# Proyecto de Análisis de Opiniones de Clientes

**Repositorio de Git hub:**

<https://github.com/Dextron03/Proyecto_Final_Electiva_1>

**Documentación breve:**

Este proyecto es una aplicación de consola .NET que implementa un proceso ETL (Extracción, Transformación y Carga) para archivos CSV relacionados con clientes, productos y encuestas. El objetivo principal es analizar información de distintas fuentes para obtener insights sobre clientes, productos y sus opiniones, respetando un modelo de datos bien definido.

**Script SQL con la definición de las tablas y sus relaciones:**

CREATE DATABASE OpinionDB

ON PRIMARY

(

NAME = N'OpinionDB\_Primary',

FILENAME = 'C:\Users\PC\Desktop\Tareas ITLA\Electiva 1 - Big Data\Proyecto Final\SAOCP\OpinionDB\_Primary.mdf',

SIZE = 10MB,

MAXSIZE = UNLIMITED,

FILEGROWTH = 5MB

),

FILEGROUP FG\_2024

(

NAME = 'OpinionDB\_2024',

FILENAME = 'C:\Users\PC\Desktop\Tareas ITLA\Electiva 1 - Big Data\Proyecto Final\SAOCP\Opinion2024\OpinionDB\_2024.ndf',

SIZE = 10MB,

MAXSIZE = UNLIMITED,

FILEGROWTH = 5MB

),

FILEGROUP FG\_2025

(

NAME = 'OpinionDB\_2025',

FILENAME = 'C:\Users\PC\Desktop\Tareas ITLA\Electiva 1 - Big Data\Proyecto Final\SAOCP\Opinion2025\OpinionDB\_2025.ndf',

SIZE = 10MB,

MAXSIZE = UNLIMITED,

FILEGROWTH = 5MB

),

FILEGROUP FG\_2026

(

NAME = 'OpinionDB\_2026',

FILENAME = 'C:\Users\PC\Desktop\Tareas ITLA\Electiva 1 - Big Data\Proyecto Final\SAOCP\Opinion2026\OpinionDB\_2026.ndf',

SIZE = 10MB,

MAXSIZE = UNLIMITED,

FILEGROWTH = 5MB

)

LOG ON

(

NAME = N'OpinionDB\_Log',

FILENAME = 'C:\Users\PC\Desktop\Tareas ITLA\Electiva 1 - Big Data\Proyecto Final\SAOCP\Logs\OpinionDB.ldf',

SIZE = 10MB,

MAXSIZE = 2GB,

FILEGROWTH = 5MB

);

USE OpinionDB;

CREATE TABLE Clientes (

IdCliente INT IDENTITY(1,1) PRIMARY KEY,

Nombre NVARCHAR(100) NOT NULL,

Email NVARCHAR(150) UNIQUE NOT NULL

);

CREATE TABLE Categorias (

IdCategoria INT PRIMARY KEY IDENTITY(1,1),

Nombre NVARCHAR(100) UNIQUE NOT NULL

);

CREATE TABLE Productos (

IdProducto INT PRIMARY KEY,

Nombre NVARCHAR(100) NOT NULL,

IdCategoria INT,

CONSTRAINT FK\_Productos\_Categoria FOREIGN KEY (IdCategoria) REFERENCES Categorias(IdCategoria)

);

CREATE TABLE TiposFuente (

IdTipoFuente INT PRIMARY KEY IDENTITY(1,1),

Nombre NVARCHAR(50) UNIQUE NOT NULL

);

CREATE TABLE FuenteDatos (

IdFuente VARCHAR(10) PRIMARY KEY,

FechaCarga DATE NOT NULL,

IdTipoFuente INT,

CONSTRAINT FK\_FuenteDatos\_TipoFuente FOREIGN KEY (IdTipoFuente) REFERENCES TiposFuente(IdTipoFuente)

);

CREATE TABLE Clasificaciones (

IdClasificacion INT PRIMARY KEY IDENTITY(1,1),

Nombre NVARCHAR(50) UNIQUE NOT NULL

);

CREATE PARTITION FUNCTION pf\_FechaRango (DATE)

