

# Crear una gran base de datos que conecte a los productores de café colombiano con clientes internacionales.

Leonardo Enrique Alvarado<sup>1</sup>, Jisel Otavo Madrigal<sup>2</sup>

<sup>1-2</sup>Facultad de Ingeniería y Ciencias Básicas

Universidad Central

Maestría en Analítica de Datos

Curso de Bases de Datos

Bogotá, Colombia

{<sup>1</sup>lalvaradoa1@ucentral.edu.co, <sup>2</sup>ootavom@ucentral.edu.co}

November 19, 2022

## Contents

<b>1</b>	<b>Introducción ((Max 250 Palabras) - (Primera entrega)</b>	<b>3</b>
<b>2</b>	<b>Características del proyecto de investigación (Max 500 Palabras)</b>	<b>3</b>
	- ()	
2.1	Estudio de nichos de venta de Cafe en el exterior. (Max 100 Palabras) - (Primera entrega) . . . . .	4
2.1.1	Información de variedades de café. (Max 100 Palabras) - (Primera entrega) . . . . .	4
2.1.2	Identificar los diferentes gustos de los consumidores de café en el exterior.(Max 100 Palabras) - (Primera entrega) .	4
2.1.3	Entregar la trazabilidad del café desde el productor hasta el consumidor.(Max 100 Palabras) - (Primera entrega) . . .	4
2.2	Alcance (Max 200 Palabras) - (Primera entrega) . . . . .	4
2.3	Pregunta de investigación (Max 100 Palabras) - (Primera entrega) .	4
2.4	Hipotesis (Max 100 Palabras) - (Primera entrega) . . . . .	5
<b>3</b>	<b>Reflexiones sobre el origen de datos e información (Max 400 Palabras) - (Primera entrega)</b>	<b>6</b>
3.1	¿Cual es el origen de los datos e información ? (Max 100 Palabras) - (Primera entrega) . . . . .	6
3.2	¿Cuales son las consideraciones legales o eticas del uso de la información? (Max 100 Palabras) - (Primera entrega) . . . . .	6

3.3	¿Cuales son los retos de la información y los datos que utilizara en la base de datos en terminos de la calidad y la consolidación? (Max 100 Palabras) - ( <i>Primera entrega</i> ) . . . . .	6
3.4	¿Que espera de la utilización de un sistema de Bases de Datos para su proyecto? (Max 100 Palabras) - ( <i>Primera entrega</i> ) . . . . .	6
<b>4</b>	<b>Diseño del Modelo de Datos del SMBD (Sistema Manejador de Bases de Datos)(<i>Primera entrega</i>)</b>	<b>7</b>
4.1	Características del SMBD (Sistema Manejador de Bases de Datos) para el proyecto ( <i>Primera entrega</i> ) . . . . .	7
4.2	Diagrama modelo de datos ( <i>Primera entrega</i> ) . . . . .	7
4.3	Imágenes de la Base de Datos ( <i>Primera entrega</i> ) . . . . .	7
4.4	Código SQL - lenguaje de definición de datos (DDL) ( <i>Primera entrega</i> ) . . . . .	10
4.5	Código SQL - Manipulación de datos (DML) ( <i>Primera entrega</i> ) . .	10
4.6	Código SQL + Resultados: Vistas ( <i>Primera entrega</i> ) . . . . .	10
4.7	Código SQL + Resultados: Triggers ( <i>Primera entrega</i> ) . . . . .	11
4.8	Código SQL + Resultados: Funciones ( <i>Primera entrega</i> ) . . . . .	12
4.9	Código SQL + Resultados: procedimientos almacenados ( <i>Primera entrega</i> ) . . . . .	13
<b>5</b>	<b>Bases de Datos No-SQL (<i>Segunda entrega</i>)</b>	<b>15</b>
5.1	Diagrama Bases de Datos No-SQL ( <i>Segunda entrega</i> ) . . . . .	15
5.2	SMBD utilizado para la Base de Datos No-SQL ( <i>Segunda entrega</i> )	15
<b>6</b>	<b>Aplicación de ETL (Extract, Transform, Load) y Bodega de Datos (<i>Tercera entrega</i>)</b>	<b>18</b>
6.1	Ejemplo de aplicación de ETL y Bodega de Datos ( <i>Tercera entrega</i> )	18
<b>7</b>	<b>Lecciones aprendidas (<i>Tercera entrega</i>)</b>	<b>19</b>
<b>8</b>	<b>Bibliografía</b>	<b>20</b>

## 1 Introducción ((Max 250 Palabras) - (Primera entrega)

El café en Colombia ha sido y seguirá siendo importante motor de desarrollo. Se estima que para el año 2022 el valor que recibirá el país equivale a 14 billones de pesos, una cifra nunca vista ni siquiera en la época de oro del sector. El Espectador. Valor de la cosecha del café colombiano será histórica este 2022. Roberto Velez.10 octubre 2022. El mercado exterior reconoce a Colombia como un productor de café de excelente calidad y con un crecimiento en los denominados cafés especiales. Entonces no se puede hablar de café colombiano sino de un amplio rango de productos que cautivan el paladar de los consumidores, no solo localmente sino internacionalmente. No se parecen en nada la tasa de un café de la Sierra Nevada con un café de Nariño, un café de Norte de Santander con un café del Huila, un café de putumayo con un café de Antioquia. Son cosas totalmente distintas estando dentro de un paraguas de altísima calidad.

