Stephanie Haidenberg

Módulo 1

Abstracto

El presente proyecto se realizó con el objetivo de demostrar el manejo de MySQL y MongoDB  
Se hizo uso de una base de datos ficticia con datos de una tienda en línea en Brasil

Brazilian ecommerce

Checkpoint 1 – Data Analysis

**Definición de proyecto**

1. **Problema a resolver**

* Mejorar la satisfacción del cliente por medio de la detección de áreas de oportunidad en las ordenes enviadas.
* Objetivo a cierre de año: calificación > 4.5

1. **Definición de la base de datos**
   1. *Fuentes de BD*

La base de datos fue descargada del sitio kaggle.com

<https://www.kaggle.com/olistbr/brazilian-ecommerce?select=olist_order_payments_dataset.csv>

La base de datos está compuesta de las siguientes tablas:

* Products: enlista todos los productos. Llave primaria product\_id.
* Sellers: enlista todos los vendedores, así como su ciudad, estado y código postal.

Llave primaria: seller\_id.

* Orders: enlista todas las órdenes y su status.

Llave primaria: order\_id. Llave foránea: customer\_id

* Order\_items: enlista todas las órdenes y nos da información sobre el producto, su vendedor, límite de entrega, precio y peso.

Llave primaria: order\_id. Llaves foráneas: product\_id, seller\_id.

* Order\_reviews: se compone de las calificaciones, título de comentario, mensaje, fecha de creación, fecha de respuesta.

Llave primaria: review\_id. Llave foránea: order\_id

* Customers: Información acerca de los clientes. Se compone de un id de cliente, su código postal, ciudad y estado.

Llave primaria: customer\_id. Llave foránea: customer\_zip\_code\_prefix

* Geolocation: Se compone del código postal, latitud, longitud, ciudad y estado.

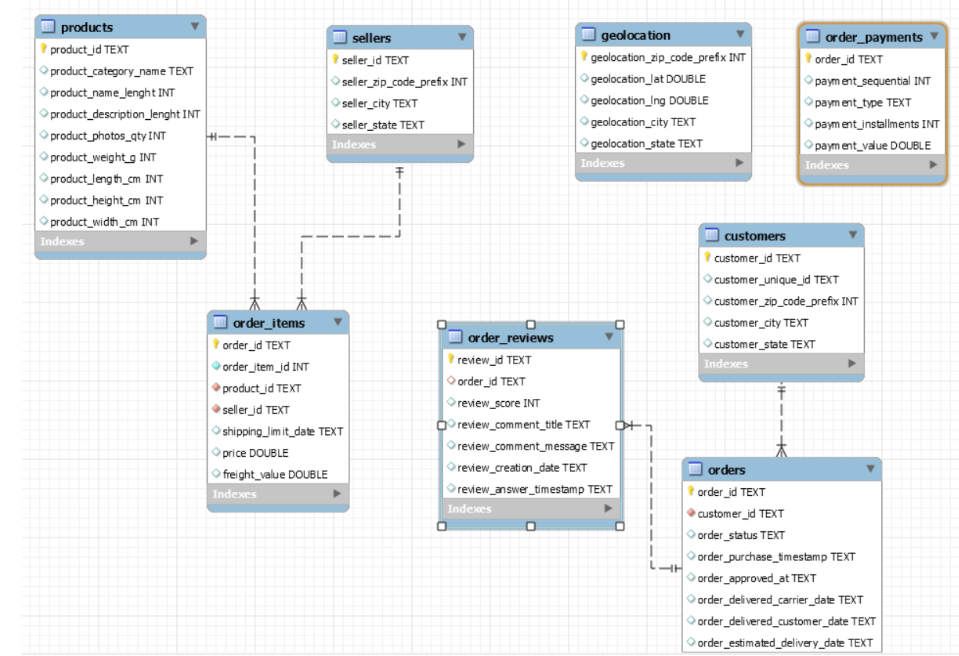
Llave primaria: geolocation\_id. Llave foránea: geolocation\_zip\_code\_prefix

* Order paymets: Enlista la orden, el número de pago, tipo de pago, y el valor.

Llave primaria: order\_id.

