

Checkpoint1

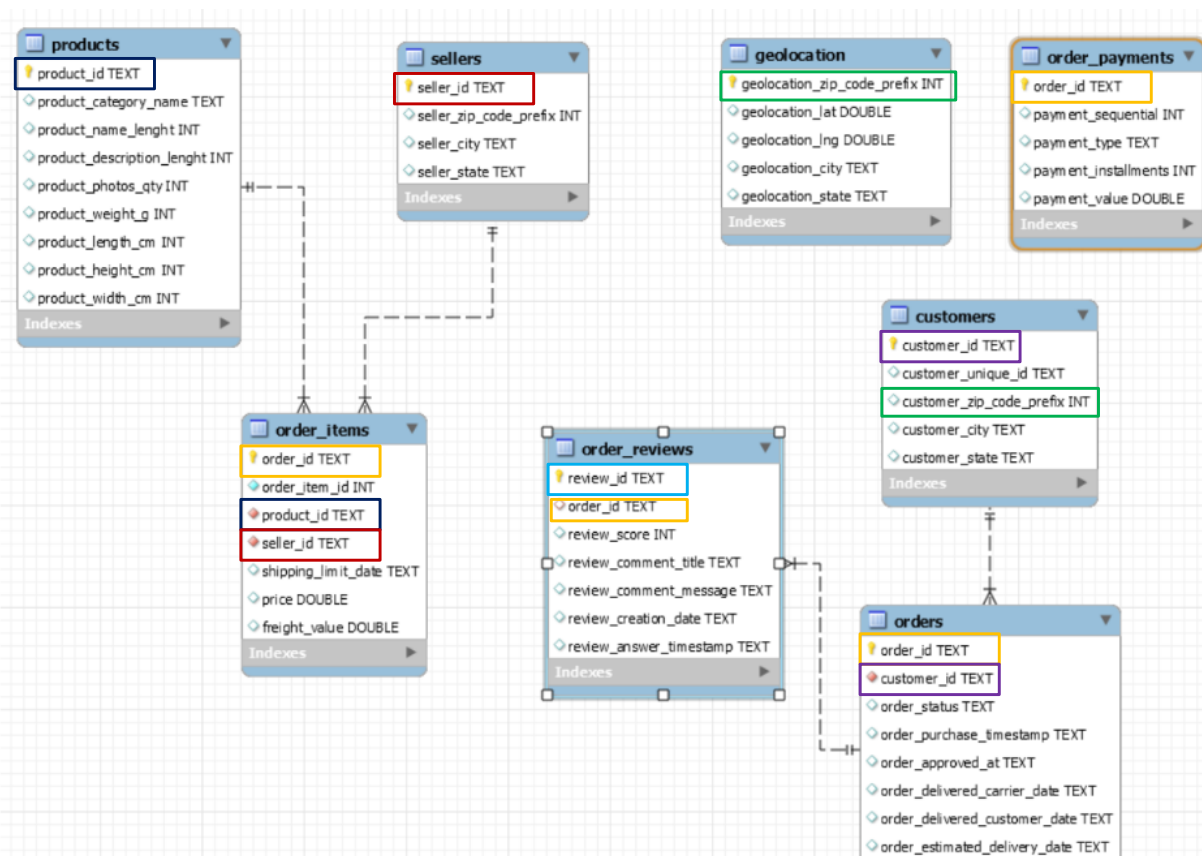
Brazilian Ecommerce

Stephanie Haidenberg
Módulo I – Data Analysis

Agenda

- Definición del proyecto
 - Estructura base de datos
 - Problema a resolver
- Consultas (MySQL / MongoDB)
 - Calificación promedio
 - Comentarios negativos
 - Calidad de descripciones y fotografías
 - Análisis vendedores
 - Análisis categorías
- Conclusión

Estructura Base de Datos



- Base de datos de un ecommerce brasileño.
- La base de datos cuenta con información sobre vendedores, clientes, órdenes, geolocalización y productos.

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

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



¿Cuál es la calificación promedio?

- La calificación promedio de todas las órdenes del ecommerce fue de 4.09 en una escala de 5.



```
SELECT avg(review_score) FROM order_reviews;
```

	avg(review_score)
▶	4.0943



```
[{$group: {  
  _id: null,  
  Promedio: {  
    $avg: "$review_score"  
  }  
}}]
```

```
_id: null  
Promedio: 4.094209420942094
```

¿Cuáles son los comentarios que tienen calificación negativa?

