28 de enero de 2022

isai GUTIÉRREZ GUTIÉRREZ

Iowa Liquor Sales

INTRODUCCIÓN A BASES DE DATOS

Índice

[1. **Definición del Proyecto** 2](#_Toc94147737)

[**1.1.** **Fuentes de BD** 3](#_Toc94147738)

[**1.2.** **Problema a resolver** 3](#_Toc94147739)

[**2.** **Definición de la Base de Datos** 3](#_Toc94147740)

[**2.1. Tablas** 3](#_Toc94147741)

[**3.** **Glosario de campos** 4](#_Toc94147742)

[**4.** **Creación de Base de datos y tablas** 4](#_Toc94147743)

[**4.1.** **MySQL** 5](#_Toc94147744)

[**4.1.1.** **Category** 5](#_Toc94147745)

[**4.1.2.** **County** 5](#_Toc94147746)

[**4.1.3.** **Items** 5](#_Toc94147747)

[**4.1.4.** **Store** 5](#_Toc94147748)

[**4.1.5** **Vendors** 5](#_Toc94147749)

[**4.1.6** **Sales** 5](#_Toc94147750)

[**4.2.** **MongoDB** 5](#_Toc94147751)

[**4.2.1.** **Category** 5](#_Toc94147752)

[**4.2.2.** **County** 5](#_Toc94147753)

[**4.2.3.** **Items** 5](#_Toc94147754)

[**4.2.4.** **Store** 5](#_Toc94147755)

[**4.2.5.** **Vendors** 5](#_Toc94147756)

[**4.2.6.** **Sales** 5](#_Toc94147757)

[**5.** **Consultas MySQL** 5](#_Toc94147758)

[5.1. ¿Qué productos generan mayor volumen de ventas? 5](#_Toc94147759)

[5.2. ¿Cuáles son los 5 productos con menor volumen de venta? 5](#_Toc94147760)

[5.3. ¿Cuáles son las 10 ciudades donde se consume mayor volumen de alcohol? 6](#_Toc94147761)

[5.4. ¿En qué ciudad se consume el mayor volumen de ventas? 6](#_Toc94147762)

[5.5. ¿En qué ciudad no conviene tener un punto de venta de acuerdo al monto de venta? 7](#_Toc94147763)

[5.6. ¿Cuáles las categorías que más venden? 7](#_Toc94147764)

[5.7. ¿Cuál es la categoría que más vende por ciudad? 8](#_Toc94147765)

[5.8. ¿En qué fecha se incrementa el volumen de ventas? 8](#_Toc94147766)

[5.9. ¿Qué producto se vende más por cantidad? 9](#_Toc94147767)

[5.10. ¿Qué producto se vende menos por cantidad? 9](#_Toc94147768)

[5.11. ¿En qué sucursal se muestra mayor volumen de ventas? 10](#_Toc94147769)

[5.12. ¿En qué mes hay mayor menor consumo de alcohol? 10](#_Toc94147770)

[5.13. Cuantos paquetes se consumen por tiendas en cada mes 11](#_Toc94147771)

[5.14. Cuantos paquetes de los 3 productos más vendidos se consumen por condado en un mes 12](#_Toc94147772)

[5.15. Determina la utilidad de cada producto 13](#_Toc94147773)

[5.16. Determina los 10 productos que generan mayor volumen ganancias de utilidad y de esos 10 obtén la ganancia total de utilidad por ciudad 14](#_Toc94147774)

[5.17. Obtener el promedio del precio de todos los productos 15](#_Toc94147775)

[5.18. ¿Cuántos productos se venden por arriba de la media? 15](#_Toc94147776)

[5.19. ¿Cuántos productos se venden por abajo de la media? 15](#_Toc94147777)

[**6.** **Consultas MongoDB** 15](#_Toc94147778)

[6.1. ¿Qué productos generan mayor volumen de ventas? Muestra los últimos 5 15](#_Toc94147779)

[6.2. ¿Cuáles son los 5 productos con menor volumen de venta? 16](#_Toc94147780)

[6.3. ¿En qué tiendas se consume el mayor volumen de ventas? 18](#_Toc94147781)

[6.4. ¿Cuál es la categoría que mas vende ? 19](#_Toc94147782)

[6.5. ¿Cuál es el producto que se vende mas? 20](#_Toc94147783)

[6.6. ¿En qué ciudad se consume el mayor volumen de ventas? 21](#_Toc94147784)

[6.7. ¿Cuál es el producto que se vende menos? 22](#_Toc94147785)

[6.8. Determina la utilidad de cada producto 23](#_Toc94147786)

[6.9. ¿En qué año se incrementa el volumen de ventas? 24](#_Toc94147787)

[6.10. ¿Cuáles son las 10 tiendas donde se consume mayor volumen de alcohol? 25](#_Toc94147788)

[6.11. ¿En qué sucursal se muestra mayor volumen de ventas? 26](#_Toc94147789)

[6.12. ¿En qué mes hay menor consumo de alcohol? 27](#_Toc94147790)

[6.13. Determina los 10 productos que generan mayor volumen ganancias de utilidad 28](#_Toc94147791)

[6.14. ¿En qué mes hay mayor consumo de alcohol? 30](#_Toc94147792)

[6.15. Obtener el promedio del precio de todos los productos 32](#_Toc94147793)

# 1. **Definición del Proyecto**

## **Fuentes de BD**

La fuente de datos que se utilizaron fue de la página de Kaggle (<https://www.kaggle.com/residentmario/iowa-liquor-sales>)

## **Problema a resolver**

Determinar en que parte del estado de iowa es mejor tener un punto de venta, a partir de los datos obtenidos.

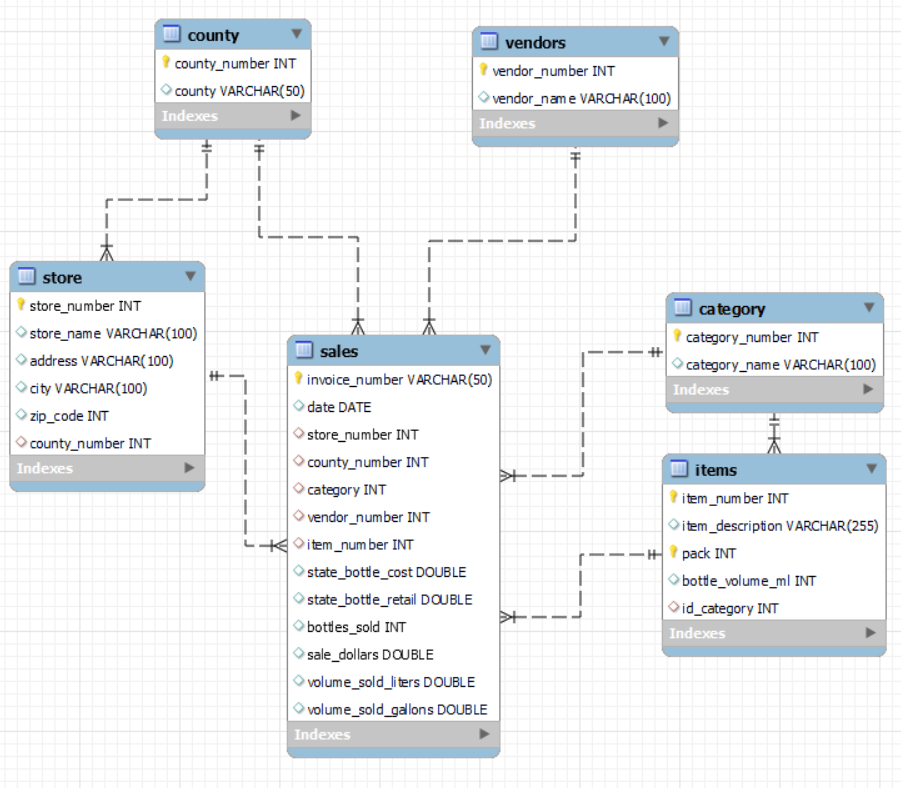
# **Definición de la Base de Datos**

## **2.1. Tablas**

Los datos están divididos por varias tablas las cuales son:

1. **Store**: Tiendas de los licores
2. **Vendors**: Vendedores de las tiendas
3. **County**: El condado donde esta cada tienda
4. **Category**: Categoría de cada producto.
5. **Items**: Productos que se venden en cada tienda
6. **Sales**: Registros de ventas de los distintos tipos de alcohol, volumen (lt, galones) en las diferentes tiendas con su precio de venta y el monto total.

**2.2. Diagrama Entidad Relación**



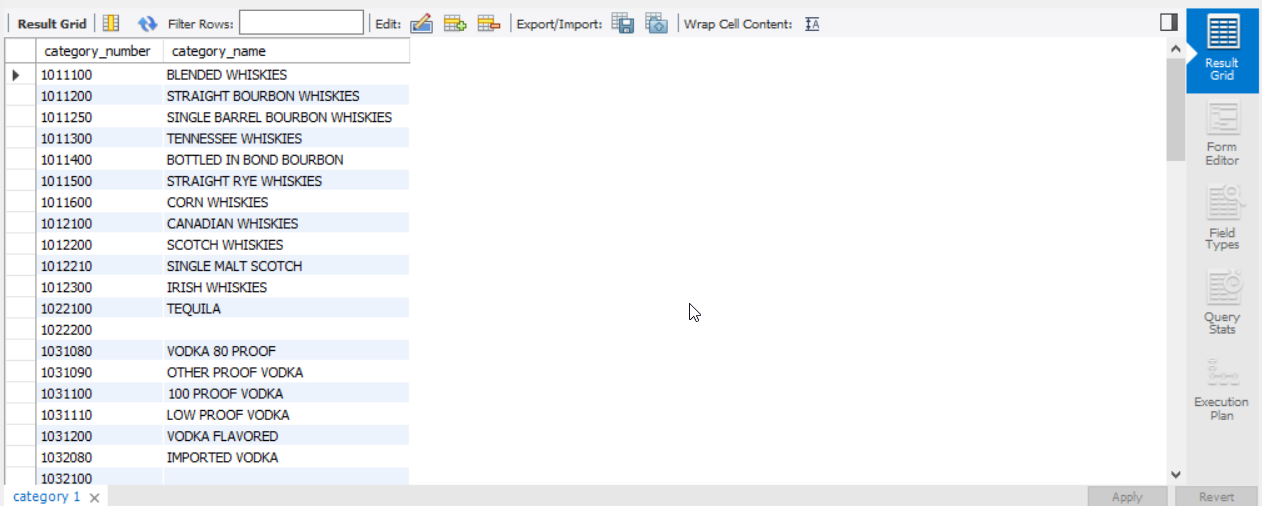
# **Glosario de campos**

* **state\_bottle\_cost:** Costo del producto
* **state\_bottle\_retail:** Precio de venta del producto
* **bottles\_sold:** Botellas vendidas
* **sale\_dollars:** Monto total de venta
* **volume\_sold\_liters:** Volumen de litros vendidos
* **volume\_sold\_gallons:** Volumen de galones vendidos

