

MONGO DB QUERIES

INTRODUCCIÓN

El propósito de este trabajo es seleccionar una colección de datos, integrarla en MongoDB y realizar un análisis exhaustivo mediante la implementación de las consultas necesarias. Para dicho análisis, he optado por utilizar un conjunto de datos que incluye información sobre vehículos, junto con detalles relacionados con sus propietarios, país, el año de fabricación, la marca, el modelo, el color y el método de pago. El archivo de datos utilizado ha sido obtenido de Kaggle.

<https://www.kaggle.com/datasets/iamsouravbanerjee/cars-dataset>

El dataset fue creado por el usuario Sourav Banerjee y cuenta con un total de 30.000 registros y 8 variables. De estas, 7 son categóricas y 1 es numérica (año de fabricación). No presenta valores nulos y tiene un tamaño de 1.8 MB.

Las variables que componen este conjunto de datos son las siguientes:

- First Name: Nombre del propietario
- Last Name: Apellido del propietario
- Country: País donde reside la persona
- Car Brand: Marca de coche
- Car Model: Modelo del coche
- Car Color: Color del coche
- Year of Manufacture: Año de fabricación del coche
- Credit Card Type: Tipo de tarjeta de crédito

1. Convertimos nuestro dataset de csv a json desde Python.

```
import pandas as pd

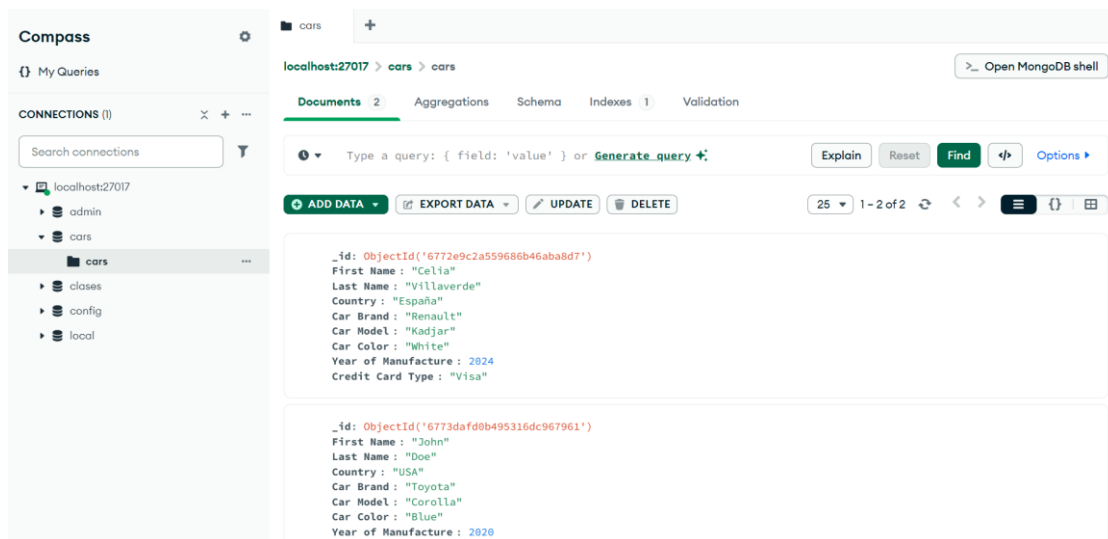
# Cargar el archivo CSV
csv_file = "C:\\Users\\celia\\Desktop\\MASTER DATA SCIENCE\\NO SQL\\Tarea\\cars.csv"
df = pd.read_csv(csv_file)

# Convertir a JSON
json_file = "C:\\Users\\celia\\Desktop\\MASTER DATA SCIENCE\\NO SQL\\Tarea\\cars.json"
df.to_json(json_file, orient="records", lines=True)

print(f"Archivo convertido a JSON y guardado en {json_file}")
```

Archivo convertido a JSON y guardado en C:\\Users\\celia\\Desktop\\MASTER DATA SCIENCE\\NO SQL\\Tarea\\cars.json

2. Cargamos el fichero JSON en MongoDB Compass.



Para cargar el fichero en MongoDB primero lo importamos en MongoDB Compass para poder utilizarlo.

