

Taller en clase

Nombre: Erick Rodas

Limpieza de países

Customers.csv

Antes:

The screenshot shows the Tableau Prep Builder interface. On the left, there's a sidebar with 'Conexiones' (Connections) listing several CSV files and a 'Tables' section with 'customers'. The main workspace shows a flow from a 'customers' source to a 'Limpiar 2' step. Below the step, four columns are displayed: 'customer_name', 'email', 'city', 'country', and 'segment'. The 'country' column contains values like 'Colombia', 'Ecuador', and 'Peru'. A context menu is open over the 'country' column, with the option 'Convertir a mayúsculas' (Convert to uppercase) highlighted.

Después:

This screenshot shows the same Tableau Prep Builder interface after the 'country' column has been converted to uppercase. The 'country' column now displays values in uppercase: 'COLOMBIA', 'ECUADOR', and 'PERU'. The 'segment' column also shows some changes, with one row now having 'Corporate' instead of 'Corporate'. The rest of the table remains the same.

Explicación: Me di cuenta de que los países se repiten al ser escritos en diferentes formatos, por lo tanto, aplique la opción de limpieza de pasarlo todo a mayúsculas para q se sume automáticamente y esté todo en un solo formato

Products.csv

Antes:

Después:

Explicación: Realizamos una división de valores para que la tabla sea mas sencilla de entender y evitar tanto problema o repetir mucho el texto

Sales q1.csv

Antes:

Limpiar 5 10 campos 120 filas | Filtrar valores... | ... | 2 Recomendaciones | Buscar

Cambios (0)

#	unit_price	country	city	region	channel
0	29,9	Colombia	Arequipa	andina	Marketplace
29,9	49,9	Ecuador	Bogotá	Andina	Online
49,9	79	Peru	Cuenca	costa	Tienda
79	120		Guayaquil	Costa	
120	350		Lima	sierra	
350			Loja	Sierra	
			Medellín		
			Quito		

order_id order_date customer_id product_id qty unit_price country city region channel

EC-5001 10/02/2025 1.012 2.003 2 120 Colombia Medellín Andina Marketplace

EC-5002 06/03/2025 1.009 2.007 5 0 Peru Lima Costa Tienda

Despues:

Limpiar 5 10 campos 120 filas | Filtrar valores... | ... | 2 Recomendaciones | Buscar

Cambios (1)

#	unit_price	country	city	region	channel
0	29.9	Colombia	Arequipa	ANDINA	Marketplace
29.9	49.9	Ecuador	Bogotá	COSTA	Online
49.9	79	Peru	Cuenca	SIERRA	Tienda
79	120		Guayaquil		
120	350		Lima		
350			Loja		
			Medellín		
			Quito		

order_id order_date customer_id product_id qty unit_price country city region channel

EC-5001 10/02/2025 1.012 2.003 2 120 Colombia Medellín ANDINA Marketplace

Explicacion: Normalizamos los valores de región pasandolo todo a mayúsculas, con esto tenemos una mejor visión de las regiones sin repetir y redundar los datos debido al formato

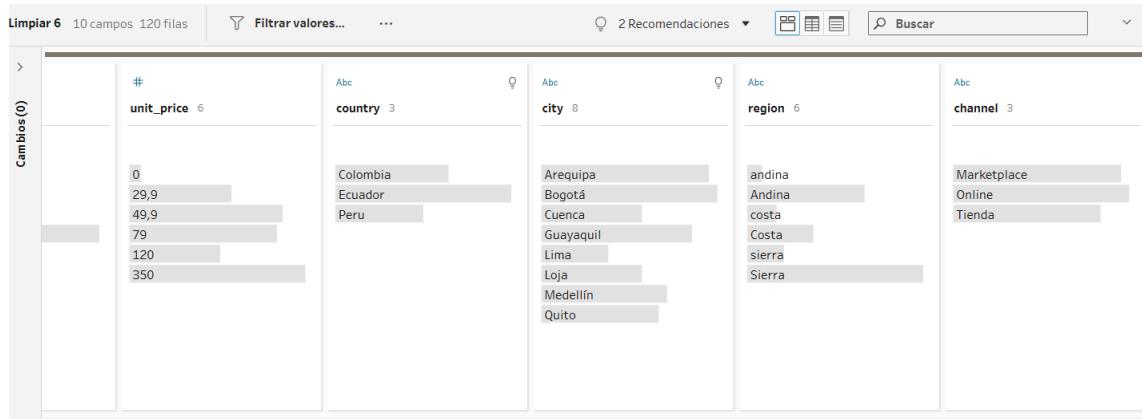
Sales Q2.csv

Antes:

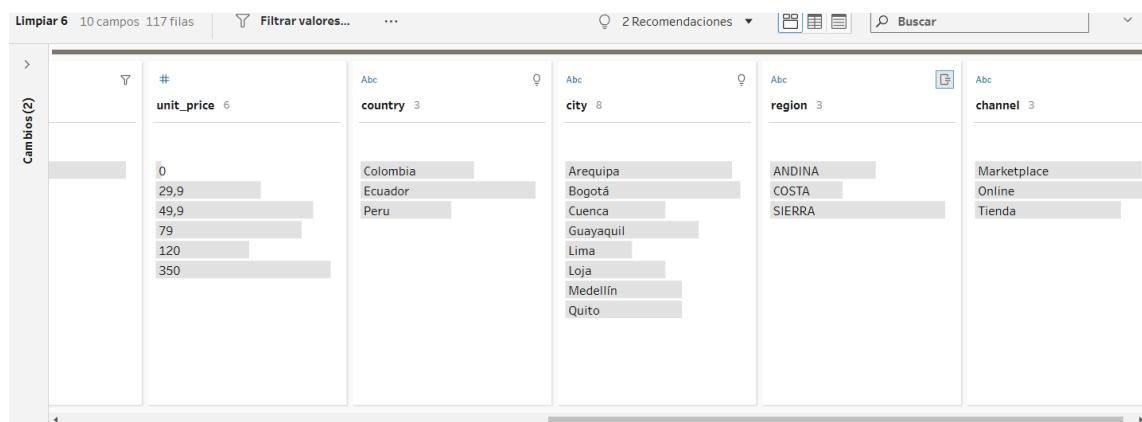
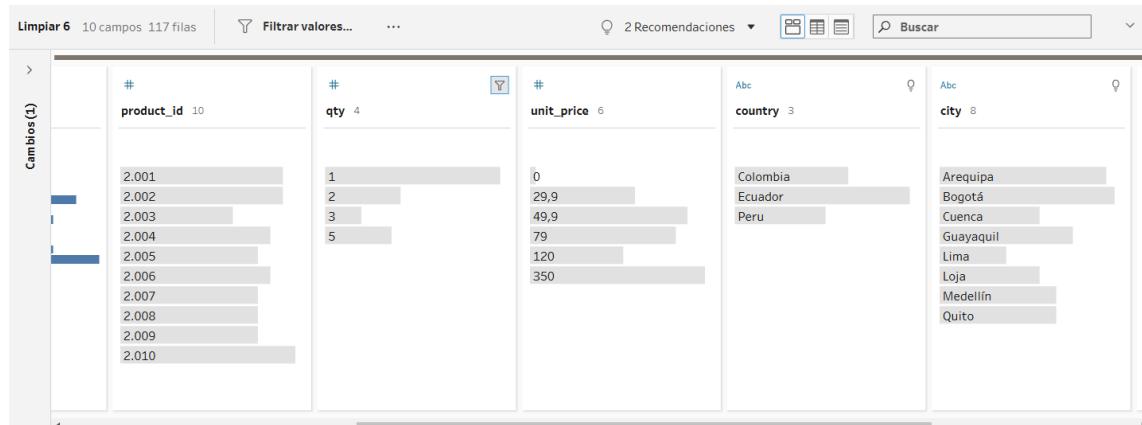
Limpiar 6 10 campos 120 filas | Filtrar valores... | ... | 2 Recomendaciones | Buscar

Cambios (0)

#	customer_id	#	product_id	#	qty	#	unit_price	#	country	#	city
1	2.001	2	2.002	3	-3	4	0	5	Colombia	6	Arequipa
2	2.002	3	2.003	4	-2	5	29,9	6	Ecuador	7	Bogotá
3	2.003	4	2.004	5	-1	6	49,9	7	Peru	8	Cuenca
4	2.004	5	2.005	6	1	7	79	8		9	Guayaquil
5	2.005	6	2.006	7	2	8	120	9		10	Lima
6	2.006	7	2.007	8	3	9	350	10		11	Loja
7	2.007	8	2.008	9	5	10		11		12	Medellín
8	2.008	9	2.009	10		11		12		13	Quito
9	2.009	10	2.010			11		12		13	



Despues:



Explicacion: Eliminamos los valores de cantidad negativos, ya que no podemos vender -1 objetos por ejemplo, por lo tanto lo mejor es filtrarlos y quedarnos con los valores que si sean mayor a 0, posteriormente realizamos la misma limpieza de regiones para evitar datos redundantes