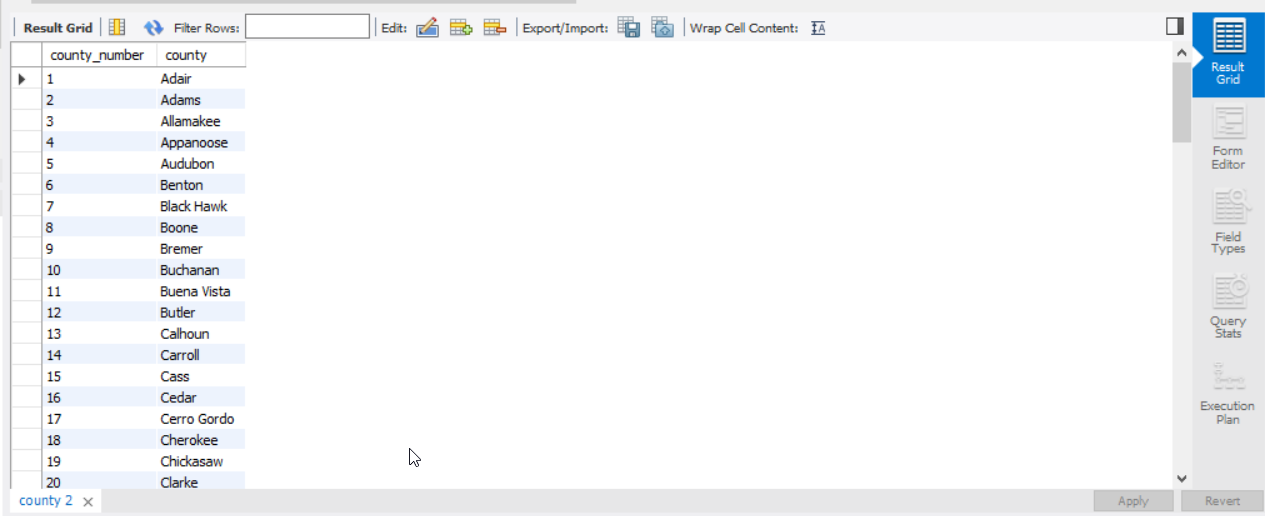
# **Creación de Base de datos y tablas**

