

# Trabajo Nociones de Arquitectura de la Información

Versión: 2021.05.30 15:07

*Observación: Cada vez que agregue nuevos elementos al documento, o que modifique algún componente del informe, revise la coherencia y consistencia con los otros elementos que hacen parte del mismo.*

## RESPONSABLES

Nombre Completo – Documento de Identificación
1.Daniela Muñoz Elorza - 1053837464
2.Mauricio Herrera Mejía - 1152436206
3.Ligia Fernanda Zabala Gamez - 1152691930
4.Blanca Isabel Florez - 1023927894
<b>REPO EN GITHUB:</b> <a href="https://github.com/lfzabalag/TrabajosAGD">https://github.com/lfzabalag/TrabajosAGD</a>

Realiza este trabajo considerando los datos que generan los sistemas transaccionales e información no estructurada de tu dominio (si trabajas por ejemplo para TCC tu dominio es la mensajería; también puedes explorar en la página <https://www.kaggle.com/datasets> o <https://arxiv.org/>). Considera tener acceso a esta información, de al menos 10 MB (puede ser uno o varios archivos de texto), y **tener al menos cuatro clases conceptuales. Este documento también debe almacenarse en el REPO. Plazo Máximo de Entrega 23 de Mayo, NO SE recibirá por correo electrónico, envío por <https://forms.gle/h7ty3yZykaUq5m7y6>**

## 1 COMPRENSIÓN DEL NEGOCIO

---

### 1.1 DESCRIPCIÓN DEL CONTEXTO DEL NEGOCIO.

Describe en máximo 250 palabras el contexto en el cual se generan los datos y cuál es el proceso que los genera.

El conjunto de datos a desarrollar se obtuvo de kaggle una página que recolecto los datos proporcionados por Olist, la cual es una de las empresas de mercado más grande de Brasil quienes ofrecen servicios de comercio electrónico, donde ayudan a pequeñas y medianas empresas a aumentar sus clientes. Olist tiene como objetivo la conexión de estas empresas de todo Brasil con sus clientes de una forma sencilla y con un solo

contrato. Lo que hacen las empresas pequeñas y medianas por medio de Olist es vender sus productos a través de su plataforma (Olist store), enviándolos directamente a sus clientes confiando en la logística que Olist dispone para hacer llegar cada uno de los productos a sus respectivos destinos. Desde el momento es que se contactan con Olist, el comerciante es guiado para su buena práctica de venta por internet, así mismo busca la mejor experiencia para los consumidores(compradores), solucionando cualquier tipo de situación que se pudiera presentar; Para finalizar Olist mide la calidad de su trabajo por medio de una encuesta de satisfacción que envían al cliente que recibe el producto por medio de un correo electrónico.

Los datos que se van a manejar tienen información de 100k pedidos realizados entre los años 2016 a 2018 realizados en múltiples mercados en Brasil

## **1.2 IDENTIFICACIÓN DEL PROBLEMA:**

Delimite en máximo 150 palabras la problemática, así como identificar los requisitos, supuestos, restricciones y beneficios de la solución de este.

El problema planteado en el grupo es ¿buscar cuáles son los productos que mas ordenan los clientes por ciudad?, para ello debemos tener en cuenta la base de datos a manejar y plantearnos una ruta con respecto a esos datos para encontrar la más idónea, integrando clientes, productos y ciudades, así mismo con estos datos identificaremos las clases con la que se deben trabajar y relacionar, donde poco a poco se construirá el esquema idóneo.

Se deberá tener en cuenta diferentes parámetros de exclusión e inclusión con respecto a todos los datos que se tienen en esta base pues no todas nos ayudan a llegar a la resolución del problema planteado, se incluirán las tablas que relaciones las 3 características principales como los son clientes, productos y ciudades, esto nos ayuda a encontrar una ruta mas específica y sencilla a resolver el problema.

## **1.3 DETERMINACIÓN DE OBJETIVOS:**

Describe en máximo 150 palabras las metas a lograr al proponer una solución basada en un modelo de datos o de analítica (cómo y qué tipo de ventaja competitiva se ganará).

La meta a lograr es tener una solución al problema planteado desde un principio, es decir, tener claro que productos son los mas vendidos por ciudad, esto en base a un modelo analítico de datos ya que la analítica de negocios incluye el proceso de examinar conjuntos de datos para extraer conclusiones sobre la misma, el cual en este caso es un modelo descriptivo pues este formato permite a un negocio conocer o ver los parámetros o características principales dentro del mismo negocio, por ejemplo, en este caso se podrían observar los productos mas vendidos en cada ciudad, determinando así ventajas en cuanto a la toma de decisiones comerciales para aumentar los ingresos, para mejorar la eficiencia operativa y optimizar las campañas de servicio a cliente y de logística.

