



Guía práctica: cómo corregir en Power Query los problemas de calidad de datos identificados

Índice	2
Paso 1 — Abrir el archivo y cargarlo en Power Query	3
Paso 2 — Corregir el ID_Estudiante (pérdida de ceros a la izquierda)	6
Paso 3 — Limpiar espacios y caracteres no imprimibles	11
Paso 4 — Unificar mayúsculas/minúsculas y corregir categorías con distinta grafía	12
Paso 4a — Estandarizar mayúsculas y minúsculas de forma automática	12
Paso 4b — Corregir a mano lo que la capitalización automática no puede resolver	13
Paso 5 — Eliminar duplicados (exactos y parciales)	14
Paso 5a — Duplicado exacto (ID 00240010)	14
Paso 5b — Duplicado parcial (ID 00240015)	15
Paso 6 — Tratar valores perdidos y códigos centinela	17
Paso 6a — Convertir los códigos centinela en nulos reales	18
Paso 6b — Decidir qué hacer con los nulos resultantes	20
Paso 7 — Estandarizar fechas con formatos mezclados y detectar fechas inválidas	21
Paso 8 — Corregir formatos que no coinciden con el tipo de dato	24
Paso 8 a — Ingreso_Familiar_Mensual (símbolo de moneda y separador de miles)	24
Paso 8b — Edad (texto en un campo numérico)	25
Paso 9 — Marcar valores fuera del diccionario de datos	26
Paso 10 — Marcar inconsistencias lógicas entre columnas	29
Paso 11 — Documentar y cargar el resultado final	29

Esta guía muestra, clic por clic, cómo corregir en Power Query cada uno de los problemas de calidad de datos identificados en el diagnóstico de la hoja Datos_Matricula del archivo Dataset_Calidad_Datos_Universidad.xlsx. Los pasos parten desde abrir el archivo por primera vez.

Paso 1 — Abrir el archivo y cargarlo en Power Query

1. Abre Microsoft Excel y luego abre el archivo Dataset_Calidad_Datos_Universidad.xlsx (Archivo > Abrir > Examinar, y navega hasta la carpeta donde lo guardaste).
2. En la parte inferior de la ventana, haz clic en la pestaña de la hoja Datos_Matricula para asegurarte de que esa hoja esté activa.
3. Haz clic en cualquier celda dentro de la tabla de datos, por ejemplo en la celda A1, donde está el encabezado ID_Estudiante.

Autoguardado Dataset_Calidad_Datos_Universidad.xlsx • Guardado en Este PC

Buscar

Archivo Inicio Insertar Dibujar Disposición de página Fórmulas Datos Revisar Vista Automatizar Ayuda

Cortar Copiar Copiar formato Portapapeles Fuente Alineación Número Estilos

Normal Incorrecto

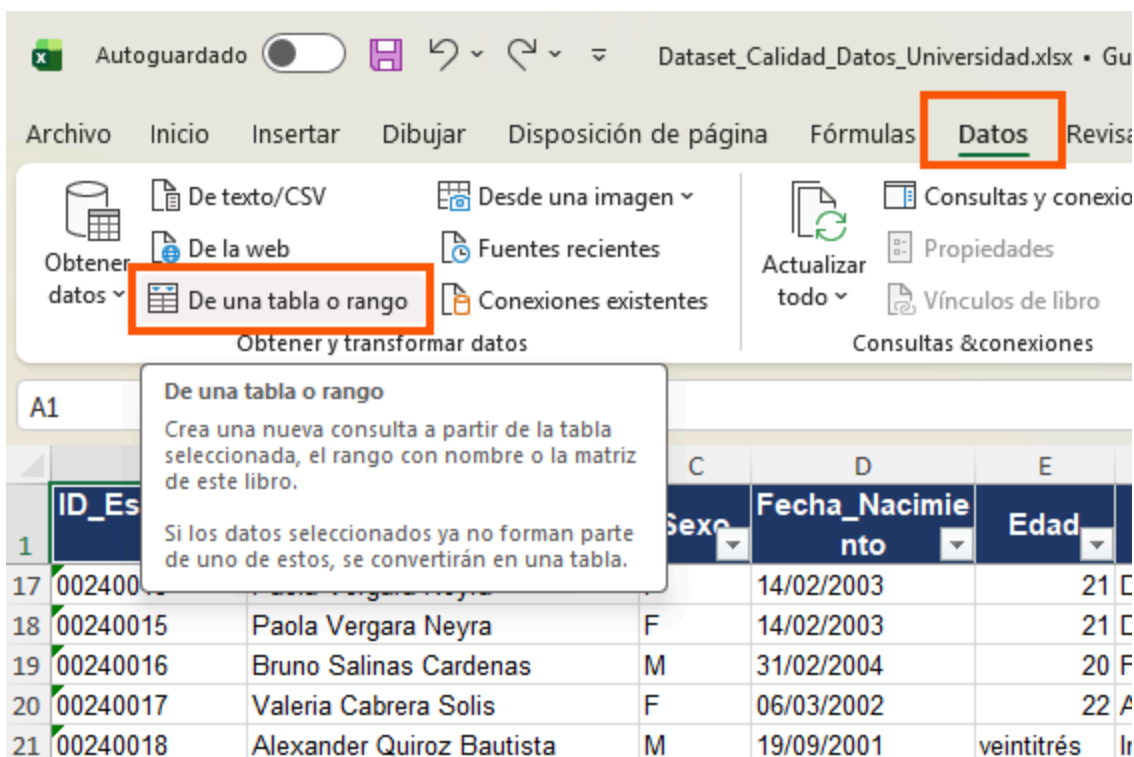
A1 ID_E estudiante

	A	B	C	D	E	F	G	H	I	J	K
	ID_Estudiant e	Nombres_y_Apellidos	Sexo	Fecha_Nacimie nto	Edad	Programa	Ciudad_Orig en	Estado_Ci vil	Modalidad	Estatura_ m	Prome er
2	00240001	Andrea Salazar Chumpitaz	F	15/04/2003	21	Ingeniería Informática	Lima	1	Presencial	1.63	
3	00240002	Jose Luis Ramirez Vega	M	22/11/2002	22	Ingeniería Civil	Arequipa	1	presencial	1.75	
4	00240003	MARIA FERNANDA TORRES LE	F	03-08-2001	23	Administración	lima	1	Virtual	1.6	
5	00240004	Carlos Mendoza Rios	M	1999/07/19	25	Ingeniería Industrial	Trujillo	2	Presencial	1.78	
6	00240005	Lucia Alvarado Quispe	F	30/01/2004	20	Psicología	Cusco	1	Presencial	1.58	
7	00240006	Fernando Gutierrez Paz	M		24	Derecho	Lima	4	Presencial	1.8	
8	00240007	Daniela Ochoa Bravo	F	05/09/2002	21	Sociología	Piura	1	Virtual	1.66	
9	00240008	Andrés Villanueva Soto	M	1998-02-14	27	Diseño Gráfico	Lima	2	Presencial	1.72	
10	00240009	Karina Espinoza Diaz	F	18/06/2003	21	Contabilidad	Chiclayo	1	Presencial	1.61	
11	00240010	Renzo Aguilar Ponce	M	09/12/2001	22	Ingeniería Informática	Lima	1	Presencial	1.74	
12	00240010	Renzo Aguilar Ponce	M	09/12/2001	22	Ingeniería Informática	Lima	1	Presencial	1.74	
13	00240011	Ximena Castro Farfan	F	27/07/2002	22	Educación Inicial	Ica	1	Presencial	1.64	
14	00240012	Sebastian Rojas Huaman	M	2002/04/30	22	Ingeniería Civil	Huancayo	1	presencial	1.77	
15	00240013	Grecia Nolasco Bermudez	999	11/10/2003	20	Marketing	Lima	1	Virtual	1.59	
16	00240014	Diego Fernandez Ugarte	M	23/05/2000	26	Ingeniería Mecánica	Tacna	2	Presencial	1.81	
17	00240015	Paola Vergara Neyra	F	14/02/2003	21	Derecho	Arequipa	1	Presencial	1.62	
18	00240015	Paola Vergara Neyra	F	14/02/2003	21	Derecho	Arequipa	1	Presencial	1.62	
19	00240016	Bruno Salinas Cardenas	M	31/02/2004	20	Filosofía	Lima	1	Presencial	1.7	
20	00240017	Valeria Cabrera Solis	F	06/03/2002	22	Antropología	Cusco	3	Virtual	1.6	
21	00240018	Alexander Quiroz Bautista	M	19/09/2001	veintitrés	Ingeniería Industrial	Trujillo	1	presencial	1.79	
22	00240019	Milagros Peña Zevallos	F	2003/12/08	20	Administración	Piura	1	Presencial	1.57	
23	00240020	Cristian Loayza Medina	M	17/01/1999	25	Derecho	Lima	5	Presencial	1.76	
24	00240021	Nicole Farfan Reyes	F	25/06/2003	21	Psicología	Lima	8	Virtual	1.65	
25	00240022	Gonzalo Prado Aliaga	M	04/04/2002	22	Ingeniería Civil	Chiclayo	1	Presencial	1.73	
26	00240023	Fiorella Chirinos Meza	F	09/09/2003	20	Diseño Gráfico	Lima	1	Presencial	-1.65	
27	00240024	Adrian Bustamante Solano	M	12/12/2000	24	Contabilidad	Huancayo	2	Virtual	1.71	
28	00240025	Camila Herrera Vasquez	F	01/05/2004	20	Educación Primaria	Ica	1	Presencial	1.6	
29	00240026	Jhonatan Rivas Cordova	M	20/03/2009	15	Ingeniería Informática	Lima	1	Presencial	1.68	
30	00240027	Estefania Nunez Carbajal	F	28/08/2002	22	Marketing	Arequipa	1	Presencial	1.63	
31	00240028	Ruben Zamudio Flores	M	15/07/2001	23	Derecho	Lima	2	presencial	1.75	
32	00240029	Antuanet Cardenas Rojas	F	-	N/A	Psicología	Lima	1	Virtual	1.62	
33	00240030	Erick Salcedo Munoz	M	22/10/2000	24	Ingeniería Mecánica	Cusco	1	Presencial	1.8	
34	00240031	Yesenia Portocarrero Diaz	F	30/04/2003	21	Sociología	Trujillo	1	Presencial	1.59	
35	00240032	Marco Antonio Leya Rios	M	30/02/2002	22	Ingeniería Industrial	Piura	1	Presencial	1.74	
36	00240033	Rosa Elena Chavez Guerra	F	03/11/1998	27	Administración	Lima	2	Virtual	1.61	

