

Taller en clase

Nombre: Erick Rodas

Limpieza de países

Customers.csv

Antes:

The screenshot shows the Tableau Prep Builder interface. On the left, the 'Conexiones' pane lists data sources: customers.csv, geo_regions.csv, products.csv, sales_2025_q1.csv, and sales_2025_q2.csv. The main workspace displays a flow with a 'Limpiar 2' step. The 'Limpiar 2' step is selected, and a context menu is open over the 'country' field. The menu options include: Filtro, Limpiar, Agrupar valores, Dividir valores, Identificar filas duplicadas, Ver estado, ✓ Detalle, Resumen, Cambiar el nombre del campo, Duplicar campo, Mantener solamente el campo, Crear campo calculado, Ocultar campo, and Eliminar. The data preview shows columns: customer_id, customer_name, email, city, country, and segment. The 'country' column contains values like Colombia, Ecuador, and Peru.

Despues:

The screenshot shows the Tableau Prep Builder interface after the 'Limpiar 2' step. The 'country' field is now uppercase, and the 'segment' field is also visible. The data preview shows columns: customer_id, customer_name, email, city, country, and segment. The 'country' column contains values like COLOMBIA, ECUADOR, and PERU. The 'segment' column contains values like Consumer, Corporate, and Home Office.

Explicacion: 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:

Limpiar 4 4 campos 10 filas

Filtrar valores...

1 recomendación

Buscar

#	product_id 10	category 3	subcategory 4	product_name 10
2.001		Hogar	Accesorios	Producto 2001
2.002		Oficina	Cocina	Producto 2002
2.003		Tech	Laptops	Producto 2003
2.004			Muebles	Producto 2004
2.005				Producto 2005
2.006				Producto 2006
2.007				Producto 2007
2.008				Producto 2008
2.009				Producto 2009
2.010				Producto 2010

product_id

category

subcategory

product_name

2.001	Tech	Accesorios	Producto 2001
2.002	Tech	Accesorios	Producto 2002

Después:

Limpiar 4 6 campos 10 filas

Identificar filas duplicadas

1 recomendación

Buscar

#	product_id 10	category 3	subcategory 4	product_name 10	product_name 1	product_name 10
2.001		Hogar	Accesorios	Producto 2001	Producto	2001
2.002		Oficina	Cocina	Producto 2002		2002
2.003		Tech	Laptops	Producto 2003		2003
2.004			Muebles	Producto 2004		2004
2.005				Producto 2005		2005
2.006				Producto 2006		2006
2.007				Producto 2007		2007
2.008				Producto 2008		2008
2.009				Producto 2009		2009
2.010				Producto 2010		2010

Explicacion: 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 6	country 3	city 8	region 6	channel 3
0		Colombia	Arequipa	andina	Marketplace
29,9		Ecuador	Bogotá	Andina	Online
49,9		Peru	Cuenca	costa	Tienda
79			Guayaquil	Costa	
120			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 6	country 3	city 8	region 3	channel 3
0		Colombia	Arequipa	ANDINA	Marketplace
29,9		Ecuador	Bogotá	COSTA	Online
49,9		Peru	Cuenca	SIERRA	Tienda
79			Guayaquil		
120			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)

order_id 20	product_id 10	qty 7	unit_price 6	country 3	city 8
	2.001	-3	0	Colombia	Arequipa
	2.002	-2	29,9	Ecuador	Bogotá
	2.003	-1	49,9	Peru	Cuenca
	2.004	1	79		Guayaquil
	2.005	2	120		Lima
	2.006	3	350		Loja
	2.007	5			Medellín
	2.008				Quito
	2.009				
	2.010				

Limpiar 6 10 campos 120 filas

Filtrar valores...

2 Recomendaciones

Buscar

#	unit_price 6	country 3	city 8	region 6	channel 3
0		Colombia	Arequipa	andina	Marketplace
29,9		Ecuador	Bogotá	Andina	Online
49,9		Peru	Cuenca	costa	Tienda
79			Guayaquil	Costa	
120			Lima	sierra	
350			Loja	Sierra	
			Medellín		
			Quito		

Despues:

Limpiar 6 10 campos 117 filas

Filtrar valores...

2 Recomendaciones

Buscar

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

Limpiar 6 10 campos 117 filas

Filtrar valores...

2 Recomendaciones

Buscar

#	unit_price 6	country 3	city 8	region 3	channel 3
0		Colombia	Arequipa	ANDINA	Marketplace
29,9		Ecuador	Bogotá	COSTA	Online
49,9		Peru	Cuenca	SIERRA	Tienda
79			Guayaquil		
120			Lima		
350			Loja		
			Medellín		
			Quito		

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 limpieza de regiones para evitar datos redundantes