Todas las tablas de Orders se unen por medio de la llave primaria: order\_id

**Diagrama Entidad - Relación**





**Consultas**

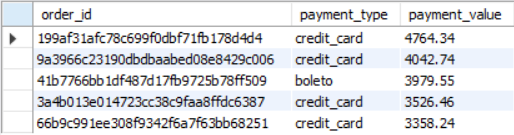
1. **¿Cuáles son las ordenes que tienen una calificación mayor a 3?**

SELECT order\_id, review\_score FROM order\_reviews WHERE review\_score > 3;



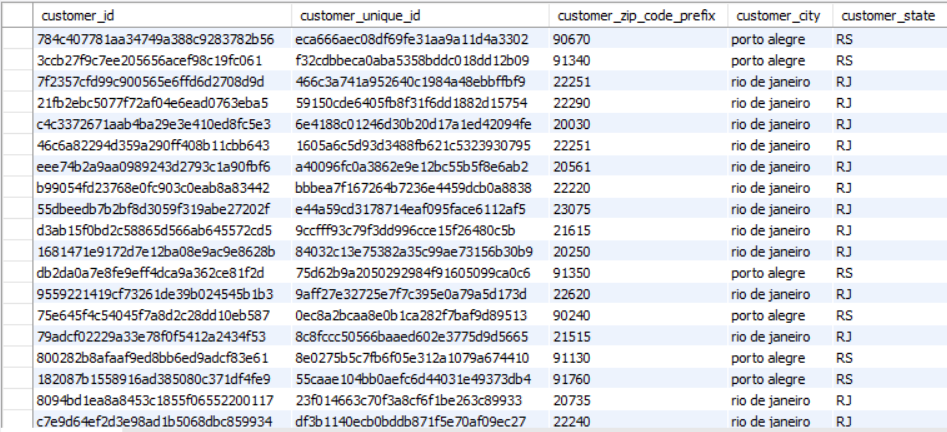
1. **¿Cuáles son las 5 órdenes con el valor de transaccion más alto y cuál fue su método de pago?**

SELECT order\_id, payment\_type, payment\_value FROM order\_payments ORDER BY payment\_value DESC LIMIT 5;



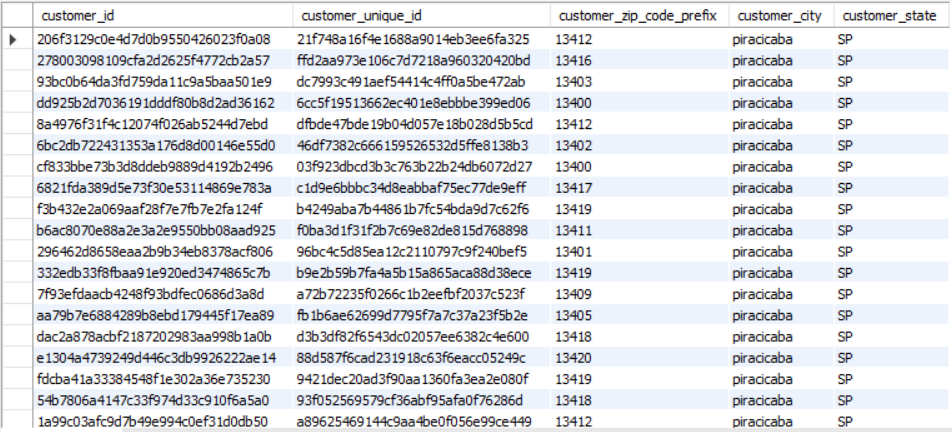
1. **¿Cuáles son todos los clientes que hicieron pedido de Rio de Janeiro o Porto Alegre?**

SELECT \* FROM customers WHERE customer\_city IN ("Rio de Janeiro", "Porto alegre");



1. **¿Cuáles son todos los clientes que compraron del estado de SP y de la ciudad de Piracicaba?**

SELECT \* FROM customers WHERE (customer\_state = "SP") AND (customer\_city = "piracicaba");



1. **¿Cuál es la calificación promedio de todas las ordenes realizadas?**

SELECT avg(review\_score) FROM order\_reviews;



1. **¿Cuáles son todos los comentarios que mencionan que el producto es de mala calidad (inferior) y que el producto llegó con retraso (atrasado)?**

SELECT order\_id, review\_score, review\_comment\_message, review\_creation\_date FROM order\_reviews WHERE review\_comment\_message LIKE "%inferior%" OR review\_comment\_message LIKE "%atrasado%";



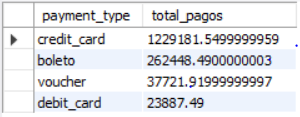


1. **¿Cuál es el método de pago más utilizado por total facturación y total pedidos?**

