

AIM: Extract and load data using Pentaho Data integration tool

THEORY:

Kettle (K.E.T.T.L.E - Kettle ETL Environment) has been recently acquired by the Pentaho group and renamed to Pentaho Data Integration. Kettle is a leading open source ETL application on the market. It is classified as an ETL tool, however the concept of classic ETL process (extract, transform, load) has been slightly modified in Kettle as it is composed of four elements, ETTL, which stands for:

- ✓ Data extraction from source databases
- ✓ Transport of the data
- ✓ Data transformation
- ✓ Loading of data into a data warehouse Kettle is a set of tools and applications which allows data manipulations across multiple sources.

The main components of Pentaho Data Integration are:

Spoon – It is a graphical tool which make the design of an ETTL process transformations easy to create. It performs the typical data flow functions like reading, validating, refining, transforming, writing data to a variety of different data sources and destinations. Transformations designed in Spoon can be run with Kettle Pan and Kitchen.

Pan – It is an application dedicated to run data transformations designed in Spoon.

Chef – It is a tool to create jobs which automate the database update process in a complex way.

Kitchen – It is an application which helps execute the jobs in a batch mode, usually using a schedule which makes it easy to start and control the ETL processing.

Carte – It is a web server which allows remote monitoring of the running Pentaho Data Integration ETL processes through a web browser.

ETL Process Transformation:

As the name suggested ETL stands for Extract Transform and Load. Just like the name applies ETL tool Extracts data from the source. Transforms the data while in transit and then it loads the data in to Specified database.

➤ **Sort Transformation:**

The Sort transformation sorts input data in ascending or descending order and copies the sorted data to the transformation output. You can apply multiple sorts to an input; each sort is identified by a numeral that determines the sort order. The column with the lowest number is sorted first, the sort column with the second lowest number is sorted next, and so on.

➤ **Add Sequence:**

The Add sequence step adds a sequence to the stream. A sequence is an everchanging integer value with a specific start and increment value. You can either use a database sequence to determine the value of the sequence, or have it generated by Kettle.

➤ **Concat:**

The Concat Fields step is used to concatenate multiple fields into one target field. The fields can be separated by a separator and the enclosure logic is completely compatible with the Text File Output step.

➤ **Unique Rows:**

The Unique rows step removes duplicate rows from the input stream(s).

➤ **Split:**

The Split Fields step allows you to split fields based on delimiter information

Features of Pentaho:

Pentaho Reporting primarily includes a Reporting Engine, a Report Designer, a Business Intelligence (BI) Server. It comes loaded with the following features – • Report Designer – Used for creating pixel perfect report.

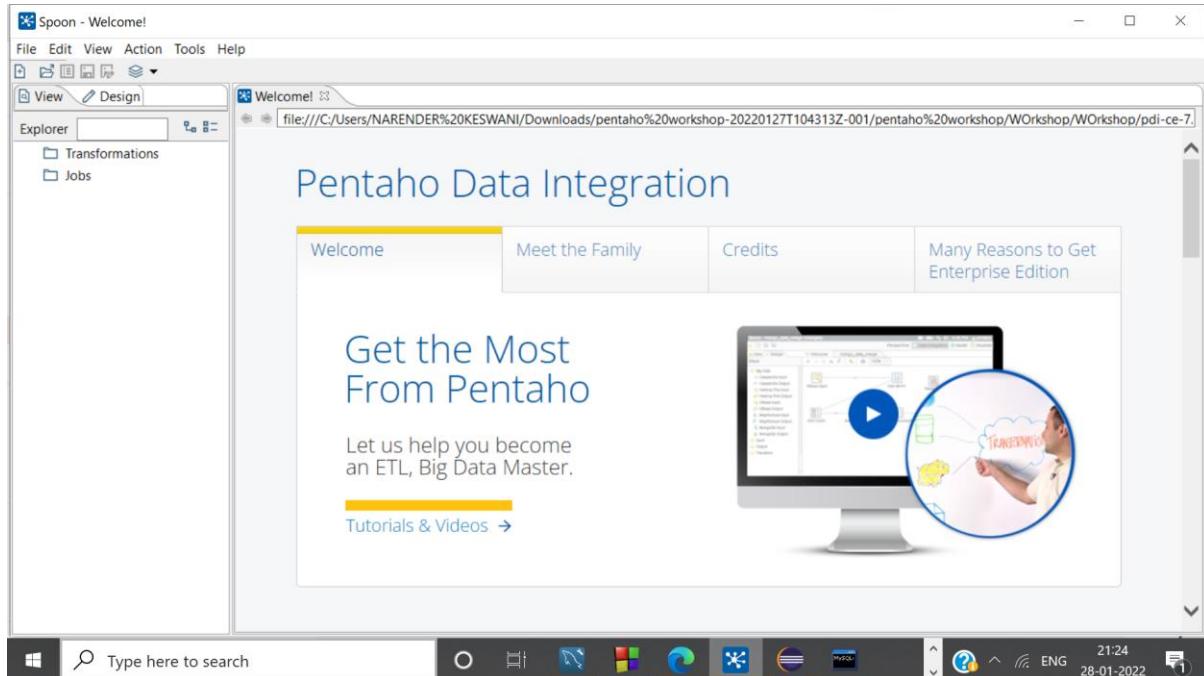
- Metadata Editor – Allows to add user-friendly metadata domain to a data source.
- Report Designer and Design Studio – Used for fine-tuning of reports and ad-hoc reporting.
- Pentaho user console web interface – Used for easily managing reports and analyzing views.
- Ad-Hoc reporting interface – Offers a step-by-step wizard for designing simple reports. Output formats include PDF, RTF, HTML, and XLS.
- A complex scheduling sub-system – Allows users to execute reports at given intervals.
- Mailing – Users can email a published report to other users.

A) Demo of reading and writing csv file.

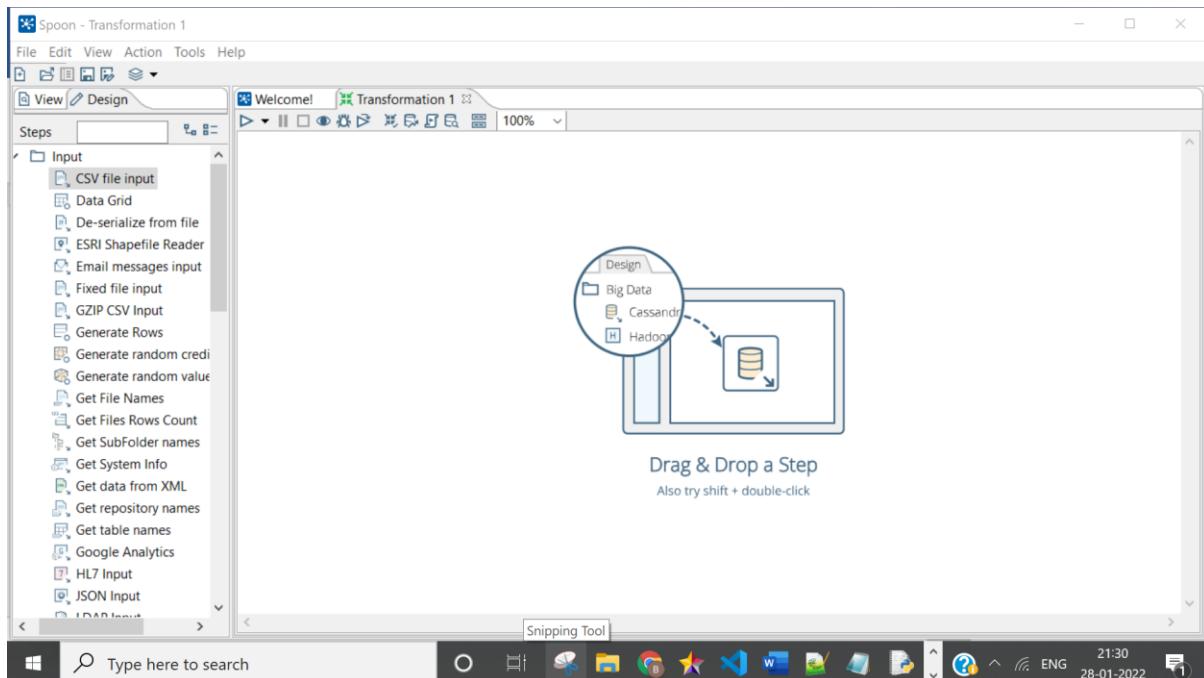
Step 1. Open Data Integration folder

Step 2. Double click on spoon(Windows Batch file)

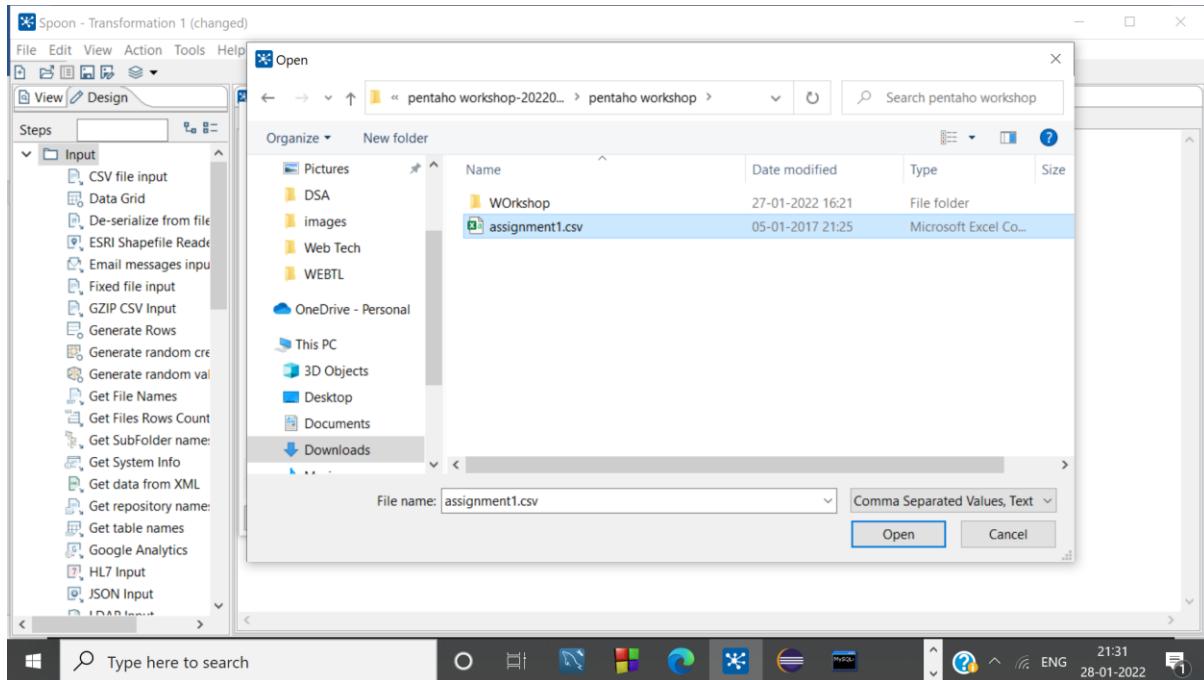
Step 3. A welcome! window appears with some useful links for you to see.



