



OLIST ECOMMERCE-BRASIL

Análisis ficticio para módulo BDD – Bedu

Santiago Cortina



ÍNDICE

- **1. Objetivo**
- **2. Contexto**
- **3. Análisis**
- **4. Conclusiones**

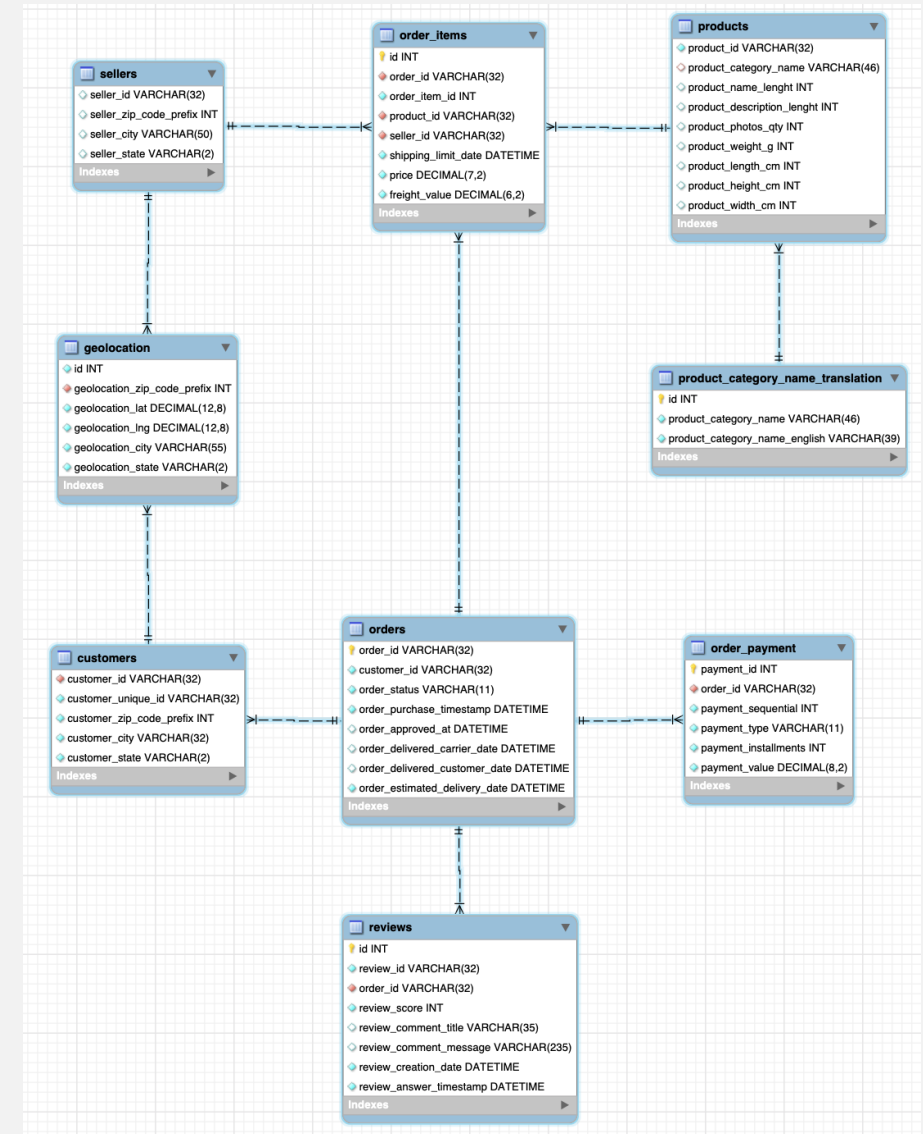
I. OBJETIVO

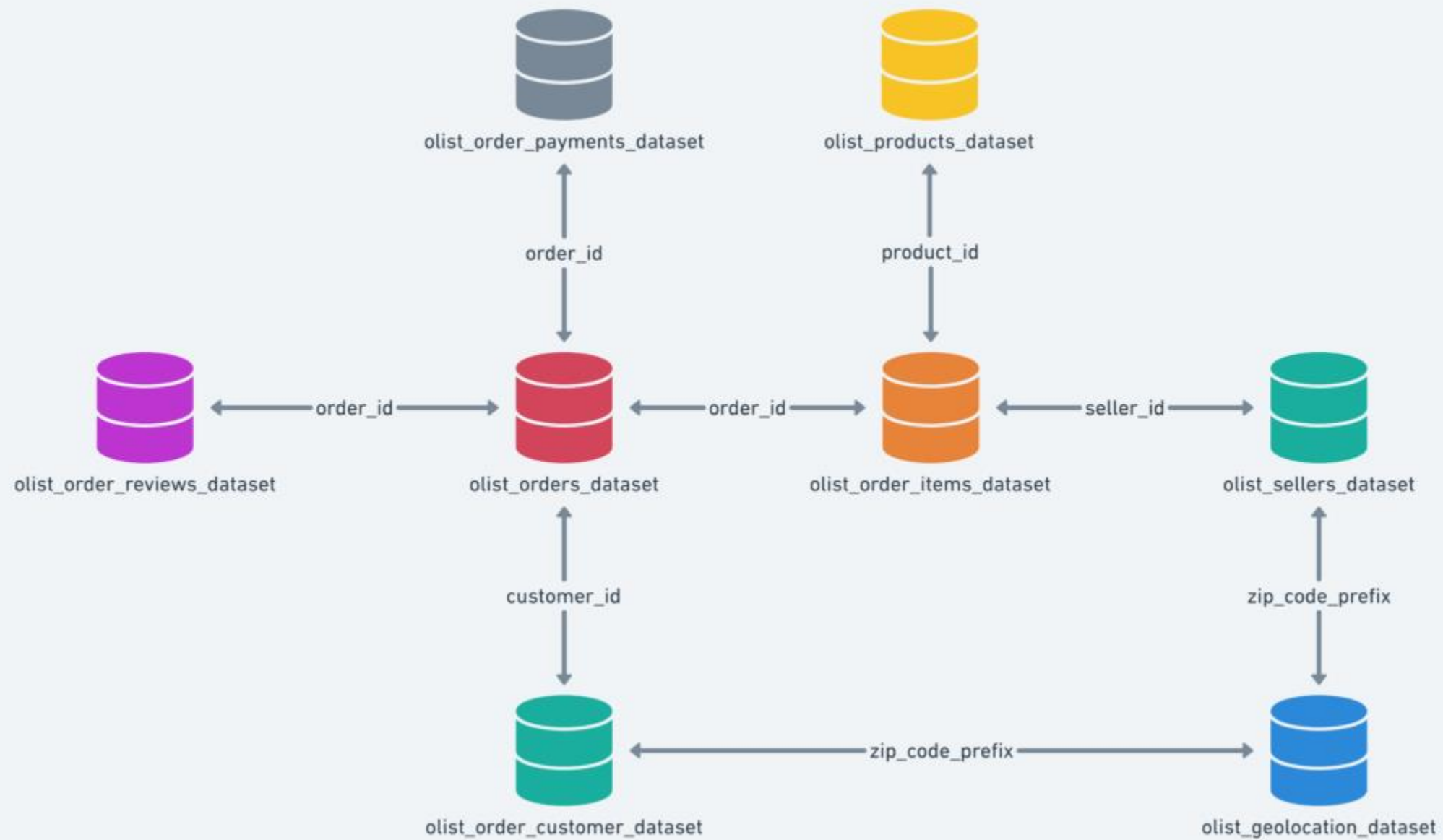
El objetivo del presente trabajo es familiarizarnos con los datos y posibles métricas de un Marketplace a través del uso de MySQL y MongoDB.

2. CONTEXTO

DATOS

- Se usó un Dataset gratuito de Olist, un Ecommerce de Brasil con la siguiente estructura de datos:
 - Información de 2016 a 2018
 - 9 tablas
 - Ordenes
 - Reseñas
 - Pago de orden
 - Clientes
 - Geolocalización
 - Vendedores
 - Ítems de orden
 - Productos
 - Traducción de producto
- Para la realización de tablas y gráficos, se generaron consultas y exportaron los resultados como CSV para manipular.

