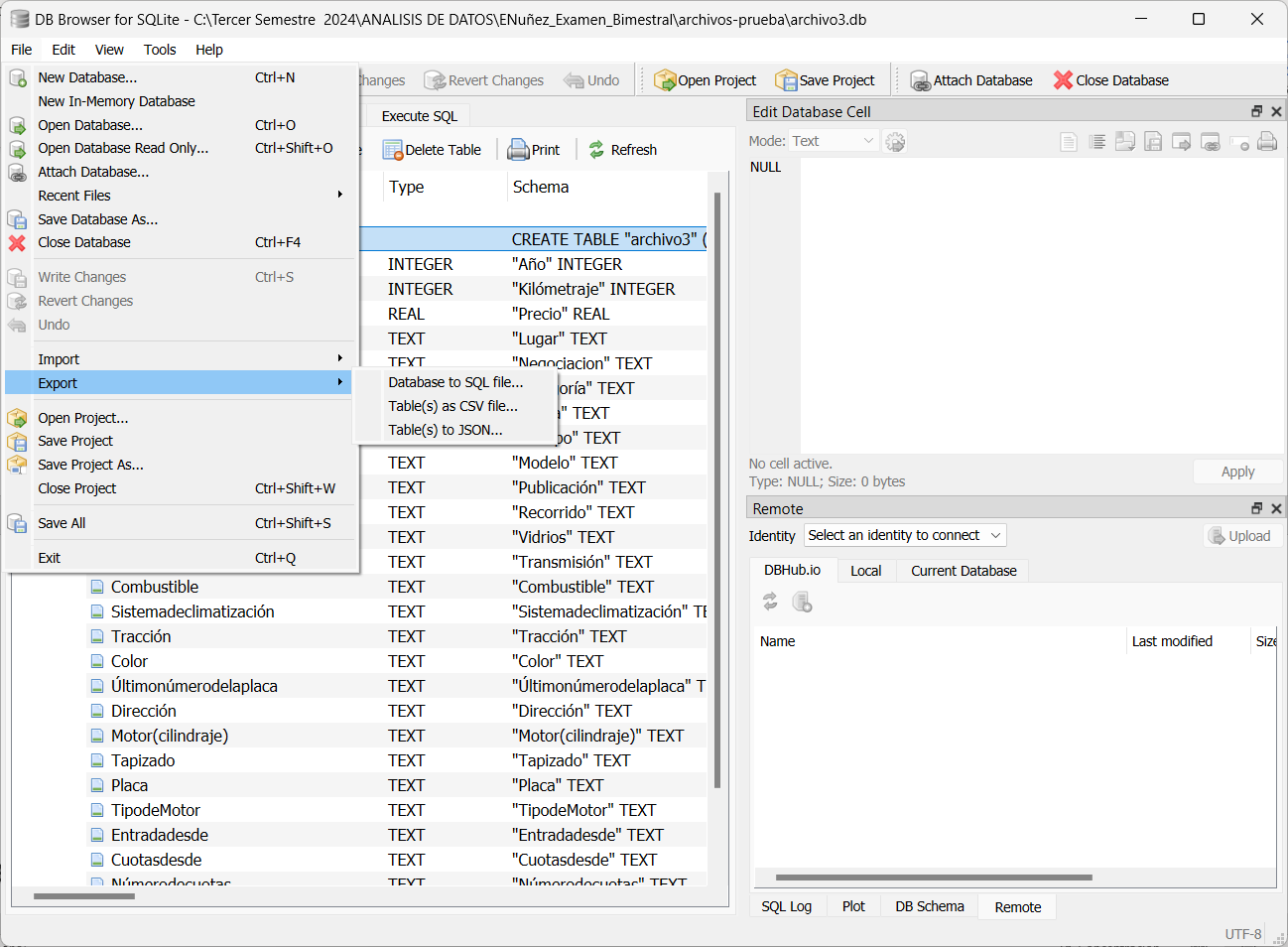
**Exportamos la Base de Datos en formato (.csv)**

