



## **Presentación**

### **Nombre:**

Xavier Casilla |Mat. 2023-0995

### **Asignatura:**

Electiva 1: Big Data

### **Docente:**

Francis Ramírez

### **Tema:**

Diseño Preliminar: Sistema de Análisis de Opiniones de Clientes

### **Fecha:**

21/09/2025

# Sistema de Análisis de Opiniones de Clientes

## Definición del problema

Una empresa de comercio electrónico desea analizar las opiniones que los clientes dejan sobre sus productos en diferentes canales:

- **Archivos CSV:** encuestas internas de satisfacción.
- **Base de datos relacional:** reseñas publicadas en su sitio web.
- **API REST:** comentarios en redes sociales.

Actualmente, la información está dispersa y es difícil integrarla para generar indicadores de satisfacción y tendencias de opiniones.

El objetivo es desarrollar un sistema ETL en .NET que consolide, procese y analice estos datos, permitiendo generar indicadores clave como:

- Total de comentarios procesados.
- Clasificación de opiniones (positivas, negativas, neutras).
- Tendencia de satisfacción por producto.
- Cantidad de reseñas por producto y porcentaje de satisfacción

GitHub: [Xavi1324/SDOC--BigData](https://github.com/Xavi1324/SDOC--BigData)

## 1. Modelado de la Base de Datos

### Dimensiones:

#### Producto

- Product\_Id INT (PK, identity)
- ProductName VARCHAR(150) NOT NULL
- Categoria\_Id INT NOT NULL (FK → Categoria.Categoria\_Id)

### **Categoría**

- Categoria\_Id INT (PK, identity)
- Nombre VARCHAR(100) NOT NULL UNIQUE

### **Cliente**

- Client\_Id INT (PK, identity)
- ClientName VARCHAR(100) NOT NULL
- LastName VARCHAR(100) NOT NULL
- Email VARCHAR(200) NULL UNIQUE
- Pais\_Id INT NULL (FK → Pais.Pais\_Id)

### **Pais**

- Pais\_Id INT (PK, identity)
- Nombre VARCHAR(100) NOT NULL UNIQUE

### **Tiempo**

- Time\_Id INT (PK, identity)
- Date DATE NOT NULL
- Year INT NOT NULL
- Month TINYINT NOT NULL
- Day TINYINT NOT NULL
- Hour TINYINT NULL

### **TipoFuente**

- TipoFuente\_Id INT (PK, identity)
- Descripcion VARCHAR(100) NOT NULL UNIQUE

## **Fuente**

- Fuente\_Id INT (PK, identity)
- TipoFuente\_Id INT NOT NULL (FK → TipoFuente.TipoFuente\_Id)
- Nombre VARCHAR(100) NOT NULL
- UrlPath VARCHAR(300) NULL
- FechaRegistro DATE NOT NULL

## **Clasificación**

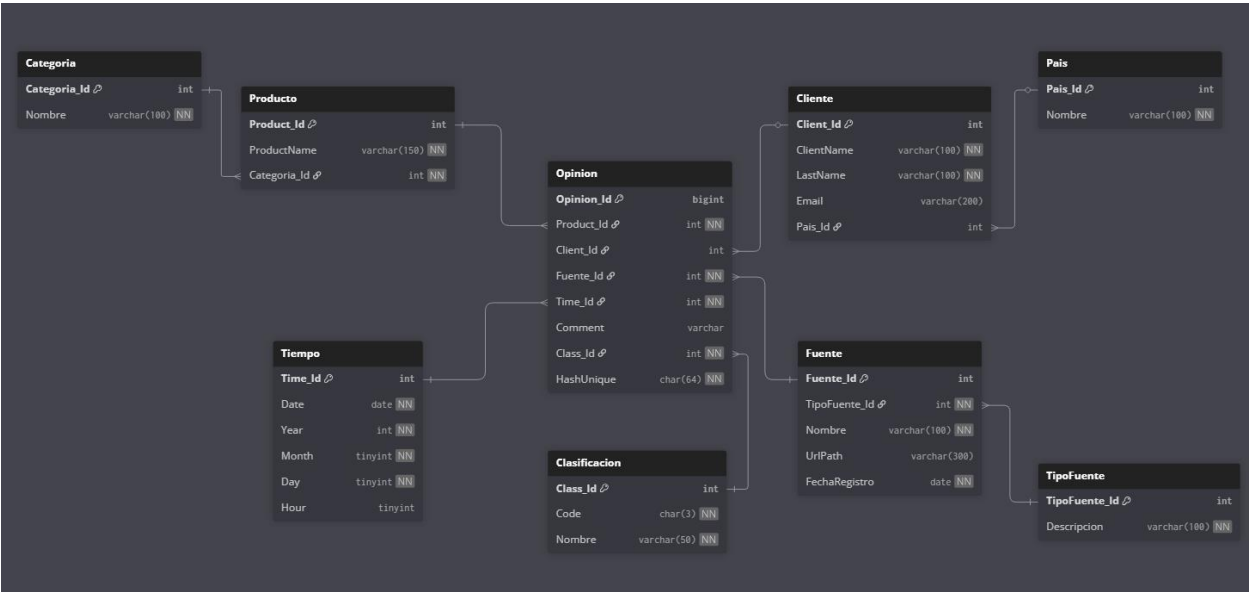
- Class\_Id SMALLINT (PK, identity)
- Code CHAR(3) NOT NULL UNIQUE (ej: POS, NEG, NEU)
- Nombre VARCHAR(50) NOT NULL

