

## **LABORATORIO 15**

### **Parte 1**

En la primera parte se aplicarán los temas vistos en la sesión como es la limpieza de los datos, se deberá realizar el paso a paso con las respectivas capturas de pantalla, esta aplicación es de acuerdo con los escenarios planteados.

#### **1. Ejercicio de práctica 1.**

##### **1. Escenario datos de ventas en línea**

Se tiene un conjunto de datos de una empresa de ventas online que contiene información de clientes, productos vendidos y fechas de transacciones. El archivo tiene problemas comunes como valores faltantes en la columna de códigos de productos, precios con formato incorrecto y nombres de clientes en mayúsculas y minúsculas mezcladas.

Data: data\_ventas\_para\_limpieza 1

Paso a paso para la limpieza

fx = Table.TransformColumns(#"Tipo cambiado1",{{"Nombre\_Cliente", Text.Proper, type text}})

	A <sub>C</sub> ID_Venta	A <sub>C</sub> Nombre_Cliente	A <sub>C</sub> Codigo_Producto	\$ Precio	Fecha_Transaccion
1	1	Isabel	P004	73,99	1/01/2023
2	2	Pedro	P001	52,23	2/01/2023
3	3	Luis	P003	99,09	3/01/2023
4	4	Lucía	P002	33,92	4/01/2023
5	5	María	P005	35,46	5/01/2023
6	6	Pedro	P002	16,08	6/01/2023
7	7	María	P005	49,89	7/01/2023
8	8	Pedro	P003	65,67	8/01/2023
9	9	Pedro	P001	13,43	9/01/2023
10	10	Ana	P003	81,04	10/01/2023
11	11	Luis	P005	31,00	11/01/2023
12	12	Carlos	P005	73,26	12/01/2023
13	13	Isabel	P003	67,41	13/01/2023
14	14	Luis	P003	58,38	14/01/2023
15	15	María	P004	51,49	15/01/2023
16	16	María	P005	24,11	16/01/2023
17	17	Ana	P001	80,18	17/01/2023
18	18	Isabel	P003	49,81	18/01/2023
19	19	Ana	P003	95,46	19/01/2023
20	20	Pedro	P005	24,50	20/01/2023
21	21	Luis	P001	95,90	21/01/2023
22	22	Luis	P003	40,39	22/01/2023
23	23	María	P005	79,22	23/01/2023
24	24	Ana	P003	77,06	24/01/2023
25	25	Lucía	P004	65,39	25/01/2023
26	26	María	sin codigo	28,87	26/01/2023
27	27	María	P005	82,57	27/01/2023
28	28	María	sin codigo	70,53	28/01/2023
29	29	Lucía	P002	12,89	29/01/2023
30	30	Carlos	P003	28,21	30/01/2023

Generación de perfiles de columnas basada en las 1000 primeras filas

## CONCLUSIONES

- . Manejamos valores faltantes en la columna
- . Normalizamos el valor con el formato de nombre
- . Normalizamos el formato de fecha

## 2. Escenario: Limpieza de datos de un servicio de atención al cliente

Trabajas en un centro de atención al cliente y tienes un conjunto de datos que contiene información sobre las interacciones con los clientes. Este dataset incluye el ID del cliente, Nombre del cliente, Categoría de consulta, Duración de la llamada y Fecha de la interacción. Algunos registros tienen problemas como nombres en minúsculas y mayúsculas inconsistentes, valores faltantes en la categoría de la consulta, y duraciones de llamada con errores de formato.

Data: data\_atencion\_cliente\_para\_limpieza 2

## Paso a paso para la limpieza

< X ✓ *fx* = Table.TransformColumns(#"Tipo cambiado2",{{"Nombre\_Cliente", Text.Proper, type text}})

