

Proyecto “Olist - Brazilian E-Commerce”

1. Definición del Proyecto (Fuentes y Problema a Resolver)

1. Fuentes de BD

La fuente de datos utilizada fue provista por Andrés Ramírez, el experto BEDU.

2. Problema a Resolver

Olist, una e-commerce brasileña, quisiera aumentar sus ventas, sobre todo en las categorías con menos productos comprados. Para ello, y como primera instancia, es necesario reconocer tales categorías. Posteriormente, y como parte de la propuesta a hacer, se ubicarán los meses con menos ventas. Una vez contando con dichos datos, se le hará la recomendación a la empresa para que ofrezca descuentos en esas categorías y en los meses identificados.

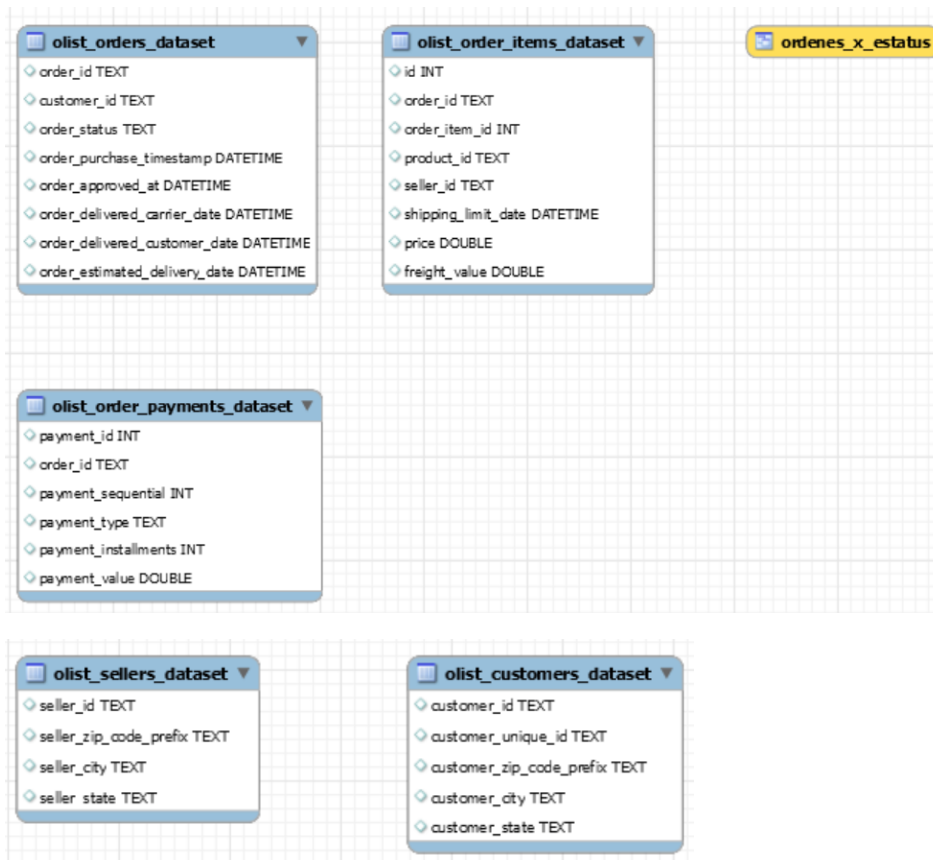
2. Definición de la Base de Datos

Diagrama Entidad Relación

Los datos provienen de 9 archivos CSV:

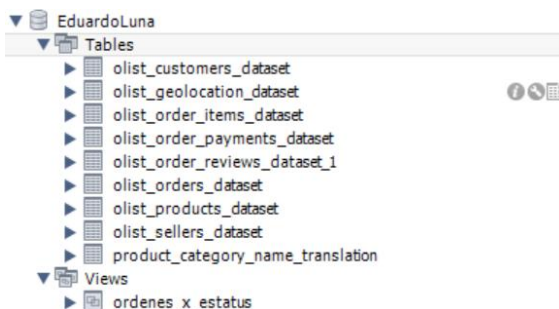
- olist_customers_dataset
- olist_geolocation_dataset
- olist_order_items_dataset
- olist_order_payments_dataset
- olist_order_reviews_dataset_1
- olist_orders_dataset
- olist_products_dataset
- olist_sellers_dataset_1
- product_category_name_translation_1





Creación de BBDD en MySQL

Las tuplas fueron reducidas a 2000, ya que MySQL toma demasiado tiempo para cargar la información. Adicional a esto, me di cuenta de que unos archivos no podían ser cargados, dado que algunos campos de fechas no aceptaban registros en blanco. En algunos casos, y como solución más rápida, decidí omitirlos.



Creación de BBDD en Mongo DB

A diferencia de MySQL, en Mongo todo fue mucho más rápido. Lo único necesario fue llevar a cabo la reducción de los datos a cargar, debido a que la versión gratuita tiene un cierto límite de memoria.