## 1.4 EVALUACIÓN DE LA SITUACIÓN ACTUAL:

Describe en máximo 150 palabras el estado actual antes de implementar la solución de analítica, a fin de tener un punto de comparación que permita medir el grado de éxito de la solución.

En principio los datos a tratar se encuentran en tablas de Excel individualizadas, algunas de estas tablas tienen espacios en blanco que dificultan la obtención de los resultados, pues sacan un error al encontrarse vacíos. Los datos no se encuentran relacionados en cuanto a cliente, producto y ciudad. La idea es sacar resultados de los productos más vendidos en cada una de las ciudades, esto con un parámetro anual, de los cuales los datos están organizados por mes.

## 2 COMPRENSIÓN DE LOS DATOS

---

### 2.1 RECOLECCIÓN DE DATOS

Describe en máximo 150 palabras los datos a utilizar identificando las fuentes, las técnicas empleadas en su recolección, los problemas encontrados en su obtención y la forma como se resolvieron los mismos. Además, adjunte los datos (archivos de texto, etc.) agréguelos en el github (**REPO EN GITHUB**) en un solo archivo, por favor compáralo(s). Llame el archivo T1.2.1.Datos.zip

Los datos que utilizamos para la solución del problema fueron recolectados desde la página <https://www.kaggle.com/olistbr/brazilian-ecommerce>; en la dataset de olist\_products\_dataset.csv en las columnas product\_category\_name, product\_name\_lenght, product\_description\_lenght, product\_photos\_qty se encontraron algunas celdas vacías, así que para que no generará error estas celdas vacías se llenaron con los datos Null, 0, 0, 0, respectivamente. En la dataset olist\_orders\_dataset.csv se encontraron celdas vacías en algunas fechas de entrega entre el cliente y el vendedor, para subsanar estos problemas se puso la fecha así 00/00/0000 00:00 para evitar problemas en la tabla. En la dataset de olist\_order\_reviews\_dataset.csv, en la cual se responden las encuestas por parte de los clientes, hay muchos espacios en blanco por ende para subsanar éstos se puso en las celdas Null y así evitar errores.

### 2.2 DESCRIPCIÓN DE DATOS (DICCIONARIO):

Diligencia la siguiente tabla, puede agregar otra columna si lo considera necesario.

**Tabla 1:** olist\_products\_dataset

Nombre del atributo / variable	Formato o Tipo de Dato	Descripción
product_id	char	Número de identificación del producto
product_category_name	char	Categoría a la que pertenece el producto
product_name_lenght	int	Número de caracteres extraídos del nombre del producto

product_description_length	int	Número de caracteres extraídos de la descripción del producto
product_photos_qty	int	Número de fotos publicadas del producto
product_weight_g	int	Medida de peso en gramos del producto
product_length_cm	int	Medida de longitud del producto
product_height_cm	int	Medida de la altura del producto
product_width_cm	int	Medida del ancho del producto

**Tabla 2:** olist\_order\_items\_dataset

Nombre del atributo / variable	Formato o Tipo de Dato	Descripción
order_id	char	Número de identificación de la orden del pedido
order_item	int	Cantidad de productos pedidos en la misma orden
product_id	char	Número de identificación del producto
seller_id	char	Número de identificación del vendedor
shipping_limit_date	date	Fecha límite de envío por parte del vendedor al socio logístico
price	float	Precio del producto
freight_value	float	Valor del flete por artículo

**Tabla 3:** olist\_orders\_dataset

Nombre del atributo / variable	Formato o Tipo de Dato	Descripción
order_id	char	Número de identificación de la orden del pedido
customer_id	char	Clave para el conjunto de datos del cliente. Por cada pedido es una clave única
order_status	string	Referencia al estado del producto
order_purchase_timestamp	date/time	Fecha en la cual se realizó la compra
order_approved_at	date/time	Fecha en la cual se aprobó el pedido
order_delivered_carrier_date	date/time	Fecha en la cual se entregó el pedido al socio logístico
order_delivered_customer_date	date/time	Fecha en la cual se entregó el pedido al cliente
order_estimated_delivery_date	date/time	Fecha estimada de entrega que se le informó al cliente en el momento de la compra