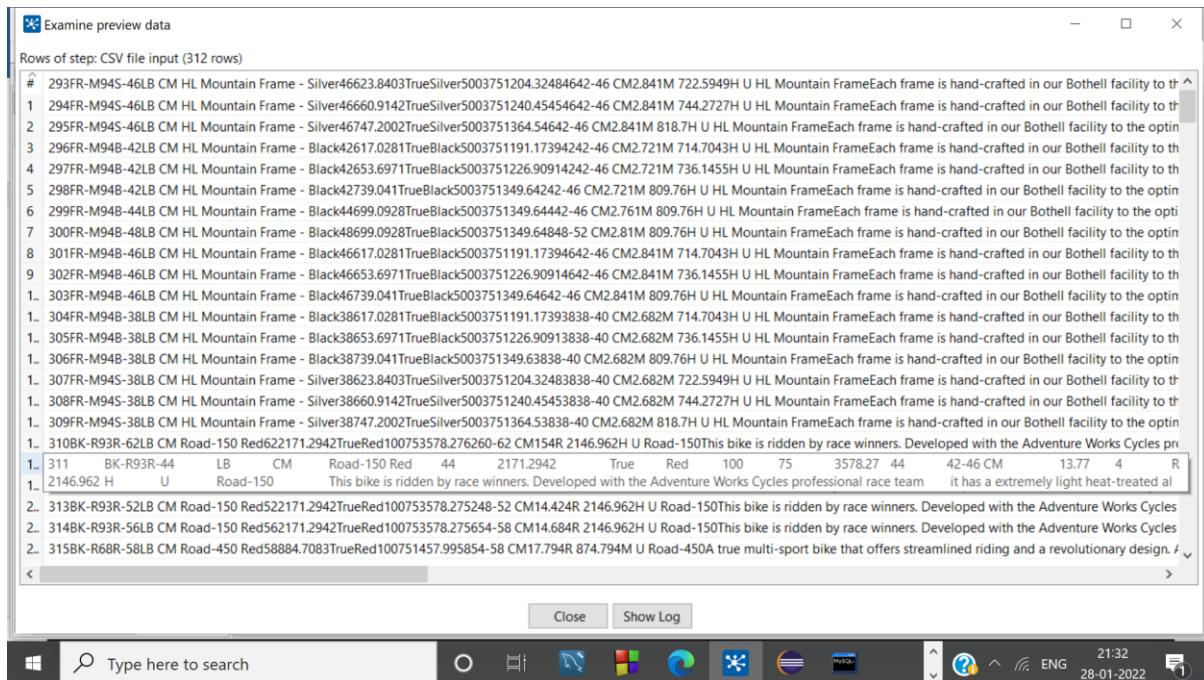
Step 4: Create new transformation, in input drag & drop the csv file.



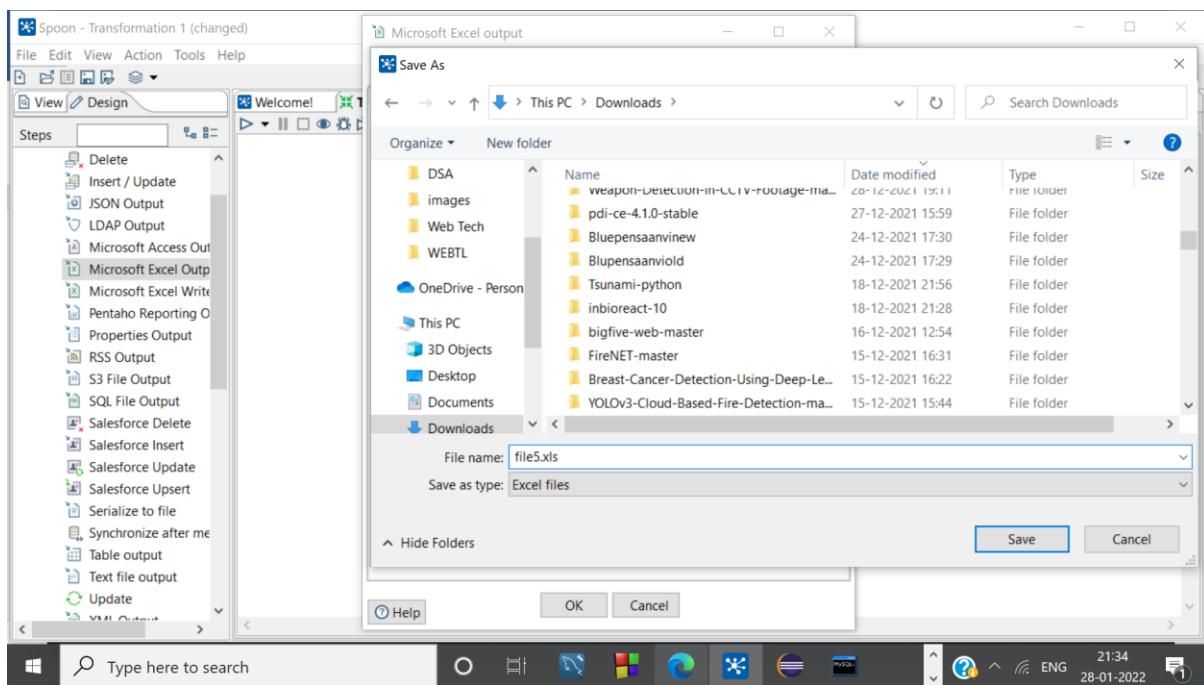
Step 5: Specify the path of input file



Step 6: Check Preview Data

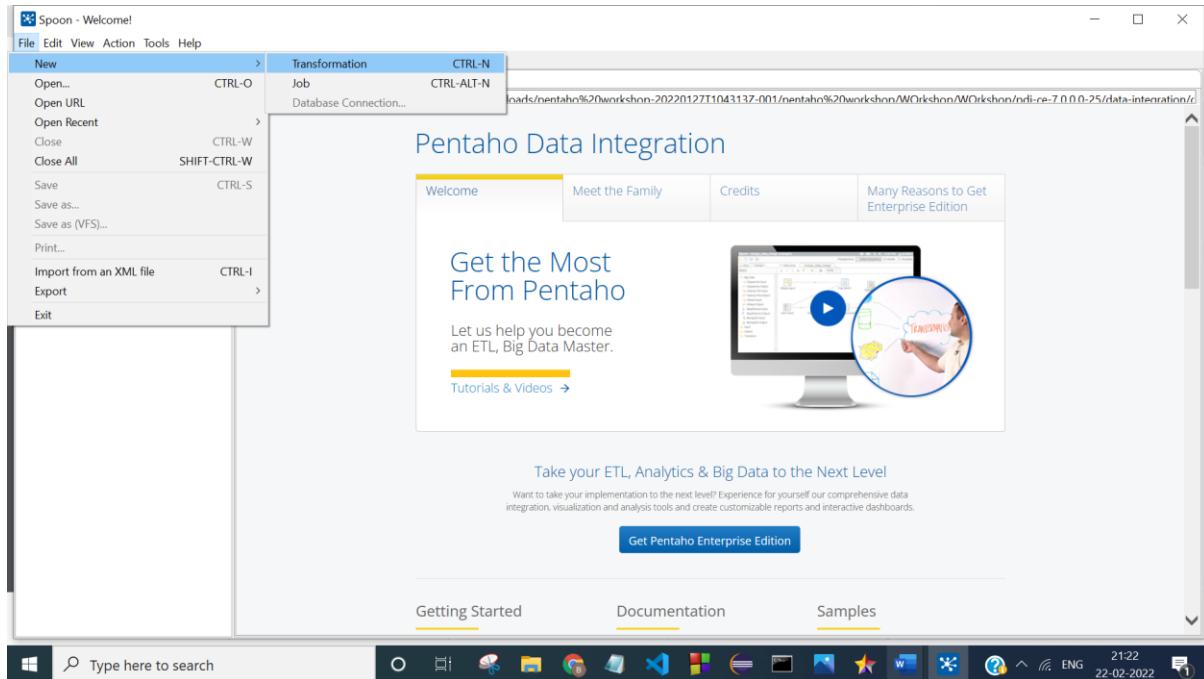


Step 7: Now, drag & drop csv output file from output section

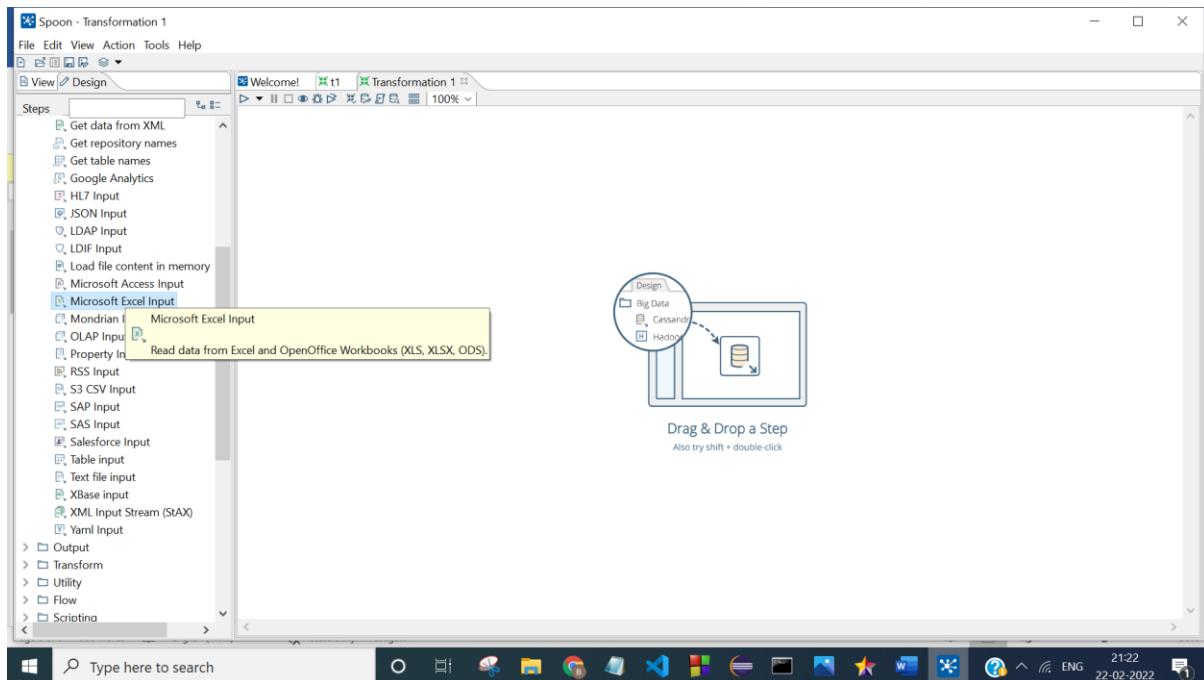


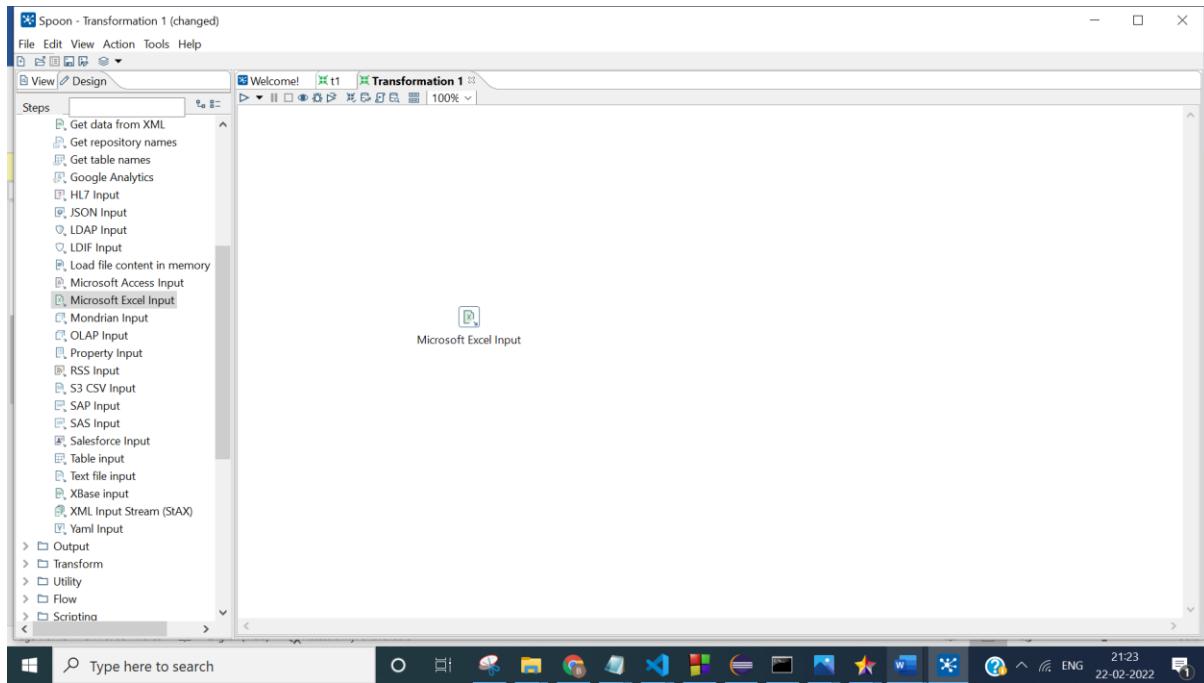
B) SORT ROWS TRANSFORMATIOON:

Step 1. In the Pentaho Data Integration tools (Spoon), Pull down the File menu and select the New menu item followed by Transformation.