3. Queries

Después de cargar el documento en MongoDB, se procede a analizar su estructura y el tipo de información que contiene.

A continuación, se detallan las consultas necesarias para extraer los datos deseados:

- Proyectamos todos los datos de la colección de cars

```
db.cars0.find({})  
  
.projection({})  
  
.sort({})  
  
.limit(0)
```

_id	First Name	Last Name	Country	Car Brand	Car Model	Car Color	Year of Manu	Credit Card Type
6783b59128069a8379a95794	Yetty	Arghent	Indonesia	Ford	Club Wagon	Teal	1993 (2.0K)	mastercard
6783b59128069a8379a95795	Crystal	Bosworth	China	Cadillac	Escalade ESV	Fuscia	2007 (2.0K)	mastercard

Como podemos observar nuestro archivo cars0 está formado por 30000 documentos, podemos observar los distintos tipos de variables de los que disponemos.

- Saber cuántos registros de automóviles se encuentran en la colección:

```
db.cars0.find().count()
```

30000

Nuestro dataset dispone de 30000 documentos

- Mostramos sólo los campos del nombre y apellidos

```
db.cars0.find({}, { "First Name": 1, "Last Name": 1 });
```

11

db.cars0.find({}, { "First Name": 1, "Last Name": 1 });

cars0

0.026 s











30.000 Docs

50

p. 1

1 - 50

Table

	_id	First Name	Last Name
1	 6783b59128069a8379a95794	Yetty	Arghent
2	 6783b59128069a8379a95795	Crystal	Bosworth
3	 6783b59128069a8379a95796	Monro	Houdhury
4	 6783b59128069a8379a95797	Bowie	Clair
5	 6783b59128069a8379a95798	Myrvyn	McAllister
6	 6783b59128069a8379a95799	Jarid	Son
7	 6783b59128069a8379a9579a	Griffin	Letham
8	 6783b59128069a8379a9579b	Kristin	Hubbock
9	 6783b59128069a8379a9579c	Nannie	Rainsdon
10	 6783b59128069a8379a9579d	Eloisa	Thick

Observamos el nombre y apellido de cada propietario que ha comprado un coche.

- Filtramos por la marca “Toyota” y por modelo específico “Corolla”:

```
db.cars0.find({ "Car Brand": "Toyota", "Car Model": "Corolla" });
```

cars0

0.081 s

82 Docs

50

p. 1

1 - 50

Table

_id	First Name	Last Name	Country	Car Brand	Car Model	Car Color	Year of Manu	Credit Card Type
6783b59128069a8379a95845	Salomon	Zanetto	United States	Toyota	Corolla	Crimson	2007 (2.0K)	jcb
6783b59128069a8379a958b4	Alvinia	Bedenham	China	Toyota	Corolla	Green	2002 (2.0K)	jcb
6783b59128069a8379a959fd	Levey	Cockaday	China	Toyota	Corolla	Red	2008 (2.0K)	jcb
6783b59228069a8379a95bfc	Ethyl	Slora	Sweden	Toyota	Corolla	Mauv	2010 (2.0K)	jcb
6783b59228069a8379a95efd	Michellina	Silwood	Canada	Toyota	Corolla	Fuscia	2006 (2.0K)	jcb
6783b59228069a8379a95f7b	Ernaline	Pakeman	Guatemala	Toyota	Corolla	Indigo	2002 (2.0K)	mastercard
6783b59228069a8379a963a8	Cam	Mathis	Argentina	Toyota	Corolla	Aquamarine	1996 (2.0K)	diners-club-carte

La consola sólo los muestra aquellos coches cuya marca sea Toyota y modelo Corolla