## **MySQL**

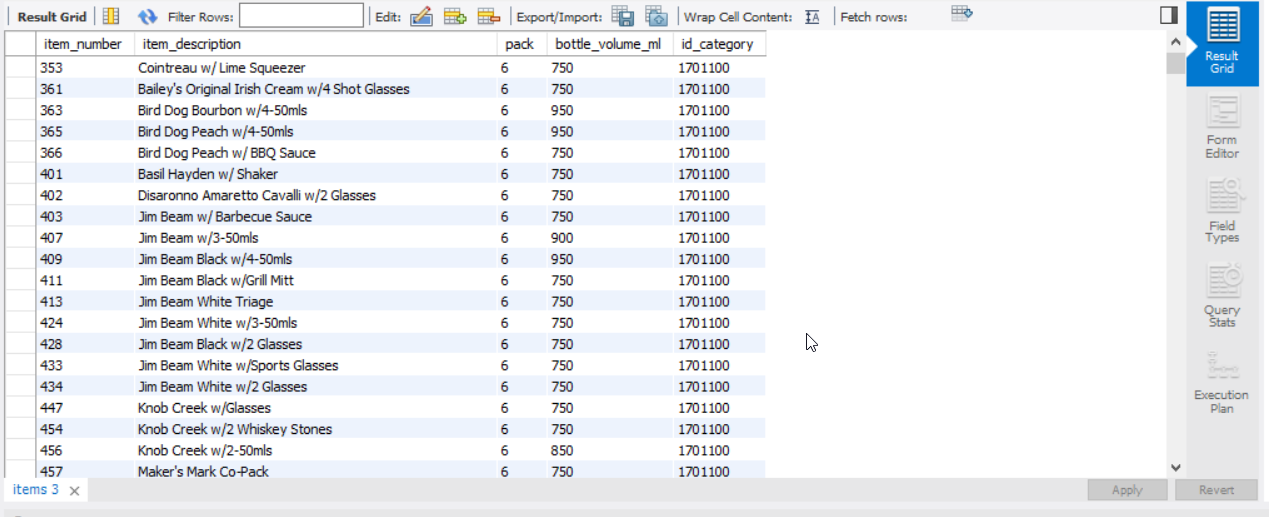
### **Category**

****

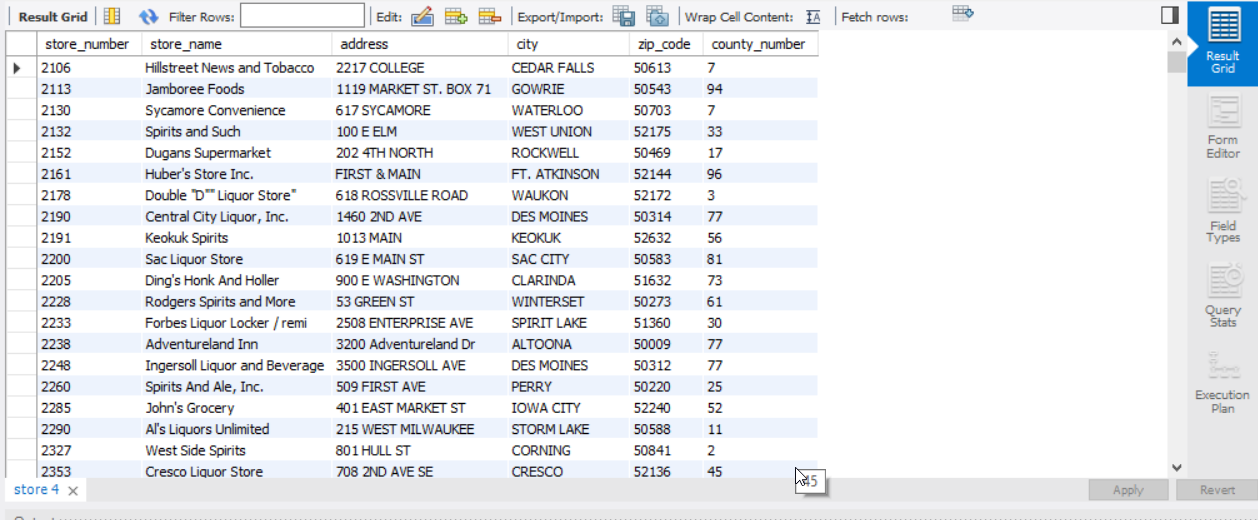
### **County**

****

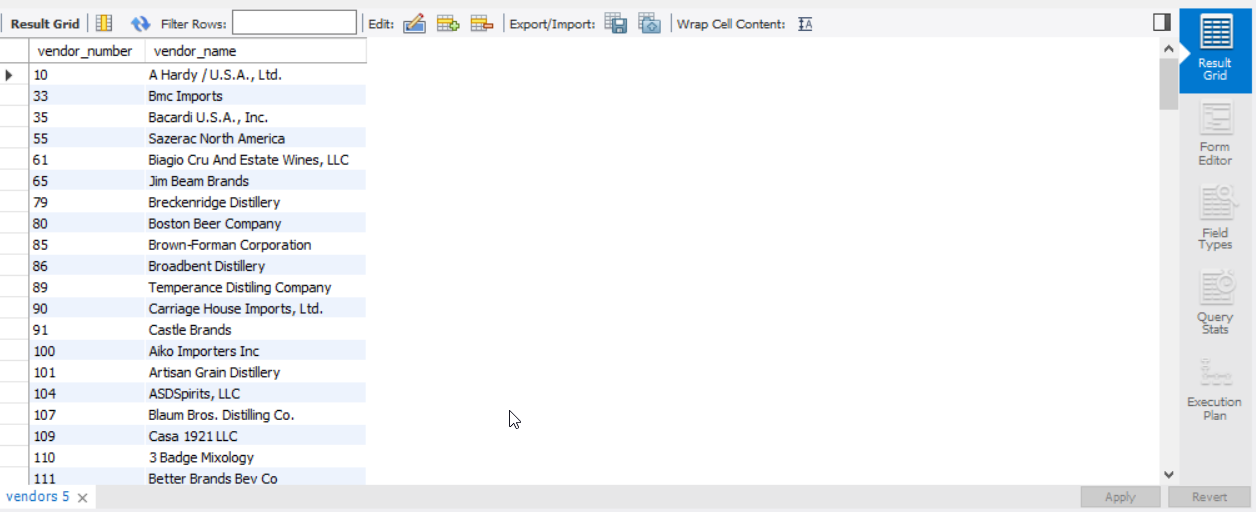
### **Items**

****

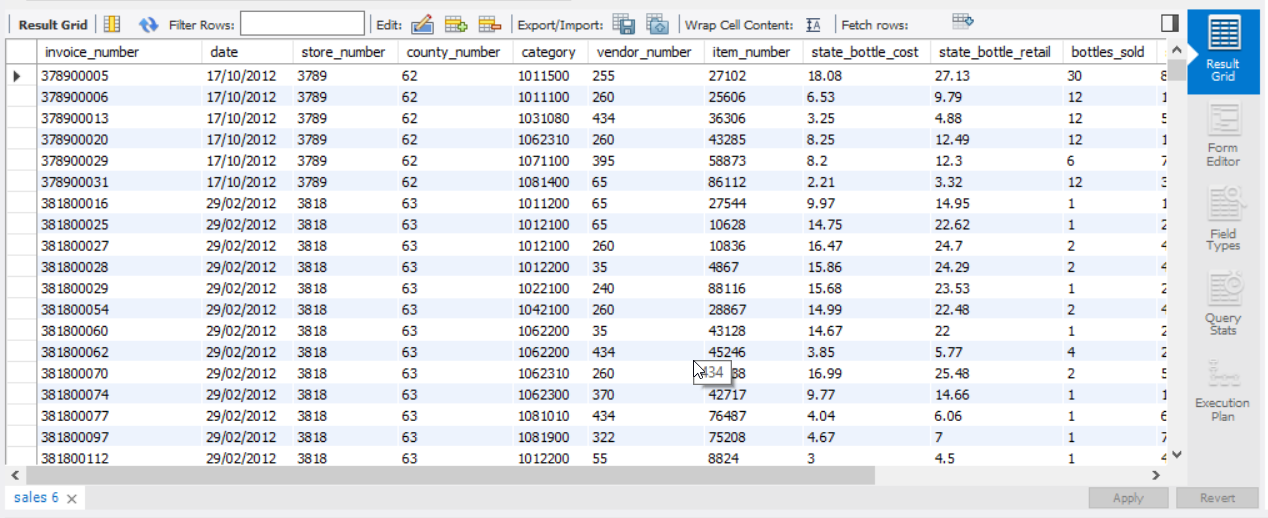
### **Store**

****

### **Vendors**

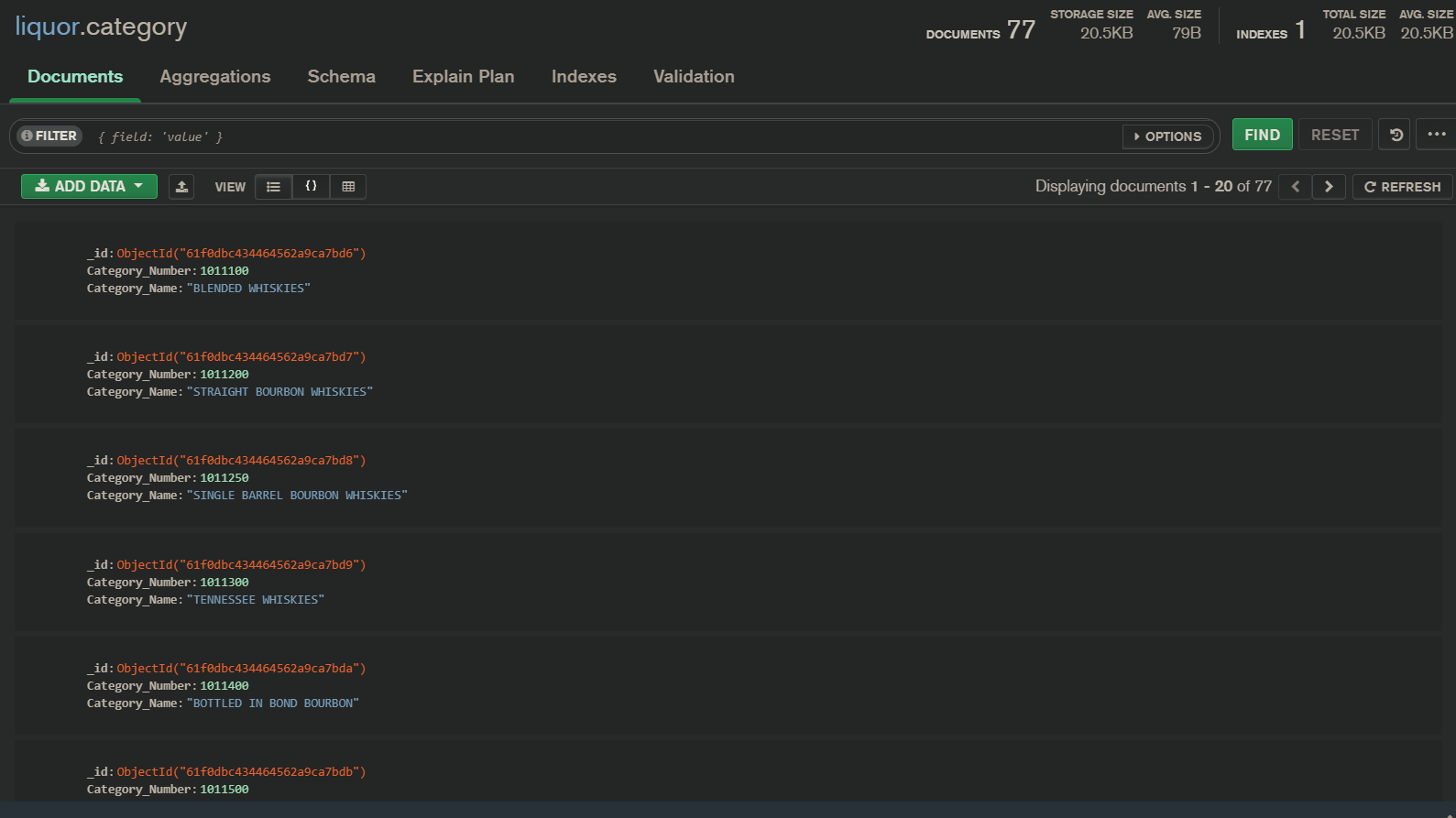
****

### **Sales**

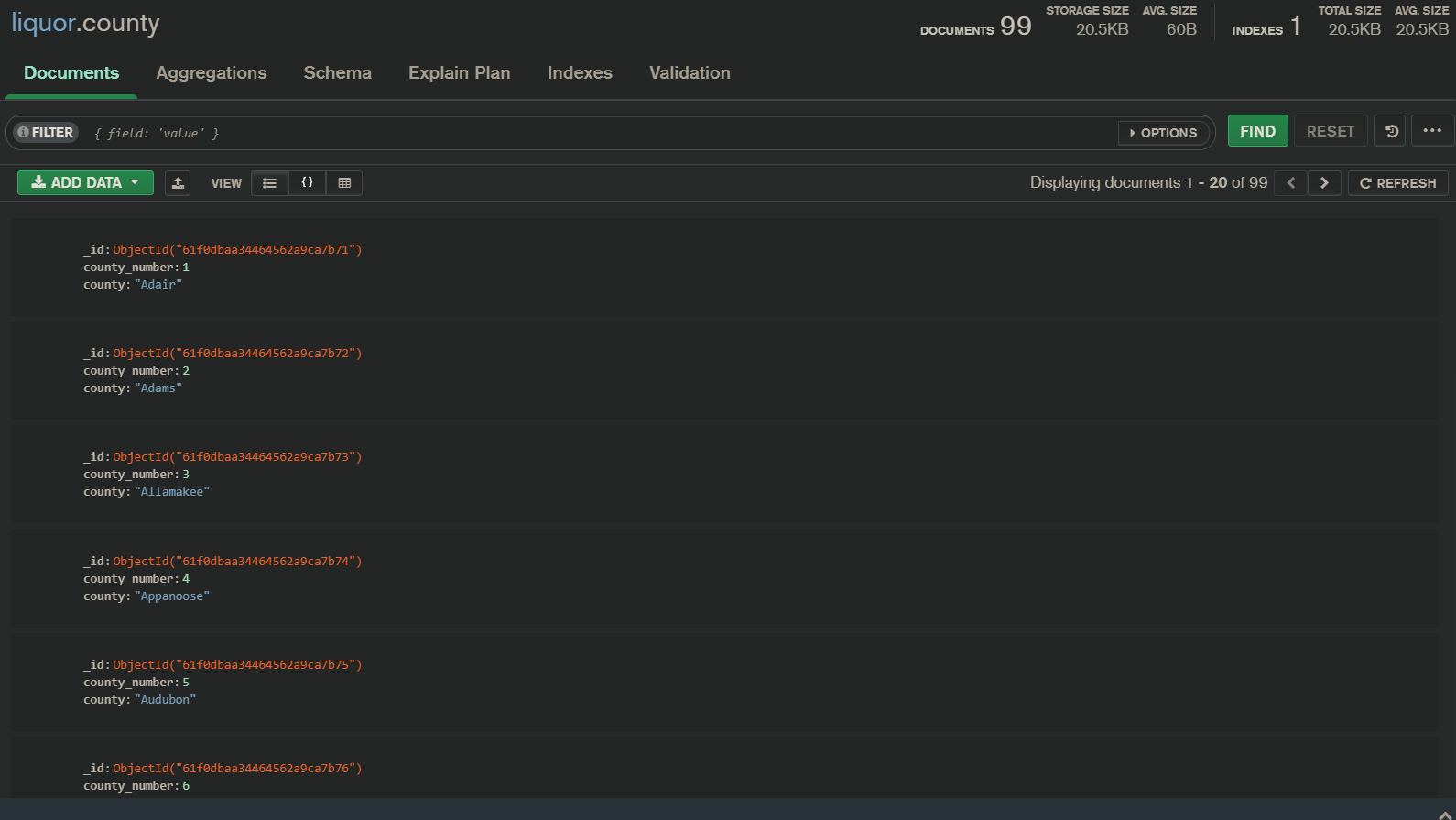


## **MongoDB**

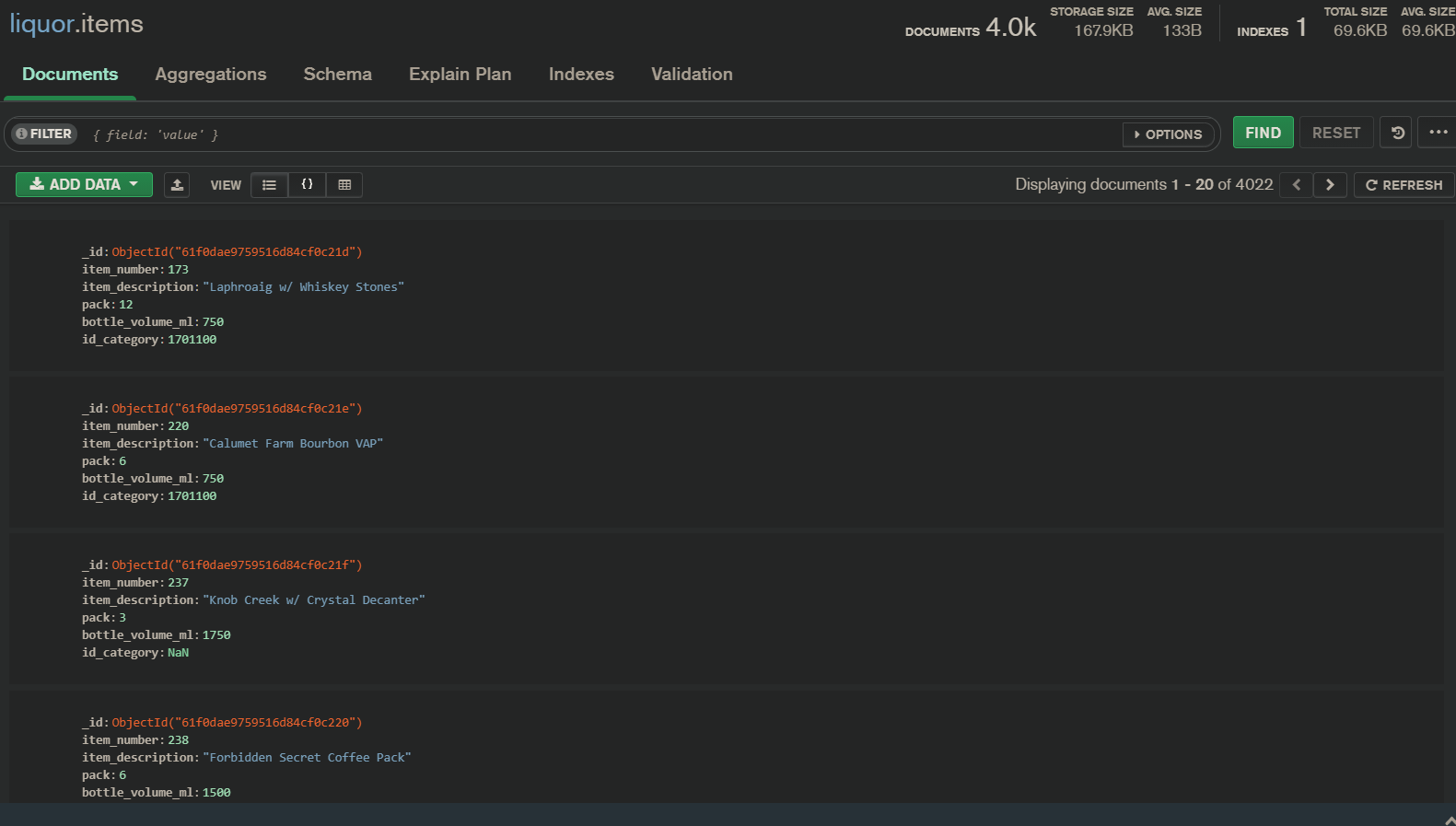
### **Category**

****

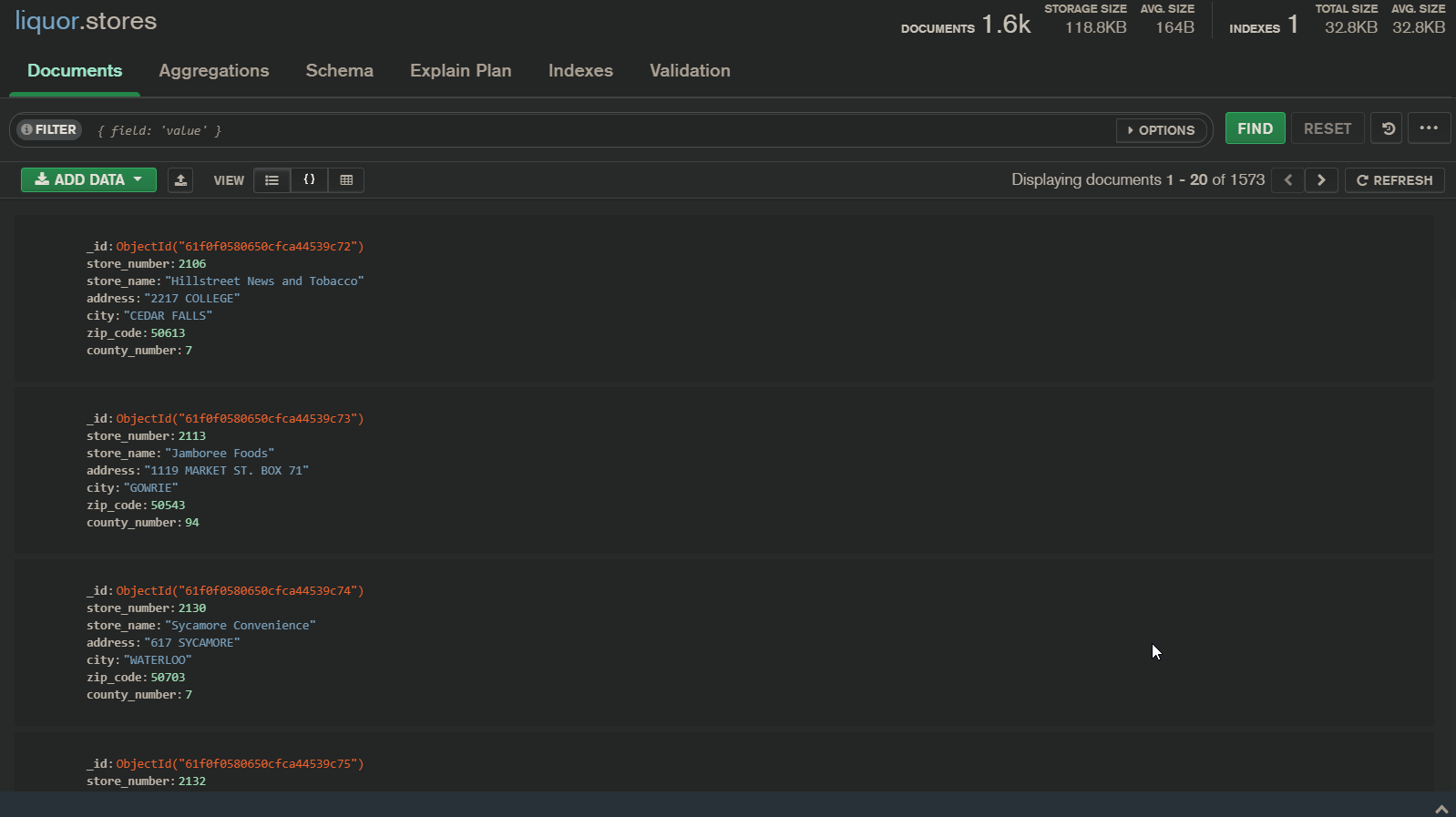
### **County**

****

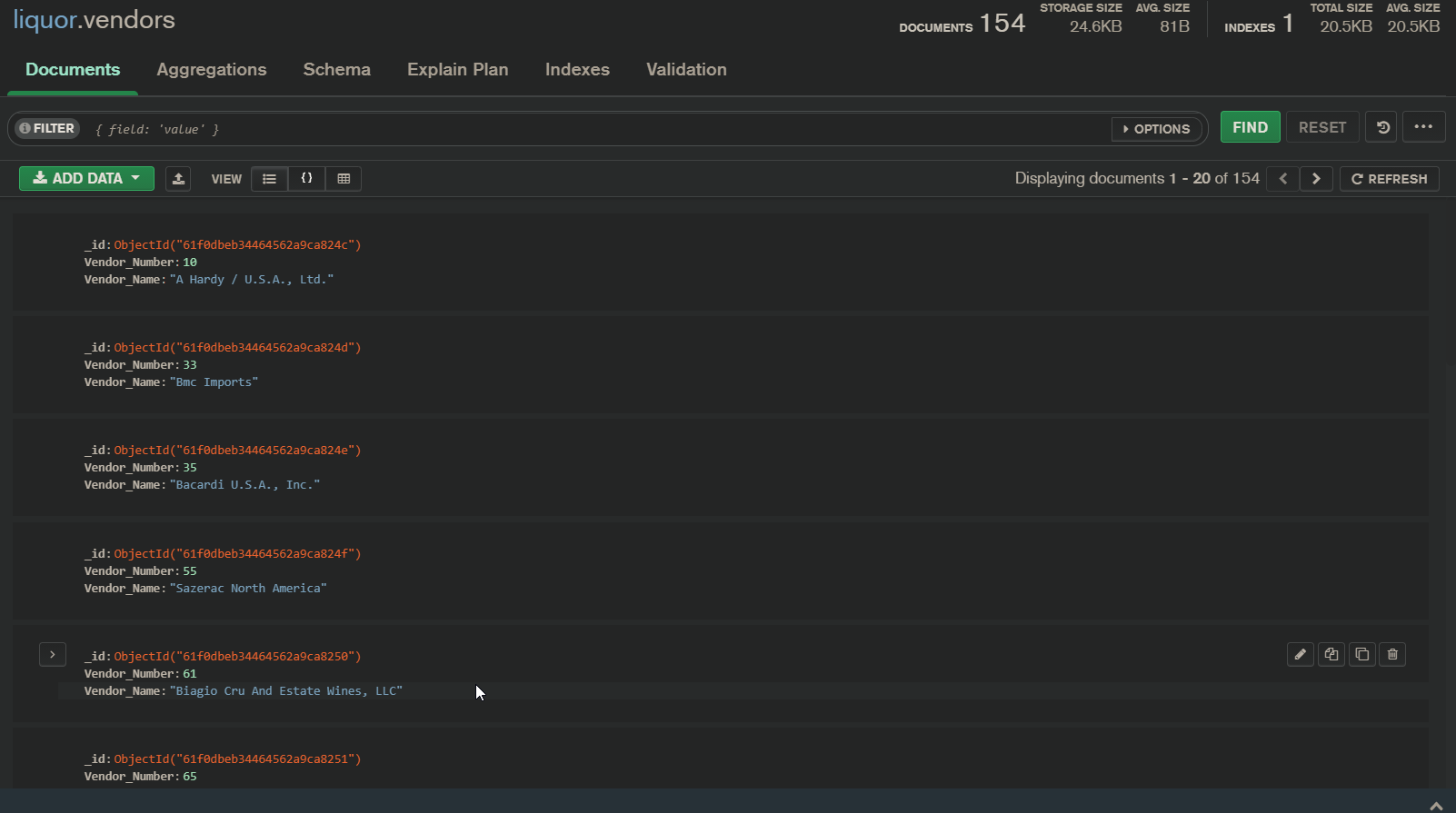
### **Items**

****

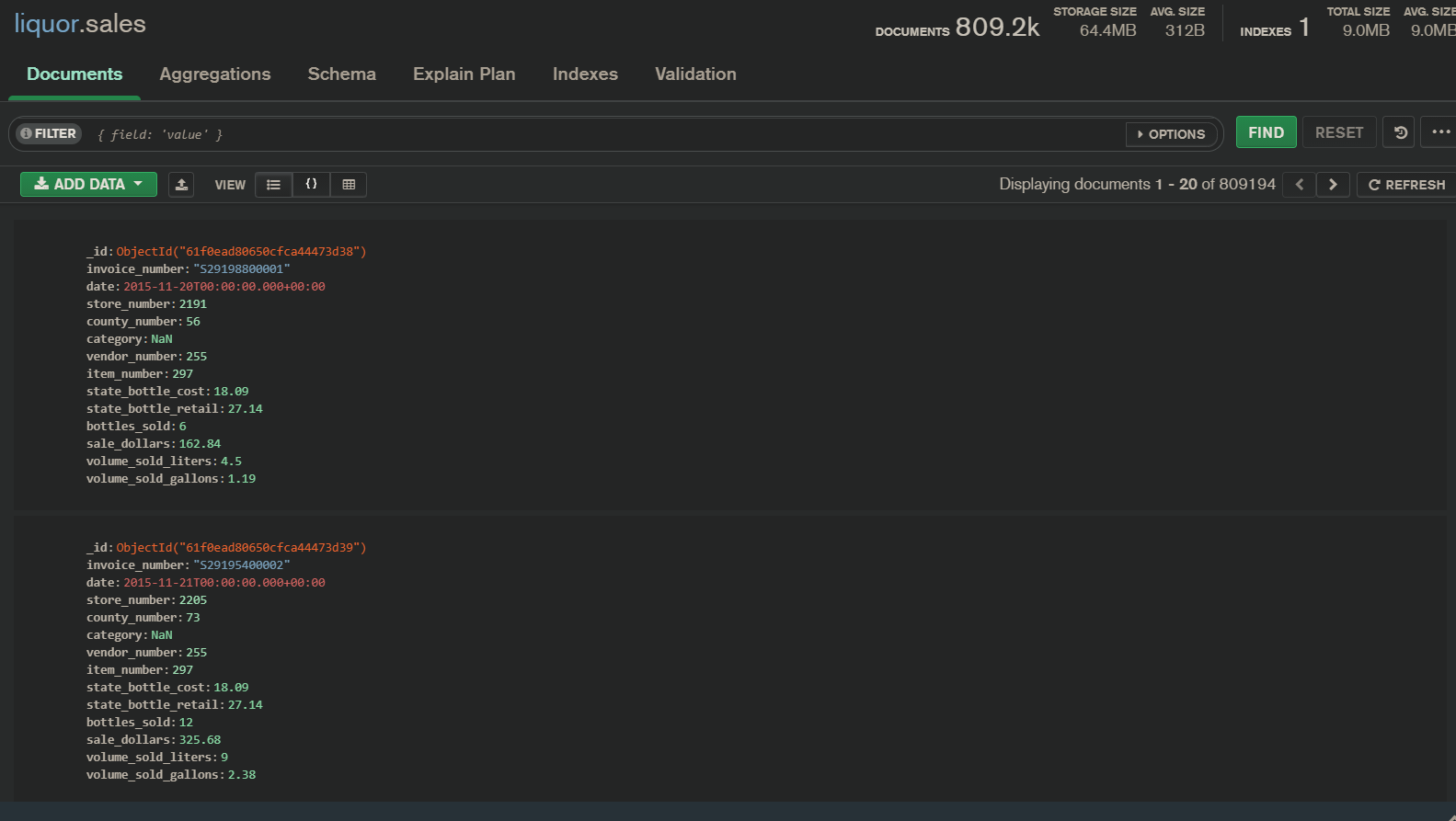
### **Store**

****

### **Vendors**

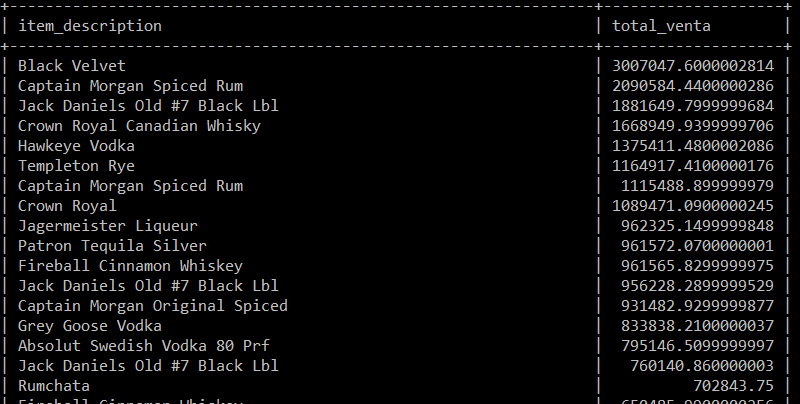
****

### **Sales**

****

# **Consultas MySQL**

## ¿Qué productos generan mayor volumen de ventas?



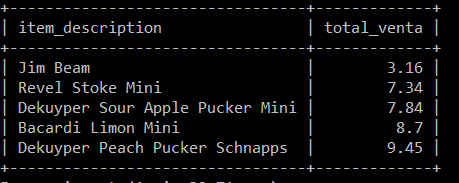
select i.item\_description, sum(sale\_dollars) as total\_venta

from items i join sales s on i.item\_number = s.item\_number

group by (item\_description)

order by total\_venta desc;