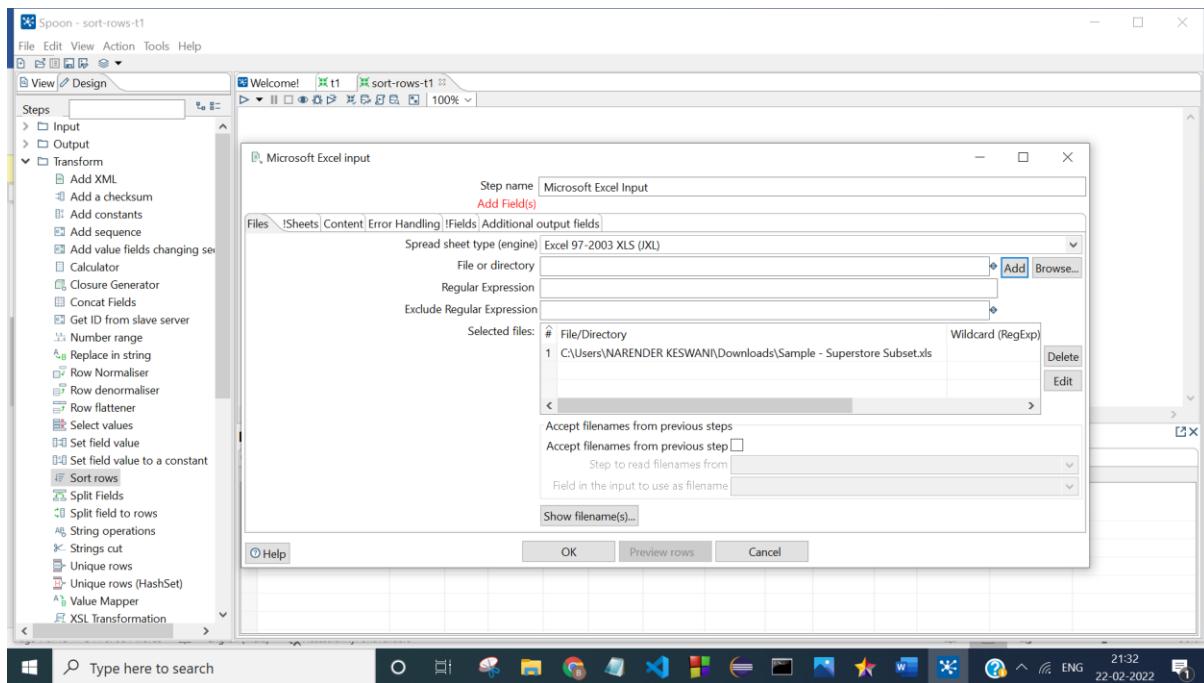


Step 2. Open up the Input folder and drag and drop the Microsoft excel input step on to the transformation window.

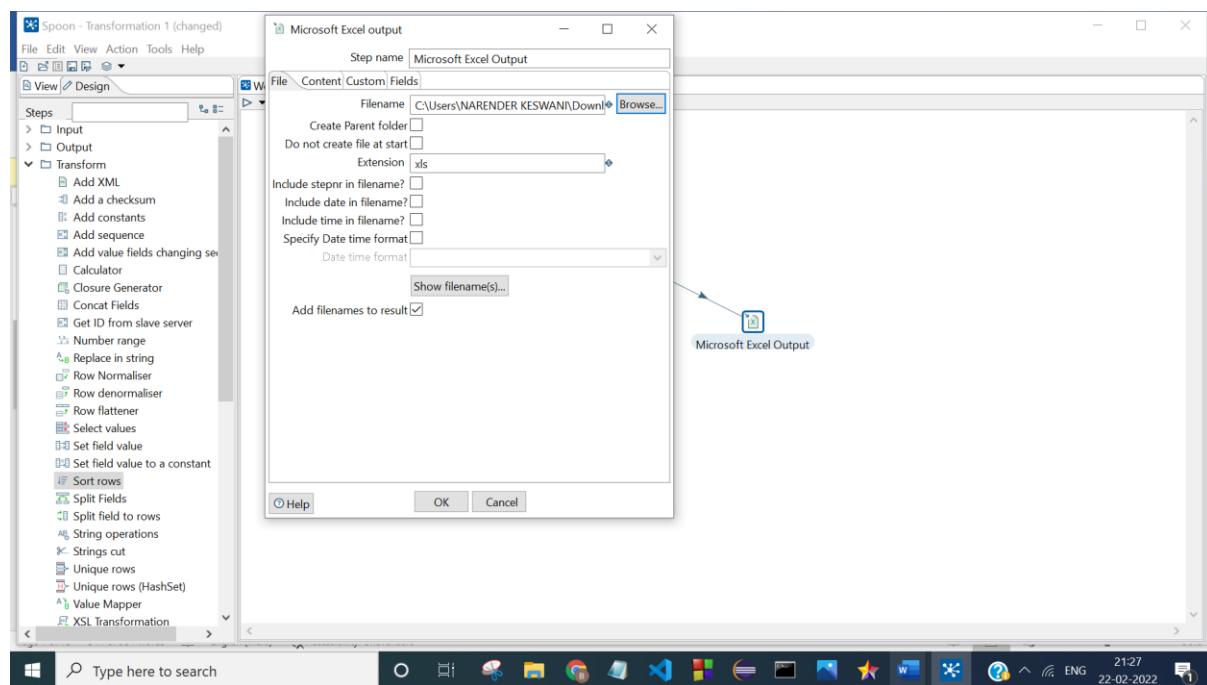
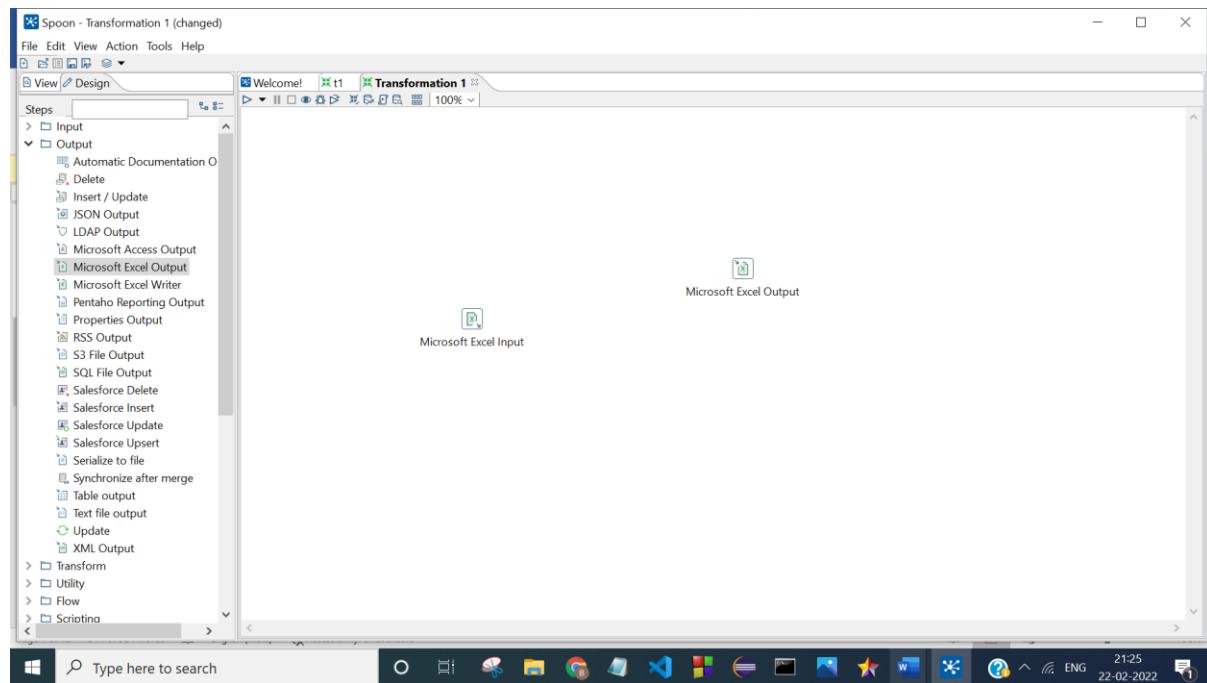




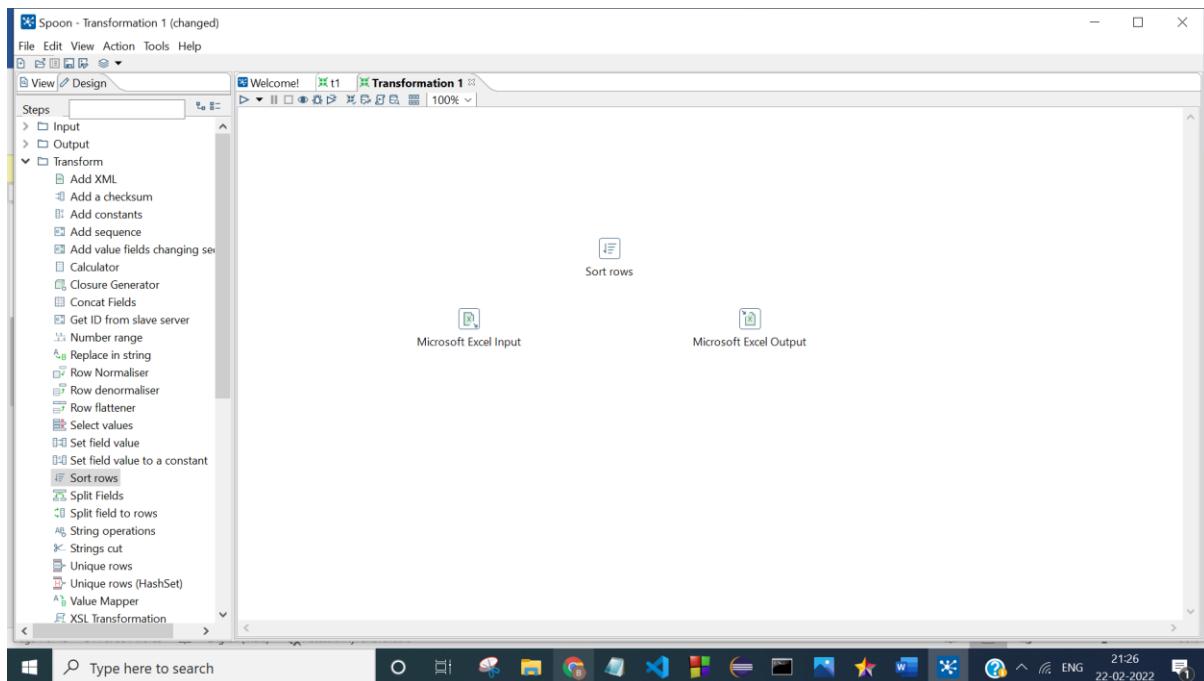
Step 3. Double-click on the Microsoft Excel Input step to view its properties Click on the Browse button next to Filename field and navigate to the folder with the Excel files. Select the Excel file as shown below and then click the Open button to add excel file .



