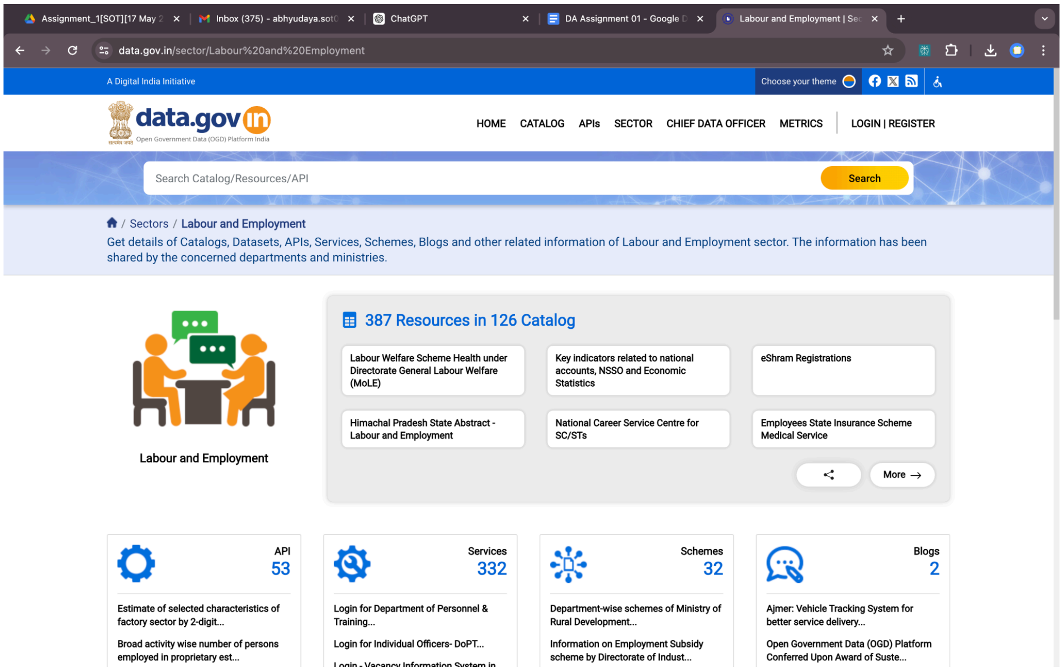
1.Data Selection


Assignment_1[SOT][17 May 2
Inbox (375) - abhyudaya.sot
ChatGPT
DA Assignment 01 - Google D
Labour and Employment | Sec
data.gov.in/sector/Labour%20and%20Employment
A Digital India Initiative
Choose your theme
data.gov.in
Open Government Data (OGD) Platform India
HOME
CATALOG
APIs
SECTOR
CHIEF DATA OFFICER
METRICS
LOGIN | REGISTER
Search Catalog/Resources/API
Search
Sectors / Labour and Employment
Labour and Employment
More

Get details of Catalogs, Datasets, APIs, Services, Schemes, Blogs and other related information of Labour and Employment sector. The information has been shared by the concerned departments and ministries.

387 Resources in 126 Catalog

- Labour Welfare Scheme Health under Directorate General Labour Welfare (MoLE)
- Key indicators related to national accounts, NSSO and Economic Statistics
- eShram Registrations
- Himachal Pradesh State Abstract - Labour and Employment
- National Career Service Centre for SC/STs
- Employees State Insurance Scheme Medical Service

API 53

- Estimate of selected characteristics of factory sector by 2-digit...
- Broad activity wise number of persons employed in proprietary est...

Services 332

- Login for Department of Personnel & Training...
- Login for Individual Officers- DoPT...
- Login - Vacancy Information System in

Schemes 32

- Department-wise schemes of Ministry of Rural Development...
- Information on Employment Subsidy scheme by Directorate of Indust...

Blogs 2

- Ajmer: Vehicle Tracking System for better service delivery...
- Open Government Data (OGD) Platform Conferred Upon Award of Suste...