## ¿Cuáles son los 5 productos con menor volumen de venta?



select i.item\_description, sum(sale\_dollars) as total\_venta

from items i join sales s on i.item\_number = s.item\_number

group by (item\_description)

order by total\_venta asc

limit 5;

## ¿Cuáles son las 10 ciudades donde se consume mayor volumen de alcohol?



select s.city, sum(sl.volume\_sold\_liters) as volumen\_alcohol\_lt

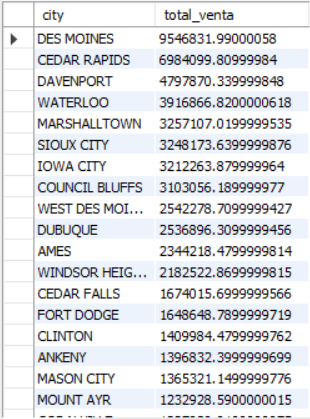
from store s join sales sl on s.store\_number = sl.store\_number

group by (city)

order by volumen\_alcohol\_lt desc

limit 10;

## ¿En qué ciudad se consume el mayor volumen de ventas?



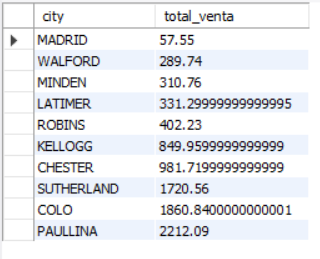
select s.city, sum(sl.sale\_dollars) as total\_venta

from store s join sales sl on s.store\_number = sl.store\_number

group by (city)

order by total\_venta desc;

## ¿En qué ciudad no conviene tener un punto de venta de acuerdo al monto de venta?



select s.city, sum(sl.sale\_dollars) as total\_venta

from store s join sales sl on s.store\_number = sl.store\_number

group by (city)

order by total\_venta asc