SELECT payment\_type, sum(payment\_value) as total\_pagos

FROM order\_payments

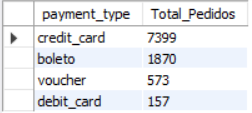
GROUP BY payment\_type;



SELECT payment\_type, count(\*) AS Total\_Pedidos

FROM order\_payments

GROUP BY payment\_type;



1. **¿Cuántas ordenes no han sido entregadas?**

SELECT order\_status, count(\*) FROM orders WHERE order\_status != "delivered";



1. **Una buena descripción de los productos ayuda mucho a mejorar la calificación de la orden, ¿cuántos son los caracteres mínimos y máximos que hay en las descripciones?**

SELECT min(product\_description\_lenght) AS MinimoCaracts, max(product\_description\_lenght) AS MaximoCaracts FROM products;



1. **Se debe de tener un mínimo de 700 caracteres en las descripciones ¿cuáles son las categorías que tienen menos de 700 caracteres en sus descripciones?**

SELECT product\_category\_name, product\_description\_lenght

FROM products

WHERE product\_description\_lenght < 700

ORDER BY product\_description\_lenght ASC;



1. **¿Cuáles son las 10 órdenes con facturación más alta?**

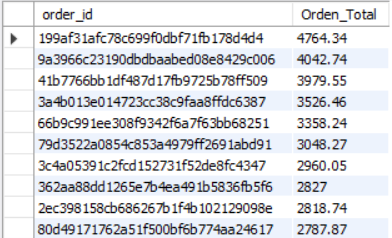
SELECT order\_id, sum(payment\_value) AS Orden\_Total

FROM order\_payments

GROUP BY order\_id

ORDER BY Orden\_Total DESC

LIMIT 10;



1. **¿Cuál es el número de orden, numero de producto y numero de vendedor del producto con el precio más alto?**

SELECT order\_id, product\_id, seller\_id, price

FROM order\_items

WHERE price = (SELECT max(price) FROM order\_items);



1. **¿De qué estado y ciudad son los 10 vendedores con peor calificación?**

SELECT s.seller\_id as seller, r.review\_score as calificacion, s.seller\_state, s.seller\_city

FROM sellers s

JOIN order\_items oi

ON s.seller\_id = oi.seller\_id

JOIN order\_reviews r

ON r.order\_id = oi.order\_id

GROUP BY s.seller\_id

ORDER BY calificacion ASC

LIMIT 10;



1. **¿Cuáles son las 10 ciudades con más ventas?**

SELECT c.customer\_city as ciudad, sum(op.payment\_value) as pago

FROM order\_payments op

JOIN orders o

ON op.order\_id = o.order\_id

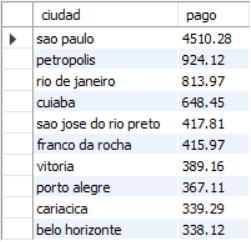
JOIN customers c

ON o.customer\_id = c.customer\_id

GROUP BY c.customer\_city

ORDER BY pago DESC

LIMIT 10;



1. **¿Cuáles son las 10 categorías de producto con mejor calificación y cuáles son las 10 que tienen la peor calificación?**

*La mejor calificación*

SELECT p.product\_category\_name as Categoria, avg(o.review\_score) as Calificacion\_Promedio

FROM products p

JOIN order\_items oi

ON p.product\_id = oi.product\_id

JOIN order\_reviews o

ON o.order\_id = oi.order\_id

GROUP BY Categoria

ORDER BY Calificacion\_Promedio DESC

LIMIT 10;



*La peor calificación*

create view ProyectoSH AS

(SELECT p.product\_category\_name as Categoria, avg(o.review\_score) as Calificacion\_Promedio

