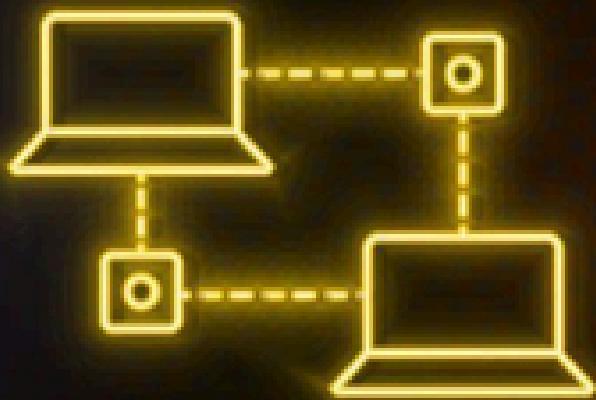




USER GUIDE



DEVELOPED BY :

Joy

21je0430@iitism.ac.in

Lalith Chatala

21je0508@iitism.ac.in

Table Of Contents



Access the Platform



Dashboard Overview



Dataset Tab



Result Tab



**Transformation
Operations**



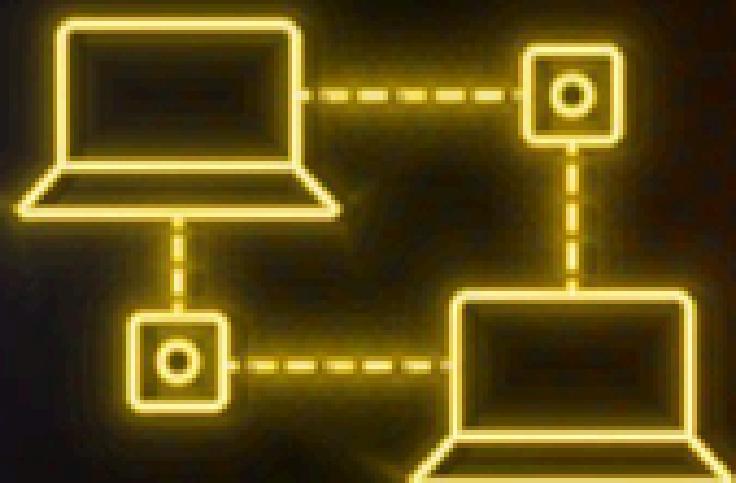
Conversion Operations



Agentic AI



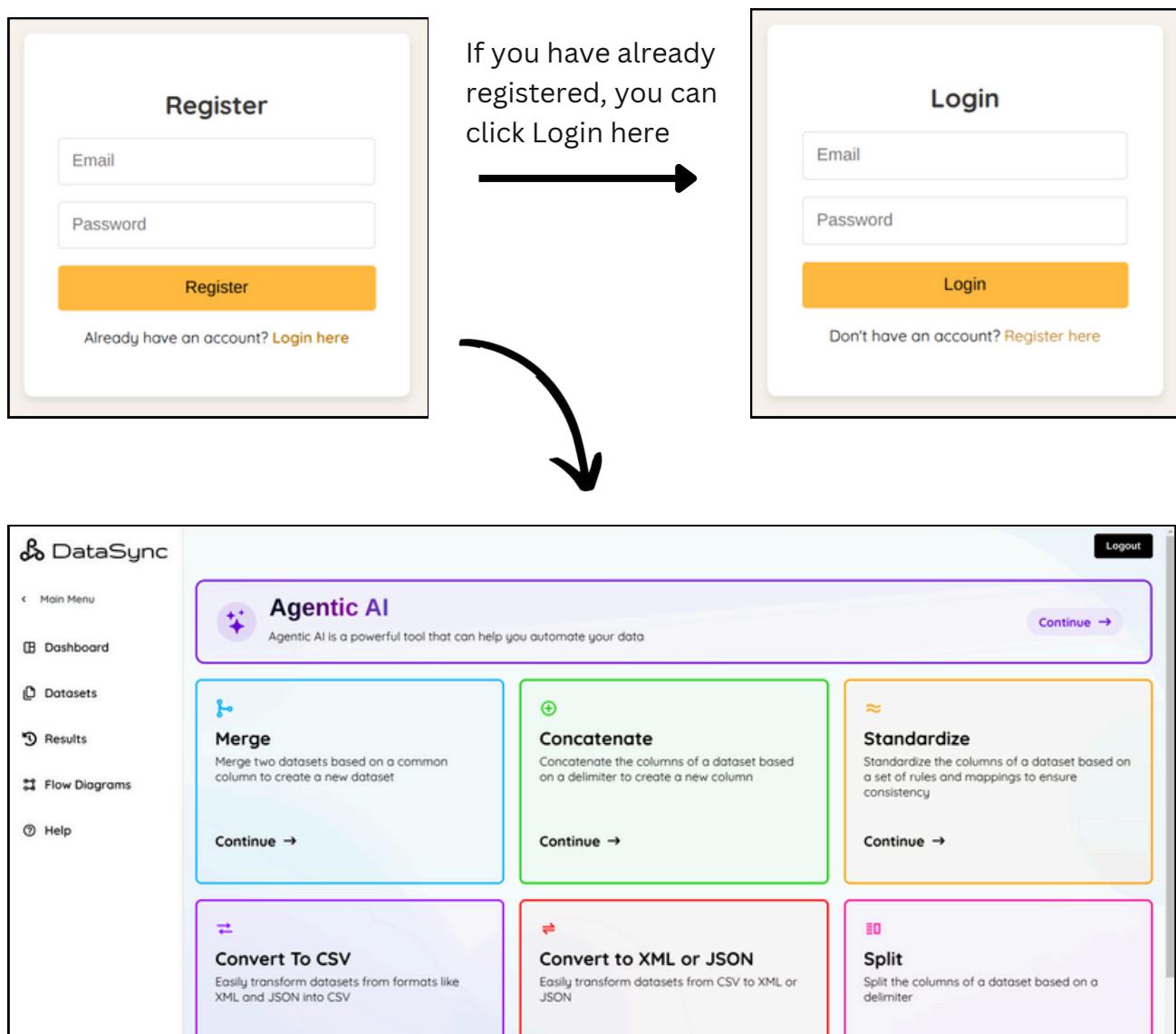
Flow Diagrams



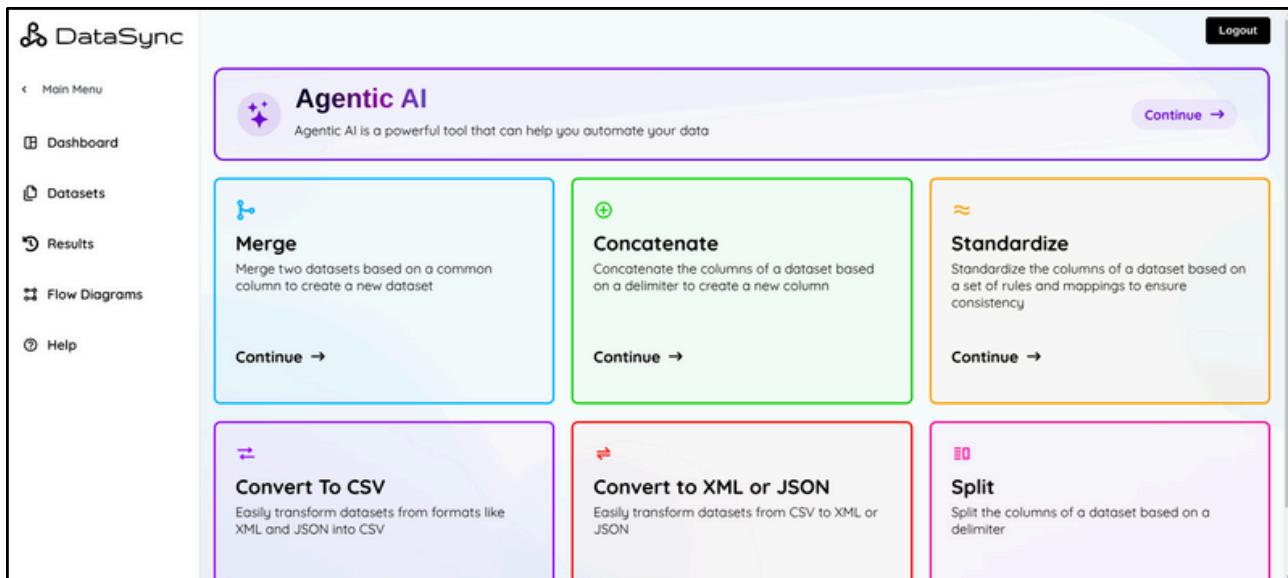
Access the Platform

Register

Enter your details to create a new account. Upon successful registration, you will be directed to the main dashboard, which provides a comprehensive overview of the available data management tools and features.



Dashboard Overview



Main Menu

The left sidebar serves as your navigation hub, offering quick access to all the main areas of the application:

- **Dashboard:** Your central overview page.
- **Datasets:** Where you can manage and view all your datasets.
- **Results:** Access and review the results of completed data operations.
- **Flow Diagrams:** Design and manage your data workflows.
- **Help:** Find tutorials and get support.

Dataset Tab

The Datasets tab is crucial for managing and manipulating datasets. This guide will cover how to add, view, and manipulate datasets, as well as how to utilize AI-generated suggestions for dataset operations.

Name	Size (bytes)	Type	Date Created	Actions	Export
standardize.csv	570	text/csv	17/10/2024, 16:36:40		