I selected the data from a public repository on the government website Data.gov.

2.Task Overview:

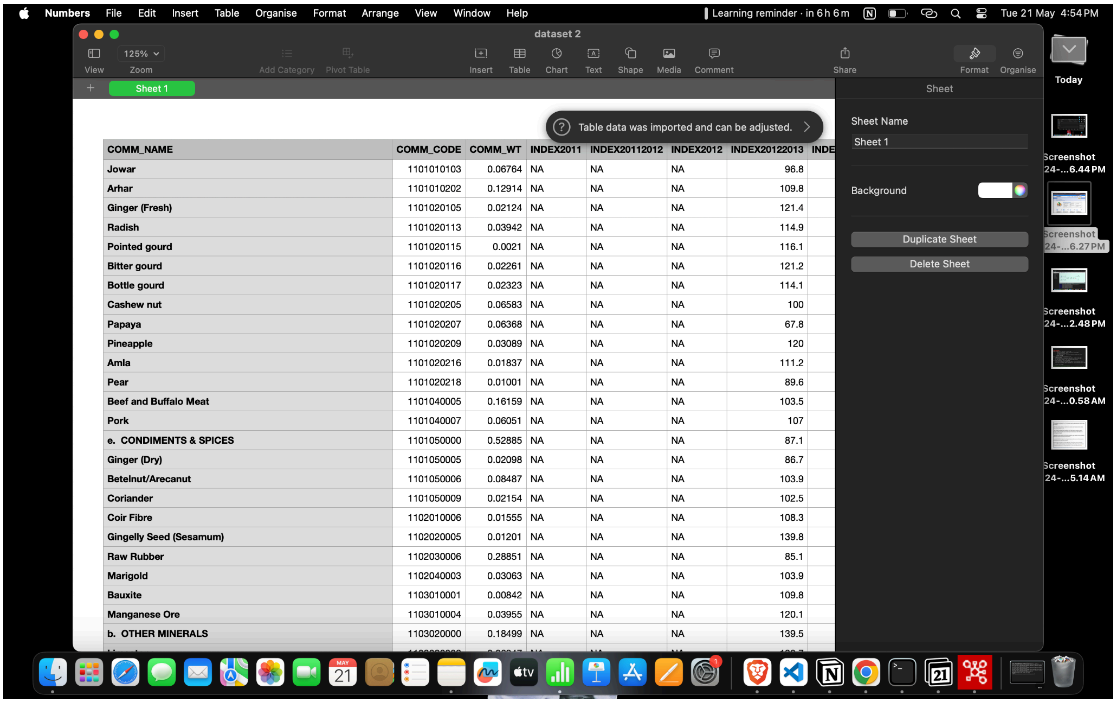


Table data was imported and can be adjusted.

COMM_NAME	COMM_CODE	COMM_WT	INDEX2011	INDEX20112012	INDEX2012	INDEX20122013	INDE
Jowar	1101010103	0.06764	NA	NA	NA	96.8	
Arhar	1101010202	0.12914	NA	NA	NA	109.8	
Ginger (Fresh)	1101020105	0.02124	NA	NA	NA	121.4	
Radish	1101020113	0.03942	NA	NA	NA	114.9	
Pointed gourd	1101020115	0.0021	NA	NA	NA	116.1	
Bitter gourd	1101020116	0.02261	NA	NA	NA	121.2	
Bottle gourd	1101020117	0.02323	NA	NA	NA	114.1	
Cashew nut	1101020205	0.06583	NA	NA	NA	100	
Papaya	1101020207	0.06368	NA	NA	NA	67.8	
Pineapple	1101020209	0.03089	NA	NA	NA	120	
Amla	1101020216	0.01837	NA	NA	NA	111.2	
Pear	1101020218	0.01001	NA	NA	NA	89.6	
Beef and Buffalo Meat	1101040005	0.16159	NA	NA	NA	103.5	
Pork	1101040007	0.06051	NA	NA	NA	107	
e. CONDIMENTS & SPICES	1101050000	0.52885	NA	NA	NA	87.1	
Ginger (Dry)	1101050005	0.02098	NA	NA	NA	86.7	
Betelnut/Arecanut	1101050006	0.08487	NA	NA	NA	103.9	
Coriander	1101050009	0.02154	NA	NA	NA	102.5	
Coir Fibre	1102010006	0.01555	NA	NA	NA	108.3	
Gingelly Seed (Sesamum)	1102020005	0.01201	NA	NA	NA	139.8	
Raw Rubber	1102030006	0.28851	NA	NA	NA	85.1	
Marigold	1102040003	0.03063	NA	NA	NA	103.9	
Bauxite	1103010001	0.00842	NA	NA	NA	109.8	
Manganese Ore	1103010004	0.03955	NA	NA	NA	120.1	
b. OTHER MINERALS	1103020000	0.18499	NA	NA	NA	139.5	