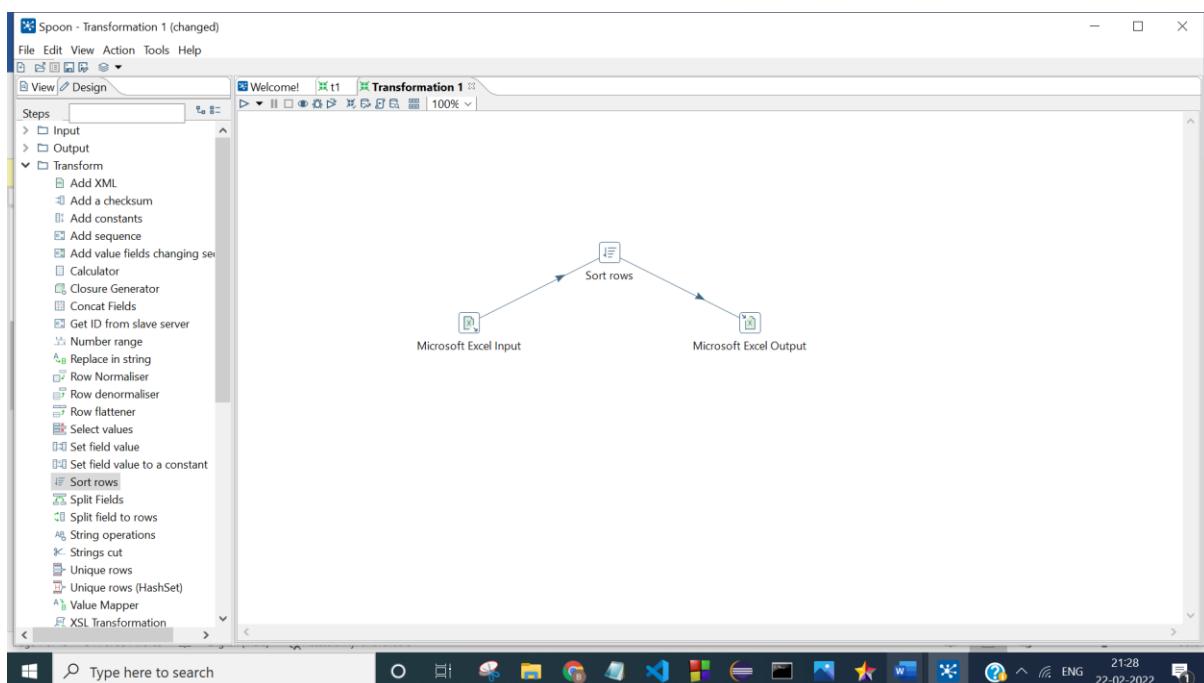
Step 4: Double-click on the Microsoft Excel output step to view its properties Click on the Browse button next to Filename field and navigate to the folder to store the output.



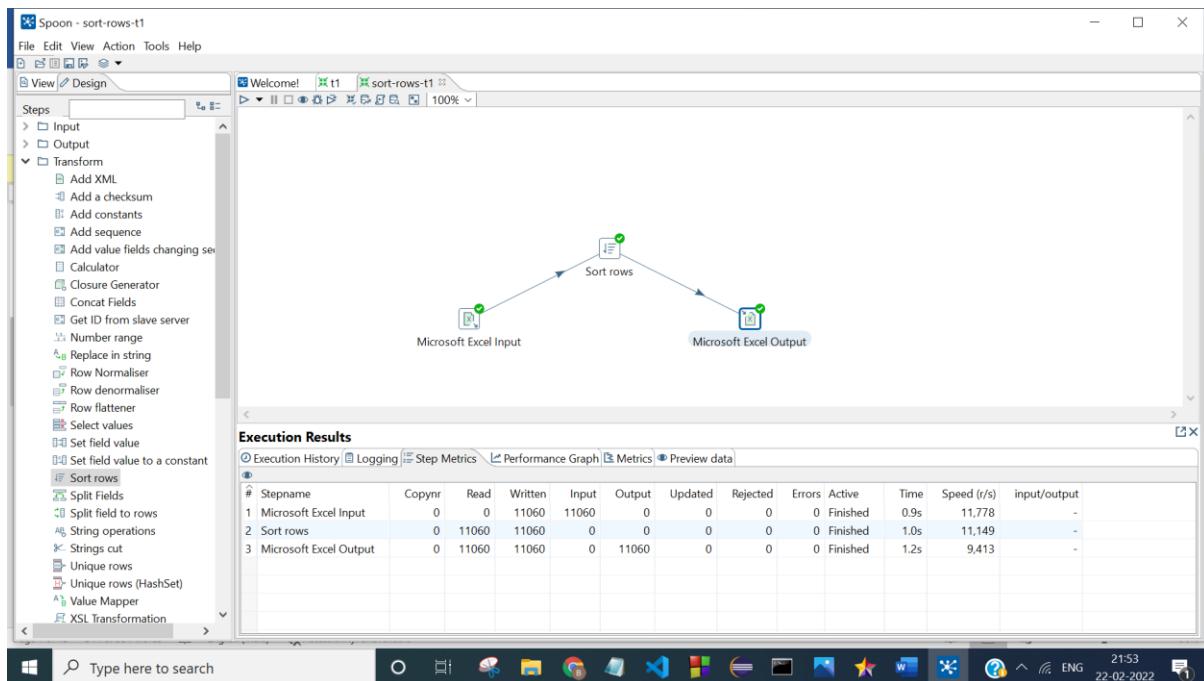
Step 5: Drag & Drop sort rows from the transformation sections.



Step 6: Create connections among input, sort & output using hops.



Step 7: Execute the program:



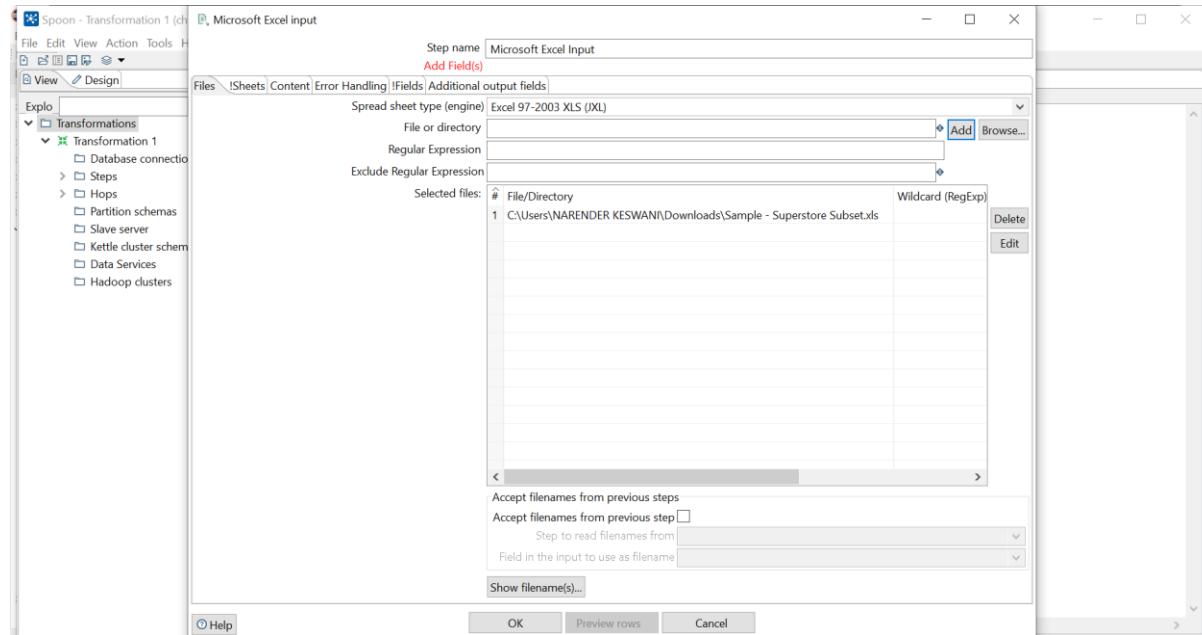
OUTPUT:

#	Row ID	Order Priority	Discount	Unit Price	Shipping Cost	Customer ID	Customer Name	Ship Mode	Customer Segment
1	2.00	Not Specified	.01	2.08	2.56	2867.0	Dana Teague	Regular Air	Corporate
2	27.00	Critical	.06	12.44	6.27	1821.0	Vanessa Boyer	Regular Air	Consumer
3	52.00	Critical	.08	155.99	8.08	1402.0	Wesley Tate	Regular Air	Corporate
4	53.00	Critical	.10	6.48	10.05	1402.0	Wesley Tate	Regular Air	Corporate
5	62.00	High	.02	48.58	54.11	2747.0	Brian Grady	Delivery Truck	Corporate
6	63.00	High	.07	39.48	1.99	2747.0	Brian Grady	Regular Air	Corporate
7	64.00	Medium	.08	124.49	51.94	553.00	Kristine Connolly	Delivery Truck	Corporate
8	65.00	Returned	<null>	<null>	<null>	<null>	<null>	<null>	<null>
9	66.00	High	.02	3.69	.50	3289.0	Emily Britt	Regular Air	Corporate
10	67.00	High	.09	3.85	.70	3289.0	Emily Britt	Regular Air	Corporate
11	68.00	Not Specified	.06	11.70	6.96	1630.0	Jimmy Han	Regular Air	Home Office
12	78.00	Low	.09	6.08	1.82	898.00	Harriet Hodges	Regular Air	Small Business
13	87.00	Critical	.04	3.08	.99	3106.0	Alexander O'Brien	Regular Air	Home Office
14	88.00	Critical	.04	125.99	4.20	3106.0	Alexander O'Brien	Regular Air	Home Office
15	95.00	Medium	.10	5.28	6.26	3011.0	Tammy Raynor	Regular Air	Corporate
16	96.00	Medium	.01	65.99	2.50	3011.0	Tammy Raynor	Regular Air	Corporate
17	97.00	High	.03	7.30	7.72	563.00	Marjorie P. Goodwin	Regular Air	Home Office
18	98.00	High	.09	42.76	6.22	1106.0	Maxine Collier Grady	Regular Air	Small Business
19	106.00	High	.01	9.31	3.98	1,106.00	Maxine Collier Grady	Regular Air	Small Business
20	112.00	Not Specified	.04	80.98	4.50	607.00	Clara Hauser	Regular Air	Corporate
21	113.00	Not Specified	.08	6.48	5.14	607.00	Clara Hauser	Regular Air	Corporate
22	119.00	Medium	.06	65.99	8.99	1488.00	Anthony Goodwin	Regular Air	Home Office
23	132.00	Not Specified	.02	6.48	5.14	1106.00	Maxine Collier Grady	Regular Air	Small Business