## **Hecho:**

## **Opinión**

- Opinion\_Id BIGINT (PK, identity)
- Product\_Id INT NOT NULL (FK → Producto.Product\_Id)
- Client\_Id INT NULL (FK → Cliente.Client\_Id)
- Fuente\_Id INT NOT NULL (FK → Fuente.Fuente\_Id)
- Time\_Id INT NOT NULL (FK → Tiempo.Time\_Id)
- Comment NVARCHAR(MAX) NULL
- Class\_Id SMALLINT NOT NULL (FK → Clasificacion.Class\_Id)
- HashUnique CHAR(64) NOT NULL UNIQUE

# Diagrama entidad relación



## Desarrollo del Pipeline ETL

### Extracción y transformación:

```
482 | 3485 | 474 | 1/22/2025 | Negativa | 1 | EncuestaInterna | No era lo que esperaba, muy decepcionado.
483 | 6852 | 287 | 1/2/2025 | Neutra | 3 | EncuestaInterna | Ni malo ni excelente, simplemente regular.
484 | 3626 | 80 | 1/24/2025 | Negativa | 1 | EncuestaInterna | El producto dejó de funcionar al poco tiempo de usarlo.
485 | 5102 | 67 | 5/6/2025 | Negativa | 1 | EncuestaInterna | El producto dejó de funcionar al poco tiempo de usarlo.
486 | 1585 | 584 | 12/26/2024 | Negativa | 2 | EncuestaInterna | El producto dejó de funcionar al poco tiempo de usarlo.
487 | 3087 | 969 | 6/27/2025 | Negativa | 1 | EncuestaInterna | Mala calidad, no lo recomiendo para nada.
488 | 6971 | 659 | 7/24/2025 | Negativa | 2 | EncuestaInterna | El producto dejó de funcionar al poco tiempo de usarlo.
489 | 9147 | 858 | 8/7/2025 | Negativa | 2 | EncuestaInterna | El producto llegó dañado y no cumplió con lo prometido.
490 | 3187 | 271 | 4/26/2025 | Negativa | 2 | EncuestaInterna | No era lo que esperaba, muy decepcionado.
491 | 4808 | 844 | 3/24/2025 | Neutra | 3 | EncuestaInterna | El envío fue normal, dentro del tiempo esperado.
492 | 1707 | 637 | 9/17/2024 | Positiva | 4 | EncuestaInterna | Me encanta, es justo lo que necesitaba.
493 | 23 | 989 | 12/27/2024 | Negativa | 1 | EncuestaInterna | El producto llegó dañado y no cumplió con lo prometido.
494 | 1829 | 702 | 3/29/2025 | Positiva | 4 | EncuestaInterna | Excelente calidad, definitivamente volvería a comprar.
495 | 8255 | 550 | 1/12/2025 | Negativa | 1 | EncuestaInterna | El producto llegó dañado y no cumplió con lo prometido.
496 | 2257 | 480 | 3/26/2025 | Positiva | 4 | EncuestaInterna | El producto superó mis expectativas, muy recomendable.
497 | 2487 | 609 | 4/5/2025 | Negativa | 2 | EncuestaInterna | El producto llegó dañado y no cumplió con lo prometido.
498 | 6079 | 90 | 1/28/2025 | Positiva | 4 | EncuestaInterna | El servicio fue rápido y eficiente, estoy muy satisfecho.
499 | 9732 | 140 | 6/13/2025 | Negativa | 1 | EncuestaInterna | Mala calidad, no lo recomiendo para nada.
500 | 9572 | 96 | 3/1/2025 | Neutra | 3 | EncuestaInterna | El producto está bien, pero esperaba un poco más de calidad.

[Transformación] Surveys
Leídos: 500 | Limpios: 6 | Descartados: 494
Archivo transformado guardado en carpeta 'output/'.
```