**Tabla 4:** olist\_customers\_dataset

Nombre del atributo / variable	Formato o Tipo de Dato	Descripción
--------------------------------	------------------------	-------------

customer_id	char	Clave para el conjunto de datos del cliente. Por cada pedido es una clave única
customer_unique_id	char	Identificador único del cliente
customer_zip_code_prefix	int	Primeros cinco dígitos del código postal del cliente
customer_city	string	Ciudad del cliente
customer_state	string	Estado del cliente

**Tabla 5:** olist\_geolocation\_dataset

Nombre del atributo / variable	Formato o Tipo de Dato	Descripción
geolocation_zip_code_prefix	int	Primeros cinco dígitos del código de área
geolocation_at	int	Latitud
geolocation_lng	int	Longitud
geolocation_city	string	Nombre de la Ciudad
geolocation_state	string	Estado de la localización

**Tabla 6:** olist\_order\_payments\_dataset

Nombre del atributo / variable	Formato o Tipo de Dato	Descripción
order_id	char	Número de identificación de la orden del pedido
payment_sequential	int	Pago del pedido por más de un medio de pago
payment_type	string	Método de pago elegido por el cliente
payment_installments	int	Número de cuotas elegidas por el cliente
payment_value	float	Valor de la transacción

**Tabla 7:** olist\_order\_reviews\_dataset

Nombre del atributo / variable	Formato o Tipo de Dato	Descripción
review_id	char	Número de identificación de la encuesta
order_id	char	Número de identificación de la orden del pedido
review_score	int	Encuesta de satisfacción de una a cinco
review_comment_title	string	Título del comentario dejado por el cliente
review_comment_message	string	Mensaje del comentario dejado por el cliente
review_creation_date	date/time	Fecha en la cual se creó la encuesta de satisfacción
review_answer_timestamp	date/time	Fecha en la cual se recibió respuesta de la encuesta

**Tabla 8:** olist\_sellers\_dataset

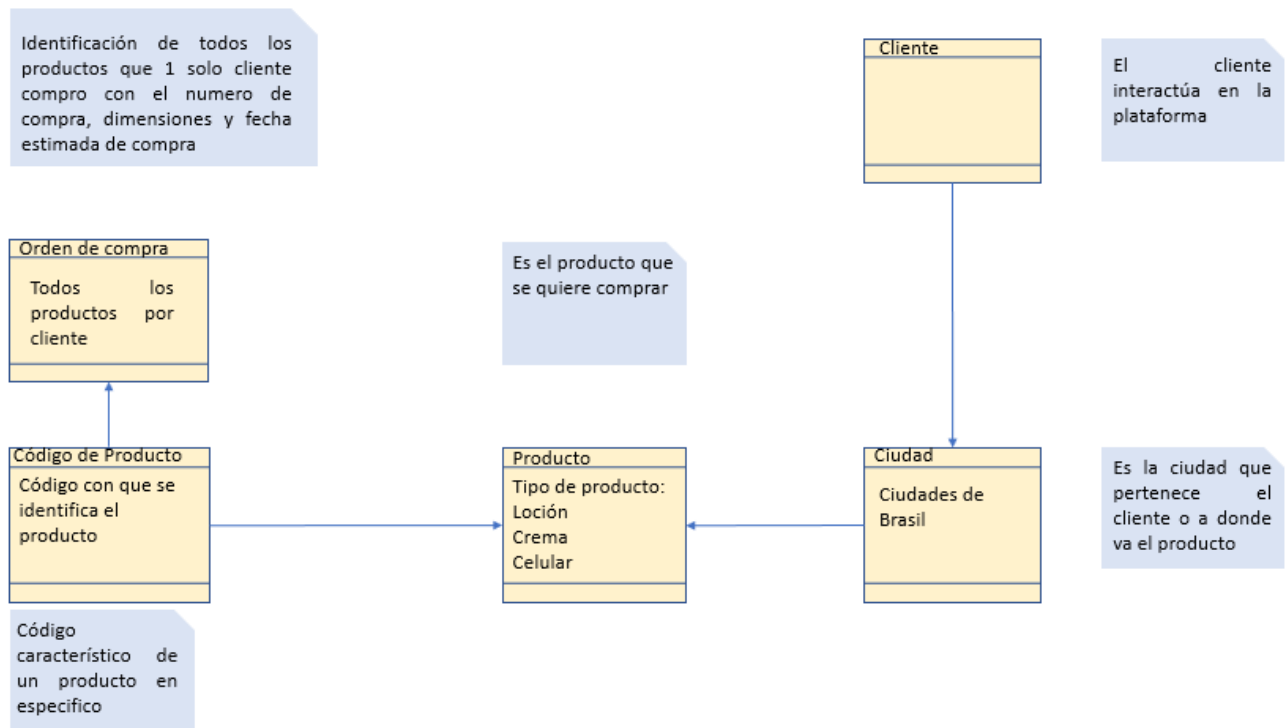
Nombre del atributo / variable	Formato o Tipo de Dato	Descripción
seller_id	char	Número de identificación del vendedor
seller_zip_code_id	int	Código postal del vendedor
seller_city	string	Ciudad del vendedor
seller_state	string	Estado del vendedor