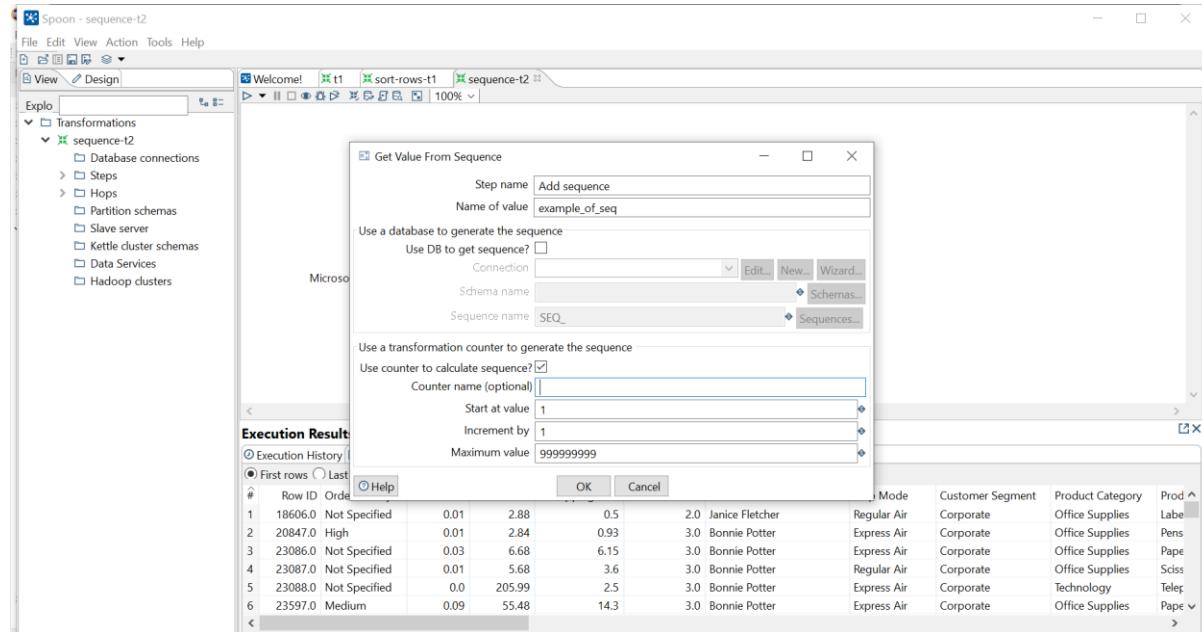
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	Row ID	Order Prio	Discount	Unit Price	Shipping C	Customer	Customer	Ship Mode	Customer Segment									
2	2.00	Not Specif	.01	2.08	2.56	2,867.00	Dana Teag	Regular Ai	Corporate									
3	27.00	Critical	.06	12.44	6.27	1,821.00	Vanessa Boyer	Regular Ai	Consumer									
4	52.00	Critical	.08	155.99	8.08	1,402.00	Wesley Ta	Regular Ai	Corporate									
5	53.00	Critical	.10	6.48	10.05	1,402.00	Wesley Ta	Regular Ai	Corporate									
6	62.00	High	.02	48.58	54.11	2,747.00	Brian Grac	Delivery Ti	Corporate									
7	63.00	High	.07	39.48	1.99	2,747.00	Brian Grac	Regular Ai	Corporate									
8	64.00	Medium	.08	124.49	51.94	553.00	Kristine Connolly	Delivery Ti	Corporate									
9	65.00	Returned																
10	66.00	High	.02	3.69	.50	3,289.00	Emily Britt	Regular Ai	Corporate									
11	67.00	High	.09	3.85	.70	3,289.00	Emily Britt	Regular Ai	Corporate									
12	68.00	Not Specif	.06	11.70	6.96	1,630.00	Jimmy Han	Regular Ai	Home Office									
13	78.00	Low	.09	6.08	1.82	898.00	Harriet Hodges	Regular Ai	Small Business									
14	87.00	Critical	.04	3.08	.99	3,106.00	Alexander O'Brien	Regular Ai	Home Office									
15	88.00	Critical	.02	6.48	5.90	3,106.00	Alexander O'Brien	Regular Ai	Home Office									
16	89.00	Critical	.04	125.99	4.20	3,106.00	Alexander O'Brien	Regular Ai	Home Office									
17	95.00	Medium	.10	5.28	6.26	3,011.00	Tammy Raynor	Regular Ai	Corporate									
18	96.00	Medium	.01	65.99	2.50	3,011.00	Tammy Raynor	Regular Ai	Corporate									
19	97.00	High	.03	7.30	7.72	563.00	Marjorie P. Goodwin	Regular Ai	Corporate									
20	98.00	High	.09	42.76	6.22	563.00	Marjorie P. Goodwin	Regular Ai	Corporate									
21	106.00	High	.01	9.31	3.98	1,106.00	Maxine Collier Grady	Regular Ai	Small Business									
22	112.00	Not Specif	.04	80.98	4.50	607.00	Clara Hauser	Regular Ai	Corporate									

C) SEQUENCE TRANSFORMATION:

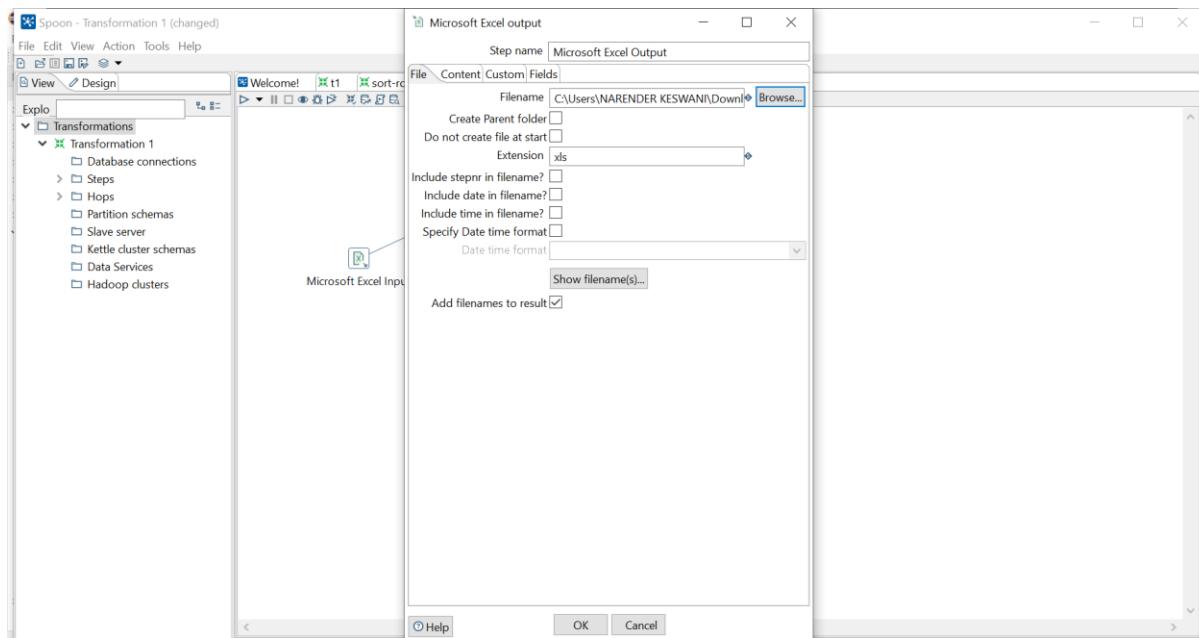
Step 1. Open up the Input folder and drag and drop the Microsoft excel input step on to the transformation window.



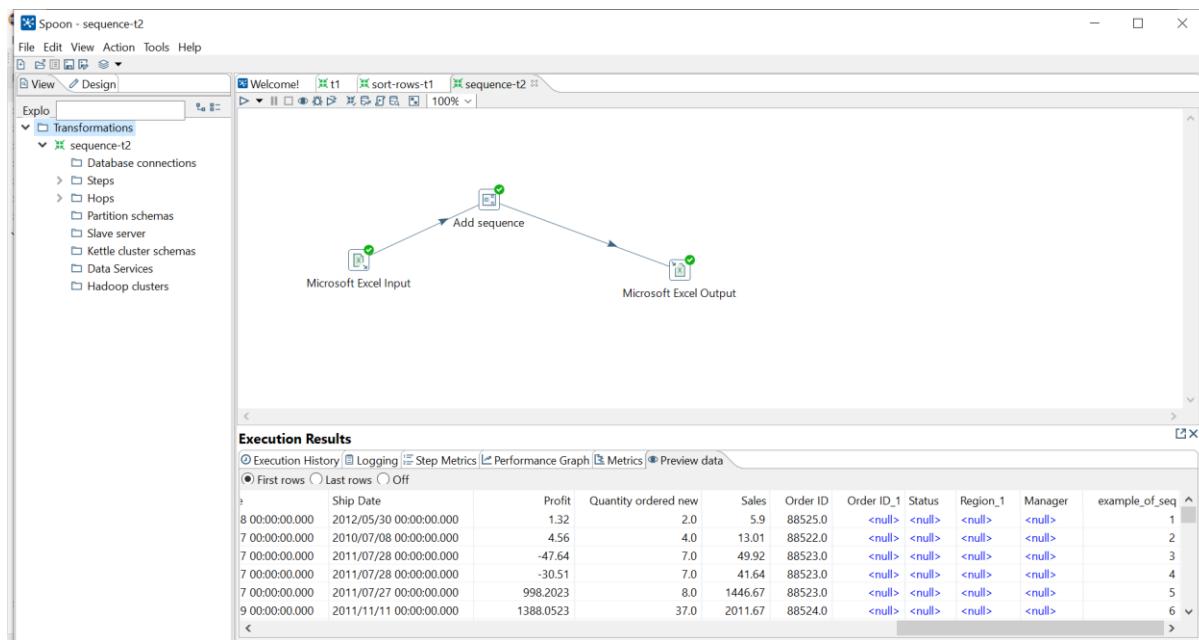
Step 2. Open up the Transformation folder and drag and drop the Add Sequence and config according to requirements



Step 3. Open up the Output folder and drag and drop the Microsoft excel output and config it.



Step 4. Execute the transformation.