Entender estas diferencias es de vital importancia en el mercado exterior, desde el punto de vista del gusto consumidor para tratar de llegar con un café que cumpla con las expectativas de los clientes.

Aquí es donde el manejo adecuado de la base de datos es fundamental, ya que contar con las referencias del café producido, ubicación de las fincas, tipo de suelo y características del café, nos proporcionan la ayuda necesaria para conectar las fincas colombianas con clientes del exterior.

Esto significa ir en el sentido de los cafés especiales, tan en boga en el mercado internacional. según palabras del gerente de la federación de cafeteros Roberto Vélez ” Yo tengo una frase como de directorio telefónico que dice: hay un café colombiano para cada calidad y hay un paladar para cada café colombiano”.

Con ese objetivo nuestro proyecto pretende específicamente buscar esos nichos de mercado donde podríamos destacarnos y que toda la producción que tenemos hoy en Colombia pueda ser mercadeada en el exterior generando mayores ingresos para el país, ya que actualmente muchas cosechas no son vendidas en el exterior sino en Colombia con un diferencial importante de utilidad.

## 2 Características del proyecto de investigación (Max 500 Palabras) - ()

Primera entrega El proyecto tiene como característica principal la búsqueda de nuevos mercados para las diversas variedades de café producido en Colombia. Las diferentes condiciones de los suelos y por ende de las plantas de café que allí se cultivan son características únicas que los caficultores colombianos deberían aprovechar.

Dar a conocer a los clientes extranjeros los diferentes cafés tipo exportación tales como: café verde, café tostado, café soluble o Instantáneo, Café Liofilizado, las diferentes variedades y las fincas en las cuales están son producidas, estrechando lazos entre el consumidor y el productor, realizando fidelización y ayudando al productor a conocer a su cliente, prediciendo su comportamiento de compra y sus gustos.

## **2.1 Estudio de nichos de venta de Café en el exterior. (Max 100 Palabras) - (*Primera entrega*)**

Con la información referente a las preferencias de los consumidores o compradores en el exterior se puede ofrecer el tipo de café que mejor cumpla con esas expectativas. De esta forma se generan oportunidades de negocio que buscamos sean eficientemente aprovechadas.

### **2.1.1 Información de variedades de café. (Max 100 Palabras) - (*Primera entrega*)**

Identificar las variedades de café disponibles y las regiones donde existe mayor producción de una variedad que de otra para construir una base organizada con dicha información.

### **2.1.2 Identificar los diferentes gustos de los consumidores de café en el exterior.(Max 100 Palabras) - (*Primera entrega*)**

Conocer el gusto o las preferencias de los mercados consumidores de café para diseñar una base de datos que permita su adecuada relación con determinado tipo de café.

### **2.1.3 Entregar la trazabilidad del café desde el productor hasta el consumidor.(Max 100 Palabras) - (*Primera entrega*)**

Conseguir un manejo de la información desde el origen del café hasta su destino final.

## **2.2 Alcance (Max 200 Palabras) - (*Primera entrega*)**

Consideramos que el alcance cubre a todos los caficultores sin importar su tamaño. Sin embargo, es muy probable que sea de mayor utilidad para los pequeños cultivadores que por su volumen de producción se ven forzados a entregar su producción a un intermediario. Esto abre las puertas del mercado internacional a productores que consideran hoy impensable exportar su café. El proyecto podría aumentar las exportaciones de café colombiano en 10 por ciento a todo el mundo, llevando al consumidor un producto direccionado según sus gusto, evitando la pérdida de tiempo en largas esperas y cataciones.

## **2.3 Pregunta de investigación (Max 100 Palabras) - (*Primera entrega*)**

¿Cuáles son los diferentes gustos de los consumidores de café en el exterior? ¿Es posible que con la información que se tiene llegar a relacionar determinada variedad de café con el gusto de los consumidores en el exterior?

## **2.4 Hipotesis (Max 100 Palabras) - (*Primera entrega*)**

Si contamos con información suficiente para determinar el gusto del consumidor podremos llegar con un café que satisfaga sus expectativas.