**Tabla 9:** product\_category\_name\_traslation

Nombre del atributo / variable	Formato o Tipo de Dato	Descripción
product_name_category	string	Nombre de la categoría en portugués
product_name_category_english	string	Nombre de la categoría en inglés

## 2.3 MODELO DEL DOMINIO

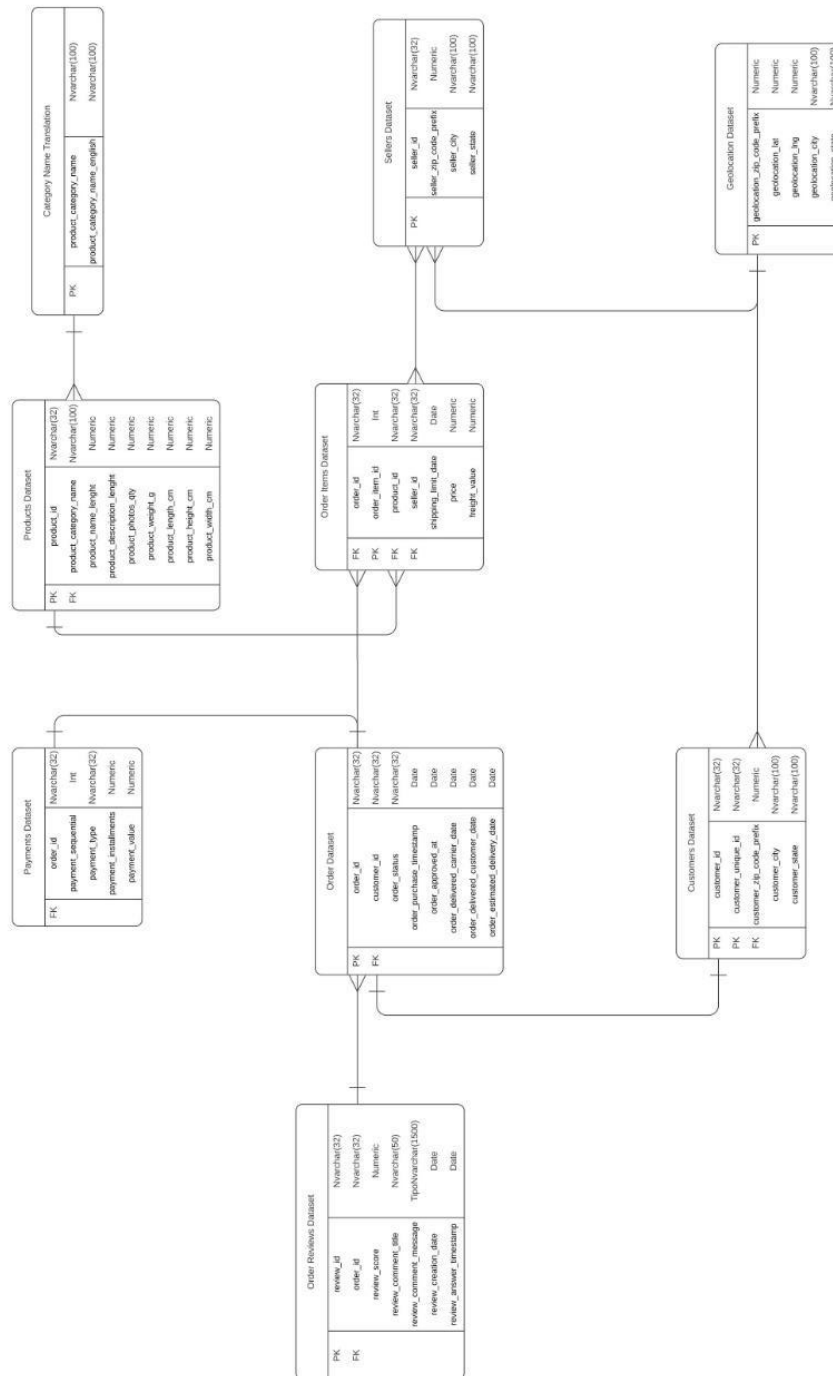
**Observación:** Incluya el gráfico del modelo del dominio que representa la estructura de datos de su problema.