OUTPUT:

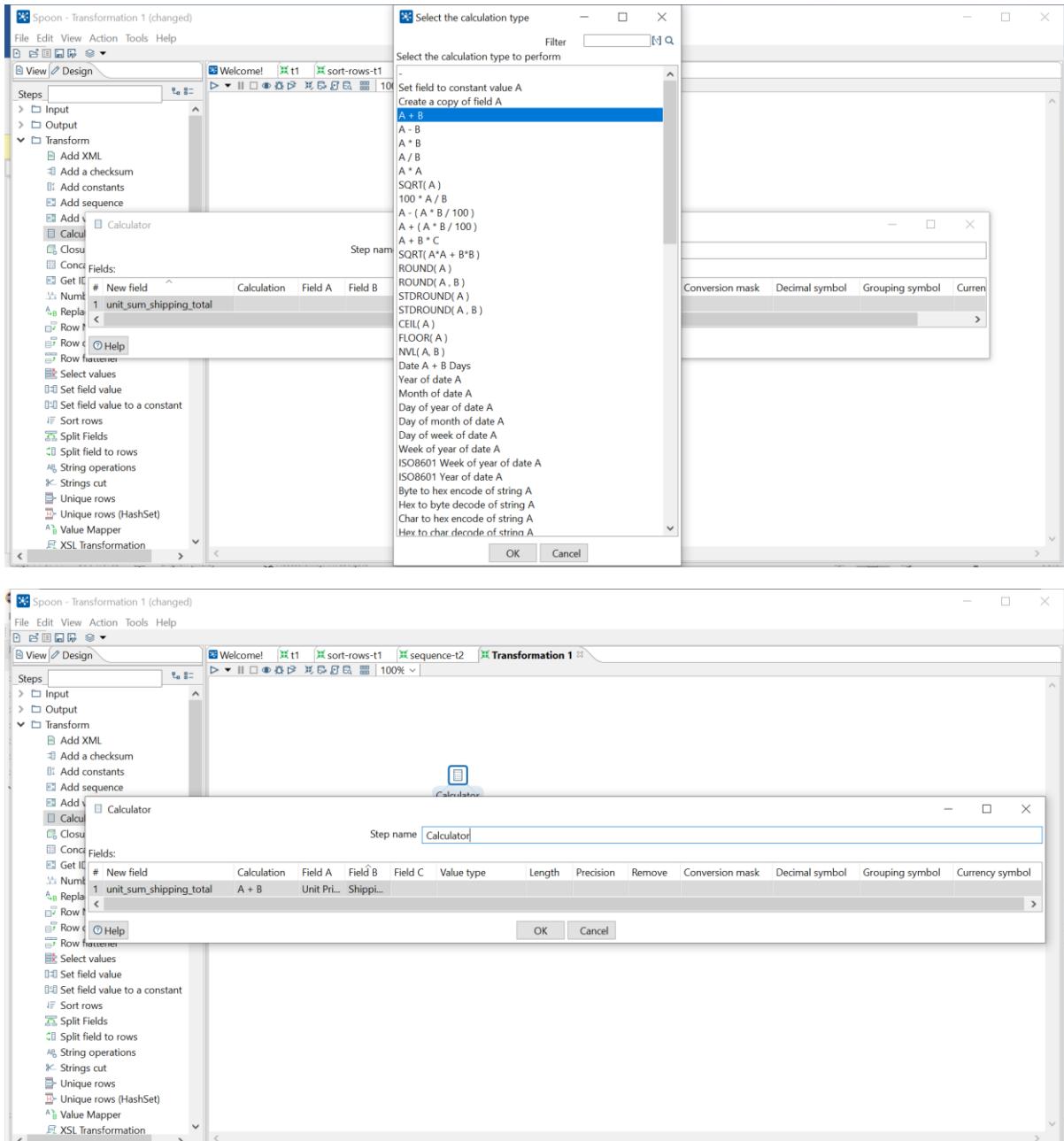
The screenshot shows a Microsoft Excel spreadsheet titled "sequence-output.xls". The data is organized into columns labeled Q through AM. Column Q contains city names, and column AC contains a sequence of numbers from 1 to 25. The data includes columns for Postal Code, Order Date, Ship Date, Profit, Quantity or Sales, Order ID, Order ID_1, Status, Region_1, Manager, and example_of_seq. A yellow banner at the top of the sheet reads: "GET GENUINE OFFICE Your license isn't genuine, and you may be a victim of software counterfeiting. Avoid interruption and keep your files safe with genuine Office today." with buttons for "Get genuine Office" and "Learn more".

Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM
1	City	Postal Cod	Order Date	Ship Date	Profit	Quantity or Sales	Order ID	Order ID_1	Status	Region_1	Manager	example_of_seq										
2	Addison	60101.			1.32	2.	5.9	88525				1										
3	Anacortes	98221			4.56	4.	13.01	88522				2										
4	Anacortes	98221			-47.64	7.	49.92	88523				3										
5	Anacortes	98221			-30.51	7.	41.64	88523				4										
6	Anacortes	98221			996.202	8.	1446.67	88523				5										
7	Anacortes	98221			1388.052	37.	2011.67	88524				6										
8	Anacortes	98221			1001.445	12.	1451.37	88526				7										
9	San Gabri	91776			4390.367	12.	6302.85	90193				8										
10	San Gabri	91776			-141.26	18.	113.25	90197				9										
11	San Jose	95123			1045.467	16.	1515.17	90194				10										
12	San Jose	95123			-13.86	4.	28.61	90200				11										
13	San Jose	95123			57.581	17.	83.45	90200				12										
14	San Jose	95123			1176.505	24.	1705.99	90200				13										
15	All River	27274			72.93	19.	209.99	90198				14										
16	Bedford	3110			-158.74	5.	705.47	90199				15										
17	Camden	8101			-346.615	8.	1794.27	90199				16										
18	Pennsauke	8109			142.796	14.	206.95	90195				17										
19	Roselle	7203			-53.81	22.	211.15	90192				18										
20	Cranston	2907			23.12	8.	90.39	90196				19										
21	Prior Lake	55372			803.471	16.	1164.45	86838				20										
22	Prior Lake	55372			-24.03	7.	22.23	86838				21										
23	Prior Lake	55372			-37.03	4.	13.99	86838				22										
24	Prior Lake	55372			-71	4.	14.26	86838				23										
25	Prior Lake	55372			-44.54	4.	29.55	86845				24										
26	Smithtown	11787			-69.82	7.	33.47	86837				25										

D) CALCULATOR (SUM):

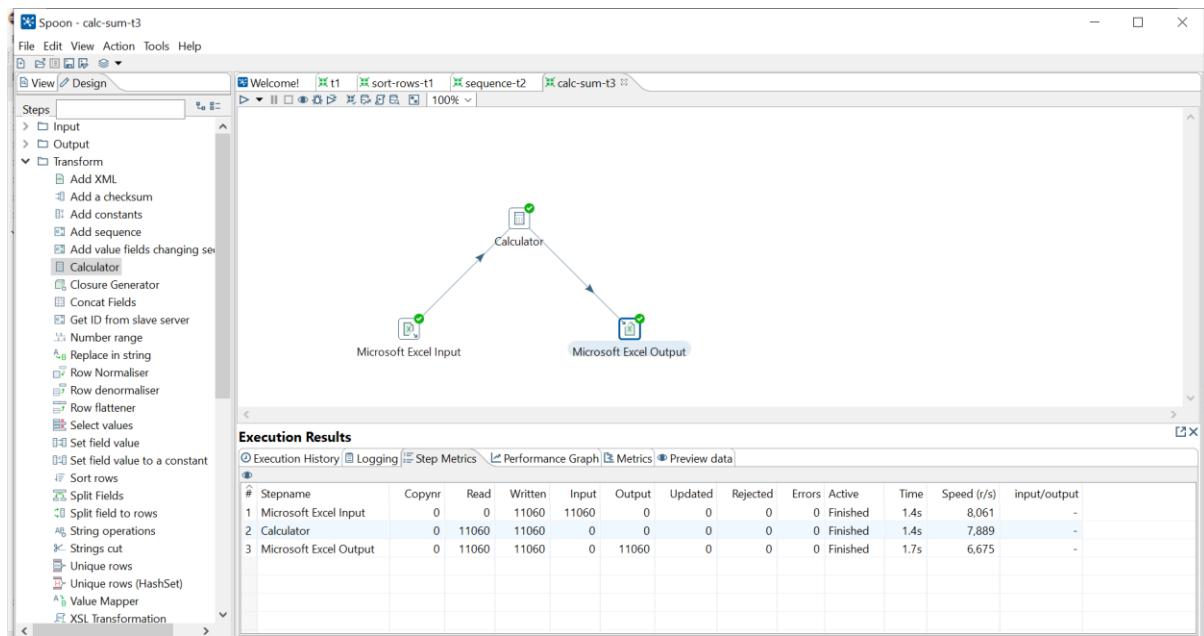
Step 1. Open up the Input folder and drag and drop the Microsoft excel input step on to the transformation window.

Step 2. Open up the Transformation folder and drag and drop the calculator and config it according to requirements.



Step 3. Open up the output folder and drag and drop the Microsoft excel output step on to the transformation window.

Step 4. Execute the transformation



OUTPUT:

#	Row ID	Order Priority	Discount	Unit Price	Shipping Cost	Customer ID	Customer Name	unit_sum_shipping_total
1	18606.0	Not Specified	.01	2.88	.50	2.0	Janice Fletcher	3.38
2	20847.0	High	.01	2.84	.93	3.0	Bonnie Potter	3.77
3	23086.0	Not Specified	.03	6.68	6.15	3.0	Bonnie Potter	12.83
4	23087.0	Not Specified	.01	5.68	3.6	3.0	Bonnie Potter	9.28
5	23088.0	Not Specified	.00	205.99	2.5	3.0	Bonnie Potter	208.49
6	23597.0	Medium	.09	55.48	14.3	3.0	Bonnie Potter	69.78
7	25549.0	Low	.08	120.97	26.3	3.0	Bonnie Potter	147.27
8	20228.0	Not Specified	.02	500.98	26.0	5.0	Ronnie Proctor	526.98
9	19483.0	Low	.08	6.48	6.81	5.0	Ronnie Proctor	13.29
10	24782.0	High	.01	90.24	.99	6.0	Dwight Hwang	91.23
11	24563.0	Critical	.07	6.48	6.6	6.0	Dwight Hwang	13.08
12	24564.0	Critical	.01	4.84	.71	6.0	Dwight Hwang	5.55
13	24565.0	Critical	.10	85.99	.99	6.0	Dwight Hwang	86.98
14	21866.0	High	.05	12.28	4.86	7.0	Leon Gill	17.14
15	20876.0	Medium	.08	140.98	36.09	8.0	Melanie Garner	177.07
16	20877.0	Medium	.10	286.85	61.76	9.0	Lorraine Houston	348.61
17	22241.0	Critical	.06	15.57	1.39	10.0	Meredith Norrie Thomas	16.96
18	21776.0	Critical	.06	9.48	7.29	11.0	Marcus Dot	16.77
19	21776.0	Critical	.06	9.48	7.29	12.0	Kara Pace	14.35
20	23328.0	High	.04	10.98	3.37	14.0	Gwendolyn F Tyson	98.68
21	24844.0	Medium	.09	78.69	19.99	14.0	Gwendolyn F Tyson	98.68
22	24846.0	Medium	.08	32.8	2.31	14.0	Gwendolyn F Tyson	5.59
23	24847.0	Medium	.05	32.8	4.2	14.0	Gwendolyn F Tyson	7.48
24	24848.0	Medium	.05	3.58	1.63	14.0	Gwendolyn F Tyson	5.21
25	24845.0	Medium	.01	6.48	7.86	14.0	Gwendolyn F Tyson	14.34
26	21866.0	High	.00	12.28	4.86	15.0	Tonya D.	16.50