- The wholesale data provided has been sourced from government records, detailing prices of daily essentials such as oils, crops, and vegetables.

- This dataset includes both numerical and categorical columns, offering a comprehensive view of wholesale prices across various categories.

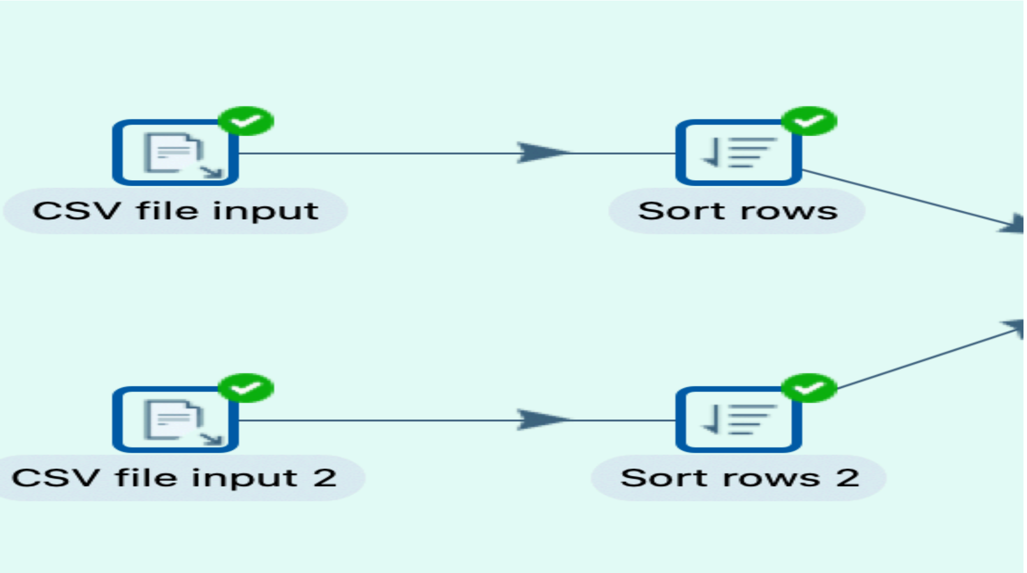
- The provided dataset has been edited to address missing values and remove duplicate entries.

- Additionally, the dataset has been divided into two parts.

- When these two parts are merged together, access to complete rows is ensured, allowing for comprehensive data analysis.

The task I am currently undertaking involves:

1. **Initially, I'll organize the rows in ascending order for both segments of the incomplete dataset.**



The **"Sort Rows"** step allows you to sort the rows of data based on one or more fields within those rows.

Here's how it typically works:

1. **Select Fields to Sort**: You specify which fields or columns you want to sort your data by.

2. **Ascending or Descending Order**: You define whether the sorting should be in ascending or descending order for each field.