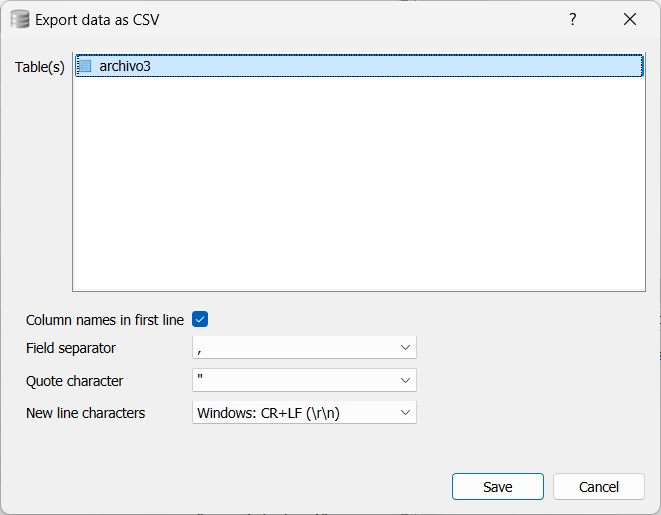


1. El archivo llamado archivo3.db debe ser abierto con SQLite.

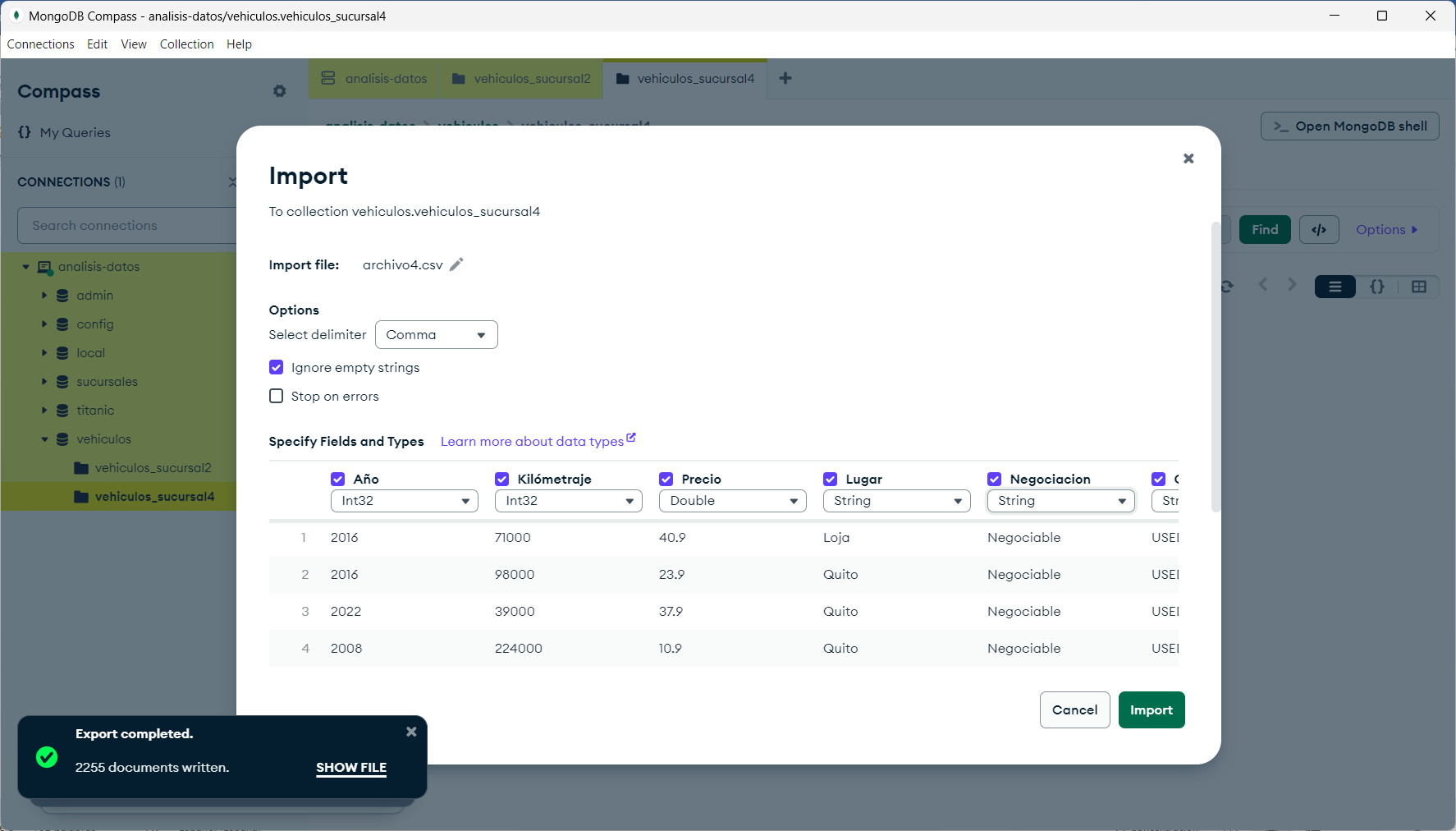


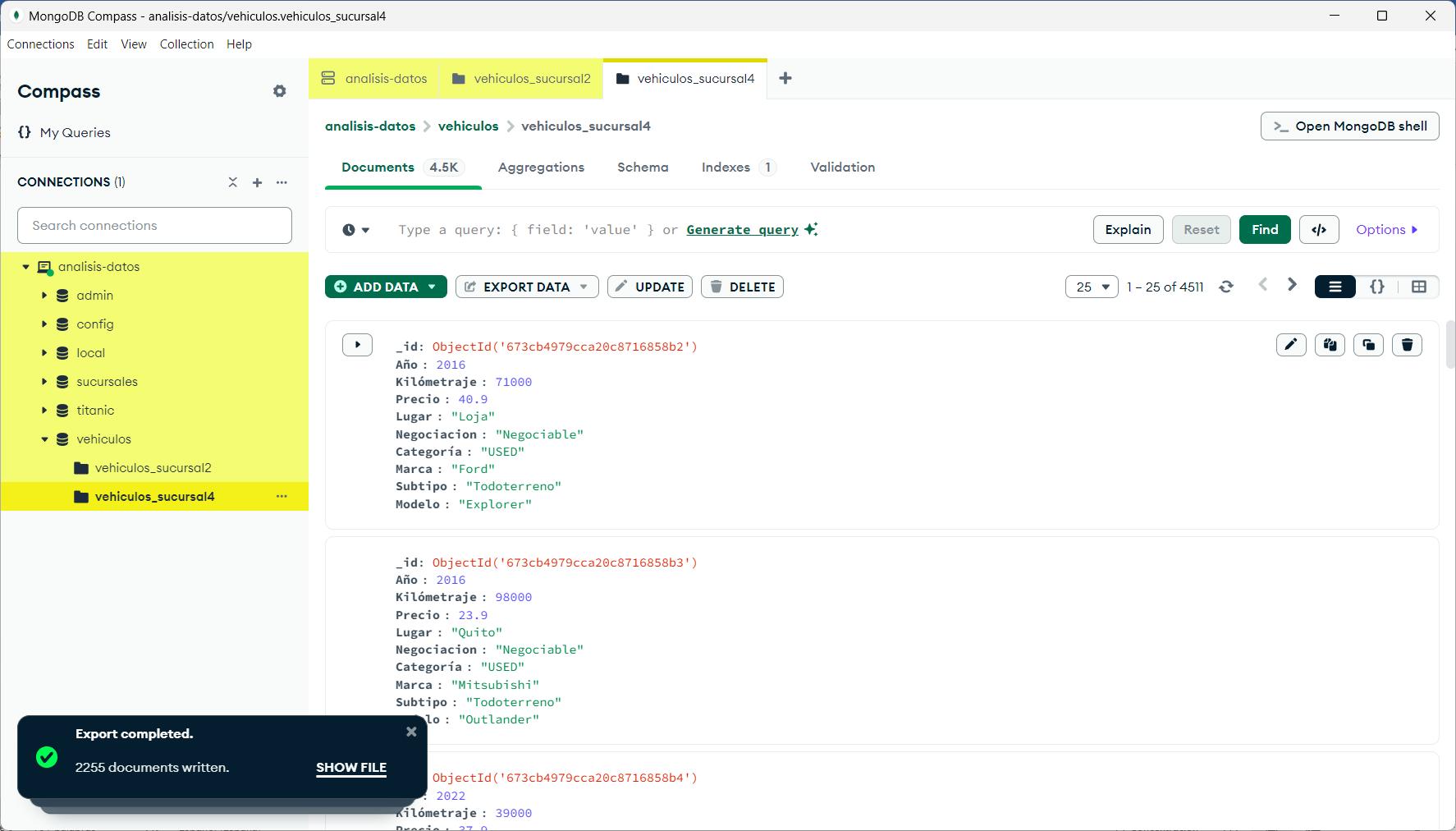
**Exportamos la Base de Datos en formato (.csv)**



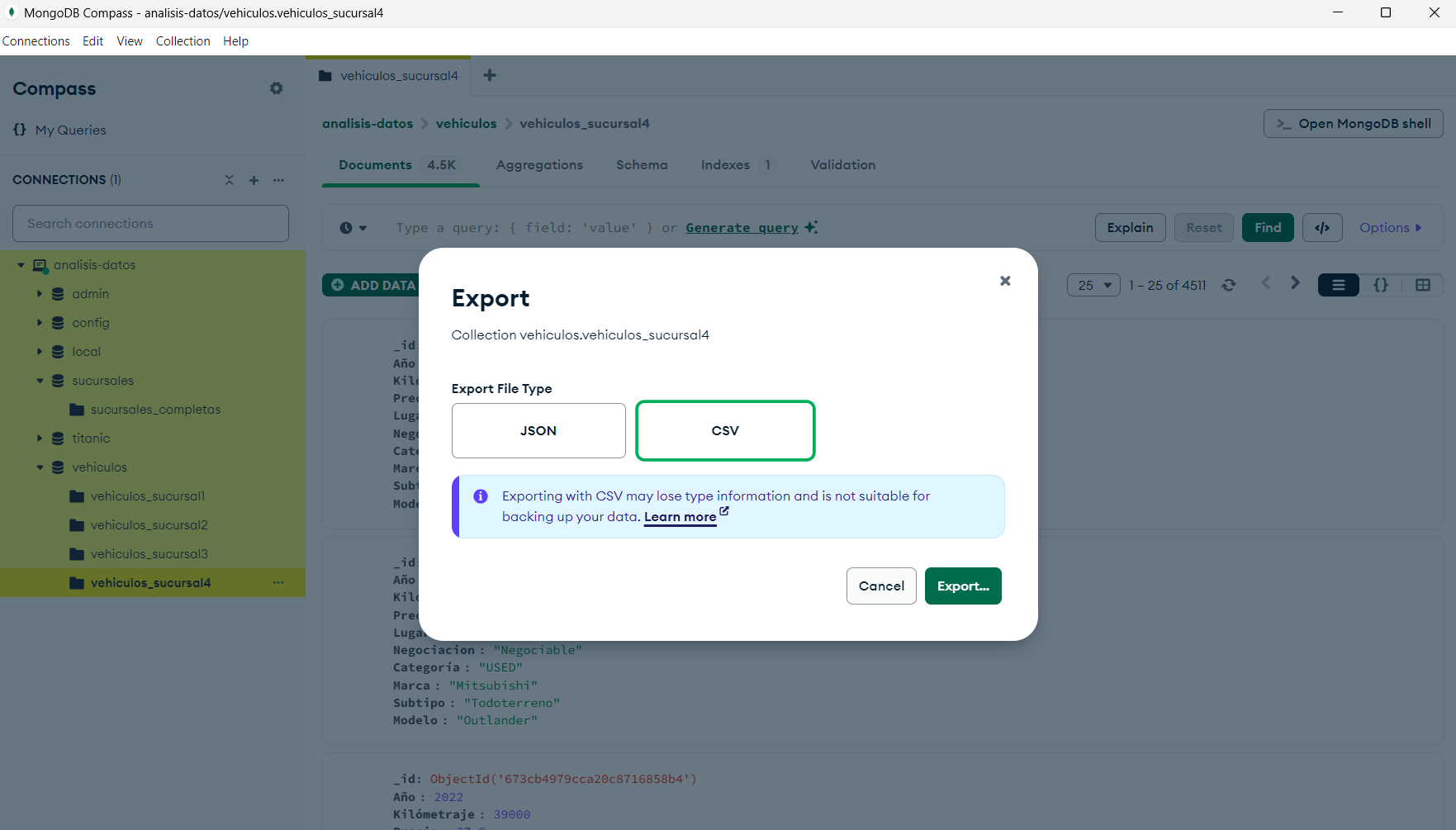


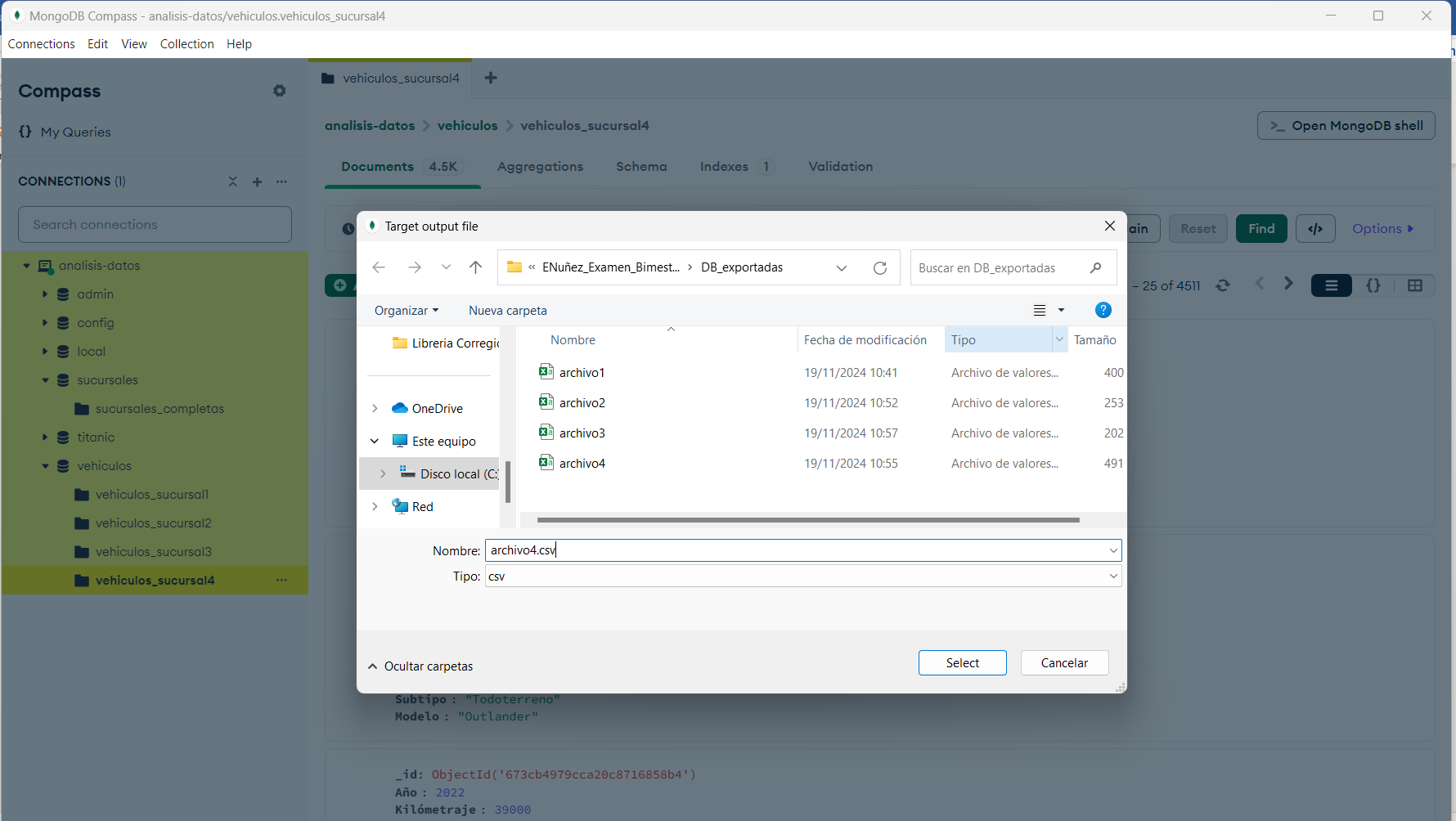
1. El archivo llamado archivo4.csv debe ser importado a mongoDB.



