pentaho®

SOFTWARE EVALUATION FOR IS7034 (001)

A → O Intelligence!

Anusha Gona

Vamsi Krishna Kalepu

Krunal Dilip Shigavan

Nageswara Rao Ch

TABLE OF CONTENTS

INTRODUCTION

"Above all else show the data" - Edward Tufte

is Business Intelligence (BI) software that provides Data Integration, Online Analytical Processing (OLAP), Reporting, Dashboard, Data Mining and Extract, Transform, Load (ETL) services. The head office is in Orlando, Florida. Pentaho was acquired by Hitachi Data Systems in 2015 and became part of Hitachi Vantara in 2017.

Pentaho is a Java framework for building business intelligence solutions. While Pentaho is best known for its Business Analysis Server (formerly known as Business Intelligence Server), Pentaho software consists of several Java classes with specific functions. In addition to these Java classes, any BI solution can be created.

The only exception to this pattern is Pentaho Data Integration's ETL tool – PDI (formerly known as Kettle).

Kettle was a powerful ETL tool based on Java. Kettle itself represents its meaning; KETTLE stands for Kettle Extraction Transformation Transport Load Environment. Matt Casters, an independent BI consultant developed Kettle and it was open sourced in 2005.
It was acquired by Pentaho in 2006 and the name was changed to 'Pentaho Data Integration.

There are many components such as Spoon, Pan, Kitchen, Carte – all these names are culinary metaphors given to these offerings.
PDI is a set of software for designing data streams that can run both on the server and in standalone processes. PDI includes Kitchen, a tool for managing tasks and transformations, and Spoon, a graphical user interface for designing these tasks and transformations.

Features such as reporting and OLAP are achieved by integrating sub-projects into Pentaho such as the Mondrian OLAP engine and jFree Report. These projects have been curated by Pentaho for some time. Some of these sub-projects also have standalone clients, such as Pentaho Report Designer, a frontend for jFree Reports, and Pentaho Schema Workbench, a graphical user interface for writing XML used by Mondrian to manage OLAP cubes.

Pentaho offers corporate and community editions of this software. The enterprise software comes with an annual subscription and includes additional features and support not included in the Community Edition. Pentaho's core offering is often complemented by additional products, mostly in the form of plugins, from the company and the broader user community.

USE CASE:

Common Use Cases of Pentaho Data Integration

Though PDI is an ETL and integration tool, it can be used for:

- Data migration between different databases and applications
- Integration of real-time ETL as a data source
- Data cleansing with steps
- Hadoop job execution and scheduling
- Rapid prototyping of ROLAP schemas
- Data warehouse population with built-in support for changing dimensions.

Scenario:
John is an analytics Manager at a furniture sales company. He wants to create a sales dashboard to help the company understand its sales performance. To do this, he needs to collect sales data from the company's database. However, the file he receives is incomplete, and he needs to clean it up before he can use it.

Process:

Data Collection: John collects the sales data from the company's database and receives a file in a CSV format.

Cleaning the Data:
John cleans the data by removing any null values and replacing a wrong string with the right one. To do this, he uses an ETL process that performs the following steps:

- Extract: John extracts the data from the Excel file using an ETL tool.
- Transform: John uses the ETL tool to remove any null values and replace the wrong string with the right one. He also adds a flag to the records that are missing important information, indicating that the records need further details from the manager before they can be included in the sales dashboard.
- Load: John loads the cleaned data into a new CSV file.

The whole process is carried out as a job in the PDI where the complete file is used to make Sales dashboard using Power BI and incomplete file is sent to manager for further details.

SOFTWARE DETAILS

Pentaho Reporting is a suite (collection of tools) for creating relational and analytical reporting. Using Pentaho, we can transform complex data into meaningful reports and draw information out of them. Pentaho supports creating reports in various formats such as HTML, Excel, PDF, Text, CSV, and xml.
Pentaho can accept data from different data sources including SQL databases, OLAP data sources, and even the Pentaho Data Integration ETL tool.

Components of the Software

Pentaho is a comprehensive data integration and business intelligence software suite that consists of several different components. Here are some of the key components of the Pentaho software:

1. **Pentaho Data Integration (PDI):** This is the data integration component of the Pentaho suite, which is used to extract, transform, and load (ETL) data from various sources into a target system or data warehouse. It includes a drag-and-drop graphical interface, as well as a powerful scripting language for more advanced transformations.
2. **Pentaho Business Analytics (BA):** This component provides advanced reporting and analytics capabilities, allowing users to create interactive dashboards and reports, perform ad hoc analysis, and share insights with others.
3. **Pentaho Metadata:** This component allows users to define and manage metadata, which provides a layer of abstraction between the physical data sources and the business logic of the reporting and analysis tools.
4. **Pentaho Report Designer:** This is a standalone tool that allows users to create professional-looking reports using a variety of data sources.
5. **Pentaho Schema Workbench:** This tool is used to design and manage multidimensional data models for use in OLAP (online analytical processing) applications.
6. **Pentaho Aggregation Designer:** This tool provides a graphical interface for designing and testing aggregate tables, which can improve query performance in OLAP applications.
7. **Pentaho Dashboard Designer:** This component allows users to create interactive dashboards, which can include charts, tables, and other data visualizations.
8. **Pentaho Analyzer:** This tool provides ad-hoc data analysis capabilities, allowing users to explore data in a self-service manner.

Overall, the Pentaho software suite is designed to provide end-to-end data integration and business intelligence capabilities, from data integration and transformation to reporting, analysis, and visualization.

KEY ALTERNATIVES

There are several alternatives to Pentaho in the data integration and business intelligence space. Here are a few key alternatives:

Talend is a data integration platform that provides ETL, data quality, and data governance capabilities. It also includes a suite of business intelligence tools for reporting, analytics, and data visualization.

Microsoft Power BI is a cloud-based business intelligence platform that includes data visualization, reporting, and analytics capabilities. It integrates with a variety of data sources and provides a user-friendly interface for creating dashboards and reports.

QlikView is a business intelligence tool that provides data visualization, reporting, and analytics capabilities. It includes a proprietary data engine that allows for rapid data analysis and visualization.

Apache Nifi is an open-source data integration platform that allows users to easily build data pipelines for moving, transforming, and processing data. It includes a web-based interface for creating and managing data flows.

Tableau is a powerful data visualization tool that allows users to create interactive dashboards, reports, and charts. It includes a wide range of data connectors to various data sources, including big data platforms.

IBM Cognos Analytics is a business intelligence platform that includes reporting, analysis, and dashboarding capabilities. It can be deployed on-premises or in the cloud.

Oracle Business Intelligence is a suite of tools for reporting, analysis, and data visualization. It includes a web-based interface for creating dashboards and reports.

SAP BusinessObjects is a suite of business intelligence tools that includes reporting, analysis, and data visualization capabilities. It can be deployed on-premises or in the cloud.

These are just a few of the many alternatives to Pentaho in the data integration and business intelligence space. Each of these products has its own strengths and weaknesses, and the best solution for a particular use case will depend on the specific needs of the organization and the data being analyzed.

SWOT ANALYSIS

Pentaho is a comprehensive data integration and business intelligence suite that has several strengths and weaknesses. Here are some of the key strengths and weaknesses of Pentaho:

Strengths

1. Comprehensive: Pentaho offers a comprehensive set of tools for data integration, ETL, reporting, analysis, and visualization, which makes it a one-stop solution for all data-related needs.
2. Open-source: Pentaho is an open-source software suite, which means that users can customize and extend the platform as needed.
3. Scalable: Pentaho can handle large volumes of data and can scale up or down as needed.
4. User-friendly: Pentaho includes a user-friendly, web-based interface that allows users to create and manage data workflows, reports, and visualizations without needing to write code.
5. Integration: Pentaho integrates with a wide range of data sources, including traditional databases, big data platforms, and cloud-based data sources.

Weakness

1. Support: The support for Pentaho is not as extensive as some of the other proprietary tools in the market, which may be a concern for some users.
2. Complexity: Pentaho is a powerful platform, but it can be complex to set up and configure, especially for those with limited technical expertise.
3. Resource Intensive: Pentaho requires a significant number of resources to run, which may be a challenge for organizations with limited resources.
4. Limited advanced analytics: While Pentaho offers basic reporting and analysis capabilities, it may not have the same advanced analytics features as some of the other platforms in the market, which may be a drawback for organizations that require advanced predictive or prescriptive analytics.

Opportunities

6. Growing Demand for Business Intelligence.
7. Adoption of Big Data Technologies
8. Expansion of Cloud-based Offerings

Threats

5. Competition from Established Players
6. Rapidly Evolving Market
7. Dependence on Open-Source Community.

Overall, Pentaho is a powerful and comprehensive platform that offers a lot of flexibility and customization options, but it may not be the best fit for all organizations, especially those with limited technical resources or a need for more advanced analytics capabilities.

SOFTWARE RECOMMENDATION

Pentaho is a comprehensive Business Intelligence (BI) suite that provides all the necessary components for BI and analytics. It includes data integration, reporting, dashboarding, data mining, and analytics. While learning Pentaho was challenging due to limited resources, the BI suite is a great choice for those who want a complete solution from a single software, but it's important to evaluate use cases and consider alternative solutions based on specific needs.

However, the community edition of Pentaho has limited capabilities, and it may not be sufficient for all use cases. We initially faced challenges with the software's user interface and documentation.

Despite these challenges, Pentaho has several benefits, including its ability to integrate with various data sources, its powerful ETL (extract, transform, load) capabilities. We could not explore the flexible features of reporting and dashboards as it is part of enterprise edition.

FUTURE DEVELOPMENTS

Based on the current market trends and industry needs, here are some potential future developments of Pentaho:

- **More AI and Machine Learning Capabilities:** As AI and machine learning continue to gain popularity, Pentaho could incorporate more advanced analytics capabilities, such as predictive and prescriptive analytics, to help businesses make data-driven decisions.
- **Enhanced Cloud Capabilities:** With the increasing adoption of cloud-based solutions, Pentaho may look to improve its cloud capabilities and expand its integrations with cloud platforms, such as AWS, Microsoft Azure, and Google Cloud Platform.
- **Increased Data Security:** With the growing importance of data privacy and security, Pentaho could focus on enhancing its security features, such as improving data encryption and access control, to better protect sensitive data.
- **Improved User Interface:** Pentaho could continue to enhance its user interface to make it more user-friendly and intuitive, allowing even non-technical users to easily create and manage data workflows and visualizations.
- **Continued Focus on Data Integration:** Pentaho may continue to strengthen its data integration capabilities, making it easier for organizations to connect to a wide range of data sources and integrate data from disparate systems.

These are a few potential future developments of Pentaho, but as with any software platform, the actual development roadmap may vary based on market needs, customer feedback, and technological advancements.

CONCLUSION

Pentaho is a comprehensive data integration and business analytics platform that provides a range of tools to manage and analyze data, from data integration to reporting and data visualization. The platform is used by organizations of all sizes to improve their decision-making processes, gain insights into their data, and optimize their operations.

Based on the features and capabilities of Pentaho, it can be concluded that:

- Pentaho is a highly flexible platform that can be customized to meet the specific needs of organizations in various industries.
- The platform offers a range of powerful data integration tools, including data extraction, transformation, and loading (ETL), data cleansing, and data profiling.
- Pentaho provides advanced analytics capabilities, such as predictive analytics, data mining, and machine learning, which can help organizations gain valuable insights into their data.
- The platform also offers a comprehensive set of reporting and data visualization tools, which enable organizations to create dashboards, reports, and other visualizations that can be shared with stakeholders.
- Pentaho has a strong community of developers, users, and contributors, who provide support, share knowledge, and create plugins and extensions to enhance the platform's functionality.

Overall, Pentaho is a robust and versatile platform that can help organizations manage and analyze their data, make better decisions, and achieve their business goals.

REFERENCES

[1.] https://www.slideshare.net/XpandIT/real-use-cases-pentaho-big-data-ecosystem

[2.] https://help.hitachivantara.com/Documentation/Pentaho/9.0/Setup/Pentaho_Data_Integration_(PDI)_tutorial

[3.] https://www.spec-india.com/blog/pentaho-data-integration-kettle

[4.] https://en.wikipedia.org/wiki/Pentaho

[5.] https://learn.microsoft.com/en-us/power-bi/

APPENDIX A: TUTORIAL

HOW TO GET THE SOFTWARE AND DATA

PRE-REQUISITES:

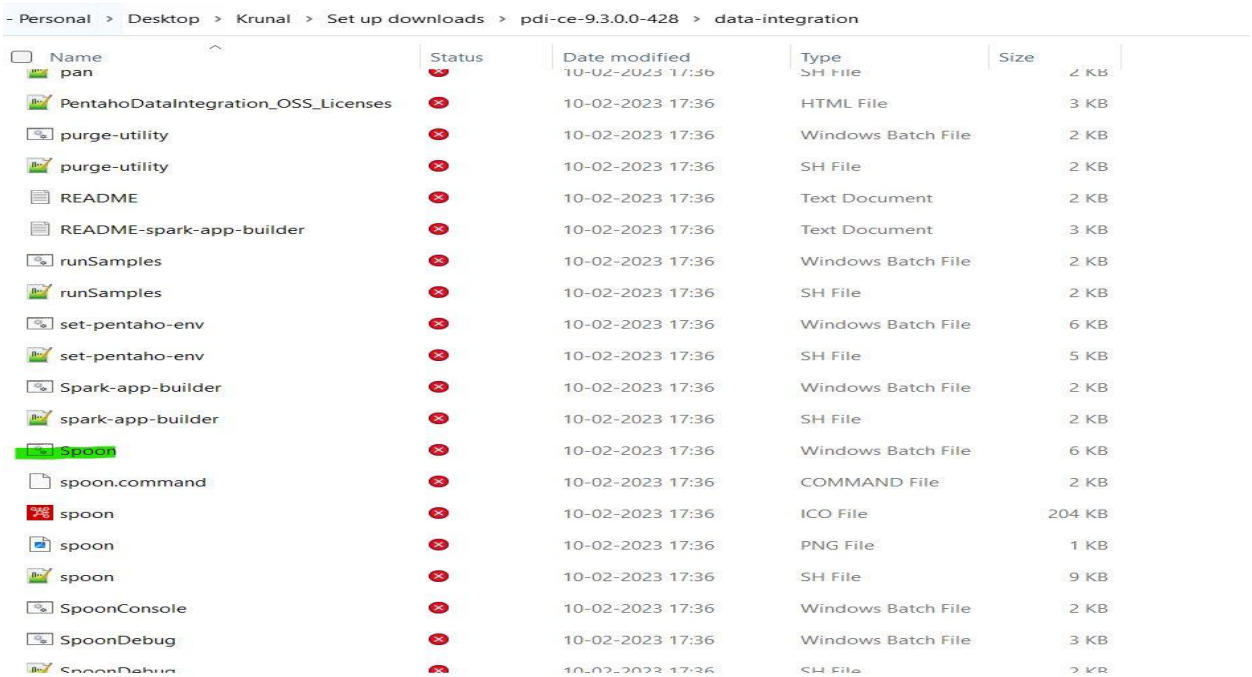
1. Check if Java is installed in your system.
2. If not installed, Java latest version & set the jdk library path in Environment Variable
3. If Java is installed, check java version in command prompt using "java -version"
4. If Java version is not the latest, update the Java version.

INSTALLATION

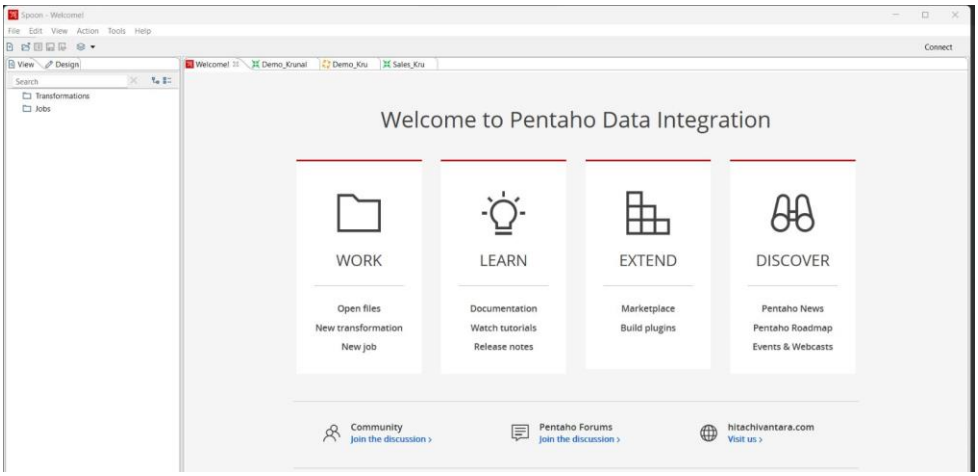
DOWNLOAD LINK

https://sourceforge.net/projects/pentaho/

- Download the zip file (Current latest version is 9.3.0.0)
- Unzip it in a folder of your choice.
- Extract the zip file.
- Run Spoon Batch File



- This is the community edition. All the things taught in this course can be implemented on enterprise edition also. Enterprise editions some additional features, you learn more about them in this video: https://www.hitachivantara.com/en-us/video/pentaho-community-edition-vs-enterprise-edition.html
- If your office or your client is using an older version of PDI, you can find the older versions in the 'files' tab.



HOW TO LOAD DATA

DATA SOURCE:

- Sample data files are available in the downloaded zip file.
- Available data have file types - Text File (.txt), JSON File (.js), XML File(.xml), CSV File(.csv)
- The downloaded zip file even has sample transformation & job files.

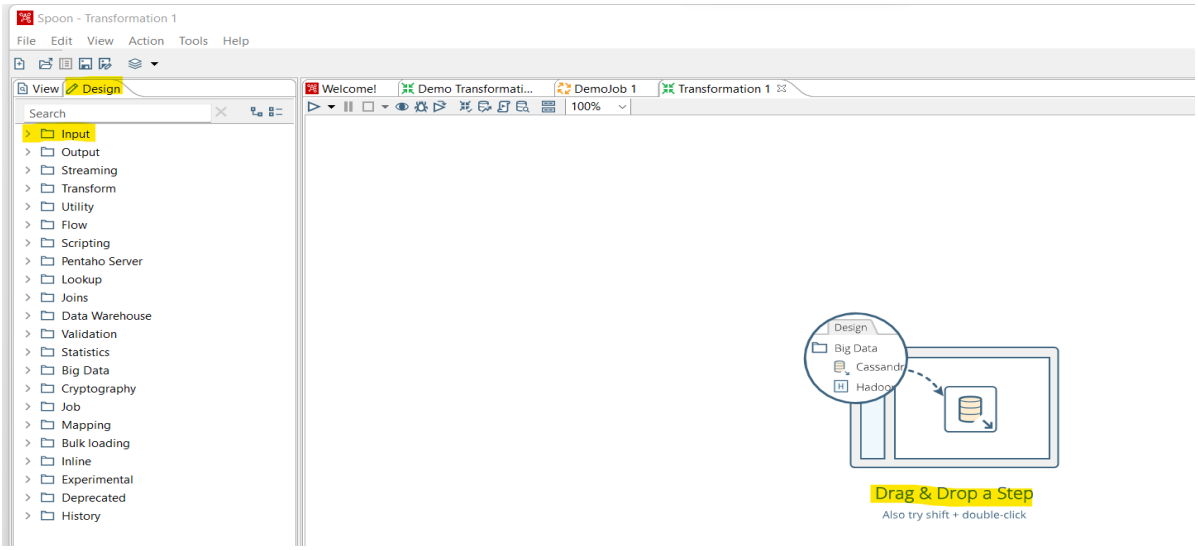
DATA LOADING:

Before understanding, how pentaho loads data for transforming or cleaning we should know about below concepts:

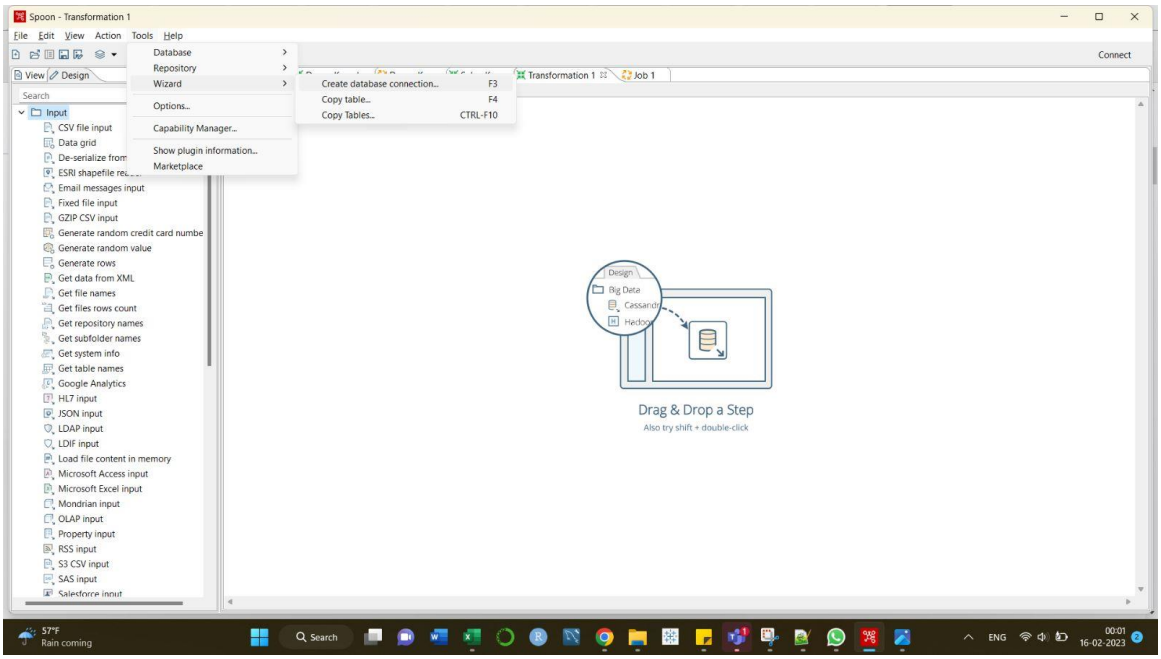
1. Transformation - Data transformation is part of an ETL process and refers to preparing data for analysis. This involves cleaning (removing duplicates, fill-in missing values), reshaping (converting currencies, pivot tables), and computing new dimensions and metrics.
2. Jobs - These are activities that supports transformation process like file management, mail, aborting the process.