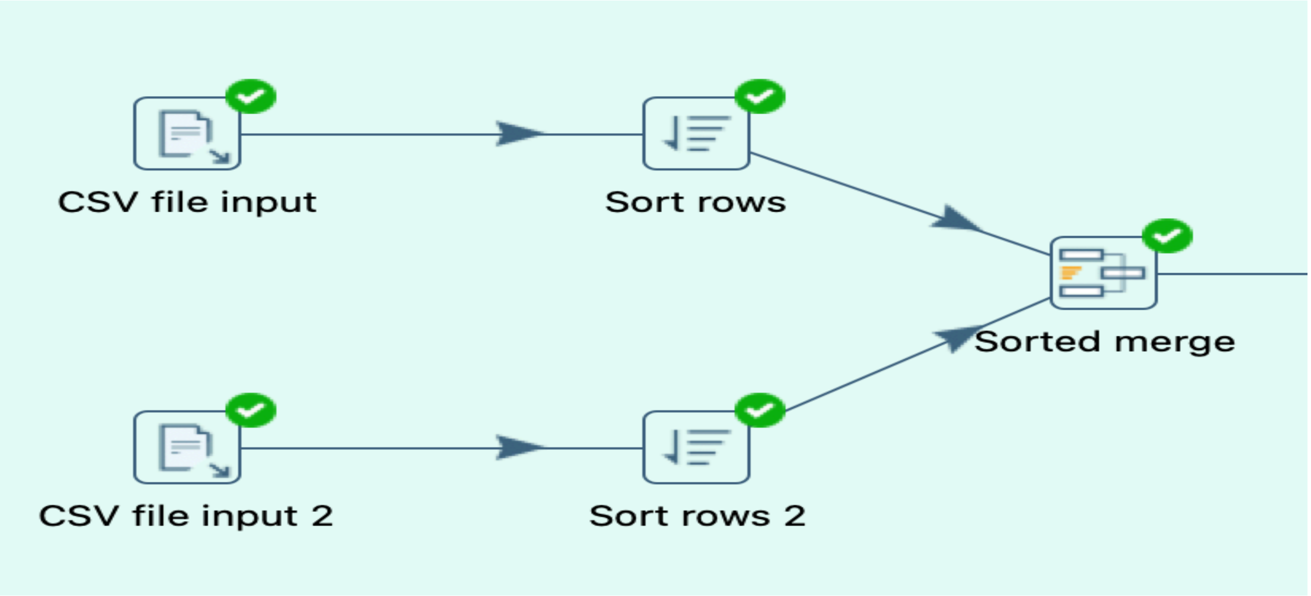
3. **Sort Algorithm**: You may also have options for the sorting algorithm to be used, depending on the version of Pentaho you're using. Common options include quicksort or mergesort.

4. **Sorting Options:** You might have additional options for handling null values or case sensitivity during sorting.

5. **Output**: The step outputs the sorted rows of data, ready for further processing or output to a destination.

This step is often used in data transformation processes where you need to arrange data in a specific order before performing further operations, such as joining with another dataset or aggregating values. Sorting data is a fundamental operation in data processing, and the "Sort Rows" step in Pentaho provides a convenient way to achieve this within your data pipelines.