Collection Name ^	Documents	Avg. Document Size	Total Document Size	Num. Indexes	Total Index Size	Properties
olist_customers_dataset	2,000	218.1 B	426.0 KB	1	32.0 KB	
olist_geolocation_dataset	2,000	176.1 B	344.0 KB	1	32.0 KB	
olist_order_items_dataset	2,000	258.7 B	505.3 KB	1	32.0 KB	
olist_order_payments_dataset	2,000	191.2 B	373.4 KB	1	32.0 KB	
olist_order_reviews_dataset	2,000	294.4 B	575.1 KB	1	32.0 KB	
olist_orders_dataset	2,000	324.0 B	632.8 KB	1	32.0 KB	
olist_products_dataset	2,000	284.7 B	556.0 KB	1	32.0 KB	

3. Consultas

SQL

1. Cantidad de órdenes por ciudad

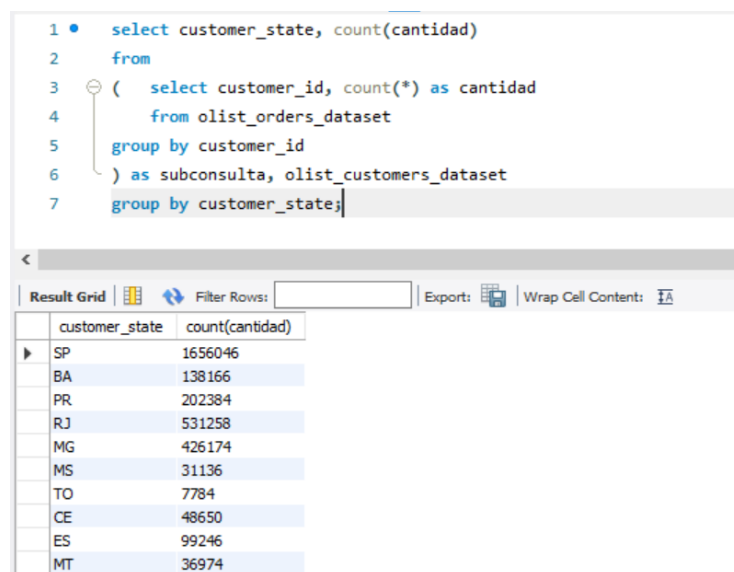
```
select customer_city, count(cantidad)
from
(
    select customer_id, count(*) as cantidad
    from olist_orders_dataset
    group by customer_id
) as subconsulta, olist_customers_dataset
group by customer_city;
```

1 •	select customer_city, count(cantidad)
2	from
3	(select customer_id, count(*) as cantidad
4	from olist_orders_dataset
5	group by customer_id
6) as subconsulta, olist_customers_dataset
7	group by customer_city;

customer_city	count(cantidad)
sao paulo	640234
alagoinhas	1946
londrina	5838
piracicaba	17514
nova iguacu	23352
conceicao dos ouros	1946
belo horizonte	110922
campo grande	9730
ananas	1946
itabira	9730

2. Cantidad de órdenes por estado

```
select customer_state, count(cantidad)
from
(
    select customer_id, count(*) as cantidad
    from olist_orders_dataset
group by customer_id
) as subconsulta, olist_customers_dataset
group by customer_state;
```



The screenshot shows a SQL query editor with a query window and a results window. The query window contains the following SQL code:

```
1 • select customer_state, count(cantidad)
2   from
3   ( select customer_id, count(*) as cantidad
4     from olist_orders_dataset
5     group by customer_id
6   ) as subconsulta, olist_customers_dataset
7   group by customer_state;
```

The results window displays a table with two columns: customer_state and count(cantidad). The table contains 10 rows of data, with the first row highlighted in blue.

customer_state	count(cantidad)
SP	1656046
BA	138166
PR	202384
RJ	531258
MG	426174
MS	31136
TO	7784
CE	48650
ES	99246
MT	36974

3. Cuáles son los 5 productos más y menos vendidos

```
select product_id, count(*) as productos
from olist_order_items_dataset
group by product_id
order by productos desc limit 5;
```

```
1 • select product_id, count(*) as productos
2   from olist_order_items_dataset
3  group by product_id
4  order by productos desc limit 5;
```

product_id	productos
422879e10f46682990de24d770e7f83d	11
e7cc48a9daff5436f63d3aad9426f28b	10
b532349fe46b38fbc7bb3914c1bdae07	8
5769ef0a239114ac3a854af00df129e4	8
aca2eb7d00ea1a7b8ebd4e68314663af	8

4. Cantidad de órdenes por método de pago

```
select payment_type, count(*)
from olist_order_payments_dataset
group by payment_type;
```

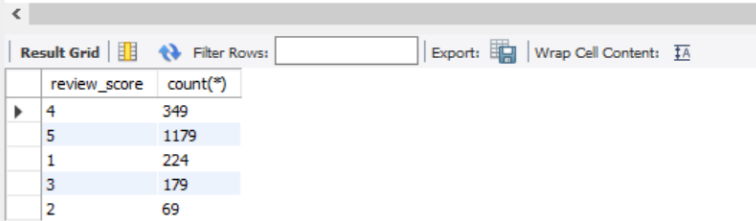
```
1 • select payment_type, count(*)
2   from olist_order_payments_dataset
3  group by payment_type;
```

payment_type	count(*)
credit_card	1493
boleto	353
voucher	117
debit_card	37

5. Cantidad de órdenes por nivel de satisfacción de los clientes

```
select review_score, count(*)
from olist_order_reviews_dataset_1
group by review_score;
```

```
1 • select review_score, count(*)
2   from olist_order_reviews_dataset_1
3   group by review_score;
```




review_score	count(*)
4	349
5	1179
1	224
3	179
2	69

6. Cantidad de comentarios recibidos

```
select count(*)
```

```
from olist_order_reviews_dataset_1;
```

```
1 • select count(*)
2   from olist_order_reviews_dataset_1;
```



count(*)
2000

7. Cantidad de órdenes por mes

```
select year(order_purchase_timestamp), month(order_purchase_timestamp), count(*)
```

```
from olist_orders_dataset
```

```
group by year(order_purchase_timestamp), month(order_purchase_timestamp);
```

```
1 • select year(order_purchase_timestamp), month(order_purchase_timestamp), count(*)
2   from olist_orders_dataset
3   group by year(order_purchase_timestamp), month(order_purchase_timestamp);
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: |

	year(order_purchase_timestamp)	month(order_purchase_timestamp)	count(*)
▶	2017	10	86
	2018	7	136
	2018	8	120
	2017	11	163
	2018	2	138
	2017	7	83

8. Cantidad de productos por categoría

