



## *Ejercicio de Laboratorio 7: ETL – P2*

Instituto Politécnico Nacional.  
Escuela Superior de Cómputo.  
Licenciatura en ciencia de datos.  
Bases de Datos Avanzadas

Emiliano López Méndez.  
Hinostroza Loera Leonardo  
Antonio Aguilar Velazquez  
Escudero Robles Arantxa  
Lugo Acosta Paulina

# Introduccion

En el presente laboratorio, abordaremos un flujo completo de proceso ETL (Extracción, Transformación y Carga), utilizando herramientas avanzadas para el manejo y limpieza de datos. Este ejercicio práctico se enfoca en asegurar la calidad de los datos, eliminando valores faltantes y resolviendo códigos postales incompletos mediante la integración de información adicional. La finalidad de esta actividad es fortalecer nuestras competencias en la manipulación y transformación de datos, con el objetivo de prepararlos para análisis profundos o su almacenamiento en bases de datos, procesos fundamentales en el campo de la ciencia de datos..

## Desarrollo de la Actividad:

**Extracción de Datos:** Se inicia el proceso con la carga de datos desde un archivo CSV. En esta fase, se usa un paso de entrada de archivo de texto para acceder al archivo `sales_data.csv`, visualizar su estructura y preparar los datos para los pasos siguientes.

**Filtrado de Registros con Códigos Postales Faltantes:** Se aplica un filtro para retener solo aquellos registros completos en los que se disponga de un código postal, evitando cargar datos incompletos en la base de datos.

**Resolución de Códigos Faltantes:** Para los registros sin código postal, se integran datos adicionales de un segundo archivo que contiene información de ciudades, estados y códigos postales. Esto permite enriquecer los datos originales y completar los campos vacíos.

**Limpieza y Normalización de Datos:** Una vez que los datos cuentan con la información necesaria, se efectúan ajustes adicionales. En el campo país, se unifican los valores, reemplazando términos como "Estados Unidos" por "EE. UU." y categorizando las ventas en pequeñas, medianas y grandes.

**Ejecución y Orquestación:** Finalmente, se utiliza la herramienta Pentaho para coordinar y ejecutar los trabajos de transformación de manera programada. Este proceso asegura la consistencia y la actualización periódica de los datos, simulando un escenario en el cual un sistema externo coloca el archivo de entrada en una ubicación específica, y el trabajo de ETL se ejecuta automáticamente.

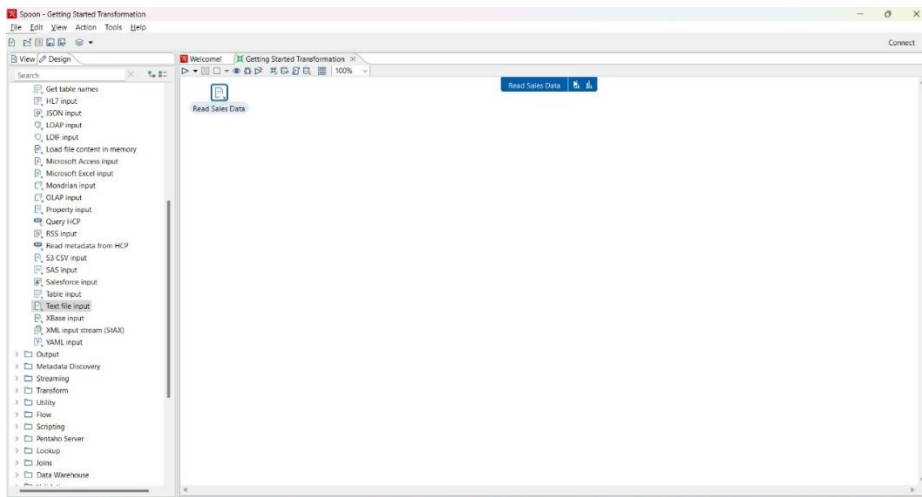
# Parte 1 Extract and load data

En el Paso 1, recuperará datos de un archivo plano CSV y utilizará el paso Entrada de archivo de texto para conectarse a un repositorio, ver el esquema del archivo y recuperar el contenido de los datos.

