

Herramientas de Perfilado de Datos

Carlos Martínez, Ph.D.



Perfilado de Datos

El perfilado de datos es esencial para entender la composición de nuestros datos y cómo podemos mejorar su calidad antes de proceder al análisis.

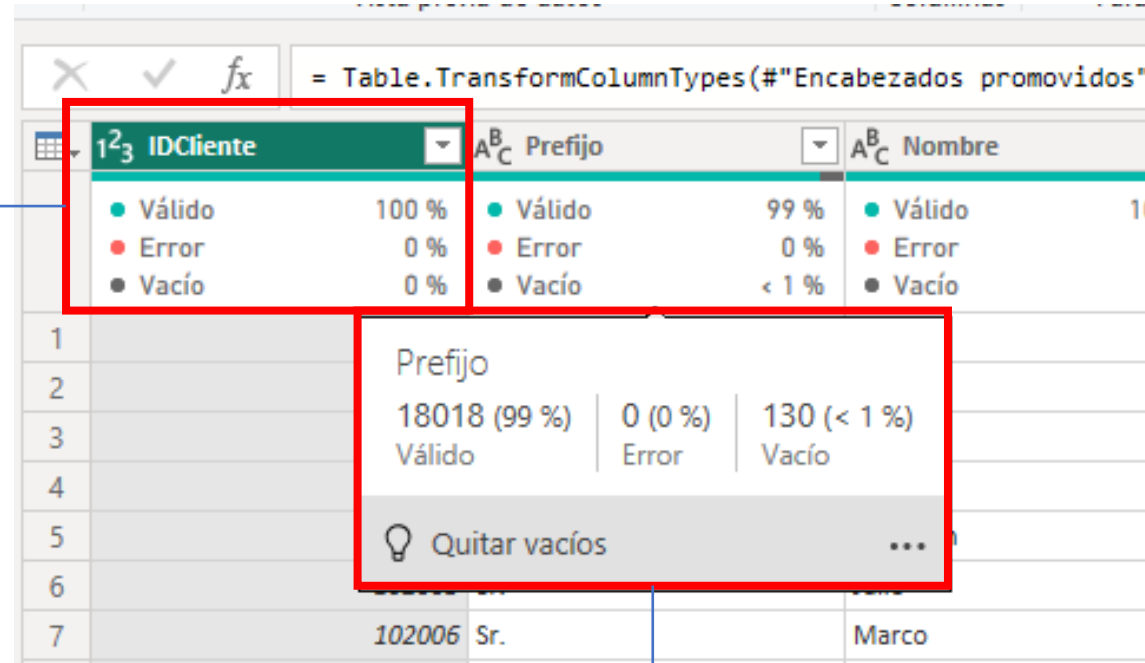
Cada herramienta de perfilado que revisaremos ofrece opciones, diseños y propósitos ligeramente diferentes, diseñados para brindarnos una comprensión visual e intuitiva de nuestros datos.

No solo veremos cómo funcionan estas herramientas, sino también cómo aplicarlas eficazmente para garantizar la integridad de nuestros datos.



Perfilado de Datos: Calidad de Columna

Muestra el porcentaje de valores válidos, con errores y los que están vacíos



	IDCliente	Prefijo	Nombre
Válido	100 %	99 %	100 %
Error	0 %	0 %	0 %
Vacío	0 %	< 1 %	0 %

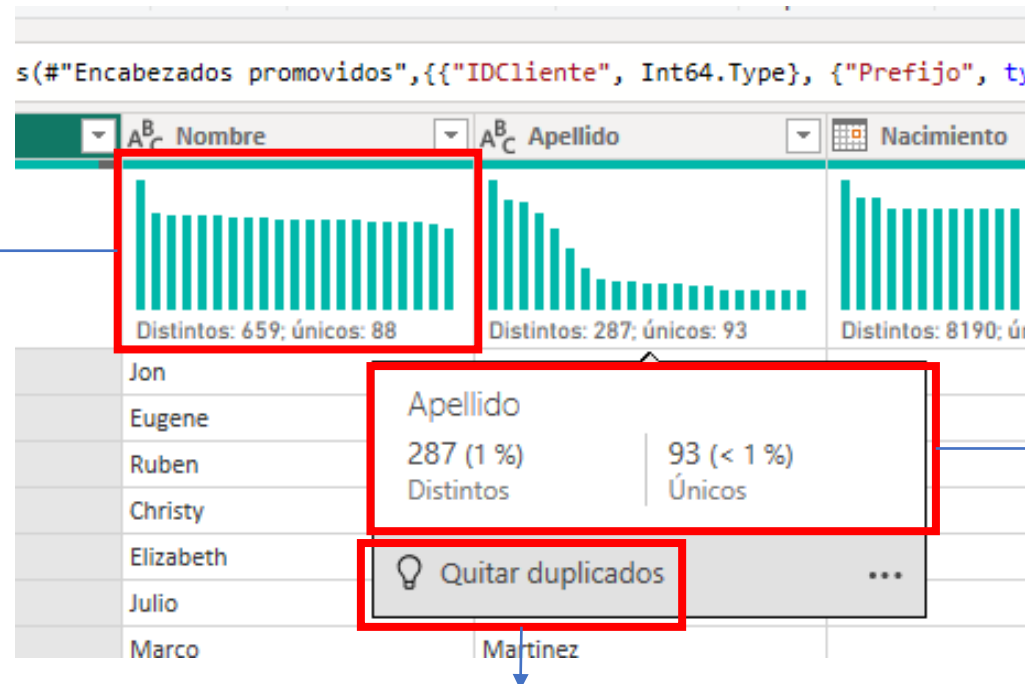
Válido	Error	Vacío
18018 (99 %)	0 (0 %)	130 (< 1 %)