# Carga de datos

## Clientes

	Client_Id	ClientName	LastName	Email	Pais_Id
482	482	Cliente_4...	N/A	cliente482@mail.c...	NULL
483	483	Cliente_4...	N/A	cliente483@mail.c...	NULL
484	484	Cliente_4...	N/A	cliente484@mail.c...	NULL
485	485	Cliente_4...	N/A	cliente485@mail.c...	NULL
486	486	Cliente_4...	N/A	cliente486@mail.c...	NULL
487	487	Cliente_4...	N/A	cliente487@mail.c...	NULL
488	488	Cliente_4...	N/A	cliente488@mail.c...	NULL
489	489	Cliente_4...	N/A	cliente489@mail.c...	NULL
490	490	Cliente_4...	N/A	cliente490@mail.c...	NULL
491	491	Cliente_4...	N/A	cliente491@mail.c...	NULL
492	492	Cliente_4...	N/A	cliente492@mail.c...	NULL
493	493	Cliente_4...	N/A	cliente493@mail.c...	NULL
494	494	Cliente_4...	N/A	cliente494@mail.c...	NULL
495	495	Cliente_4...	N/A	cliente495@mail.c...	NULL
496	496	Cliente_4...	N/A	cliente496@mail.c...	NULL
497	497	Cliente_4...	N/A	cliente497@mail.c...	NULL
498	498	Cliente_4...	N/A	cliente498@mail.c...	NULL
499	499	Cliente_4...	N/A	cliente499@mail.c...	NULL
500	500	Cliente_5...	N/A	cliente500@mail.c...	NULL

## Opinión

	Opinion_Id	Product_Id	Client_Id	Fuente_Id	Time_Id	Comment	Class_Id	HashUnique
188	188	7	3	102	141	Producto recibido, cumple su función. OK.	1	adcb3e6090cb1f1faf0e8c936f2919e5750bdcd4f51b6170b0...
189	189	17	26	102	142	Satisface lo básico. OK.	1	663566ce82695644ba7d84740ac2da1c32a12e539ad0dafd...
190	190	2	13	102	92	Me encanta este producto, excelente calidad. Muy ...	1	628b83cbcd494d15648f52753b12086d498ab768af29048f...
191	191	9	8	102	143	Entrega correcta, sin comentarios adicionales. OK.	3	4bf17ee6338a3e948bd5285e50ae65405c03159ba6e0df312...
192	192	13	22	102	144	Entrega correcta, sin comentarios adicionales. OK.	3	e4ee6b807596688f0069594a1d0b9beefb0ab3171a672839f...
193	193	12	34	102	17	Me encanta este producto, excelente calidad. Muy ...	2	4b35d3559292be5477bee530460f6eb08487ac521fb6a49f...
194	194	14	39	102	138	Información suficiente, sin mayor novedad. OK.	1	7faf31fe05686d4e04cde421543f9f967a65e5a07b7563894c...
195	195	11	7	102	16	Producto llegó rápido y funciona perfecto. Muy reco...	3	5e5a4b5e15d188d8f2154b1705362f0fa7b531c67dfc0a654...
196	196	20	28	102	99	Calidad superior, muy contento. Muy recomendable.	3	339a7cad9785d337c64244df6e7c005bde64ab341445b540...
197	197	2	23	102	66	No cumple con lo anunciado, insatisfecho. No volv...	2	5a979db6432d2856aad2a55d3af1e3802ab1a23c7644e4ba...
198	198	3	26	102	145	Muy satisfecho con la compra, lo recomiendo. Muy ...	2	8cfa93032a225a2247eb66441d737fa94756776a8cdfb85e6f...
199	199	20	29	102	57	Me encanta este producto, excelente calidad. Muy ...	1	6cd0159c83bc6bd0ddce3d893398abc9f5733ba8891e3041...
200	200	12	25	102	146	Producto recibido, cumple su función. OK.	2	b84f7079618dcdff73f0a53b81cb1aed402f05b079bf6889ec6...
201	201	15	24	102	5	Me encanta este producto, excelente calidad. Muy ...	2	6ccc424cd8b6a4ff13f63d9102d44e35213a9ff7da1beb144d...
202	202	17	24	102	147	Producto llegó rápido y funciona perfecto. Muy reco...	2	48969224d3aa2a76555a893220a39cfce2b4477f9cdf9eb51...
203	203	1	12	102	97	No cumple con lo anunciado, insatisfecho. No volv...	2	ae215c72487f59c922d1b4d13d45924c21481ec89c1b77c7...
204	204	9	17	102	148	Información suficiente, sin mayor novedad. OK.	1	412c6fd8855a4f422410f8a571c7a0d1cd616e40c630fe25c6...
205	205	16	13	102	149	Me encanta este producto, excelente calidad. Muy ...	2	714ad9c19dc824b1fca6a64955dffe873d5e92f6d81e2eb691...
206	206	15	6	102	150	Es lo que esperaba, nada excepcional. OK.	1	644c485a942fcae9ab6889ca6a1fbb770fa844aa374710e2ff...