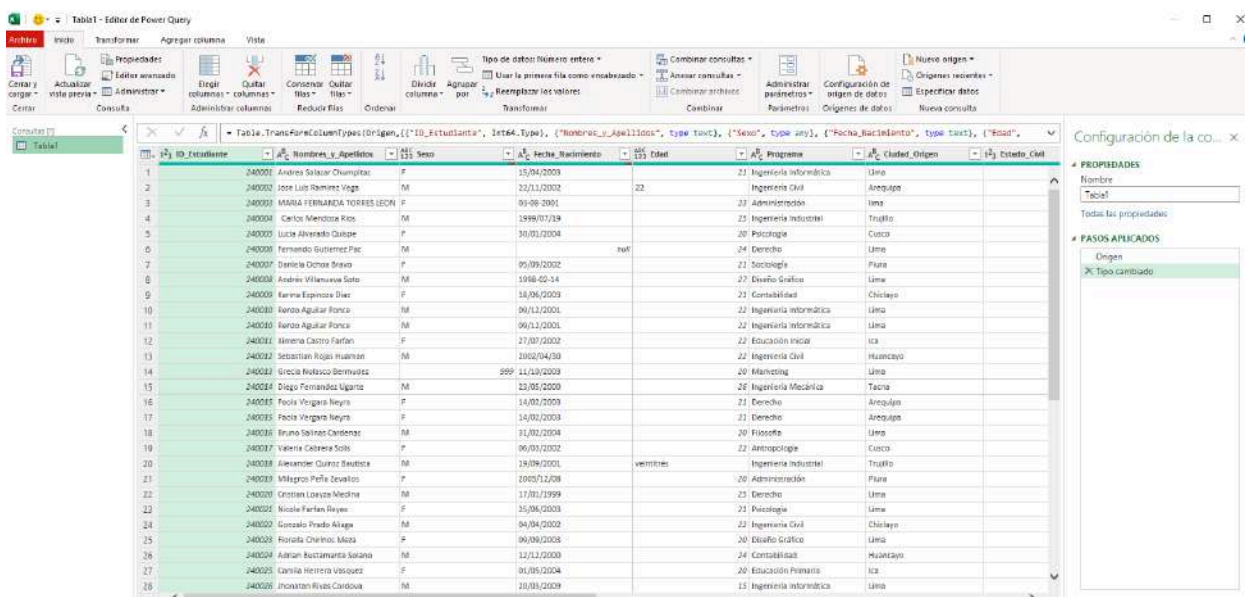
Datos_Matricula Diccionario_Datos Guia_Problemas (uso docente) +

4. Ve a la pestaña Datos (Data) en la cinta de opciones.

5. Dentro del grupo "Obtener y transformar datos" (Get & Transform Data), haz clic en el botón Desde tabla/rango (From Table/Range).



6. Si Excel muestra un aviso preguntando si los datos tienen encabezados ("Mi tabla tiene encabezados" / "My table has headers"), verifica que las casillas de los títulos estén marcadas — como ID_Estudiante, Nombres_y_Apellidos, etc. — están en la primera fila — y haz clic en Aceptar (OK).



Con esto, Excel convierte el rango en una Tabla (si todavía no lo era) y abre automáticamente el Editor de Power Query con los datos ya cargados. A partir de aquí continúa con el Paso 2.

Paso 2 — Corregir el ID_Estudiante (pérdida de ceros a la izquierda)

La columna ID_Estudiante tiene un valor guardado como número (240044) en vez de texto con ceros a la izquierda (00240044). Para forzar que todos los códigos tengan 8 caracteres:

1. Selecciona la columna ID_Estudiante haciendo clic en su encabezado.
2. Ve a la pestaña Transformar (Transform) > grupo "Cualquier columna" (Any Column) > botón "Tipo de dato" (Data Type) > elige Texto (Text). Esto convierte primero el valor numérico en texto simple, sin ceros todavía.

Tabla1 - Editor de Power Query

Archivo Inicio Transformar Agregar columna Vista

Transponer
Invertir filas
Contar filas

Tabla

Consultas [1]
Tabla1

Tipo de datos: Número entero

- Número decimal
- Moneda
- Número entero
- Porcentaje
- Fecha/Hora
- Fecha
- Hora
- Fecha/Hora/Zona horaria
- Duración
- Texto
- Verdadero/Falso
- Binario

Reemplazar los valores
Rellenar
Columna dinámica
Cualquier columna

Anular dinamicación de columnas
Mover
Convertir en lista

ID	Nombre y Apellidos	Sexo	Fecha_Nacimiento
1			
2			
3			
4			
5			
6			
7			
8	240008 Andrés Villanueva Soto	M	15/04/2003
9	240009 Karina Espinoza Díaz	F	22/11/2002
10	240010 Renzo Aguilar Ponce	M	03-08-2001
11	240010 Renzo Aguilar Ponce	M	1999/07/19
12	240011 Ximena Castro Farfan	F	30/01/2004
13	240012 Sebastian Rojas Huaman	M	
14	240013 Grecia Nolasco Bermudez		999 05/09/2002
15	240014 Diego Fernandez Ugarte	M	1998-02-14
16	240015 Paola Vergara Neyra	F	18/06/2003
17	240015 Paola Vergara Neyra	F	09/12/2001
18	240016 Bruno Salinas Cardenas	M	09/12/2001
19	240017 Valeria Cabrera Solis	F	27/07/2002
20	240018 Alexander Quiroz Bautista	M	2002/04/30
21	240019 Milagros Peña Zevallos	F	11/10/2003
22	240020 Cristian Loayza Medina	M	23/05/2000
23	240021 Nicole Farfan Reyes	F	14/02/2003
24	240022 Gonzalo Prado Aliaga	M	14/02/2003
25	240023 Fiorella Chirinos Meza	F	31/02/2004
26	240024 Adrian Bustamante Solano	M	06/03/2002
27	240025 Camila Herrera Vasquez	F	19/09/2001
28	240026 Jhonatan Rivas Cordova	M	2003/12/08
29			17/01/1999

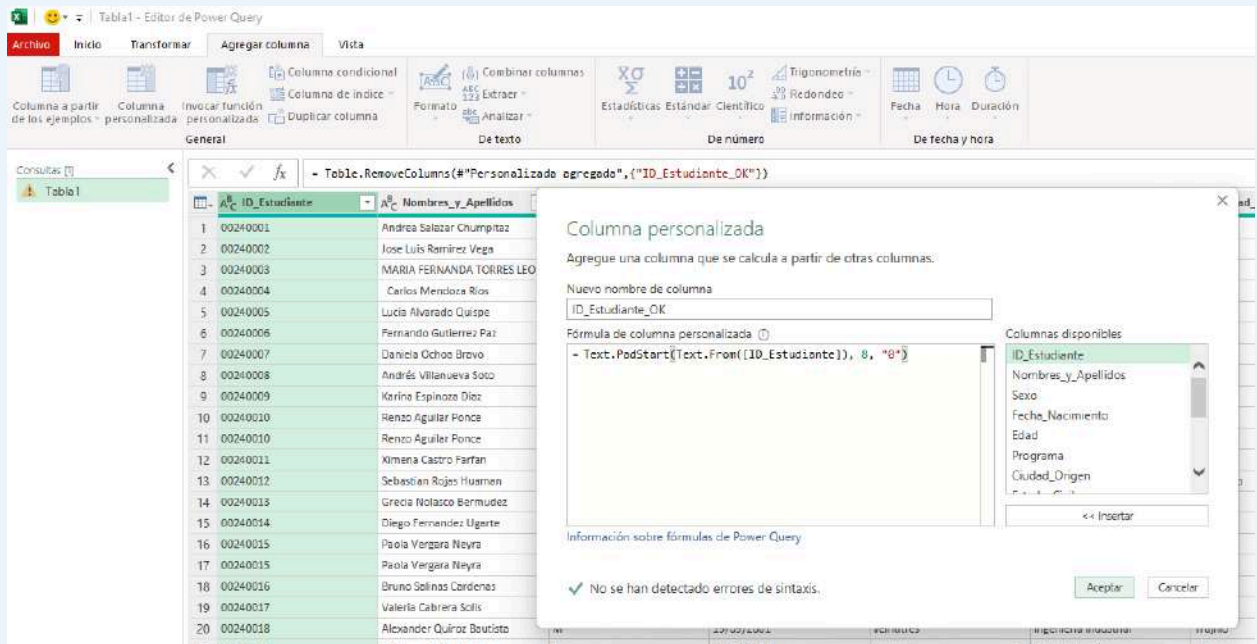
3. 15 COLUMNAS, 50 FILAS Generación de perfiles de columnas basada en las 1000 primeras filas