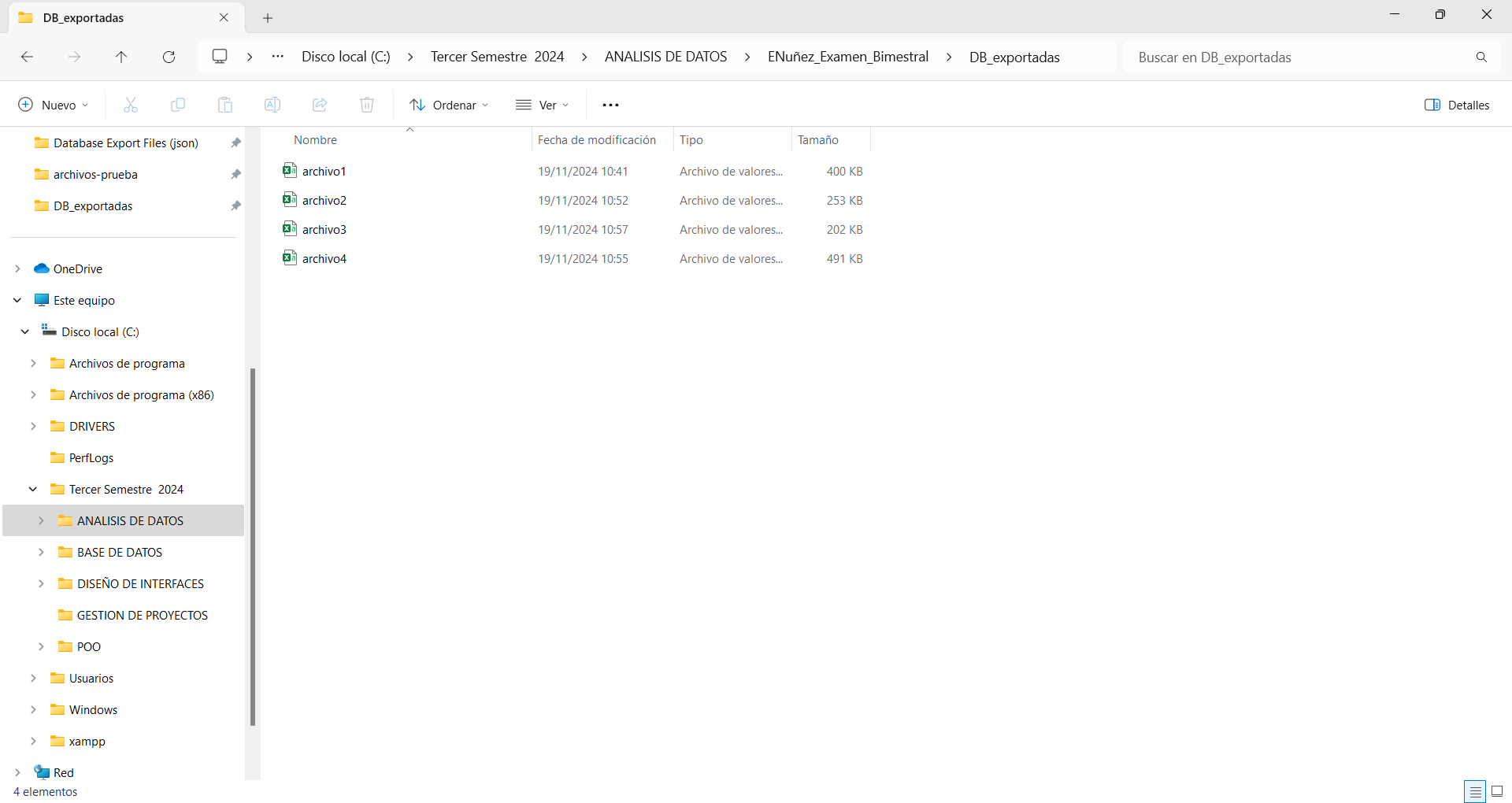


**Exportamos la Base de Datos en formato (.csv)**

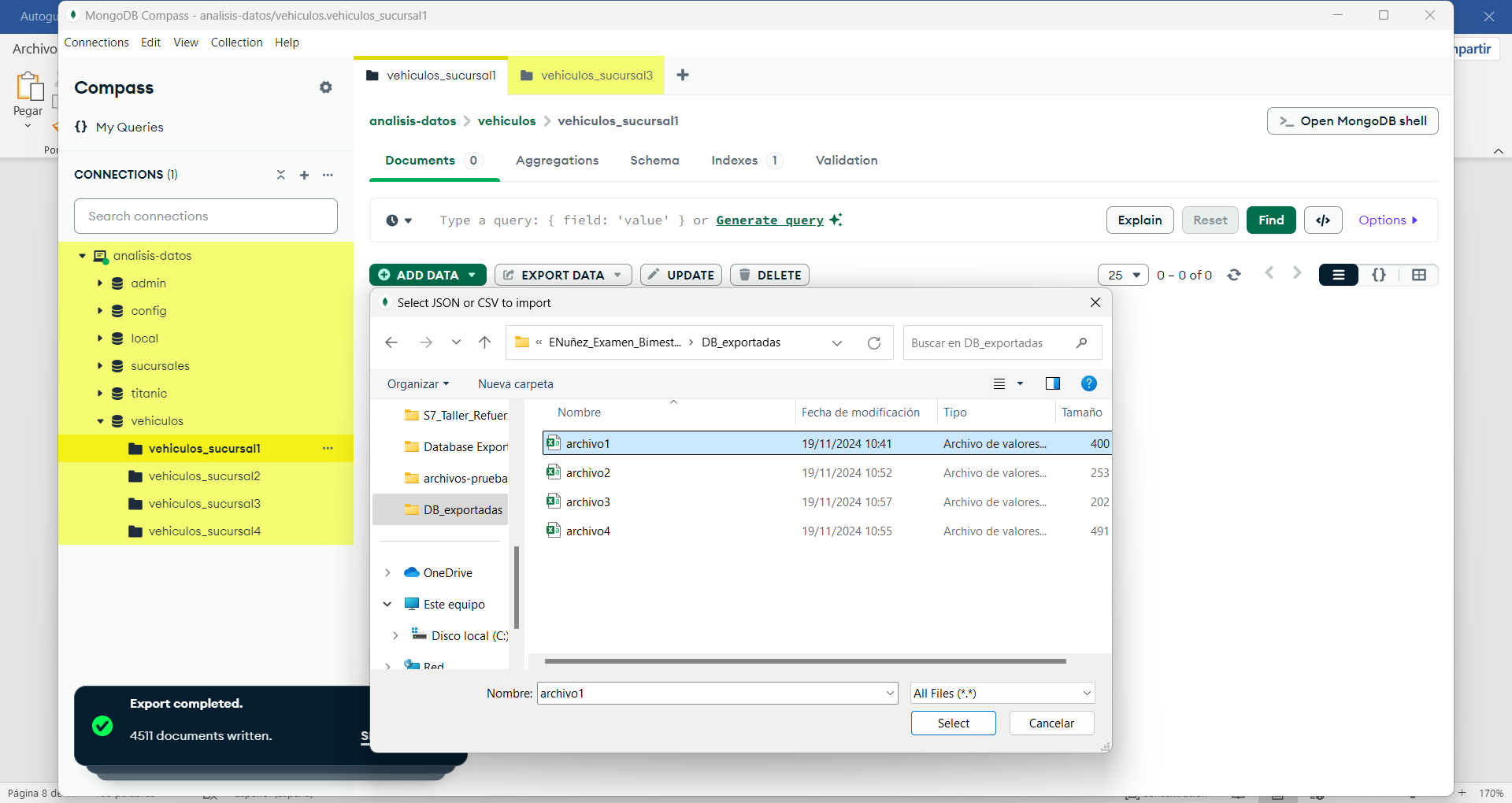


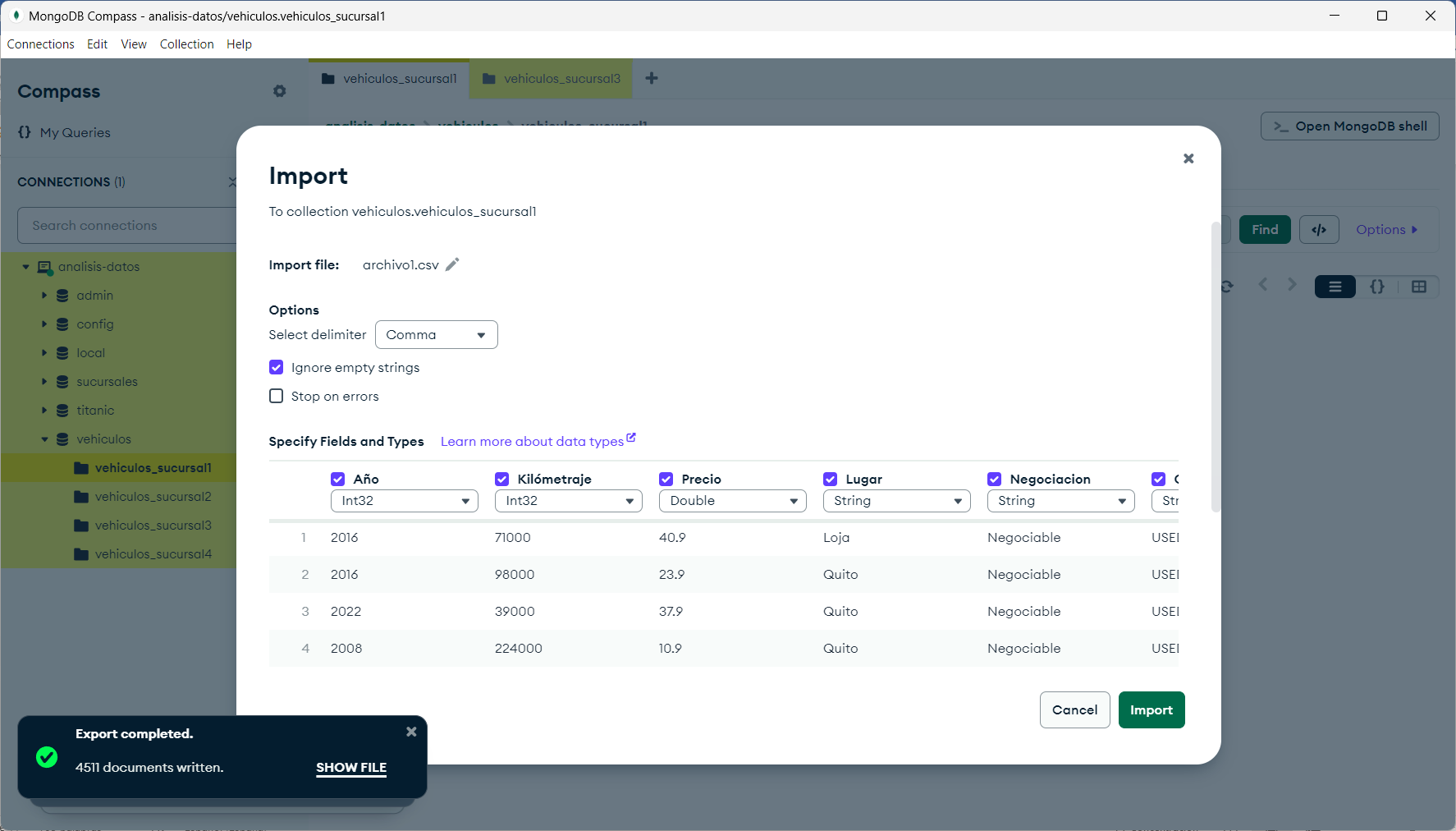