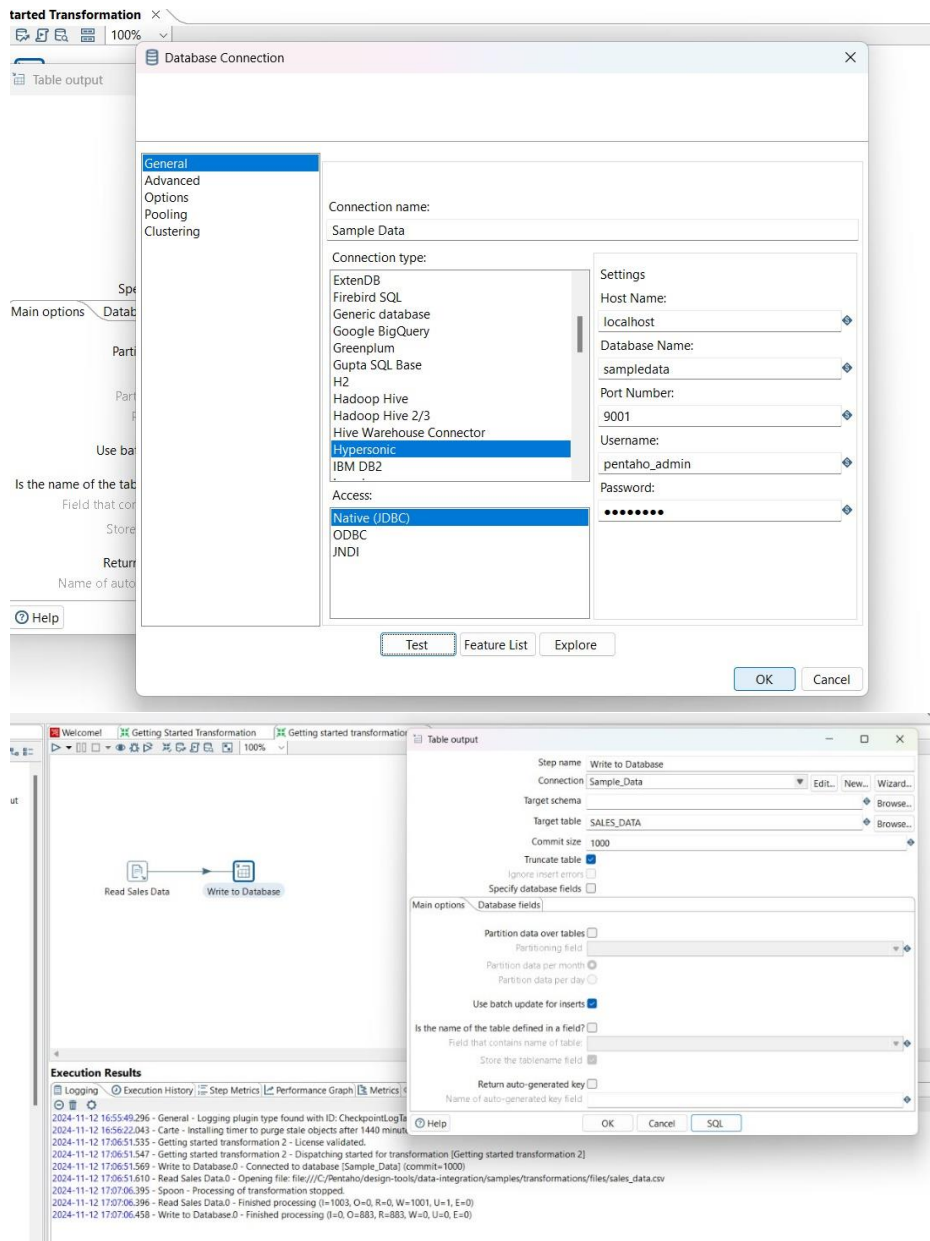
The screenshot shows the 'Text file input' step configuration in a data integration tool. The 'File or directory' field is set to 'C:\Users\emlcom\OneDrive\Documents\Asamblea\Bases-de-Datos-Avanzadas\Laboratorio 6\sales\_data.csv'. The 'Preview rows' button is visible at the bottom.

Below the configuration window, a preview of the data is shown. The data is a CSV file with the following columns: C/CODE, CUSTOMERNAME, PHONE, ADDRESSLINE1, ADDRESSLINE2, CITY, STATE, POSTALCODE, COUNTRY, TERRITORY, CONTACTLASTNAME, CONTACTFIRSTNAME.