- Ve a la pestaña Agregar columna (Add Column) en la cinta de opciones.
- Haz clic en el botón Columna personalizada (Custom Column).
- En el cuadro de diálogo, en "Nuevo nombre de columna" escribe: ID_Estudiante_OK.

- En el cuadro de fórmula, escribe exactamente (recuerda que este cuadro siempre exige el signo = al inicio):

CÓDIGO M — RELLENAR CEROS A LA IZQUIERDA

```
=Text.PadStart(Text.From([ID_Estudiante]), 8, "0")
```



Columna personalizada

Agregue una columna que se calcula a partir de otras columnas:

Nuevo nombre de columna
ID_Estudiante_OK

Fórmula de columna personalizada
=Text.PadStart(Text.From([ID_Estudiante]), 8, "0")

Columnas disponibles
ID_Estudiante
Nombres_y_Apellidos
Sexo
Fecha_Nacimiento
Edad
Programa
Ciudad_Origen

Información sobre fórmulas de Power Query

✓ No se han detectado errores de sintaxis.

Aceptar Cancelar

- Haz clic en Aceptar (OK). Aparece una columna nueva al final de la tabla, ID_Estudiante_OK, ya con los ceros a la izquierda (ej. "00240044").

▼	ABC 123	ID_Estudiante_OK	▼
		00240001	
		00240002	
		00240003	
		00240004	
		00240005	
		00240006	
		00240007	
		00240008	
		00240009	
		00240010	
		00240010	
		00240011	
		00240012	
		00240013	
		00240014	
		00240015	
		00240015	
		00240016	
		00240017	
		00240018	
		00240019	
		00240020	
		00240021	
		00240022	
		00240023	
		00240024	
		00240025	
		00240026	

- Borra la columna original: clic derecho sobre el encabezado ID_Estudiante > Quitar (Remove).

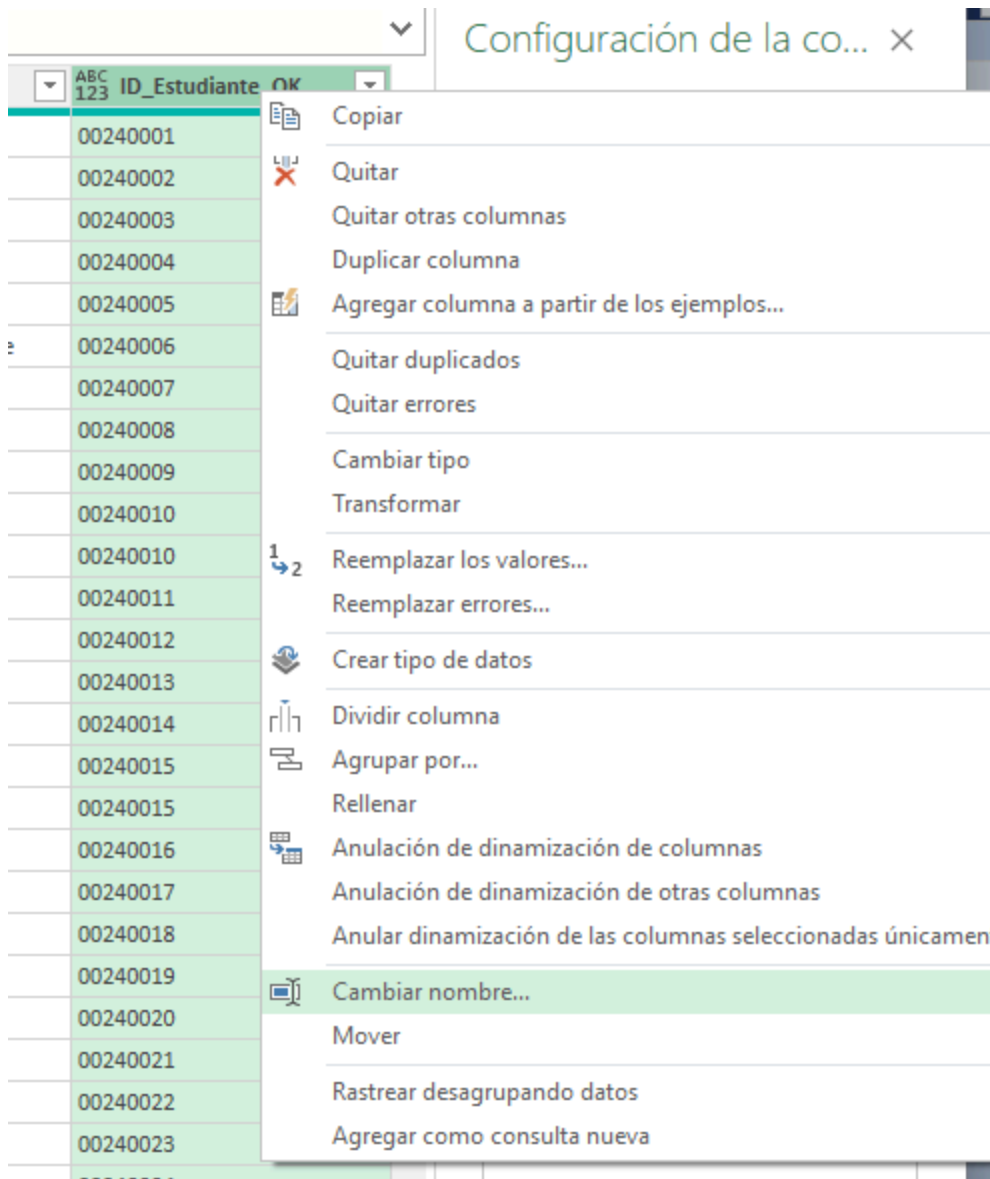
General De texto D

< X ✓ fx = Table.AddColumn("#Columnas quitadas", "ID_Estudiante_OK", e

ABC 123	ID_Es
1 00240001	
2 00240002	
3 00240003	
4 00240004	
5 00240005	
6 00240006	
7 00240007	
8 00240008	

- Copiar
- Quitar
- Quitar otras columnas
- Duplicar columna
- Agregar columna a partir de los ejemplos...
- Quitar duplicados
- Quitar errores

10. Renombra la columna nueva: doble clic sobre el encabezado ID_Estudiante_OK (o clic derecho > Renombrar / Cambiar nombre) y escribe ID_Estudiante.



11. Arrastra esa columna hasta que quede primera en la tabla, para conservar el orden original de las columnas.
12. Confirma el tipo de dato: haz clic en el ícono de tipo a la izquierda del encabezado (ABC/123) y selecciona Texto (Text), por si Power Query la dejó como "Cualquiera" (Any).

Resultado esperado

Todas las filas de ID_Estudiante muestran 8 caracteres, incluyendo el registro que antes aparecía como 240044 (ahora 00240044).

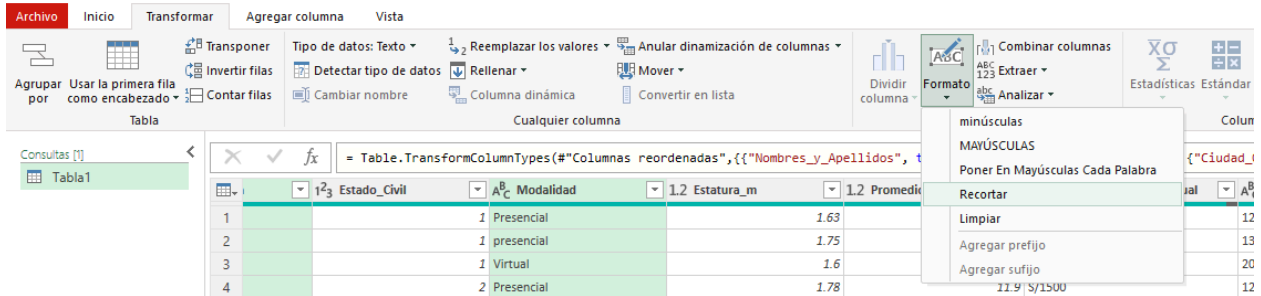
Paso 3 – Limpiar espacios y caracteres no imprimibles

Varias filas tienen espacios extra al inicio o al final del texto (por ejemplo, " Carlos Mendoza Rios" o "Sandra Palomino Quintana " y "Ingenieria " con espacio final).