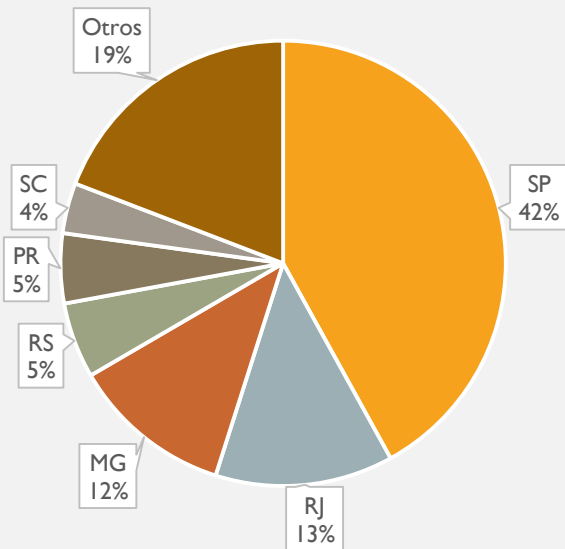
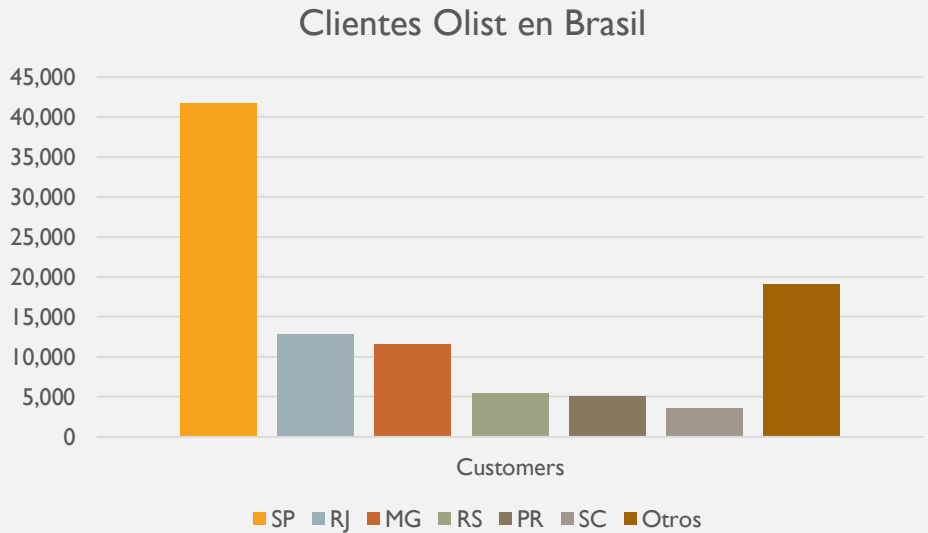




3. ANÁLISIS

ANÁLISIS GENERAL DE CLIENTES

- Olist tiene cobertura en los 27 estados de Brasil.
- El 81% de sus clientes están registrados en 6 estados



Estado	Clientes	Share
SP	41,746	41,98%
RJ	12,852	12,92%
MG	11,635	11,70%
RS	5,466	5,50%
PR	5,045	5,07%
SC	3,637	3,66%
BA	3,380	3,40%
DF	2,140	2,15%
ES	2,033	2,04%
GO	2,020	2,03%
PE	1,652	1,66%
CE	1,336	1,34%
PA	975	0,98%
MT	907	0,91%
MA	747	0,75%
MS	715	0,72%
PB	536	0,54%
PI	495	0,50%
RN	485	0,49%
AL	413	0,42%
SE	350	0,35%
TO	280	0,28%
RO	253	0,25%
AM	148	0,15%
AC	81	0,08%
AP	68	0,07%
RR	46	0,05%
Total	99,441	

CONSULTAS

MYSQL

```
SELECT customer_state, COUNT(*) AS  
CustomersXState  
FROM customers  
GROUP BY customer_state  
ORDER BY 2 DESC;
```

-

MONGODB

```
MongoDB  
[{$group: {  
  _id: {customer: "$customer_id", state:  
    "$customer_state"}  
}}, {$unwind: {  
  path: "$_id"  
}}, {$addFields: {  
  customer: "$_id.customer",  
  state: "$_id.state"  
}}, {$group: {  
  _id: "$state",  
  customerXstate: {  
    $sum: 1  
  }  
}}, {$sort: {  
  customerXstate: -1  
}}]
```