## Tiempo

	Time_Id	Date	Year	Month	Day	Hour	DateKey
132	132	2024-09-22	2024	9	22	NULL	20240922
133	133	2025-09-04	2025	9	4	NULL	20250904
134	134	2024-12-17	2024	12	17	NULL	20241217
135	135	2025-08-20	2025	8	20	NULL	20250820
136	136	2025-05-13	2025	5	13	NULL	20250513
137	137	2025-03-06	2025	3	6	NULL	20250306
138	138	2024-12-08	2024	12	8	NULL	20241208
139	139	2024-10-25	2024	10	25	NULL	20241025
140	140	2025-08-26	2025	8	26	NULL	20250826
141	141	2025-04-05	2025	4	5	NULL	20250405
142	142	2025-01-14	2025	1	14	NULL	20250114
143	143	2025-02-21	2025	2	21	NULL	20250221
144	144	2024-09-24	2024	9	24	NULL	20240924
145	145	2025-08-29	2025	8	29	NULL	20250829
146	146	2024-11-08	2024	11	8	NULL	20241108
147	147	2025-05-09	2025	5	9	NULL	20250509
148	148	2024-11-04	2024	11	4	NULL	20241104
149	149	2025-01-25	2025	1	25	NULL	20250125
150	150	2024-10-07	2024	10	7	NULL	20241007

Query executed successfully

## TipoFuente

150 %

Results

Messages

	TipoFuente_Id	Descripcion
1	2	Csv
2	5	Ecommerce
3	4	Encuesta
4	3	Red Social
5	1	Web

## Productos

	Product_Id	ProductName	Categoria_Id
182	82	Producto_82	1
183	83	Producto_83	5
184	84	Producto_84	3
185	85	Producto_85	4
186	86	Producto_86	4
187	87	Producto_87	3
188	88	Producto_88	1
189	89	Producto_89	1
190	9	Producto_9	4
191	90	Producto_90	1
192	91	Producto_91	1
193	92	Producto_92	1
194	93	Producto_93	5
195	94	Producto_94	1
196	95	Producto_95	3
197	96	Producto_96	1
198	97	Producto_97	1
199	98	Producto_98	5
200	99	Producto_99	1