### **3 Reflexiones sobre el origen de datos e información** (Max 400 Palabras) - (*Primera entrega*)

Base de datos de variedades de café y región.

#### **3.1 ¿Cual es el origen de los datos e información ?** (Max 100 Palabras) - (*Primera entrega*)

La información se toma de varias fuentes de información como lo son las estadísticas de la federación nacional de cafeteros, datos abiertos Colombia, plataforma kaggle. Además de trabajo de campo en diferentes departamentos de Colombia.

#### **3.2 ¿Cuales son las consideraciones legales o eticas del uso de la información?** (Max 100 Palabras) - (*Primera entrega*)

Teniendo en cuenta que los datos a trabajar corresponde a datos publicos y que no comprometen el manejo de información sensible, consideramos que no existe ningun tipo de falta a la ética o al manejo de la información.

#### **3.3 ¿Cuales son los retos de la información y los datos que utilizara en la base de datos en terminos de la calidad y la consolidación?** (Max 100 Palabras) - (*Primera entrega*)

Encontrar bases de datos idoneas que nos permitan diferenciar los gustos de los consumidores de café en el exterior. Conseguir conectar de manera adecuada la base de datos de variedades de café con la base de gustos de los consumidores en el exterior.

#### **3.4 ¿Que espera de la utilización de un sistema de Bases de Datos para su proyecto?** (Max 100 Palabras) - (*Primera entrega*)

Lograr conectar adecuadamente la información para predecir que variedad cumple con las expectativas de cada mercado o cliente.

## 4 Diseño del Modelo de Datos del SMBD (Sistema Manejador de Bases de Datos) (Primera entrega)

### 4.1 Características del SMBD (Sistema Manejador de Bases de Datos) para el proyecto (Primera entrega)

### 4.2 Diagrama modelo de datos (Primera entrega)

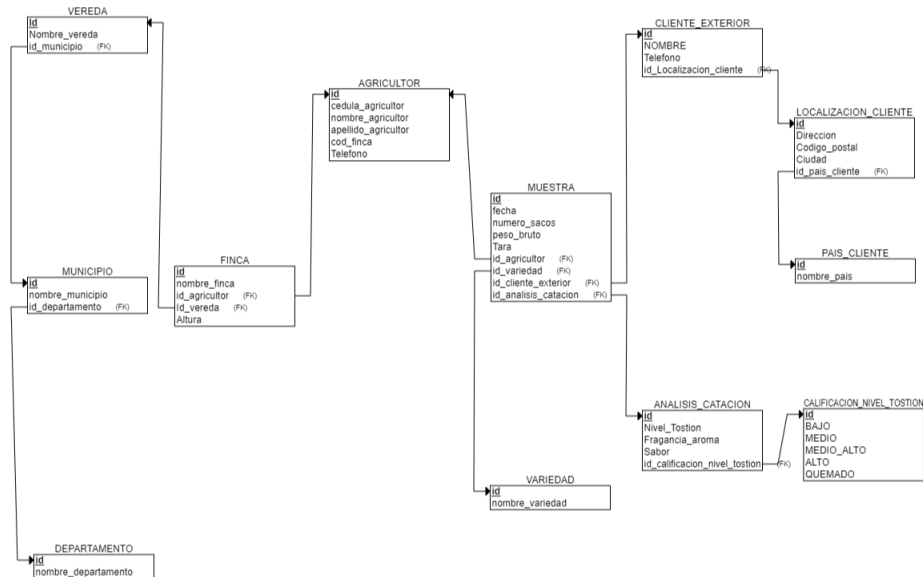


Figure 1: Caption

### 4.3 Imágenes de la Base de Datos (Primera entrega)

```
✓ CREATE TABLE public.cafe_departamento (  
    id bigint NOT NULL,  
    nombre character varying(45) NOT NULL  
);
```

Figure 2: Caption

```
INSERT INTO public.cafe_departamento (id, nombre) VALUES (1, 'CAUCA');  
INSERT INTO public.cafe_departamento (id, nombre) VALUES (2, 'HUILA');  
INSERT INTO public.cafe_departamento (id, nombre) VALUES (3, 'IBAGUE');  
INSERT INTO public.cafe_departamento (id, nombre) VALUES (4, 'NARIÑO');  
INSERT INTO public.cafe_departamento (id, nombre) VALUES (5, 'TOLIMA');  
INSERT INTO public.cafe_departamento (id, nombre) VALUES (6, 'VALLE DEL CAUCA');
```