AS RANGE LEFT FOR VALUES ('2024-12-31', '2025-12-31');

CREATE PARTITION SCHEME ps\_FechaRango

AS PARTITION pf\_FechaRango

TO (FG\_2024, FG\_2025, FG\_2026);

CREATE TABLE Comentarios (

IdComment VARCHAR(10) NOT NULL,

IdCliente INT NOT NULL,

IdProducto INT NOT NULL,

IdFuente VARCHAR(10) NOT NULL,

Fecha DATE NOT NULL,

Comentario NVARCHAR(MAX),

PRIMARY KEY NONCLUSTERED (IdComment, Fecha),

CONSTRAINT FK\_Comentarios\_Cliente FOREIGN KEY (IdCliente) REFERENCES Clientes(IdCliente),

CONSTRAINT FK\_Comentarios\_Producto FOREIGN KEY (IdProducto) REFERENCES Productos(IdProducto),

CONSTRAINT FK\_Comentarios\_Fuente FOREIGN KEY (IdFuente) REFERENCES FuenteDatos(IdFuente)

)

ON ps\_FechaRango(Fecha);

CREATE TABLE Encuestas (

IdOpinion INT NOT NULL,

IdCliente INT NOT NULL,

IdProducto INT NOT NULL,

Fecha DATE NOT NULL,

Comentario NVARCHAR(MAX),

IdClasificacion INT,

PuntajeSatisfaccion INT CHECK (PuntajeSatisfaccion BETWEEN 1 AND 5),

IdFuente VARCHAR(10) NOT NULL,

PRIMARY KEY NONCLUSTERED (IdOpinion, Fecha),

CONSTRAINT FK\_Encuestas\_Clasificacion FOREIGN KEY (IdClasificacion) REFERENCES Clasificaciones(IdClasificacion),

CONSTRAINT FK\_Encuestas\_Cliente FOREIGN KEY (IdCliente) REFERENCES Clientes(IdCliente),

CONSTRAINT FK\_Encuestas\_Producto FOREIGN KEY (IdProducto) REFERENCES Productos(IdProducto),

CONSTRAINT FK\_Encuestas\_Fuente FOREIGN KEY (IdFuente) REFERENCES FuenteDatos(IdFuente)

)

ON ps\_FechaRango(Fecha);

CREATE TABLE WebReviews (

IdReview VARCHAR(10) NOT NULL,

IdCliente INT NOT NULL,

IdProducto INT NOT NULL,

Fecha DATE NOT NULL,

Comentario NVARCHAR(MAX),

Rating INT CHECK (Rating BETWEEN 1 AND 5),

IdFuente VARCHAR(10) NOT NULL,

PRIMARY KEY NONCLUSTERED (IdReview, Fecha),

CONSTRAINT FK\_WebReviews\_Cliente FOREIGN KEY (IdCliente) REFERENCES Clientes(IdCliente),

CONSTRAINT FK\_WebReviews\_Producto FOREIGN KEY (IdProducto) REFERENCES Productos(IdProducto),

CONSTRAINT FK\_WebReviews\_Fuente FOREIGN KEY (IdFuente) REFERENCES FuenteDatos(IdFuente)

)

ON ps\_FechaRango(Fecha);

