



## **Proyecto “Olist”**

Uso de Mongo DB Compass

Módulo 1: Introducción a bases de datos

Luis Guillermo Jiménez Bojórquez

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

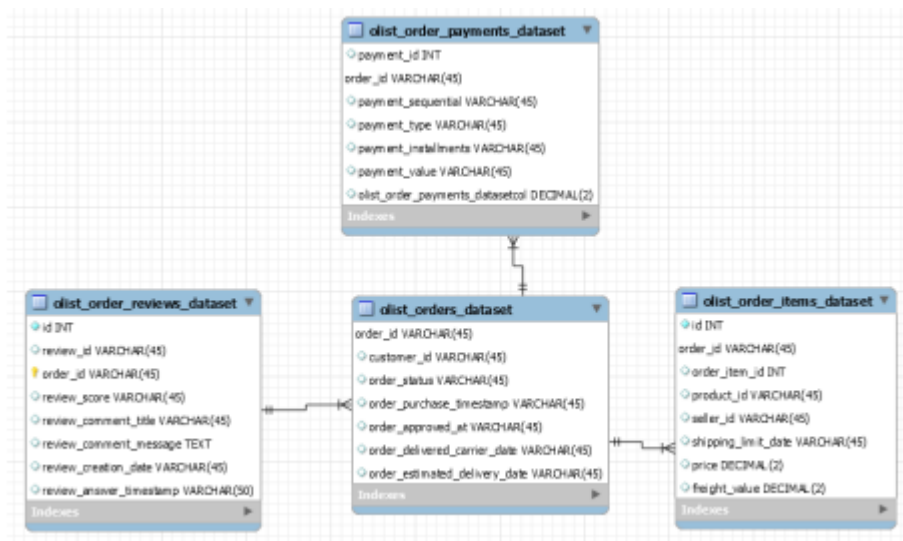
### a. Problema a resolver:

Se quiere conocer los principales KPI's con respecto a ventas esto con el propósito de poder elaborar la estrategia que se utilizará el próximo año. Le gustaría tomar en cuenta el performance de los principales vendedores y el volumen de venta de los productos que tienen.

### b. Fuentes de la BD: La información ha sido por el experto Andrés Ramírez del módulo 1: Introducción a bases de datos del bootcamp.

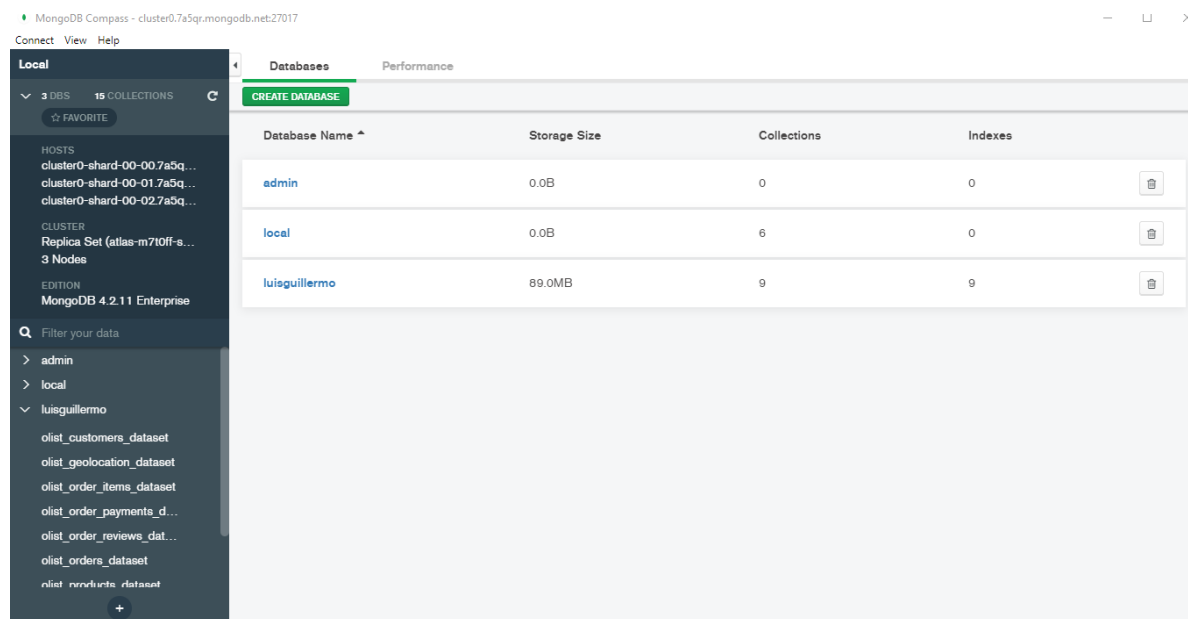
## 2. Definición de la base de datos:

- 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 Mongo DB Compass:

Cargar las bases de datos en CSV a MongoDB es muy rápido y fácil solo es necesario definir muy bien desde un principio el tipo de datos que tendrá cada columna.



### 3. Consultas en MongoDB Compass:

- a. ¿Cuál es la cantidad de órdenes por ciudad?

```
{{ $lookup: {
```

Eduardo Luna Álvarez - #data-analysis-gdl-20-07

```
  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
}
}}

```



b. ¿Cuál es la 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
  }
}
}}

```



c. ¿Cuáles son el top 5 productos más y menos vendidos?

```

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

```



d. ¿Cuál es la cantidad de órdenes por método de pago?

```

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

```



e. ¿Cuál es cantidad de órdenes por nivel de satisfacción de los clientes?

```

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

```



f. ¿Cuál es la cantidad de comentarios recibidos?

```

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

```

```
}}
```



g. ¿Cuál es la 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
}}]

```

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

```
1 * {
2   year: '$_id.year',
3   month: '$_id.month',
4   total: 1
5 }
```

```
{ "_id": "Object", "total": 29, "year": 2017, "month": 6 }
{ "_id": "Object", "total": 1, "year": 2016, "month": 7 }
```

h. ¿Cuál es la cantidad de productos por categoría?

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

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

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

```
{ "_id": "market_place", "total": 8 }
{ "_id": "fashion_calçados", "total": 12 }
```

i. ¿Cuáles productos que pesan más del promedio?

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



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

```

```

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

```

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_lenght: 56
product_description_lenght: 1272
product_photos_qty: 4
product_weight_g: 18350
product_length_cm: 70

```

```

_id: ObjectId("5fc3b7e3c6a2b4731848fb4e")
product_id: "03b63c5fc16691"
product_category_name: "bet"
product_name_lenght: 49
product_description_lenght: 1272
product_photos_qty: 4
product_weight_g: 7150
product_length_cm: 50

```

i. ¿Cuál es la cantidad de órdenes por estatus?

```

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

```

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

```

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

```

```

_id: "delivered"
total: 1946

```

```

_id: "canceled"
total: 9

```

j. ¿Cuáles son los clientes que han dado su opinión?

```

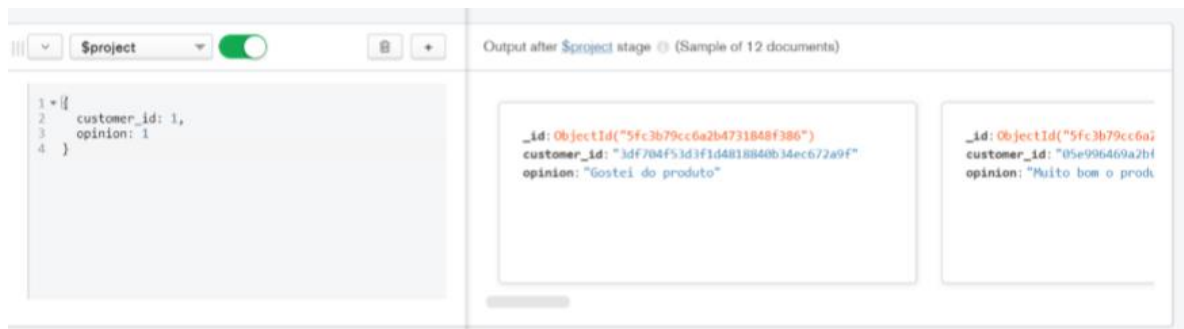
[{$lookup: {

```

```

from: 'olist_order_reviews_dataset',
localField: 'order_id',
foreignField: 'order_id',
as: 'reviewArr'
}}, {$addFields: {
  reviewObj: {
    $arrayElemAt: [
      '$reviewArr',
      0
    ]
  }
}}, {$match: {
  reviewObj: {
    $exists: true
  }
}}, {$addFields: {
  opinion: "$reviewObj.review_comment_message"
}}, {$match: {
  opinion: {$ne: ""}
}}, {$project: {
  customer_id: 1,
  opinion: 1
}}