Figure 3: Caption



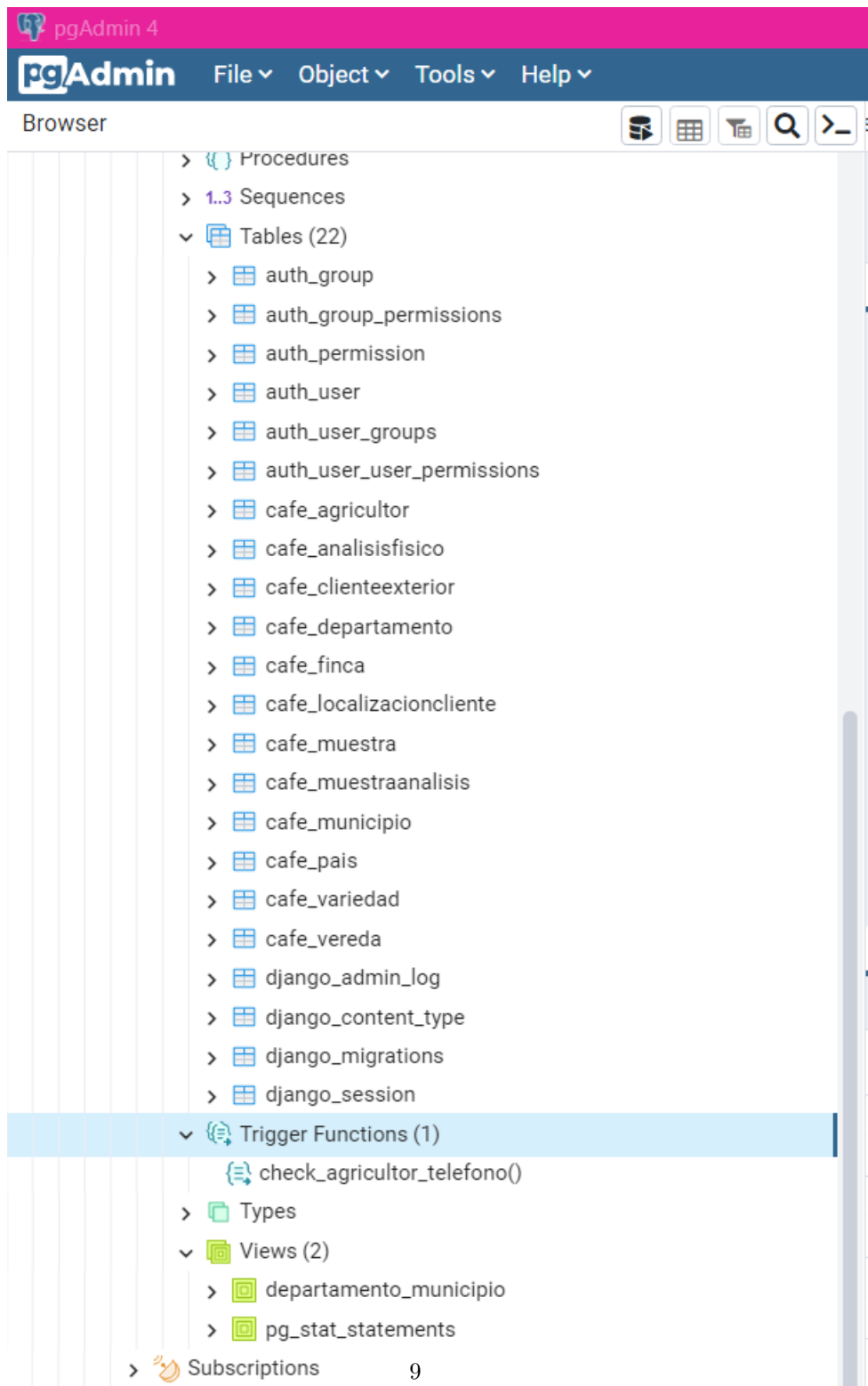


Figure 4: Caption

#### 4.4 Código SQL - lenguaje de definición de datos (DDL) (Primera entrega)

```
CREATE DATABASE ojnwbomj WITH TEMPLATE = template0
ENCODING = 'UTF8' LOCALE = 'en_US.UTF-8';
```

```
CREATE TABLE public.cafe_departamento (
    id bigint NOT NULL,
    nombre character varying(45) NOT NULL
);
```

#### 4.5 Código SQL - Manipulación de datos (DML) (Primera entrega)

```
INSERT INTO public.cafe_departamento (id, nombre) VALUES (1, 'CAUCA');
INSERT INTO public.cafe_departamento (id, nombre) VALUES (2, 'HUILA');
INSERT INTO public.cafe_departamento (id, nombre) VALUES (3, 'IBAGUE');
INSERT INTO public.cafe_departamento (id, nombre) VALUES (4, 'NARIÑO');
INSERT INTO public.cafe_departamento (id, nombre) VALUES (5, 'TOLIMA');
INSERT INTO public.cafe_departamento (id, nombre) VALUES (6, 'VALLE DEL CAUCA');
```