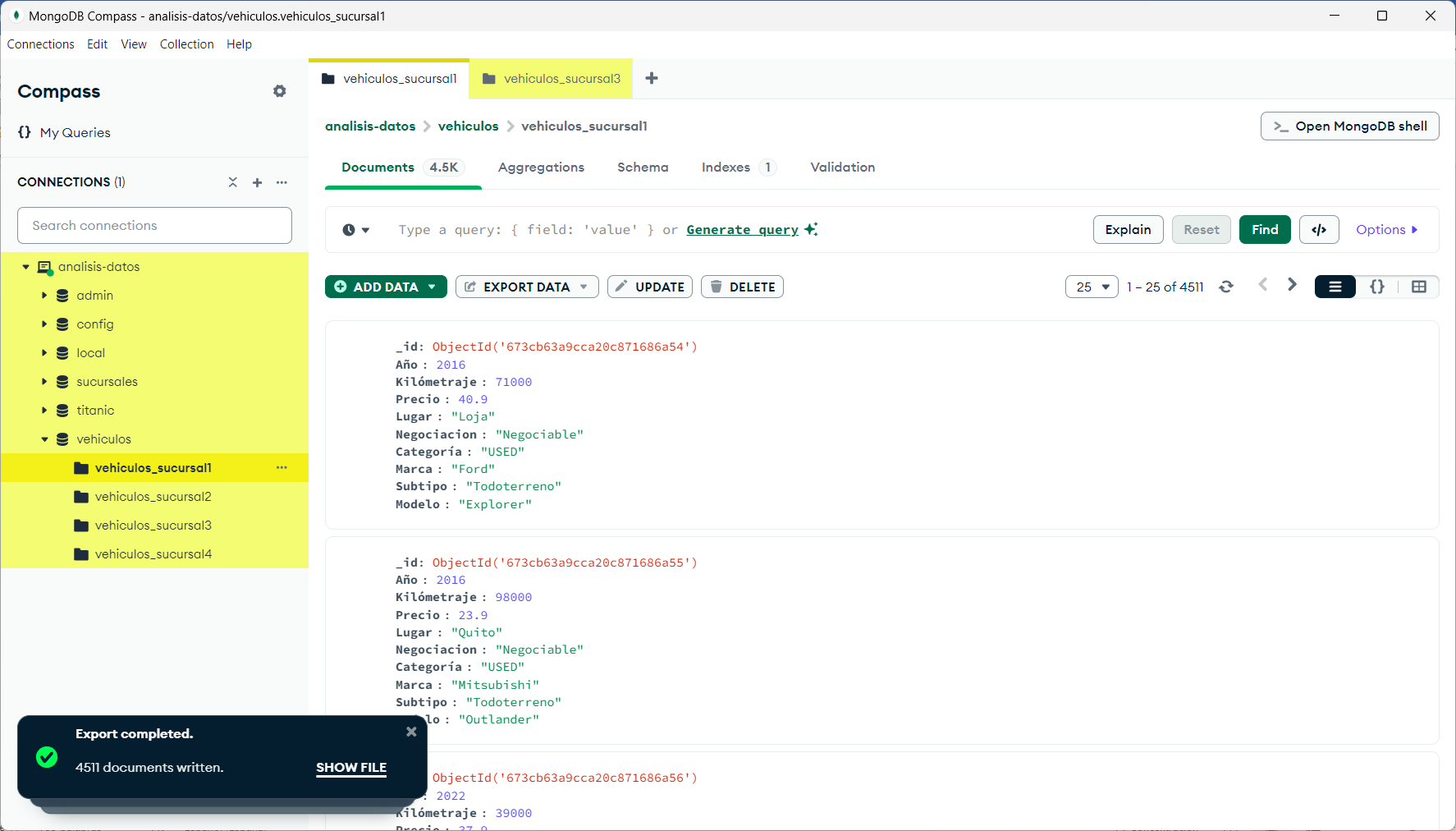
**Bases de Datos Exportadas**



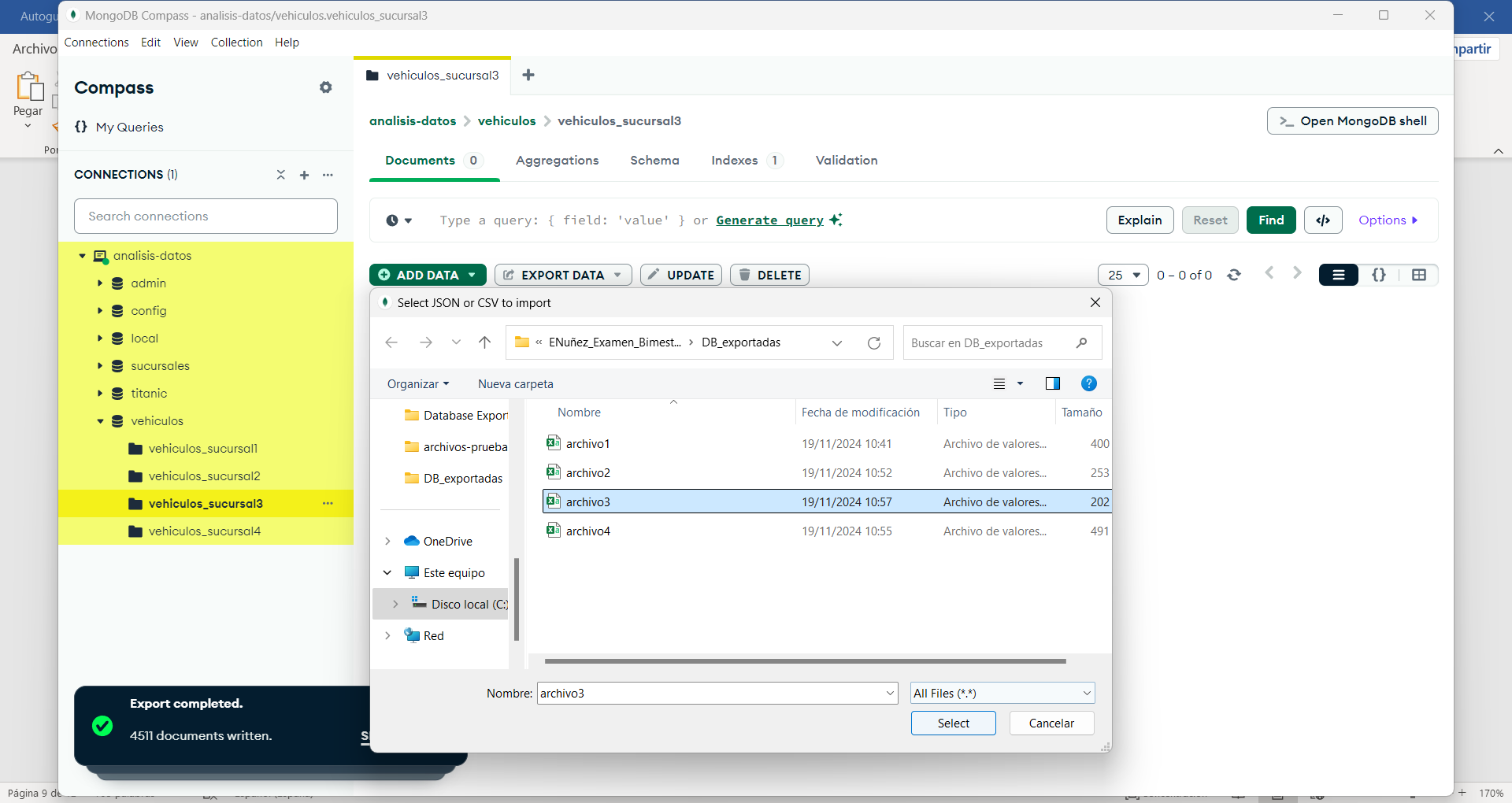
1. El dataframe generado en el paso 1 se lo debe exportar a mongoDB