1. Selecciona todas las columnas de texto a limpiar: clic en el encabezado de Nombres_y_Apellidos, luego mantén presionada la tecla Ctrl y haz clic también en los encabezados de Programa, Ciudad_Origen, Modalidad y Becado (deben quedar resaltadas al mismo tiempo).

ID_Estudiante	Nombres_y_Apellidos	Sexo	Fecha_Nacimiento	Edad	Programa	Ciudad_Origen	Estado_Civil
00240021	Nicole Farfen Reyes	F	25/06/2003		21 Psicología	Lima	
00240022	Gonzalo Prado Allaga	M	04/04/2002		22 Ingeniería Civil	Chiclayo	
00240023	Florella Chirinos Meza	F	09/09/2003		20 Diseño Gráfico	Lima	
00240024	Adrian Bustamante Solano	M	12/12/2000		24 Contabilidad	Huancayo	
00240025	Cemilo Herrera Vasquez	F	01/05/2004		20 Educación Primaria	Ica	
00240026	Jhonatan Rivas Cordova	M	20/03/2009		15 Ingeniería Informática	Lima	
00240027	Estefanía Nuñez Carbajal	F	28/08/2002		22 Marketing	Arequipa	
00240028	Ruben Zamudio Flores	M	15/07/2001		23 Derecho	Lima	
00240029	Antuñet Cardenas Rojas	F	-	N/A	21 Psicología	Lima	
00240030	Erick Salcedo Munoz	M	22/10/2000		24 Ingeniería Mecánica	Cusco	
00240031	Yessenia Portocarrero Diaz	F	30/04/2003		21 Sociología	Trujillo	
00240032	Marco Antonio Leyva Rios	M	30/02/2002		22 Ingeniería Industrial	Piura	
00240033	Rosa Elena Chavez Guerra	F	03/11/1998		27 Administración	Lima	
00240034	Kevin Salas Terreros	M	09/06/2003		21 Contabilidad	Chiclayo	
00240035	Brenda Ibarra Colque	F	14/01/2004		20 Diseño Gráfico	Lima	
00240036	Hector Maldonado Pinto	M	19/09/1999		25 Derecho	Arequipa	
00240037	Ariana Torres Benavides	F	07/07/2002		22 Psicología	Lima	
00240038	Franco Requejo Alva	M	2001-08-25		23 Ingeniería Informática	Trujillo	
00240039	Sandra Palomino Quintana	F	12/12/2003		20 Ingeniería Civil	Lima	
00240040	Jorge Luis Vidal Campos	M	28/02/2000		24 Ingeniería Civil	Cusco	
00240041	Katherine Rios Fernandez	F	16/05/2003		21 Marketing	Lima	
00240042	Luis Fernando Aparicio	M	10/10/2005		40 Derecho	Lima	
00240043	Marlene Solórzano Vega	F	21/03/2002		22 Sociología	Iquitos	
00240044	Willy Choque Apaza	M	14/07/2001		23 Ingeniería Industrial	Arequipa	
00240045	Gabriele Montoya Pileto	F	05/12/2003		20 Educación Inicial	Lima	
00240046	Tomas Espejo Rondon	M	08/08/2002		22 Contabilidad	Lima	
00240047	Alessandra Chuquilla	F	03 mar 2004		20 Diseño Gráfico	Trujillo	
00240048	Percy Guevara Cotrino		09/09/2000		24 Ingeniería Informática	Piura	

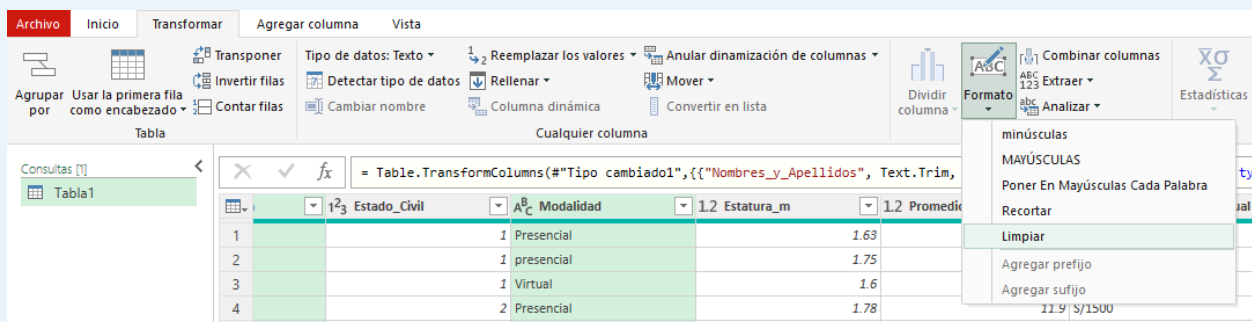
2. Ve a la pestaña Transformar (Transform) > grupo "Texto" (Text Column) > botón Formato (Format).
3. En el menú desplegable, elige Recortar (Trim). Esto quita los espacios al inicio y al final de las siete columnas seleccionadas en un solo paso.



4. Con las mismas siete columnas todavía seleccionadas, repite: Transformar > Formato (Format) > Limpiar (Clean). Esto elimina caracteres no imprimibles ocultos que a veces vienen de copiar y pegar desde otros sistemas.

Nota

Este paso todavía no unifica mayúsculas y minúsculas (por ejemplo, "presencial" sigue distinto de "Presencial"); eso se corrige en el Paso 4.



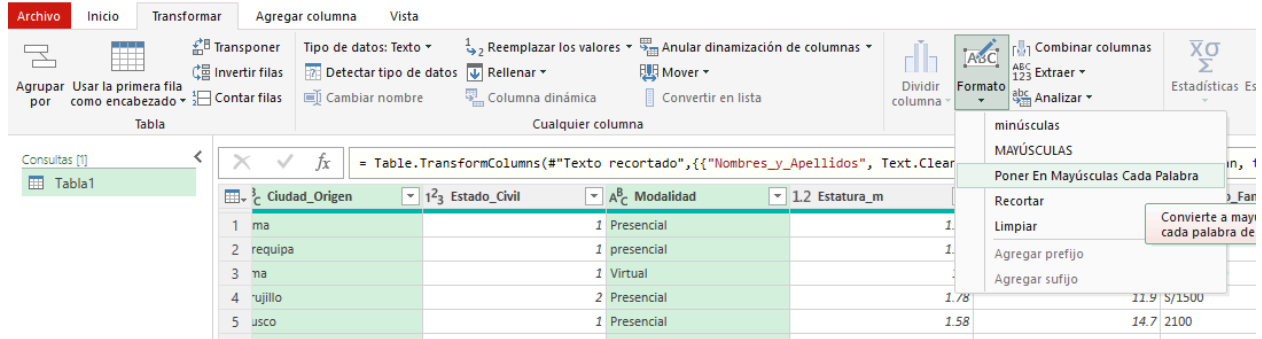
Paso 4 — Unificar mayúsculas/minúsculas y corregir categorías con distinta grafía

En vez de corregir manualmente cada variante ("sí", "SI", "sÍ", "si"...), primero se estandariza el uso de mayúsculas de forma automática, y solo al final se corrigen a mano los casos que sigan distintos por una tilde faltante o por ser directamente una palabra diferente.

Paso 4a — Estandarizar mayúsculas y minúsculas de forma automática

1. Selecciona las columnas categóricas: clic en el encabezado de Ciudad_Origen, luego Ctrl + clic en Modalidad y Becado (no incluyas Nombres_y_Apellidos aquí, esa columna no tiene este problema).

2. Ve a Transformar (Transform) > grupo "Texto" (Text Column) > botón Formato (Format).
3. Elige Poner en mayúscula cada palabra (Capitalize Each Word).

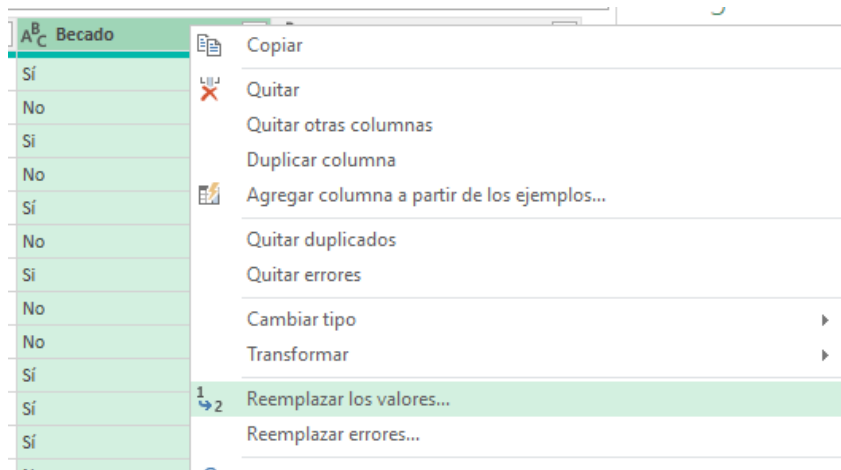


Con este único paso, valores como "lima" quedan como "Lima", "presencial" como "Presencial", "alto"/"bajo" como "Alto"/"Bajo", y "NO"/"no" como "No": ya coinciden con el resto del dataset y no necesitan corrección adicional.