```
select product_category_name, count(*)
from olist_products_dataset
group by product_category_name;
```

```
1 • select product_category_name, count(*)
2   from olist_products_dataset
3   group by product_category_name;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: |

	product_category_name	count(*)
▶	perfumaria	54
	artes	5
	esporte_lazer	175
	bebes	57
	utilidades_domesticas	135

9. Cantidad de órdenes por cliente

```
select customer_id, count(*) as ordenes
from olist_orders_dataset
group by customer_id;
```

```
1 • select customer_id, count(*) as ordenes
2   from olist_orders_dataset
3  group by customer_id;
```

customer_id	ordenes
9ef432eb6251297304e76186b10a928d	1
b0830fb4747a6c6d20dea0b8c802d7ef	1
41ce2a54c0b03bf3443c3d931a367089	1
f88197465ea7920adcdbec7375364d82	1
8ab97904e6daea8866dbdbc4fb7aad2c	1

10. ¿Cuáles productos pesan más del promedio?

```
select product_id, product_weight_g
from olist_products_dataset
where product_weight_g > (select avg(product_weight_g) from olist_products_dataset);
```

```
1 • select product_id, product_weight_g
2   from olist_products_dataset
3  where product_weight_g > (select avg(product_weight_g) from olist_products_dataset);
```

product_id	product_weight_g
732bd381ad09e530fe0a5f457d81becb	18350
03b63c5fc16691530586ae020c345514	7150
1c1890ba1779090cd54008a3c3302921	2550
f6574524b07d5bb753bab6f35abe2d06	3600
051b9ff13dd55c0a6655a15ff296f80d	3600

11. En una consulta, mostrar la cantidad de órdenes por estatus

```
create view ordenes_x_estatus as
(select order_status, count(*)
from olist_orders_dataset
group by order_status);
```



```
1 • create view ordenes_x_estatus as
2   (select order_status, count(*))
3   from olist_orders_dataset
4   group by order_status);
```

Output

#	Time	Action	Message
✓ 57	09:56:21	select customer_id, count(*) as ordenes from olist_orders_dataset group by custome...	1000 row(s) returned
✓ 58	09:57:31	select product_id, product_weight_g from olist_products_dataset where product_w...	416 row(s) returned
✓ 59	09:58:10	create view ordenes_x_estatus as (select order_status, count(*) from olist_orders_d...	0 row(s) affected

select * from EduardoLuna.ordenes_x_estatus;

```
1 • SELECT * FROM EduardoLuna.ordenes_x_estatus;
```

Result Grid

order_status	count(*)
delivered	1946

12. Cantidad de veces que cada cliente ha dado su opinión

```
select customer_id, count(review_id)
from olist_orders_dataset as o
join olist_order_reviews_dataset_1 as orv
on o.order_id = orv.order_id
group by customer_id;
```

```
1 • select customer_id, count(review_id)
2   from olist_orders_dataset as o
3  join olist_order_reviews_dataset_1 as orv
4   on o.order_id = orv.order_id
5  group by customer_id;
```

customer_id	count(review_id)
20b5aae6a3e31111009f9a7ecc31a232	1
ca08ac90cf5b8ff42d87a9d754c2374b	1
a260ca264a1787a1842ced2735bff664	1
bd13582471af02ce718568d9cf8bbd93	1
26b32bdec947c658c37dcfa2859bdf0	1
d29935fdaf4a76f34653da3def4f2a24	1

13. Los productos que ha comprado cada cliente

```
select customer_id, product_id
from olist_orders_dataset as o
left join olist_order_items_dataset as oi
on o.order_id = oi.order_id
where product_id is not null
order by customer_id;
```

```
1 • select customer_id, product_id
2   from olist_orders_dataset as o
3  left join olist_order_items_dataset as oi
4   on o.order_id = oi.order_id
5  where product_id is not null
6  order by customer_id;
```

customer_id	product_id
05428560b11124d146bc86a2c4d5bf35	59f17be62604541797d4df87ae80cf73
076dcea2eb6c8dbe9354f7c0d8670a00	8aa6223e400af9c97b07c75993142721
0c7a95280b3f80eadd81c829e093dbec	4cb33e9787e8a38b82bd68e70f5f4700
13a546131c3ba5360cefc5abbb428a01	3b0f7951038b105522c2d566b54421f7
16c0be52bf7bf9069c6fa0d850c660ac	115e606ac8059e4cfa1c6c3310fb365c

14. El total de lo que ha pagado cada cliente

```
select customer_id, sum(price)
from olist_orders_dataset as o
left join olist_order_items_dataset as oi
```

on o.order_id = oi.order_id

group by customer_id;

```
1 • select customer_id, sum(price)
2   from olist_orders_dataset as o
3  left join olist_order_items_dataset as oi
4    on o.order_id = oi.order_id
5  group by customer_id;
```

customer_id	sum(price)
57ee2ef64f17a5f9a4bf30489b06326c	42.9
2e068340b3037f667ce3224bd59260af	78.98
3c628393675b42c6b5ef89461f68ecef	359.8
076dcea2eb6c8dbe9354f7c0d8670a00	174.9
cf4604a56d4e374de013f1ec35280d00	102.9

15.Cantidad de productos vendidos por categoría

select product_category_name, count(o.product_id)

from olist_products_dataset as op

left join olist_order_items_dataset as o

on op.product_id = o.product_id

group by product_category_name;

```
1 • select product_category_name, count(o.product_id)
2   from olist_products_dataset as op
3  left join olist_order_items_dataset as o
4    on op.product_id = o.product_id
5  group by product_category_name;
```

product_category_name	count(o.product_id)
esporte_lazer	12
cool_stuff	10
moveis_decoracao	8
beleza_saude	8
utilidades_domesticas	13