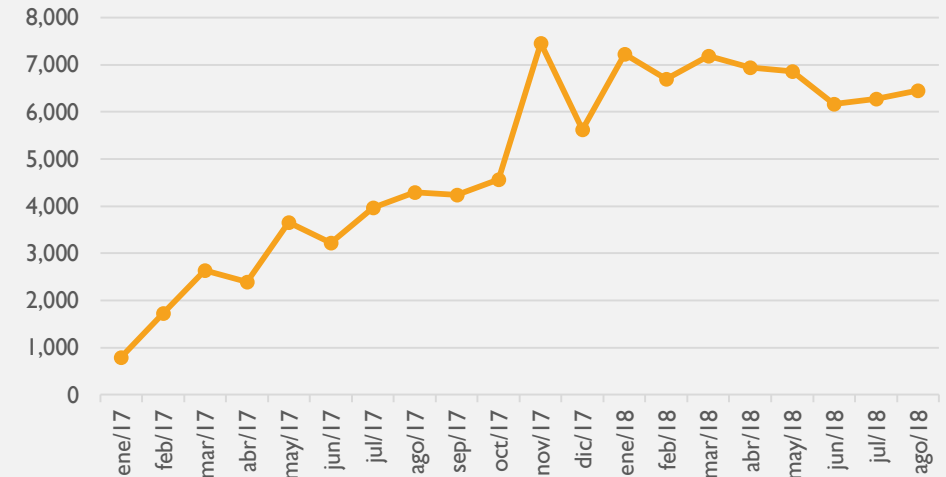
USUARIOS ACTIVOS POR MES

Mes	Cientes
sep/16	3
oct/16	308
dic/16	1
ene/17	789
feb/17	1,733
mar/17	2,641
abr/17	2,391
may/17	3,660
jun/17	3,217
jul/17	3,969
ago/17	4,293
sep/17	4,243
oct/17	4,568
nov/17	7,451
dic/17	5,624
ene/18	7,220
feb/18	6,694
mar/18	7,188
abr/18	6,934
may/18	6,853
jun/18	6,160
jul/18	6,273
ago/18	6,452

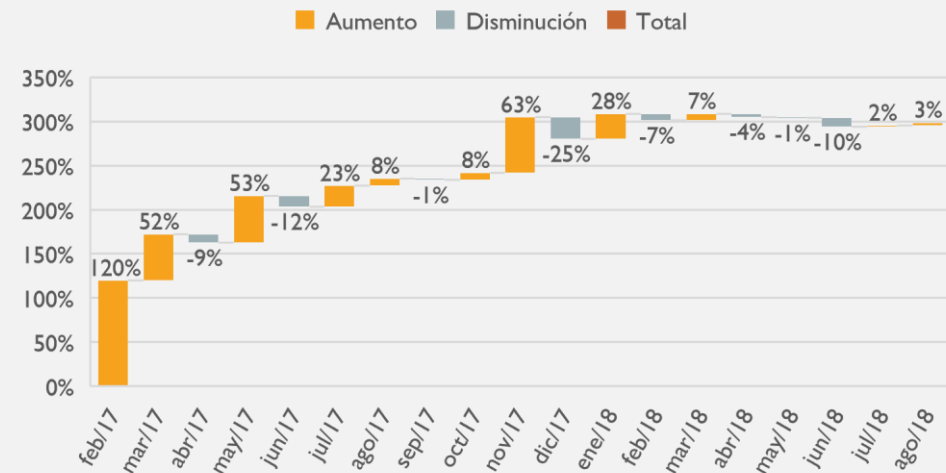


Mes	Cientes	Delta	Delta (%)
ene/17	789	789	NA
feb/17	1,733	944	120%
mar/17	2,641	908	52%
abr/17	2,391	-250	-9%
may/17	3,660	1,269	53%
jun/17	3,217	-443	-12%
jul/17	3,969	752	23%
ago/17	4,293	324	8%
sep/17	4,243	-50	-1%
oct/17	4,568	325	8%
nov/17	7,451	2,883	63%
dic/17	5,624	-1,827	-25%
ene/18	7,220	1,596	28%
feb/18	6,694	-526	-7%
mar/18	7,188	494	7%
abr/18	6,934	-254	-4%
may/18	6,853	-81	-1%
jun/18	6,160	-693	-10%
jul/18	6,273	113	2%
ago/18	6,452	179	3%
Promedio	4,918	323	16%