Paso 4b — Corregir a mano lo que la capitalización automática no puede resolver

Queda un caso que la capitalización automática no arregla, porque no son un problema de mayúsculas sino de una tilde faltante o de una palabra distinta:

1. Clic derecho sobre el encabezado de Becado > Reemplazar valores (Replace Values).



2. En "Valor que se va a buscar" (Value To Find) escribe: Si

3. En "Reemplazar con" (Replace With) escribe: Sí

×

Reemplazar los valores

Reemplace un valor con otro de las columnas seleccionadas.

Valor que buscar

Reemplazar con

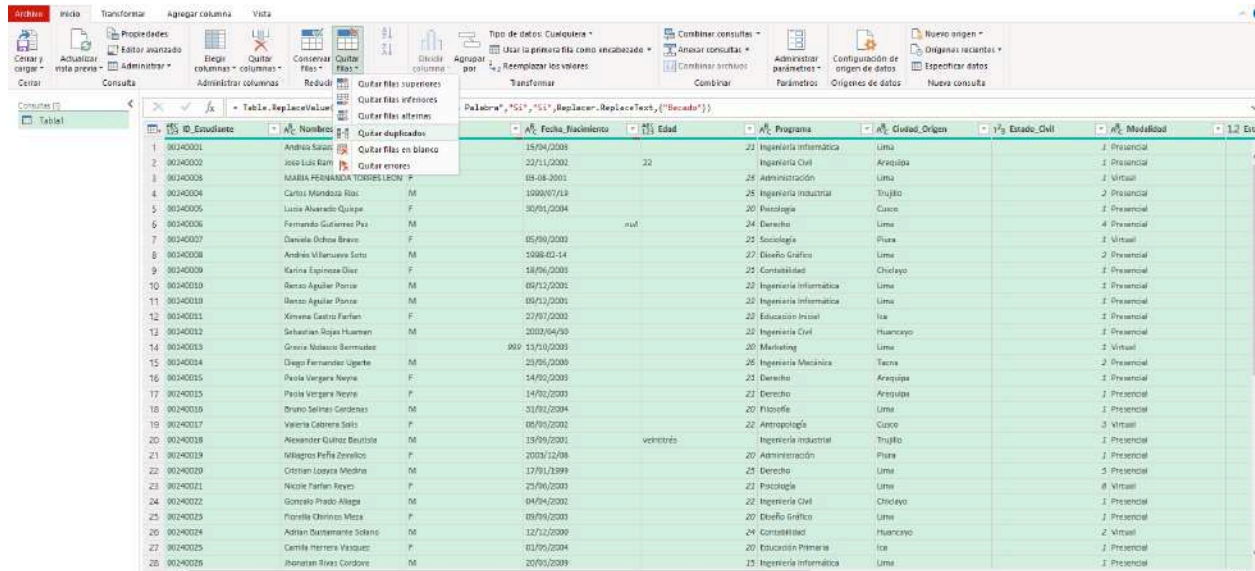
▸ Opciones avanzadas

4. Clic en Aceptar (OK).

Paso 5 — Eliminar duplicados (exactos y parciales)

Paso 5a — Duplicado exacto (ID 00240010)

1. Haz clic en el encabezado de la primera columna de la tabla.
2. Mantén presionada la tecla Mayús (Shift) y haz clic en el encabezado de la última columna, para seleccionar todas las columnas a la vez.
3. Ve a la pestaña Inicio (Home) > grupo "Quitar filas" (Reduce Rows) > Quitar filas > Quitar duplicados (Remove Rows > Remove Duplicates).



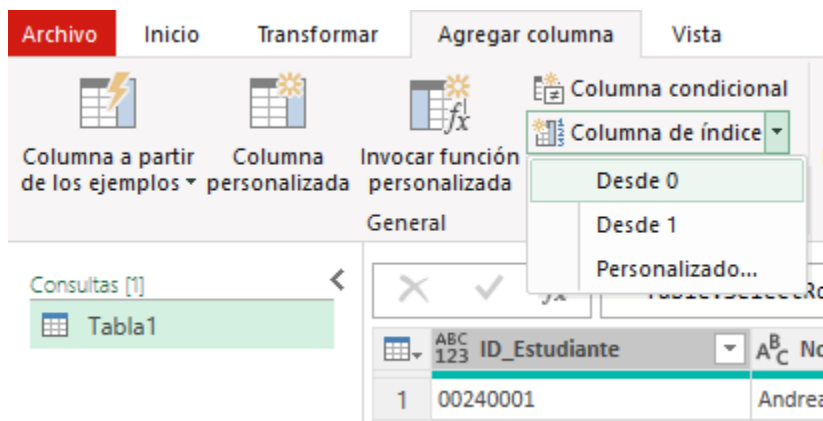
ID_Enudiante	Nombre	Fecha_Nacimiento	Edad	Programa	Ciudad_Origen	Ys	Estado_Civil	Modalidad
00240001	Andrea Salas	15/04/2008		Ingeniería Informática	Lima			1. Presencial
00240002	José Luis Ramírez	22/11/2002	22	Ingeniería Civil	Arequipa			1. Presencial
00240003	MARISA PERIBARDA YORRES LARON	05-08-2001		Administración	Lima			1. Virtual
00240004	Carlos Monreal Ríos	19/08/1971		Ingeniería Industrial	Trujillo			2. Presencial
00240005	Luzia Alvarado Quijpe	30/01/2004		Psicología	Cusco			1. Presencial
00240006	Fernando Gutiérrez Piza		nul	Derecho	Lima			4. Presencial
00240007	Daniela Ordaz Bravo	05/06/2002		Sociología	Piura			1. Virtual
00240008	Andrés Villaverde Soto	10/08/02-14		Diseño Gráfico	Lima			2. Presencial
00240009	Karina Espinoza Olivo	18/06/2001		Contabilidad	Chiclayo			1. Presencial
00240010	Banco Aguilar Ponce	08/12/2001		Ingeniería Informática	Lima			1. Presencial
00240011	Banco Aguilar Ponce	08/12/2001		Ingeniería Informática	Lima			1. Presencial
00240012	Kimberly Castro Parlan	22/07/2002		Educación Inicial	Ica			1. Presencial
00240013	Sebastian Rojas Huaman	20/02/04/90		Ingeniería Civil	Huancayo			1. Presencial
00240014	Gracia Molinos Bermudez	9/09-11/10/2003		Marketing	Lima			2. Virtual
00240015	Diego Fernandez Ugarte	23/05/2000		Ingeniería Mecánica	Tarma			2. Presencial
00240016	Perla Vergara Neyra	14/02/2003		Derecho	Arequipa			1. Presencial
00240017	Perla Vergara Neyra	14/02/2003		Derecho	Arequipa			1. Presencial
00240018	Bruno Salinas Cardenas	31/02/2004		Filosofía	Lima			1. Presencial
00240019	Valeria Cabrera Salto	08/09/2002		Antropología	Cusco			3. Virtual
00240020	Alexander Gutierrez Beutista	19/09/2001	veintitrés	Ingeniería Industrial	Trujillo			1. Presencial
00240021	Milagros Palla Zentillos	20/03/12/88		Administración	Piura			1. Presencial
00240022	Cristian Lozano Medina	17/01/1999		Derecho	Lima			5. Presencial
00240023	Nicole Parlan Reyes	25/06/2003		Psicología	Lima			8. Virtual
00240024	Gonzalo Prado Alaya	04/04/2002		Ingeniería Civil	Chiclayo			1. Presencial
00240025	Patricio Chirinos Mesa	08/09/2003		Diseño Gráfico	Lima			1. Presencial
00240026	Adrian Duenkemann Sotelo	12/12/2000		Contabilidad	Huancayo			2. Virtual
00240027	Carmela Herrera Vazquez	01/05/2004		Educación Primaria	Ica			1. Presencial
00240028	Jonathan Rojas Cardozo	20/05/2009		Ingeniería Informática	Lima			1. Presencial

Como las dos filas del ID 00240010 son idénticas en todas las columnas, Power Query conserva solo una.