JSON

1. Cantidad de órdenes por ciudad

[\$lookup: {

```
from: 'olist_customers_dataset',
localField: 'customer_id',
foreignField: 'customer_id',
as: 'clienteArr'
}}, {$addFields: {
  clienteObj: {
    $arrayElemAt: [ "$clienteArr", 0]
  }
}}, {$match: {
  clienteObj: { $exists: true }
}}, {$addFields: {
  ciudad: "$clienteObj.customer_city"
}}, {$group: {
  _id: "$ciudad",
  total: {
    $sum: 1
  }
}}}
```

Olist.olist_orders_dataset

DOCUMENTS 2.0k TOTAL SIZE 632.8KB AVG. SIZE 324B INDEXES 1 TOTAL SIZE 32.0KB AVG. SIZE 32.0KB

Documents Aggregations Schema Explain Plan Indexes Validation

COLLATION Untitled- Modified SAVE SAMPLE MODE AUTO PREVIEW

order_purchase_timestamp: 2018-08-08T13:58:00.000+00:00
order_approved_at: 2018-08-08T13:55:00.000+00:00
order_delivered_carrier_date: 2018-08-08T18:50:00.000+00:00
order_delivered_customer_date: 1970-01-01T00:00:00.000+00:00
order_approved_at: 2018-09-04T09:00:00.000+00:00

order_purchase_timestamp: 2018-09-04T09:00:00.000+00:00
order_approved_at: 2018-09-04T09:00:00.000+00:00
order_delivered_carrier_date: 2018-09-04T09:00:00.000+00:00
order_delivered_customer_date: 2018-09-04T09:00:00.000+00:00

\$group

Output after \$group stage (Sample of 20 documents)

```
1 - {
2   _id: '$ciudad',
3   total: {
4     $sum: 1
5   }
6 }
```

_id: "belo horizonte"
total: 2

_id: "sorocaba"
total: 1

ADD STAGE

2. Cantidad de órdenes por estado

```
{{$lookup: {  
  from: 'olist_customers_dataset',  
  localField: 'customer_id',  
  foreignField: 'customer_id',  
  as: 'clienteArr'  
}}, {$addFields: {  
  clienteObj: {  
    $arrayElemAt: [ "$clienteArr", 0]  
  }  
}}, {$match: {  
  clienteObj: { $exists: true }  
}}, {$addFields: {  
  estado: "$clienteObj.customer_state"  
}}, {$group: {  
  _id: "$estado",  
  total: {  
    $sum: 1  
  }  
}}}
```

Olist.olist_orders_dataset

DOCUMENTS 2.0k TOTAL SIZE 632.8KB AVG. SIZE 324B INDEXES 1 TOTAL SIZE 32.0KB AVG. SIZE 32.0KB

Documents Aggregations Schema Explain Plan Indexes Validation

COLLATION Untitled- Modified SAVE SAMPLE MODE AUTO PREVIEW

Output after \$group stage (Sample of 12 documents)

```
1 = {  
2   _id: '$estado',  
3   total: {  
4     $sum: 1  
5   }  
6 }
```

ADD STAGE

order_purchase_timestamp: 2018-08-08T13:30:00.000+00:00
order_approved_at: 2018-08-08T13:55:00.000+00:00
order_delivered_carrier_date: 2018-08-08T18:50:00.000+00:00
order_delivered_customer_date: 2018-01-01T00:00:00.000+00:00
order_estimated_delivery_date: 2018-04-00T00:00:00.000+00:00

order_purchase_timestamp: 2018-09-08T13:30:00.000+00:00
order_approved_at: 2018-09-08T13:55:00.000+00:00
order_delivered_carrier_date: 2018-09-08T18:50:00.000+00:00
order_delivered_customer_date: 2018-01-01T00:00:00.000+00:00
order_estimated_delivery_date: 2018-04-00T00:00:00.000+00:00

_id: "SP"
total: 19

_id: "SC"
total: 1

3. Cuáles son los 5 productos más y menos vendidos

```
[{$group: {  
  _id: "$product_id",  
  total: {  
    $sum: 1  
  }  
}}, {$sort: {  
  total: -1  
}}, {$limit: 5}]
```

Olist.olist_order_items_dataset

DOCUMENTS 2.0k TOTAL SIZE 505.3KB AVG. SIZE 259B INDEXES 1 TOTAL SIZE 32.0KB AVG. SIZE 32.0KB

Documents Aggregations Schema Explain Plan Indexes Validation

COLLATION Untitled SAVE SAMPLE MODE AUTO PREVIEW

1 5

Output after \$limit stage (Sample of 5 documents)

```
{  
  "_id": "422879e10f46682990de24d770e7f83d"  
  "total": 11  
}
```

```
{  
  "_id": "e7cc48a9daff5436f63d3"  
  "total": 10  
}
```

ADD STAGE

Olist.olist_order_items_dataset

DOCUMENTS 2.0k TOTAL SIZE 505.3KB AVG. SIZE 259B INDEXES 1 TOTAL SIZE 32.0KB AVG. SIZE 32.0KB

Documents Aggregations Schema Explain Plan Indexes Validation

COLLATION Untitled - Modified SAVE SAMPLE MODE AUTO PREVIEW

Output after \$limit stage (Sample of 5 documents)

```
1 5
```

```
{
  "_id": "77a399638eca91f6fa1a508a30e01565",
  "total": 1
}
```

```
{
  "_id": "909b87db6cb3a7ab26bd6",
  "total": 1
}
```