#### 4.6 Código SQL + Resultados: Vistas (Primera entrega)

```
CREATE OR REPLACE VIEW public.departamento_municipio
AS
SELECT cafe_departamento.id AS departamento_id,
    cafe_departamento.nombre AS departamento,
    cafe_municipio.id AS municipio_id,
    cafe_municipio.nombre AS municipio
FROM cafe_departamento
    JOIN cafe_municipio ON cafe_departamento.id = cafe_municipio.departamento_id;
```

ojnwbomj/ojn... ojnwbomj/ojn... ojnwbomj/ojnwbomj@curso bd\*

ojnwbomj/ojnwbomj@curso bd

Query Query History

```

1  -- View: public.departamento_municipio
2
3  -- DROP VIEW public.departamento_municipio;
4
5  CREATE OR REPLACE VIEW public.departamento_municipio
6  AS
7  SELECT cafe_departamento.id AS departamento_id,
8         cafe_departamento.nombre AS departamento,
9         cafe_municipio.id AS municipio_id,
10        cafe_municipio.nombre AS municipio
11  FROM cafe_departamento
12        JOIN cafe_municipio ON cafe_departamento.id = cafe_municipio.departamento_id;
13
14  ALTER TABLE public.departamento_municipio
15        OWNER TO ojnwbomj;
16
17  select * from departamento_municipio

```

Data output Messages Notifications

	departamento_id bigint	departamento character varying (45)	municipio_id bigint	municipio character varying (45)
1	2	HUILA	1	ACEVEDO
2	2	HUILA	2	COLOMBIA
3	1	CAUCA	3	CORINTO/CALOTO
4	4	NARIÑO	4	EL TAMBO
5	5	TOLIMA	5	IBAGUE
6	2	HUILA	6	PALESTINA
7	2	HUILA	7	PITAL
8	2	HUILA	8	PITALITO

Total rows: 14 of 14 Query complete 00:00:01.674

Figure 5: Caption

#### 4.7 Código SQL + Resultados: Triggers (*Primera entrega*)

```

CREATE TRIGGER add_agricultor_telefono
BEFORE INSERT OR UPDATE
ON cafe_agricultor

```

```

FOR EACH ROW
EXECUTE PROCEDURE check_agricultor_telefono();

insert into cafe_agricultor (nombre, apellido, telefono)
values
('jisel', 'otavo', '3197117438'),
('sergio', 'otavo', '8979297');

select * from cafe_agricultor;

```

#### 4.8 Código SQL + Resultados: Funciones (*Primera entrega*)

```

CREATE OR REPLACE FUNCTION check_agricultor_telefono()
RETURNS trigger AS
$$
BEGIN
    IF position('+ ' in NEW.telefono) = 0 THEN
    IF position('3' in NEW.telefono) = 1 THEN
    NEW.telefono = concat('+57', NEW.telefono);
    ELSE
    NEW.telefono = concat('+1', NEW.telefono);
    END IF;
    END IF;
    RETURN NEW;
END;
$$
LANGUAGE 'plpgsql';

```

The screenshot shows a PostgreSQL IDE interface. The top menu bar includes Dashboard, Properties, SQL, Statistics, Dependencies, Dependents, and Processes. The current session is named 'ojnwbomj/ojnwbomj@cafe\*'. Below the menu bar is a toolbar with various icons for file operations, query execution, and settings. The main area is divided into two tabs: 'Query' and 'Query History'. The 'Query' tab is active, displaying the following SQL code:

```

1 CREATE OR REPLACE FUNCTION check_agricultor_telefono()
2 RETURNS trigger AS
3 $$
4 BEGIN
5 IF position('+ ' in NEW.telefono) = 0 THEN
6 IF position('3 ' in NEW.telefono) = 1 THEN
7 NEW.telefono = concat('+57', NEW.telefono);
8 ELSE
9 NEW.telefono = concat('+1', NEW.telefono);
10 END IF;
11 END IF;
12 RETURN NEW;
13 END;
14 $$
15 LANGUAGE 'plpgsql';
16
17 CREATE TRIGGER add_agricultor_telefono
18 BEFORE INSERT OR UPDATE
19 ON cafe_agricultor

```

Below the SQL editor, there are three tabs: 'Data output', 'Messages', and 'Notifications'. The 'Data output' tab is active, showing a table with the following data:

	id [PK] bigint	nombre character varying (45)	apellido character varying (45)	telefono character varying (45)
1	1	jisel	otavo	+573197117438
2	2	sergio	otavo	+18979297

Figure 6: Caption

#### 4.9 Código SQL + Resultados: procedimientos almacenados (Primera entrega)

```

create or replace procedure nombre_agricultor_mayuscula(_id int)
language plpgsql
as $$
begin
    update cafe_agricultor set nombre=upper(nombre), apellido=upper(apellido)
    where id=_id;

    commit;
end;$$