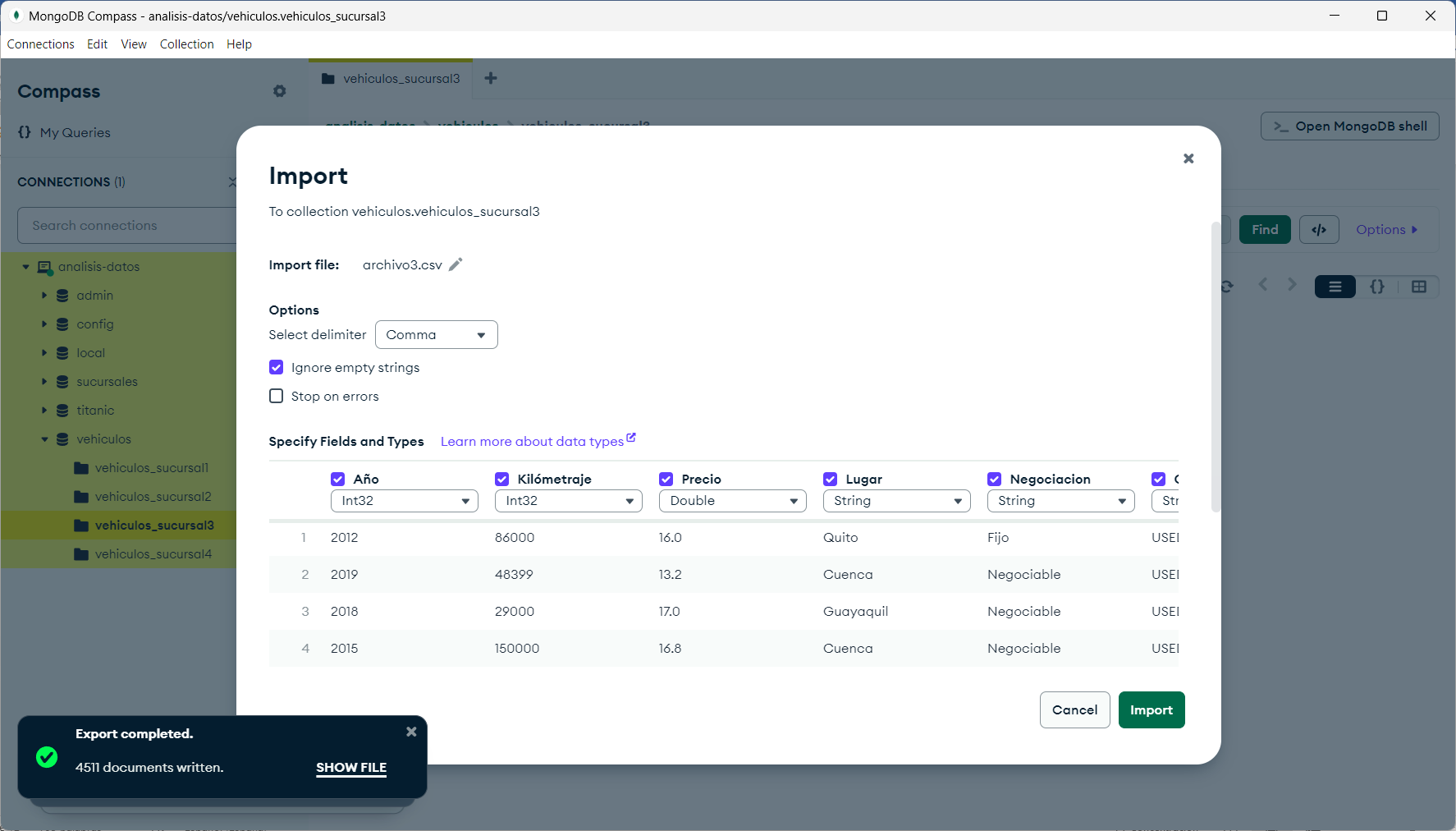


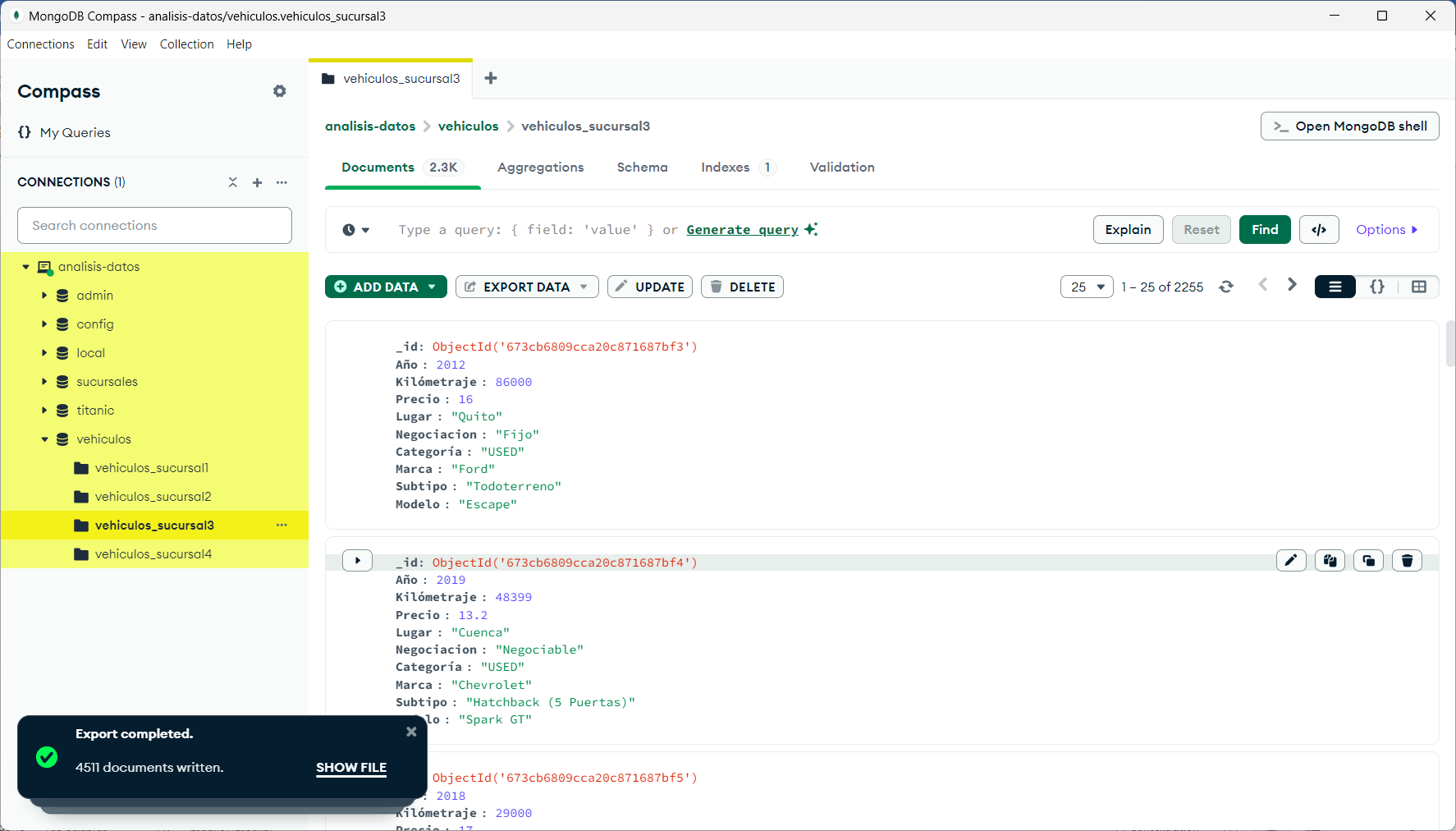


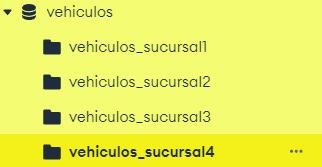


1. El dataframe generado en el paso 3 se lo debe exportar a mongoDB

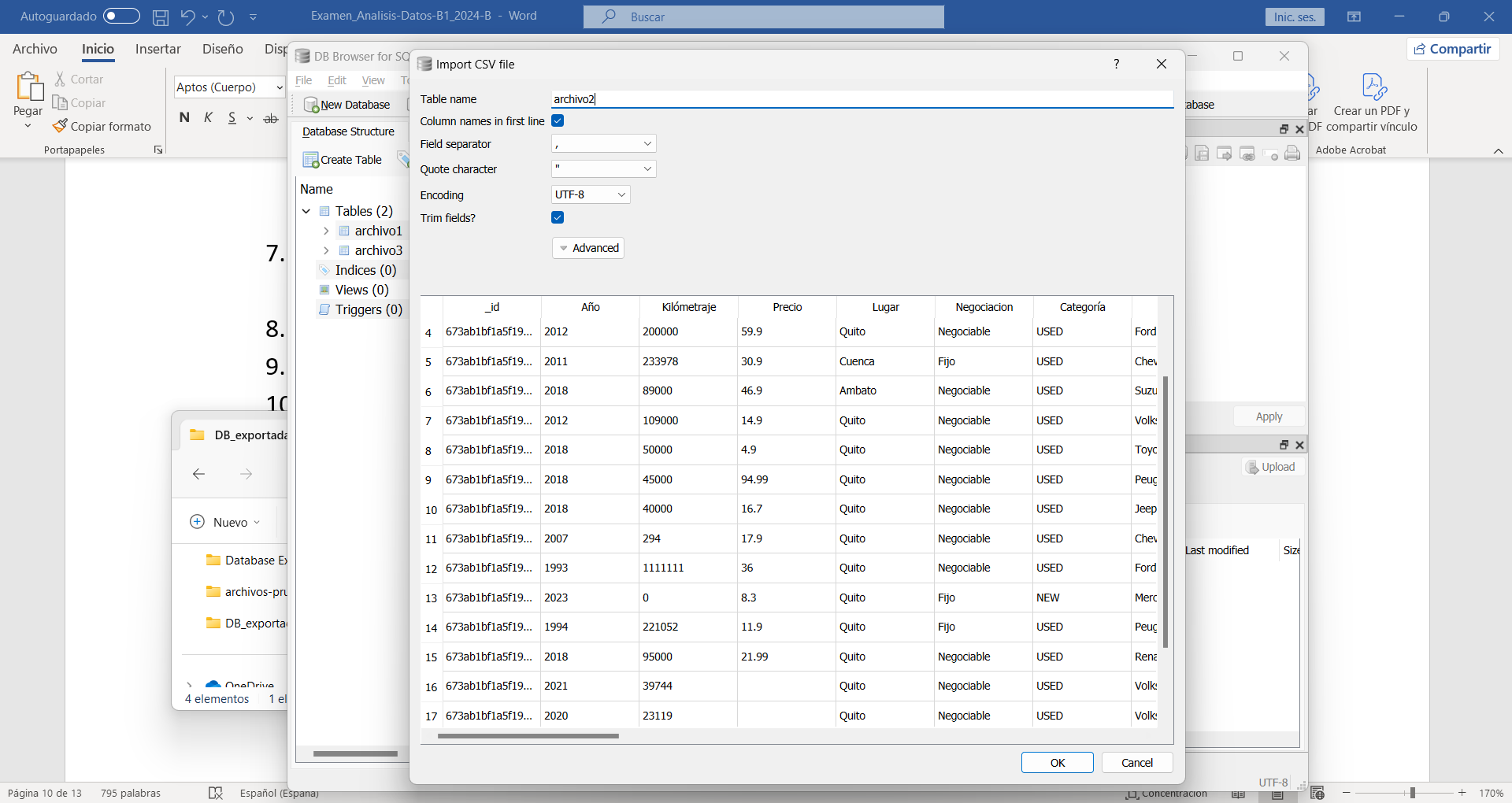