## 3 MODELO ENTIDAD-RELACIÓN

### 3.1 TOMA DE PANTALLA DEL MODELO E-R

**Observación:** lo que se pide, puede usar <https://draw.io> o Microsoft Visio® y modele usando la notación de Barker.



### 3.2 SENTENCIA O CONSULTA DE CREACIÓN DEL TABLA(S)

**Observación:** Escriba el código en el Sistema de Gestión de Bases de Datos Relacionales de su elección (se recomienda SQLite por simplicidad, mediante <https://sqlitebrowser.org/>) para crear las tablas que corresponda con su conjunto de datos específico. Almacene en el repositorio **(REPO EN GITHUB)** el script con el nombre de T1.3.2.Creacion\_Tablas.sql

Se recomienda repasar SQL en <https://www.w3schools.com/sql/default.asp>

### 3.3 SENTENCIAS PARA INSERTAR DATOS

**Observación:** Escriba el código para insertar los datos en cada una de las tablas creadas. Almacene en el repositorio **(REPO EN GITHUB)** el script con el nombre de T1.3.3.Insertar\_Datos.sql

### 3.4 SENTENCIA DE CONSULTA

**Observación:** realice la exploración básica de los datos, conteos totales y por categorías, máximos, promedio y mínimos. Es decir, aplique estadística descriptiva con el fin de conocer las propiedades de los datos y entenderlos lo mejor posible. Use solamente sentencias SQL. Anexe las tomas de pantalla donde evidencie la sentencia SQL y su correspondiente ejecución. Además, Almacene en el repositorio **(REPO EN GITHUB)** el script con el nombre de T1.3.4.Consultar\_Datos.sql

## Exploración de datos

- Exploración básica

The screenshot shows the SQLiteOnline.com web interface. The browser address bar displays 'sqliteonline.com'. The interface includes a menu bar with 'File', 'Owner DB', 'Run', 'Export', and 'Import'. On the left, a sidebar lists database engines: SQLite, MariaDB, PostgreSQL, and MS SQL. The 'MS SQL' engine is selected, and a table named 'categoria' is chosen. The main area shows a SQL query: 'SELECT \* FROM categoria'. Below the query, the results are displayed in a table with two columns: 'product\_category\_name' and 'product\_category\_name\_english'. The results list various product categories such as 'beleza\_saude', 'informatica\_acessorios', 'automotivo', etc. An 'Activar Windows' watermark is visible in the bottom right corner.

product_category_name	product_category_name_english
beleza_saude	health_beauty
informatica_acessorios	computers_accessories
automotivo	auto
cama_mesa_banho	bed_bath_table
moveis_decoracao	furniture_decor
esporte_lazer	sports_leisure
perfumaria	perfumery
utilidades_domesticas	housewares
telefonica	telephony



- Conteos totales

The screenshot shows the sqLiteonline.com interface. The left sidebar lists databases: SQLite, MariaDB, PostgreSQL, and MS SQL. The MS SQL database is selected, showing a list of tables: categoria, ciudad, clientes, demo, encuestas, envios, orden, ordendelproducto, pagodelaorden, and producto. The 'ciudad' table is selected. The SQL editor contains the query: `1 SELECT COUNT (*) FROM ciudad`. The results pane shows a single row with the value 2999. The bottom of the screen shows a Windows taskbar with a Word document titled 'primer trabajo de la materia - Word'.

- Conteo por categoría

The screenshot shows the sqLiteonline.com interface. The left sidebar lists databases: SQLite, MariaDB, PostgreSQL, and MS SQL. The MS SQL database is selected, showing a list of tables: categoria, ciudad, clientes, demo, encuestas, envios, orden, ordendelproducto, pagodelaorden, and producto. The 'categoria' table is selected. The SQL editor contains the query: `1 SELECT * FROM categoria`, `2 WHERE product_category_name like 'b%'`, and `3`. The results pane shows a table with two columns: 'product\_category\_name' and 'product\_category\_name\_english'. The data rows are: 

product_category_name	product_category_name_english
beleza_saude	health_beauty
bebés	baby
brinquedos	toys
bebidas	drinks

. The bottom of the screen shows a Windows taskbar with a Word document titled 'primer trabajo de la materia - Word'.