Data can be loaded in Pentaho in two ways:

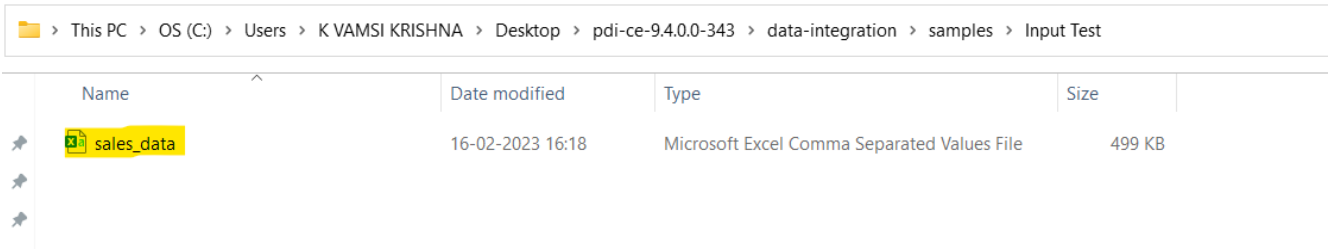
1. We can use the Input option.



2. We can connect Pentaho to the database.



Sample data files are available in the downloaded zip file.



USE THE SOFTWARE

HOW TO USE PENTAHO DATA INTEGRATION:

USE CASE:

John is an analytics Manager at a furniture sales company. He wants to create a sales dashboard to help the company understand its sales performance. To do this, he needs to collect sales data from the company's database. However, the file he receives is incomplete, and he needs to clean it up before he can use it.

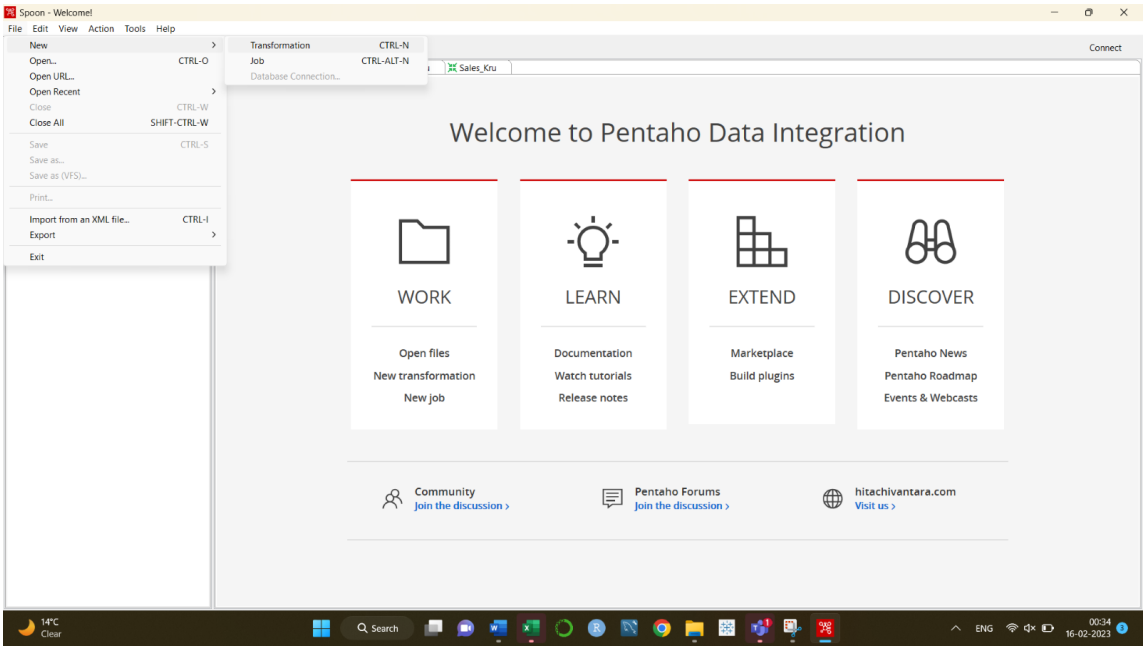
Process:

1. Check whether there is a sales file in the sales folder.
2. If the file is available, import data from the file. Extract
3. Identify the rows with missing data and create a separate file for them. Transform
4. Upload the complete file as an Excel file in dashboard folder. Loading
5. Send the incomplete file to the sales manager for rectification.

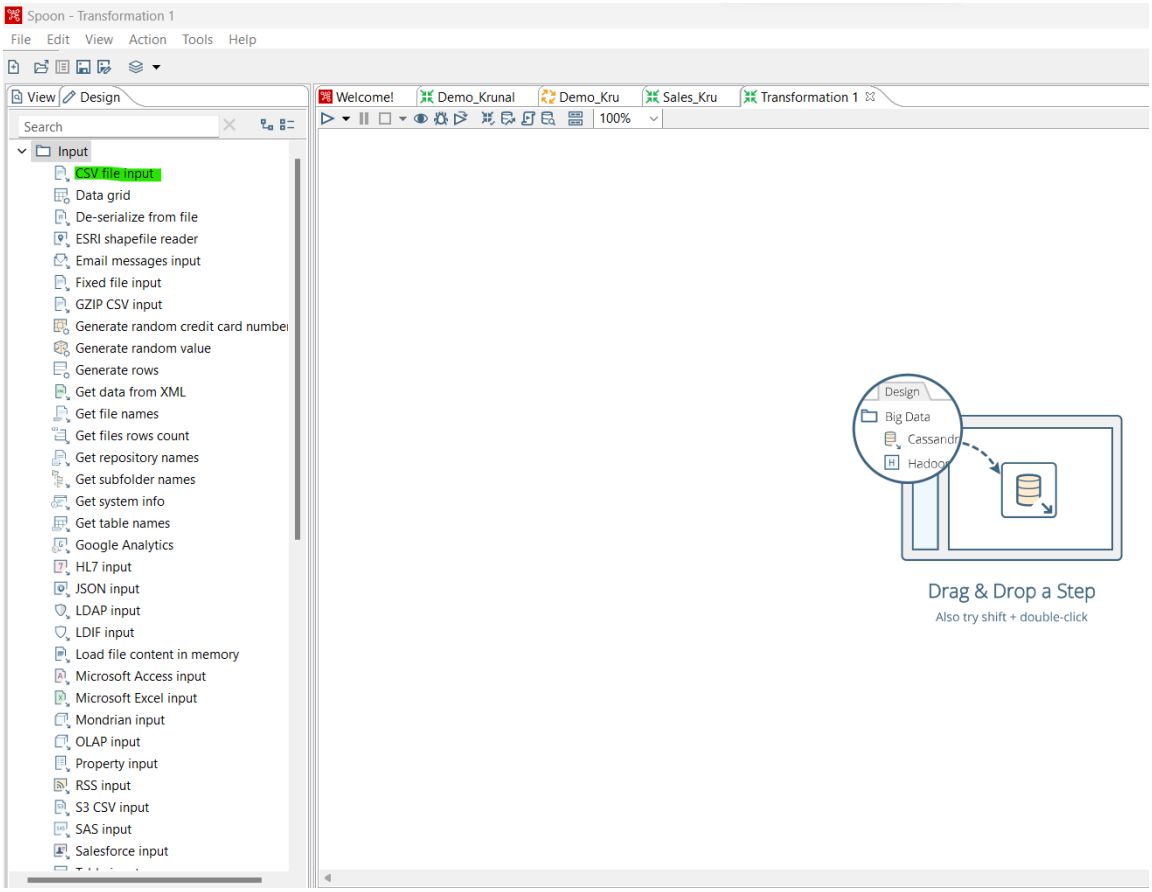
Steps 2,3 and 4 will be done in a PDI Transformation. Steps 1, Transformation, and step 5 will be run as part of PDI Job

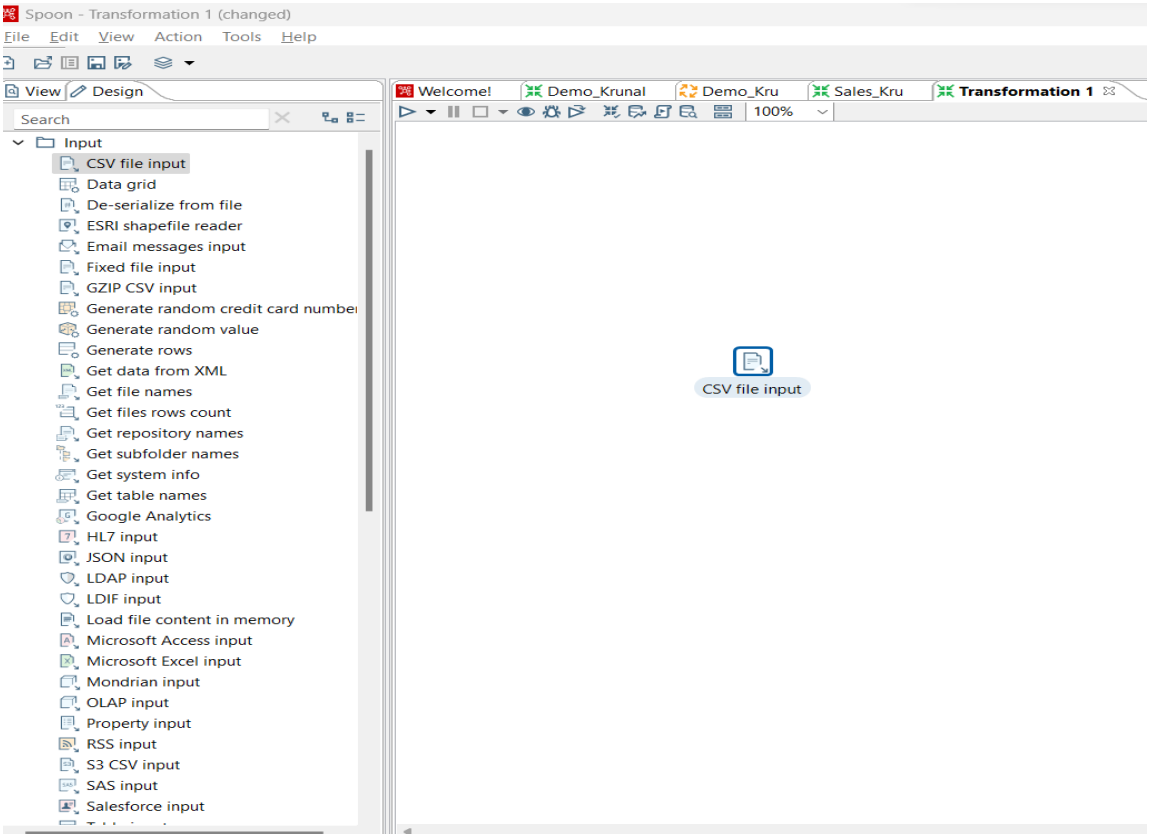
Step-1: **Transformation Process**

a. Create New Transformation File

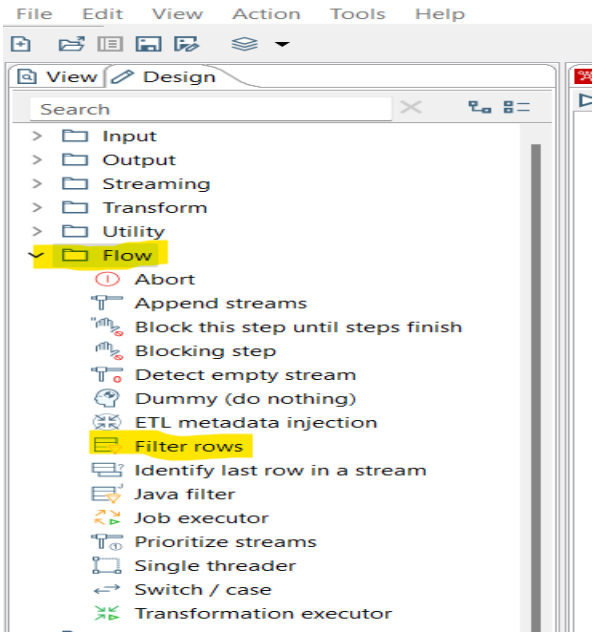


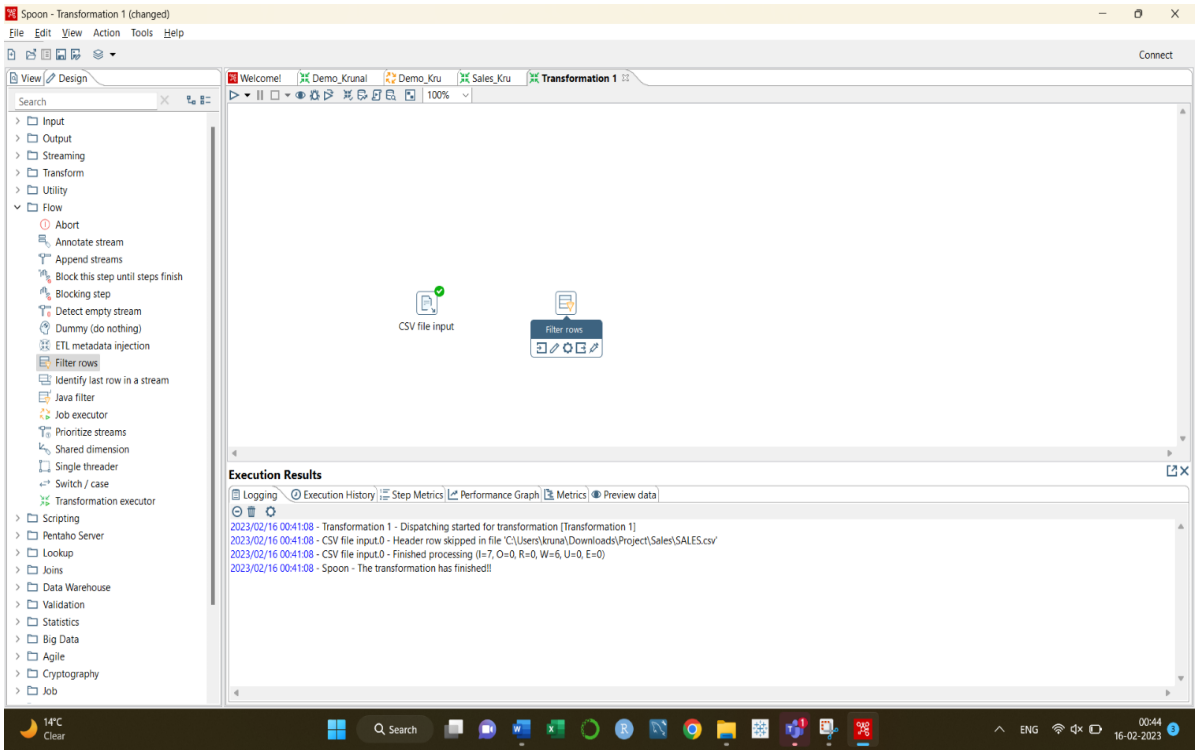
b. Select CSV file Input option from Input Section





c. To filter out the null values in the CSV File, we will apply Filter Rows (Flow >> Filter Rows)

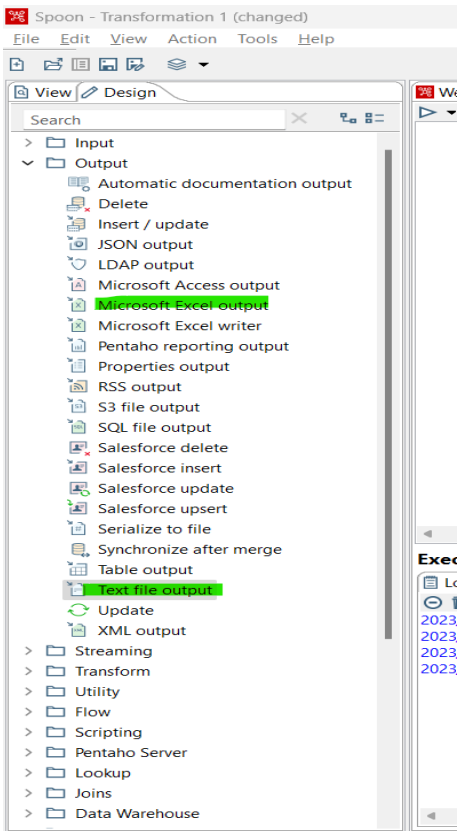




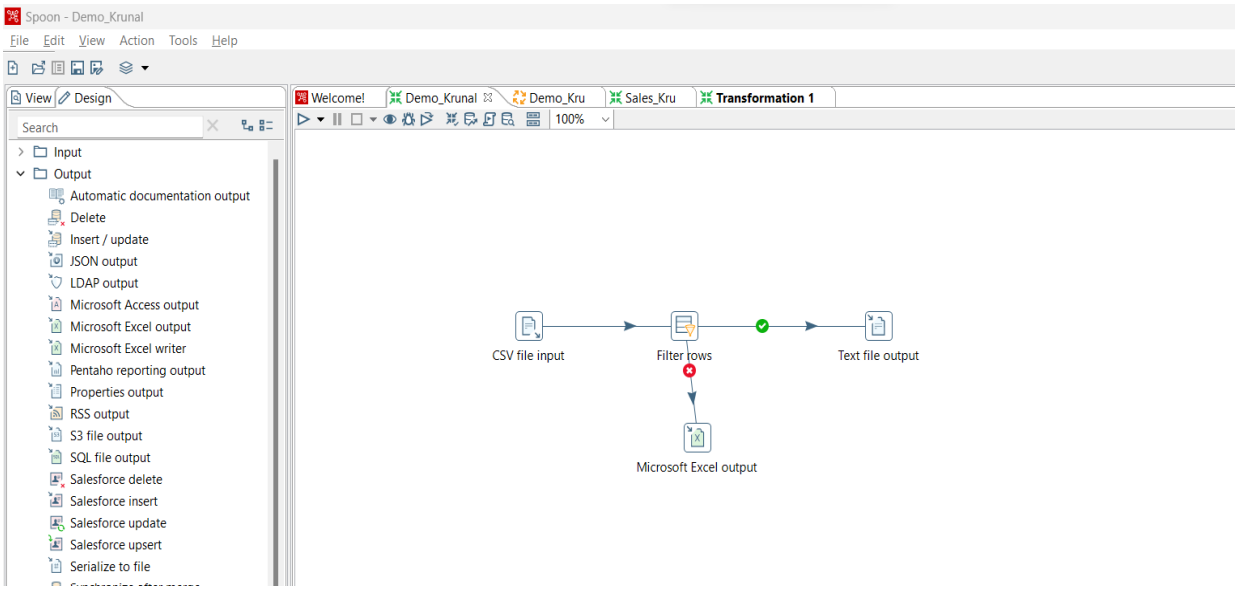
d. We will add two Steps from Output Section:

   i. Text File Output: It will contain data which has no Null values.

   ii. Microsoft Excel File: It will contain data which as Null values.