C/CODE	CUSTOMERNAME	PHONE	ADDRESSLINE1	ADDRESSLINE2	CITY	STATE	POSTALCODE	COUNTRY	TERRITORY	CONTACTLASTNAME	CONTACTFIRSTNAME
78	Land of Toys Inc.	2125537818	897 Long Airport Avenue	<null>	NYC	NY	10022	United States	NA	Yu	Kwai
78	Reims Collectables	26.47.1555	59 rue de l'Abbaye	<null>	Reims	<null>	51130	France	EMEA	Hennart	Paul
78	Lyon Souvenirs	+33 1 48 60 7555	27 rue du Colonel Pierre Jura	<null>	Paris	<null>	75336	France	EMEA	Da Cunha	Daniel
78	ThornElectronics.com	6055557405	7854 Holmdel Dr	<null>	Roseland	CA	90003	USA	NA	Young	Julie
78	Corporate Gift Ideas Co.	6055551186	7734 Strong St.	<null>	San Francisco	CA	<null>	United States	NA	Brown	Julie
78	Technix Stores Inc.	6055556809	9408 Furr Circle	<null>	Burlingame	CA	94017	USA	NA	Hirano	Jun
78	Dandel Designs Imports	20.16.1555	184, Chemise de Journal	<null>	Ville	<null>	59000	France	EMEA	Ranc	Martine
78	Henrik Gifts	+47 2267 3215	Diamanten 121, 98 744 Sentrum	<null>	Bergen	<null>	N 5804	Norway	EMEA	Oestun	Veyel
78	Mini Wheels Co.	6055555787	5507 North Rendale Street	<null>	San Francisco	CA	<null>	United States	NA	Murphy	Julie
78	Auto Camp - Petit	(1) 47 554555	25 rue Lauriston	<null>	Paris	<null>	75016	France	EMEA	Jenier	Dominique
78	Australian Collectors Co.	03 9525 4555	436 St Kilda Road	Level 3	Melbourne	Victoria	3004	Australia	APAC	Ferguson	Peter
78	Vitaforme Inc.	2125551500	2678 Kingston Rd.	Suite 101	NYC	NY	10022	USA	NA	Trick	Michael
78	Teen Collectables Inc.	2015559100	7476 Moss Rd.	<null>	Newark	NJ	94019	USA	NA	Brown	William
78	Gift Depot Inc.	2035552570	25585 South Bay Ln.	<null>	Bridgewater	CT	07520	United States	NA	King	Julie
78	La Rochelle Gifts	40.67.8555	67, rue des Croquante Otages	<null>	Nantes	<null>	44000	France	EMEA	Labrune	Janine
78	Marta's Replicas Co.	6175558555	39323 Spinnaker Dr.	<null>	Cambridge	MA	51247	USA	NA	Hernandez	Marta
78	Toys of Finland Co.	90-224 8555	Keskikatu 45	<null>	Helsinki	<null>	21240	Finland	EMEA	Karttunen	Matti
78	Reims Mini Imports	07 48 8555	Erving Steiner gate 78	<null>	Stavem	<null>	4113	Norway	EMEA	Bergfjell	Jens
78	Discreet Classics Inc.	2155551555	7566 Remington St.	<null>	Albiontown	PA	70307	USA	NA	Yu	Kyang
78	Land of Toys Inc.	2125557818	897 Long Airport Avenue	<null>	NYC	NY	10022	United States	NA	Yu	Kwai
78	Schizony Collectables	6502-9335	Gesleweg 14	<null>	Schizony	NOW	5020	Austria	EMEA	Pipps	Georg
78	Southernmost Adult Things Co.	+61 2 7480 8555	Monter Money Building, 815 Pacific Hwy	Level 6	Cherrybrook	NOW	2067	Australia	APAC	Hendley	Adrian
78	La Rochelle Gifts	40.67.8555	67, rue des Croquante Otages	<null>	Nantes	<null>	44000	France	EMEA	Labrune	Janine
78	FunGiftsInc.com	5085552555	1785 First Street	<null>	New Bedford	MA	50553	USA	NA	Bentley	Violeta



Documentos > 4semestre > Bases-de-Datos-Avanzadas >					
Ordenar Ver					
Nombre		Fecha de modificación	Tipo	Tamaño	
Laboratorio 6		11/11/2024 3:59 PM	Carpeta de archivos		
Laboratorio 7		11/11/2024 4:22 PM	Carpeta de archivos		
Laboratorio4		10/21/2024 7:16 PM	Carpeta de archivos		
Getting Started Transformation .ktr		11/11/2024 4:23 PM	Archivo KTR	24 KB	
README.md		10/17/2024 4:18 PM	Archivo de origen ...	1 KB	



## Parte 2 Filter for missing codes

Después de completar el Paso 1: Extraer y cargar datos, estará listo para agregar un componente de transformación a su canalización de datos. El archivo fuente contiene varios registros a los que les faltan códigos postales. Esta sección del tutorial filtra aquellos registros a los que les faltan códigos postales, donde el POSTALCODE no es nulo (la condición verdadera) y garantiza que solo se carguen registros completos en la tabla de la base de datos.

is Help

Welcome! Getting Started Transformation

Rows of step: Read Sales Data (1000 rows)