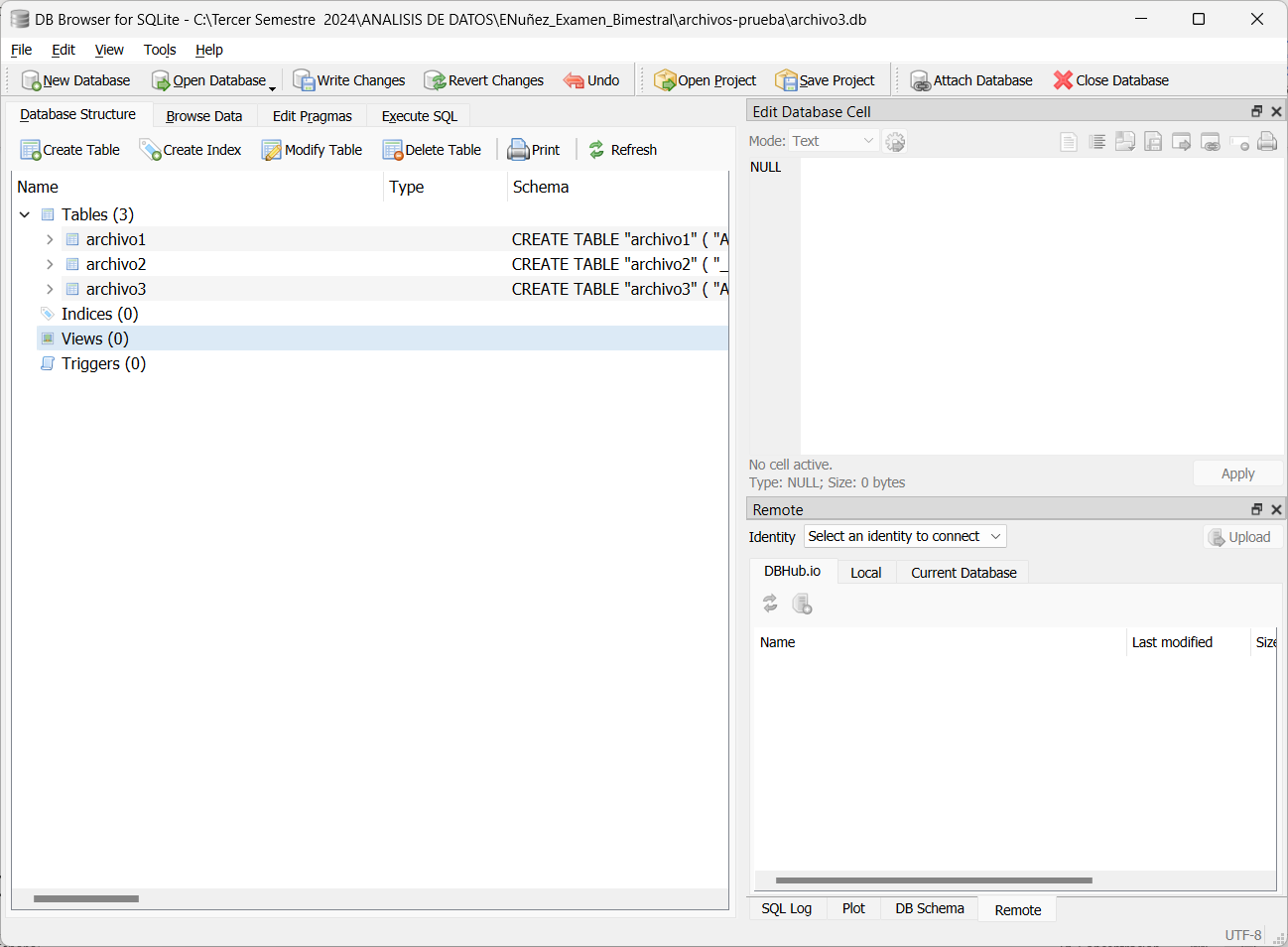




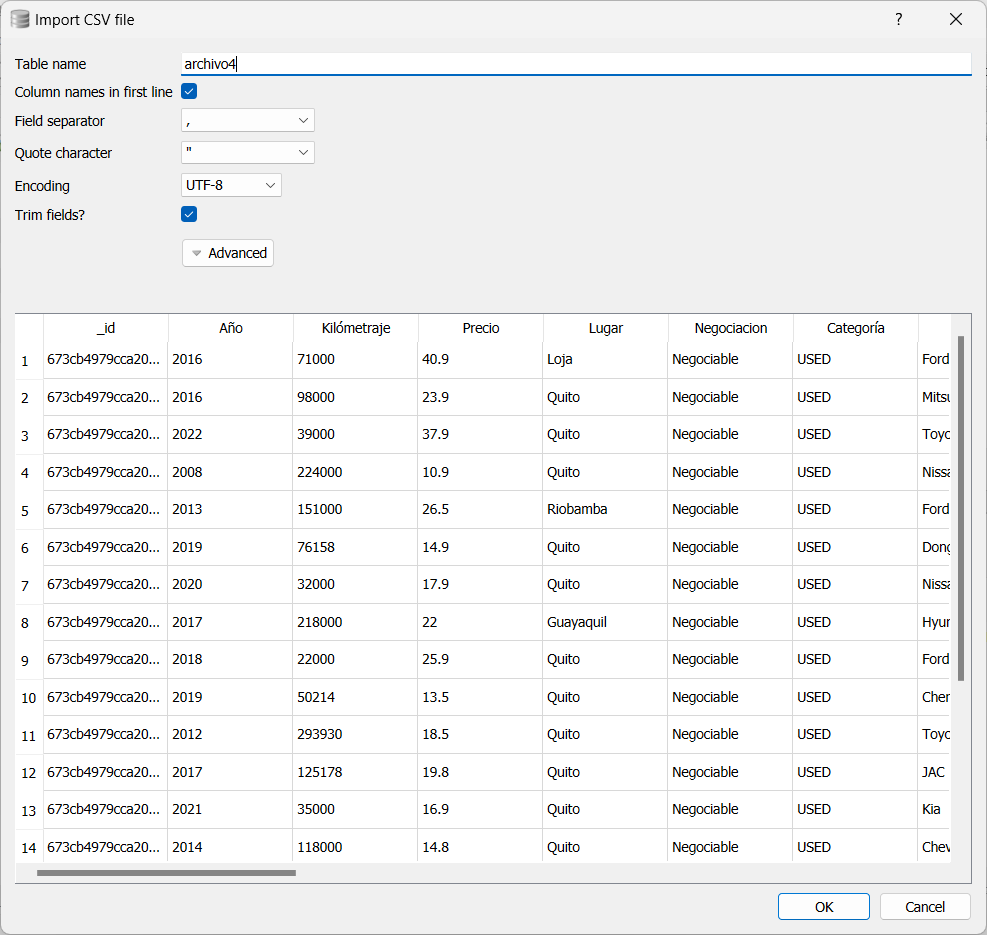


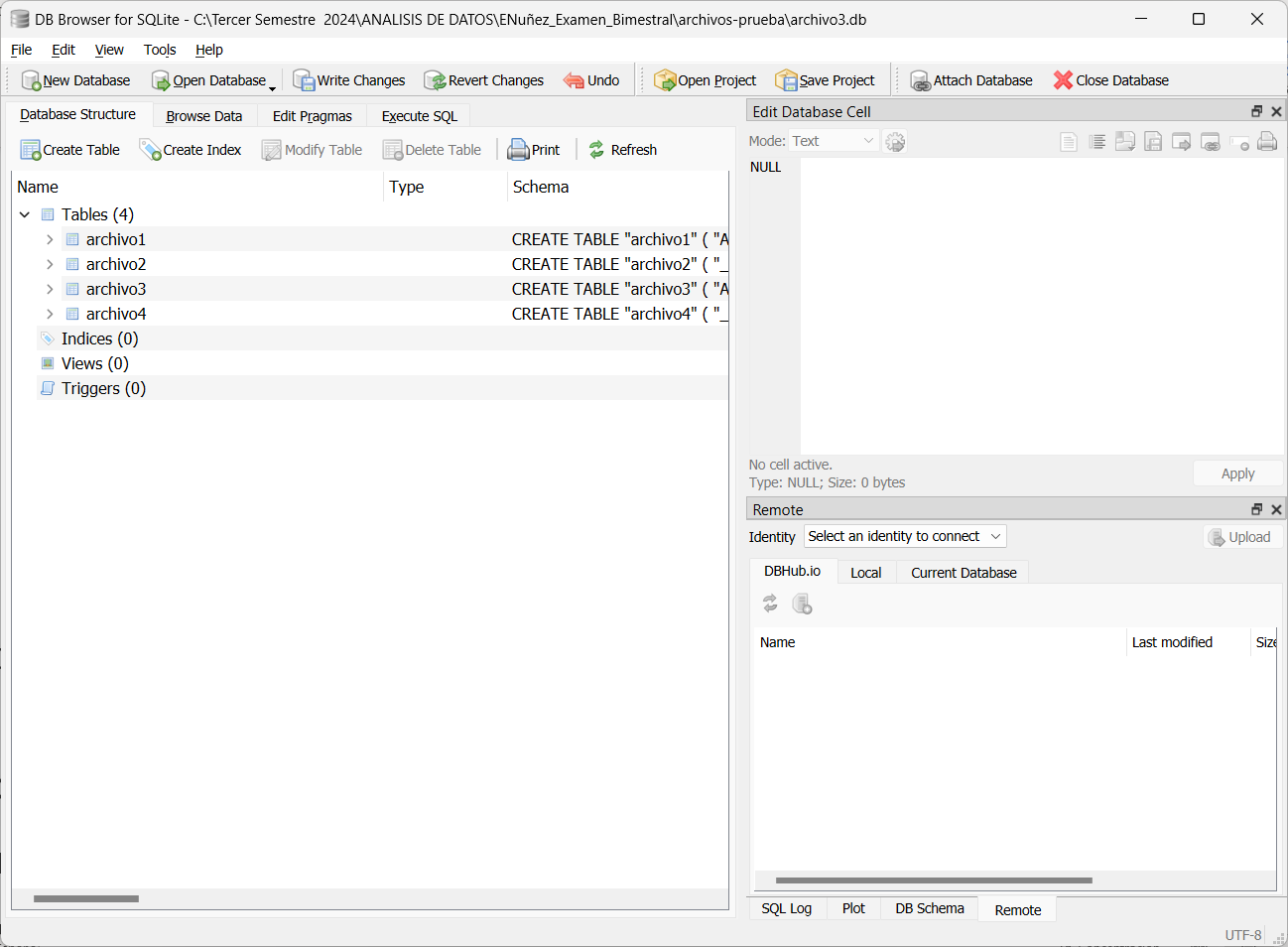
1. El dataframe generado en el paso 2 se lo debe exportar a SQLite