FROM products p

JOIN order\_items oi

ON p.product\_id = oi.product\_id

JOIN order\_reviews o

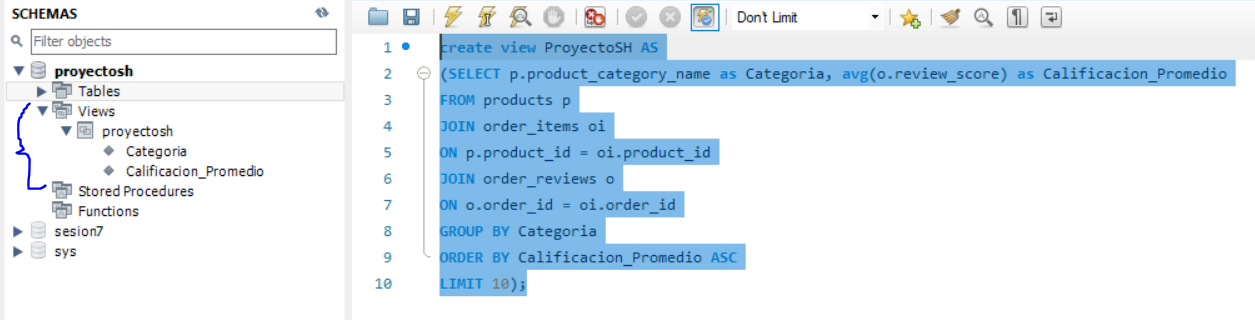
ON o.order\_id = oi.order\_id

GROUP BY Categoria

ORDER BY Calificacion\_Promedio ASC

LIMIT 10);







**Consultas**

1. **¿Cuál es la calificación promedio de todas las ordenes realizadas?**

[{$group: {

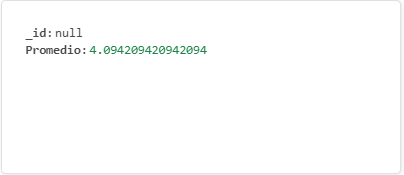
\_id: null,

Promedio: {

$avg: "$review\_score"

}

}}]



1. **¿cuáles son las categorías que tienen más fotos en sus publicaciones?**

[{$match: {

product\_photos\_qty: {$ne: NaN}

}}, {$group: {

\_id: "$product\_category\_name",

Fotos: {

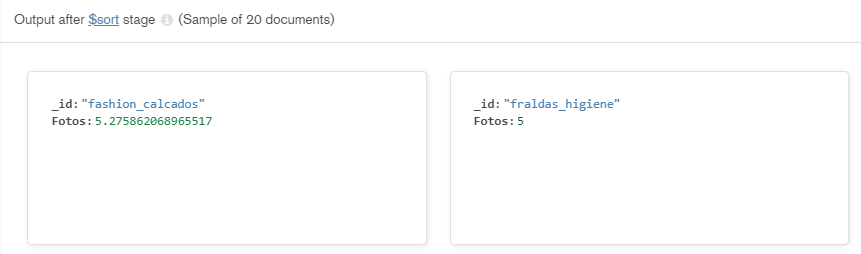
$avg: "$product\_photos\_qty"

}

}}, {$sort: {

Fotos: -1

}}]



1. **¿cuáles son las categorías que tienen menos fotos en sus publicaciones?**

[{$match: {

product\_photos\_qty: {$ne: NaN}

}}, {$group: {

\_id: "$product\_category\_name",

Fotos: {

$avg: "$product\_photos\_qty"

}

}}, {$sort: {

Fotos: 1

}}]



1. **¿Cuáles son las ordenes que tienen una calificación mayor a 3?**

[{$match: {

review\_score: {$gte: 3}

}}]



1. **Se debe de tener un mínimo de 700 caracteres en las descripciones ¿cuáles son las categorías que tienen menos de 700 caracteres en sus descripciones?**

[{$match: {

product\_description\_lenght: {$lte: 700}

}}, {$sort: {

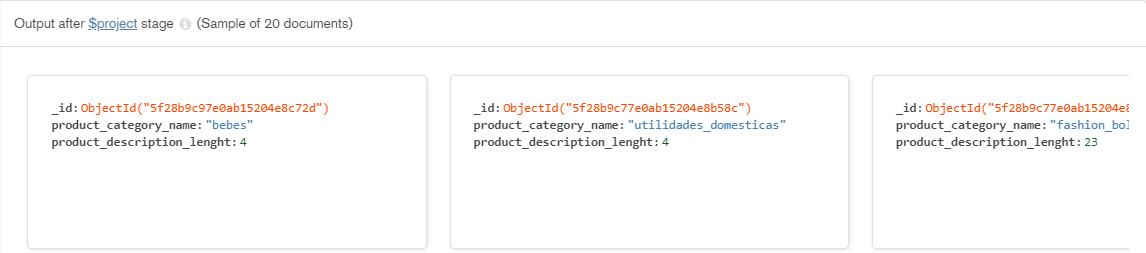
product\_description\_lenght: 1

}}, {$project: {

product\_category\_name:1,

product\_description\_lenght: 1

}}]