```

```
call nombre_agricultor_mayuscula(1);
```

```
select * from cafe_agricultor;
```

The screenshot shows a database management interface with a top navigation bar (Dashboard, Properties, SQL, Statistics, Dependencies, Dependents, Processes) and a user profile (ojnwbomj/ojn...). Below the navigation bar is a toolbar with icons for file operations, filters, and execution. The main area is divided into two tabs: 'Query' and 'Query History'. The 'Query' tab is active, displaying a SQL script with line numbers 1 through 13. The script defines a procedure 'nombre\_agricultor\_mayuscula' and calls it with the argument 1, followed by a select statement. Below the query editor is a 'Data output' section with a toolbar for data manipulation. It displays a table with 5 columns: 'id [PK] bigint', 'nombre character varying (45)', 'apellido character varying (45)', and 'telefono character varying (45)'. The table contains two rows of data.

```
1 create or replace procedure nombre_agricultor_mayuscula(_id int)
2 language plpgsql
3 as $$
4 begin
5     update cafe_agricultor set nombre=upper(nombre), apellido=upper(apellido)
6     where id=_id;
7
8     commit;
9 end;$$
10
11 call nombre_agricultor_mayuscula(1);
12
13 select * from cafe_agricultor;
```

	id [PK] bigint	nombre character varying (45)	apellido character varying (45)	telefono character varying (45)
1	2	sergio	otavo	+18979297
2	1	JISEL	OTAVO	+573197117438

Figure 7: Caption

## 5 Bases de Datos No-SQL (*Segunda entrega*)

### 5.1 Diagrama Bases de Datos No-SQL (*Segunda entrega*)

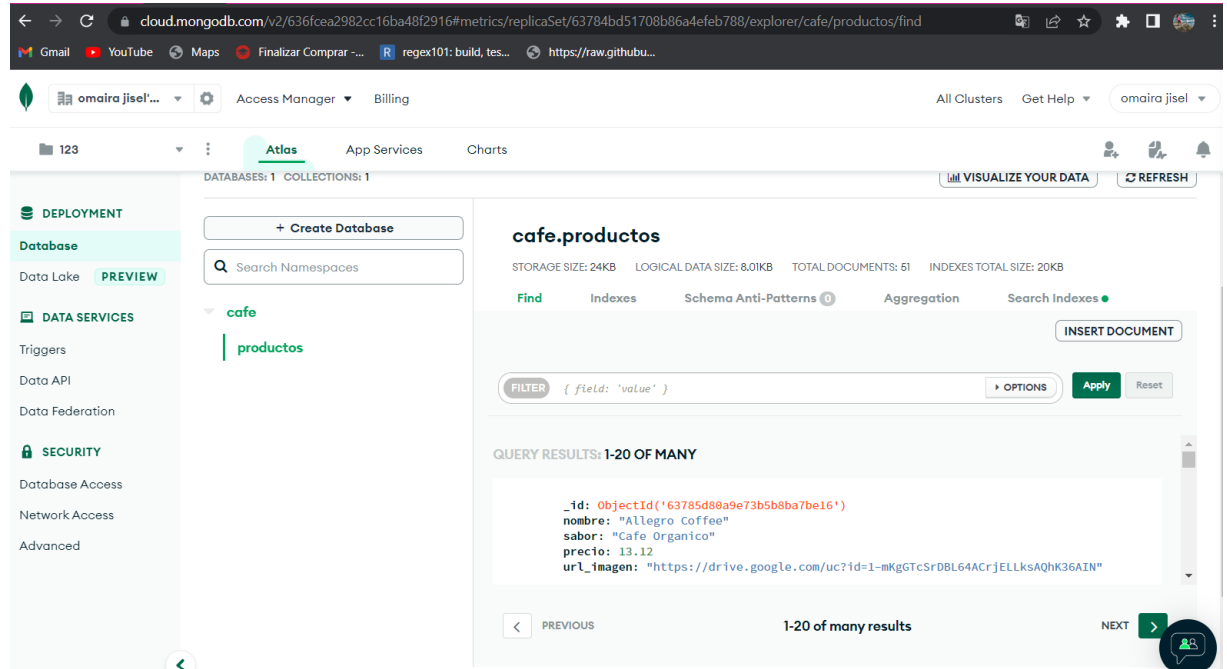


Figure 8: Caption

### 5.2 SMBD utilizado para la Base de Datos No-SQL (*Segunda entrega*)

Conexión y creación de la base de datos noSql en MongoDB

```
import pymongo
```

```
client = pymongo.MongoClient('mongodb+srv://curso:12341234@cursodb.zkdm5xs.mongodb.net/?retryWrites=true')
db = client.cafe
```

```
for product in db.productos.find({'nombre': {'$regex' : 'coffee', '$options' : 'i'}}):
    print(product)
```