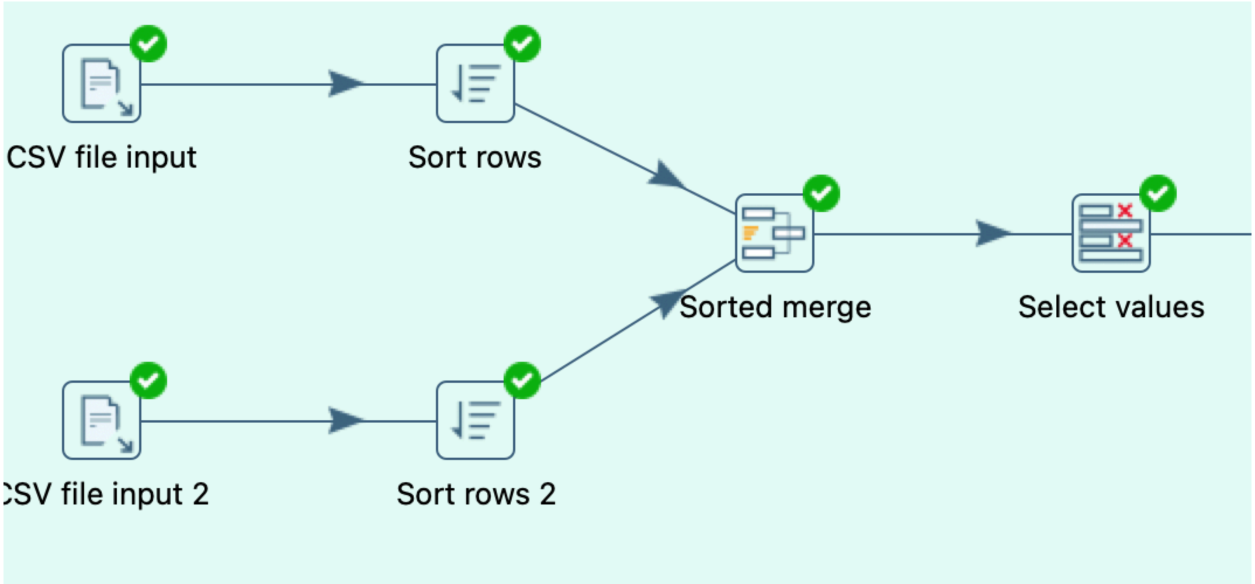
2.**Merging the two sorted dataset:-**



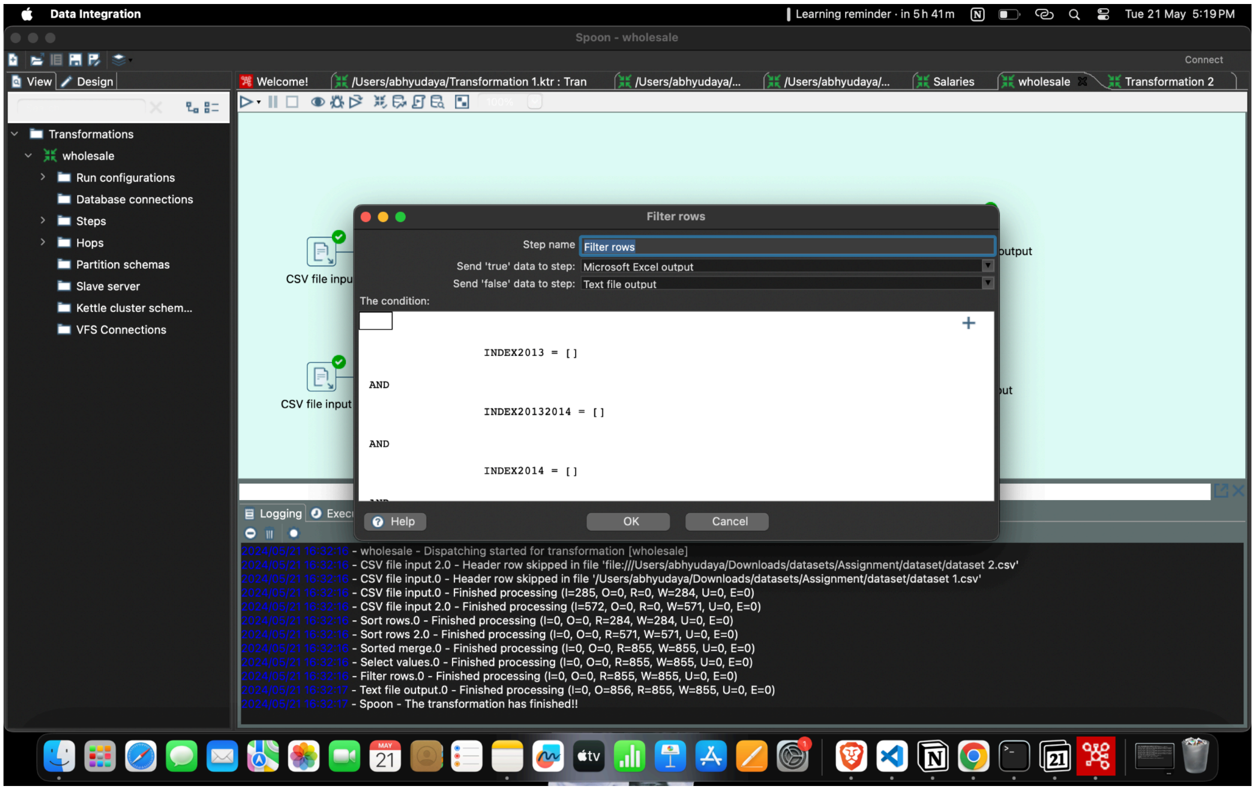
The "**Sorted Merge"** step in Pentaho, also known as "Merge Join" in some versions, is used to merge two sorted datasets based on a common key field or fields. This step is similar to a database join operation, but it requires that both input streams are sorted based on the join key beforehand.