1. **¿Qué órdenes recibieron una calificación entre 4 y 5?**

[{$match:

{

review\_score: {$in: [4,5]}

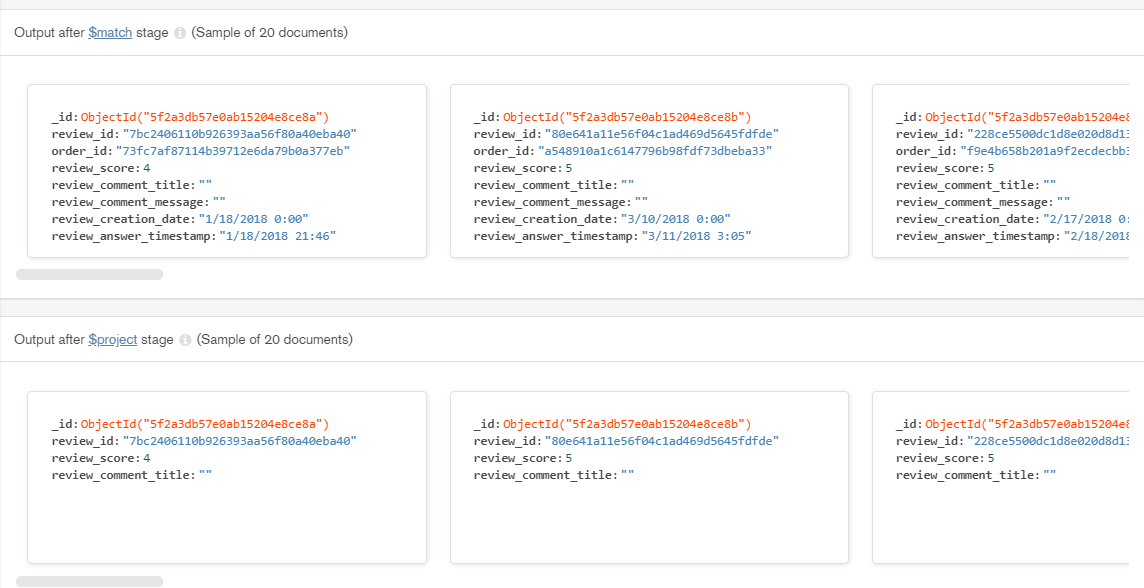
}}, {$project: {

review\_id: 1,

review\_comment\_title: 1,

review\_score: 1

}}]



1. **¿Cuáles son todos los comentarios que mencionan que el producto es de mala calidad (inferior) y que el producto llegó con retraso (atrasado)?**

[{$match:

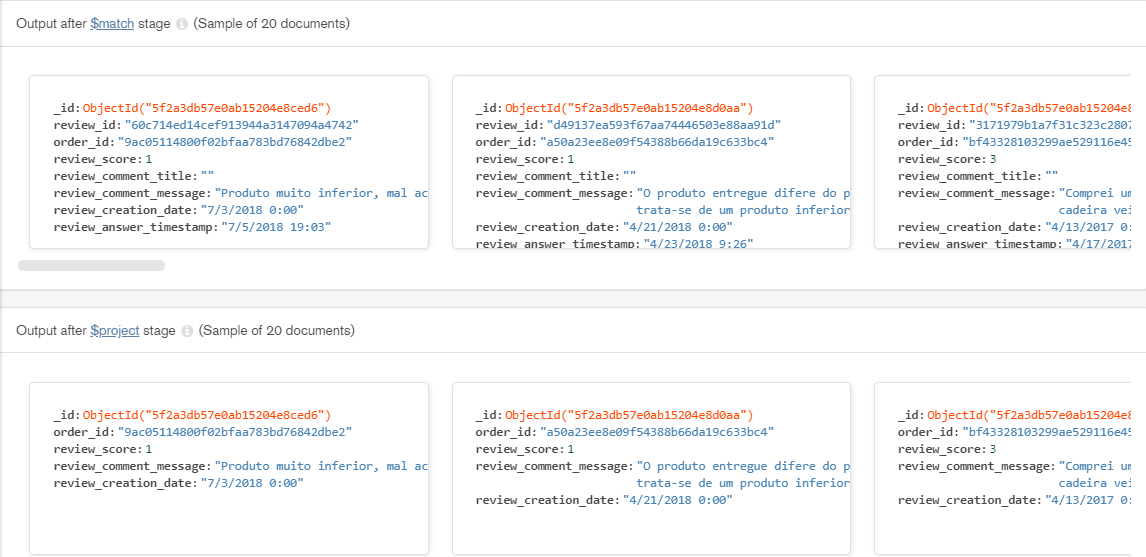
{

review\_comment\_message: {$in: [ /inferior/i, /atrasado/i]},

}}, {$project: {

order\_id: 1, review\_score:1, review\_comment\_message:1, review\_creation\_date:1