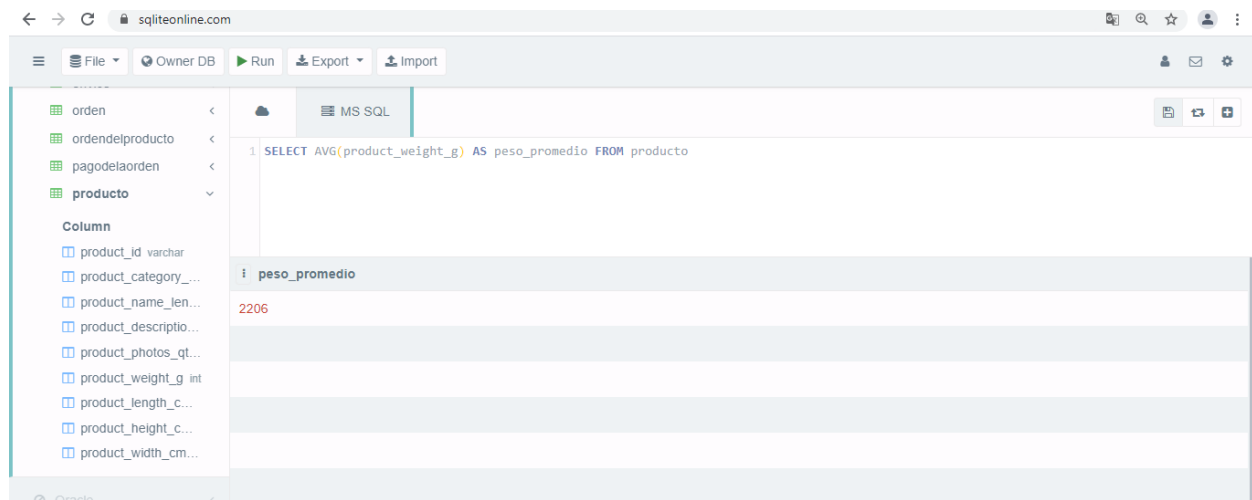
- Valor máximo

The screenshot shows the sqLiteonline.com interface. The left sidebar lists databases: SQLite, MariaDB, PostgreSQL, and MS SQL. The MS SQL database is selected, showing a table list including 'encuestas'. The main editor contains the SQL query: `SELECT Max (review_score) AS maximo_valor FROM encuestas`. The results pane shows a single row with the column 'maximo\_valor' and the value '5'. An 'Activar Windows' watermark is visible in the bottom right corner.

- Valor mínimo

The screenshot shows the sqLiteonline.com interface. The left sidebar lists databases: SQLite, MariaDB, PostgreSQL, and MS SQL. The MS SQL database is selected, showing a table list including 'producto'. The main editor contains the SQL query: `SELECT Min (product_weight_g) AS minimo_peso FROM producto`. The results pane shows a single row with the column 'minimo\_peso' and the value '50'. An 'Activar Windows' watermark is visible in the bottom right corner.

- Valor promedio



## 4 MONGODB

---

### 4.1 SENTENCIA O CONSULTA DE CREACIÓN DEL DOCUMENTO(S)

**Observación:** Escriba el código en MongoDB para crear al menos 20 documentos que correspondan a su conjunto de datos específico. Almacene en el repositorio **(REPO EN GITHUB)** el script con el nombre de T1.4.1.Creacion\_Documentos.sql

### 4.2 SENTENCIA DE CONSULTA

**Observación:** Realice la exploración básica de los datos, conteos totales y por categorías, máximos, promedio y mínimos. Es decir, aplique estadística descriptiva con el fin de conocer las propiedades de los datos y entenderlos lo mejor posible. Use solamente sentencias SQL. Anexe las tomas de pantalla donde evidencie la sentencia SQL y su correspondiente ejecución. Además, Almacene en el repositorio **(REPO EN GITHUB)** el script con el nombre de T1.4.2.Consultar\_Datos.sql

## Exploración datos mongo

- Exploración básica

```
C:\mongo\bin\mongo.exe
> show databases
admin      0.000GB
analitica  0.002GB
config     0.000GB
local      0.000GB
> use analitica
switched to db analitica
> show collections;
categoria
ciudades
clientes
encuestas
envios
ordendelpago
ordendelproducto
ordenes
productos
> db.categoria.find().pretty();
{
  "_id" : ObjectId("60a97d9b03cdf3267c980d64"),
  "product_category_name" : "beleza_saude",
  "product_category_name_english" : "health_beauty"
}
{
  "_id" : ObjectId("60a97d9b03cdf3267c980d65"),
  "product_category_name" : "informatica_acessorios",
  "product_category_name_english" : "computers_accessories"
}
{
  "_id" : ObjectId("60a97d9b03cdf3267c980d66"),
  "product_category_name" : "automotivo",
  "product_category_name_english" : "auto"
}
{
  "_id" : ObjectId("60a97d9b03cdf3267c980d67"),
  "product_category_name" : "cama_mesa_banho",
  "product_category_name_english" : "bed_bath_table"
}
{
  "_id" : ObjectId("60a97d9b03cdf3267c980d68"),
  "product_category_name" : "moveis_decoracao",
  "product_category_name_english" : "furniture_decon"
```