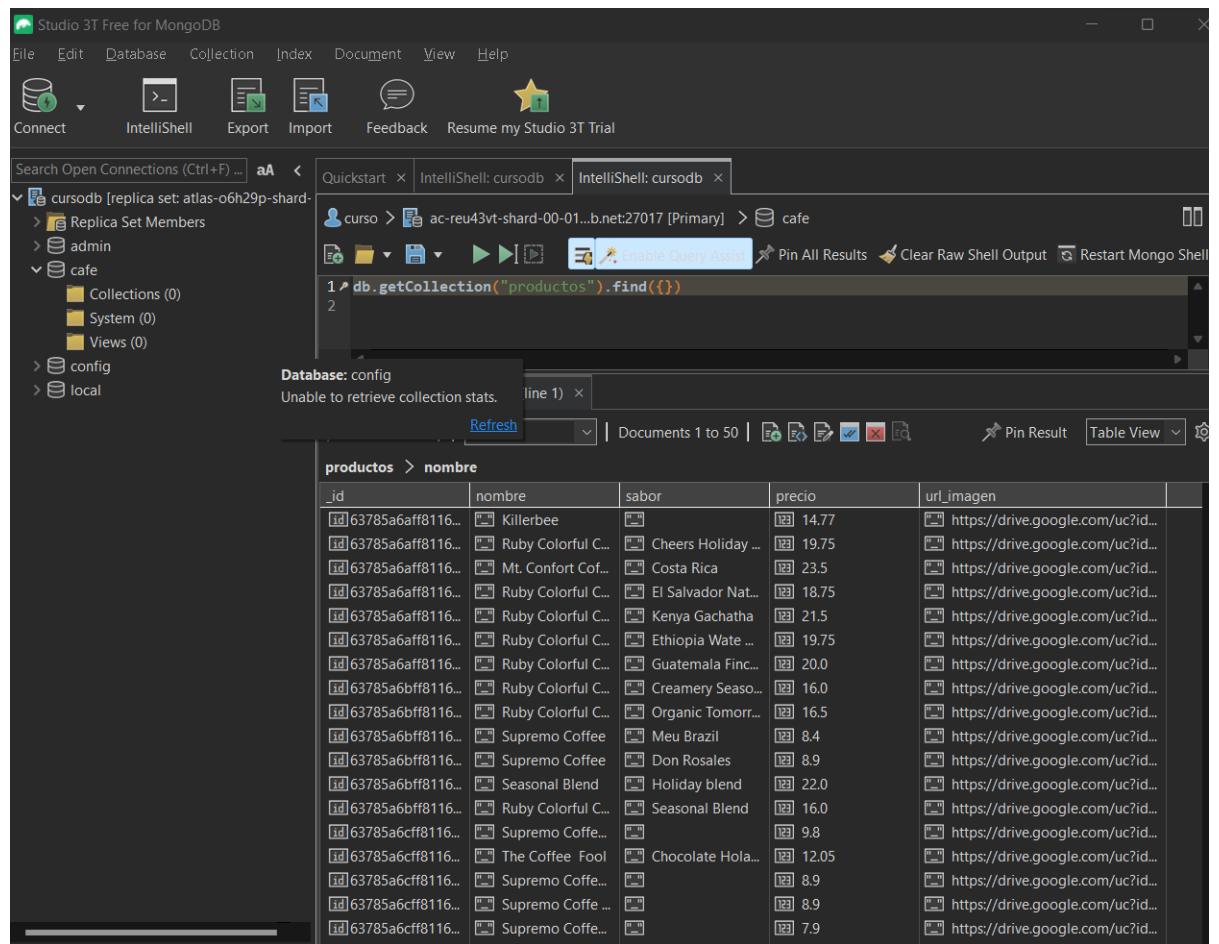


Figure 9: Caption

## Consultas NoSql

```
print('total productos: ', db.productos.count_documents({}))
print('con sabor: ', db.productos.count_documents({'sabor': {'$exists': True}}))
print('sin sabor: ', db.productos.count_documents({'sabor': {'$exists': False}}))
```