integration workflows to combine data from different sources before further processing or analysis**.**

3.Utilizing the "Select Values" method offers a versatile approach for refining data, encompassing crucial tasks such as meticulous data type management, precise field selection, and efficient removal of duplicate values.



4.Utilizing the "Filter Rows" method to meticulously refine the dataset by systematically eliminating duplicate entries and null values, ensuring the data integrity remains pristine.



Filter rows

Step name: Filter rows

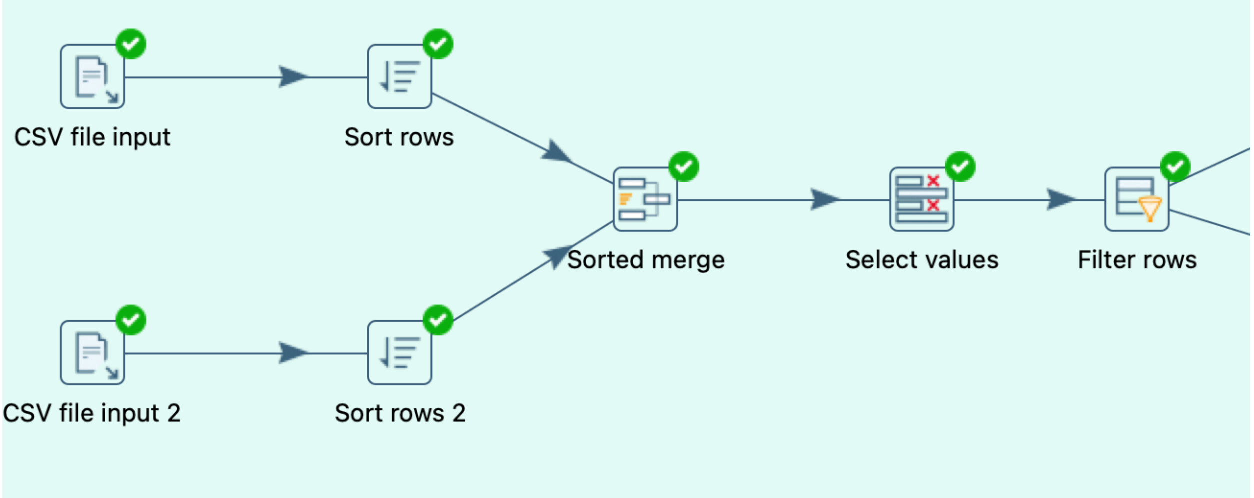
Send 'true' data to step: Microsoft Excel output

