



# Tecnológico de Monterrey

Diccionario de Datos

Paulina Leal Mosqueda, A01659576

Santiago Nava Figueroa A01174557

Uziel Heredia Estrada, A01667072

4 de diciembre de 2025

Aplicación de métodos multivariados en Ciencias de Datos (Gpo 601)

Profesores: Juliho Davida Castillo Colmenares y Jonathan Montalvo Urquizo

Instituto Tecnológico y de Estudios Superiores de Monterrey CCM

# Cientes

**Archivo:** clientes.csv

**Observaciones:** 326 clientes

**Periodo de tiempo:** un año, 2024

**Contexto:** son los datos sobre las compras de clientes en una empresa de e-commerce

**Tipo de datos:** utilizar métodos no supervisados, no existen grupos de clientes predefinidos

**Descripción general:** no hay valores nulos, ni outliers. La variable Fecha\_Resgistro tiene un typo.

## Descripción de Variables

### Variables de Identificación e información personal

Variable	Tipo	Descripción	Valores
ID_Cliente	entero	Identificador del cliente	1, 2, etc.
Nombre	objeto	Nombre del cliente	Nombre
Apellido	objeto	Apellido del cliente	Apellido
Email	objeto	Correo del cliente	email@.algo

### Variables Adicionales

Variable	Tipo	Descripción	Valores
Fecha_Resgistro	objeto	Fechas en que se registró o realizó su primer compra el cliente	día/mes/año
Región	objeto	Ubicación o región del cliente	Zona en Argentina

# Productos

**Archivo:** productos.csv

**Observaciones:** 38 productos existentes

**Periodo de tiempo:** un año, 2024

**Contexto:** son los datos sobre los productos disponibles en una empresa de e-commerce

**Tipo de datos:** utilizar métodos no supervisados, no existen grupos de productos predefinidos

**Descripción general:** no hay valores nulos. Hay outliers en la columna Precio\_Unitario. La columna Precio\_Unitario no está como entero, se debe de modificar.

## Descripción de Variables

Variable	Tipo	Descripción	Valores
ID_Producto	entero	Identificador único del producto	1, 2, etc.
Nombre_producto	string	Nombre descriptivo del producto	Nombre
Categoría	string	ID de categoría al que pertenece el producto. 1 - Lácteos, 2 - Carnicería, 3 - Panadería, 4 - Frutas y Verduras, 5 - Congelados, 6 - Bebidas, 7 - Galletitas y Snacks, 8 - conservas	1, 2, 3, 4, 5, 6, 7, 8
Precio_Unitario	objeto	Precio por unidad de producto	precio de cada producto
Stock	int	cantidad de productos disponibles en el inventario	cantidades

# Ventas

**Archivo:** ventas.csv

**Observaciones:** 3029 ventas

**Periodo de tiempo:** un año, 2024

**Contexto:** son todas las ventas generadas en una empresa de e-commerce durante un año

**Tipo de datos:** utilizar métodos no supervisados, no existen grupos de grupos predefinidos

**Descripción general:** no hay valores nulos. No hay outliers

## Descripción de Variables

Variable	Tipo	Descripción	Valores
ID_Venta	entero	Identificador único de la venta	1, 2, etc.
ID_Cliente	entero	Identificador único del cliente que hizo la compra	1, 2, etc.
ID_Producto	entero	Identificador único del producto	1, 2, etc.
Cantidad	entero	Número de unidades vendidas	1, 2, etc.
Método de Pago	objeto	Forma de pago utilizada. 1- Efectivo, 2 - Tarjeta de Crédito, 3 - Tarjeta de Débito, 4 - Mercado Pago, 5 - Transferencia	1, 2, 3, 4, 5
Estado	objeto	Estado de la transacción.	Completa, pendiente, cancelada

**Nota:** es posible juntar las tres bases de datos mediante ID\_Cliente, ID\_Producto. Sin embargo hay dos cosas a considerar. La primera es que la fecha de venta se confunde con la

etiqueta de Fecha\_Registro, entonces, se le debe de renombrar como Fecha\_Venta. La segunda es que hay registros de venta por cada producto que compró el cliente, entonces al juntarlas, habrá duplicados.

**df**

**Archivo:** es la unión de todos los archivos anteriores

**Observaciones:** 3029 registros de venta

**Periodo de tiempo:** un año, 2024

**Contexto:** son todos los datos sobre las compras de clientes en una empresa de e-commerce, incluye a información de los productos y de los clientes

**Tipo de datos:** utilizar métodos no supervisados, no existen grupos de clientes predefinidos

**Descripción general:** no hay valores nulos, ni outliers. Hay outliers en la columna Precio\_Unitario. Hay duplicados.

En este caso, la base de datos incluye a toda las bases de datos anteriores, solo que se generan nuevas variables, por ello solo será necesario explicar estas últimas.

Variable	Tipo	Descripción	Valores
Año_Venta	entero	Año en el que se vendió	2024
Dia_Venta	entero	Día en el que el producto se vendió	1, 2, 3, etc.
Mes_Venta	entero	Mes en el que el producto se vendió	1, 2, etc.
Temporada	categoría	Temporada en la que el producto se vendió. Basado en Argentina.	Invierno, Otoño, Primavera, Verano
Ingreso	float	Ingreso generado en la venta. Es cantidad por precio unitario	1, 2, 3, 4, 5, etc.

Recencia	int	Número de días desde la última compra del cliente hasta la fecha. Sirve para ver si hay clientes que tienen tiempo sin ir a la tienda o si aún siguen viniendo.	1, 2, 3, etc.
Frecuencia	int	Frecuencia con la que compra el cliente. Es el número total de compras que hizo cada cliente	1, 2, 3, etc.
Valor Monetario	int	Cantidad de dinero gastado por cada cliente. Sirve para saber cuánto dinero está invirtiendo el cliente en la tienda.	1, 2, 3, etc.