}}]



1. **¿Cuál es el método de pago más utilizado por total pedidos?**

[{$group: {

\_id: "$payment\_type",

Total: {

$sum: 1

}

}}, {$sort: {

Total: -1

}}]



1. **¿Cuál es la participación (share%) del principal método de pago?**

[{$group: {

\_id: "$payment\_type",

Pago: {

$sum: 1

}

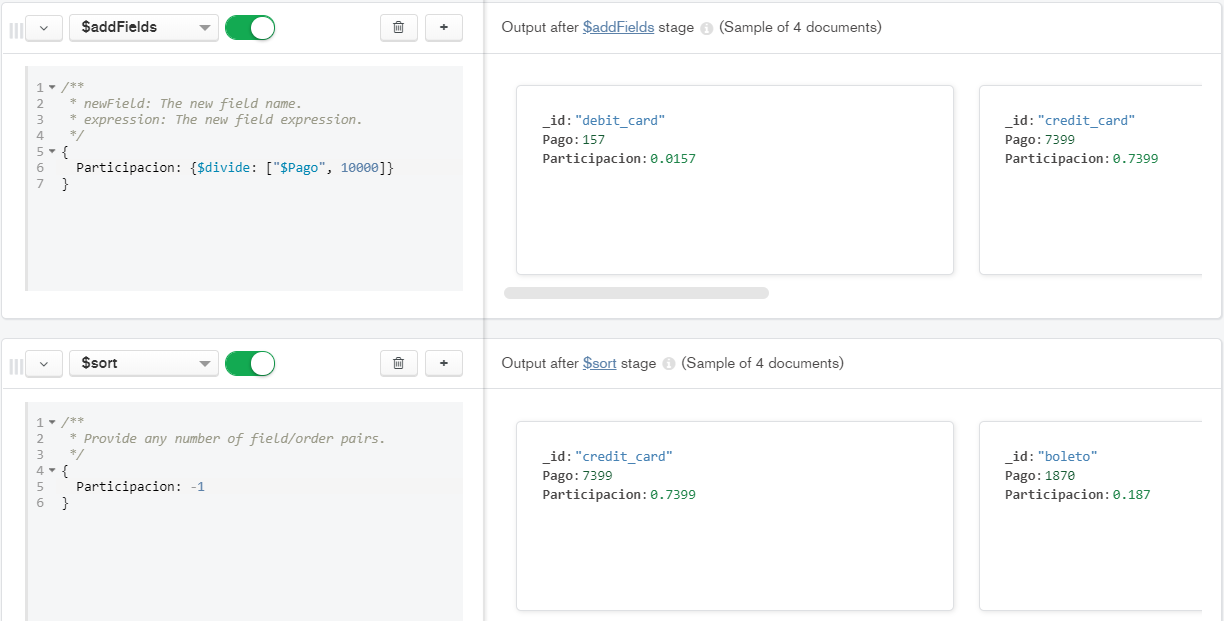
}}, {$addFields: {

Participacion: {$divide: ["$Pago", 10000]}

}}, {$sort: {

Participacion: -1

}}]



1. **¿Cuáles son las 5 órdenes con el valor de transacción más alto?**

[{$group: {

\_id: "$order\_id",

Ventas: {

$sum: "$payment\_value"

}

}}, {$sort: {

Ventas: -1

}}, {$limit: 5}]