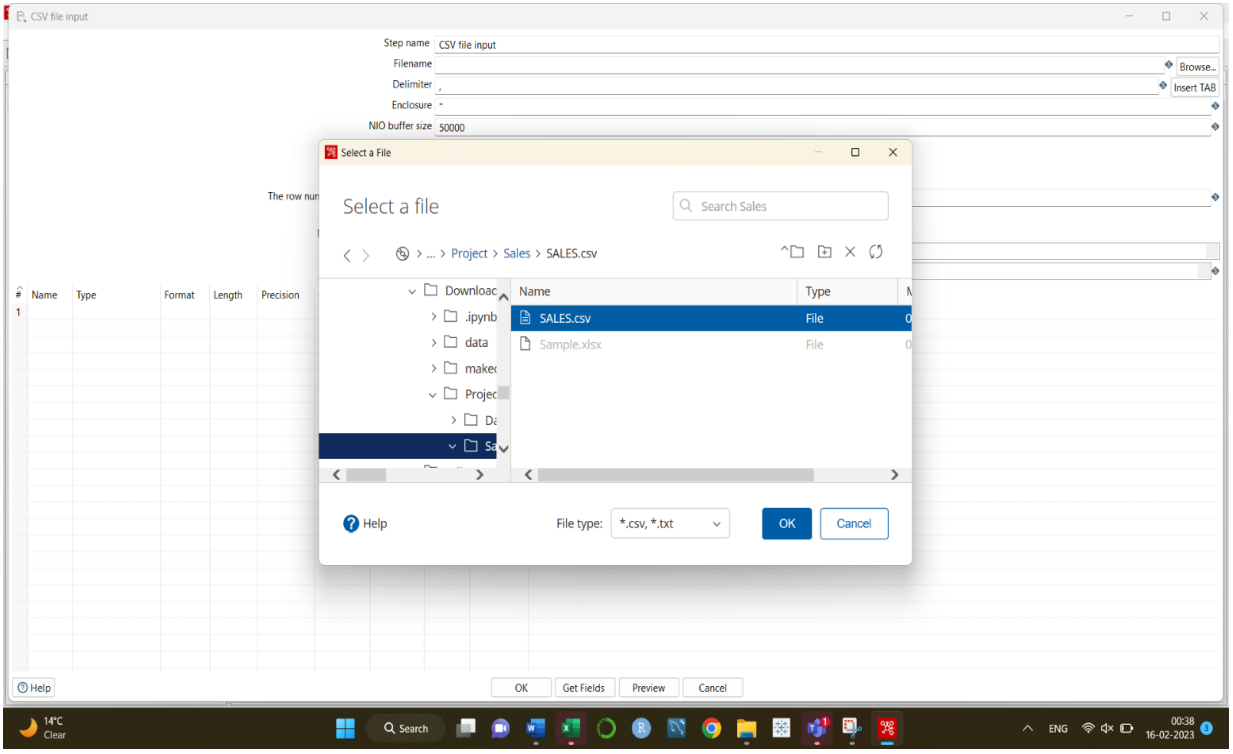


e. Connect all the STEPS using HOPS and set the flow of the Automation Process:

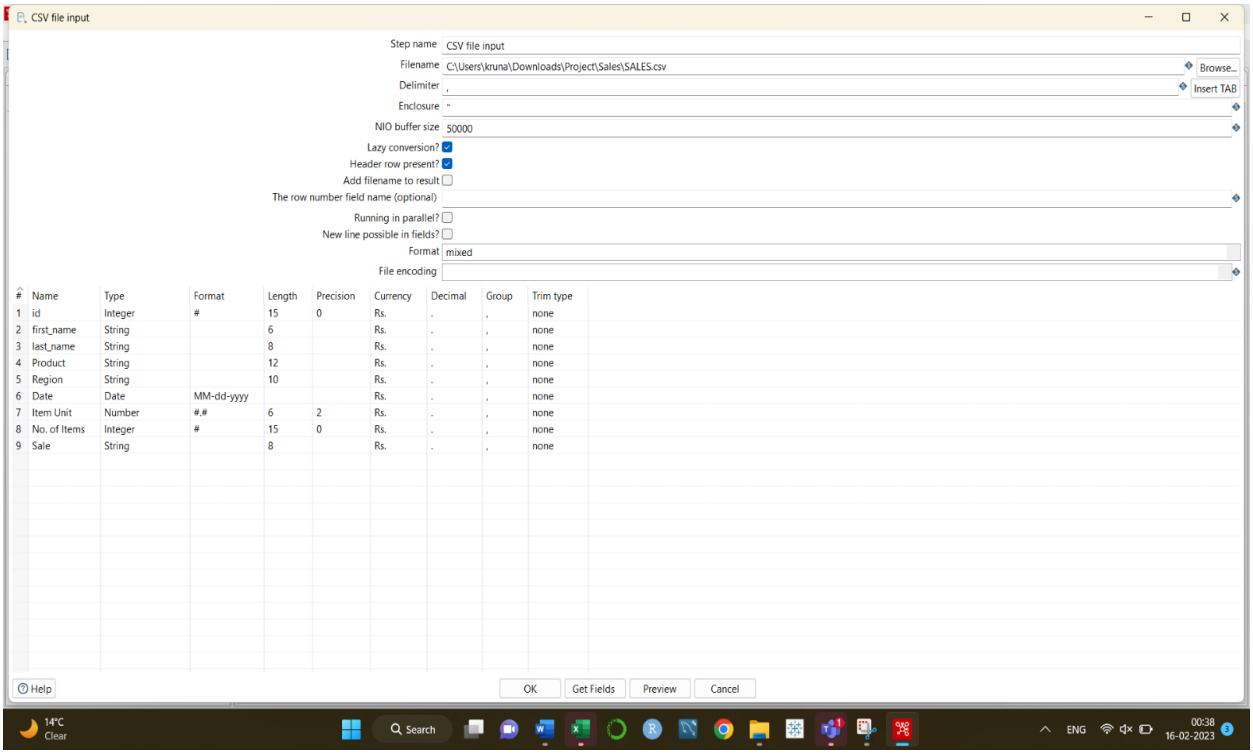


Step-2: **Transformation Process Configuration**

a. Set CSV File path:

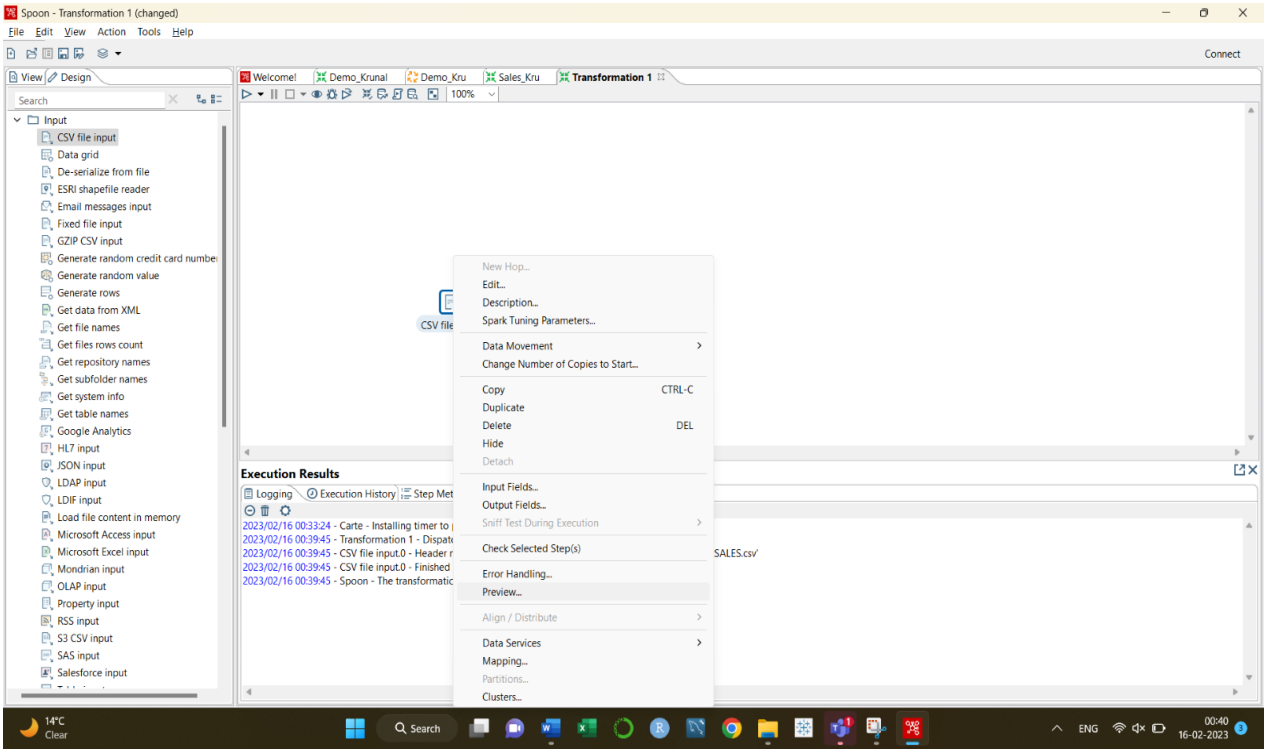


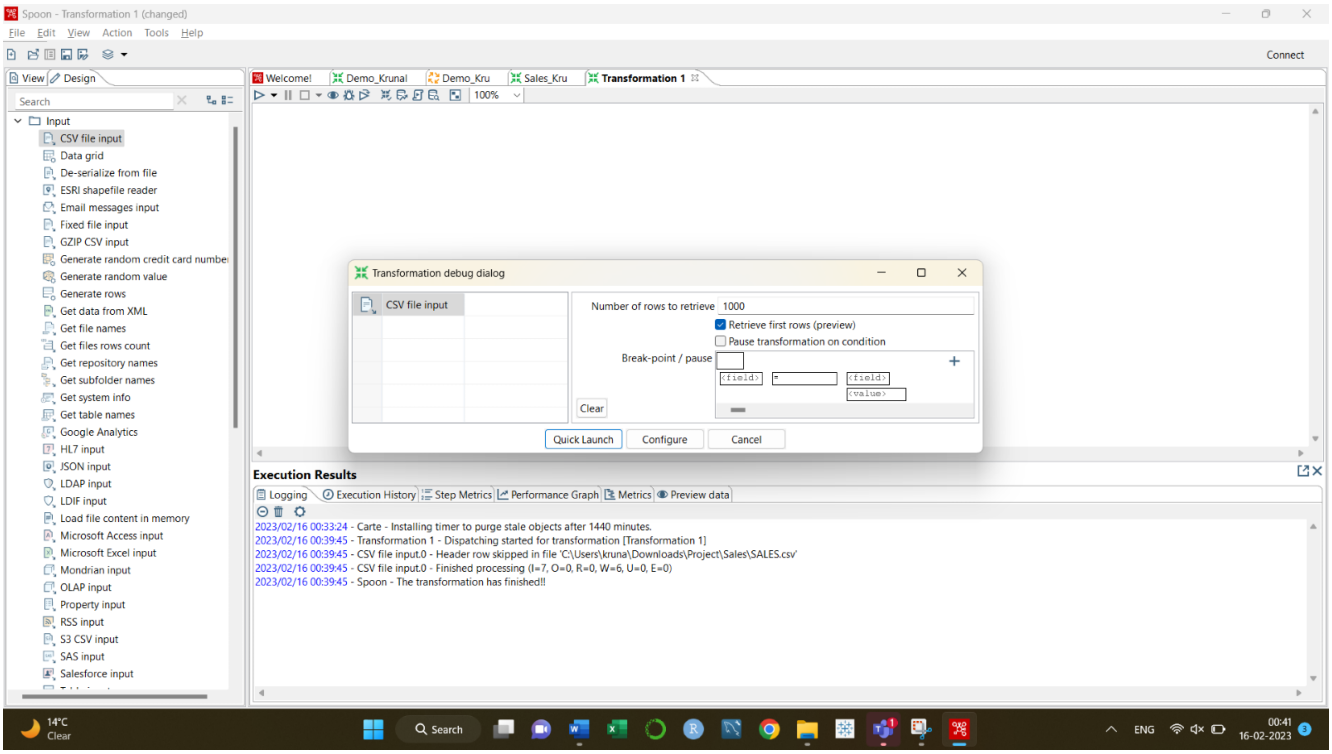
b. Click on Get Fields: This will extract the headers of CSV File



#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim type
1	id	Integer	#	15	0	Rs.	.	,	none
2	first_name	String		6		Rs.	.	,	none
3	last_name	String		8		Rs.	.	,	none
4	Product	String		12		Rs.	.	,	none
5	Region	String		10		Rs.	.	,	none
6	Date	Date	MM-dd-yyyy			Rs.	.	,	none
7	Item Unit	Number	#.#	6	2	Rs.	.	,	none
8	No. of Items	Integer	#	15	0	Rs.	.	,	none
9	Sale	String		8		Rs.	.	,	none

c. To check the data of CSV File, right-click on CSV file Input >> Preview >> Quick Launch





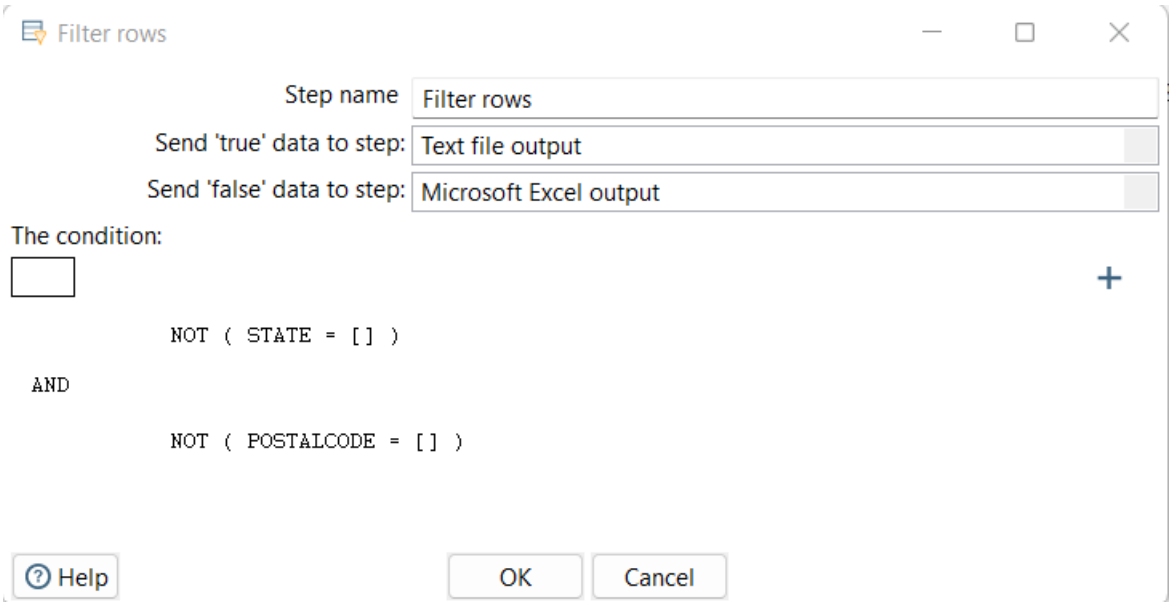
Examine preview data

Rows of step: CSV file input (1000 rows)

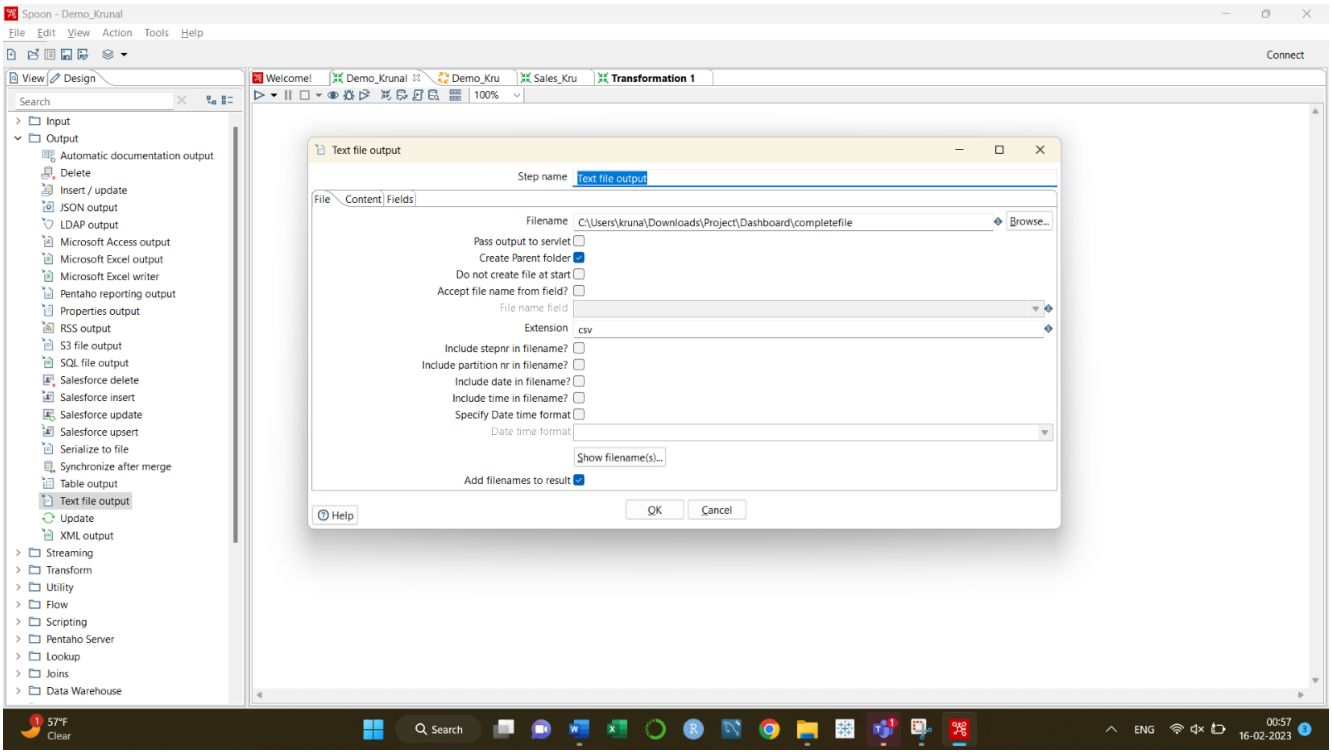
MSRP	PRODUCTCODE	CUSTOMERNAME	PHONE	ADDRESSLINE1	CITY	STATE	POSTALCODE
95	S10_1678	Land of Toys Inc.	2125557818	897 Long Airport Avenue	NYC	NY	10022
95	S10_1678	Reims Collectables	26.47.1555	59 rue de l'Abbaye	Reims	<null>	51100
95	S10_1678	Lyon Souveniers	+33 1 46 62 7555	27 rue du Colonel Pierre Avia	Paris	<null>	75508
95	S10_1678	Toys4GrownUps.com	6265557265	78934 Hillside Dr.	Pasadena	CA	90003
95	S10_1678	Corporate Gift Ideas Co.	6505551386	7734 Strong St.	San Francisco	CA	<null>
95	S10_1678	Technics Stores Inc.	6505556809	9408 Furth Circle	Burlingame	CA	94217
95	S10_1678	Daedalus Designs Imports	20.16.1555	184, chauss,e de Tournai	Lille	<null>	59000
95	S10_1678	Herkku Gifts	+47 2267 3215	Drammen 121, PR 744 Sentrum	Bergen	<null>	N 5804
95	S10_1678	Mini Wheels Co.	6505555787	5557 North Pendale Street	San Francisco	CA	<null>
95	S10_1678	Auto Canal+ Petit	(1) 47.55.6555	25, rue Lauriston	Paris	<null>	75016
95	S10_1678	Australian Collectors, Co.	03 9520 4555	636 St Kilda Road	Melbourne	Victoria	3004
95	S10_1678	Vitachrome Inc.	2125551500	2678 Kingston Rd.	NYC	NY	10022
95	S10_1678	Tekni Collectables Inc.	2015559350	7476 Moss Rd.	Newark	NJ	94019
95	S10_1678	Gift Depot Inc.	2035552570	25593 South Bay Ln.	Bridgewater	CT	97562
95	S10_1678	La Rochelle Gifts	40.67.8555	67, rue des Cinquante Otages	Nantes	<null>	44000
95	S10_1678	Marta's Replicas Co.	6175558555	39323 Spinnaker Dr.	Cambridge	MA	51247
95	S10_1678	Toys of Finland, Co.	90-224 8555	Keskuskatu 45	Helsinki	<null>	21240
95	S10_1678	Baane Mini Imports	07-98 9555	Erling Skakkes gate 78	Stavern	<null>	4110
95	S10_1678	Diecast Classics Inc.	2155551555	7586 Pompton St.	Allentown	PA	70267
95	S10_1678	Land of Toys Inc.	2125557818	897 Long Airport Avenue	NYC	NY	10022
95	S10_1678	Salzburg Collectables	6562-9555	Geislweg 14	Salzburg	<null>	5020
95	S10_1678	Souveniers And Things Co.	+61 2 9495 8555	Monitor Money Building, 815 Pacific Hwy	Chatswood	NSW	2067
95	S10_1678	La Rochelle Gifts	40.67.8555	67, rue des Cinquante Otages	Nantes	<null>	44000
95	S10_1678	FunGiftIdeas.com	5085552555	1785 First Street	New Bedford	MA	50553

Close Show Log

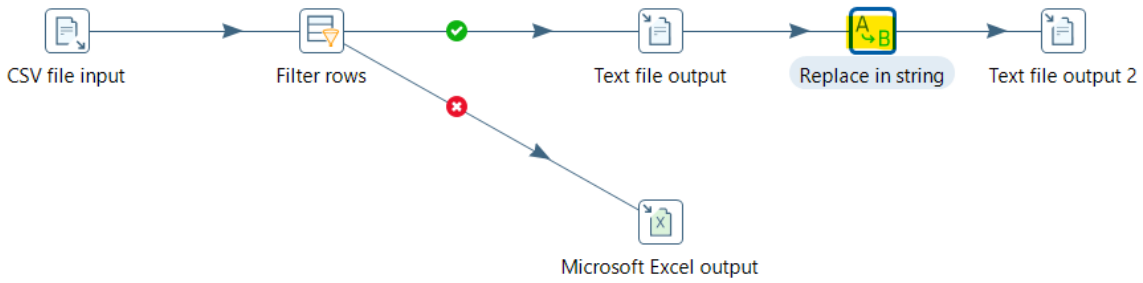
d. To filter out data having Null values in Product, Region & Date Column, we have set the conditions.



e. Set the path and file extension where we want to save the file having correct data.



f. Replace string USA with United states.