Paso 5b — Duplicado parcial (ID 00240015)

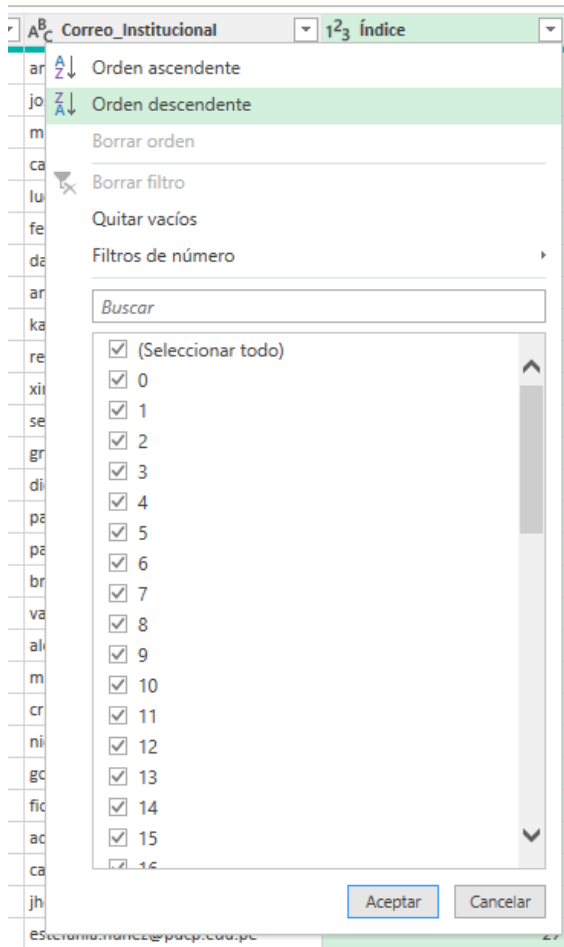
Este caso no se puede resolver con "Quitar duplicados" directamente, porque las dos filas no son idénticas (difieren en Promedio_Ponderado, Ingreso_Familiar_Mensual y Becado). Se aplica una regla de negocio simple: conservar el registro que apareció más tarde en la fuente original, asumiendo que representa la actualización más reciente.

1. Ve a la pestaña Agregar columna (Add Column) > Columna de índice (Index Column) > Desde 0 (From 0). Se agrega una columna llamada Índice que refleja el orden original de las filas.

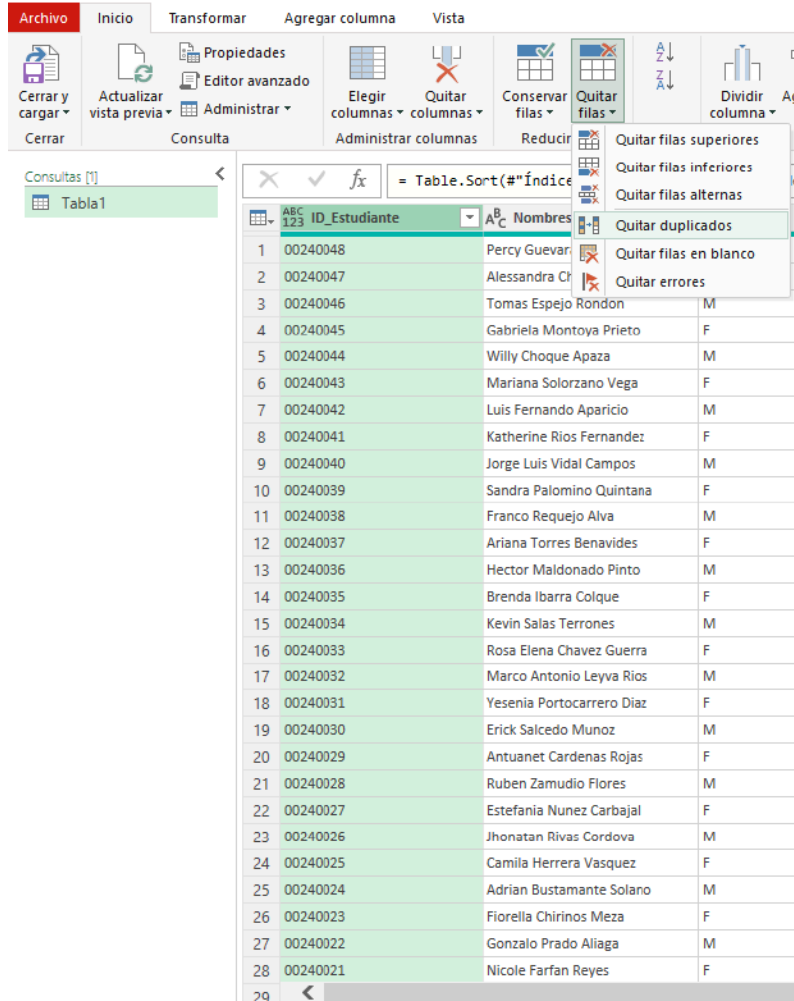


Índice	ID_Enudiante	Nombre
1	00240001	Andrea
2	00240002	José Luis Ramírez

- Haz clic en la flecha de filtro/orden del encabezado Índice y elige Orden descendente (Sort Descending), para que la fila que apareció más tarde en el archivo original quede primero.



- Haz clic únicamente en el encabezado de la columna ID_Estudiante (verifica que ninguna otra columna quede seleccionada al mismo tiempo).
- Ve a Inicio (Home) > Quitar filas > Quitar duplicados (Remove Rows > Remove Duplicates). Como solo la columna ID_Estudiante está seleccionada, Power Query compara únicamente ese campo y conserva la primera fila de cada ID repetido según el orden actual — que ahora es la más reciente, gracias al ordenamiento por Índice.



The screenshot shows the Danalytics application window. The top menu bar includes 'Archivo', 'Inicio', 'Transformar', 'Agregar columna', and 'Vista'. Below the menu is a ribbon with various icons for data manipulation. The main area displays a table with columns 'ID_Estudiante' and 'Nombres'. A right-click context menu is open over the 'ID_Estudiante' column header, showing options like 'Quitar duplicados' and 'Quitar errores'.

ID_Estudiante	Nombres
1 00240048	Percy Guevar
2 00240047	Alessandra Ch
3 00240046	Tomas Espejo Rondon
4 00240045	Gabriela Montoya Prieto
5 00240044	Willy Choque Apaza
6 00240043	Mariana Solorzano Vega
7 00240042	Luis Fernando Aparicio
8 00240041	Katherine Rios Fernandez
9 00240040	Jorge Luis Vidal Campos
10 00240039	Sandra Palomino Quintana
11 00240038	Franco Requejo Alva
12 00240037	Ariana Torres Benavides
13 00240036	Hector Maldonado Pinto
14 00240035	Brenda Ibarra Colique
15 00240034	Kevin Salas Terrones
16 00240033	Rosa Elena Chavez Guerra
17 00240032	Marco Antonio Leyva Rios
18 00240031	Yessenia Portocarrero Diaz
19 00240030	Erick Salcedo Munoz
20 00240029	Antuanet Cardenas Rojas
21 00240028	Ruben Zamudio Flores
22 00240027	Estefania Nunez Carbajal
23 00240026	Jhonatan Rivas Cordova
24 00240025	Camila Herrera Vasquez
25 00240024	Adrian Bustamante Solano
26 00240023	Fiorella Chirinos Meza
27 00240022	Gonzalo Prado Aliaga
28 00240021	Nicole Farfan Reyes

- Quita la columna Índice, que ya no se necesita: clic derecho sobre su encabezado > Quitar (Remove).

Importante

Este criterio de "quedarse con el registro más reciente" es una decisión de negocio, no una regla automática de Power Query. Documenta el paso (renómbralo, por ejemplo, a "Quitar duplicados parciales – se conserva el registro más reciente") para que cualquier persona que revise el flujo entienda por qué se descartó esa fila.

Paso 6 – Tratar valores perdidos y códigos centinela

Antes de decidir qué hacer con los datos faltantes, hay que convertir los códigos que se usaron como sustituto ("NA", "N/A", "-", 999) en valores nulos reales, para que Power Query los reconozca como datos faltantes de verdad.

Paso 6a — Convertir los códigos centinela en nulos reales

1. Clic derecho sobre el encabezado de Ingreso_Familiar_Mensual > Reemplazar valores (Replace Values).
2. En "Valor que se va a buscar": NA — deja "Reemplazar con" vacío — clic en Aceptar.

1.2 Estatura_m	1.2 Promedio_Ponderado	A ^B _C Ingreso_Familiar_Mensual	A ^B _C Fecha_Matricula	A ^B _C Becado
1.8	11.8	null	06/04/2024	No
1.61	14.1	S/ 1,600.00	05/04/2024	Sí

Reemplazar los valores

Reemplace un valor con otro de las columnas seleccionadas.

Valor que buscar

Reemplazar con

▸ Opciones avanzadas

Aceptar Cancelar

1.74	13.3	S/ 1,600.00	28/03/2024	No
1.59	16	S/ 2,000.00	28/03/2024	Sí

3. Repite de nuevo, pero usando N/A.
4. Repite sobre Fecha_Nacimiento con "Valor que se va a buscar": - (el guion usado como marcador de dato faltante).