ADD STAGE

4. Cantidad de órdenes por método de pago

```
{{ $group: {
  _id: "$payment_type",
  total: {
    $sum: 1
  }
}}
```

Olist.olist_order_payments_dataset

DOCUMENTS 2.0k TOTAL SIZE 373.4KB AVG. SIZE 191B INDEXES 1 TOTAL SIZE 32.0KB AVG. SIZE 32.0KB

Documents Aggregations Schema Explain Plan Indexes Validation

COLLATION Untitled SAVE SAMPLE MODE AUTO PREVIEW

aggregation pipeline stages. [Learn more](#)

```
1 {{
2   _id: '$payment_type',
3   total: {
4     $sum: 1
5   }
6 }}
```

Output after \$group stage (Sample of 4 documents)

```
{
  "payment_type": "credit_card",
  "payment_installments": 8,
  "payment_value": 99.33
}
```

```
{
  "payment_type": "credit_card",
  "payment_installments": 1,
  "payment_value": 24.39
}
```

```
{
  "_id": "voucher",
  "total": 117
}
```

```
{
  "_id": "boleto",
  "total": 353
}
```

ADD STAGE

5. Cantidad de órdenes por nivel de satisfacción de los clientes

```
[{$group: {  
  _id: "$review_score",  
  total: {  
    $sum: 1  
  }  
}}]
```

Olist.olist_order_reviews_dataset

DOCUMENTS 2.0k TOTAL SIZE 575.1KB AVG. SIZE 294B INDEXES 1 TOTAL SIZE 32.0KB AVG. SIZE 32.0KB

Documents Aggregations Schema Explain Plan Indexes Validation

aggregation pipeline stages. [Learn more](#)

review_comment_message: ""
review_creation_date: 1970-01-01T00:00:00.000+00:00
review_answer_timestamp: 1970-01-01T00:00:00.000+00:00

review_score: 5
review_comment_title: ""
review_comment_message: ""
review_creation_date: 2018-

Output after \$group stage (Sample of 5 documents)

_id: 4
total: 349

_id: 3
total: 179

1 ▾
2 {
3 _id: '\$review_score',
4 total: {
5 \$sum: 1
6 }
7 }

ADD STAGE

6. Cantidad de comentarios recibidos

```
[{$match: {  
  review_comment_message: {  
    $ne: ""  
  }  
}}, {$group: {  
  _id: null,  
  total: {  
    $sum: 1  
  }  
}]
```


}}

Olist.olist_order_reviews_dataset

DOCUMENTS 2.0k TOTAL SIZE 575.1KB AVG. SIZE 294B INDEXES 1 TOTAL SIZE 32.0KB AVG. SIZE 32.0KB

Documents Aggregations Schema Explain Plan Indexes Validation

COLLATION Untitled- Modified SAVE SAMPLE MODE AUTO PREVIEW

review_score: 3
review_comment_title: ""
review_comment_message: "Recebi bem antes do prazo esti
review_creation_date: 1970-01-01T00:00:00.000+00:00
review_creation_date: 1970-01-01T00:00:00.000+00:00
review_creation_date: 1970-01-01T00:00:00.000+00:00

review_score: 3
review_comment_title: ""
review_comment_message: "Pe
review_creation_date: 1970-01-01T00:00:00.000+00:00

\$group

Output after \$group stage (Sample of 1 document)

```
1 {  
2   _id: null,  
3   total: {  
4     $sum: 1  
5   }  
6 }
```

ADD STAGE

_id: null
total: 867

7. Cantidad de órdenes por mes

```
{{$group: {  
  _id: {  
    year: {  
      $year: '$order_purchase_timestamp'  
    },  
    month: {  
      $month: '$order_purchase_timestamp'  
    }  
  },  
  total: {  
    $sum: 1  
  }  
}}, {$project: {  
  year: "$_id.year",  
  month: "$_id.month",  
  total: 1
```

}}

Olist.olist_orders_dataset

DOCUMENTS 2.0kTOTAL SIZE 632.8KB AVG. SIZE 324BINDEXES 1TOTAL SIZE 32.0KB AVG. SIZE 32.0KB

DocumentsAggregationsSchemaExplain PlanIndexesValidation

COLLATION

Untitled- Modified

SAVE

SAMPLE MODE

AUTO PREVIEW

8

9

10

11

12

13

},

total: {

sum: 1

}

\$project

Output after \$project stage (Sample of 20 documents)

_id: Object

total: 29

year: 2017

month: 6

_id: Object

total: 1

year: 2016

month: 7

ADD STAGE

8. Cantidad de productos por categoría

```
[{$group: {
  _id: "$product_category_name",
  total: {
    $sum: 1
  }
}}
```

Olist.olist_products_dataset

DOCUMENTS 2.0k TOTAL SIZE 556.0KB AVG. SIZE 285B INDEXES 1 TOTAL SIZE 32.0KB AVG. SIZE 32.0KB

Documents Aggregations Schema Explain Plan Indexes Validation

aggregation pipeline stages: [Learn more](#)

product_description_lenght: 28/
product_photos_qty: 1
product_weight_g: 225
product_length_cm: 16

product_description_lenght
product_photos_qty: 1
product_weight_g: 1000
product_length_cm: 30

\$group

```
1- {  
2  _id: '$product_category_name',  
3  total: {  
4    $sum: 1  
5  }  
6 }
```

Output after \$group stage (Sample of 20 documents)