- La mayoría de las calificaciones bajas mencionaron que el producto fue de mala calidad y que se atrasó su pedido.
- Existe una oportunidad en mejorar la calidad de los productos y la logística de envío.



```
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%";
```

order_id	review_score	review_comment_message	review_creation_date
9ac05114800f02bfaa783bd76842dbe2	1	Producto muito inferior, mal acabado.	7/3/2018 0:00
a50a23ee8e09f54388b66da19c633bc4	1	O produto entregue difere do pedido trata-se d...	4/21/2018 0:00
bf43328103299ae529116e4521e82191	3	Comprei uma cadeira, recebi no prazo, porém a ...	4/13/2017 0:00
96123b9ae5a9cdfd4efe5fe402a94675	1	Fornecedor targar y en me mandou um message...	10/28/2016 0:00
2c23086c23ba12b76e8312efa418c62d	4	Muito fácil e rápido. Tempo de entrega no meno...	4/18/2017 0:00
cc6e97ae2828f8e0e3d687628b51c155	1	Material inferior ao que demonstrava na imagem...	4/25/2018 0:00
eb23b9fbd39df83ffdb7e7dd6a00fe	3	Na descrição do produto é indicado que o mem...	8/31/2018 0:00
5981ae276b32c4b5d69ed39efd08d0a8	1	como posso avaliar o produto? só posso avaliar ...	5/17/2017 0:00
318abae1b1ee9847e237ac1501e1d0e1	4	O produto cumpre o que promete, mas achei a ...	9/2/2017 0:00
a85ad381ff410bb40159e498feb30a13	1	produto inferior ao esperado. Desejo devolve lo.	6/8/2018 0:00
88791555f9832d0bee1f87f2074392ef	3	O pedido chegou no tempo previsto.O produto,...	12/20/2017 0:00



```
{{$match:
{
  review_comment_message: {$in: [ /inferior/i, /atrasado/i]},
}}, {$project: {
  order_id: 1, review_score:1, review_comment_message:1,
  review_creation_date: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?

- No existe la buena práctica de tener un mínimo de caracteres promedio en las descripciones de los productos.



```
SELECT product_category_name, product_description_lenght
FROM products
WHERE product_description_lenght < 700
ORDER BY product_description_lenght ASC;
```

	product_category_name	product_description_lenght
▶	utilidades_domesticas	4
	bebes	4
	fashion_bolsas_e_acessorios	23
	brinquedos	28
	bebidas	30
	utilidades_domesticas	33
	fashion_bolsas_e_acessorios	34
	malas_acessorios	38
	papelaria	39
	esporte_lazer	40
	utilidades_domesticas	42



```
[{$match: { product_description_lenght: {$lte: 700}}},
{$sort: { product_description_lenght: 1}}, {$project: {
product_category_name:1,
product_description_lenght: 1}}]
```

Output after [\\$project](#) stage ⓘ (Sample of 20 documents)

```
_id: ObjectId("5f28b9c97e0ab15204e8c72d")
product_category_name: "bebes"
product_description_lenght: 4
```

```
_id: ObjectId("5f28b9c77e0ab15204e8b58c")
product_category_name: "utilidades_domesticas"
product_description_lenght: 4
```

```
_id: ObjectId("5f28b9c77e0ab15204e8b58c")
product_category_name: "fashion_bolsas_e_acessorios"
product_description_lenght: 23
```

Se debe de tener un mínimo de 3 fotografías mostrando el producto,

¿cuáles son las categorías que tienen menos de 3 fotografías?

- No existe la buena práctica de tener un mínimo de 3 fotografías por publicación de producto. Poner atención especial en flores, casa conforto y la cuisine.



```
[{$match: {
  product_photos_qty: {$ne: NaN}
}}, {$group: {
  _id: "$product_category_name",
  Fotos: {
    $avg: "$product_photos_qty"
  }
}}, {$sort: {
  Fotos: 1
}}]
```

<pre>_id: "flores" Fotos: 1</pre>	<pre>_id: "casa_conforto_2" Fotos: 1</pre>	<pre>_id: "la_cuisine" Fotos: 1</pre>
-----------------------------------	--	---------------------------------------

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

- Se deberá monitorear y dar un seguimiento más detallado a los vendedores del estado de SP, especialmente de la ciudad Ibatinga, así como el estado PR.



```
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;
```

	seller	calificacion	seller_state	seller_city
►	7ad32824caee82087b3e2e5f33b1bf32	1	SP	ibitinga
	e9bc59e7b60fc3063eb2290deda4cced	1	PR	maringa
	0df3984f9dfb3d49ac6366acbd3bbb85	1	MG	betim
	2eb70248d66e0e3ef83659f71b244378	1	SP	campinas
	01cf7e3d21494c41fb86034f2e714fa1	1	PR	francisco beltrao
	391fc6631aebcf3004804e51b40bcf1e	1	SP	ibitinga
	8b321bb669392f5163d04c59e235e066	1	SP	sao paulo
	897060da8b9a21f655304d50fd935913	1	SP	ribeirao preto
	d3674f271c91f824f82d24d92011f669	1	DF	brasilvia
	d2374cbcb3ca4ab1086534108cc3ab7	1	SP	ibitinga

¿Cuáles son las 10 categorías con peor calificación?

- Las 3 categorías que deberán ser analizadas y optimizadas son climatizacao, pcs y electrodomésticos.



```
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);
```

	Categoria	Calificacion_Promedio
►	climatizacao	1.0000
	pcs	1.0000
	eletrodomesticos_2	1.0000
	construcao_ferramentas_seguranca	3.0000
	eletrodomesticos	3.2500
	papelaria	3.5000
	informatica_acessorios	3.6000
	pet_shop	3.6667
	telefonica	3.7000
	brinquedos	3.7143

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 (>700 caracteres) y fotografías (>3 fotografías) de los productos para mejorar la satisfacción del cliente.
- Hacer un seguimiento 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.