Diferentes compradores



Comportamiento de usuario activos



CONSULTA MYSQL CON VISTA

CREAR VISTA

```
CREATE VIEW orden AS
(SELECT o.order_id AS Order_id,
o.order_purchase_timestamp AS Purchase_Date,
o.order_delivered_customer_date AS
Delivered_Date, o.customer_id AS customer_id,
op.payment_type AS Payment_Type,
op.payment_value AS Payment_Value,
s.seller_id AS seller
FROM orders o
LEFT JOIN order_payment op
ON o.order_id = op.order_id
RIGHT JOIN order_items s
ON o.order_id = s.order_id
ORDER BY s.order_id ASC);
```

CONSULTA

```
SELECT
DATE_FORMAT(Purchase_Date, '%Y %m') AS date,
COUNT(DISTINCT(customer_id)) AS dist_customers
FROM orden
GROUP BY DATE_FORMAT(Purchase_Date, '%Y %m')
ORDER BY 1;
```

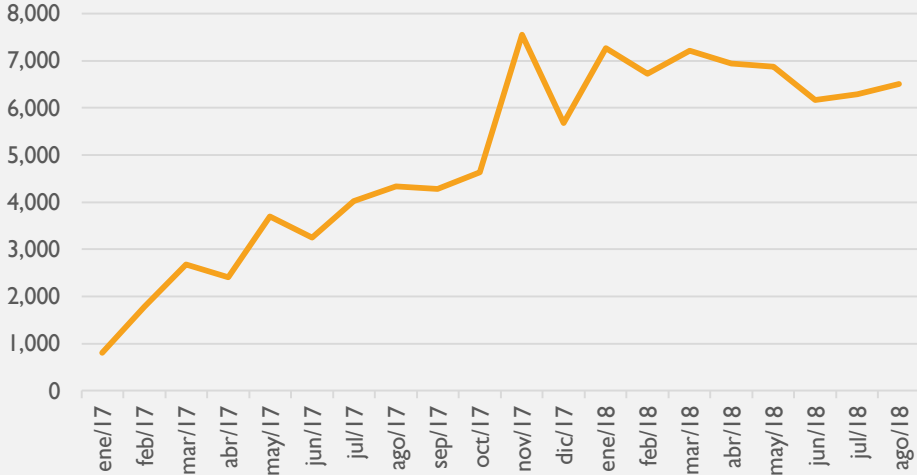
ORDENES MENSUALES

Mes	Orders
dic/16	1
oct/16	324
sep/16	4
dic/17	5,673
nov/17	7,544
oct/17	4,631
sep/17	4,285
ago/17	4,331
jul/17	4,026
jun/17	3,245
may/17	3,700
abr/17	2,404
mar/17	2,682
feb/17	1,780
ene/17	800
oct/18	4
sep/18	16
ago/18	6,512
jul/18	6,292
jun/18	6,167
may/18	6,873
abr/18	6,939
mar/18	7,211
feb/18	6,728
ene/18	7,269