1. **¿En cuántos pagos se realizan el mayor número de ventas?**

[{$group: {

\_id: "$payment\_installments",

Ventas: {

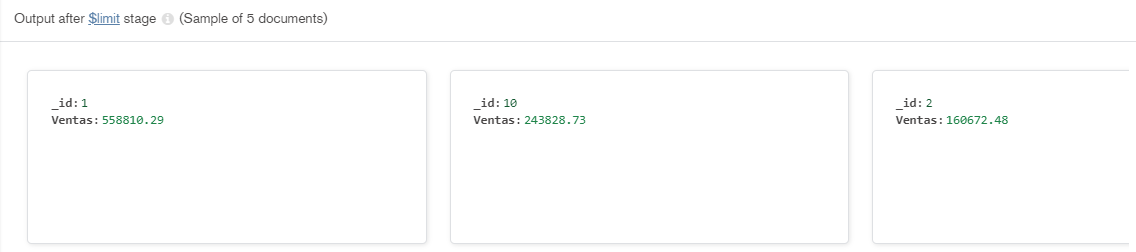
$sum: "$payment\_value"

}

}}, {$sort: {

Ventas: -1

}}, {$limit: 5}]



1. **¿Cuáles es el precio promedio de todos los productos?**

[{$group: {

\_id: null,

PrecioPromedio: {

$avg: "$price"

}

}}]



1. **¿Cuál es el precio de cada uno de los productos?**

[{$lookup: {

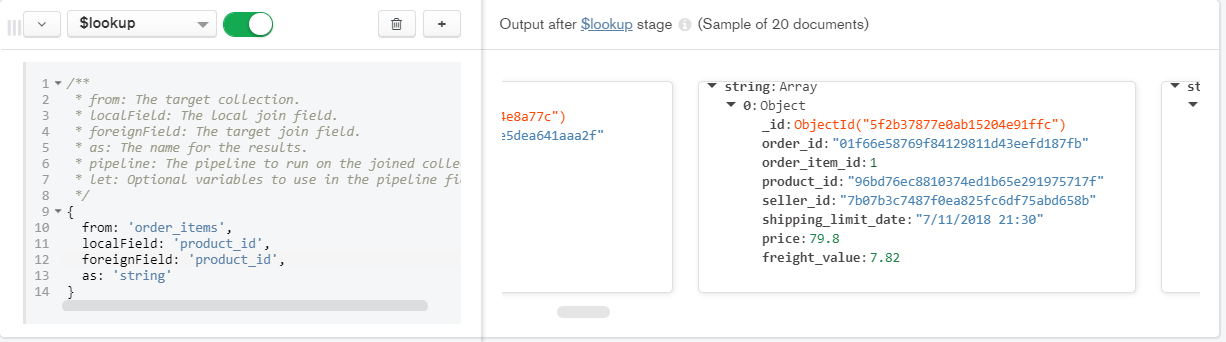
from: 'order\_items',

localField: 'product\_id',

foreignField: 'product\_id',

as: 'string'

}}]



1. **¿Qué órdenes tienen más de 3 items?**

[{$match: {

order\_item\_id: {$gte: 3}

}}, {$group: {

\_id: "$order\_id",

items: {

$sum: 1

}

}}]



1. **¿De qué ciudades son los clientes?**

[{$group: {

\_id: "$customer\_city",

Ciudades: {

$sum: 1

}

}}, {$lookup: {

from: 'Customers',

localField: '\_id',

foreignField: 'customer\_city',

as: 'Ciudad'

}}, {$addFields: {

Ciudad\_obj: { $arrayElemAt: ["$Ciudad",0]}

}}, {$addFields: {

Ciudad: "$Ciudad\_obj.customer\_city"