- Filtramos del total de coches, aquellos que hayan sido fabricados después de 1980

```
db.cars0.find({ "Year of Manufacture": { $gt: 1980 } });
```

cars0		0.051 s	29,282 Docs				50		p. 1		1 - 50		Table			
_Id		First Name	Last Name	Country	Car Brand	Car Model	Car Color	Year of Manu	Credit Card Type							
6783b59128069a8379a95794		Yetty	Arghent	Indonesia	Ford	Club Wagon	Teal	1993 (2.0K)	mastercard							
6783b59128069a8379a95795		Crystal	Bosworth	China	Cadillac	Escalade ESV	Fuscia	2007 (2.0K)	mastercard							
6783b59128069a8379a95796		Monro	Houdhury	Indonesia	Mazda	Miata MX-5	Orange	2009 (2.0K)	maestro							
6783b59128069a8379a95797		Bowie	Clair	China	Audi	A4	Orange	2005 (2.0K)	instapayment							

Como podemos observar son 29282 la cantidad de coches que fueron fabricados después de 1980.

- Insertamos más información

```
db.cars0.insertOne({
    "First Name": "Celia",
    "Last Name": "Villaverde",
    "Country": "España",
    "Car Brand": "Renault",
    "Car Model": "Kadjar",
    "Car Color": "White",
    "Year of Manufacture": 2020,
    "Credit Card Type": "Visa"
});
```

Key	Value	Type
(1)	{ acknowledged : true, insertedId : ObjectId("6783fce3c294732329952363") }	Object
acknowledged	true	Bool
insertedId	6783fce3c294732329952363	ObjectId

Se ha añadido la información de manera correcta.

- Actualizamos el color del coche a negro “Black” para el registro de la persona cuyo nombre es "Celia"

```
db.cars0.updateOne(
  { "First Name": "Celia" },
  { $set: { "Car Color": "Black" } }
);
```

Key	Value	Type
acknowledged	true	Bool
matchedCount	1	Int32
modifiedCount	1	Int32

- Saber cuáles son los 10 nombres más frecuentes de los propietarios, ordenados de mayor a menor según su cantidad de ocurrencias.


```
db.cars0.aggregate([
  { $group: { _id: "$First Name", Count: { $sum: 1 } } },
  { $sort: { Count: -1 } },
  { $limit: 10 }
]);
```

cars0 0.153 s 10 Docs		
	_id	Count
1	Tally	15
2	Teddie	15
3	Sonny	15
4	Dell	14
5	Wally	14
6	Ted	13
7	Elisha	13
8	Micky	13
9	Perry	12
10	Lorne	12

Estos son los 10 nombres más comunes entre nuestros propietarios, además observamos la cantidad de veces que se repiten.

- Saber cuáles son los 3 apellidos más frecuentes entre los propietarios, ordenados de mayor a menor según su cantidad de ocurrencias.





```
db.cars0.aggregate([
  { $group: { _id: "$Last Name", Count: { $sum: 1 } } },
  { $sort: { Count: -1 } },
  { $limit: 3 }
]);
```

cars0	 0.151 s	3 Docs
_id *		Count
Teggin		6
Stenners		6
Carillo		6

Teggin, Stenners y Carillo, destacan como los 3 apellidos más comunes.

- Conocer las 5 marcas de coches más frecuentes en la colección, ordenadas de mayor a menor según su cantidad de registros.

```
db.cars0.aggregate([
  { $group: { _id: "$Car Brand", Total: { $sum: 1 } } },
  { $sort: { Total: -1 } },
  { $limit: 5 }
]);
```

	cars0	 0.082 s	5 Docs
	<u>_id</u> * 	Total 	
1	Ford	2562 (2.6K)	
2	Chevrolet	2487 (2.5K)	
3	Toyota	1445 (1.4K)	
4	Dodge	1438 (1.4K)	
5	GMC	1296 (1.3K)	

Como se observa en la consola, esas son las 5 marcas mas comunes de nuestra colección.