	A <sup>B</sup> C ID_Cliente	A <sup>B</sup> C Nombre_Cliente	A <sup>B</sup> C Categoria_Conсульта	A <sup>B</sup> C Duracion_Llamada	A <sup>B</sup> C Fecha_Interaccion
1	2	José	Soporte Técnico	24.36	2/01/2023
2	3	Carlos	Facturación	58.86	3/01/2023
3	4	José	Facturación	33.57	4/01/2023
4	5	Pedro	Facturación	18.28	5/01/2023
5	7	Pedro	Soporte Técnico	29.0	7/01/2023
6	8	Luis	Consulta General	39.77	8/01/2023
7	9	Lucía	Soporte Técnico	23.21	9/01/2023
8	10	Ana	Consulta General	3.2	10/01/2023
9	11	María	Quejas	2.37	11/01/2023
10	12	José	Soporte Técnico	49.46	12/01/2023
11	13	Isabel	Quejas	26.74	13/01/2023
12	14	Pedro	Soporte Técnico	4.82	14/01/2023
13	15	Isabel	Consulta General	42.04	15/01/2023
14	16	Luis	Quejas	20.38	16/01/2023
15	17	Carlos	Facturación	19.76	17/01/2023
16	18	Isabel	Quejas	51.29	18/01/2023
17	19	Ana	Facturación	10.43	19/01/2023
18	20	María	Sugerencias	31.22	20/01/2023
19	21	Luis	Soporte Técnico	21.74	21/01/2023
20	22	Pedro	Quejas	48.28	22/01/2023
21	23	Isabel	Soporte Técnico	28.45	23/01/2023
22	24	Pedro	Soporte Técnico	23.27	24/01/2023
23	25	Luis	Facturación	39.42	25/01/2023
24	26	Carlos	Consulta General	18.8	26/01/2023
25	27	Luis	Quejas	39.98	27/01/2023
26	28	Lucía	Quejas	16.79	28/01/2023
27	29	Carlos	Sugerencias	32.1	29/01/2023
28	30	Carlos	Sugerencias	17.72	30/01/2023
29	31	Ana	Quejas	23.5	31/01/2023
30	32	Lucía	Soporte Técnico	57.76	1/02/2023

Generación de perfiles de columnas basada en las 1000 primeras filas

## CONCLUSIONES

- . En ID\_Cliente lo pasamos de numero entero a texto
- . En Nombre\_Cliente lo que hicimos fu tranformar los las mayusculas y minusculas que habian por "poner en mayuscula cada palabra"
- . Tambien cambiamos la duracion de llamada a numero decimal fijo
- . y por ultimo la fecha de interaccion lo pasamos de ser numero entero a FECHA

## 3. Escenario: Tienda minoritaria de tecnología

Trabajas para una tienda minoritaria tecnología, el equipo de inventario tiene dificultades para gestionar adecuadamente los datos de productos debido a la calidad inconsistente de

los registros. El conjunto de datos de inventario incluye información sobre ID de producto, Nombre del producto, Cantidad disponible, Precio unitario y Fecha de ingreso al inventario. Sin embargo, se han detectado varios problemas en los datos que impiden su uso eficaz para análisis de ventas y reposición de stock.

Los problemas para solucionar incluyen:

- Nombres de productos faltantes.
- Precios de algunos productos faltantes.
- Cantidades de productos con texto adicional "unidades".

El objetivo es optimizar la calidad de los datos de inventario mediante la corrección de errores comunes, asegurando que los registros sean consistentes, completos y adecuados para análisis de inventarios y toma de decisiones estratégicas en la tienda.

Data: data\_inventario\_para\_limpieza 3

### **Paso a paso para la limpieza**

< = Table.TransformColumnTypes(#"Valor reemplazado1",{{"Cantidad", Int64.Type}, {"Fecha\_Ingreso", type date}, {"Prec

	A <sub>C</sub> ID_Producto	A <sub>C</sub> Nombre_Producto	1 <sup>2</sup> <sub>3</sub> Cantidad	\$ Precio_Unitario	Fecha_Ingreso
1	1	Teclado	47	669,54	1/06/2022
2	2	Laptop	42	325,79	2/06/2022
3	3	Impresora	29	516,47	3/06/2022
4	4	Teclado	24	210,03	4/06/2022
5	5	Mouse	22	114,99	5/06/2022
6	6	Mouse	46	838,36	6/06/2022
7	7	Laptop	37	520,16	7/06/2022
8	8	Cámara	8	652,74	8/06/2022
9	9	Laptop	4	82,43	9/06/2022
10	10	Impresora	10	772,89	10/06/2022
11	11	Mouse	34	125,02	11/06/2022
12	12	Mouse	17	407,14	12/06/2022
13	13	Mouse	23	0,00	13/06/2022
14	14	Cámara	5	439,30	14/06/2022
15	15	Teléfono	44	800,79	15/06/2022
16	16	Teclado	18	746,90	16/06/2022
17	17	Teléfono	10	0,00	17/06/2022
18	18	Teclado	31	901,05	18/06/2022
19	20	Impresora	14	191,20	20/06/2022
20	21	Teclado	20	632,71	21/06/2022
21	22	Teléfono	40	197,23	22/06/2022
22	23	Teléfono	27	690,51	23/06/2022
23	24	Mouse	4	887,56	24/06/2022
24	25	Cámara	28	140,63	25/06/2022
25	26	Mouse	49	53,22	26/06/2022
26	27	Impresora	36	464,48	27/06/2022
27	28	Mouse	45	601,35	28/06/2022
28	29	Teléfono	5	382,72	29/06/2022
29	31	Mouse	7	832,21	1/07/2022
30	32	Mouse	26	135,46	2/07/2022

versión de perfil de columnas basada en los 1000 primeros filas