#	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES	ORDERDATE	STATUS	QTR_ID
1	10107	30	95.7	2	2871	2/24/2003 0:00	Shipped	1
2	10121	34	81.35	5	2765.9	5/7/2003 0:00	Shipped	2
3	10134	41	94.74	2	3884.34	7/1/2003 0:00	Shipped	3
4	10145	45	83.26	6	3746.7	8/25/2003 0:00	Shipped	3
5	10159	49	100	14	5205.27	10/10/2003 0:00	Shipped	4
6	10168	36	96.66	1	3479.76	10/28/2003 0:00	Shipped	4
7	10180	29	86.13	9	2497.77	11/11/2003 0:00	Shipped	4
8	10188	48	100	1	5512.32	11/18/2003 0:00	Shipped	4
9	10201	22	98.57	2	2168.54	12/1/2003 0:00	Shipped	4
1.	10211	41	100	14	4708.44	1/15/2004 0:00	Shipped	1
1.	10223	37	100	1	3965.66	2/20/2004 0:00	Shipped	1
1.	10237	23	100	7	2333.12	4/5/2004 0:00	Shipped	2
1.	10251	28	100	2	3188.64	5/18/2004 0:00	Shipped	2
1.	10263	34	100	2	3676.76	6/28/2004 0:00	Shipped	2
1.	10275	45	92.83	1	4177.35	7/23/2004 0:00	Shipped	3
1.	10285	36	100	6	4099.68	8/27/2004 0:00	Shipped	3
1.	10299	23	100	9	2597.39	9/30/2004 0:00	Shipped	3
1.	10309	41	100	5	4394.38	10/15/2004 0:00	Shipped	4
1.	10318	46	94.74	1	4358.04	11/2/2004 0:00	Shipped	4
2.	10329	42	100	1	4396.14	11/15/2004 0:00	Shipped	4
2.	10341	41	100	9	7737.93	11/24/2004 0:00	Shipped	4
2.	10361	20	72.55	13	1451	12/17/2004 0:00	Shipped	4
2.	10375	21	34.91	12	733.11	2/3/2005 0:00	Shipped	1
2.	10388	42	76.36	4	3207.12	3/3/2005 0:00	Shipped	1

Execution Results

Logging Execution History Step Metrics

2024-11-12 17:08:43.521 - Getting started transformation  
2024-11-12 17:08:43.521 - Getting started transformation  
2024-11-12 17:08:43.522 - Write to Database.0 - Connect  
2024-11-12 17:08:43.524 - Read Sales Data.0 - Opening fi

Close Stop Get more rows



```
graph LR; A[Read Sales Data] --> B[Filter rows]; B --> C[Write to Database];
```

Filter rows

Step name: Filter Missing Zips

Send 'true' data to step: Write to Database

Send 'false' data to step:

The condition:

POSTALCODE IS NOT NULL

Help OK Cancel

2 17:08:43.522 - Write to Database.0 - Connected to database [Sample\_Data] (commit=1000)

## Parte 3 Resolve missing codes

Después de completar el Paso 2: Filtrar los códigos faltantes, estará listo para resolver los códigos postales faltantes. En esta sección, aprenderá a utilizar un segundo archivo de texto que contiene una lista de ciudades, estados y códigos postales, para buscar los códigos postales de aquellos registros en los que faltan campos, que es la rama falsa de su Filtro. paso de filas.

Primero, utilizará un paso de entrada de archivo de texto para leer el archivo fuente. Luego, utilizará un paso de búsqueda de secuencia para incorporar los códigos postales resueltos a la secuencia. Por último, utilizará el paso Seleccionar valores para cambiar el nombre de los campos en la secuencia, eliminar campos innecesarios y más.

Step name: Read Postal Codes

Additional output fields

File or directory:

Regular Expression:

Exclude Regular Expression:

Selected files:

#	File/Directory	Wildcard (RegExp)	Exclude wildcard	Required	Include as	Delete	Edit
1	C:\Pentaho\design-tools\data-integration\samples\transformations\files\zipsortedbycitystate.csv			N	N		

Accept filenames from previous step ☐

Pass through fields from previous step ☐

Step to read filenames from:

Field in the input to use as filename:

Show filename(s)

Step name: Read Postal Codes

File | Content | Error handling | Filters | Fields | Additional output fields

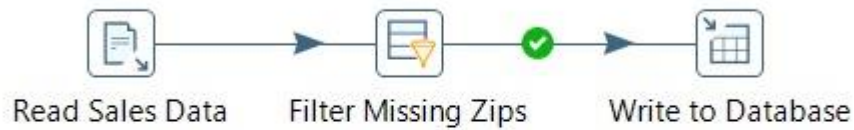
#	Name	Type	Format	Position	Length	Precision	Currency	Decimal	Group	Null if	Default	Trim type	Repeat
1	CITY	String			25		\$	.	-	-		none	N
2	STATE	String			2		\$	.	-	-		none	N
3	POSTALCODE	Integer	#		15	0	\$	.	-	-		none	N