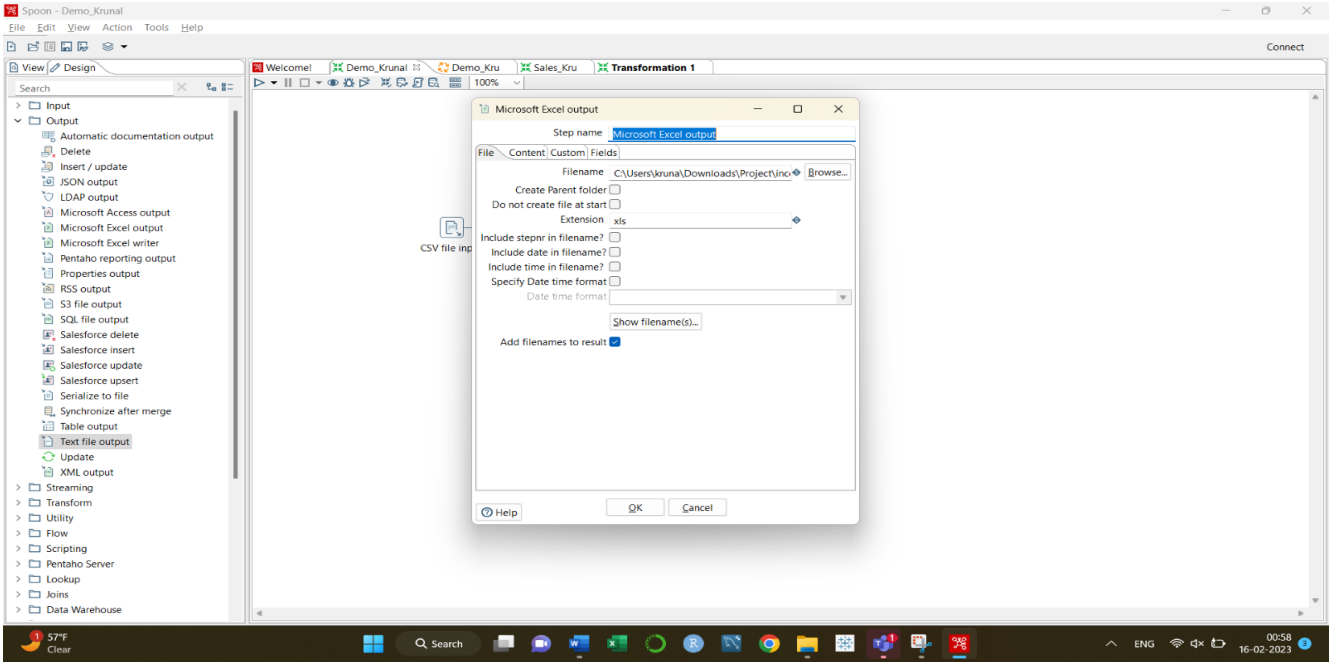
Replace in string

Step name: Replace in string

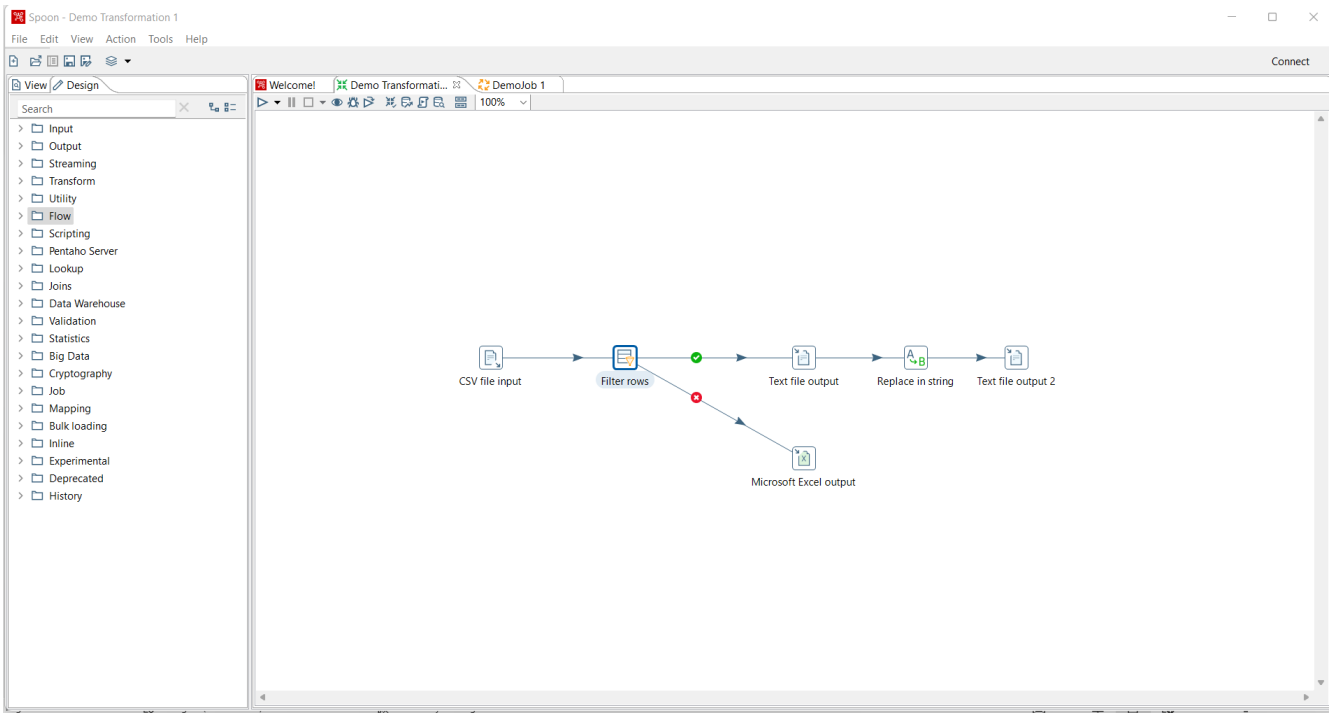
Fields string

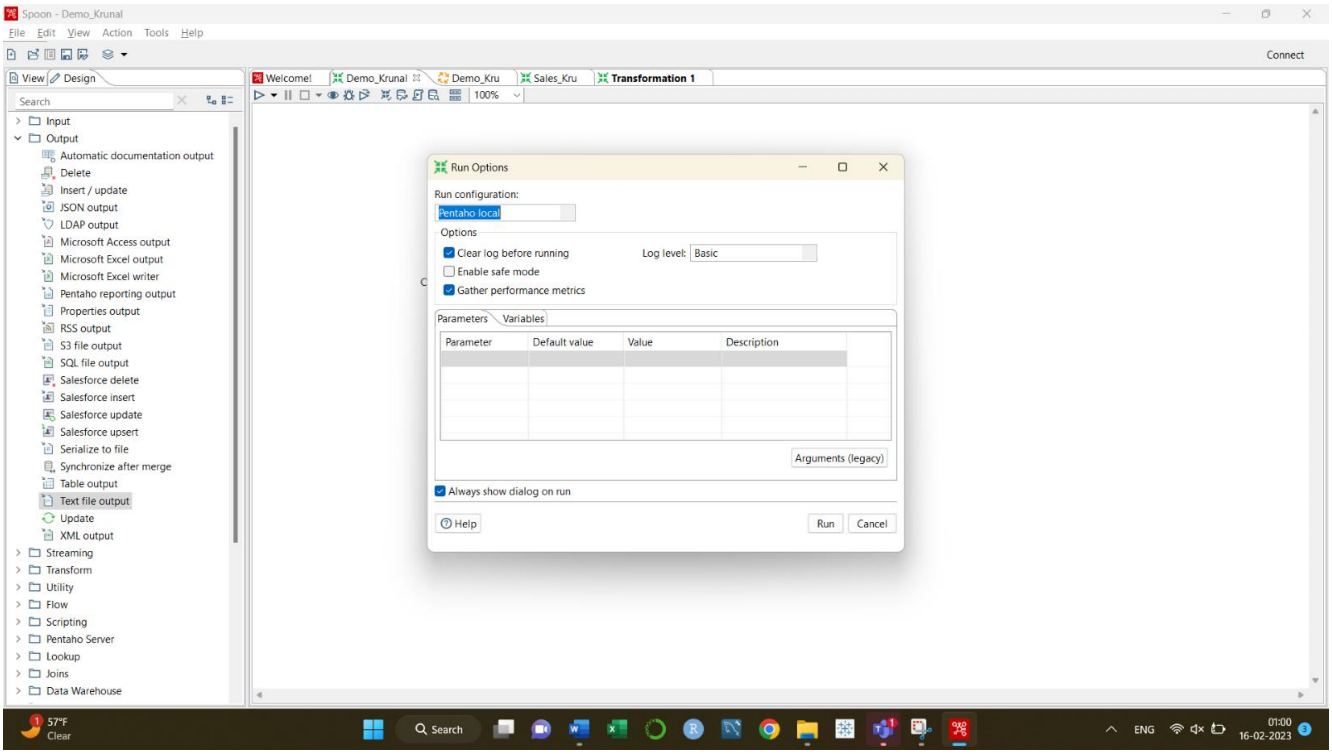
#	In stream field	Out stream field	use RegEx	Search	Replace with	Set empty string?	Replace with field	Whole Word	Case sensitive	Is Unic
1	ORDERDATE		N			N		N	N	N
2	STATUS		N			N		N	N	N
3	PRODUCTLINE		N			N		N	N	N
4	PRODUCTCODE		N			N		N	N	N
5	CUSTOMERNAME		N			N		N	N	N
6	PHONE		N			N		N	N	N
7	ADDRESSLINE1		N			N		N	N	N
8	CITY		N			N		N	N	N
9	STATE		N			N		N	N	N
1..	POSTALCODE		N			N		N	N	N
1..	COUNTRY		N	USA	United States	N		N	N	N
1..	TERRITORY		N			N		N	N	N
1..	CONTACTLASTNAME		N			N		N	N	N
1..	CONTACTFIRSTNAME		N			N		N	N	N

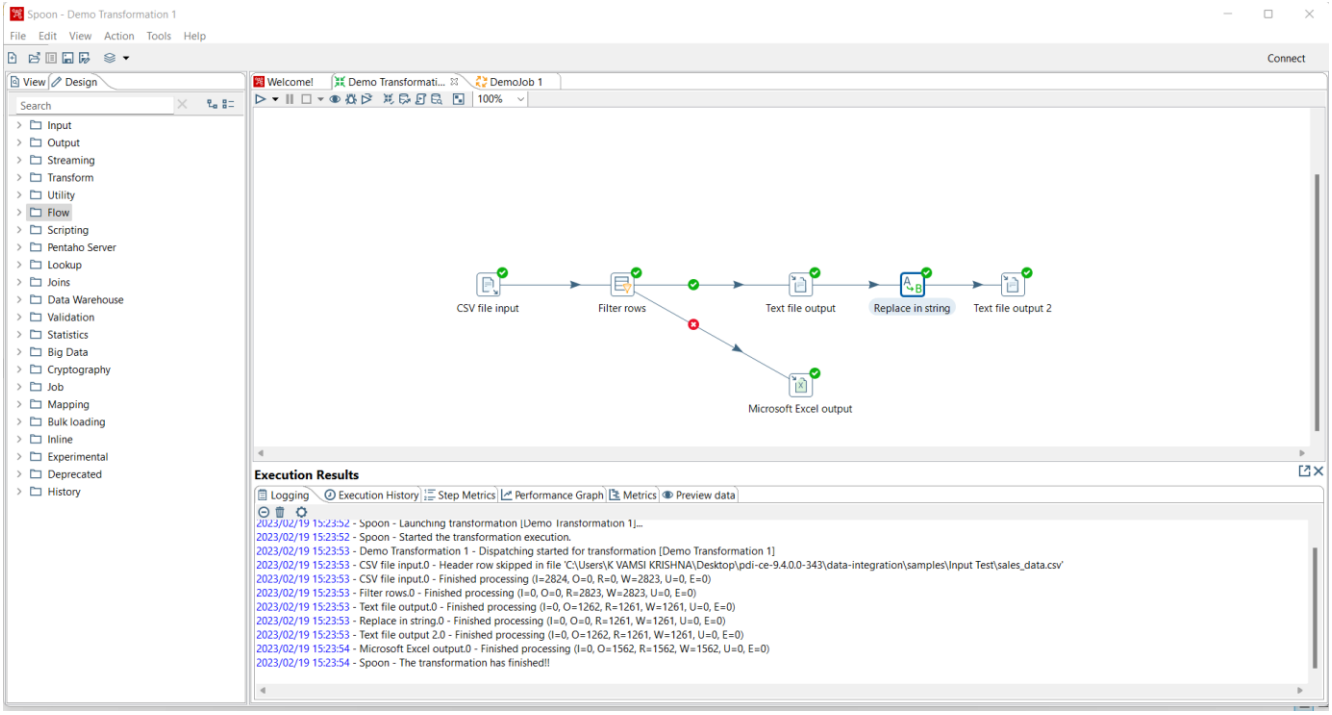
g. Set the path where we want to save the file having data with Null values.



Step-3: **Run the Transformation Process**

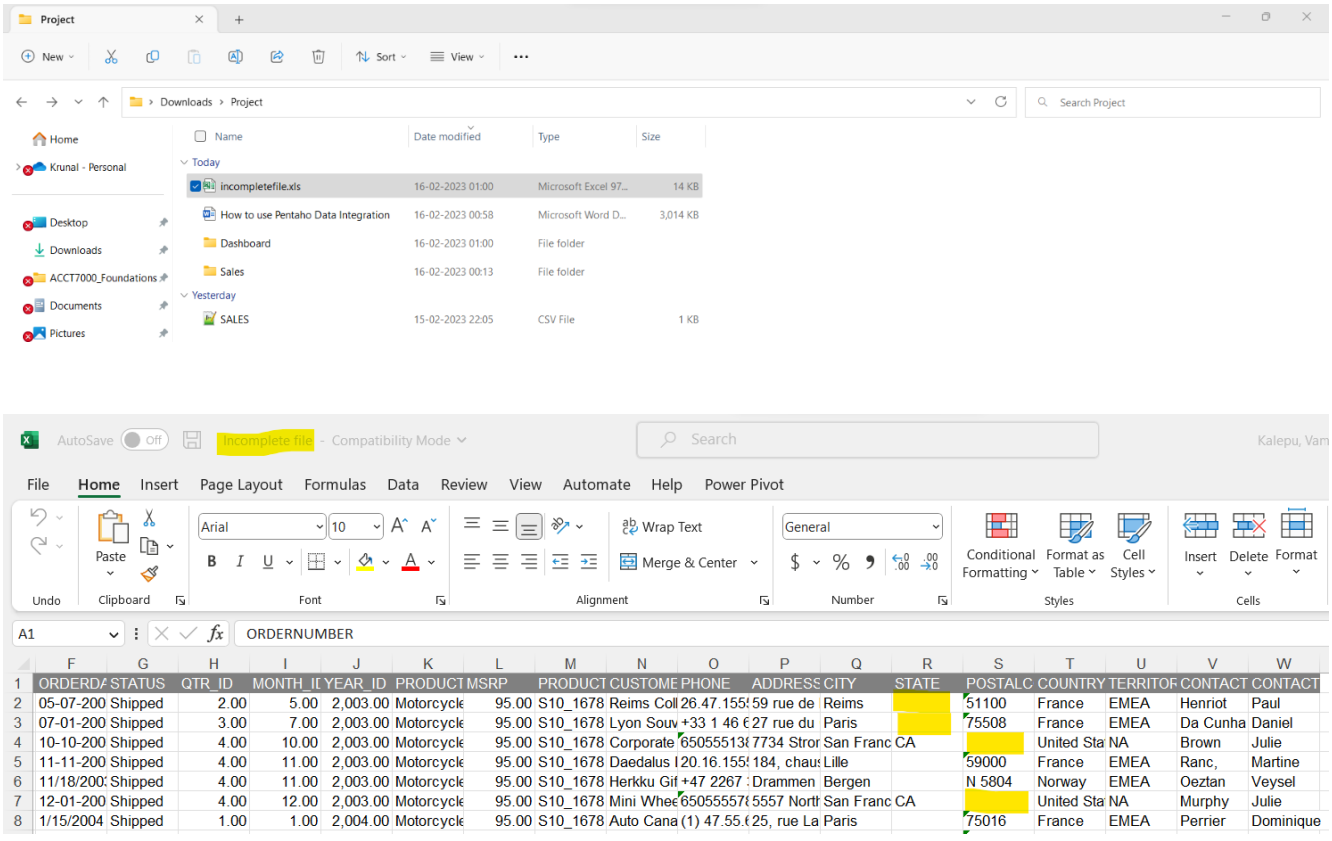




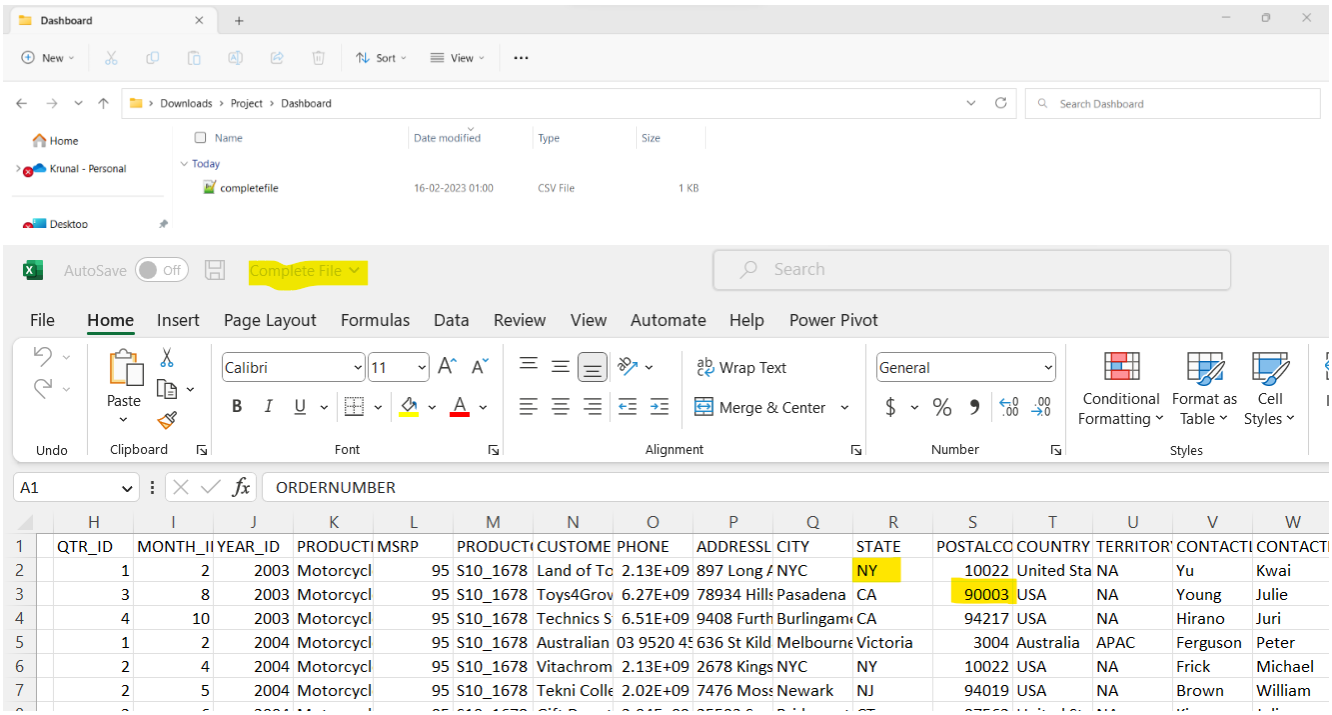


Check that the two files are created on the configured path:

Incomplete file



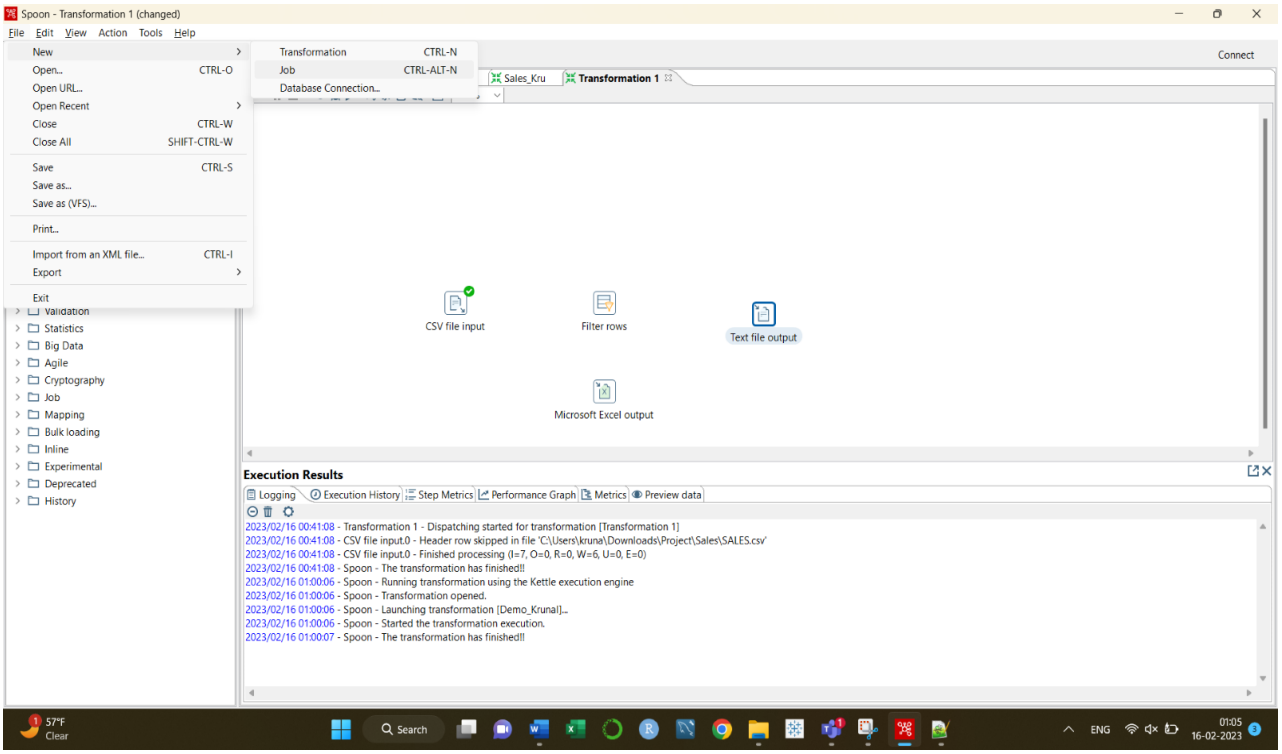
Complete file



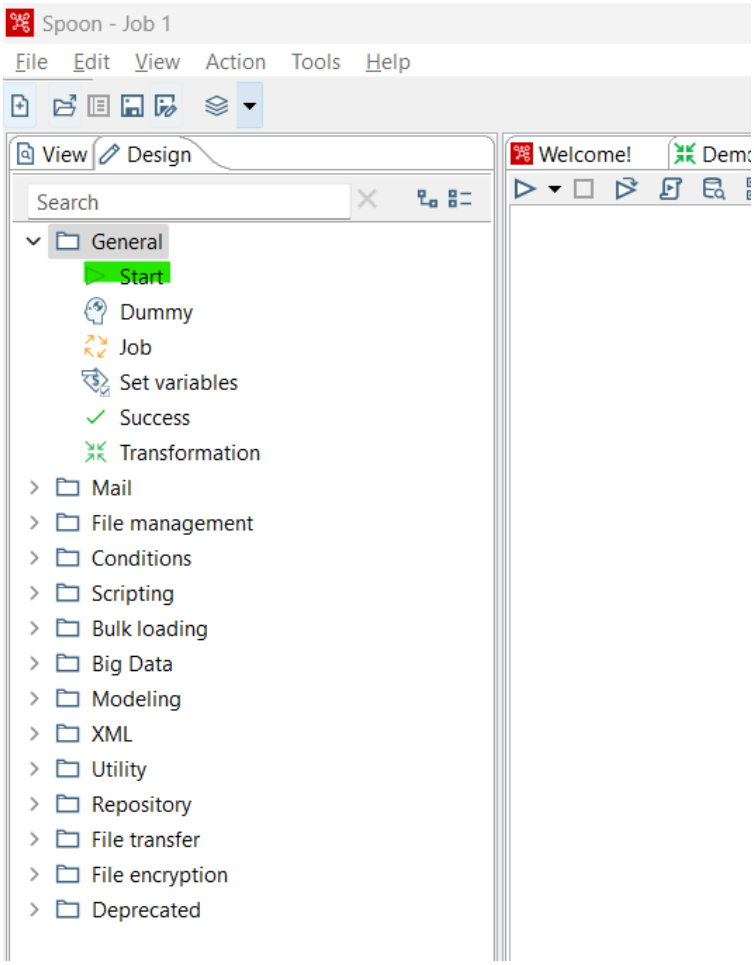
	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	QTR_ID	MONTH_ID	YEAR_ID	PRODUCTI	MSRP	PRODUCT	CUSTOME	PHONE	ADDRESSL	CITY	STATE	POSTALCO	COUNTRY	TERRITOR	CONTACTI	CONTACTI
2	1	2	2003	Motorcycl	95	S10_1678	Land of To	2.13E+09	897 Long A	NYC	NY	10022	United Sta	NA	Yu	Kwai
3	3	8	2003	Motorcycl	95	S10_1678	Toys4Grov	6.27E+09	78934 Hills	Pasadena	CA	90003	USA	NA	Young	Julie
4	4	10	2003	Motorcycl	95	S10_1678	Technics S	6.51E+09	9408 Furth	Burlingame	CA	94217	USA	NA	Hirano	Juri
5	1	2	2004	Motorcycl	95	S10_1678	Australian	03 9520 45	636 St Kild	Melbourne	Victoria	3004	Australia	APAC	Ferguson	Peter
6	2	4	2004	Motorcycl	95	S10_1678	Vitachrom	2.13E+09	2678 Kings	NYC	NY	10022	USA	NA	Frick	Michael
7	2	5	2004	Motorcycl	95	S10_1678	Tekni Colle	2.02E+09	7476 Moss	Newark	NJ	94019	USA	NA	Brown	William

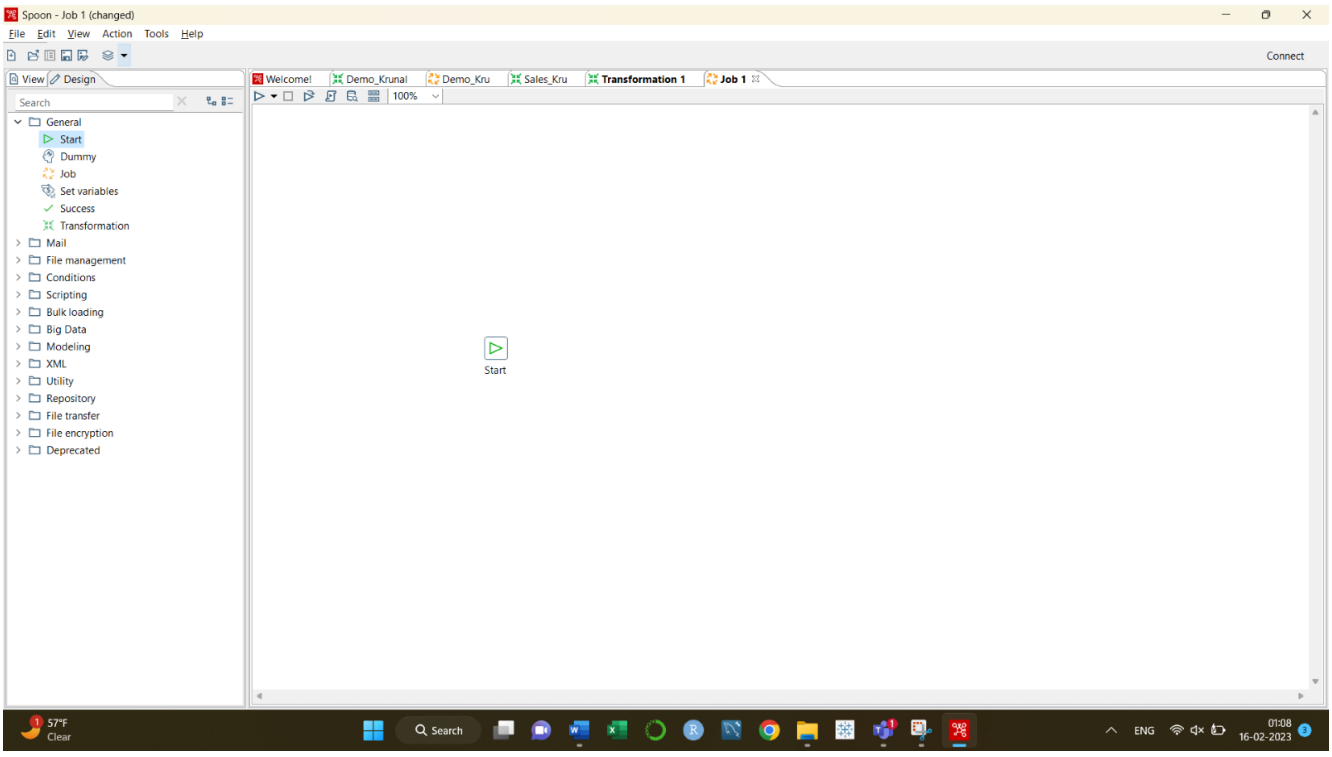
Step-4: **Job Process**

a. Create Job File



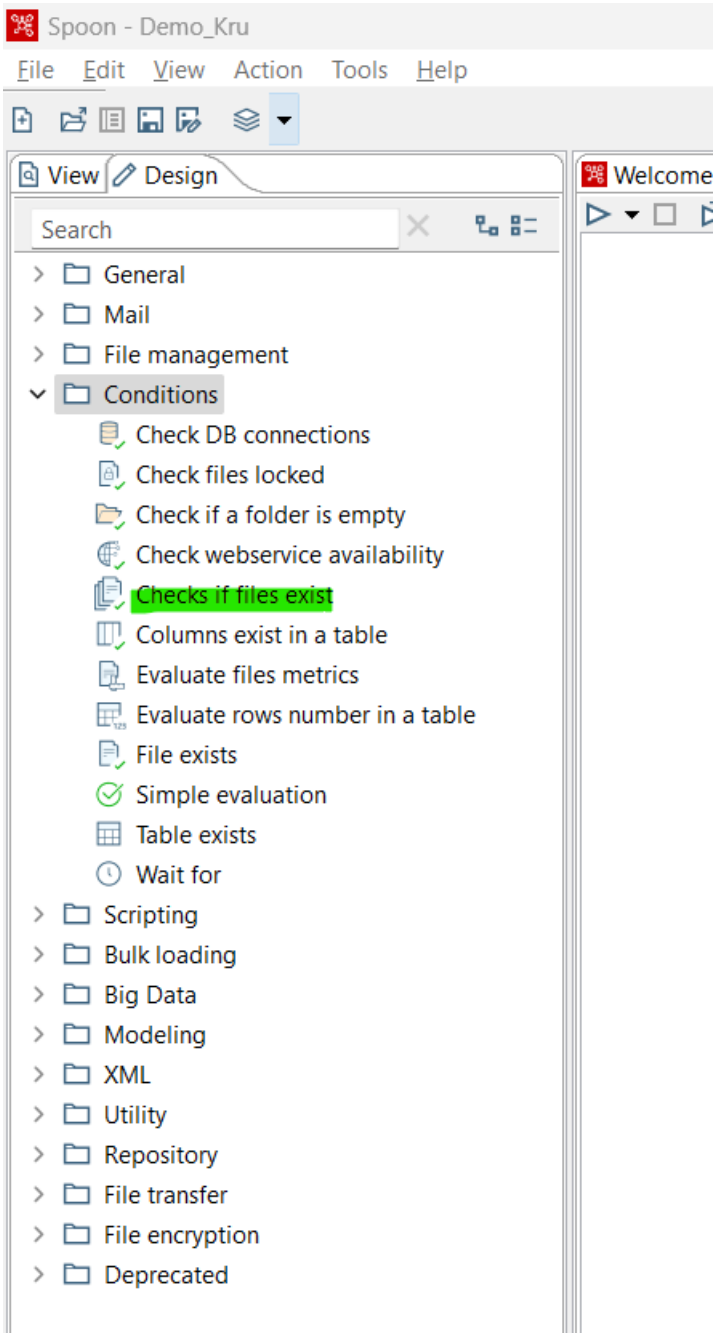
b. Select General >> Start!



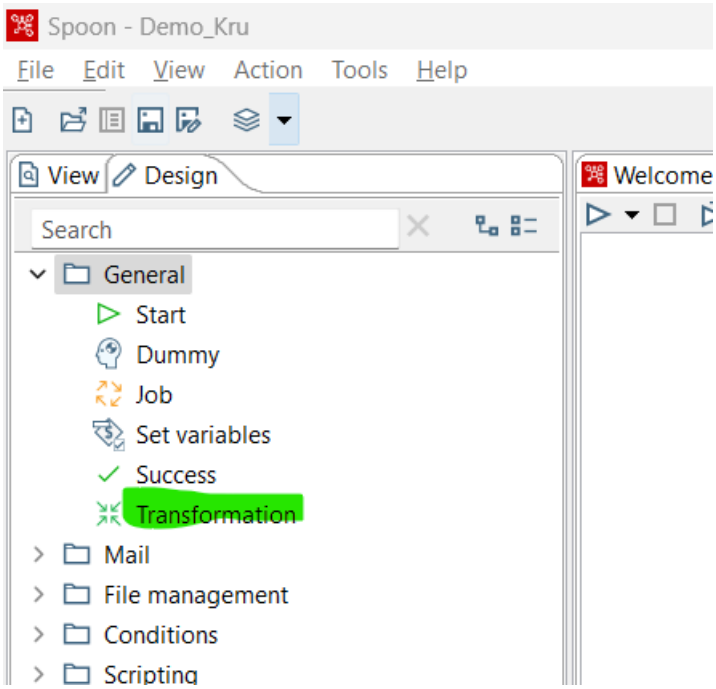


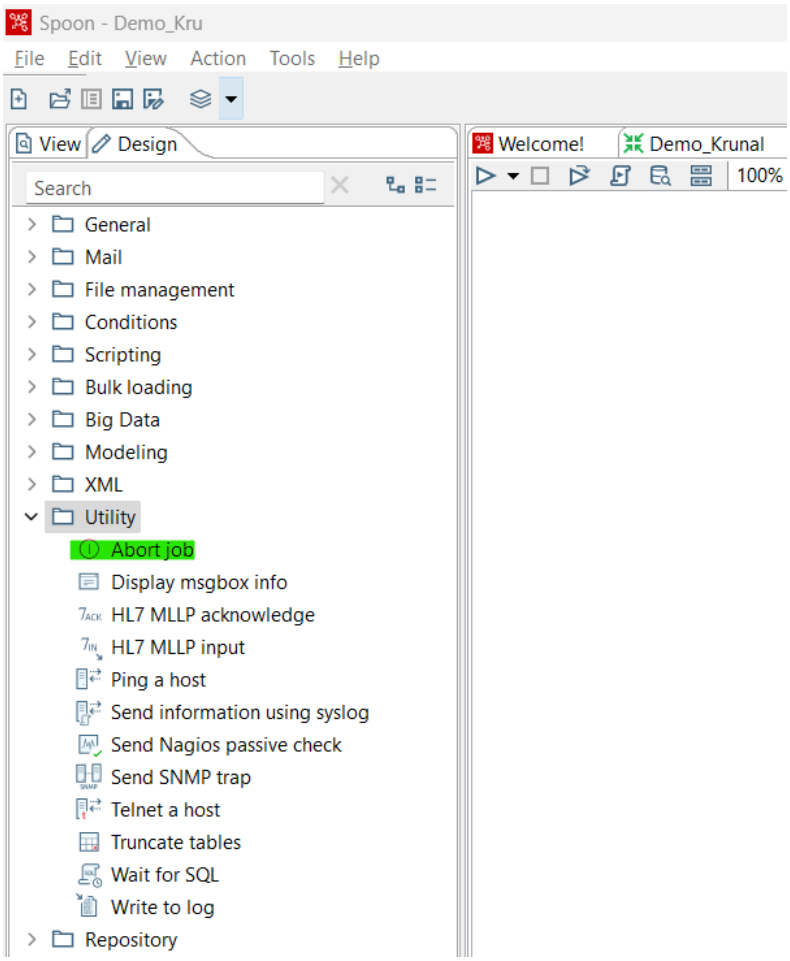
c. We need to check if the Input file exists at the configured path.

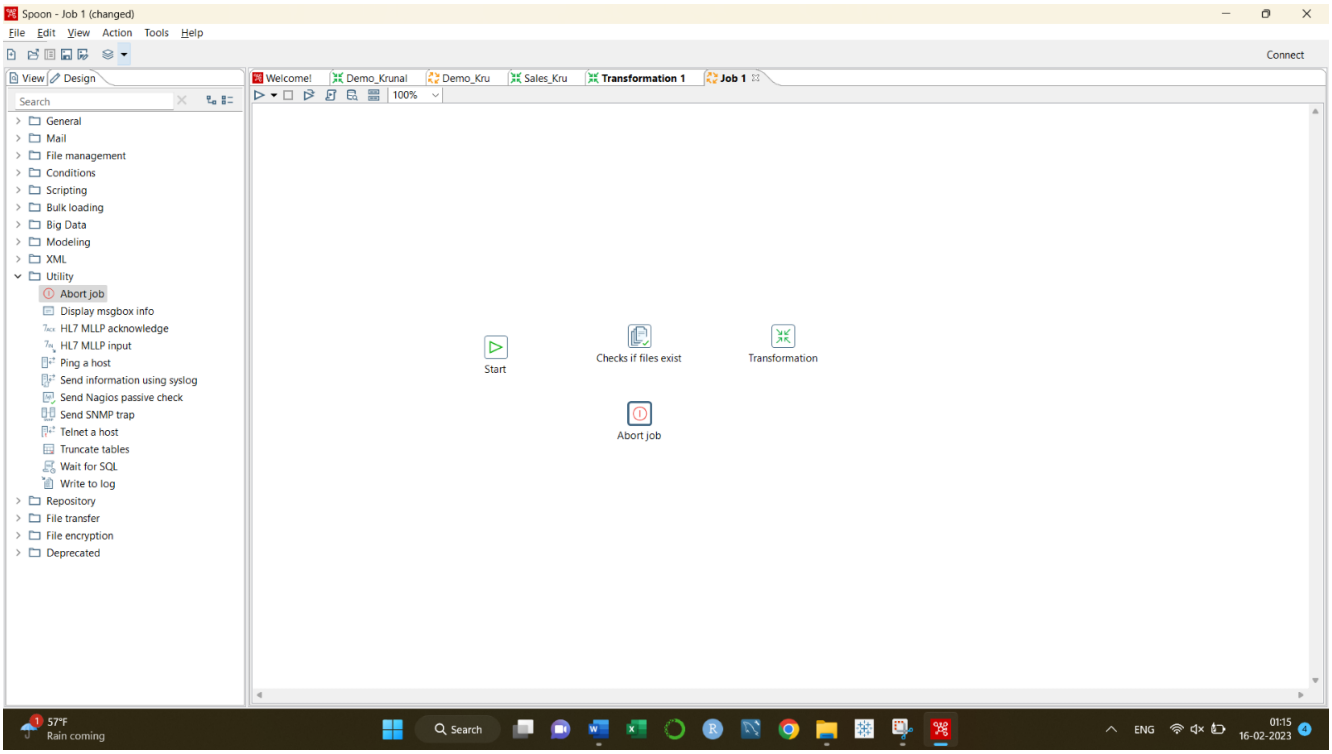
Select Conditions >> Checks if files exist!



d. Based on the availability of file, we can either move ahead with the Transformation Process or Abort the Job:

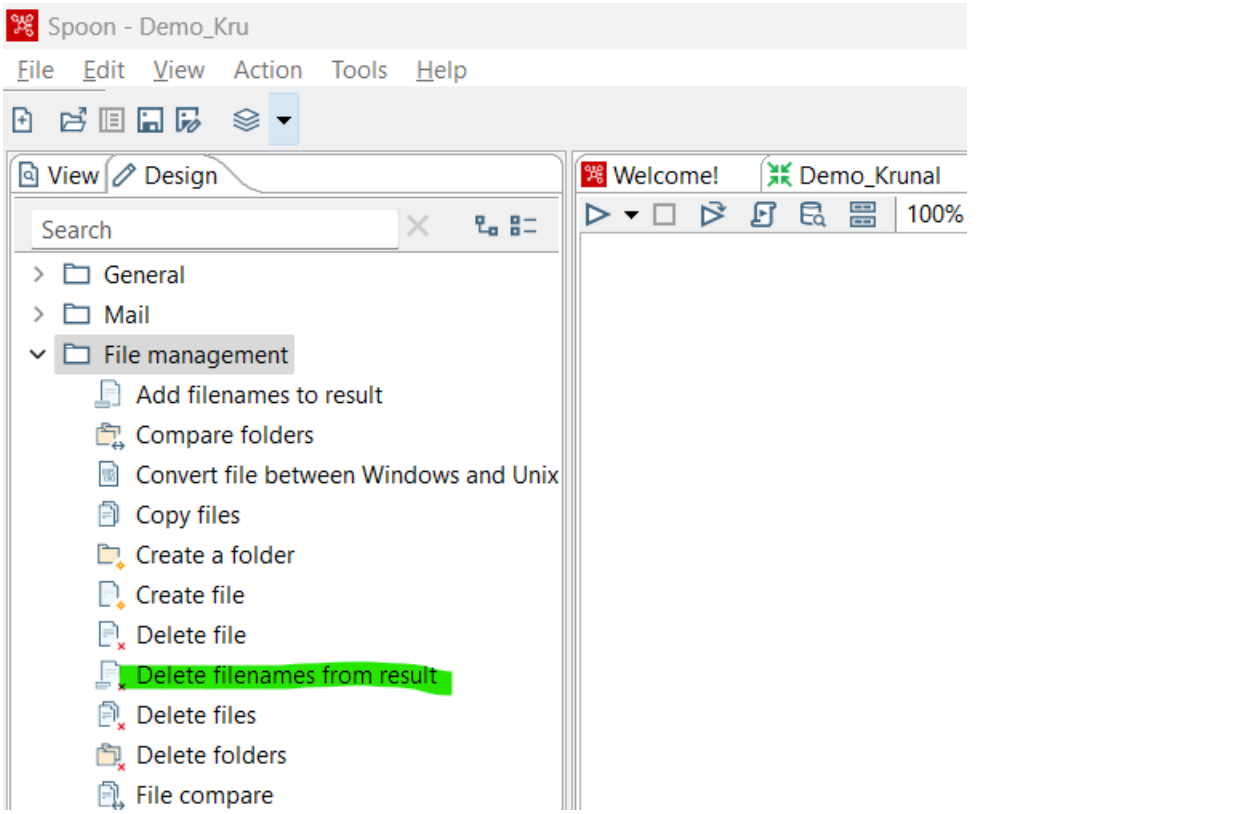




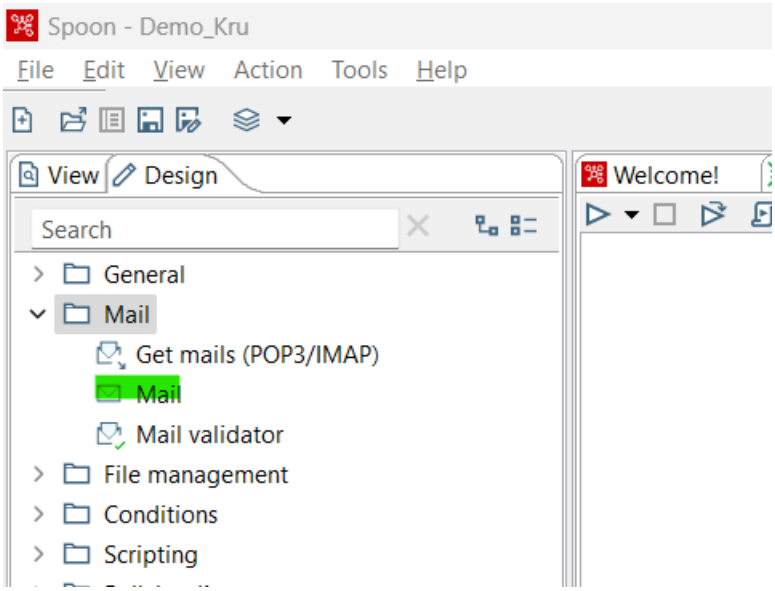


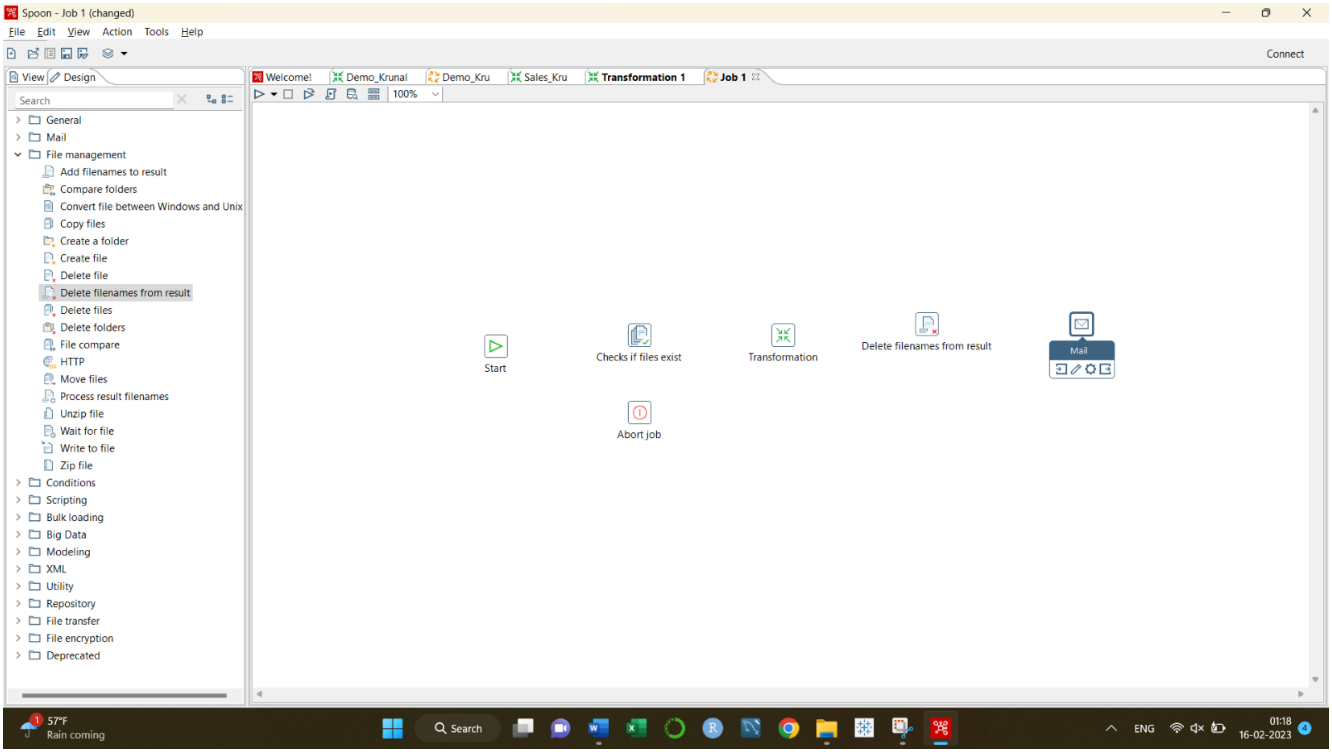
e. We need to mail the Manager only that file which has data with Null values.

This can be configured using File Management >> Delete filenames from result.

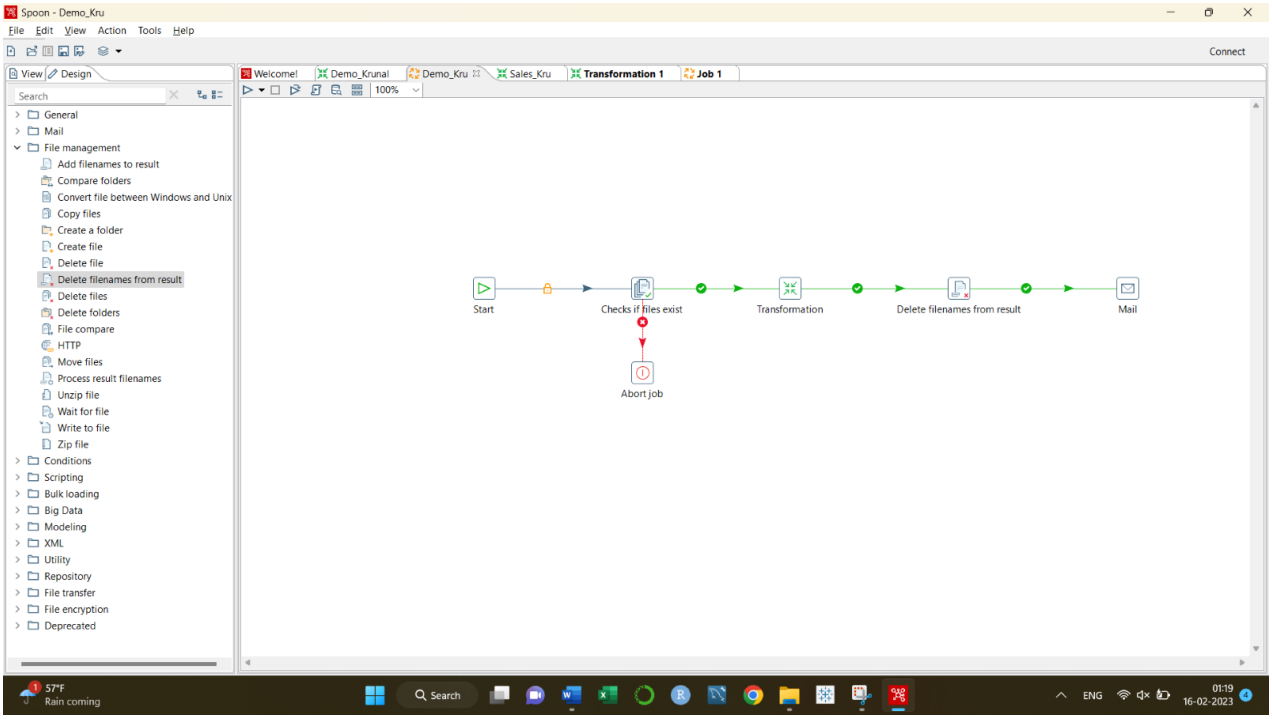


We can configure the Mail job below option:



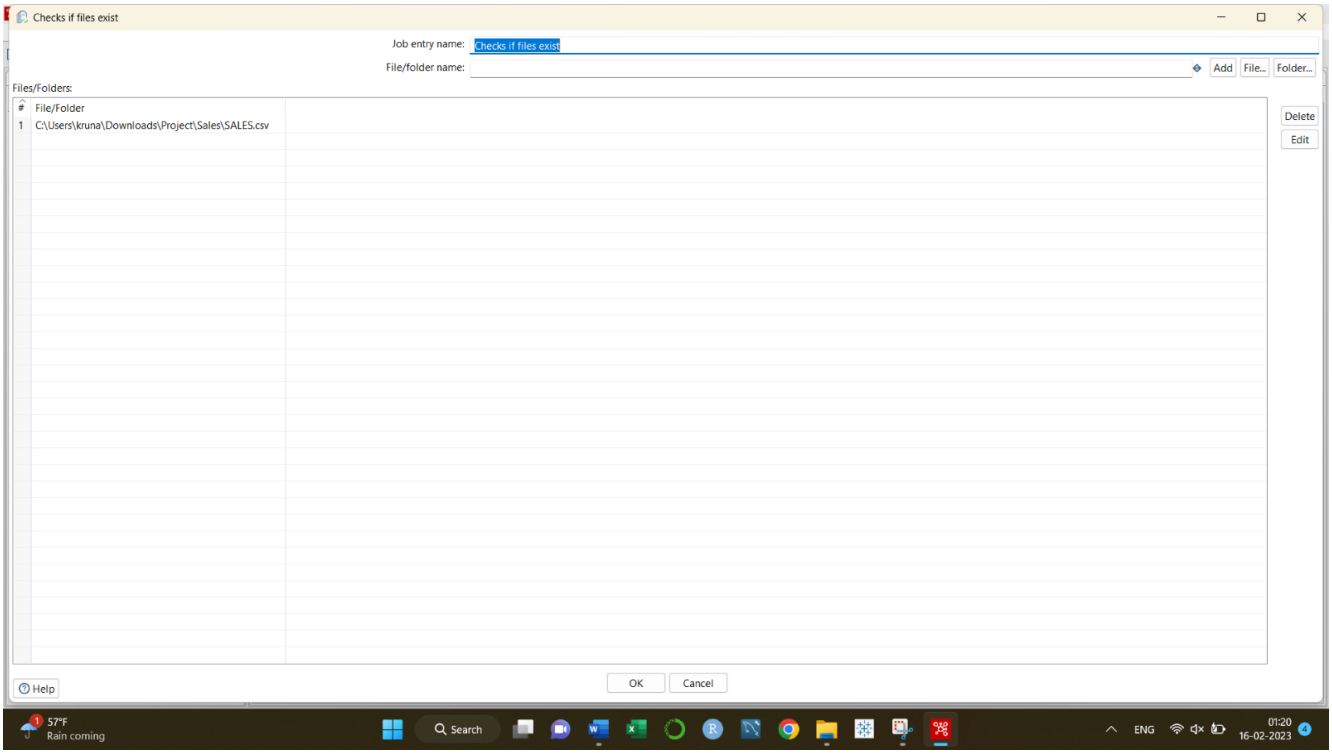


f. We can connect the STEPS using HOPS to set the flow of Job Process

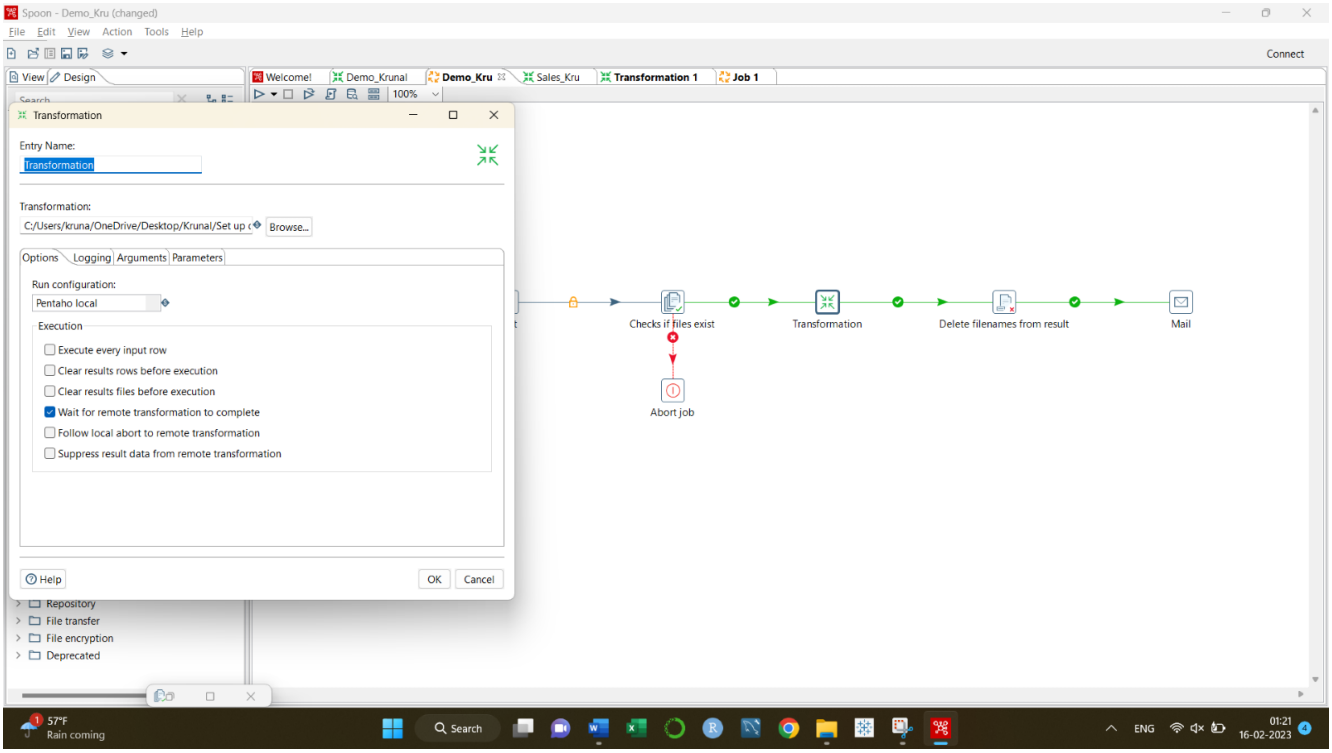


Step-5: **Job Process Configuration**

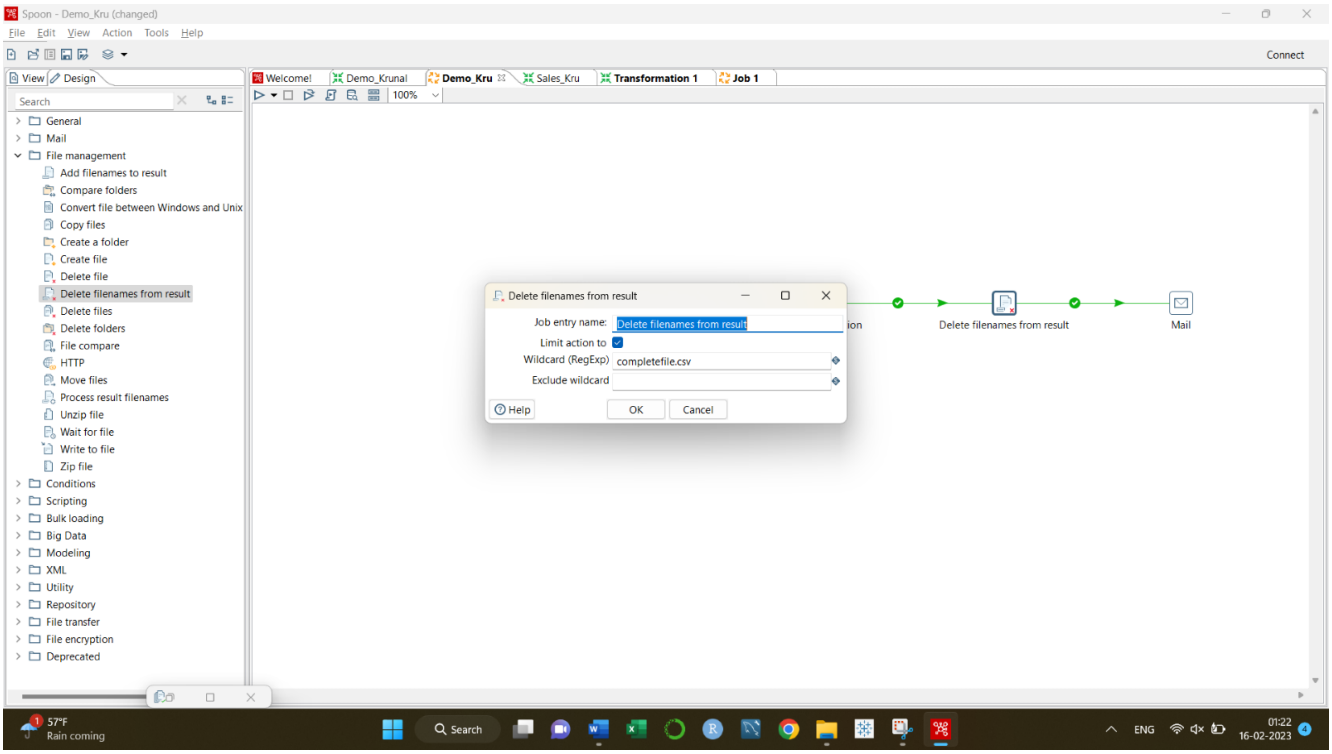
a. Adding the path where Input file will be available in Checks if file exist.



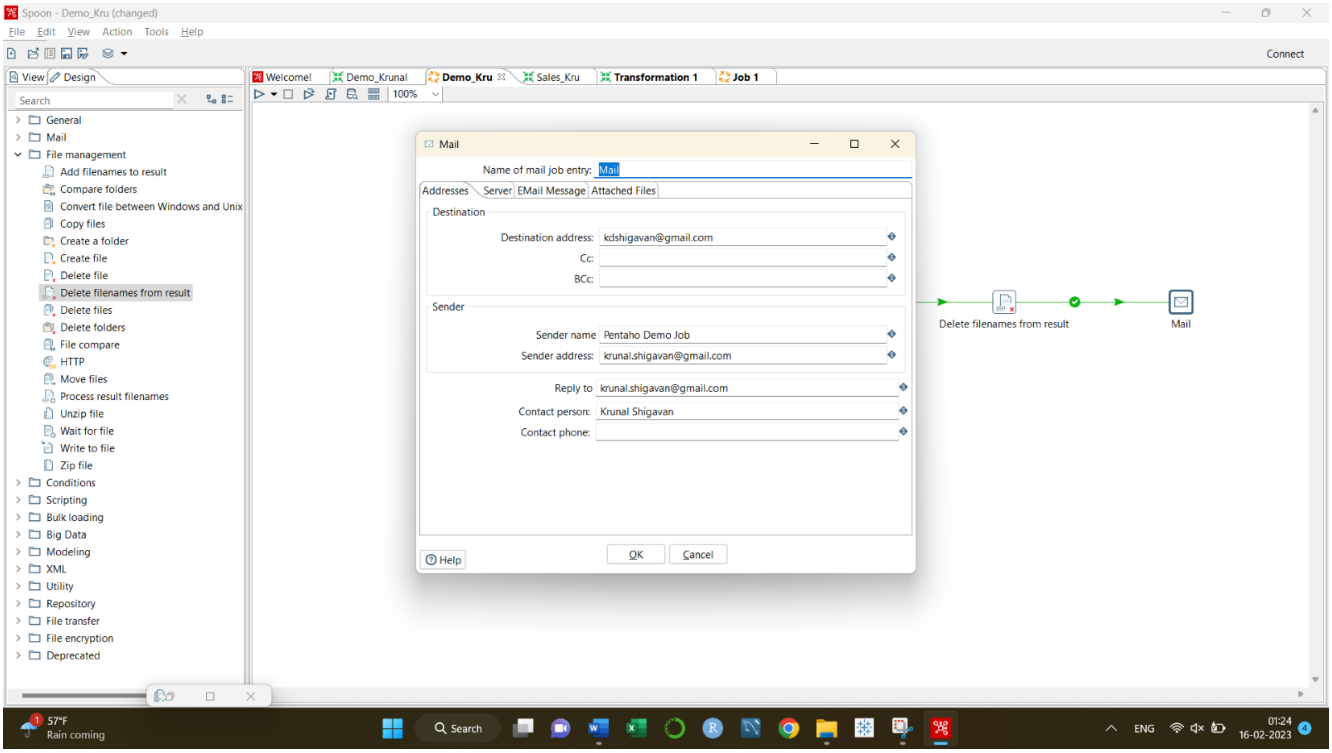
b. Adding the path where the Transformation File is available in Transformation.