- Conteo total

```
C:\mongo\bin\mongo.exe
{
  "_id" : ObjectId("60a97f8603cdf3267c9816b5"),
  "geolocation_zip_code_prefix" : 1039,
  "geolocation_lat" : -23.54188301,
  "geolocation_lng" : -46.63991947,
  "geolocation_city" : "sao paulo",
  "geolocation_state" : "SP"
}
{
  "_id" : ObjectId("60a97f8603cdf3267c9816b6"),
  "geolocation_zip_code_prefix" : 1024,
  "geolocation_lat" : -23.54138952,
  "geolocation_lng" : -46.62989909,
  "geolocation_city" : "sao paulo",
  "geolocation_state" : "SP"
}
{
  "_id" : ObjectId("60a97f8603cdf3267c9816b7"),
  "geolocation_zip_code_prefix" : 1009,
  "geolocation_lat" : -23.5460354,
  "geolocation_lng" : -46.63658793,
  "geolocation_city" : "sao paulo",
  "geolocation_state" : "SP"
}
{
  "_id" : ObjectId("60a97f8603cdf3267c9816b8"),
  "geolocation_zip_code_prefix" : 1046,
  "geolocation_lat" : -23.54588428,
  "geolocation_lng" : -46.64316319,
  "geolocation_city" : "sao paulo",
  "geolocation_state" : "SP"
}
Type "it" for more
> db.ciudades.find().count();
2999
>
```

- Conteo por categoría

```
C:\mongo\bin\mongo.exe
{
  "_id" : ObjectId("60a97f8603cdf3267c9816b6"),
  "geolocation_zip_code_prefix" : 1024,
  "geolocation_lat" : -23.54138952,
  "geolocation_lng" : -46.62989909,
  "geolocation_city" : "são paulo",
  "geolocation_state" : "SP"
}
{
  "_id" : ObjectId("60a97f8603cdf3267c9816b7"),
  "geolocation_zip_code_prefix" : 1009,
  "geolocation_lat" : -23.5460354,
  "geolocation_lng" : -46.63658793,
  "geolocation_city" : "sao paulo",
  "geolocation_state" : "sp"
}
{
  "_id" : ObjectId("60a97f8603cdf3267c9816b8"),
  "geolocation_zip_code_prefix" : 1046,
  "geolocation_lat" : -23.54588428,
  "geolocation_lng" : -46.64316319,
  "geolocation_city" : "sao paulo",
  "geolocation_state" : "sp"
}
Type "it" for more
> db.ciudades.find().count();
2999
> db.categoria.find({campo:/^b/}).count();
0
> db.categoria.find({product_category_name:/^b/}).count();
4
>
=
```

- Valor máximo

```
C:\mongo\bin\mongo.exe
> db.encuestas.aggregate([{$group:{review_score:{review_score:"$review_score"},maximo:{$max:'$review_score'}}}]);
uncaught exception: Error: command failed: {
  "ok" : 0,
  "errmsg" : "The field 'review_score' must be an accumulator object",
  "code" : 40234,
  "codeName" : "Location40234"
} : aggregate failed :
_getErrorWithCode@src/mongo/shell/utils.js:25:13
doassert@src/mongo/shell/assert.js:18:14
assertCommandWorked@src/mongo/shell/assert.js:639:17
assert.commandWorked@src/mongo/shell/assert.js:729:16
DB.prototype._runAggregate@src/mongo/shell/db.js:266:5
DBCollection.prototype.aggregate@src/mongo/shell/collection.js:1058:12
@(shell):1:1
> db.encuestas.aggregate([{$group:{_id:{review_score:"$review_score"},maximo:{$max:'$review_score'}}}]);
{ "_id" : { "review_score" : 4 }, "maximo" : 4 }
{ "_id" : { "review_score" : 3 }, "maximo" : 3 }
{ "_id" : { "review_score" : 1 }, "maximo" : 1 }
{ "_id" : { "review_score" : 5 }, "maximo" : 5 }
{ "_id" : { "review_score" : 2 }, "maximo" : 2 }
> db.encuestas.aggregate([{$group:{_id:"",last:{$max:'$review_score'}}}]);
{ "_id" : "", "last" : 5 }
```