merge1(id,email,full_name).csv	870	text/csv	24/10/2024, 15:08:29		
merge2(id,company).csv	238	text/csv	24/10/2024, 15:10:07		
concatenate(id,first_name,last_name).csv	385	text/csv	24/10/2024, 15:15:36		
split(id,full name).csv	359	text/csv	24/10/2024, 16:03:13		

Adding New Datasets

1. **Navigate** to the Datasets tab from the Main Menu.
2. Click on the **+ Add Dataset** button at the top-right corner of the page.

3. In the **Upload Dataset** modal:

- **Drag and drop** your CSV files into the designated area, or
- Click within the outlined area to **browse** your computer and select the files you wish to upload.

4. After selecting your files, click **Upload** to add them to your datasets.

Viewing and Managing Datasets

- **View a Dataset:** Click the blue eye icon under the Actions column next to the dataset you want to view. This will open the dataset in a detailed view where you can see all records.

standardize.csv			
CSV Data			
Id	First_name	Last_name	Course
1	Luce	Powlett	Master of Technology
2	Herminia	Say	M.Tech.
3	Flora	Tommasuzzi	phd
4	Ennis	Lawlie	m.tech.
5	Cindra	Rowson	Master of Technology
6	Elie	Overthrow	b.tech.
7	Gus	Frisch	Bachelor of Technoloau

- **Delete a Dataset:** Click the red trash bin icon under the Actions column to remove the dataset from your list.

- **Export Options:**

- To export the dataset as a CSV, click the green CSV button under the Export column.
- To export the dataset as a PDF, click the red PDF button under the Export column.

✨ AI Suggestions

- When viewing a dataset, locate and click the AI Suggestions button at the bottom-right of the dataset view.
- The AI will analyze the dataset and suggest potential actions you can perform, such as standardizing data or merging columns. Each suggestion includes:
 - **Action Type:** What operation could be beneficial (e.g., standardize, merge, split).
 - **Parameters:** Specific details about the action (e.g., which columns to standardize and suggested standard values).

The screenshot shows a dataset named "standardize.csv" in a software interface. The main view displays a table with columns: Id, First_name, Last_name, and Course. The data consists of 21 rows, with rows 1 through 7 highlighted in grey. An arrow points from the "AI Suggestions" button at the bottom right of the main view towards a floating "Suggestions" panel on the right side of the screen.

Suggestions Panel Content:

- Thoughts ^**
- Suggestion 1**
- Concatenate first_name and last_name**
- Concatenates the 'first_name' and 'last_name' columns to create a new 'Full Name' column.
- Recommended Action**
- Concatenate**
- Parameters**
- columns array:

```
[ "first_name",
```

I Results Tab

The Results tab in DataSync provides access to datasets generated from various operations performed on the platform. It functions similarly to the Datasets tab with some specific differences related to the types of data displayed.