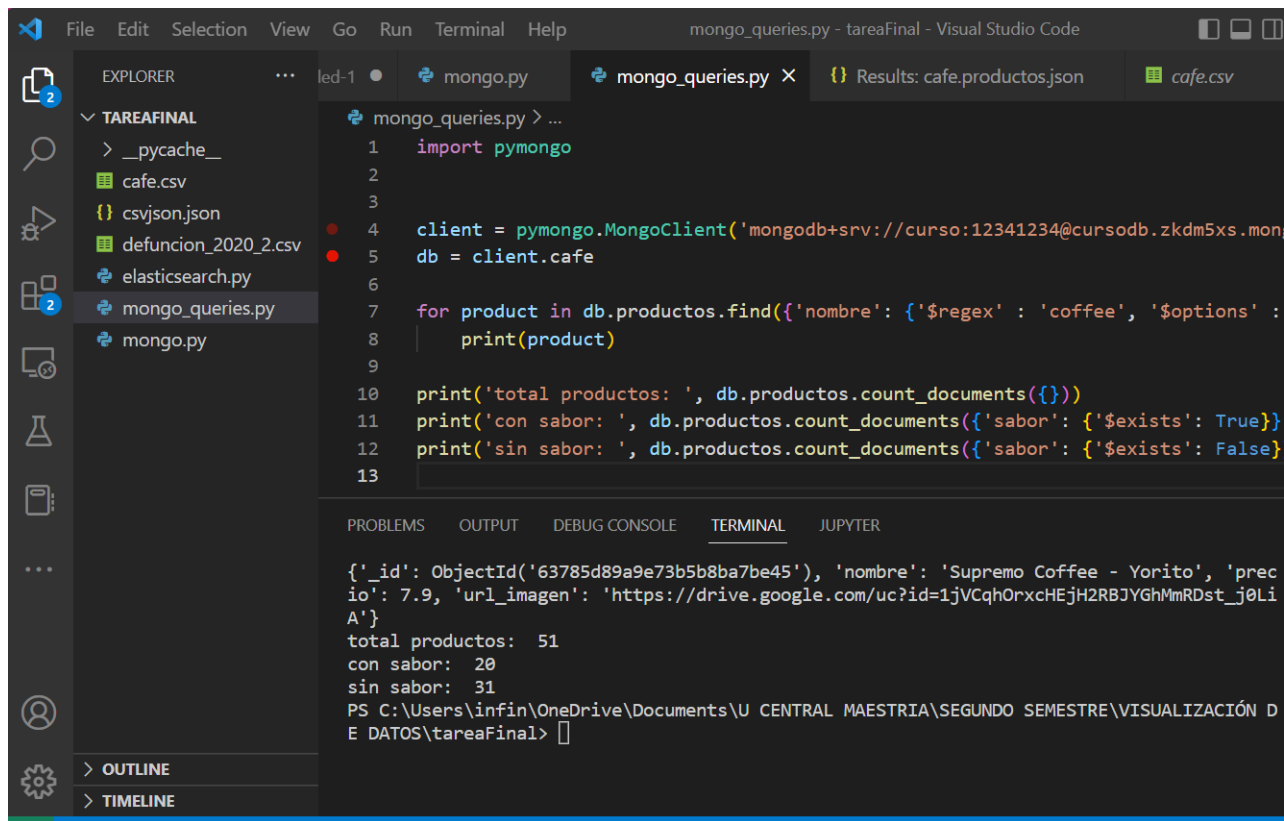


Figure 10: Caption

- 6    Aplicación de ETL (Extract, Transform, Load)  
     y Bodega de Datos** *(Tercera entrega)*
- 6.1   Ejemplo de aplicación de ETL y Bodega de Datos**  
     *(Tercera entrega)*

## 7 Lecciones aprendidas (*Tercera entrega*)

Principalmente podemos decir que aprendimos mucho sobre el lenguaje de consultas SQL, nos vamos con conocimientos en MySQL, PostgreSQL y Mongo. Además, mejoramos nuestro nivel de inglés y también nos llevamos un avance en nuestro trabajo de grado. Esta materia aparte de aprender bases de datos también nos deja muchas enseñanzas y más confianza en nosotros mismos, ya que en un comienzo pensábamos que era un reto muy difícil de alcanzar, pero a la fecha que se esta entregando el presente documento estamos muy felices del resultado y de satisfacción porque si se pudo con el reto y es lo que nos motiva a seguir avanzando en nuestro camino como magísteres en análisis de datos. Nuestros agradecimientos al profesor Wilmer Mesias Lopez Lopez por su entrega y paciencia a lo largo de este curso y a nuestros compañeros que con sus aportes nos ayudaron a mejorar nuestro trabajo.

## 8 Bibliografía

<https://federaciondecafeteros.org/wp/estadisticas-cafeteras/>  
<https://docs.djangoproject.com/en/4.1/ref/contrib/admin/>  
<https://www.todopostgresql.com/ejemplos-de-join-en-postgresql/>  
<https://erdplus.com/>  
<https://nokyotsu.com/latex/figuras.html>  
<https://www.elephantsql.com/>  
<https://cloud.mongodb.com/>  
<https://www.mongodb.com/docs/atlas/pause-terminate-cluster/>  
<https://realpython.com/python-csv/>  
<https://nisorness.com/data/2017/7/26/how-to-connect-to-mongodb-atlas-using-robo-3t-robomongo/>  
<https://studio3t.com/download-thank-you/?OS=win64>