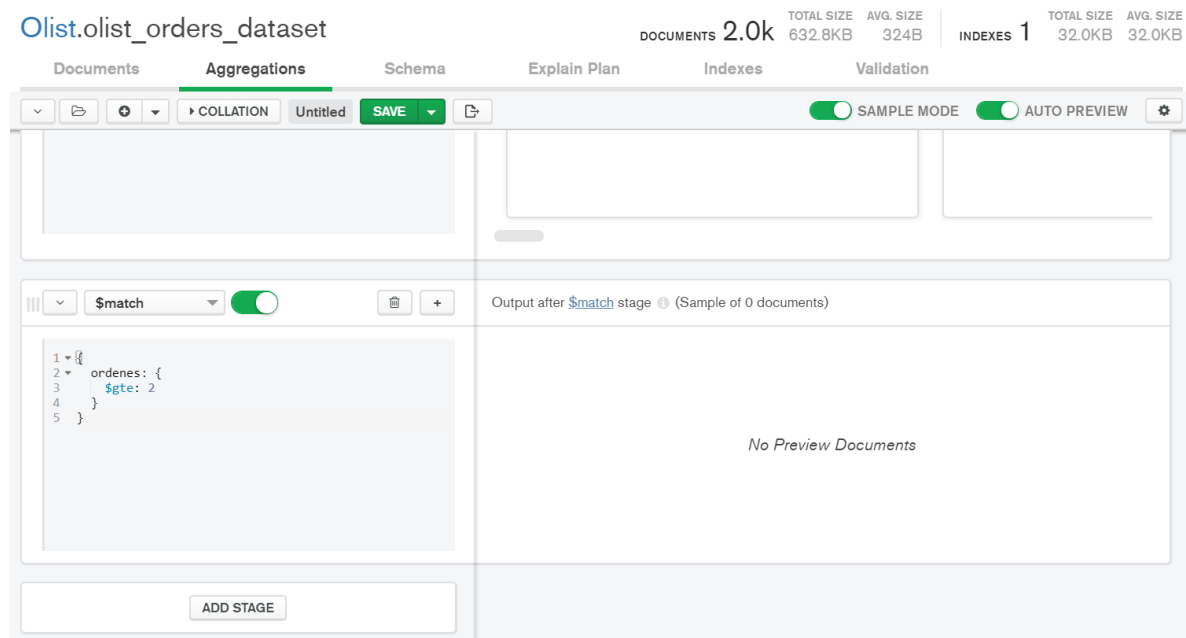
_id: "market_place"
total: 8

_id: "fashion_calçados"
total: 12

ADD STAGE

9. ¿Cuántos clientes han llevado a cabo 2 o más órdenes?

```
[{$group: {  
  _id: "$customer_id",  
  ordenes: {  
    $sum: 1  
  }  
}}, {$match: {  
  ordenes: {  
    $gte: 2  
  }  
}}]
```



10. ¿Cuáles productos pesan más del promedio?

```
{
  $group: {
    _id: null,
    items: {
      $sum: 1
    },
    total: {
      $sum: '$product_weight_g'
    }
  },
  $addFields: {
    pesoPromedio: {
      $divide: [
        '$total',
        '$items'
      ]
    }
  }
}
```

Olist.olist_products_dataset

DOCUMENTS 2.0k TOTAL SIZE 556.0KB AVG. SIZE 285B INDEXES 1 TOTAL SIZE 32.0KB AVG. SIZE 32.0KB

Documents Aggregations Schema Explain Plan Indexes Validation

COLLATION Untitled- Modified SAVE SAMPLE MODE AUTO PREVIEW

Output after \$addFields stage (Sample of 1 document)

```
1 {
2   pesoPromedio: {
3     $divide: [
4       '$total',
5       '$items'
6     ]
7   }
8 }
```

```
{
  "_id": null,
  "items": 2000,
  "total": 4412047,
  "pesoPromedio": 2206.0235
}
```

ADD STAGE

```
[{$match: {
  product_weight_g: {$gte: 2206.0235}
}}]
```

Olist.olist_products_dataset

DOCUMENTS 2.0k TOTAL SIZE 556.0KB AVG. SIZE 285B INDEXES 1 TOTAL SIZE 32.0KB AVG. SIZE 32.0KB

Documents Aggregations Schema Explain Plan Indexes Validation

COLLATION Untitled- Modified SAVE SAMPLE MODE AUTO PREVIEW

aggregation pipeline stages. [Learn more](#)

Output after \$match stage (Sample of 20 documents)

```
1 {
2   product_weight_g: {
3     $gte: 2206.0235
4   }
5 }
```

```
{
  "_id": ObjectId("5fc3b7e3c6a2b4731848fb4e"),
  "product_id": "732bd381ad09e530fe0a5f457d81becb",
  "product_category_name": "cool_stuff",
  "product_name_length": 56,
  "product_description_length": 1272,
  "product_photos_qty": 4,
  "product_weight_g": 18350,
  "product_length_cm": 70
}
```

```
{
  "_id": ObjectId("5fc3b7e3c6a2b4731848fb4e"),
  "product_id": "03b63c5fc16691",
  "product_category_name": "bel",
  "product_name_length": 49,
  "product_description_length": 1272,
  "product_photos_qty": 4,
  "product_weight_g": 7150,
  "product_length_cm": 50
}
```

ADD STAGE

11. Cantidad de órdenes por estatus

```
[{$group: {
```

```
_id: "$order_status",

total: {

  $sum: 1

}

}}
```

Olist.olist_orders_dataset

DOCUMENTS 2.0k TOTAL SIZE 632.8KB AVG. SIZE 324B INDEXES 1 TOTAL SIZE 32.0KB AVG. SIZE 32.0KB

Documents Aggregations Schema Explain Plan Indexes Validation

aggregation pipeline stages. [Learn more](#)

order_purchase_timestamp: 2017-02-10T16:56:00.000+00:00
order_approved_at: 2017-02-10T17:07:00.000+00:00
order_delivered_carrier_date: 2017-04-11T00:55:00.000+00:00
order_delivered_customer_date: 2017-10-11T02:25:00.000+00:00

order_purchase_timestamp: 1970-01-01T00:00:00.000+00:00
order_approved_at: 1970-01-01T00:00:00.000+00:00
order_delivered_carrier_date: 1970-01-01T00:00:00.000+00:00
order_delivered_customer_date: 1970-01-01T00:00:00.000+00:00