Send 'false' data to step: Text file output

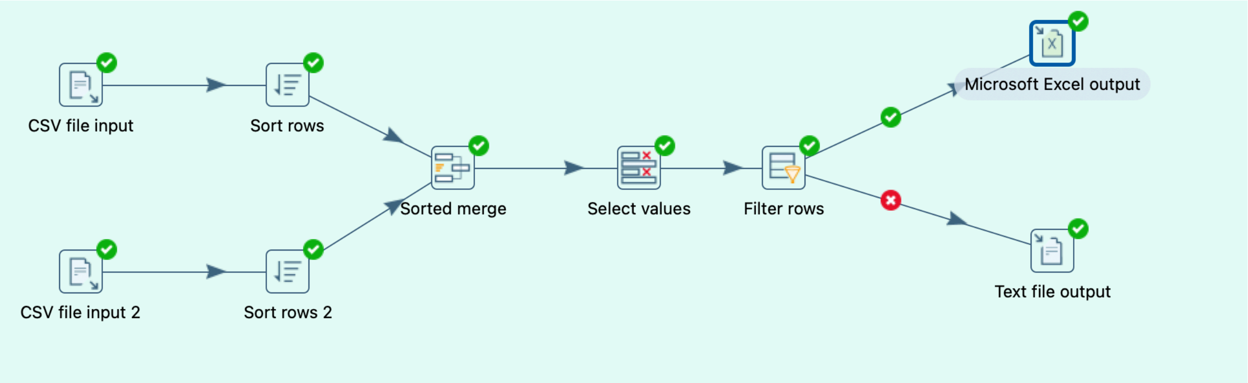
The condition:

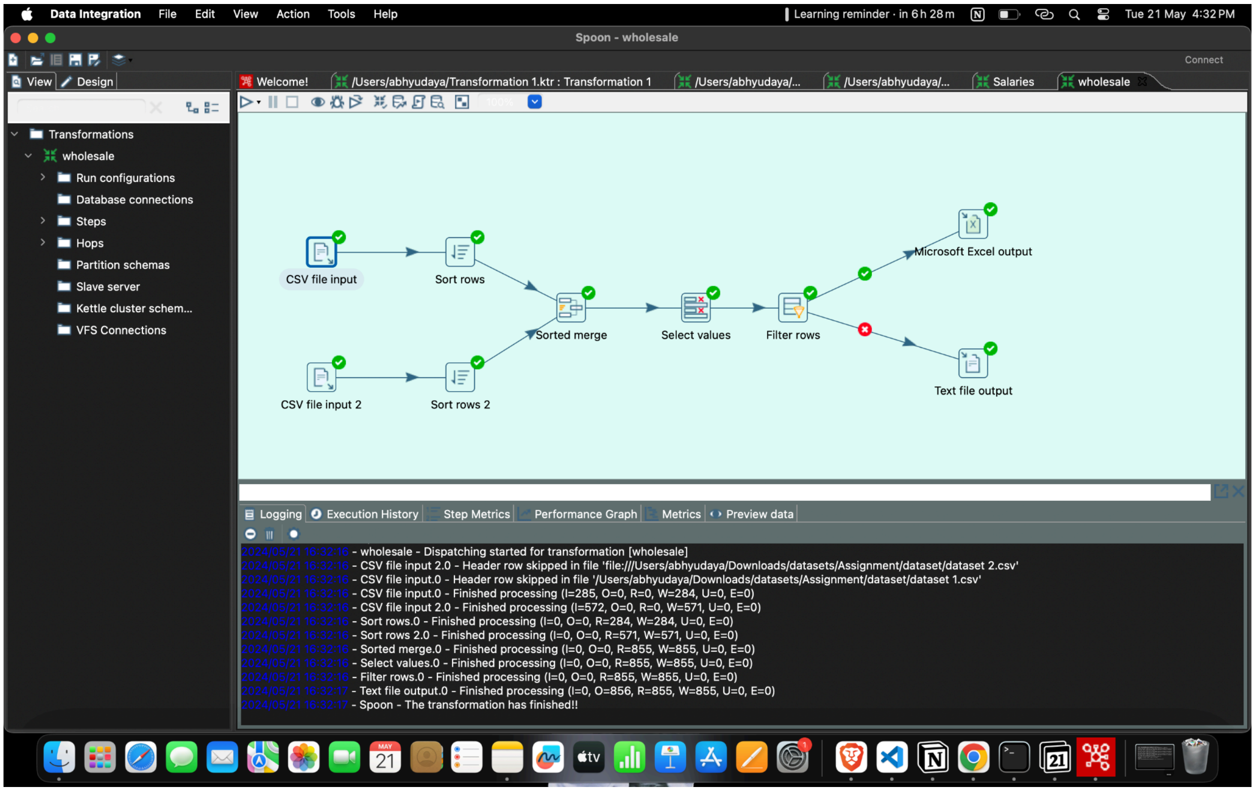
```
                INDEX2013 = []

AND

                INDEX20132014 = []

AND

                INDEX2014 = []
```

```
2024/05/21 16:32:16 - wholesale - Dispatching started for transformation [wholesale]
2024/05/21 16:32:16 - CSV file input 2.0 - Header row skipped in file 'file:///Users/abhyudaya/Downloads/datasets/Assignment/dataset/dataset 2.csv'
2024/05/21 16:32:16 - CSV file input.0 - Header row skipped in file '/Users/abhyudaya/Downloads/datasets/Assignment/dataset/dataset 1.csv'
2024/05/21 16:32:16 - CSV file input.0 - Finished processing (I=285, O=0, R=0, W=284, U=0, E=0)
2024/05/21 16:32:16 - CSV file input 2.0 - Finished processing (I=572, O=0, R=0, W=571, U=0, E=0)
2024/05/21 16:32:16 - Sort rows.0 - Finished processing (I=0, O=0, R=284, W=284, U=0, E=0)
2024/05/21 16:32:16 - Sort rows 2.0 - Finished processing (I=0, O=0, R=571, W=571, U=0, E=0)
2024/05/21 16:32:16 - Sorted merge.0 - Finished processing (I=0, O=0, R=855, W=855, U=0, E=0)
2024/05/21 16:32:16 - Select values.0 - Finished processing (I=0, O=0, R=855, W=855, U=0, E=0)
2024/05/21 16:32:16 - Filter rows.0 - Finished processing (I=0, O=0, R=855, W=855, U=0, E=0)
2024/05/21 16:32:17 - Text file output.0 - Finished processing (I=0, O=856, R=855, W=855, U=0, E=0)
2024/05/21 16:32:17 - Spoon - The transformation has finished!!
```