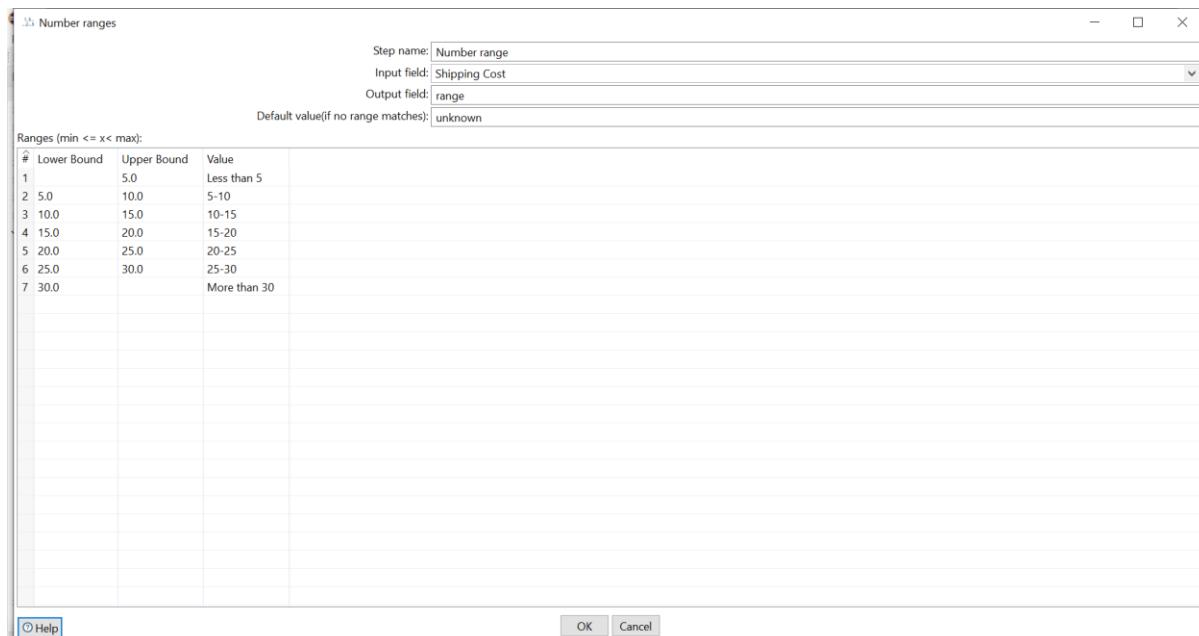
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	F
1	Row ID	Order Prio	Discount	Unit Price	Shipping CCustomer	Customer ID	Customer Name	unit_sum_shipping_total										
2	18,606.0	Not Specif	.01	2.88	.50	2.0	Janice Fiel	3.38										
3	20,847.0	High	.01	2.84	.93	3.0	Bonnie Pol	3.77										
4	23,086.0	Not Specif	.03	6.68	6.15	3.0	Bonnie Pol	12.83										
5	23,087.0	Not Specif	.01	5.68	3.60	3.0	Bonnie Pol	9.28										
6	23,088.0	Not Specif	.00	205.99	2.50	3.0	Bonnie Pol	208.49										
7	23,597.0	Medium	.09	55.48	14.30	3.0	Bonnie Pol	69.78										
8	25,549.0	Low	.08	120.97	26.30	3.0	Bonnie Pol	147.27										
9	20,228.0	Not Specif	.02	500.98	26.00	5.0	Ronnie Prc	526.98										
10	19,483.0	Low	.08	6.48	6.81	5.0	Ronnie Prc	13.29										
11	24,782.0	High	.01	90.24	.99	6.0	Dwight Hw	91.23										
12	24,563.0	Critical	.07	6.48	6.60	6.0	Dwight Hw	13.08										
13	24,564.0	Critical	.01	4.84	.71	6.0	Dwight Hw	5.55										
14	24,565.0	Critical	.10	85.99	.99	6.0	Dwight Hw	86.98										
15	21,866.0	High	.05	12.28	4.86	7.0	Leon Gill	17.14										
16	20,876.0	Medium	.08	140.98	36.09	8.0	Melanie G.	177.07										
17	20,877.0	Medium	.10	286.85	61.76	9.0	Lorraine H	348.61										
18	22,241.0	Critical	.06	15.57	1.39	10.0	Meredith N	16.96										
19	21,776.0	Critical	.06	9.48	7.29	11.0	Marcus Dt	16.77										
20	23,328.0	High	.04	10.98	3.37	12.0	Kara Pace	14.35										
21	24,844.0	Medium	.09	78.69	19.99	14.0	Gwendolyn F Tyson	98.68										
22	24,846.0	Medium	.08	32.8	2.34	14.0	Gwendolyn F Tyson	5.59										

E) Number Range:

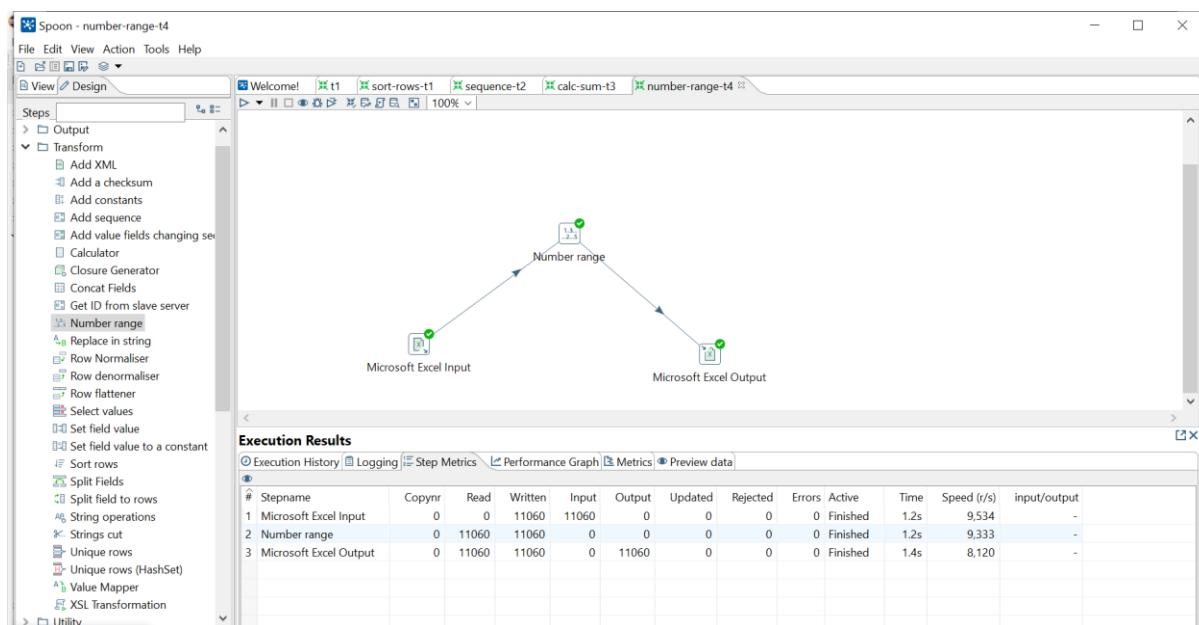
Step 1. Open up the Input folder and drag and drop the Microsoft excel input step on to the transformation window.

Step 2. Open up the Transformation folder and drag and drop the number range and config it according to requirements.

Step 3. Open up the output folder and drag and drop the Microsoft excel output.



Step 4. Execute the transformation



OUTPUT:

The screenshot shows the Apache Nifi interface for a flow named 'number-range-t4'. The 'Execution Results' tab is selected, displaying a table of data. The columns include Row ID, Order Priority, Discount, Unit Price, Shipping Cost, Customer ID, Customer Name, and range. The data consists of 24 rows of customer information with their respective order details and ranges.

#	Row ID	Order Priority	Discount	Unit Price	Shipping Cost	Customer ID	Customer Name	range
1	18606.0	Not Specified	.01	2.88	.50	2.0	Janice Fletcher	Less than 5
2	20847.0	High	.01	2.84	.93	3.0	Bonnie Potter	Less than 5
3	23086.0	Not Specified	.03	6.68	6.15	3.0	Bonnie Potter	5-10
4	23087.0	Not Specified	.01	5.68	3.6	3.0	Bonnie Potter	Less than 5
5	23088.0	Not Specified	.0	205.99	2.5	3.0	Bonnie Potter	Less than 5
6	23597.0	Medium	.09	55.48	14.3	3.0	Bonnie Potter	10-15
7	25549.0	Low	.08	120.97	26.3	3.0	Bonnie Potter	25-30
8	20228.0	Not Specified	.02	500.98	26.0	5.0	Ronnie Proctor	25-30
9	19483.0	Low	.08	6.48	6.81	5.0	Ronnie Proctor	5-10
10	24782.0	High	.01	90.24	0.99	6.0	Dwight Hwang	Less than 5
11	24563.0	Critical	.07	6.48	6.6	6.0	Dwight Hwang	5-10
12	24564.0	Critical	.01	4.84	0.71	6.0	Dwight Hwang	Less than 5
13	24565.0	Critical	.1	85.99	0.99	6.0	Dwight Hwang	Less than 5
14	21866.0	High	.05	12.28	4.86	7.0	Leon Gill	Less than 5
15	20876.0	Medium	.08	140.98	36.09	8.0	Melanie Garner	More than 30
16	20877.0	Medium	.1	286.85	61.76	9.0	Lorraine Houston	More than 30
17	22241.0	Critical	.06	15.57	1.39	10.0	Meredith Norrie Thomas	Less than 5
18	21776.0	Critical	.06	9.48	7.29	11.0	Marcus Dunlap	5-10
19	23328.0	High	.04	10.98	3.37	12.0	Kara Pace	Less than 5
20	24844.0	Medium	.09	78.69	19.99	14.0	Gwendolyn F Tyson	15-20
21	24846.0	Medium	.08	3.28	2.31	14.0	Gwendolyn F Tyson	Less than 5
22	24847.0	Medium	.05	3.28	4.2	14.0	Gwendolyn F Tyson	Less than 5
23	24848.0	Medium	.05	3.58	1.63	14.0	Gwendolyn F Tyson	Less than 5
24	24845.0	Medium	.01	6.48	7.86	14.0	Gwendolyn F Tyson	5-10

The screenshot shows the 'number-range-output.xls.xls' file in Microsoft Excel. The data is presented in a table format across multiple rows. The columns include Row ID, Order Prio, Discount, Unit Price, Shipping, Customer ID, and Customer Name. The data corresponds to the rows shown in the Spoon interface.