Output after \$group stage (Sample of 6 documents)

```
1 = {
2   _id: '$order_status',
3   total: {
4     $sum: 1
5   }
6 }
```

ADD STAGE

total: 1946

total: 9

12. Los clientes que han dado su opinión

```
[{$lookup: {
  from: 'olist_order_reviews_dataset',
  localField: 'order_id',
  foreignField: 'order_id',
  as: 'reviewArr'
}}, {$addFields: {
  reviewObj: {
    $arrayElemAt: [
      '$reviewArr',
      0
    ]
  }
}
```

```
}}, {$match: {
  reviewObj: {
    $exists: true
  }
}}, {$addField: {
  opinion: "$reviewObj.review_comment_message"
}}, {$match: {
  opinion: {$ne: ""}
}}, {$project: {
  customer_id: 1,
  opinion: 1
}}
```

Olist.olist_orders_dataset

DOCUMENTS 2.0k TOTAL SIZE 632.8KB AVG. SIZE 324B INDEXES 1 TOTAL SIZE 32.0KB AVG. SIZE 32.0KB

Documents Aggregations Schema Explain Plan Indexes Validation

COLLATION Untitled - Modified SAVE SAMPLE MODE AUTO PREVIEW

order_purchase_timestamp: 2018-01-03T20:14:00.000+00:00
order_approved_at: 2018-01-03T21:10:00.000+00:00
order_delivered_carrier_date: 2018-02-04T03:09:00.000+00:00
order_delivered_customer_date: 2018-12-04T05:36:00.000+00:00

order_purchase_timestamp: 2017-09-03T12:14:00.000+00:00
order_approved_at: 2017-09-03T12:14:00.000+00:00
order_delivered_carrier_date: 2017-09-03T12:14:00.000+00:00
order_delivered_customer_date: 2017-09-03T12:14:00.000+00:00

\$project

1 {
2 customer_id: 1,
3 opinion: 1
4 }

Output after \$project stage (Sample of 12 documents)

_id: ObjectId("5fc3b79cc6a2b4731848f386")
customer_id: "3df704f53d3f1d4818840b34ec672a9f"
opinion: "Gostei do produto"

_id: ObjectId("5fc3b79cc6a2b4731848f386")
customer_id: "05e996469a2bf"
opinion: "Muito bom o produto"

ADD STAGE

13. ¿Cuántos productos ha comprado cada cliente?

```
[{$lookup: {
  from: 'olist_order_items_dataset',
  localField: 'order_id',
  foreignField: 'order_id',
  as: 'itemArr'
}]
```

```
}}, {$addFields: {  
  itemObj: {  
    $arrayElemAt: [  
      '$itemArr',  
      0  
    ]  
  }  
}}, {$match: {  
  itemObj: {  
    $exists: true  
  }  
}}, {$addFields: {  
  producto: '$itemObj.product_id'  
}}, {$group: {  
  _id: "$customer_id",  
  total: {  
    $sum: 1  
  }  
}}}
```

Olist.olist_orders_dataset

DOCUMENTS 2.0k TOTAL SIZE 632.8KB AVG. SIZE 324B INDEXES 1 TOTAL SIZE 32.0KB AVG. SIZE 32.0KB

Documents Aggregations Schema Explain Plan Indexes Validation

COLLATION Untitled - Modified SAVE SAMPLE MODE AUTO PREVIEW

Output after \$group stage (Sample of 20 documents)

```
1 {  
2   _id: '$customer_id',  
3   total: {  
4     $sum: 1  
5   }  
6 }
```

ADD STAGE

order_purchase_timestamp: 1970-01-01T00:00:00.000+00:00
order_approved_at: 1970-01-01T00:00:00.000+00:00
order_delivered_carrier_date: 1970-01-01T00:00:00.000+00:00
order_delivered_customer_date: 2018-07-02T19:59:00.000+00:00
order_estimated_delivery_date: 1970-01-01T00:00:00.000+00:00

_id: "076dcea2eb6c8dbe9354f7c0d8670a00"
total: 1

_id: "625784c6aad212dc7d245"
total: 1

14. El total de lo que ha pagado cada cliente

```
[{$lookup: {
  from: 'olist_order_items_dataset',
  localField: 'order_id',
  foreignField: 'order_id',
  as: 'itemArr'
}}, {$addFields: {
  itemObj: {
    $arrayElemAt: [ "$itemArr", 0]
  }
}}, {$match: {
  itemObj: { $exists: true }
}}, {$addFields: {
  precio: "$itemObj.price"
}}, {$group: {
  _id: "$customer_id",
  total: {
    $sum: "$precio"
  }
}}]
```

Olist.olist_orders_dataset

DOCUMENTS 2.0k TOTAL SIZE 632.8KB AVG. SIZE 324B INDEXES 1 TOTAL SIZE 32.0KB AVG. SIZE 32.0KB

Documents Aggregations Schema Explain Plan Indexes Validation

COLLATION Untitled- Modified SAVE SAMPLE MODE AUTO PREVIEW

Output after \$group stage (Sample of 20 documents)

```

1 {
2   _id: '$customer_id',
3   total: {
4     $sum: '$precio'
5   }
6 }

```

ADD STAGE

order_purchase_timestamp: 1970-01-01T00:00:00.000+00:00
order_approved_at: 1970-01-01T00:00:00.000+00:00
order_delivered_carrier_date: 1970-01-01T00:00:00.000+00:00
order_delivered_customer_date: 2018-07-02T19:59:00.000+00:00
order_estimated_delivery_date: 2070-01-01T00:00:00.000+00:00