```



k. ¿Cuántos productos ha comprado cada cliente?

```

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

```

Eduardo Luna Álvarez - #data-analysis-gdl-20-07

```

}}, {$addFields: {
  itemObj: {
    $arrayElemAt: [
      '$itemArr',
      0
    ]
  }
}}

```

```

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

```



- I. ¿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"
  }
}}

```

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

```

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

```

```

_id: "3c628393675b42c6b5ef89461f68ecef"
total: 179.9

```

```

_id: "748632357c64f888b1f65"
total: 89.99

```

j. ¿Cuál es la 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
  }
}
}]

```

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

```

## Vistas

```

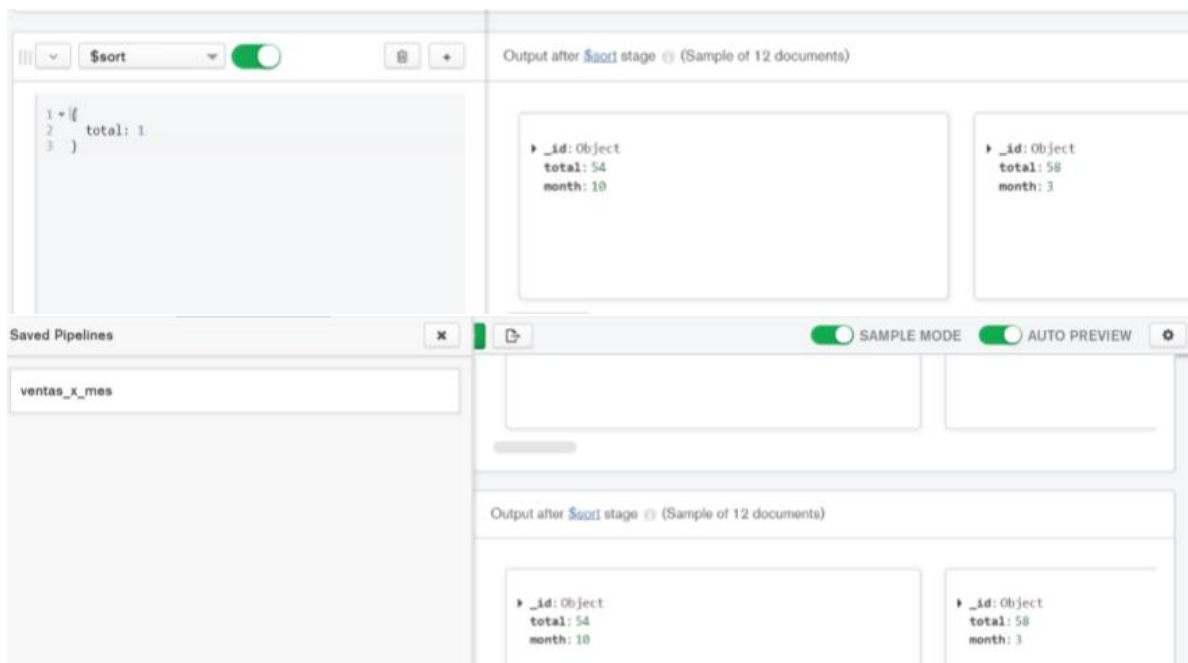
[{$group: {

```

```

_id: {
  month: {
    $month: '$order_purchase_timestamp'
  }
},
total: {
  $sum: 1
}
}}, {$project: {
  month: '$_id.month',
  total: 1
}}, {$sort: {
  total: 1
}}

```



#### 4. Conclusiones

1. La importación de a MongoDB al ser una base de datos no relacional es mucho más rápida que en una que si es relacional y trabajar con grandes cantidades de datos de hace mucho más fácil.
2. Es recomendable incrementar la publicidad sobre los productos dedicados a la industria agraria ya que de esta forma podrán incrementar sus ventas. Esto se pudo buscar intentando tener una mucho mejor calidad de leads por parte del equipo de marketing.
3. los meses con menos ventas es recomendable hacer campañas con descuentos, considerar el no cobrar los envíos de los productos o poner promociones especiales por referidos.