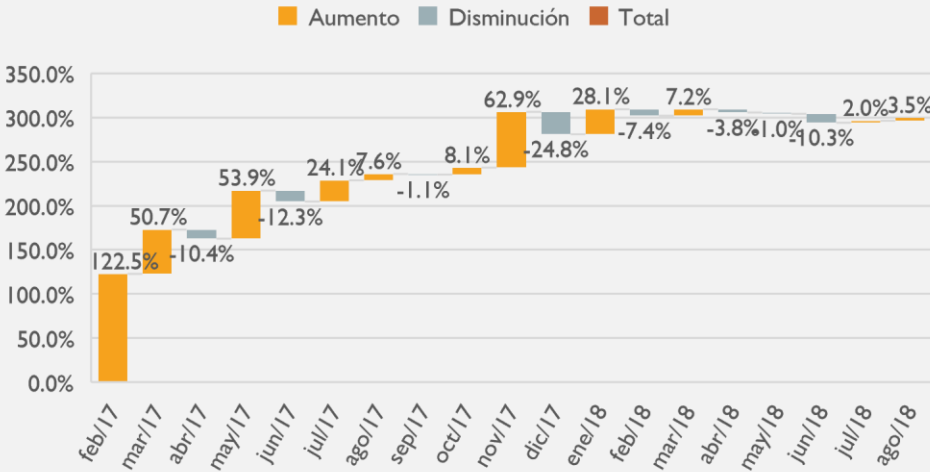


Mes	Orders	Delta	Delta (%)
ene/17	800	800	NA
feb/17	1,780	980	122.5%
mar/17	2,682	902	50.7%
abr/17	2,404	-278	-10.4%
may/17	3,700	1,296	53.9%
jun/17	3,245	-455	-12.3%
jul/17	4,026	781	24.1%
ago/17	4,331	305	7.6%
sep/17	4,285	-46	-1.1%
oct/17	4,631	346	8.1%
nov/17	7,544	2,913	62.9%
dic/17	5,673	-1,871	-24.8%
ene/18	7,269	1,596	28.1%
feb/18	6,728	-541	-7.4%
mar/18	7,211	483	7.2%
abr/18	6,939	-272	-3.8%
may/18	6,873	-66	-1.0%
jun/18	6,167	-706	-10.3%
jul/18	6,292	125	2.0%
ago/18	6,512	220	3.5%

Ordenes



Comportamiento de ordenes mensuales



CONSULTAS

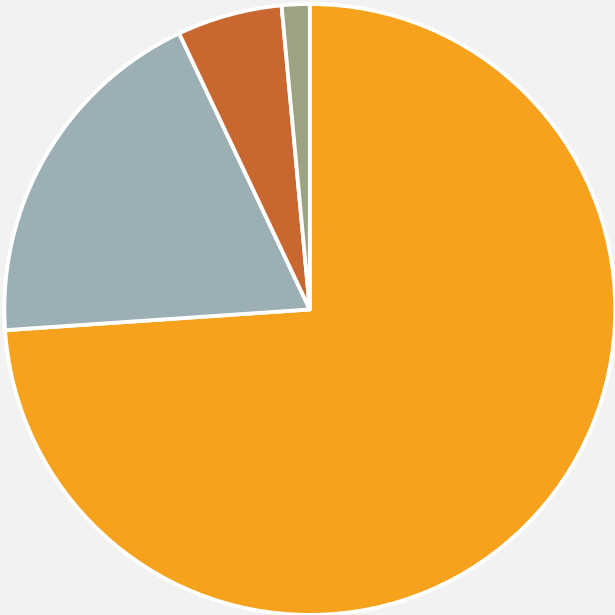
MYSQL

```
SELECT MONTH(order_purchase_timestamp)  
AS Month, YEAR(order_purchase_timestamp) AS  
Year, COUNT(*) AS Sales  
FROM orders  
GROUP BY 1, 2  
ORDER BY 2 ASC, 1 DESC;
```

MONGODB

- Por la estructura de la base de datos, el rendimiento de MongoDB para éste tipo de consultas es bajo, se necesitan muchas agregaciones las cuáles tardan mucho en responder.

MÉTODOS DE PAGO



<u>Método de pago</u>	<u>Share</u>
<u>Credit card</u>	77.23%
<u>Boleto</u>	19.90%
<u>Voucher</u>	5.81%
<u>Debit Card</u>	1.54%
	104.48%

- Credit card
- Boleto
- Voucher
- Debit Card

CONSULTAS

MYSQL

- ```
SELECT DISTINCT(payment_type),
COUNT(*)/
 (SELECT COUNT(DISTINCT(order_id))
 FROM order_payment)
 AS payment_method_share
FROM order_payment
GROUP BY payment_type
ORDER BY payment_method_share DESC;
```

## MONGODB

1. Número total de órdenes:  

```
[{$group: {
 _id: "$order_id",
 ordenes: {
 $sum: 1
 }
}}, {$count: 'order_id'}]
```
2. Cálculo de share:  

```
[{$group: {
 _id: "$payment_type",
 ordenes: {
 $sum: 1
 }
}}, {$addFields: {
 Share: {$divide: ["$ordenes", 99441]}
}}]
```

## 4. CONCLUSIONES



# BASE DE DATOS



Por la estructura de la base de datos, era mucho más sencillo manejarla en MySQL Workbench.



Para la carga de las bases de datos, es mejor primero generar las tablas vacías con llaves primarias y foráneas para luego sólo cargar los datos.



Cuando se cargan más de 100K registros, la velocidad de carga a MongoDB es notablemente mayor.



# OLIST DATASET

- Tiene cobertura en todo Brasil.
- Las ciudades más grandes son las que más usuarios tienen.
- Vale la pena meternos más en temas de limpieza de bases de datos para problemas como:
  - Normalización de distribuciones
  - Limpia de datos
  - Duplicados
- El método de pago favorito es tarjeta de crédito