- Se agrupan los coches por marca y se calcula la cantidad de vehículos por cada marca, ordenándolos de mayor a menor según su frecuencia.

```
db.cars0.aggregate([
  { $group: { _id: "$Car Brand", Total: { $sum: 1 } } },
  { $sort: { Total: -1 } }
]);
```

cars0 0.070 s 78 Docs		
	_id *	Total
1	Ford	2562 (2.6K)
2	Chevrolet	2487 (2.5K)
3	Toyota	1445 (1.4K)
4	Dodge	1438 (1.4K)
5	GMC	1296 (1.3K)
6	Mazda	1165 (1.2K)
7	Mercedes-Benz	1136 (1.1K)
8	Mitsubishi	1119 (1.1K)
9	Pontiac	1095 (1.1K)
10	Volkswagen	983

Se observa la cantidad de coches en total que hay de cada marca de coche, destaca Ford, con más de 2600 coches, seguido de Chevrolet y Toyota.

- Se obtiene el coche más comprado por cada año de fabricación, junto con la cantidad de vehículos de esa marca adquiridos, ordenados por año y frecuencia de compra.

```
db.cars0.aggregate([
  {
    $group: {
      _id: { Year: "$Year of Manufacture", Car: "$Car Brand" },
      Count: { $sum: 1 }
    }
  },
  {
    $sort: { "_id.Year": 1, Count: -1 }
  }
]);
```

```

    },
    {
      $group: {
        _id: "$_id.Year",
        MostPopularCar: { $first: "$_id.Car" },
        Count: { $first: "$Count" }
      }
    },
    { $sort: { _id: 1 } }
  ]);

```

	_id	MostPopularCar	Count
1	1909	Ford	1
2	1926	Chrysler	4
3	1948	Citroën	5
4	1950	Hillman	1
5	1953	Chevrolet	6
6	1954	Cadillac	3
7	1955	Chevrolet	6
8	1956	Chevrolet	4





Durante los primeros años registrados en nuestro dataset, podemos comprobar que los coches más comprados durante los primeros años fueron Ford, Chrysler, después Citroën entre otros.

- Se calcula el número de coches diferentes asociados a cada modelo de coche, ordenados de mayor a menor según la cantidad de colores únicos

```

db.cars0.aggregate([
  { $group: { _id: "$Car Model", Colors: { $addToSet: "$Car Color" } } },
  { $project: { NumColors: { $size: "$Colors" } } },
  { $sort: { NumColors: -1 } }
]);

```


	cars0	 0.117 s	1007 Docs
	<u>_id</u> * 		NumColors 
1	Stratus		20
2	MX-6		19
3	900		19
4	GS		19
5	Blazer		19
6	Ranger		19
7	RAV4		19
8	S-Series		19

En cuanto a la diferenciación de colores por modelo, observamos que Stratus es el modelo con mayor rango de colores, sin embargo, el resto de modelos presentan la misma cantidad de diferentes tipos de colores, que son 19.

- Se calcula el número de coche por país y color, ordenando los resultados de mayor a menor según la cantidad de coches

```

db.cars0.aggregate([
  { $unwind: "$Car Color" },
  {
    $group: {
      _id: { country: "$Country", color: "$Car Color" },
      count: { $sum: 1 }
    }
  },
  {
    $project: {
      _id: 0,
      Country: "$_id.country",
      CarColor: "$_id.color",
      Count: "$count"
    }
  },
  { $sort: { Count: -1 } }
])

```

});

cars0		0.154 s	2709 Docs			50			p. 1	1 - 50
	Country	CarColor	Count							
1	China	Turquoise	322							
2	China	Aquamarine	316							
3	China	Goldenrod	315							
4	China	Fuscia	302							
5	China	Indigo	301							
6	China	Red	298							
7	China	Maroon	296							
8	China	Crimson	291							
	Country	CarColor	Count							
43	Russia	Turquoise	97							
44	Russia	Teal	97							
45	Philippines	Green	96							
46	Russia	Green	95							
47	Philippines	Maroon	95							
48	Philippines	Mauv	93							
49	Russia	Red	93							
50	Philippines	Aquamarine	93							

En la imagen observamos que China cuenta con 322 coches turquesa, 316 son aguamarina, y 302 fucsia. Por otro lado Rusia cuenta con tan solo 97 coches de color Turquesa, 95 verdes y 93 rojos.