ep name Read Postal Codes Examine preview data

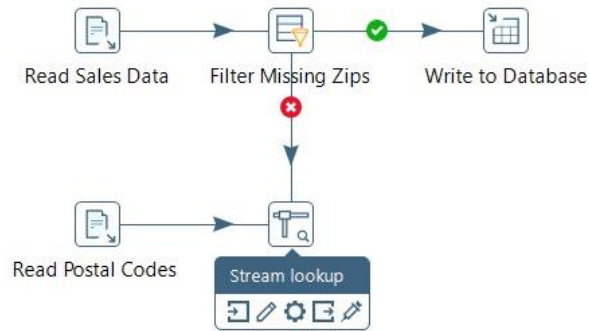
Rows of step: Read Postal Codes (1000 rows)

#	CITY	STATE	POSTALCODE
1	ABBEVILLE	AL	36310
2	ABBEVILLE	LA	70510
3	ABBEVILLE	MS	38601
4	ABBOT	ME	4406
5	ABBOTT	TX	76621
6	ABBYVILLE	KS	67510
7	ABERCROMBIE	ND	58001
8	ABERDEEN	KY	42201
9	ABERDEEN	MS	39730
1.	ABERDEEN	OH	45101
1.	ABERDEEN	SD	57402
1.	ABERDEEN PROVING GROUND	MD	21005
1.	ABERNATHY	TX	79311
1.	ABILENE	KS	67410
1.	ABILENE	TX	79602
1.	ABILENE	TX	79604
1.	ABILENE	TX	79606
1.	ABILENE	TX	79697
1.	ABILENE	TX	79699
2.	ABINGDON	MD	21009
2.	ABINGDON	VA	24211
2.	ABINGTON	CT	6230
2.	ABINGTON	PA	19001
2.	ABITA SPRINGS	LA	70420
2.	ABSARAKA	ND	58002

Close Show Log







## Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data



2024-11-12 17:08:43.521 - Getting started transformation 2 - License validated.

2024-11-12 17:08:43.521 - Getting started transformation 2 - Dispatching started for transformation [Getting started transformation 2]

nsformation Getting started transformation 2

Stream lookup

Step name: Lookup Missing Zips

Lookup step: Read Postal Codes

The key(s) to look up the value(s):

#	Field	LookupField
1	CITY	CITY
2	STATE	STATE

Specify the fields to retrieve :

#	Field	New name	Default	Type
1	POSTALCODE	ZIP_RESOLVED		String

Preserve memory (costs CPU) ☒

Key and value are exactly one integer field ☐

Use sorted list (i.s.o. hashtable) ☒

Help OK Cancel Get Fields Get lookup fields

ed transformation 2 - Dispatching started for transformation [Getting started transformation 2]

Transformation

Getting started transformation 2

100%

Read Sales Data

Filter Missing Zips

Write to Database

Read Postal Codes

Lookup Missing Zips

Examine preview data

Rows of step: Lookup Missing Zips (76 rows)

#	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES
1	10159	49	100	14	5205.27
2	10201	22	98.57	2	2168.54
3	10333	26	100	3	3003
4	10381	36	100	3	8254.8
5	10159	37	100	17	5016.83
6	10201	24	100	5	3025.92
7	10159	22	100	16	4132.7
8	10201	49	100	4	8065.89
9	10209	39	100	8	5197.92
1..	10384	34	100	4	4846.7
1..	10381	37	100	6	6231.54
1..	10159	41	100	2	8296.35
1..	10333	33	99.21	6	3273.93
1..	10381	20	100	1	2952
1..	10159	38	100	13	6238.84
1..	10201	25	100	1	4029
1..	10160	46	100	6	5294.14
1..	10159	24	73.42	3	1762.08
1..	10160	50	100	5	5182
2..	10333	29	40.25	7	1167.25
2..	10381	48	98	2	4704
2..	10111	33	99.66	6	3288.78
2..	10162	48	91.44	2	4389.12
2..	10111	48	49.06	5	2354.88

Close

Step Metrics

Performance Graph

Metrics

Preview

started transformation 2 - License validated.

started transformation 2 - Dispatching started for transformation

Database.0 - Connected to database [Sample\_Data] (commit=1000)

Postal Codes.0 - Opening file: file:///C:/Pentaho/design-tools/data-integration/samples/transformations/files/Zipssortedbycitystate.csv

Sales Data.0 - Opening file: file:///C:/Pentaho/design-tools/data-integration/samples/transformations/files/sales\_data.csv

Sales Data.0 - Finished processing (I=2824, O=0, R=0, W=2823, U=1, E=0)

Postal Codes.0 - Finished processing (I=21380, O=0, R=0, W=21379, U=1, E=0)

Missing Zips.0 - Finished processing (I=0, O=0, R=2823, W=2823, U=0, E=0)

Write to Database.0 - Finished processing (I=0, O=0, R=2747, W=2747, U=0, E=0)

Lookup Missing Zips.0 - Finished processing (I=0, O=0, R=21455, W=76, U=0, E=0)

2024-11-12 17:23:53.561 - Success - The transformation has finished!!!