limit 1;

## ¿Cuáles las categorías que más venden?



select c.category\_name, sum(s.sale\_dollars) as venta\_total

from category c join sales s on c.category\_number = s.category

group by s.category

order by venta\_total desc

limit 1;

## ¿Cuál es la categoría que más vende por ciudad?



select st.city, c.category\_name, sum(s.sale\_dollars) as venta\_total

from category c join sales s on c.category\_number = s.category

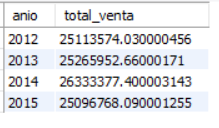
join store st on st.store\_number = s.store\_number

group by st.city, s.category

order by venta\_total desc

limit 20;

## ¿En qué fecha se incrementa el volumen de ventas?

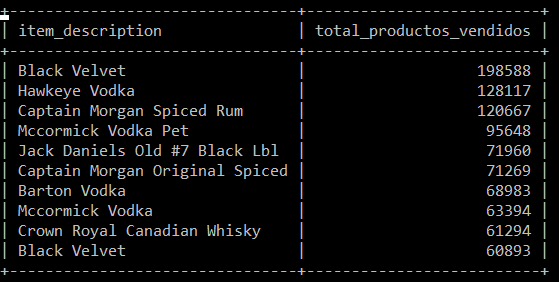


select YEAR(date) as anio, SUM(sale\_dollars) as total\_venta

from sales

group by anio;

## ¿Qué producto se vende más por cantidad?



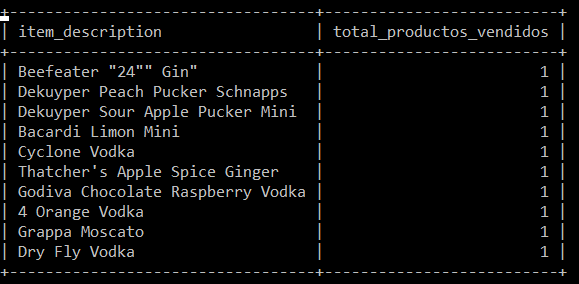
select i.item\_description, sum(pack) as total\_productos\_vendidos

from items i join sales s on i.item\_number = s.item\_number

group by item\_description

order by total\_productos\_vendidos desc;