_id: "3c628393675b42c6b5ef89461f68ecef"
total: 179.9

_id: "748632357c64f888b1f65"
total: 89.99

15.Cantidad de productos vendidos por categoría

```

{{ $lookup: {
  from: 'olist_products_dataset',
  localField: 'product_id',
  foreignField: 'product_id',
  as: 'productoArr'
}}, { $addFields: {
  productoObj: {
    $arrayElemAt: [ "$productoArr", 0 ]
  }
}}, { $match: {
  productoObj: { $exists: true }
}}, { $addFields: {
  categoria: "$productoObj.product_category_name"
}}, { $group: {
  _id: "$categoria",
  total: {
    $sum: 1

```

```
}  
}}
```

Olist.olist_order_items_dataset

DOCUMENTS 2.0k TOTAL SIZE 505.3KB AVG. SIZE 259B INDEXES 1 TOTAL SIZE 32.0KB AVG. SIZE 32.0KB

Documents Aggregations Schema Explain Plan Indexes Validation

COLLATION Untitled SAVE SAMPLE MODE AUTO PREVIEW

product_id: "8ca0e808c79136713e0076d30c007413"
seller_id: "530ec6109d11eaf87999465c6afee01"
shipping_limit_date: 1970-01-01T00:00:00.000+00:00
price: 99.9
freight_value: 37.65

product_id: "0369308a0800c"
seller_id: "218d46b86c1881c"
shipping_limit_date: 1970-01-01T00:00:00.000+00:00
price: 99
freight_value: 12.74

\$group

Output after \$group stage (Sample of 20 documents)

```
1 {  
2   _id: '$categoria',  
3   total: {  
4     $sum: 1  
5   }  
6 }
```

_id: "construcao_ferramentas_construcao"
total: 4

_id: "perfumaria"
total: 4

ADD STAGE

4. Vistas Relevantes

SQL

```
create view productos_x_categoria as  
(select product_category_name, count(o.product_id) as total  
from olist_products_dataset as op  
left join olist_order_items_dataset as o  
on op.product_id = o.product_id  
group by product_category_name  
order by total);
```

```

1 • create view productos_x_categoria as
2 (select product_category_name, count(o.product_id) as total
3 from olist_products_dataset as op
4 left join olist_order_items_dataset as o
5 on op.product_id = o.product_id
6 group by product_category_name
7 order by total);
8
9

```

Output

#	Time	Action	Message
1	13:21:43	use EduardoLuna	0 row(s) affected
2	13:22:17	select * from order_items_dataset LIMIT 0, 1000	Error Code: 1146. Table 'EduardoLuna.order_items_dataset' do...
3	13:22:30	select * from olist_order_items_dataset LIMIT 0, 1000	1000 row(s) returned
4	13:23:11	select * from olist_products_dataset LIMIT 0, 1000	1000 row(s) returned
5	13:24:39	select product_category_name, count(o.product_id) as total fro...	64 row(s) returned
6	13:27:15	create view productos_x_categoria as (select product_category...	0 row(s) affected

select * from EduardoLuna.productos_x_categoria;

```

1 • select * from EduardoLuna.productos_x_categoria;

```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: |

product_category_name	total
construcao_ferramentas_jardim	0
eletroportateis	0
agro_industria_e_comercio	0
sinalizacao_e_seguranca	0
climatizacao	0

JSON

```

[{$group: {
  _id: {
    month: {
      $month: '$order_purchase_timestamp'
    }
  },
  total: {
    $sum: 1
  }
}}, {$project: {

```

```
month: '$_id.month',  
total: 1  
}}, {$sort: {  
total: 1  
}}]
```

Olist.olist_orders_dataset

DOCUMENTS 2.0k TOTAL SIZE 632.8KB AVG. SIZE 324B INDEXES 1 TOTAL SIZE 32.0KB AVG. SIZE 32.0KB

Documents Aggregations Schema Explain Plan Indexes Validation

COLLATION Untitled- Modified SAVE SAMPLE MODE AUTO PREVIEW

Output after \$sort stage (Sample of 12 documents)

```
1 {  
  total: 1  
}
```

ADD STAGE

\$_id: Object
total: 54
month: 10

\$_id: Object
total: 58
month: 3

Olist.olist_orders_dataset

DOCUMENTS 2.0k TOTAL SIZE 632.8KB AVG. SIZE 324B INDEXES 1 TOTAL SIZE 32.0KB AVG. SIZE 32.0KB

Documents Aggregations Schema Explain Plan Indexes Validation

Saved Pipelines

ventas_x_mes

Output after \$sort stage (Sample of 12 documents)

\$_id: Object
total: 54
month: 10

\$_id: Object
total: 58
month: 3

5. Conclusiones

1. Antes que nada, cabe mencionar que los datos, en realidad, son una muestra de un total mucho más grande; no al punto para ser considerado Big Data, pero sí lo suficientemente grande como

para haber requerido ser reducido, al no contar con tiempo ni con memoria suficiente en el servidor. Como consecuencia de esto, las conclusiones no son del todo veraces.

2. Como se puede ver en la vista hecha en MySQL, son muchas las categorías que no tienen ninguna sola venta. De ahí que se recomiende poner descuentos que hagan más atractivas las secciones de construcción y herramientas de jardín, agroindustria y comercio, climatización, entre otras.
3. Tal y como se observa en la vista en MongoDB Compass, los meses con menos ventas son octubre, marzo y julio, en ese orden. Al igual que en el punto anterior, se aconseja poner descuentos atractivos que induzcan las ventas en dichos meses.
4. Otra medida para aumentar las ventas podría ser la búsqueda de costos de flete más bajos; esto con el fin de propiciar mayor cantidad de órdenes provenientes de ciudades más lejanas y con pocos clientes. Una manera para lograr esto es a través del establecimiento de una compra mínima que elimine los costos de envío.