- Se actualizan los registros de la base de datos para cambiar el color de los coches de la marca Toyota azul



```
db.cars0.updateMany(  
  { "Car Brand": "Toyota" },  
  { $set: { "Car Color": "Blue" } }  
);
```

Key	Value	Type
1 (1)	{ acknowledged : true, matchedCount : 1445, modifiedCount : 1370 }	Object
acknowledged	true	Bool
matchedCount	1445 (1.4K)	Int32
modifiedCount	1370 (1.4K)	Int32

- Se calcula la cantidad de modelos de coche únicos asociados a cada marca, ordenados de mayor a menor según el número de modelos.

```
db.cars0.aggregate([
  { $group: { _id: "$Car Brand", models: { $addToSet: "$Car Model" } } },
  { $project: { _id: 0, CarBrand: "$_id", NumModels: { $size: "$models" } } },
  { $sort: { NumModels: -1 } }
]);
```









cars0		0.093 s	78 Docs	  		50	
	CarBrand 						NumModels 
1	Chevrolet						75
2	Ford						69
3	Dodge						52
4	GMC						48
5	Mercedes-Benz						44
6	Toyota						39
7	Audi						37
8	BMW						34

	CarBrand 						NumModels 
43	Scion						6
44	Saab						6
45	Lotus						6
46	Spyker						5
47	Ram						5
48	Maybach						4
49	Eagle						4
50	Geo						4

Como podemos comprobar Chevrolet es la marca de coches que representa la mayor cantidad de modelos de coches, en este caso 75 diferentes modelos, seguido de Ford que cuenta con 69 modelos, Dodge con 52. Por el contrario, tanto Geo, Eagle y Maybach cuentan con tan solo 4 diferentes modelos.

- Se proyecta por tipo de país y tarjeta de crédito.











```
db.cars0.find({}, {"Country": 1, "Credit Card Type": 1, "_id": 0 });
```

cars0		0.023 s	30.001 Docs	  		 	50 
Country 						Credit Card Type 	
1	Indonesia					mastercard	
2	China					mastercard	
3	Indonesia					maestro	
4	China					instapayment	
5	Czech Republic					maestro	
6	Philippines					americanexpress	
7	Canada					bankcard	
8	Ethiopia					china-unionpay	

Mastercard es la tarjeta más utilizada tanto en Indonesia como en China, AmericanExpress es la más común en Filipinas y bankcard en Canadá. Es decir, que podemos observar que en cada país destacan por un amplio rango de tarjetas de crédito muy diferentes entre sí.

- Se calcula la cantidad de coches asociados a tipo de tarjeta de crédito y se ordenan de forma ascendente según el total.

```
db.cars0.aggregate([
  { $group: { _id: "$Credit Card Type", Total: { $sum: 1 } } },
  { $sort: { Total: 1 } }
]);
```

 cars0	 0.078 s	17 Docs	 cars0	 0.078 s	17 Docs
 _id * 		Total 	 _id * 		Total 
1	Visa	1	10	diners-club-enroute	1241 (1.2K)
2	diners-club-international	219	11	bankcard	1245 (1.2K)
3	solo	381	12	visa-electron	1357 (1.4K)
4	diners-club-us-ca	419	13	americanexpress	1394 (1.4K)
5	instapayment	617	14	switch	1651 (1.7K)
6	visa	827	15	maestro	2026 (2.0K)
7	laser	847	16	mastercard	2619 (2.6K)
8	china-unionpay	1218 (1.2K)	17	jcb	12.718 (12.7K)

La tarjeta menos usada a lo largo del dataset es Visa, puesto que ha sido la que yo he incluido, exceptuando esa información las menos usadas son diner-club internacional, solo y diners-club-us-ca. Por otro lado, las tarjetas de crédito que más destacan son jcb, mastercard, maestro, switch y american express.