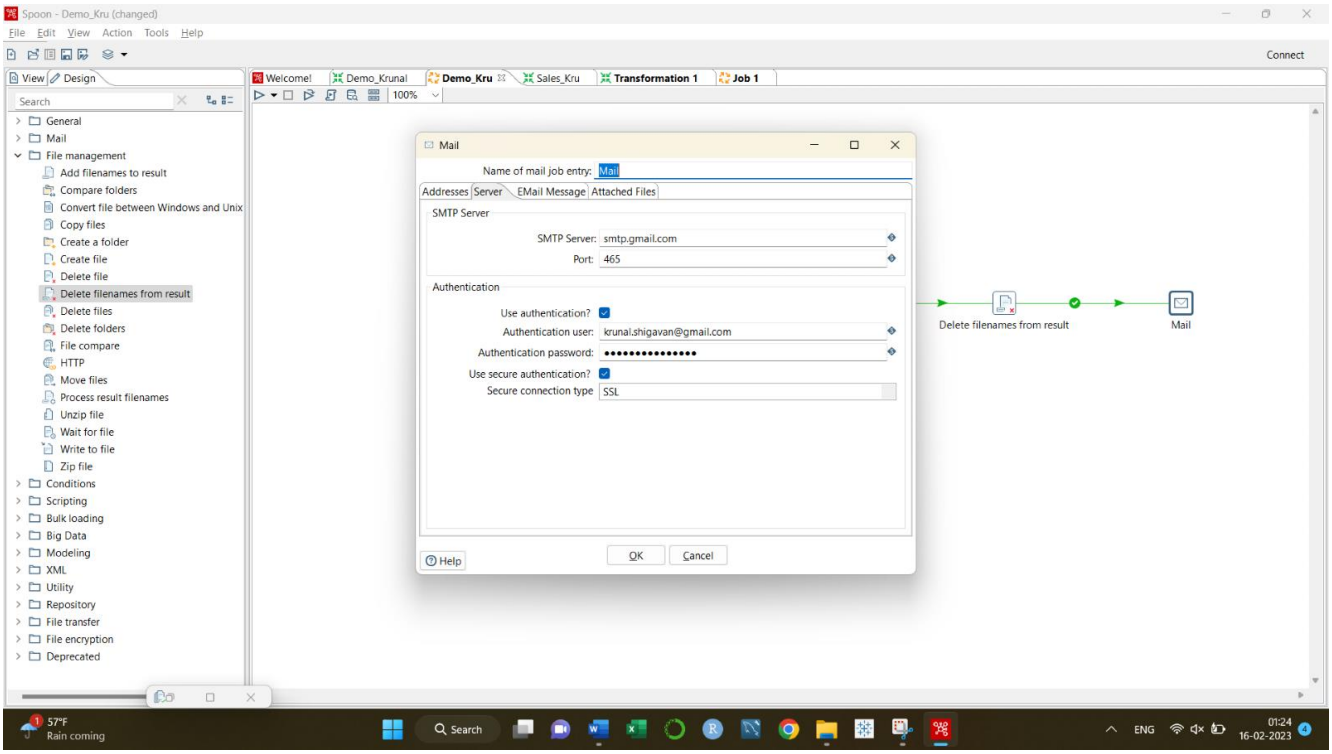


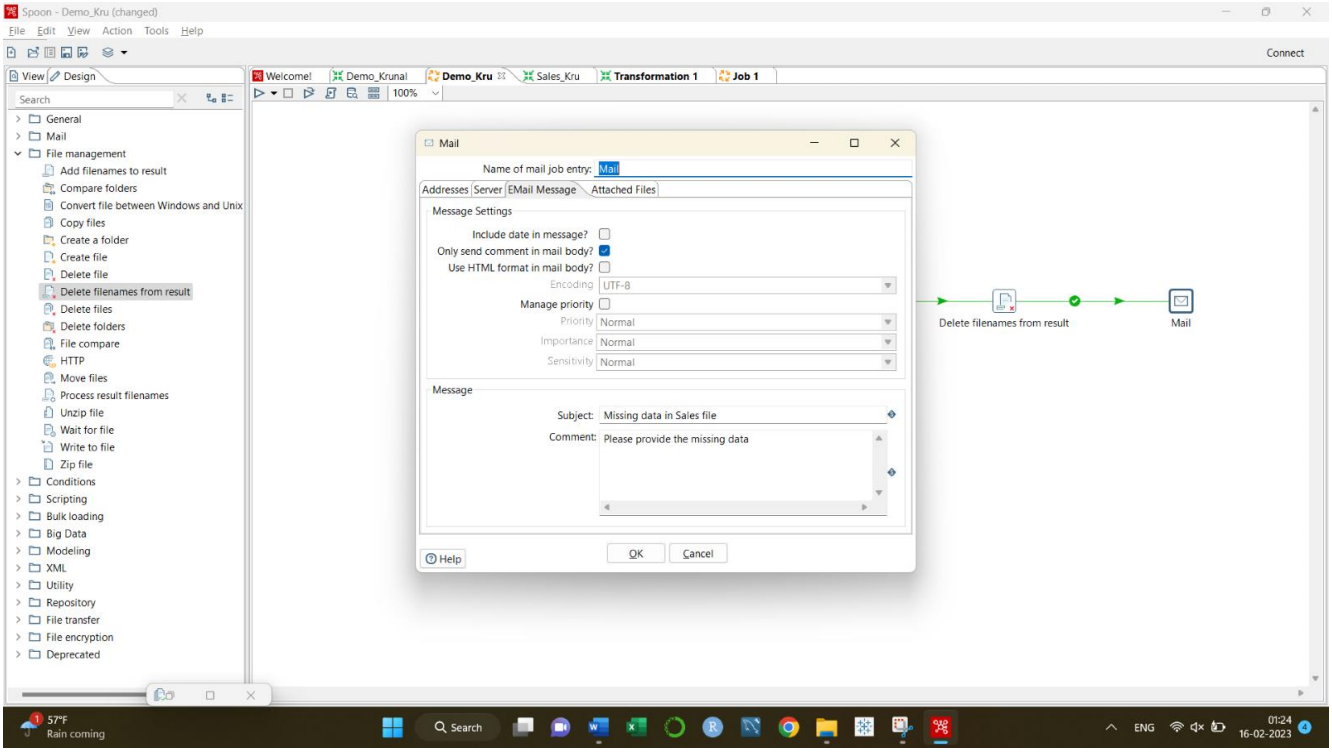
c. We provide the file name which we need to ignore from the result files in Delete filenames from result.

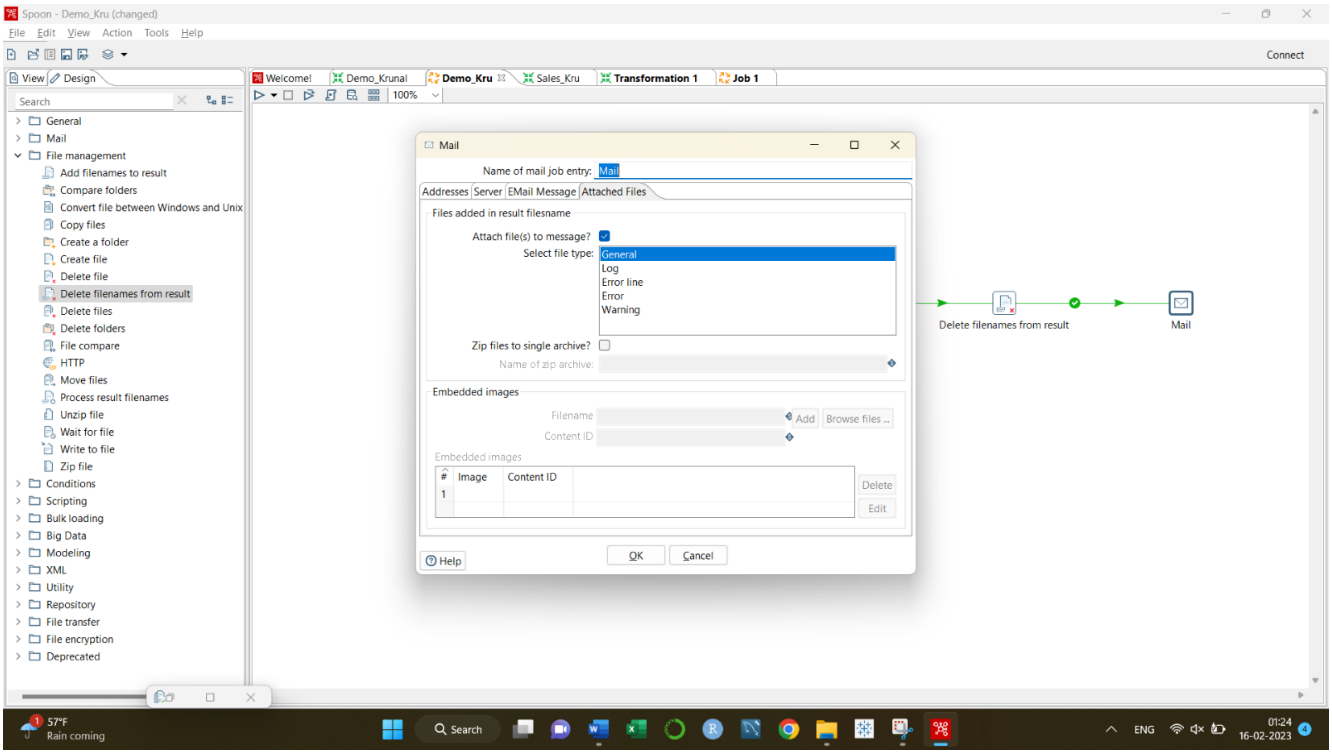


d. Mail Configuration

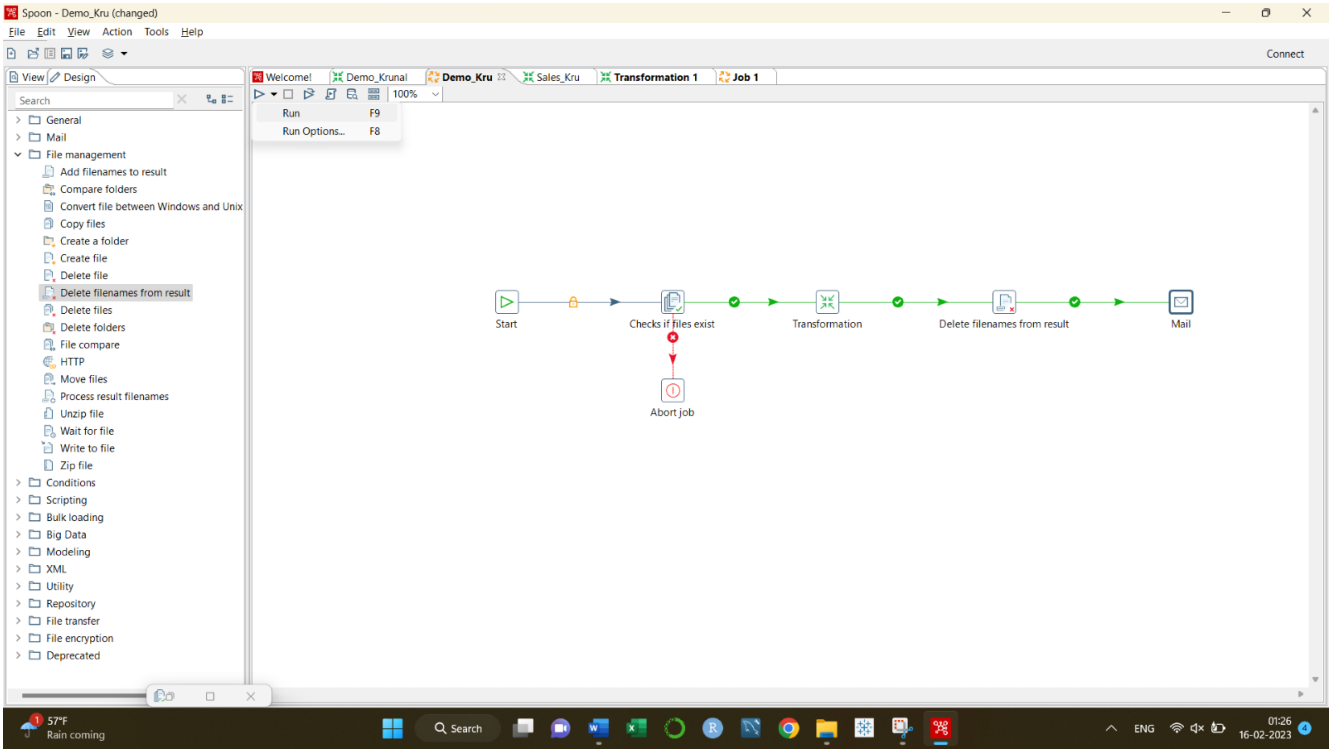


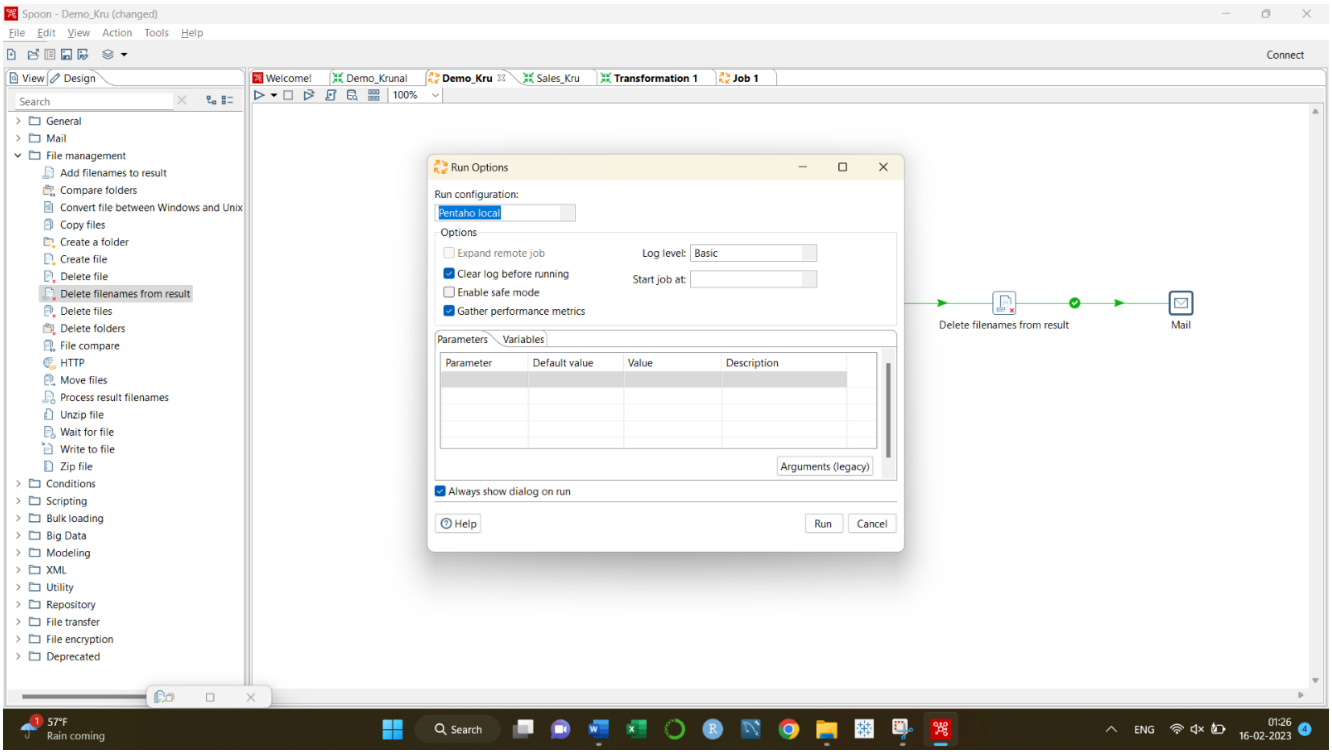


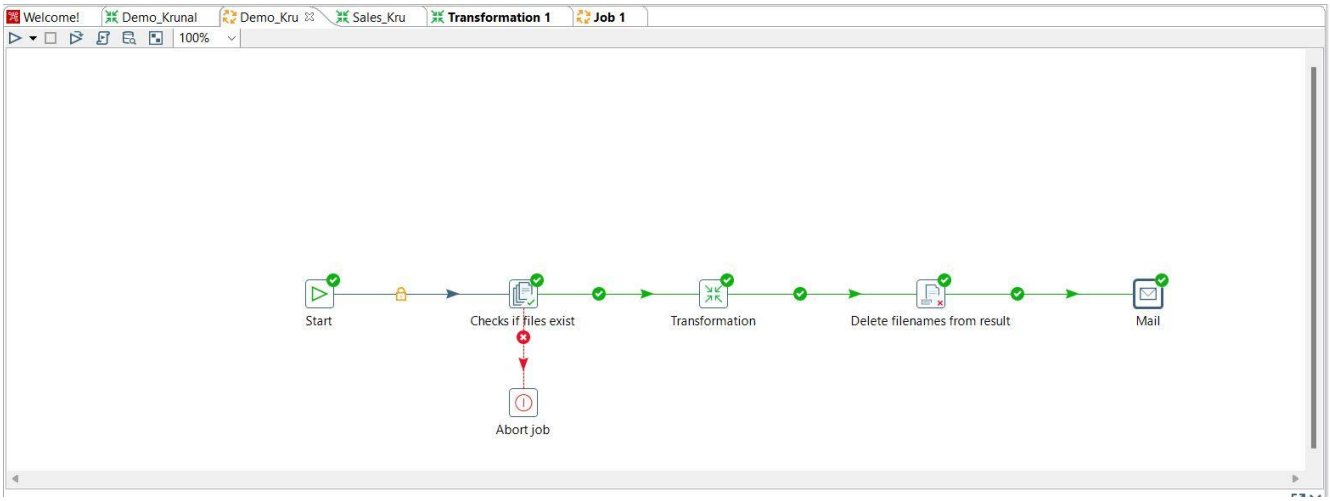

Spoon - Demo_Kru (changed)
File Edit View Action Tools Help
Connect
View Design
Search
General
Mail
File management
Add filenames to result
Compare folders
Convert file between Windows and Unix
Copy files
Create a folder
Create file
Delete file
Delete filenames from result
Delete files
Delete folders
File compare
HTTP
Move files
Process result filenames
Unzip file
Wait for file
Write to file
Zip file
Conditions
Scripting
Bulk loading
Big Data
Modeling
XML
Utility
Repository
File transfer
File encryption
Deprecated
Welcome! Demo_Krunal Demo_Kru Sales_Kru Transformation 1 Job 1
100%
Mail
Name of mail job entry: Mail
Addresses Server EMail Message Attached Files
Message Settings
Include date in message?
Only send comment in mail body?
Use HTML format in mail body?
Encoding UTF-8
Manage priority
Priority Normal
Importance Normal
Sensitivity Normal
Message
Subject: Missing data in Sales file
Comment: Please provide the missing data
Help
OK Cancel
Delete filenames from result
Mail
57°F Rain coming
Search
ENG
01:24 16-02-2023


Spoon - Demo_Kru (changed)
File Edit View Action Tools Help
Connect
View Design
Search
General
Mail
File management
Add filenames to result
Compare folders
Convert file between Windows and Unix
Copy files
Create a folder
Create file
Delete file
Delete filenames from result
Delete files
Delete folders
File compare
HTTP
Move files
Process result filenames
Unzip file
Wait for file
Write to file
Zip file
Conditions
Scripting
Bulk loading
Big Data
Modeling
XML
Utility
Repository
File transfer
File encryption
Deprecated
Welcome! Demo_Krunal Demo_Kru Sales_Kru Transformation 1 Job 1
100%
Mail
Name of mail job entry: Mail
Addresses Server EMail Message Attached Files
Files added in result filesname
Attach file(s) to message?
Select file type: General
Log
Error line
Error
Warning
Zip files to single archive?
Name of zip archive:
Embedded images
Filename
Add
Browse files ...
Content ID
Embedded images
Image Content ID
1
Delete
Edit
Help
OK Cancel
Delete filenames from result
Mail
57°F Rain coming
Search
ENG
01:24 16-02-2023