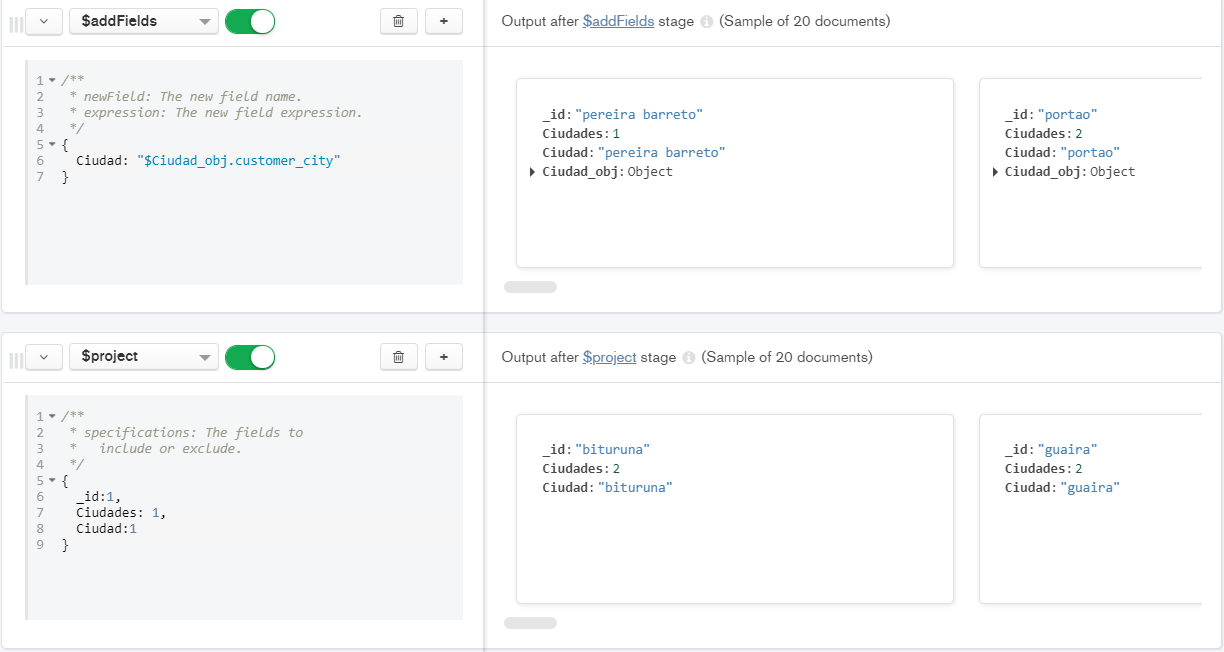
}}, {$project: {

\_id:1,

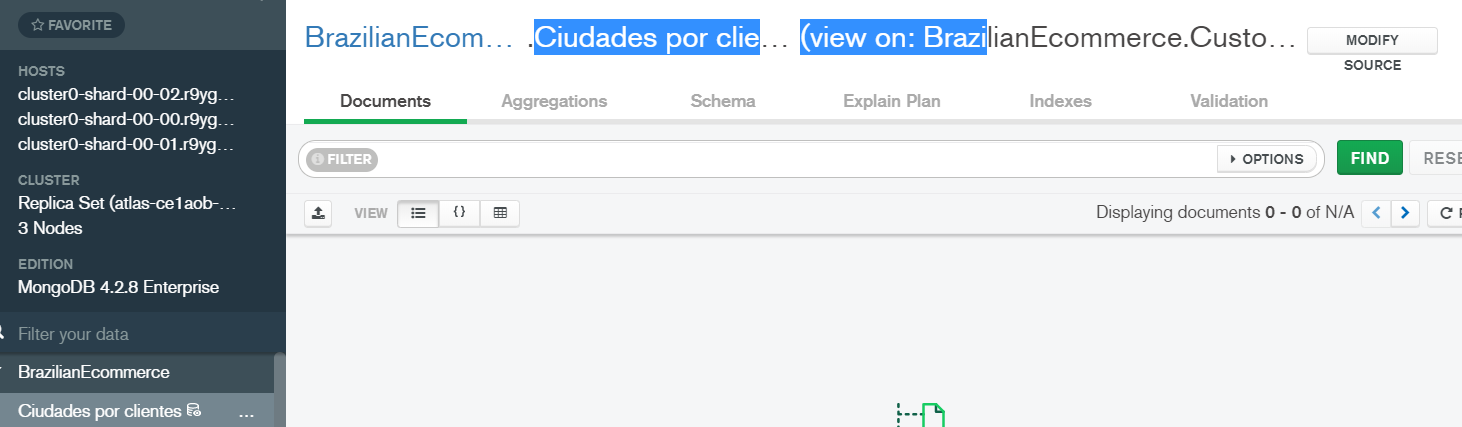
Ciudades: 1,

Ciudad:1

}}]



Creación de vista



**Conclusiones**

Conclusiones de proyecto

* Existen áreas de oportunidad en la calidad de los productos y logística de entrega de algunos vendedores.
* Se deberá estandarizar todas las descripciones y fotografías de los productos para mejorar la satisfacción del cliente.
* Hacer un análisis detallado de los vendedores y las categorías con baja calificación es esencial para cumplir el objetivo de satisfacción al cliente.

Conclusiones personales

* Me hubiera gustado enfocarme únicamente en los datos con calificación menor a 4 para sacar mejores conclusiones de cuáles eran las áreas de oportunidad.
* Me parece más sencilla la sintaxis de MySQL.
* Muy importante revisar que los datos estén bien clasificados en “string”, “double”, etc.