5. Effectively isolating accurate data from erroneous entries post-application of the 'Filter Rows' method.





```
2024/05/21 16:32:16 - wholesale - Dispatching started for transformation [wholesale]
2024/05/21 16:32:16 - CSV file input 2.0 - Header row skipped in file 'file:///Users/abhyudaya/Downloads/datasets/Assignment/dataset/dataset 2.csv'
2024/05/21 16:32:16 - CSV file input.0 - Header row skipped in file '/Users/abhyudaya/Downloads/datasets/Assignment/dataset/dataset 1.csv'
2024/05/21 16:32:16 - CSV file input.0 - Finished processing (I=285, O=0, R=0, W=284, U=0, E=0)
2024/05/21 16:32:16 - CSV file input 2.0 - Finished processing (I=572, O=0, R=0, W=571, U=0, E=0)
2024/05/21 16:32:16 - Sort rows.0 - Finished processing (I=0, O=0, R=284, W=284, U=0, E=0)
2024/05/21 16:32:16 - Sort rows 2.0 - Finished processing (I=0, O=0, R=571, W=571, U=0, E=0)
2024/05/21 16:32:16 - Sorted merge.0 - Finished processing (I=0, O=0, R=855, W=855, U=0, E=0)
2024/05/21 16:32:16 - Select values.0 - Finished processing (I=0, O=0, R=855, W=855, U=0, E=0)
2024/05/21 16:32:16 - Filter rows.0 - Finished processing (I=0, O=0, R=855, W=855, U=0, E=0)
2024/05/21 16:32:17 - Text file output.0 - Finished processing (I=0, O=856, R=855, W=855, U=0, E=0)
2024/05/21 16:32:17 - Spoon - The transformation has finished!!
```