Execution Results

Logging

Execution History

Step Metrics

Performance Graph

Metrics

Preview data

2024-11-12 17:23:53.235 - Getting started transformation 2 - License validated.

2024-11-12 17:23:53.235 - Getting started transformation 2 - Dispatching started for transformation [Getting started transformation 2]

2024-11-12 17:23:53.247 - Write to Database.0 - Connected to database [Sample\_Data] (commit=1000)

2024-11-12 17:23:53.250 - Read Postal Codes.0 - Opening file: file:///C:/Pentaho/design-tools/data-integration/samples/transformations/files/Zipssortedbycitystate.csv

2024-11-12 17:23:53.250 - Read Sales Data.0 - Opening file: file:///C:/Pentaho/design-tools/data-integration/samples/transformations/files/sales\_data.csv

2024-11-12 17:23:53.398 - Read Sales Data.0 - Finished processing (I=2824, O=0, R=0, W=2823, U=1, E=0)

2024-11-12 17:23:53.402 - Read Postal Codes.0 - Finished processing (I=21380, O=0, R=0, W=21379, U=1, E=0)

2024-11-12 17:23:53.403 - Filter Missing Zips.0 - Finished processing (I=0, O=0, R=2823, W=2823, U=0, E=0)

2024-11-12 17:23:53.516 - Write to Database.0 - Finished processing (I=0, O=0, R=2747, W=2747, U=0, E=0)

2024-11-12 17:23:53.560 - Lookup Missing Zips.0 - Finished processing (I=0, O=0, R=21455, W=76, U=0, E=0)

2024-11-12 17:23:53.561 - Success - The transformation has finished!!!

Step name

Prepare Filed Layout

Select & Alter

Remove

Meta-data

Fields :

#	Fieldname	Rename to	Length	Precision
10	YEAR_ID			
11	PRODUCTLINE			
12	MSRP			
13	PRODUCTCODE			
14	CUSTOMERNAME			
15	PHONE			
16	ADDRESSLINE1			
17	ADDRESSLINE2			
18	CITY			
19	STATE			
20	POSTALCODE			
21	ZIP_RESOLVED			
22	COUNTRY			
23	TERRITORY			
24	CONTACTLASTNAME			
25	CONTACTFIRSTNAME			

Get fields to select

Edit Mapping

Include unspecified fields, ordered by

Help

OK

Cancel

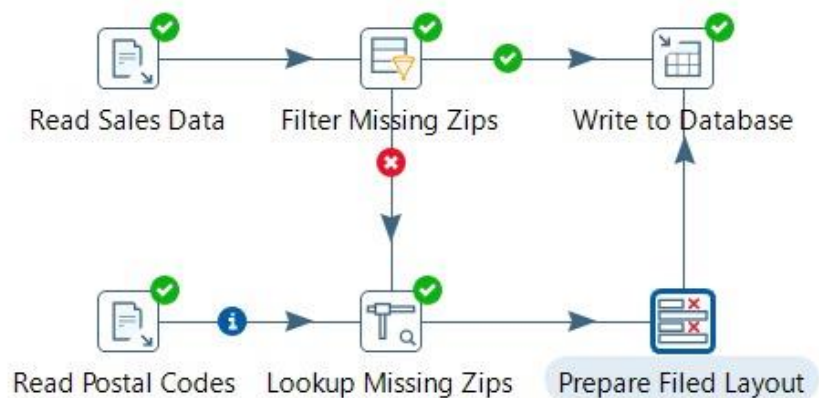
or Missing

up Missi

h Me

sted.

started for transformation [Getting started transformation 2]



Step Metrics Performance Graph Metrics Preview data

ted transformation 2 - License validated.

ted transformation 2 - Dispatching started for transformation [Getting started transformation 2]

## Parte 4 Clean the data

Después de completar el Paso 3: Resolver los datos faltantes, puede limpiar y categorizar aún más los datos en depósitos antes de cargarlos en una base de datos relacional. En esta sección, limpiará los datos del campo PAÍS asignando Estados Unidos a EE. UU. mediante el paso del asignador de valores. La limpieza de los datos garantiza que solo haya una versión de EE. UU.

Además, aprenderá a utilizar depósitos para categorizar los datos de VENTAS en categorías pequeñas, medianas y grandes mediante el paso Rango de números. Aprenderá cómo insertar estas funciones de limpieza y categorización en su transformación justo antes del paso Escribir en la base de datos en el lienzo.