**Diseño y modelado de la base de datos**

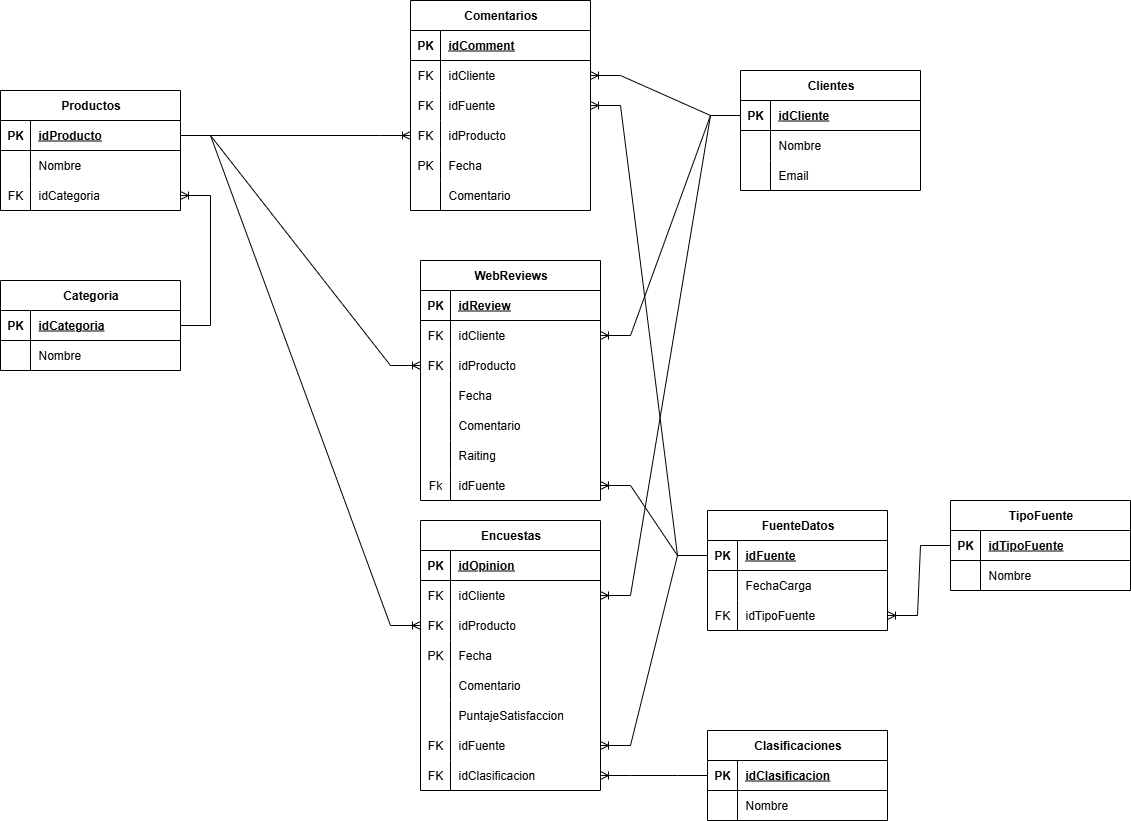
El proyecto utiliza la base de datos OpinionDB, diseñada para manejar información sobre clientes, productos, encuestas y comentarios de distintas fuentes. Se incluyen particiones por año y relaciones entre tablas para garantizar la integridad referencial.

**Tablas principales y su descripción**

* **Clientes**: Contiene la información de los clientes registrados en el sistema, incluyendo datos de contacto y referencia para encuestas y comentarios.
* **Categorias**: Almacena las categorías de productos, facilitando la clasificación y análisis por tipo de producto.
* **Productos**: Registra los productos disponibles, vinculados a su categoría correspondiente para permitir análisis segmentado.
* **TiposFuente**: Define los tipos de fuentes de datos de donde provienen las encuestas y comentarios (por ejemplo: web, redes sociales, CSV externo).
* **FuenteDatos**: Contiene la información de los diferentes conjuntos de datos cargados, incluyendo su fecha de carga y tipo de fuente.
* **Clasificaciones**: Lista los tipos de clasificación posibles para encuestas y comentarios, como “positivo”, “neutral” o “negativo”.
* **Comentarios**: Almacena los comentarios de clientes sobre productos, provenientes de distintas fuentes, con soporte para particionamiento por fecha para mejorar la gestión histórica.
* **Encuestas**: Registra los resultados de encuestas de satisfacción realizadas por los clientes sobre productos, incluyendo la clasificación y el puntaje de satisfacción.
* **WebReviews**: Contiene reseñas de productos extraídas de la web o redes sociales, asociadas a clientes y productos específicos, y soporta particionamiento por fecha.

**Particionamiento de tablas**

Se definió una función de partición por fecha (pf\_FechaRango) y un esquema de partición (ps\_FechaRango) que asigna los datos de Comentarios, Encuestas y WebReviews a los filegroups de 2024, 2025 y 2026. Esto permite un manejo eficiente de grandes volúmenes de datos históricos.

**Diagrama Entidad-Relación (ERD)**