## Categoría

	Categoria_Id	Nombre
1	5	Deportes
2	2	Electronica
3	4	Hogar
4	1	Juguetes
5	3	Ropa



Fuente

150 %

ResultsMessages

	Fuente_Id	TipoFuente_Id	Nombre	UrlPath	FechaRegistro
84	84	3	F084	NULL	2025-03-28
85	85	1	F085	NULL	2025-04-18
86	86	2	F086	NULL	2025-04-15
87	87	1	F087	NULL	2025-02-03
88	88	3	F088	NULL	2025-04-02
89	89	1	F089	NULL	2025-04-13
90	90	1	F090	NULL	2025-05-29
91	91	1	F091	NULL	2025-08-24
92	92	3	F092	NULL	2025-05-29
93	93	1	F093	NULL	2025-07-13
94	94	3	F094	NULL	2025-08-15
95	95	3	F095	NULL	2025-01-01
96	96	3	F096	NULL	2025-02-04
97	97	2	F097	NULL	2025-03-22
98	98	3	F098	NULL	2025-03-19
99	99	1	F099	NULL	2025-06-14
100	100	1	F100	NULL	2025-04-02
101	101	4	Encuestainterna	NULL	2024-12-19
102	102	5	WebReviews	NULL	2024-10-23

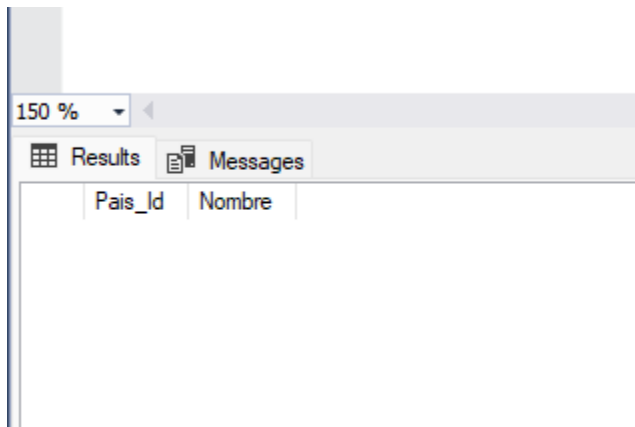
Calificación

150 %

ResultsMessages

	Class_Id	Code	Nombre
1	1	NEU	NEU
2	2	NEG	NEG
3	3	POS	POS

## País



150 %

Results Messages

Pais_Id	Nombre
---------	--------

## Script SQL – Base de dato

### Tablas:

```
CREATE TABLE dw.Categoria (  
    Categoria_Id INT IDENTITY(1,1) PRIMARY KEY,  
    Nombre VARCHAR(100) NOT NULL  
);  
  
CREATE TABLE dw.Pais (  
    Pais_Id INT IDENTITY(1,1) PRIMARY KEY,  
    Nombre VARCHAR(100) NOT NULL  
);  
  
CREATE TABLE dw.TipoFuente (  
    TipoFuente_Id INT IDENTITY(1,1) PRIMARY KEY,  
    Descripcion VARCHAR(100) NOT NULL  
);  
  
CREATE TABLE dw.Clasificacion (  
    Class_Id SMALLINT IDENTITY(1,1) PRIMARY KEY,  
    Code CHAR(3) NOT NULL,  
    Nombre VARCHAR(50) NOT NULL  
);  
  
CREATE TABLE dw.Producto (  
    Product_Id INT IDENTITY(1,1) PRIMARY KEY,  
    ProductName VARCHAR(150) NOT NULL,  
    Categoria_Id INT NOT NULL  
);
```

```

CREATE TABLE dw.Cliente (
    Client_Id INT IDENTITY(1,1) PRIMARY KEY,
    ClientName VARCHAR(100) NOT NULL,
    LastName VARCHAR(100) NOT NULL,
    Email VARCHAR(200) NULL,
    Pais_Id INT NULL
);

CREATE TABLE dw.Tiempo (
    Time_Id INT IDENTITY(1,1) PRIMARY KEY,
    [Date] DATE NOT NULL UNIQUE,
    [Year] INT NOT NULL,
    [Month] TINYINT NOT NULL,
    [Day] TINYINT NOT NULL,
    [Hour] TINYINT NULL,
    DateKey AS ([Year]*10000 + [Month]*100 + [Day]) PERSISTED
);

CREATE TABLE dw.Fuente (
    Fuente_Id INT IDENTITY(1,1) PRIMARY KEY,
    TipoFuente_Id INT NOT NULL,
    Nombre VARCHAR(100) NOT NULL,
    UrlPath VARCHAR(300) NULL,
    FechaRegistro DATE NOT NULL
);