[illegible]

19 de 30

Reemplazar los valores: Anular dinamización de columnas: Combinar columnas: Trigonon

Reemplazar: Mover: Dividir columna: Formato: Extraer: Estadísticas: Estándar: Científico: 10²: Redondear: Informac

Cualquier columna: Columna de texto: Columna de número

reValue({"Valor reemplazado3","N/A",""},Replacer.ReplaceValue,{"Edad"})

Nombres_y_Apellidos	Sexo	Fecha_Nacimiento	Edad	Programa	Ciudad_C
Y Guevara Cotrina		09/09/2000		24 Ingeniería Informática	Piura
Isandra Chuquilla	F	03 mar 2004		20 Diseño Gráfico	Trujillo
Las Espejo Rondon	M	08/08/2002		22 Contabilidad	Lima
Riela Montoya Prieto					
Y Choque Apaza					
Tania Solorzano Vega					
Fernando Aparicio					
Verine Rios Fernandez					
Le Luis Vidal Campos					
Dra Palomino Quintana					
Co Requejo Alva					
Na Torres Benavides					
Tor Maldonado Pinto					
Da Ibarra Colque					
In Salas Terrones					
A Elena Chavez Guerra	F	03/11/1998		27 Administración	Lima
Co Antonio Layva Rios	M	30/02/2002		22 Ingeniería Industrial	Piura
Enia Portocarrero Diaz	F	30/04/2003		21 Sociología	Trujillo
K Salcedo Munoz	M	22/10/2000		24 Ingeniería Mecánica	Cusco
Janet Cardenas Rojas	F			Psicología	Lima
En Zamudio Flores	M	15/07/2001		23 Derecho	Lima
Fania Nunez Carbajal	F	28/08/2002		22 Marketing	Arequipa
Tatan Rivas Cordova	M	20/03/2009		15 Ingeniería Informática	Lima
Sila Herrera Vasquez	F	01/05/2004		20 Educación Primaria	Ica
Ian Bustamante Solano	M	12/12/2000		24 Contabilidad	Huancayo
Elia Chirinos Meza	F	09/09/2003		20 Diseño Gráfico	Lima
Izalo Prado Aliaga	M	04/04/2002		22 Ingeniería Civil	Chiclayo
De Farfan Reyes	F	25/06/2003		21 Psicología	Lima

Reemplazar los valores

Reemplace un valor con otro de las columnas seleccionadas.

Valor que buscar: 999

Reemplazar con:

Aceptar Cancelar

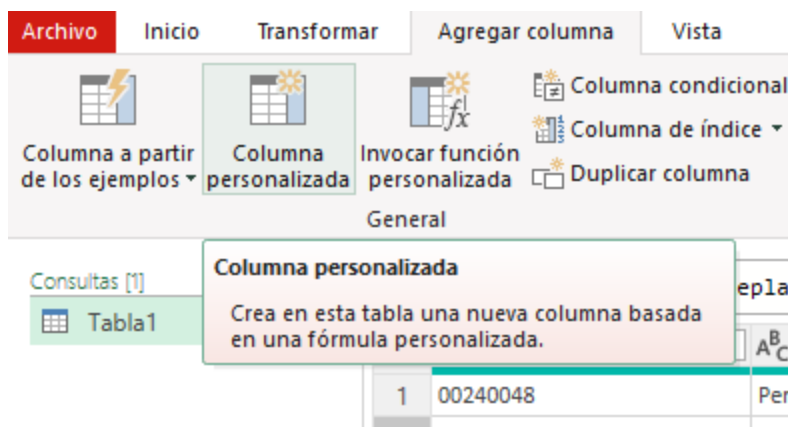
Paso 6b — Decidir qué hacer con los nulos resultantes

1. Haz clic en la flecha de filtro del encabezado de la columna con nulos (por ejemplo, Sexo).
2. En la lista de valores, revisa si aparece la opción (null) o (blank): esto te confirma cuántas filas están afectadas.
3. Si decides eliminar esas filas: desmarca todos los demás valores y deja marcado solo (null), clic en Aceptar — esto filtra y muestra solo esas filas; desde ahí puedes decidir eliminarlas con Inicio > Quitar filas, o volver a abrir el filtro y usar la opción contraria para excluirlas.
4. Si decides conservarlas para imputación posterior (con media, mediana, moda o un modelo, como se explicó en la sesión), no hagas nada más aquí: el valor null queda visible y documentado para que el equipo de negocio decida el tratamiento final.

Paso 7 — Estandarizar fechas con formatos mezclados y detectar fechas inválidas

Fecha_Nacimiento y Fecha_Matricula mezclan formatos DD/MM/AAAA, MM-DD-AAAA, AAAA/MM/DD, "12-mar-2024" y "03 mar 2004", además de dos fechas que no existen (31/02/2004 y 30/02/2002). Como estos formatos están mezclados en la misma columna, Power Query no puede convertirlos todos de una sola vez cambiando el tipo de columna: hay que aislar primero los que fallan.

1. Ve a Agregar columna (Add Column) > Columna personalizada (Custom Column).



2. Nombre de la columna: Fecha_Nacimiento_Valida.
3. En la fórmula, escribe:

CÓDIGO M — INTENTAR CONVERTIR LA FECHA Y AISLAR ERRORES

```
=try Date.From([Fecha_Nacimiento]) otherwise null
```

Columna personalizada

Agregue una columna que se calcula a partir de otras columnas.

Nuevo nombre de columna

Fecha_Nacimiento_Valida

Fórmula de columna personalizada ⓘ

```
= try Date.From([Fecha_Nacimiento]) otherwise null
```

Columnas disponibles

ID_Estudiante
Nombres_y_Apellidos
Sexo
Fecha_Nacimiento
Edad
Programa
Ciudad_Origen

<< Insertar

[Información sobre fórmulas de Power Query](#)

✓ No se han detectado errores de sintaxis.

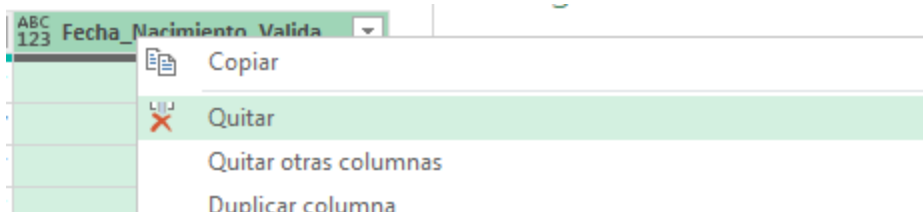
Aceptar

Cancelar

4. Clic en Aceptar (OK).
5. Haz clic en la flecha de filtro del encabezado de la nueva columna Fecha_Nacimiento_Valida y deja marcada únicamente la opción (null). Esto muestra solo las filas cuya fecha no se pudo convertir automáticamente: ahí aparecerán tanto los formatos atípicos ("12-mar-2024", "03 mar 2004") como las dos fechas inválidas.
6. Quita el filtro (vuelve a marcar "Seleccionar todo" en el encabezado de Fecha_Nacimiento_Valida) para ver la tabla completa de nuevo.
7. Para cada fila identificada como inválida en el paso 5, ve a la columna original Fecha_Nacimiento y reemplaza los valores (como los has hecho antes), escribiendo la fecha corregida en un formato simple y sin ambigüedad, por ejemplo 2024-03-12 (año-mes-día). Consulta la fuente original (el registro de matrícula) si tienes dudas sobre la fecha real de las dos fechas inválidas.



- 23 de 30



9. Repite todo este mismo procedimiento (columna personalizada con try...otherwise, filtrar por null, corregir a mano, quitar la columna auxiliar, cambiar el tipo) para la columna Fecha_Matricula.

Si tu configuración regional confunde día y mes

Si al cambiar el tipo a Fecha algunos valores como 12/03/2024 se interpretan como 3 de diciembre en vez de 12 de marzo, usa el botón de tipo de columna > Usando configuración regional (Using Locale) y elige explícitamente el idioma/formato de origen de los datos (por ejemplo, español (Perú)) antes de convertir.

Paso 8 — Corregir formatos que no coinciden con el tipo de dato

Paso 8 a — Ingreso_Familiar_Mensual (símbolo de moneda y separador de miles)

1. Clic derecho sobre el encabezado de Ingreso_Familiar_Mensual > Reemplazar valores (Replace Values).
2. "Valor que se va a buscar": S/ (incluyendo el espacio que le sigue, tal como aparece en los datos, por ejemplo "S/ ") — "Reemplazar con": deja vacío — Aceptar.

Reemplazar los valores

Reemplace un valor con otro de las columnas seleccionadas.

Valor que buscar

Reemplazar con

▸ Opciones avanzadas

3. Repite Reemplazar valores: "Valor que se va a buscar": , (la coma del separador de miles) — "Reemplazar con": deja vacío — Aceptar.



Reemplazar los valores

Reemplace un valor con otro de las columnas seleccionadas.