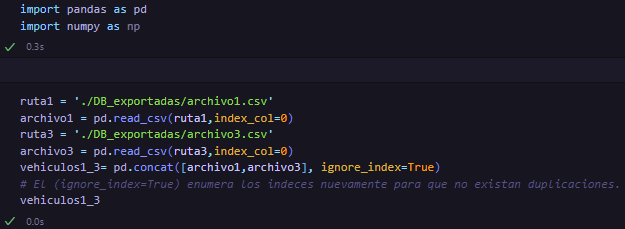


1. El dataframe generado en el paso 4 se lo debe exportar a SQLite



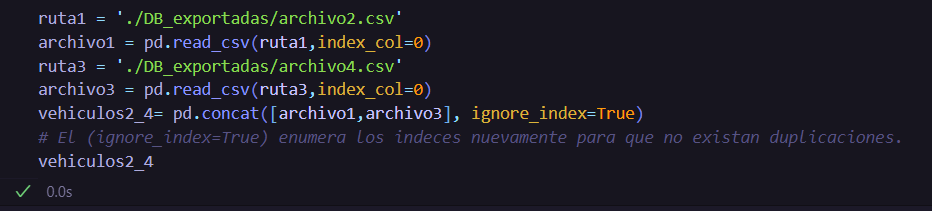


1. Se debe exportar los datos a un dataframe.





1. Se debe exportar los datos a un dataframe.





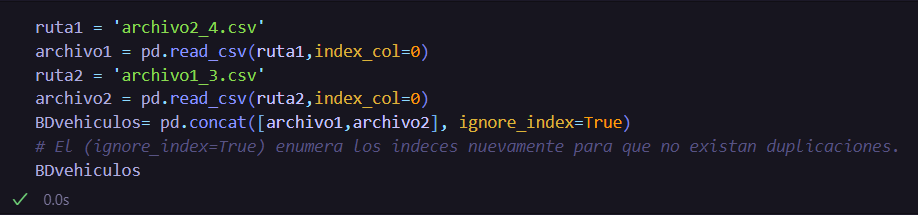


**ANALISIS DE DATOS**

**(10 puntos)**

Base de Datos con las 4 sucursales

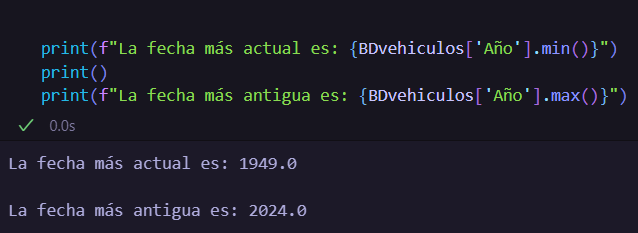
Lista para ser analizada

****

**Con pandas debe contestar a las siguientes preguntas:**

1. **Los autos que tienen los años más recientes y los autos que tienen los autos más antiguos.**

* Primero obtenemos el año más reciente y el año más antiguo.

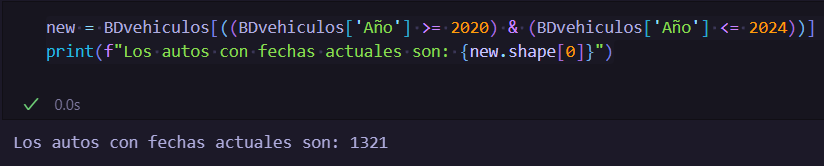


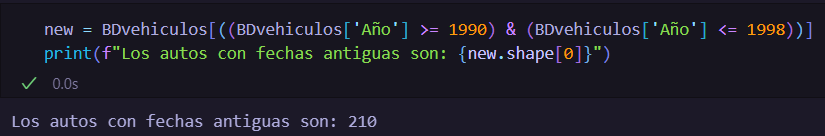
Al realizar esta consulta observamos los siguiente:  
**Año más reciente:**

2024 **Año más antiguo:**

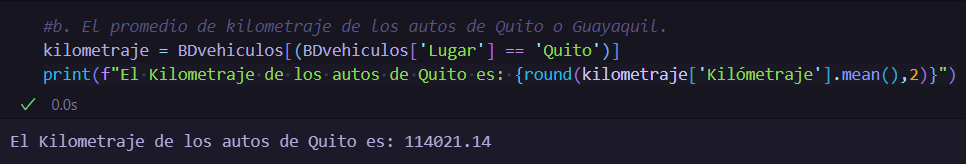
1949

Luego realizamos las consultas dando un rango que consideramos con años recientes y antiguos.

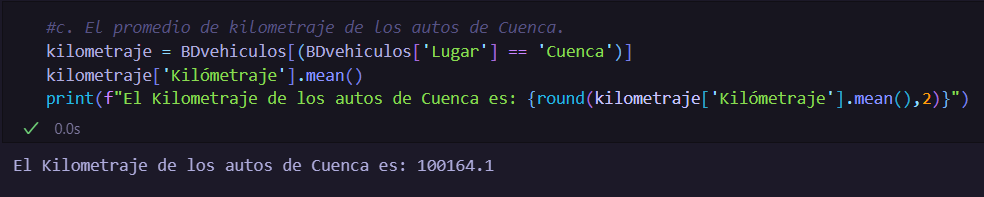




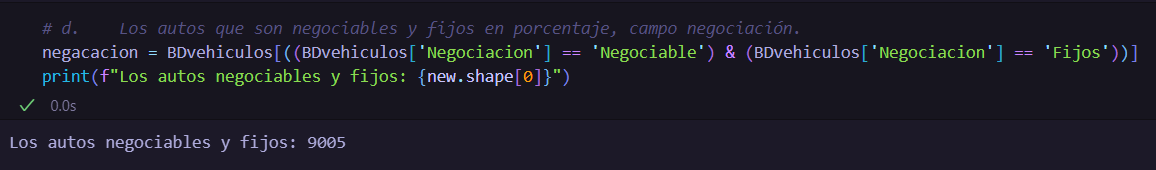
1. **El promedio de kilometraje de los autos de Quito o Guayaquil**



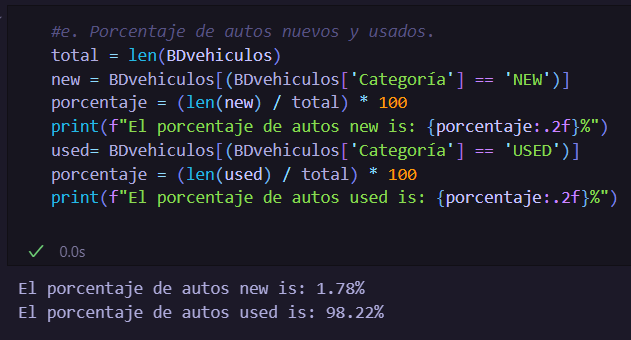
1. **El promedio de kilometraje de los autos de Cuenca.**

****

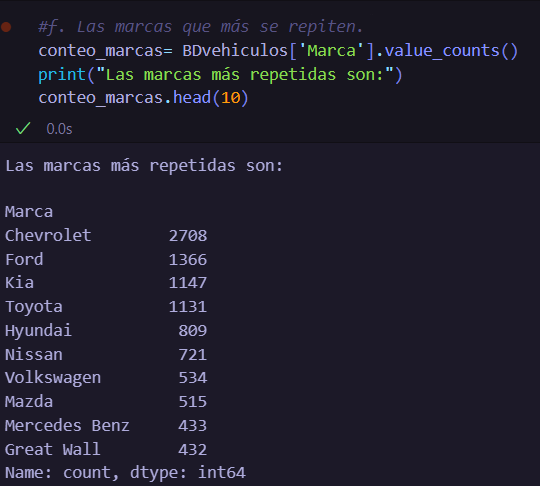
1. **Los autos que son negociables y fijos en porcentaje, campo negociación.**



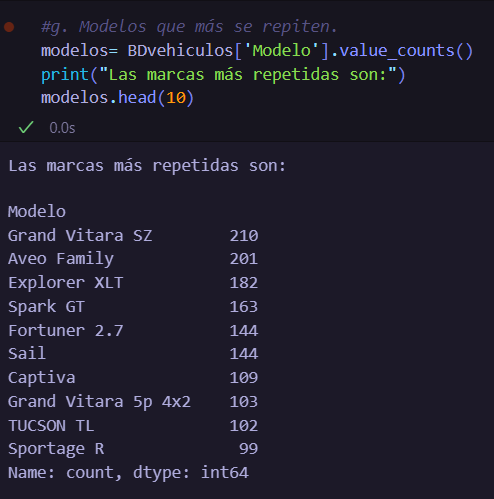
1. **Porcentaje de autos nuevos y usados.**