Value mapper

Step name : Value mapper

Fieldname to use : COUNTRY

Target field name (empty=overwrite) :

Default upon non-matching :

Field values:

#	Source value	Target value
1	United States	USA

Help OK Cancel

Number range

Step name: Number range

Input field: SALES

Output field: DEALSIZE

Default value(if no range) unknown

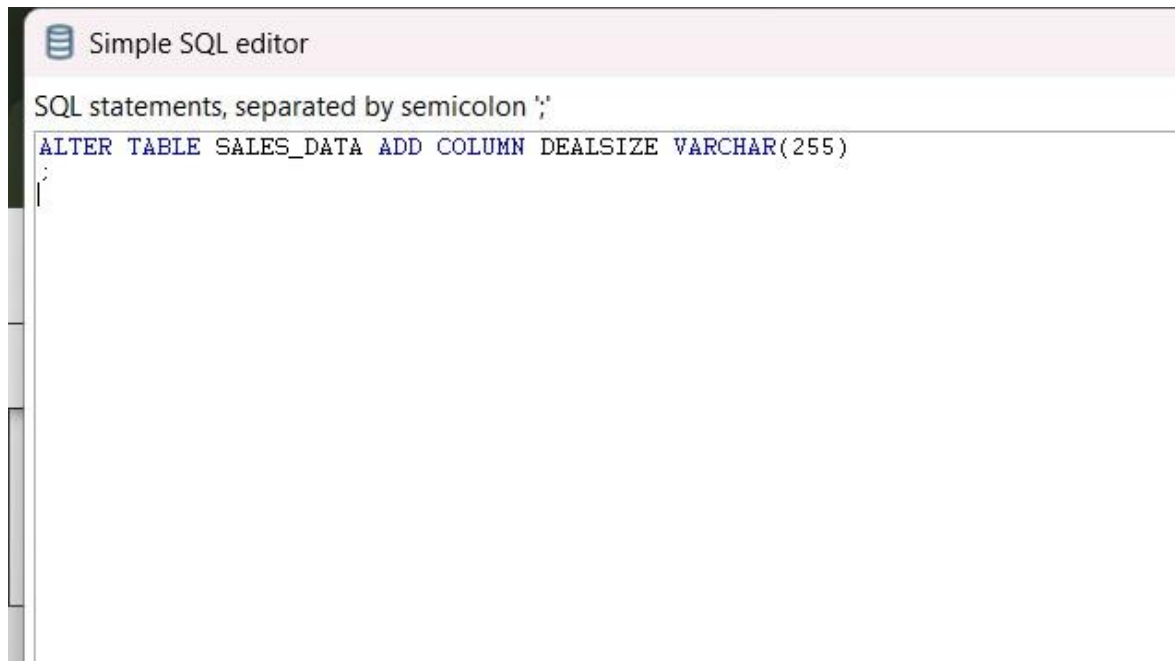
Ranges (min <= x < max):

#	Lower Bound	Upper Bound	Value
1		3000.0	Small
2	3000.0	7000.0	Medium
3	7000.0		Large

Help OK Cancel

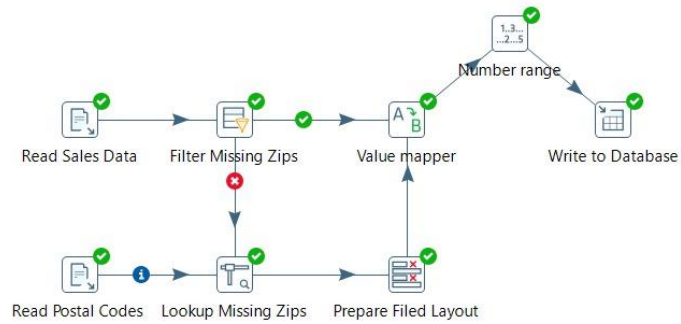
story Step Metrics Performance Graph Metrics Preview data

ing started transformation 2 - License validated.



## Parte 5 Run the transformation

Pentaho Data Integration proporciona varias opciones de implementación. La sección Ejecutar una transformación en el documento Integración de datos de Pentaho explica estas y otras opciones disponibles para su ejecución. En esta sección del tutorial, creará una transformación utilizando la opción Ejecución local.