## ¿Qué producto se vende menos por cantidad?



select i.item\_description, sum(bottles\_sold) as total\_productos\_vendidos

from items i join sales s on i.item\_number = s.item\_number

group by i.item\_number

order by total\_productos\_vendidos asc

limit 10;

## ¿En qué sucursal se muestra mayor volumen de ventas?



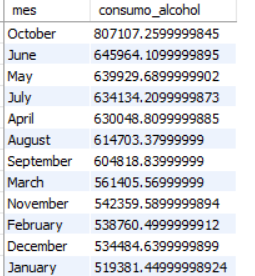
select s.store\_name, count(sl.invoice\_number) as no\_ventas

from store s join sales sl on s.store\_number = sl.store\_number

group by s.store\_number

order by no\_ventas desc;

## ¿En qué mes hay mayor menor consumo de alcohol?



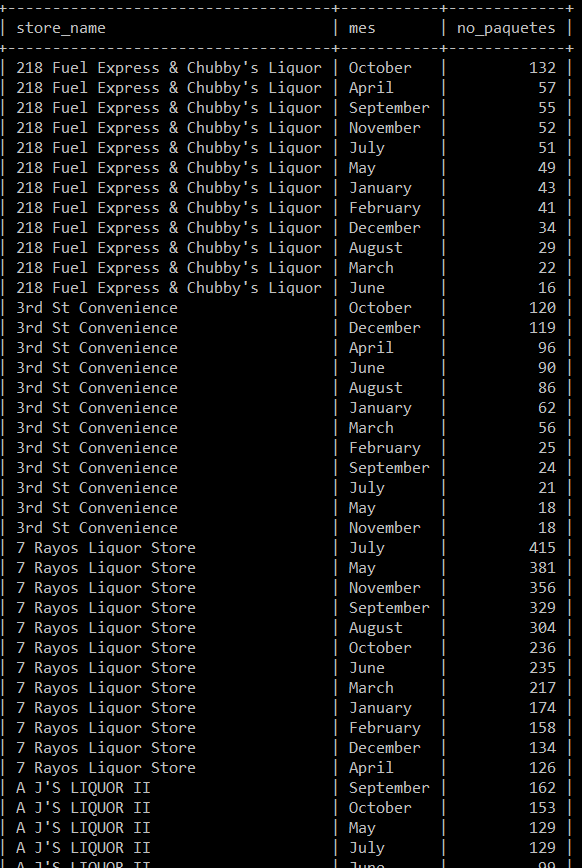
select monthname(date) as mes, sum(volume\_sold\_liters) as consumo\_alcohol

from sales

group by mes

order by consumo\_alcohol desc;

## Cuantos paquetes se consumen por tiendas en cada mes



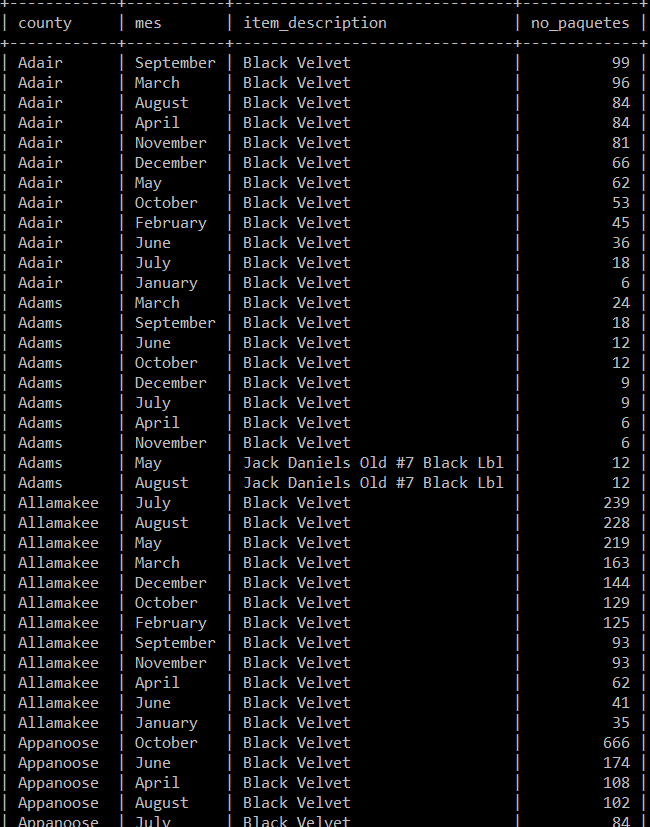
select s.store\_name, monthname(sl.date) as mes, sum(sl.bottles\_sold) as no\_paquetes

from store s join sales sl on s.store\_number = sl.store\_number

group by s.store\_number, mes

order by s.store\_name, no\_paquetes desc;

## Cuantos paquetes de los 3 productos más vendidos se consumen por condado en un mes



select c.county, monthname(s.date) as mes, i.item\_description ,sum(s.bottles\_sold) as no\_paquetes

from county c join sales s on c.county\_number = s.county\_number

join items i on i.item\_number = s.item\_number

where i.item\_number in (select number\_item FROM

(select i.item\_number as number\_item, sum(sale\_dollars) as total\_venta

from items i join sales s on i.item\_number = s.item\_number

group by (i.item\_number)

order by total\_venta desc

limit 3) no\_product\_sale

)

group by c.county, mes

order by county, i.item\_description, no\_paquetes desc;

## Determina la utilidad de cada producto

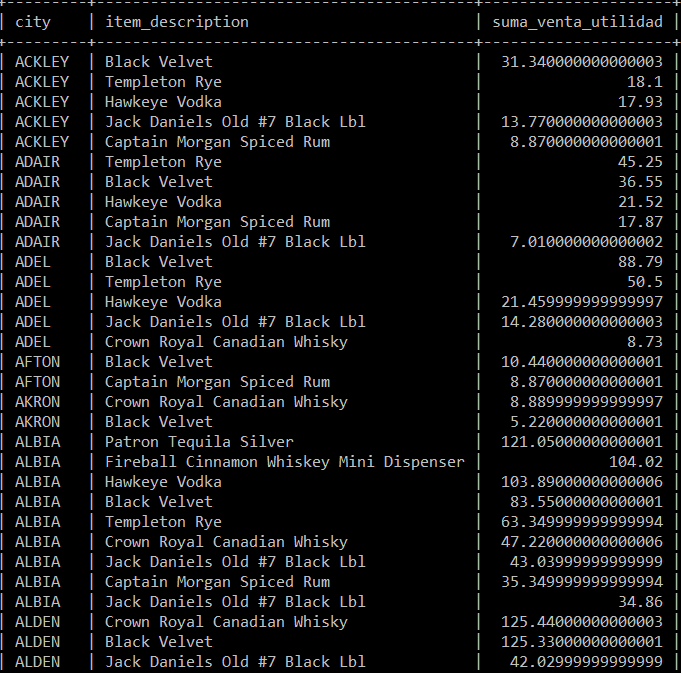


select i.item\_description, (state\_bottle\_retail-state\_bottle\_cost) as utilidad

from items i join sales s on i.item\_number = s.item\_number

group by item\_number;

## Determina los 10 productos que generan mayor volumen ganancias de utilidad y de esos 10 obtén la ganancia total de utilidad por ciudad



select st.city, i.item\_description, sum(state\_bottle\_retail-state\_bottle\_cost) as suma\_venta\_utilidad

from items i join sales s on i.item\_number = s.item\_number

join store st on st.store\_number = s.store\_number

where i.item\_number in (select number\_item FROM

(select i.item\_number as number\_item, sum(state\_bottle\_retail-state\_bottle\_cost) as total\_utilidad

from items i join sales s on i.item\_number = s.item\_number

group by (number\_item)

order by total\_utilidad desc

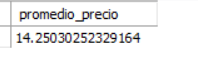
limit 10) no\_product\_sale

)

group by st.city, i.item\_number

order by st.city, suma\_venta\_utilidad desc;

## Obtener el promedio del precio de todos los productos



select avg(state\_bottle\_retail) as promedio\_precio from sales;

## ¿Cuántos productos se venden por arriba de la media?



select sum(i.pack) as product\_tot\_arriba\_media

from items i join sales s on i.item\_number = s.item\_number

where s.state\_bottle\_retail > (select avg(state\_bottle\_retail) as promedio\_precio from sales);

## ¿Cuántos productos se venden por abajo de la media?



select sum(i.pack) as product\_tot\_abajo\_media

from items i join sales s on i.item\_number = s.item\_number

where s.state\_bottle\_retail < (select avg(state\_bottle\_retail) as promedio\_precio from sales);

# **Consultas MongoDB**

## ¿Qué productos generan mayor volumen de ventas? Muestra los últimos 5



[{

$group: {

\_id: '$item\_number',

total\_venta: {

$sum: '$sale\_dollars'

},

group: {

$sum: 1

}

}

}, {

$sort: {

total\_venta: -1

}

}, {

$lookup: {

from: 'items',

localField: '\_id',

foreignField: 'item\_number',

as: 'item\_sales'

}

}, {

$addFields: {

product\_list: {

$arrayElemAt: [

'$item\_sales',

0

]

}

}

}, {

$addFields: {

item\_description: '$product\_list.item\_description'

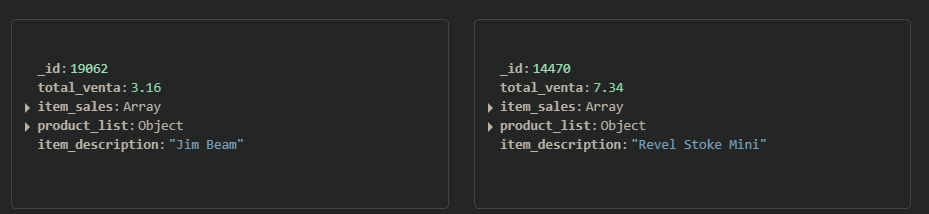
}

}, {

$limit: 5

}]

## ¿Cuáles son los 5 productos con menor volumen de venta?



[{$group: {

\_id: '$item\_number',

total\_venta: {

$sum: '$sale\_dollars'