4. Conclusiones

Gracias a MongoDB hemos logrado desarrollar el código necesario para analizar gradualmente, a través de nuestras queries, toda la información requerida. Se trata de una herramienta poderosa para la gestión de datos, permitiendo almacenar información de manera eficiente. El lenguaje es sencillo y facilita la creación de consultas de manera clara y optimizada para diversas necesidades.

Adicionalmente, con la ayuda de MongoDB Compass, hemos podido cargar el archivo en formato JSON, lo que nos ha permitido operar y trabajar directamente con los datos.

En el desarrollo de este trabajo se han implementado ejercicios de inserción, actualización, proyección, filtrado, y agregación de datos.

CONCLUSIONES DEL ANÁLISIS DEL DATASETS DE COCHES

A partir del análisis de nuestro dataset, compuesto por 30000 registros, se han obtenido las siguientes conclusiones relevantes:

1. Clientes y Apellidos más frecuentes:

- Los nombres de propietarios más repetidos entre los compradores de coches son Tally, Teddie y sonny, con 15 repeticiones cada uno. En cuanto a los apellidos, Teggin destaca como el más común.

2. Marcas y Ventas Totales:

- Las marcas más frecuentes en el dataset son Ford, Chevrolet, Toyota, Dodge y GMC. Esto se explica por la amplia disponibilidad de estas marcas en el mercado, su reputación y su enfoque como vehículos familiares y utilitarios.
- Ford lidera en un número de ventas con 2.600 unidades debido a su sólida presencia global. Por el contrario, marcas como Daewoo (43 coches) y Rolls-Royce (46 coches) presentan cifras mucho menores, probablemente porque atienden nichos de mercado más específicos. Daewoo, como marca de bajo coste podría no ser competitiva, mientras que Rolls-Royce se dirige a un público exclusivo.

3. Modelos y Variedad de coches

- El modelo de coche Status ofrece la mayor gama de colores (20 tonalidades), lo que podría ser una estrategia de marketing para

atraer a diferentes gustos y estilos de clientes. La mayoría de los demás modelos tienen una gama de 19 colores.

4. Distribución de colores por país

- Las presencias de color varían significativamente por región, puede ser por factores culturales o disponibilidad de cada mercado:
- En China, el turquesa lidera con 322 coches, en Indonesia el color Teal es el predominante (192 coches), el violeta es el color más destacado en Rusia (101 coches) y en Filipinas es el naranja el más usado (100 coches).

5. Cantidad de modelos por marca

- Chevrolet lidera en diversidad de modelos con 75 variantes, lo que refleja su estrategia de abarcar múltiples segmentos del mercado
- Ford, con 69 modelos, se posiciona como un competidor cercano.
- Marcas como Eagle, Maybach y Geo cuentan con solo 4 modelos cada uno.

6. Tarjetas de Pago Usadas

- Visa es la tarjeta menos utilizada, cuenta con 1 solo uso, esto es porque es la incorporación nueva que yo he añadido. Diners Club es la siguiente menos utilizada con 219 compras, refleja su uso limitado, probablemente por ser menos común en mercados masivos.
- Por el contrario, JCB lidera con 12700 transacciones, debido a su fuerte presencia en Asia, mientras que Mastercard (2600 transacciones) destaca por su amplia aceptación global.

El análisis de este dataset refleja cómo las preferencias de los consumidores, las estrategias de mercado de las marcas y tendencias culturales tienen un impacto significativo en las ventas, la diversidad de modelos y la elección de colores. Gracias a MongoDB, hemos podido comprobar toda esta información mediante la aplicación de los ejercicios mencionados previamente, validando la utilidad de esta herramienta para el análisis de datos complejos.