- valor mínimo

```
{ "_id" : { "review_score" : 1 }, "maximo" : 1 }
{ "_id" : { "review_score" : 5 }, "maximo" : 5 }
{ "_id" : { "review_score" : 2 }, "maximo" : 2 }
> db.encuestas.aggregate([{$group:{_id:"",last:{$max:"$review_score"}}}]);
{ "_id" : "", "last" : 5 }
> db.encuestas.aggregate([{$group:{_id:"",last:{$min:"$review_score"}}}]);
{ "_id" : "", "last" : 1 }
> db.productos.aggregate([{$group:{_id:"",last:{$min:"$product_weight_g"}}}]);
{ "_id" : "", "last" : 50 }
>
```

- valor promedio

```
C:\mongo\bin\mongo.exe
"codeName" : "Location40234"
} : aggregate failed :
_getErrorWithCode@src/mongo/shell/utils.js:25:13
doassert@src/mongo/shell/assert.js:18:14
_assertCommandWorked@src/mongo/shell/assert.js:639:17
assert.commandWorked@src/mongo/shell/assert.js:729:16
DB.prototype._runAggregate@src/mongo/shell/db.js:266:5
DBCollection.prototype.aggregate@src/mongo/shell/collection.js:1058:12
@(shell):1:1
> db.encuestas.aggregate([{$group:{_id:{review_score:"$review_score"},maximo:{$max:'$review_score'}}}]);
{ "_id" : { "review_score" : 4 }, "maximo" : 4 }
{ "_id" : { "review_score" : 3 }, "maximo" : 3 }
{ "_id" : { "review_score" : 1 }, "maximo" : 1 }
{ "_id" : { "review_score" : 5 }, "maximo" : 5 }
{ "_id" : { "review_score" : 2 }, "maximo" : 2 }
> db.encuestas.aggregate([{$group:{_id:"",last:{$max:"$review_score"}}}]);
{ "_id" : "", "last" : 5 }
> db.encuestas.aggregate([{$group:{_id:"",last:{$min:"$review_score"}}}]);
{ "_id" : "", "last" : 1 }
> db.productos.aggregate([{$group:{_id:"",last:{$min:"$product_weight_g"}}}]);
{ "_id" : "", "last" : 50 }
> db.productos.aggregate([{$group:{_id:"",last:{$avg:"$product_weight_g"}}}]);
{ "_id" : "", "last" : 2206.976988494247 }
>
```

## 5 ANÁLISIS DE LECTURA

**Observación:** Considerando el artículo: “The Definitive Guide to Graph Databases for the RDBMS Developer” de Neo4J. Compartido en las carpeta de lecturas recomendadas. Analice y responda cada pregunta en máximo 150 palabras:

1. ¿Cuáles son las limitaciones, que se pueden inferir de la lectura, para migrar los conjuntos de datos relacionales a NoSQL?

El cambio implica considerar una migración, para tener como referencia, una refactorización en código puede durar minutos, mientras que la refactorización de una base de datos puede tardar varias semanas. El factor tiempo es relevante porque se traduce en costos, dedicación de un equipo de TI para la migración, también el costo de oportunidad de hacer otras cosas con el mismo recurso disponible, agregando el retardo de poner al disfrute el cambio a quienes lo solicitaron (la evolución de la herramienta no va a la misma velocidad de la evolución del negocio).

2. ¿Cuáles limitaciones adicionales se deben considerar, a parte de las mencionadas en el artículo?

Antes de considerar limitaciones técnicas, el no identificar la pertinencia del cambio, es una limitación, independiente de cual sea la decisión (utilizar una base de datos SQL o una NoSQL).

Si se parte de que se va a incurrir en una migración de datos de SQL a NoSQL, el tiempo es un factor relevante, tiene impacto en costos.

Si se tiene una base de datos NoSQL y uno de los requisitos desde lo comercial o del negocio, implica fiabilidad y coherencia de los datos, los desarrolladores deben desarrollar el código y eso añade complejidad al sistema.

3. ¿Cuáles son las razones (criterios) que se deben considerar para migrar un conjunto de datos relacionados a NoSQL?

Partiendo de la necesidad de negocio, se pueden identificar los requerimientos tecnológicos, a partir de esta referencia se puede tomar la decisión de migrar o no migrar.

Si se requiere una consistencia robusta, una atomicidad y una integridad del dato robusta lo mejor es que sea un SQL, si en cambio se tiene una alta transaccionalidad, los datos en cierta medida se obtienen de muchas partes y no se requiere esa consistencia tan completa puede ser un escenario para migrar a NoSQL.