Al pasar el cursor, también muestra el número de observaciones



Perfilado de Datos: Distribución de Columna

Provee un gráfico de la distribución de los datos en la columna

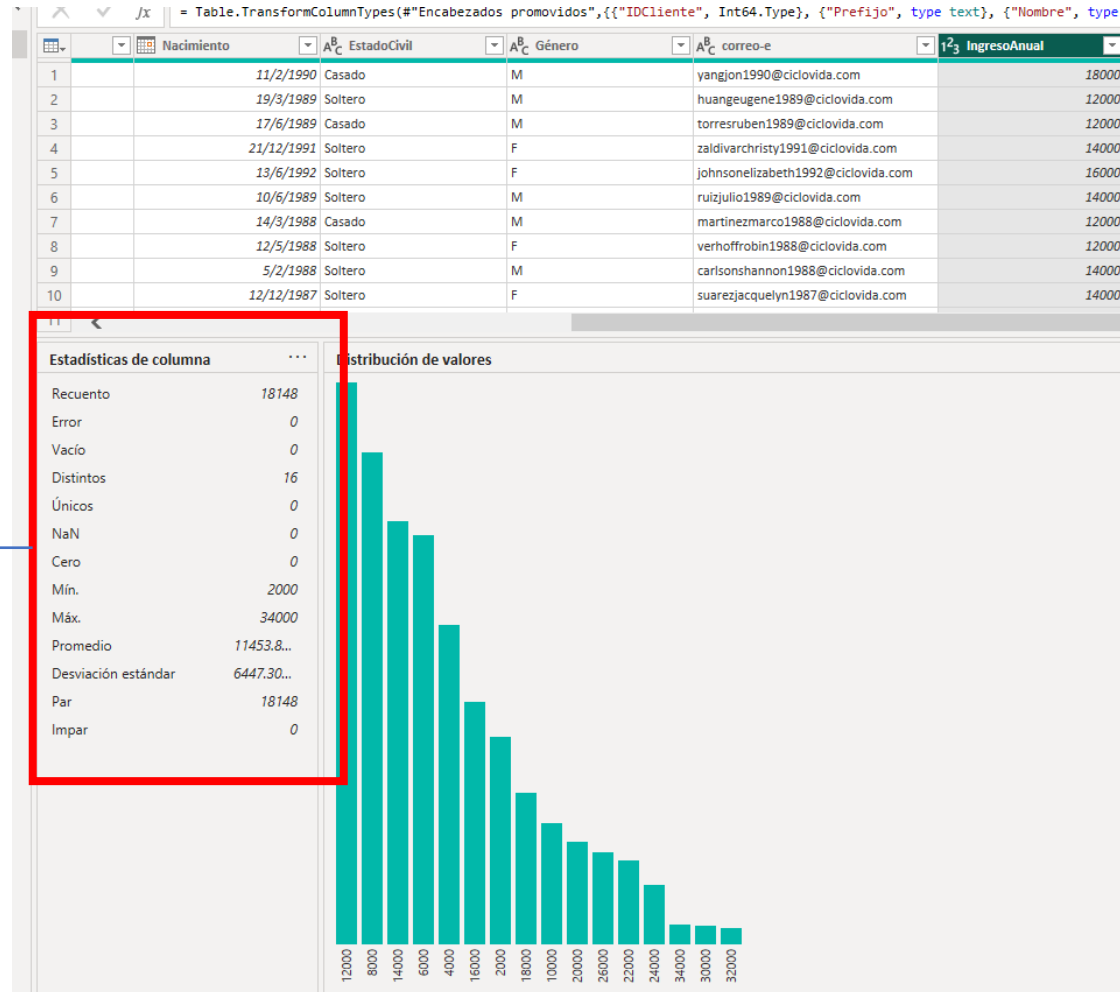


Al pasar el cursor, se pueden observar el número de valores distintos y únicos

Acción sugerida con base a los resultados de la columna



Perfilado de Datos: Perfil de Columna



Estadísticas de columna provee métricas detalladas incluyendo:
Recuento, mínimo, máximo, promedio, desviación y más.