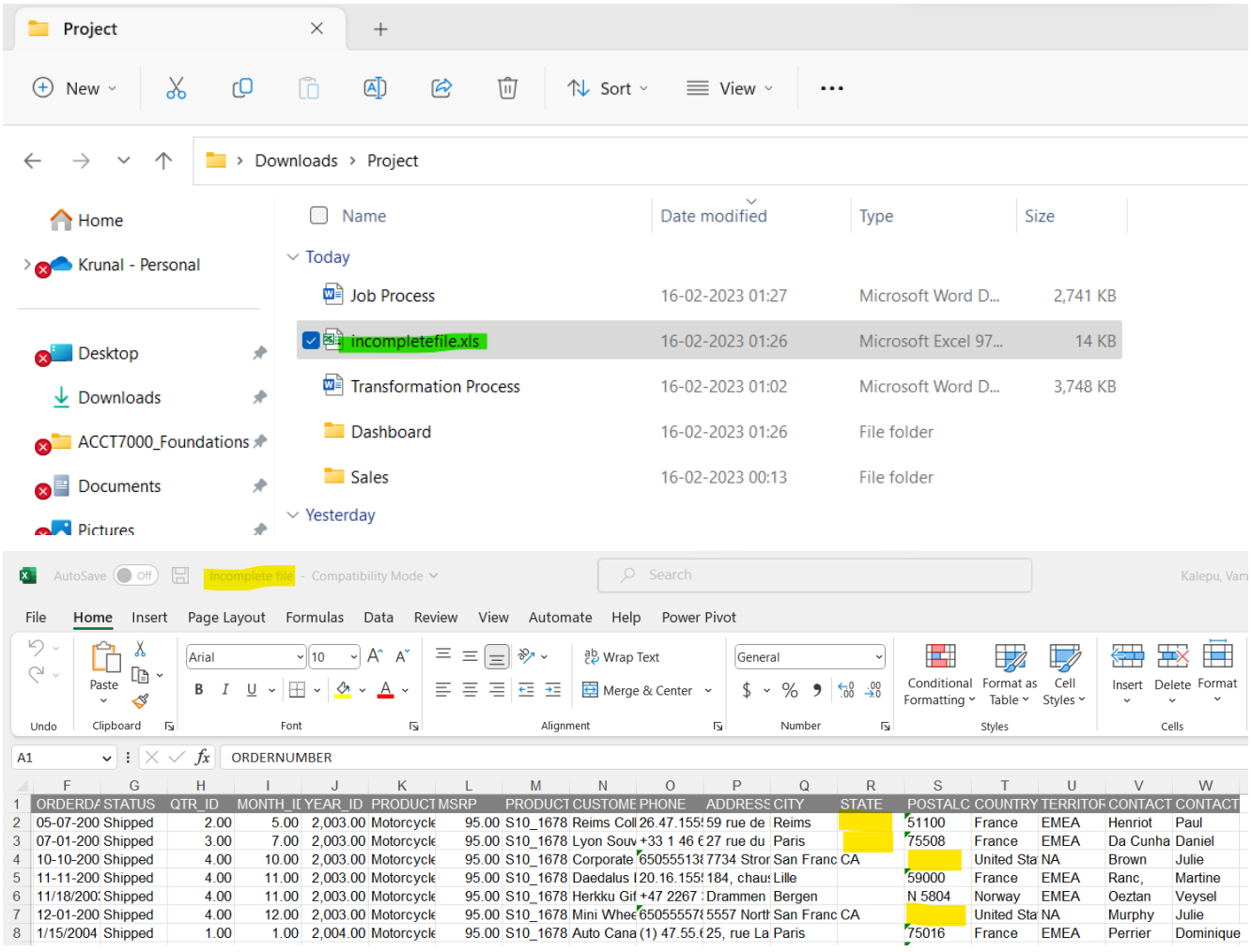
Step-6: **Run the Job Process**



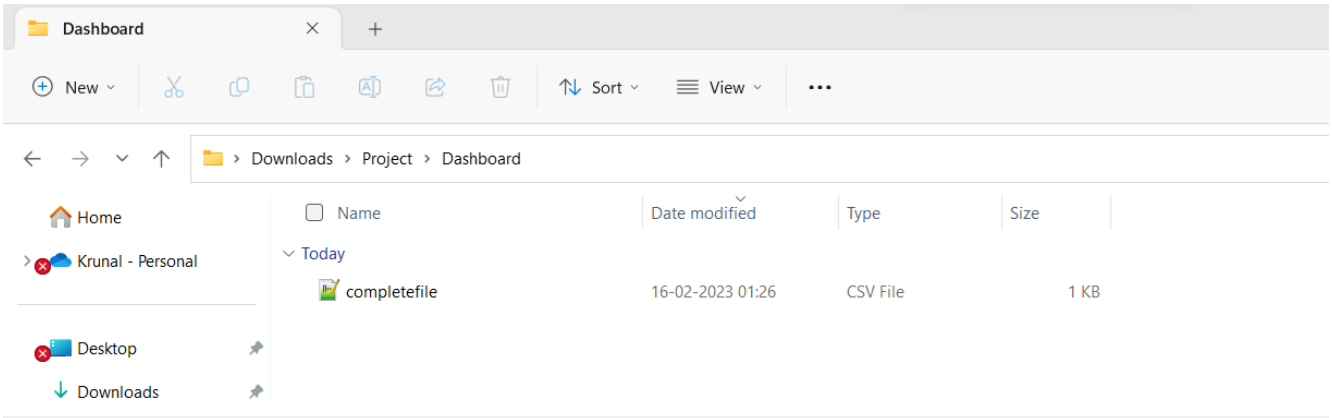


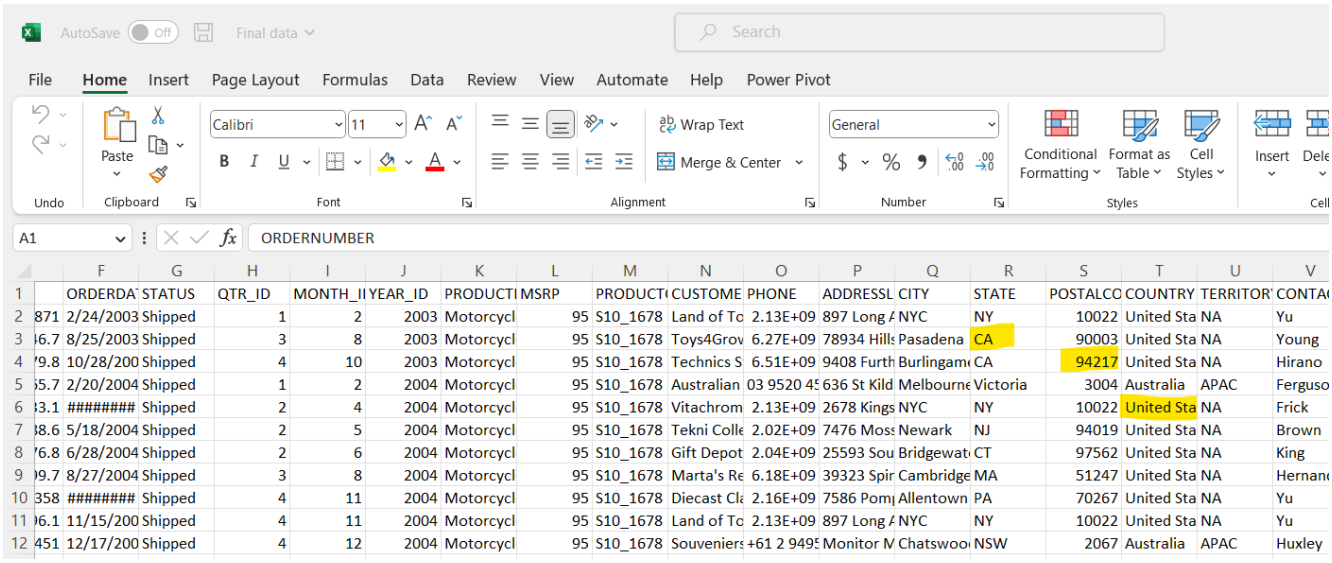


Files are created on the configured path:



	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	ORDERDA	STATUS	QTR_ID	MONTH_ID	YEAR_ID	PRODUCT	MSRP	PRODUCT	CUSTOME	PHONE	ADDRESS	CITY	STATE	POSTALC	COUNTRY	TERRITOR	CONTACT	CONTACT
2	05-07-200	Shipped	2.00	5.00	2,003.00	Motorcycle	95.00	S10_1678	Reims Coll	26.47.155	59 rue de	Reims		51100	France	EMEA	Henriot	Paul
3	07-01-200	Shipped	3.00	7.00	2,003.00	Motorcycle	95.00	S10_1678	Lyon Souv	+33 1 46 6	27 rue du	Paris		75508	France	EMEA	Da Cunha	Daniel
4	10-10-200	Shipped	4.00	10.00	2,003.00	Motorcycle	95.00	S10_1678	Corporate	650555136	7734 Stror	San Franc	CA		United Sta	NA	Brown	Julie
5	11-11-200	Shipped	4.00	11.00	2,003.00	Motorcycle	95.00	S10_1678	Daedalus I	20.16.155	184, chau	Lille		59000	France	EMEA	Ranc,	Martine
6	11/18/200	Shipped	4.00	11.00	2,003.00	Motorcycle	95.00	S10_1678	Herkku Gif	+47 2267	Drammen	Bergen		N 5804	Norway	EMEA	Oeztan	Veysel
7	12-01-200	Shipped	4.00	12.00	2,003.00	Motorcycle	95.00	S10_1678	Mini Whee	650555578	5557 North	San Franc	CA		United Sta	NA	Murphy	Julie
8	1/15/2004	Shipped	1.00	1.00	2,004.00	Motorcycle	95.00	S10_1678	Auto Cana	(1) 47.55.6	25, rue La	Paris		75016	France	EMEA	Perrier	Dominique





	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1		ORDERDA	STATUS	QTR_ID	MONTH_I	YEAR_ID	PRODUCTI	MSRP	PRODUCT	CUSTOME	PHONE	ADDRESSL	CITY	STATE	POSTALCO	COUNTRY	TERRITOR	CONTAC
2	871	2/24/2003	Shipped	1	2	2003	Motorcycl	95	S10_1678	Land of To	2.13E+09	897 Long A	NYC	NY	10022	United Sta	NA	Yu
3	6.7	8/25/2003	Shipped	3	8	2003	Motorcycl	95	S10_1678	Toys4Grov	6.27E+09	78934 Hills	Pasadena	CA	90003	United Sta	NA	Young
4	79.8	10/28/200	Shipped	4	10	2003	Motorcycl	95	S10_1678	Technics S	6.51E+09	9408 Furth	Burlingam	CA	94217	United Sta	NA	Hirano
5	5.7	2/20/2004	Shipped	1	2	2004	Motorcycl	95	S10_1678	Australian	03 9520 45	636 St Kild	Melbourne	Victoria	3004	Australia	APAC	Ferguso
6	3.1	########	Shipped	2	4	2004	Motorcycl	95	S10_1678	Vitachrom	2.13E+09	2678 Kings	NYC	NY	10022	United Sta	NA	Frick
7	8.6	5/18/2004	Shipped	2	5	2004	Motorcycl	95	S10_1678	Tekni Colle	2.02E+09	7476 Moss	Newark	NJ	94019	United Sta	NA	Brown
8	6.8	6/28/2004	Shipped	2	6	2004	Motorcycl	95	S10_1678	Gift Depot	2.04E+09	25593 Sou	Bridgewat	CT	97562	United Sta	NA	King
9	9.7	8/27/2004	Shipped	3	8	2004	Motorcycl	95	S10_1678	Marta's Re	6.18E+09	39323 Spir	Cambridge	MA	51247	United Sta	NA	Hernan
10	358	########	Shipped	4	11	2004	Motorcycl	95	S10_1678	Diecast Cla	2.16E+09	7586 Pom	Allentown	PA	70267	United Sta	NA	Yu
11	6.1	11/15/200	Shipped	4	11	2004	Motorcycl	95	S10_1678	Land of To	2.13E+09	897 Long A	NYC	NY	10022	United Sta	NA	Yu
12	451	12/17/200	Shipped	4	12	2004	Motorcycl	95	S10_1678	Souveniers	+61 2 9495	Monitor M	Chatswoo	NSW	2067	Australia	APAC	Huxley

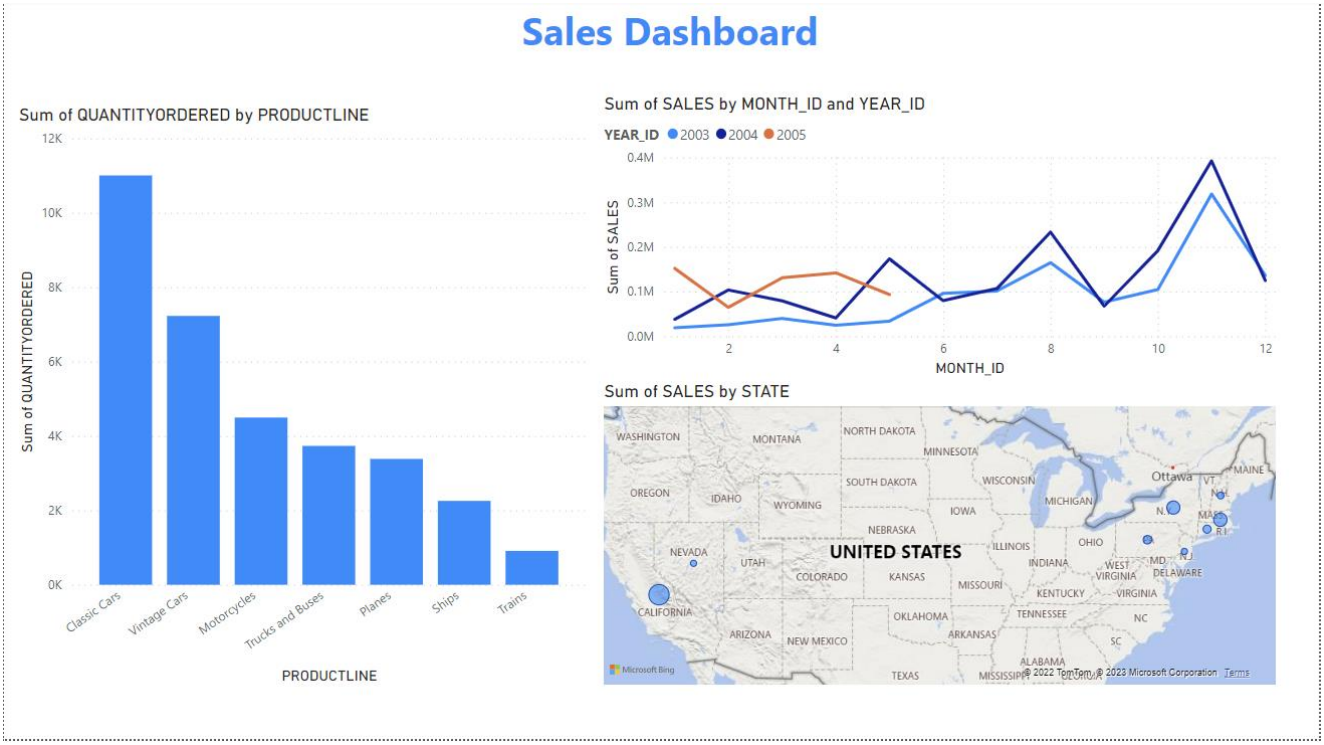
APPENDIX B: MS POWER BI

Power BI is a business analytics service by Microsoft that provides interactive visualizations and business intelligence capabilities with an interface simple enough for end-users to create their own reports and dashboards. It allows users to connect to a wide range of data sources, transform and shape the data, and create visualizations and reports that can be shared with others.

Use Case:

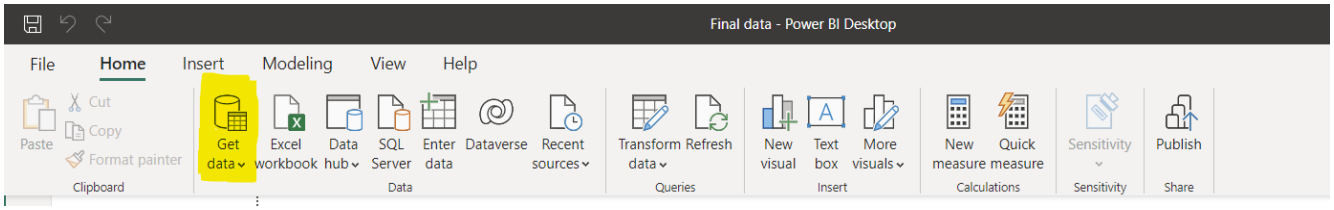
Creating Sales dashboard from Final Data created by PDI job.

Dashboard Created:

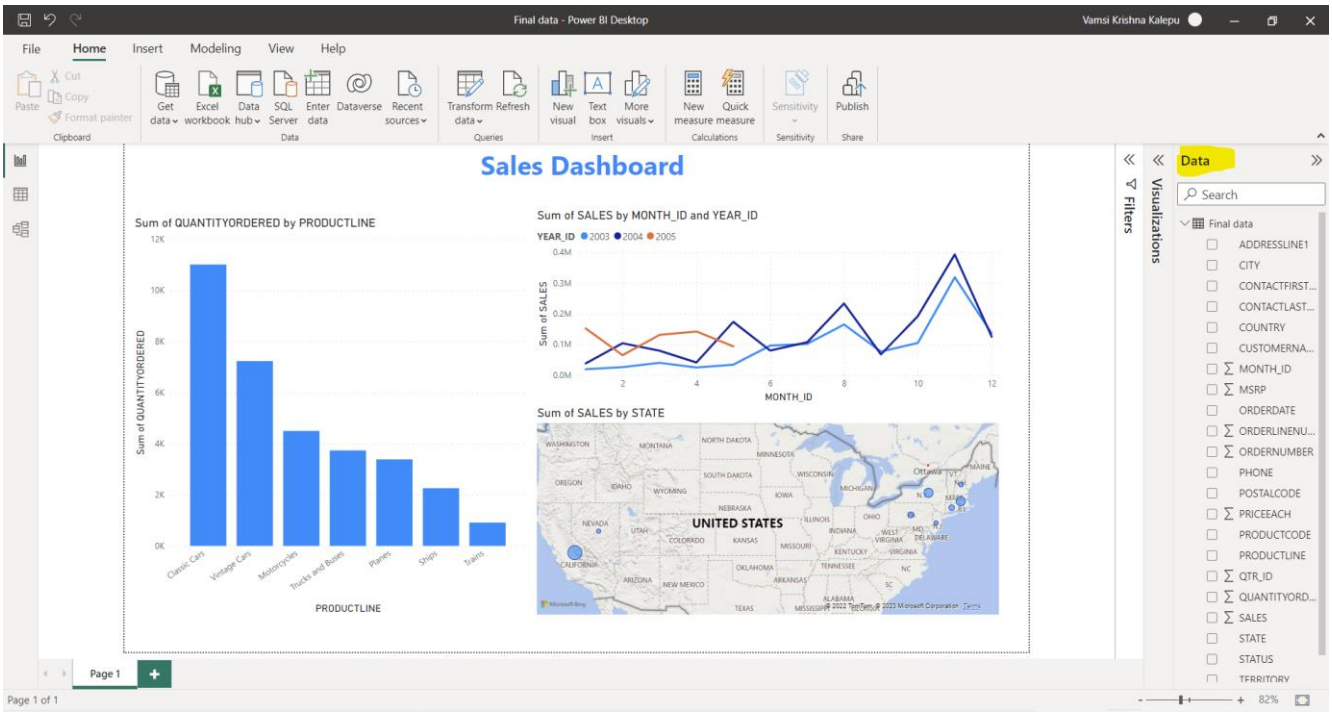


Process:

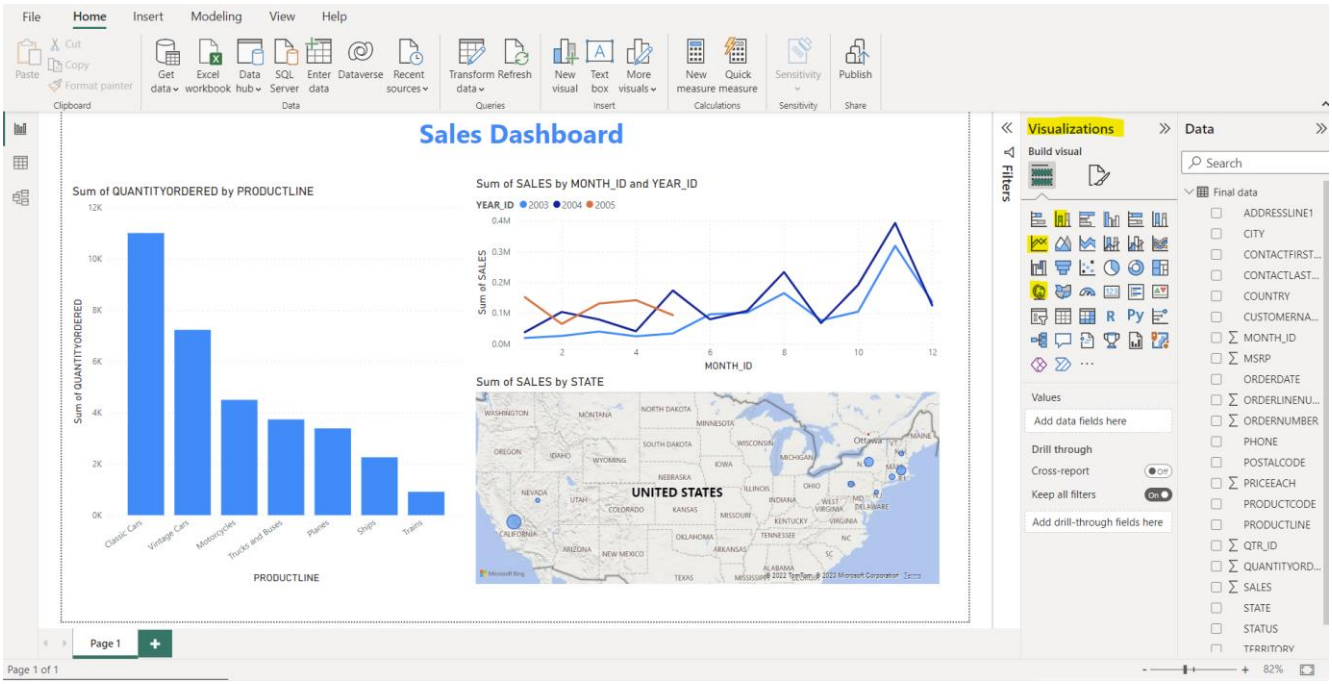
1) Add data using the Get Data option.



2) Select required attributes once data is loaded.



3) Select different graphs from the visualization section.



4) Add required filters to segregate the data as required.