### Execution Results

Logging	Execution History	Step Metrics	Performance Graph	Metrics	Preview data
<div>2024-11-12 17:41:17.075 - Read Postal Codes.0 - Opening tile: file:///C:/Pentaho/design-tools/data-integration/samples/transformations/tiles/Zipssortedbycitystat</div> <div>2024-11-12 17:41:17.075 - Read Sales Data.0 - Opening file: file:///C:/Pentaho/design-tools/data-integration/samples/transformations/files/sales_data.csv</div> <div>2024-11-12 17:41:17.150 - Read Sales Data.0 - Finished processing (I=2824, O=0, R=0, W=2823, U=1, E=0)</div> <div>2024-11-12 17:41:17.152 - Read Postal Codes.0 - Finished processing (I=21380, O=0, R=0, W=21379, U=1, E=0)</div> <div>2024-11-12 17:41:17.156 - Filter Missing Zips.0 - Finished processing (I=0, O=0, R=2823, W=2823, U=0, E=0)</div> <div>2024-11-12 17:41:17.312 - Lookup Missing Zips.0 - Finished processing (I=0, O=0, R=21455, W=76, U=0, E=0)</div> <div>2024-11-12 17:41:17.320 - Prepare Filed Layout.0 - Finished processing (I=0, O=0, R=76, W=76, U=0, E=0)</div> <div>2024-11-12 17:41:17.327 - Value mapper.0 - Finished processing (I=0, O=0, R=2823, W=2823, U=0, E=0)</div> <div>2024-11-12 17:41:17.331 - Number range.0 - Finished processing (I=0, O=0, R=2823, W=2823, U=0, E=0)</div> <div>2024-11-12 17:41:17.347 - Write to Database.0 - Finished processing (I=0, O=2823, R=2823, W=2823, U=0, E=0)</div> <div>2024-11-12 17:41:17.349 - Spoon - The transformation has finished!!</div>					

### Execution Results

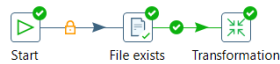
Logging														Execution History	Step Metrics	Performance Graph	Metrics	Preview data
#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output					
1	Read Sales Data	0	0	2823	2824	0	1	0	0	Finished	0.1s	36,205	-					
2	Read Postal Codes	0	0	21379	21380	0	1	0	0	Finished	0.1s	270,633	-					
3	Filter Missing Zips	0	2823	2823	0	0	0	0	0	Finished	0.1s	33,607	-					
4	Lookup Missing Zips	0	21455	76	0	0	0	0	0	Finished	0.2s	89,770	-					
5	Prepare Filed Layout	0	76	76	0	0	0	0	0	Finished	0.2s	308	-					
6	Value mapper	0	2823	2823	0	0	0	0	0	Finished	0.3s	11,114	-					
7	Number range	0	2823	2823	0	0	0	0	0	Finished	0.3s	10,942	-					
8	Write to Database	0	2823	2823	0	2823	0	0	0	Finished	0.3s	10,265	-					







datos. En un ejercicio posterior, programará el trabajo para que se ejecute todos los domingos por la mañana a las 9 a. m.



#### Execution Results

Logging History Job metrics Metrics

2024-11-12 19:13:31.906 - Filter rows.0 - Finished processing (I=0, O=0, R=2823, W=2823, U=0, E=0)  
2024-11-12 19:13:31.912 - Read Postal Codes.0 - Finished processing (I=21380, O=0, R=0, W=21379, U=1, E=0)  
2024-11-12 19:13:32.074 - Lookup Missing Zips.0 - Finished processing (I=0, O=0, R=21455, W=76, U=0, E=0)  
2024-11-12 19:13:32.082 - Prepare Field Layout.0 - Finished processing (I=0, O=0, R=76, W=76, U=0, E=0)  
2024-11-12 19:13:32.088 - Value Mapper.0 - Finished processing (I=0, O=0, R=2823, W=2823, U=0, E=0)  
2024-11-12 19:13:32.093 - Number range.0 - Finished processing (I=0, O=0, R=2823, W=2823, U=0, E=0)  
2024-11-12 19:13:33.116 - Write to Database.0 - Finished processing (I=0, O=0, R=2823, W=2823, U=0, E=0)  
2024-11-12 19:13:33.124 - Sample Job - Finished job entry [Transformation] (result=[true])  
2024-11-12 19:13:33.125 - Sample Job - Finished job entry [File exists] (result=[true])  
2024-11-12 19:13:33.125 - Sample Job - Job execution finished  
2024-11-12 19:13:33.127 - Spoon - Job has ended.

## Conclusión

La práctica realizada permitió afianzar nuestras habilidades en el proceso ETL, destacando la importancia de la calidad y completitud de los datos para evitar errores en análisis posteriores. Al resolver valores nulos y estandarizar campos, se logró un dataset coherente y listo para ser almacenado o utilizado en análisis futuros. Este ejercicio también reforzó el uso de herramientas como Pentaho para la automatización de tareas en proyectos de ciencia de datos, contribuyendo a la eficiencia y organización en el manejo de grandes volúmenes de información.