```

```

CREATE TABLE dw.Opinion (
    Opinion_Id BIGINT IDENTITY(1,1) NOT NULL PRIMARY KEY,
    Product_Id INT NOT NULL,
    Client_Id INT NULL,
    Fuente_Id INT NOT NULL,
    Time_Id INT NOT NULL,
    [Comment] NVARCHAR(MAX) NOT NULL,
    Class_Id SMALLINT NOT NULL,
    HashUnique CHAR(64) NOT NULL UNIQUE
);

```

## Claves foráneas:

```
Roles&Index.sql - ...L28HHF\xavie (77) FOREIGNKEYS.sql - ...L28HHF\xavie (75) DatabaseCreate.sql...L28HHF\xavie (54)
ALTER TABLE dw.Producto
  ADD CONSTRAINT FK_Producto_Categoria
  FOREIGN KEY (Categoria_Id) REFERENCES dw.Categoria(Categoria_Id);

ALTER TABLE dw.Cliente
  ADD CONSTRAINT FK_Cliente_Pais
  FOREIGN KEY (Pais_Id) REFERENCES dw.Pais(Pais_Id);

ALTER TABLE dw.Fuente
  ADD CONSTRAINT FK_Fuente_Tipo
  FOREIGN KEY (TipoFuente_Id) REFERENCES dw.TipoFuente(TipoFuente_Id);

ALTER TABLE dw.Opinion
  ADD CONSTRAINT FK_Opinion_Producto
  FOREIGN KEY (Product_Id) REFERENCES dw.Producto(Product_Id);

ALTER TABLE dw.Opinion
  ADD CONSTRAINT FK_Opinion_Cliente
  FOREIGN KEY (Client_Id) REFERENCES dw.Cliente(Client_Id);

ALTER TABLE dw.Opinion
  ADD CONSTRAINT FK_Opinion_Fuente
  FOREIGN KEY (Fuente_Id) REFERENCES dw.Fuente(Fuente_Id);

ALTER TABLE dw.Opinion
  ADD CONSTRAINT FK_Opinion_Tiempo
  FOREIGN KEY (Time_Id) REFERENCES dw.Tiempo(Time_Id);

ALTER TABLE dw.Opinion
  ADD CONSTRAINT FK_Opinion_Clasificacion
  FOREIGN KEY (Class_Id) REFERENCES dw.Clasificacion(Class_Id);
GO
```

## Calidad de datos y índices:

```
USE OpinionesClientesDW;  
GO
```

```
ALTER TABLE dw.Categoria      ADD CONSTRAINT UQ_Categoria_Nombre  UNIQUE (Nombre);  
ALTER TABLE dw.Pais          ADD CONSTRAINT UQ_Pais_Nombre       UNIQUE (Nombre);  
ALTER TABLE dw.TipoFuente     ADD CONSTRAINT UQ_TipoFuente_Desc   UNIQUE (Descripcion);  
ALTER TABLE dw.Clasificacion  ADD CONSTRAINT UQ_Clasificacion_Code UNIQUE (Code);  
  
CREATE UNIQUE INDEX UQ_Producto_NombreCat  
ON dw.Producto(ProductName, Categoria_Id);  
  
CREATE UNIQUE INDEX UQ_Cliente_Email_NotNull  
ON dw.Cliente(Email)  
WHERE Email IS NOT NULL;  
  
ALTER TABLE dw.Tiempo  
ADD CONSTRAINT CK_Tiempo_Month CHECK ([Month] BETWEEN 1 AND 12),  
CONSTRAINT CK_Tiempo_Day    CHECK ([Day]    BETWEEN 1 AND 31),  
CONSTRAINT CK_Tiempo_Hour   CHECK ([Hour]   IS NULL OR [Hour] BETWEEN 0 AND 23);  
  
CREATE INDEX IX_Opinion_Producto_Fecha ON dw.Opinion(Product_Id, Time_Id);  
CREATE INDEX IX_Opinion_Fuente_Fecha  ON dw.Opinion(Fuente_Id,  Time_Id);  
CREATE INDEX IX_Opinion_Cliente_Fecha ON dw.Opinion(Client_Id,  Time_Id);  
CREATE INDEX IX_Opinion_Class_Fecha   ON dw.Opinion(Class_Id,   Time_Id);  
  
GO
```

GitHub: [Xavi1324/SDOC--BigData](https://github.com/Xavi1324/SDOC--BigData)