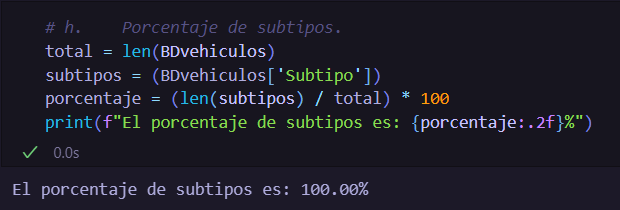
1. **Las marcas que más se repiten.**



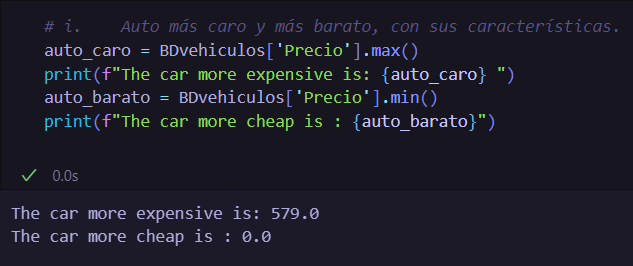
1. **Modelos que más se repiten.**



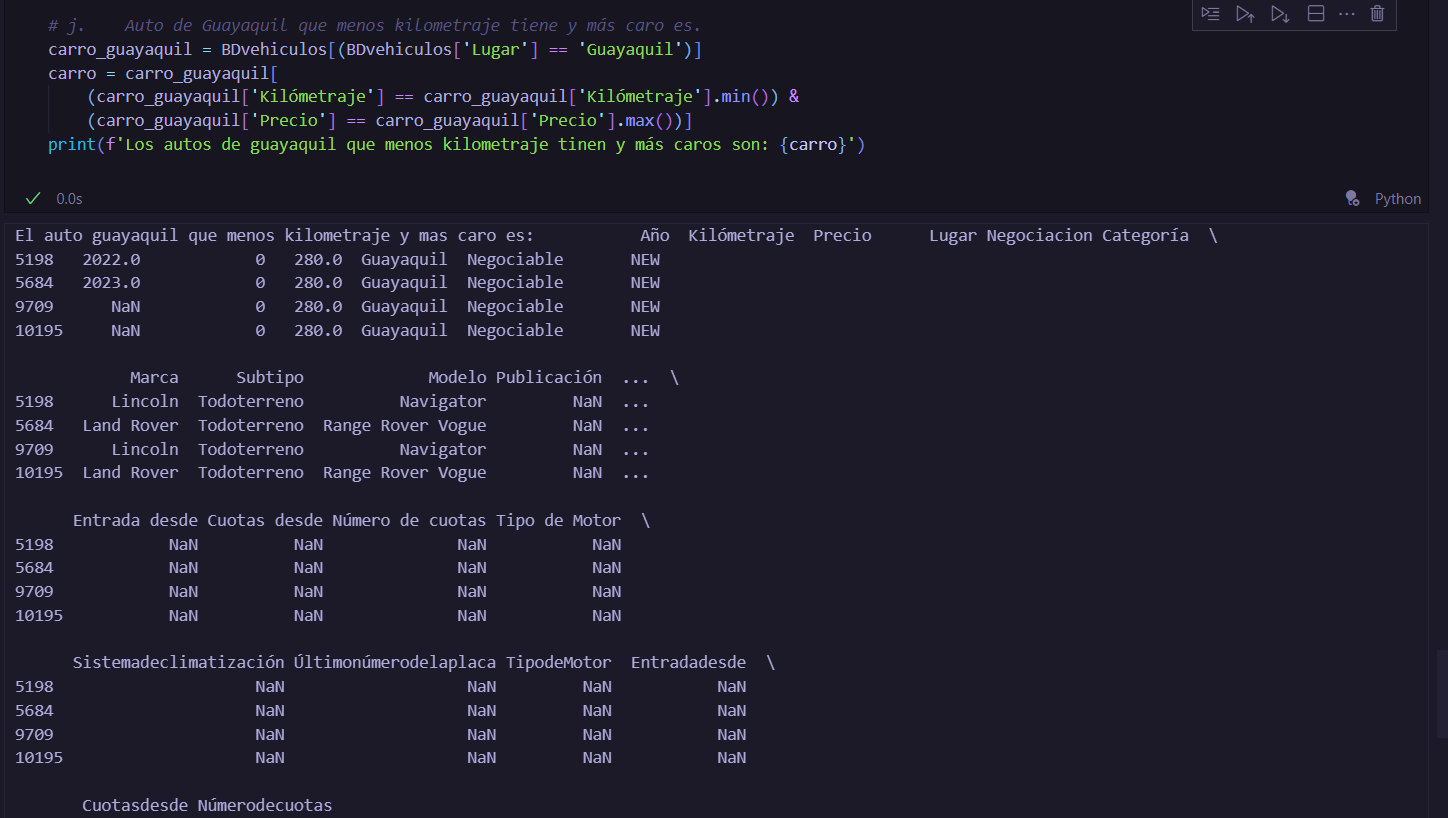
1. **Porcentaje de subtipos.**

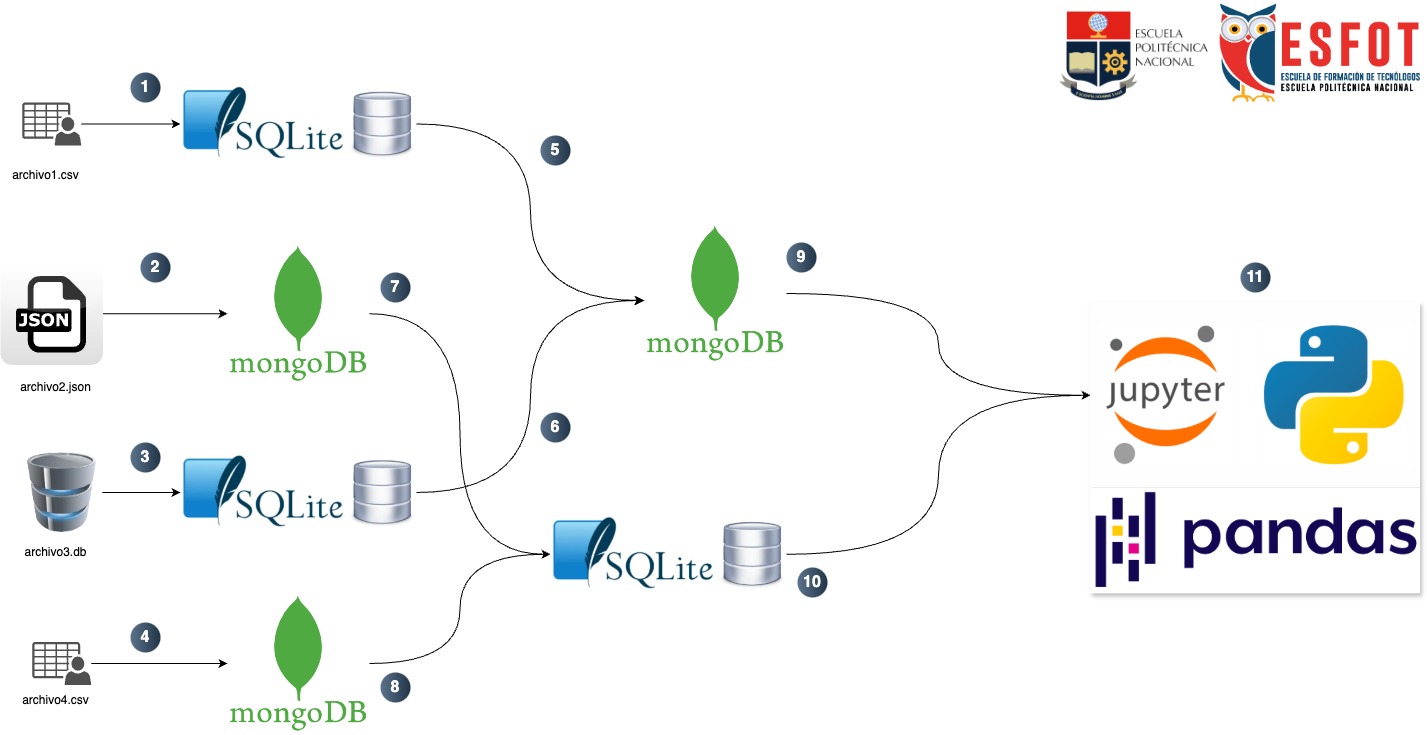


1. **Auto más caro y más barato, con sus características.**



1. **Auto de Guayaquil que menos kilometraje tiene y más caro es.**





**Parte teórica**

**(6 puntos)**

Contestar a las siguientes preguntas:

1. **Defina cada etapa de la pirámide del conocimiento.**

La **pirámide del conocimiento** es un modelo conceptual y relacional que representa cómo los datos básicos se transforman en información, conocimiento y, finalmente, en sabiduría.

**Etapas de la pirámide**

**DATOS**

Son hechos básicos y sin procesar estos no tienen contexto ni significado propio.

**INFORMACION**

Son los datos procesados y organizados, que nos brindan un contexto para entender la información que se guarda en los datos desordenados.

**CONOCIMIENTO**

Es la información analizada y entendida para posteriormente sacar conclusiones que nos ayudarán y darán paso a la sabiduría.

**SABIDURIA**

Una vez que la información se transforme en conocimiento tomamos una decisión sobre qué haremos con el conocimiento obtenido y en que nos puede ayudar.

En esta parte entra el razonamiento y la ética para decidir qué hacer con esta información.

1. **Explique que es un archivo. json**

Un archivo json es un estándar para importar archivos que contengan bases de datos, ya que, se podría decir que es el lenguaje que la mayoría de base de datos entienda porque contiene un formato que ayuda a ubicar las filas y columnas de una base de datos.

1. **Explique las diferencias entre Bases de Datos relacionales y NoSQL.**

Los dos tipos de base de datos existentes hasta el momento tienen grandes diferencias al momento de pensar en crear una base de datos para un proyecto y considero que según la necesidad del proyecto las dos tienen buenas razones para ser elegidas.

**Diferencias:**

1. **Relacionales:** Usan tablas con filas y columnas, como una hoja de cálculo.
2. **NoSQL:** Usan formatos flexibles como documentos, gráficos, claves-valor o columnas.
3. **Relacionales:** Necesitan un diseño físico o fijo (esquema) antes de agregar datos.
4. **NoSQL:** Pueden almacenar datos sin un diseño fijo, lo que las hace más flexibles.
5. **Relacionales:** Funcionan bien con datos organizados y relacionados entre sí.
6. **NoSQL:** Son mejores para datos no estructurados o que cambian mucho, como redes sociales o catálogos.
7. **Relacionales:** Usan SQL, un lenguaje estándar para trabajar con datos.
8. **NoSQL:** Usan diferentes lenguajes según el tipo de base.

**Entregables:**

Archivo ipynb y el pdf con capturas de pantalla indicando el procedimiento.

**Importante:**

Tomar en consideración:

* El uso de dataframes en cada etapa.
* MongoDB y SQLite tienen las opciones de importación y exportación.