A blue and black logo

AI-generated content may be incorrect.

**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

**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**A screenshot of a computer

AI-generated content may be incorrect.

**Desarrollo del Pipeline ETL**

**Extracción y transformación:**

**A screenshot of a computer screen

AI-generated content may be incorrect.**

**Carga de datos**

**Clientes**

**A screenshot of a computer

AI-generated content may be incorrect.**

**Opinión  
A screenshot of a computer

AI-generated content may be incorrect.**

**Tiempo  
A screenshot of a computer

AI-generated content may be incorrect.**

**TipoFuente  
A screenshot of a computer

AI-generated content may be incorrect.**

**Productos  
A screenshot of a computer

AI-generated content may be incorrect.**

**Categoría**

**A screenshot of a computer

AI-generated content may be incorrect.**

**Fuente**

**A screenshot of a computer

AI-generated content may be incorrect.**

**Calificación**

**A screenshot of a computer

AI-generated content may be incorrect.**

**País**

**A screenshot of a computer

AI-generated content may be incorrect.**

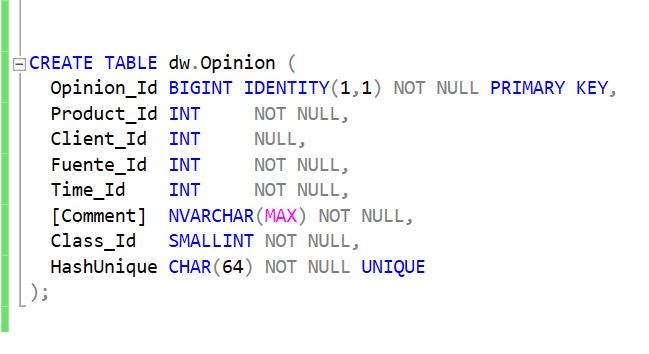
**Script SQL – Base de dato**

**Tablas:  
  
A screenshot of a computer

AI-generated content may be incorrect.**

**A screenshot of a computer program

AI-generated content may be incorrect.**

****

**Claves foráneas:  
A screenshot of a computer

AI-generated content may be incorrect.**

**A screenshot of a computer code

AI-generated content may be incorrect.**

**Calidad de datos y índices:**

**A screenshot of a computer program

AI-generated content may be incorrect.**