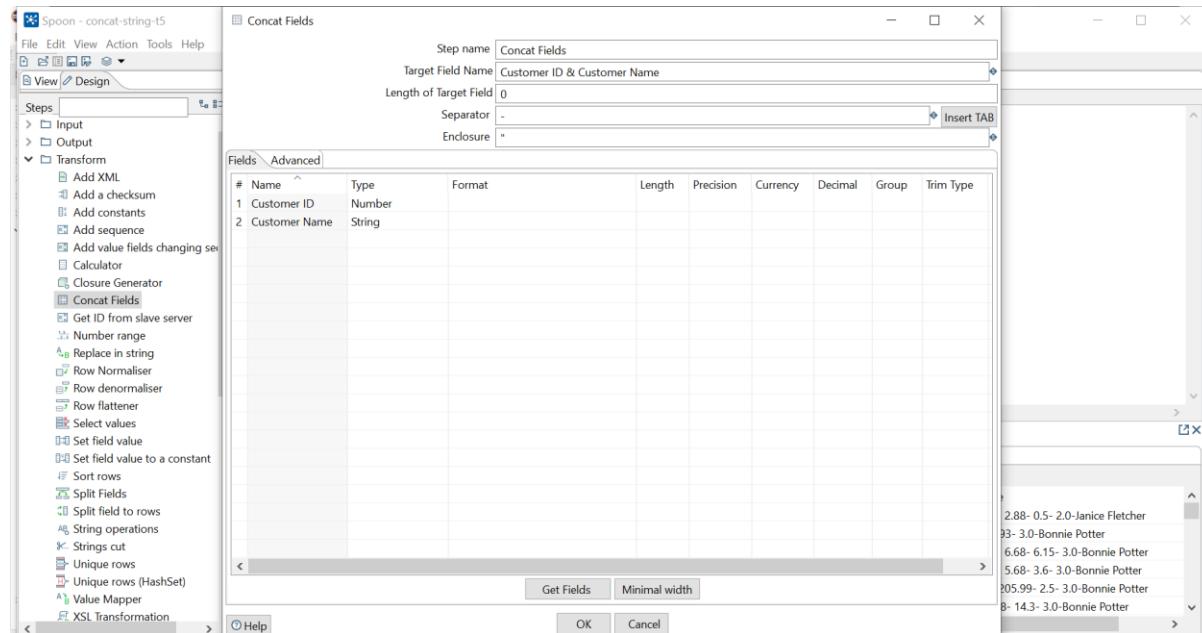
Row ID	Order Prio	Discount	Unit Price	Shipping	Customer ID	Customer Name
18,606.00	Not Specif	.01	2.88	.50	2.0	Janice Fletcher
20,847.00	High	.01	2.84	.93	3.0	Bonnie Potter
23,086.00	Not Specif	.03	6.68	6.15	3.0	Bonnie Potter
23,087.00	Not Specif	.01	5.68	3.60	3.0	Bonnie Potter
23,088.00	Not Specif	.00	205.99	2.50	3.0	Bonnie Potter
23,597.00	Medium	.09	55.48	14.30	3.0	Bonnie Potter
25,549.00	Low	.08	120.97	26.30	3.0	Bonnie Potter
20,228.00	Not Specif	.02	500.98	26.00	5.0	Ronnie Proctor
19,483.00	Low	.08	6.48	6.81	5.0	Ronnie Proctor
24,782.00	High	.01	90.24	.99	6.0	Dwight Hwang
24,563.00	Critical	.07	6.48	6.60	6.0	Dwight Hwang
24,564.00	Critical	.01	4.84	.71	6.0	Dwight Hwang
24,565.00	Critical	.10	85.99	.99	6.0	Dwight Hwang
21,866.00	High	.05	12.28	4.86	7.0	Leon Gill
20,876.00	Medium	.08	140.98	36.09	8.0	Melanie Garner
20,877.00	Medium	.10	286.85	61.76	9.0	Lorraine Houston
22,241.00	Critical	.06	15.57	1.39	10.0	Meredith Norrie Thomas
21,776.00	Critical	.06	9.48	7.29	11.0	Marcus Dunlap
23,328.00	High	.04	10.98	3.37	12.0	Kara Pace
24,844.00	Medium	.09	78.69	19.99	14.0	Gwendolyn F Tyson
24,846.00	Medium	.08	3.28	2.31	14.0	Gwendolyn F Tyson
24,847.00	Medium	.05	3.28	4.2	14.0	Gwendolyn F Tyson
24,848.00	Medium	.05	3.58	1.63	14.0	Gwendolyn F Tyson
24,845.00	Medium	.01	6.48	7.86	14.0	Gwendolyn F Tyson

F) CONCAT TRANSFORMATION:

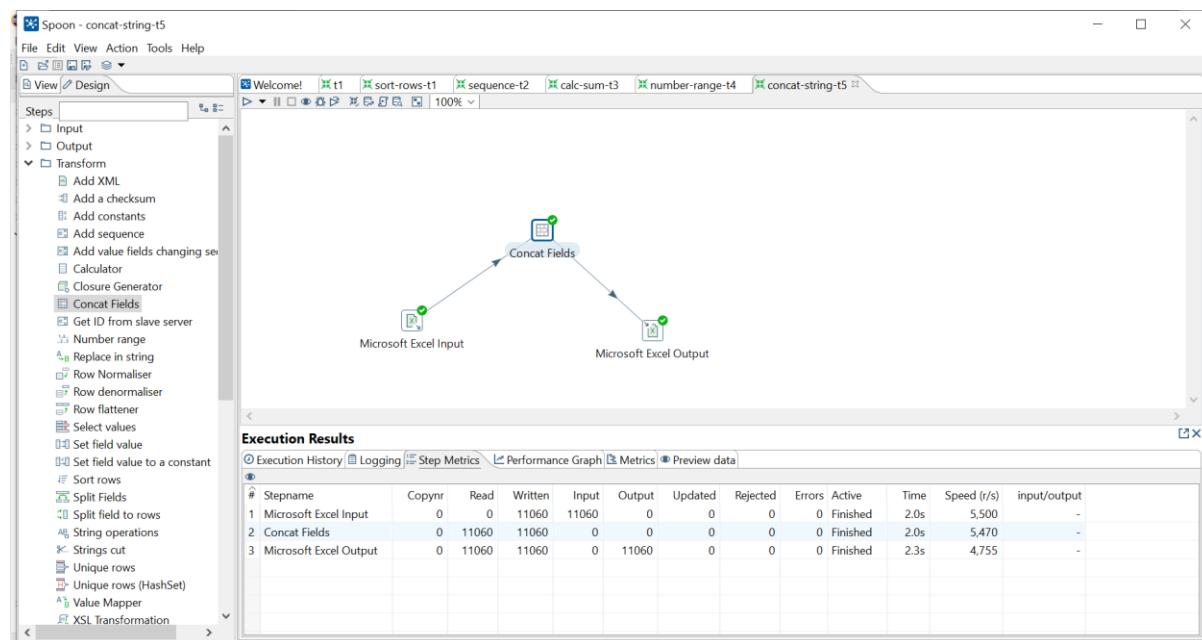
Step 1. Open up the Input folder and drag and drop the Microsoft excel input step on to the transformation window.

Step 2. Open up the Transformation folder and drag and drop the concat fields and config it according to requirements.

Step 3. Open up the output folder and drag and drop the Microsoft excel output.



Step 4. Execute the transformation



OUTPUT:

#	Row ID	Order Priority	Discount	Unit Price	Shipping Cost	Customer ID	Customer Name	Customer ID & Customer Name
1	18606.0	Not Specified	.01	2.88	.50	2.0	Janice Fletcher	2.0-Janice Fletcher
2	20847.0	High	.01	2.84	.93	3.0	Bonnie Potter	3.0-Bonnie Potter
3	23086.0	Not Specified	.03	6.68	6.15	3.0	Bonnie Potter	3.0-Bonnie Potter
4	23087.0	Not Specified	.01	5.68	3.6	3.0	Bonnie Potter	3.0-Bonnie Potter
5	23088.0	Not Specified	.00	205.99	2.5	3.0	Bonnie Potter	3.0-Bonnie Potter
6	23597.0	Medium	.09	55.48	14.3	3.0	Bonnie Potter	3.0-Bonnie Potter
7	25549.0	Low	.08	120.97	26.3	3.0	Bonnie Potter	3.0-Bonnie Potter
8	20228.0	Not Specified	.02	500.98	26.0	5.0	Ronnie Proctor	5.0-Ronnie Proctor
9	19483.0	Low	.08	6.48	6.81	5.0	Ronnie Proctor	5.0-Ronnie Proctor
10	24782.0	High	.01	90.24	0.99	6.0	Dwight Hwang	6.0-Dwight Hwang
11	24563.0	Critical	.07	6.48	6.6	6.0	Dwight Hwang	6.0-Dwight Hwang
12	24564.0	Critical	.01	4.84	0.71	6.0	Dwight Hwang	6.0-Dwight Hwang
13	24565.0	Critical	.01	85.99	0.99	6.0	Dwight Hwang	6.0-Dwight Hwang
14	21866.0	High	.05	12.28	4.86	7.0	Leon Gill	7.0-Leon Gill
15	20876.0	Medium	.08	140.98	36.09	8.0	Melanie Garner	8.0-Melanie Garner
16	20877.0	Medium	.01	286.85	61.76	9.0	Lorraine Houston	9.0-Lorraine Houston
17	22241.0	Critical	.06	15.57	1.39	10.0	Meredith Norris Thomas	10.0-Meredith Norris Thomas
18	21776.0	Critical	.06	9.48	7.29	11.0	Marcus Dunlap	11.0-Marcus Dunlap
19	23328.0	High	.04	10.98	3.37	12.0	Kara Pace	12.0-Kara Pace
20	24844.0	Medium	.09	78.69	19.99	14.0	Gwendolyn F Tyson	14.0-Gwendolyn F Tyson
21	24846.0	Medium	.08	3.28	2.31	14.0	Gwendolyn F Tyson	14.0-Gwendolyn F Tyson
22	24847.0	Medium	.05	3.28	4.2	14.0	Gwendolyn F Tyson	14.0-Gwendolyn F Tyson
23	24848.0	Medium	.05	3.58	1.63	14.0	Gwendolyn F Tyson	14.0-Gwendolyn F Tyson
24	24845.0	Medium	.01	6.48	7.86	14.0	Gwendolyn F Tyson	14.0-Gwendolyn F Tyson

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	Row ID	Order Prio	Discount	Unit Price	Shipping	Customer ID	Customer Name											
2	18,606.00	Not Specif	.01	2.88	.50	2.0	Janice Fletcher	2.0-Janice Fletcher										
3	20,847.00	High	.01	2.84	.93	3.0	Bonnie Pot	3.0-Bonnie Potter										
4	23,086.00	Not Specif	.03	6.68	6.15	3.0	Bonnie Pot	3.0-Bonnie Potter										
5	23,087.00	Not Specif	.01	5.68	3.60	3.0	Bonnie Pot	3.0-Bonnie Potter										
6	23,088.00	Not Specif	.00	205.99	2.50	3.0	Bonnie Pot	3.0-Bonnie Potter										
7	23,597.00	Medium	.09	55.48	14.30	3.0	Bonnie Pot	3.0-Bonnie Potter										
8	25,549.00	Low	.08	120.97	26.30	3.0	Bonnie Pot	3.0-Bonnie Potter										
9	20,228.00	Not Specif	.02	500.98	26.00	5.0	Ronnie Procto	5.0-Ronnie Proctor										
10	19,483.00	Low	.08	6.48	6.81	5.0	Ronnie Procto	5.0-Ronnie Proctor										
11	24,782.00	High	.01	90.24	.99	6.0	Dwight Hwang	6.0-Dwight Hwang										
12	24,563.00	Critical	.07	6.48	6.60	6.0	Dwight Hwang	6.0-Dwight Hwang										
13	24,564.00	Critical	.01	4.84	.71	6.0	Dwight Hwang	6.0-Dwight Hwang										
14	24,565.00	Critical	.10	85.99	.99	6.0	Dwight Hwang	6.0-Dwight Hwang										
15	21,866.00	High	.05	12.28	4.86	7.0	Leon Gill	7.0-Leon Gill										
16	20,876.00	Medium	.08	140.98	36.09	8.0	Melanie Garner	8.0-Melanie Garner										
17	20,877.00	Medium	.10	286.85	61.76	9.0	Lorraine Houston	9.0-Lorraine Houston										
18	22,241.00	Critical	.06	15.57	1.39	10.0	Meredith Thomas	10.0-Meredith Norris Thomas										
19	21,776.00	Critical	.06	9.48	7.29	11.0	Marcus Dunlap	11.0-Marcus Dunlap										
20	23,328.00	High	.04	10.98	3.37	12.0	Kara Pace	12.0-Kara Pace										
21	24,844.00	Medium	.09	78.69	19.99	14.0	Gwendolyn F Tyson	14.0-Gwendolyn F Tyson										

CONCLUSION:

From this practical, I have learned how to do ETL (Extract, Transform, Load) in Pentaho.