Valor que buscar

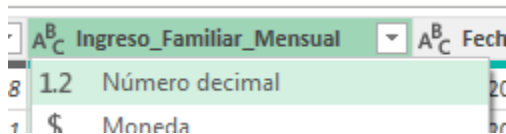
Reemplazar con

► Opciones avanzadas

Aceptar

Cancelar

4. Haz clic en el ícono de tipo de dato a la izquierda del encabezado (ABC/123) y elige Número decimal (Decimal Number).



Paso 8b — Edad (texto en un campo numérico)

1. Clic derecho sobre el encabezado de Edad > Reemplazar valores (Replace Values).
2. "Valor que se va a buscar": veintitrés — "Reemplazar con": 23 — Aceptar.

	ABC 123	Edad	ABC 123	Programa	ABC 123	Ciudad_Origen	123
	09/09/2000		24	Ingeniería Informática		Piura	
	03 mar 2004		20	Diseño Gráfico		Trujillo	
	08/08/2002		22	Contabilidad		Lima	
	05/12/2003		20	Educación Inicial		Lima	
	14/07/2001		23	Ingeniería Industrial		Arequipa	
	21/03/2002		22	Sociología		Iquitos	
	10/10/2005		40	Derecho		Lima	
	16/05/2003		21	Marketing		Lima	
	28/08/2002		22	Marketing		Arequipa	

Reemplazar los valores

Reemplace un valor con otro de las columnas seleccionadas.

Valor que buscar

Reemplazar con

Aceptar

Cancelar

3. Cambia el tipo de columna a Número entero (Whole Number). Si quedó algún valor no numérico adicional, Power Query lo mostrará como Error en la celda, lo cual te permite ubicarlo de inmediato con el filtro de la columna.

ABC 123	Edad	ABC 123	Programa
1.2	Número decimal		
\$	Moneda		
123	Número entero		

Paso 9 — Marcar valores fuera del diccionario de datos

El diccionario de datos indica que Estado_Civil solo puede tomar los valores 1 a 5, pero una fila tiene el código 8. En vez de borrar o adivinar el valor correcto, se marca para que un analista lo revise contra la fuente original.

1. Ve a Agregar columna (Add Column) > Columna personalizada (Custom Column).
2. Nombre de la columna: Validacion_EstadoCivil.

3. Fórmula:

CÓDIGO M — VALIDAR CONTRA EL DICCIONARIO DE DATOS

```
=if List.Contains({1,2,3,4,5}, [Estado_Civil]) then "OK" else "Revisar"
```



Columna personalizada

Agregue una columna que se calcula a partir de otras columnas.

Nuevo nombre de columna

Validacion_EstadoCivil

Fórmula de columna personalizada ⓘ

```
= if List.Contains({1,2,3,4,5}, [Estado_Civil]) then "OK"  
    else "Revisar"
```

Columnas disponibles

ID_Estudiante
Nombres_y_Apellidos
Sexo
Fecha_Nacimiento
Edad
Programa
Ciudad_Origen
Estado_Civil

<< Insertar

[Información sobre fórmulas de Power Query](#)

✓ No se han detectado errores de sintaxis.

Aceptar

Cancelar

4. Clic en Aceptar (OK).

BEC 123	Validacion_EstadoCivil
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK
	OK

5. Haz clic en la flecha de filtro del encabezado de Validacion_EstadoCivil y deja marcado solo "Revisar". Aparecerá la fila con el código 8.
6. Corrige el valor manualmente en la columna Estado_Civil consultando la fuente original, o dejarlo como pendiente de confirmación documentando el caso.
7. Quita el filtro y, cuando ya no lo necesites, elimina la columna auxiliar Validacion_EstadoCivil (clic derecho > Quitar).

Paso 10 — Marcar inconsistencias lógicas entre columnas

Estos valores no están "mal escritos": son números y fechas con formato correcto, pero que no tienen sentido lógico al compararlos con otra columna (una edad que no coincide con la fecha de nacimiento, una estatura negativa, etc.). Se marcan con una columna de validación en lugar de corregirse automáticamente, porque decidir el valor correcto requiere criterio humano o volver a la fuente original.

Repite el procedimiento del Paso 9 (Agregar columna > Columna personalizada > filtrar por "Revisar" > corregir o documentar > quitar la columna auxiliar) una vez por cada una de estas reglas:

CÓDIGO M — EDAD VS. FECHA_NACIMIENTO (COLUMNA: VALIDACION_EDAD)

```
=if Date.Year(DateTime.LocalNow()) - Date.Year([Fecha_Nacimiento]) <> [Edad]
then "Revisar" else "OK"
```

CÓDIGO M — ESTATURA_M FUERA DE UN RANGO HUMANO RAZONABLE (COLUMNA: VALIDACION_ESTATURA)

```
=if [Estatura_m] < 1.3 or [Estatura_m] > 2.1 then "Revisar" else "OK"
```

CÓDIGO M — PROMEDIO_PONDERADO FUERA DE LA ESCALA 0-20 (COLUMNA: VALIDACION_PROMEDIO)

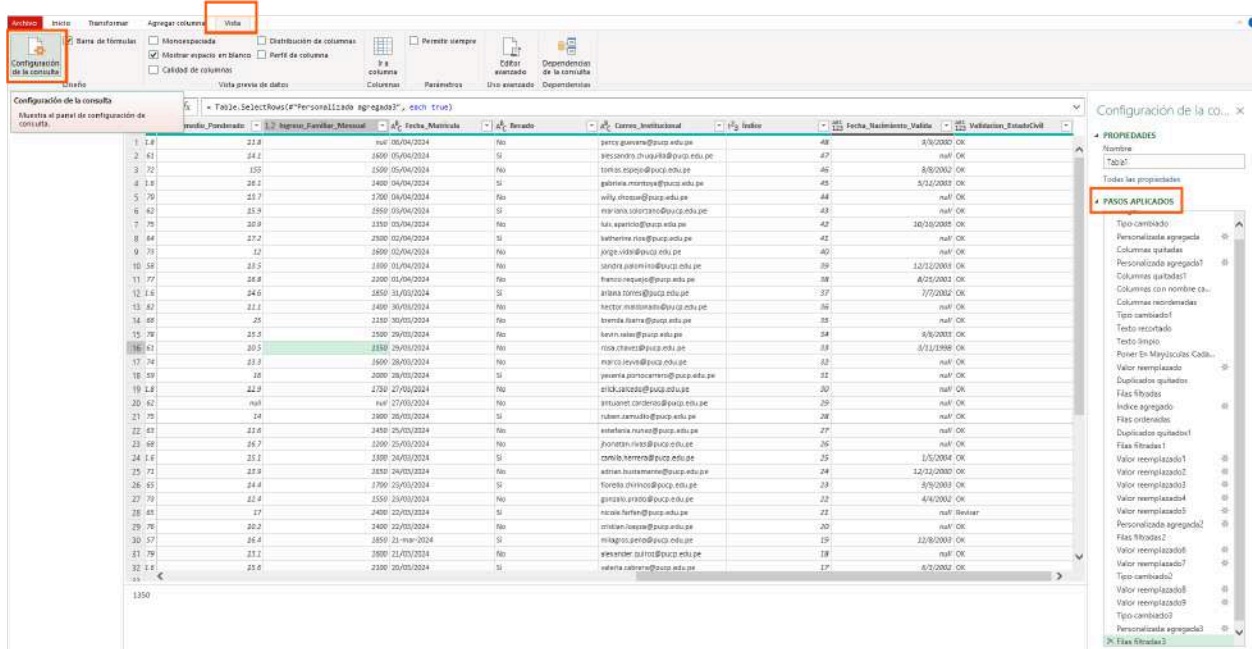
```
=if [Promedio_Ponderado] < 0 or [Promedio_Ponderado] > 20 then "Revisar"
else "OK"
```

Nota sobre la fórmula de Edad

`DateTime.LocalNow()` usa la fecha del día en que se ejecuta la consulta, así que el resultado de esta validación cambiará ligeramente con el tiempo. Si prefieres un resultado reproducible para este ejercicio, puedes reemplazar `DateTime.LocalNow()` por una fecha fija, por ejemplo `#date(2026,8,7)`.

Paso 11 — Documentar y cargar el resultado final

1. Abre el panel "Pasos aplicados" (Applied Steps) en el lado derecho del Editor de Power Query y revísalo de arriba hacia abajo: cada acción de esta guía debe aparecer como un paso independiente.



2. Renombra los pasos con nombre poco descriptivo (por ejemplo "Reemplazar valores"): clic derecho sobre el paso > Renombrar (Rename) y escribe algo claro, como "Corregir tilde en Ingenieria".
3. Verifica que ya no queden columnas auxiliares que solo se usaron para detectar errores (Índice, Fecha_Nacimiento_Valida, Validacion_EstadoCivil, etc.) salvo que decidas conservarlas a propósito como columnas de auditoría.
4. Ve a la pestaña Inicio (Home) y haz clic en Cerrar y aplicar (Close & Load). Como los datos se cargaron desde la propia hoja del libro (Paso 1), el resultado limpio se reemplaza en el mismo libro de Excel, como una tabla ya lista para usar.