}

}}, {$sort: {

total\_venta: 1

}}, {$lookup: {

from: 'items',

localField: '\_id',

foreignField: 'item\_number',

as: 'item\_sales'

}}, {$addFields: {

product\_list: {

$arrayElemAt: [

'$item\_sales',

0

]

}

}}, {$addFields: {

item\_description: '$product\_list.item\_description'

}}, {$limit: 5}]

## ¿En qué tiendas se consume el mayor volumen de ventas?



[{$group: {

\_id: '$store\_number',

total\_venta\_tienda: {

$sum: '$sale\_dollars'

}

}}, {$sort: {

total\_venta\_tienda: -1

}}, {$lookup: {

from: 'stores',

localField: '\_id',

foreignField: 'store\_number',

as: 'stores\_list'

}}, {$addFields: {

store\_info: {

$arrayElemAt: [

'$stores\_list',

0

]

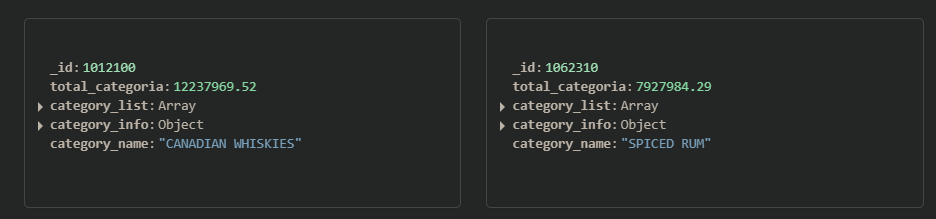
}

}}, {$addFields: {

store\_name: '$store\_info.store\_name'

}}]

## ¿Cuál es la categoría que mas vende ?



[{$group: {

\_id: '$category',

total\_categoria: {

$sum: '$sale\_dollars'

}

}}, {$sort: {

total\_categoria: -1

}}, {$lookup: {

from: 'category',

localField: '\_id',

foreignField: 'Category\_Number',

as: 'category\_list'

}}, {$addFields: {

category\_info: {

$arrayElemAt: [

'$category\_list',

0

]

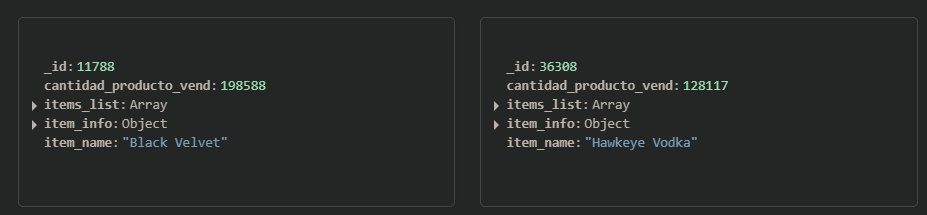
}

}}, {$addFields: {

category\_name: '$category\_info.Category\_Name'

}}]

## ¿Cuál es el producto que se vende mas?



[{$group: {

\_id: '$item\_number',

cantidad\_producto\_vend: {

$sum: '$bottles\_sold'

}

}}, {$sort: {

cantidad\_producto\_vend: -1

}}, {$lookup: {

from: 'items',

localField: '\_id',

foreignField: 'item\_number',

as: 'items\_list'

}}, {$addFields: {

item\_info: {

$arrayElemAt: [

'$items\_list',

0

]

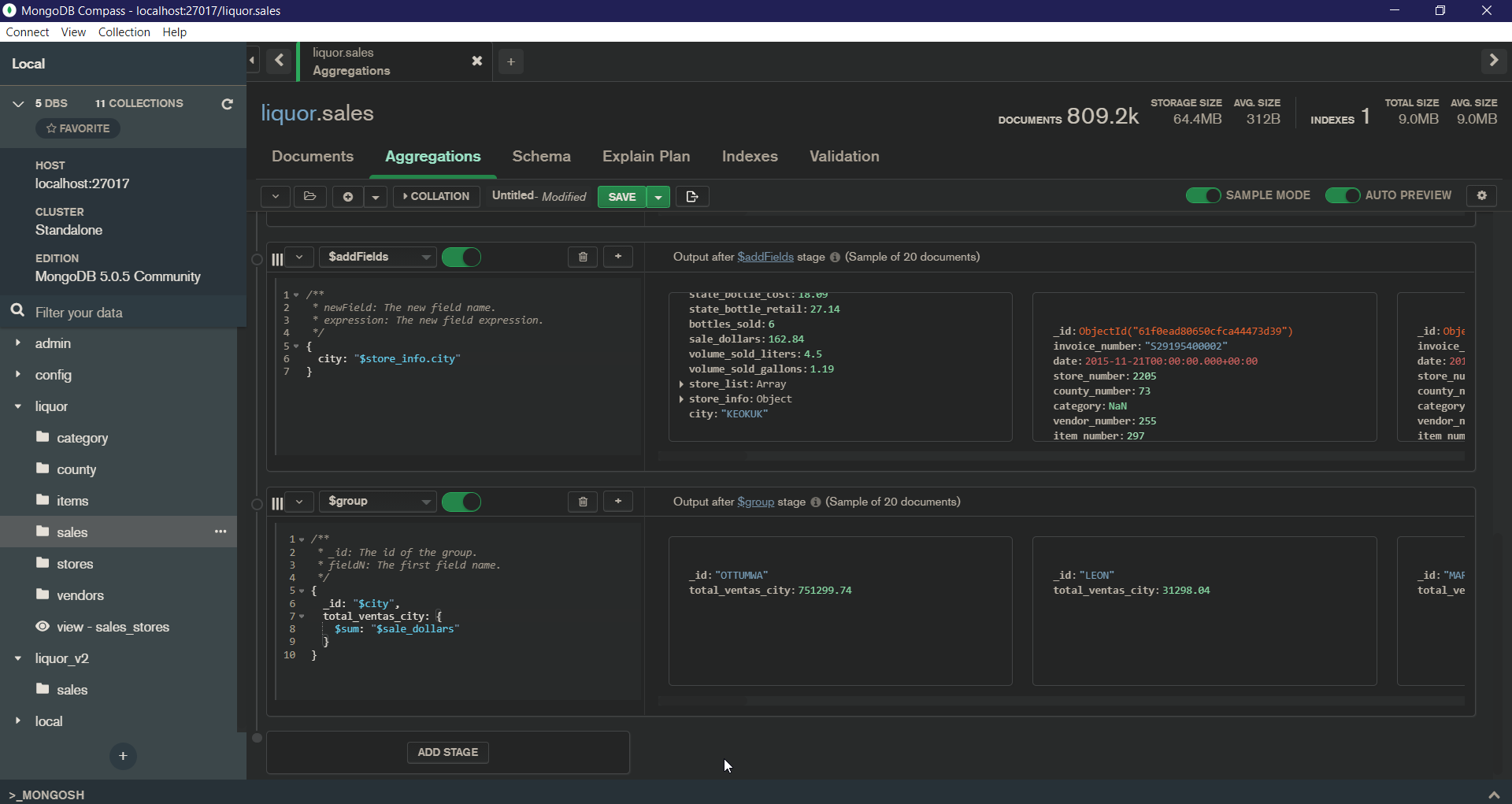
}

}}, {$addFields: {

item\_name: '$item\_info.item\_description'

}}]

## ¿En qué ciudad se consume el mayor volumen de ventas?



[{$lookup: {

from: 'stores',

localField: 'store\_number',

foreignField: 'store\_number',

as: 'store\_list'

}}, {$addFields: {

store\_info: {

$arrayElemAt: [

'$store\_list',

0

]

}

}}, {$addFields: {

city: '$store\_info.city'

}}, {$group: {

\_id: '$city',

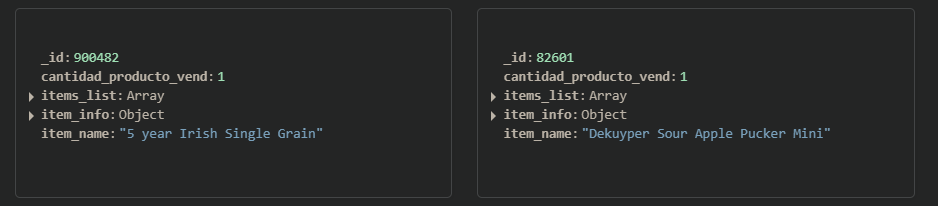
total\_ventas\_city: {

$sum: '$sale\_dollars'

}

}}]

## ¿Cuál es el producto que se vende menos?



[{$group: {

\_id: '$item\_number',

cantidad\_producto\_vend: {

$sum: '$bottles\_sold'

}

}}, {$sort: {

cantidad\_producto\_vend: 1

}}, {$lookup: {

from: 'items',

localField: '\_id',

foreignField: 'item\_number',

as: 'items\_list'

}}, {$addFields: {

item\_info: {

$arrayElemAt: [

'$items\_list',

0

]

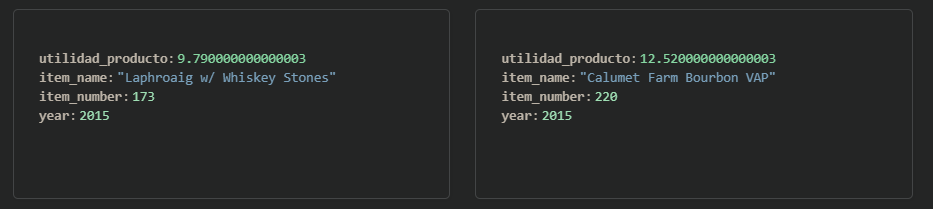
}

}}, {$addFields: {

item\_name: '$item\_info.item\_description'

}}]

## Determina la utilidad de cada producto



[{$addFields: {

utilidad\_producto: {

$subtract: [

'$state\_bottle\_retail',

'$state\_bottle\_cost'

]

}

}}, {$group: {

\_id: {

item\_number: '$item\_number',

year: {

$year: '$date'

}

},

utilidad\_producto: {

$addToSet: '$utilidad\_producto'

}

}}, {$unwind: {

path: '$utilidad\_producto'

}}, {$lookup: {

from: 'items',

localField: '\_id.item\_number',

foreignField: 'item\_number',

as: 'item\_list'

}}, {$addFields: {

item\_info: {

$arrayElemAt: [

'$item\_list',

0

]

}

}}, {$addFields: {

item\_name: '$item\_info.item\_description'

}}, {$addFields: {

item\_number: '$\_id.item\_number',

year: '$\_id.year'

}}, {$project: {

\_id: 0,

item\_name: 1,

utilidad\_producto: 1,

item\_number: 1,

year: 1

}}, {$sort: {

item\_number: 1

}}]

## ¿En qué año se incrementa el volumen de ventas?



[{$group: {

\_id: {

$year: '$date'

},

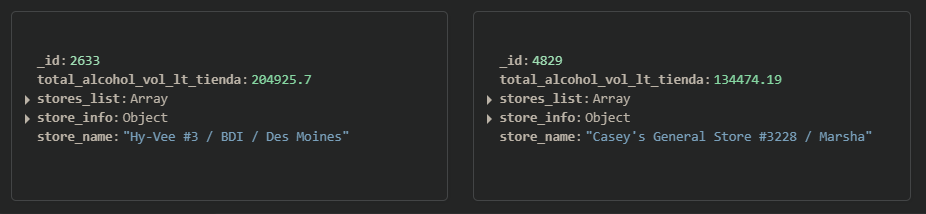
venta\_anio: {

$sum: '$sale\_dollars'

}

}}]

## ¿Cuáles son las 10 tiendas donde se consume mayor volumen de alcohol?



[{$group: {

\_id: '$store\_number',

total\_alcohol\_vol\_lt\_tienda: {

$sum: '$volume\_sold\_liters'

}

}}, {$sort: {

total\_alcohol\_vol\_lt\_tienda: -1

}}, {$lookup: {

from: 'stores',

localField: '\_id',

foreignField: 'store\_number',

as: 'stores\_list'

}}, {$addFields: {

store\_info: {

$arrayElemAt: [

'$stores\_list',

0

]

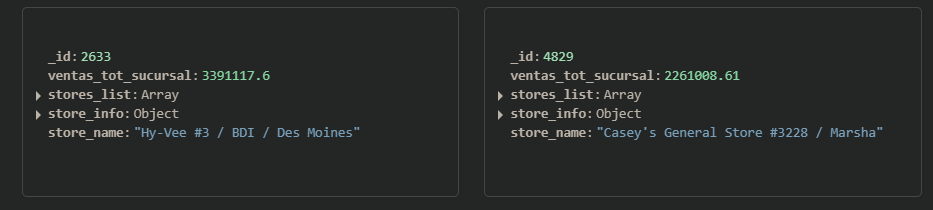
}

}}, {$addFields: {

store\_name: '$store\_info.store\_name'

}}, {$limit: 10}]

## ¿En qué sucursal se muestra mayor volumen de ventas?



[{$group: {

\_id: '$store\_number',

ventas\_tot\_sucursal: {

$sum: '$sale\_dollars'

}

}}, {$sort: {

ventas\_tot\_sucursal: -1

}}, {$lookup: {

from: 'stores',

localField: '\_id',

foreignField: 'store\_number',

as: 'stores\_list'

}}, {$addFields: {

store\_info: {

$arrayElemAt: [

'$stores\_list',

0

]

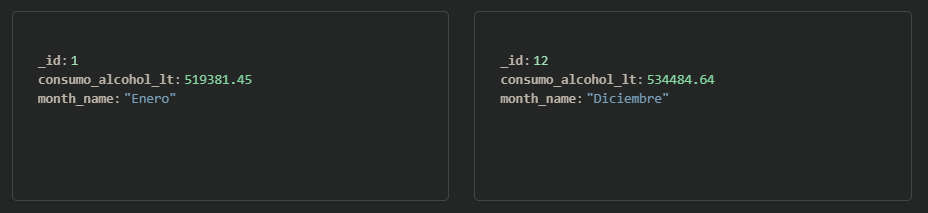
}

}}, {$addFields: {

store\_name: '$store\_info.store\_name'

}}]

## ¿En qué mes hay menor consumo de alcohol?



[{$group: {

\_id: {

$month: '$date'

},

consumo\_alcohol\_lt: {

$sum: '$volume\_sold\_liters'

}

}}, {$addFields: {

month\_name: {

$switch: {

branches: [

{ case: { $eq: [ "$\_id" ,1] }, then: "Enero" },

{ case: { $eq: [ "$\_id" ,2] }, then: "Febrero" },

{ case: { $eq: [ "$\_id" ,3] }, then: "Marzo" },

{ case: { $eq: [ "$\_id" ,4] }, then: "Abril" },

{ case: { $eq: [ "$\_id" ,5] }, then: "Mayo" },

{ case: { $eq: [ "$\_id" ,6] }, then: "Junio" },

{ case: { $eq: [ "$\_id" ,7] }, then: "Julio" },

{ case: { $eq: [ "$\_id" ,8] }, then: "Agosto" },

{ case: { $eq: [ "$\_id" ,9] }, then: "Septiembre" },

{ case: { $eq: [ "$\_id" ,10] }, then: "Octubre" },

{ case: { $eq: [ "$\_id" ,11] }, then: "Noviembre" },

{ case: { $eq: [ "$\_id" ,12] }, then: "Diciembre" }

],

default:"Did not match"

}

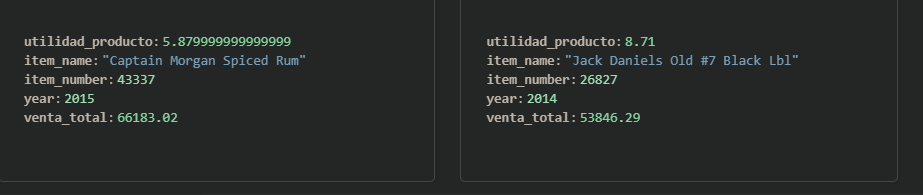
}

}}, {$sort: {

consumo\_alcohol\_lt: 1

}}]

## Determina los 10 productos que generan mayor volumen ganancias de utilidad



[{$addFields: {

utilidad\_producto: {

$subtract: [

'$state\_bottle\_retail',

'$state\_bottle\_cost'

]

}

}}, {$group: {

\_id: {

item\_number: '$item\_number',

year: {

$year: '$date'

}

},

utilidad\_producto: {

$addToSet: '$utilidad\_producto'

},

sale\_dollars: {

$addToSet: '$sale\_dollars'

}

}}, {$unwind: {

path: '$utilidad\_producto'

}}, {$lookup: {

from: 'items',

localField: '\_id.item\_number',

foreignField: 'item\_number',

as: 'item\_list'

}}, {$addFields: {

item\_info: {

$arrayElemAt: [

'$item\_list',

0

]

}

}}, {$addFields: {

item\_name: '$item\_info.item\_description'

}}, {$addFields: {

item\_number: '$\_id.item\_number',

year: '$\_id.year'

}}, {$addFields: {

venta\_total: {

$sum: '$sale\_dollars'

}

}}, {$project: {

\_id: 0,

item\_name: 1,

utilidad\_producto: 1,

item\_number: 1,

year: 1,

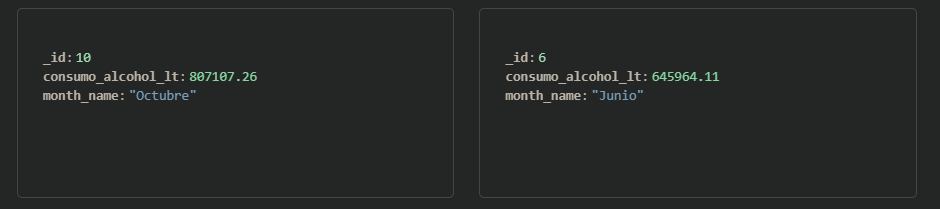
venta\_total: 1

}}, {$sort: {

venta\_total: -1

}}, {$limit: 10}]

## ¿En qué mes hay mayor consumo de alcohol?



[{$group: {

\_id: {

$month: '$date'

},

consumo\_alcohol\_lt: {

$sum: '$volume\_sold\_liters'

}

}}, {$addFields: {

month\_name: {

$switch: {

branches: [

{ case: { $eq: [ "$\_id" ,1] }, then: "Enero" },

{ case: { $eq: [ "$\_id" ,2] }, then: "Febrero" },

{ case: { $eq: [ "$\_id" ,3] }, then: "Marzo" },

{ case: { $eq: [ "$\_id" ,4] }, then: "Abril" },

{ case: { $eq: [ "$\_id" ,5] }, then: "Mayo" },

{ case: { $eq: [ "$\_id" ,6] }, then: "Junio" },

{ case: { $eq: [ "$\_id" ,7] }, then: "Julio" },

{ case: { $eq: [ "$\_id" ,8] }, then: "Agosto" },

{ case: { $eq: [ "$\_id" ,9] }, then: "Septiembre" },

{ case: { $eq: [ "$\_id" ,10] }, then: "Octubre" },

{ case: { $eq: [ "$\_id" ,11] }, then: "Noviembre" },

{ case: { $eq: [ "$\_id" ,12] }, then: "Diciembre" }

],

default:"Did not match"

}

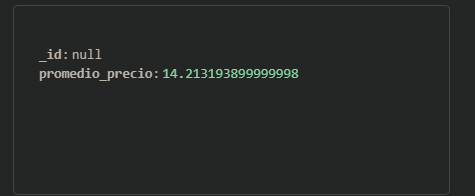
}

}}, {$sort: {

consumo\_alcohol\_lt: -1

}}]

## Obtener el promedio del precio de todos los productos



[{$group: {

\_id: null,

promedio\_precio: {

$avg: '$state\_bottle\_retail'

}

}}]

# **Conclusiones**

Para realizar un estudio de mercado con estos datos es necesario aun profundizar más sobre los mismos, pero da una vista sobre qué productos nos interesan y en qué ciudad sería conveniente tener un punto de venta, además de saber en qué fechas es posible que se solicite más un producto y así mismo saber en qué mes se consume más alcohol.

MongoDB y MySQL son herramientas muy potentes sin embargo cada una tiene sus cualidades para trabajar dependiendo el tipo de proyecto que se requiera.