Utilize the Search datasets... box to quickly locate specific results by name or descriptive keywords.

The screenshot shows the DataSync application interface with the 'Results' tab selected. The left sidebar includes links for Main Menu, Dashboard, Datasets, Results (which is highlighted), Flow Diagrams, and Help. The main content area is titled 'Results Datasets' and features a search bar labeled 'Search datasets...'. Below the search bar is a table with the following data:

Name	Size (bytes)	Type	Date Created	Actions	Export
result_xyz this is desc.	988	text/csv	11/11/2024, 16:08:06	Edit Delete	CSV PDF
add_result	1849	text/csv	11/11/2024, 16:08:48	Edit Delete	CSV PDF
Res res	1221	text/csv	07/04/2025, 18:35:34	Edit Delete	CSV PDF
add add	1849	text/csv	07/04/2025, 20:09:33	Edit Delete	CSV PDF
standardized_file.csv Standardizing course column	666	text/csv	07/04/2025, 22:44:59	Edit Delete	CSV PDF

Transformation Operations

Merge

Combine two datasets based on a common column to create a new dataset. Ideal for integrating related information from separate sources.



How to Merge Datasets

Click on the Merge on the dashboard, it will open a form like below

A screenshot of a "Merge Datasets" form. It has two main sections: "Select Dataset" and "Select Column from Dataset 2".

- Select Dataset:** A dropdown menu showing "Standard" (selected) and "Result". Below it is a search bar "Q Search datasets..." and a list of datasets:
 - standardize.csv
 - merge2(id,company).csv
 - concatenate(id,first_name,last_name).csv
 - split(id,full name).csv
 - Address.csv
 - Unstandardized_Gender_Example.csv
- Select Column from Dataset 2:** A dropdown menu showing "id".

1. Select Datasets:

Click on the drop-down menu under 'Select Dataset' to choose the first dataset you want to merge.

2. Choose Columns to Merge On:

Click the drop-down menu labeled 'Select Column for Dataset' to select the column from the first dataset that you will use as the key for merging.

Then Click on Merge:

When you click on Merge Button, a modal will appear displaying input fields for Name of your result dataset i.e the merged dataset and an optional description about it.

The diagram illustrates the process of merging datasets. It starts with a 'Merge Datasets' interface where two CSV files are selected: 'merge1(id,email(full_name).csv)' and 'merge2(id,company).csv'. Each file has a dropdown menu next to its name, likely for selecting merge keys. A large black arrow points from this interface down to a second window titled 'Enter Output File Details'. This modal contains fields for 'Output File Name' and 'Description (Optional)', along with 'Submit' and 'Cancel' buttons.

The merged dataset will then be created, combining the data from both datasets based on the selected key columns. This new dataset can be used for further analysis or reporting as required.

Concatenate

It combine multiple columns within a dataset into a single column, using a specified delimiter. This is useful for creating new data fields that consolidate information from various parts of the dataset.



Concatenate

Concatenate the columns of a dataset based on a delimiter to create a new column

[Continue →](#)

How to Concatenate columns inside Dataset

Click on the Concatenate on the dashboard, it will open a form like below

The screenshot shows a user interface for concatenating columns. At the top, there's a title 'Concatenate Columns'. Below it are several input fields and dropdown menus:

- 'Select Dataset' dropdown menu.
- 'Select Columns' dropdown menu.
- 'Delimiter' dropdown menu set to 'Comma (,)'. Other options include 'Tab (t)', 'Semicolon (;)', and 'None'.
- 'Final Column Name' input field.
- A large grey button at the bottom labeled 'Concatenate'.

1. Select Datasets:

Click on the 'Select Dataset' dropdown to choose the dataset you wish to manipulate.

2. Select Columns:

Use the 'Select Columns' dropdown to pick the columns you want to concatenate. You can select two or more columns as needed for your data requirements.

3. Choose a Delimiter:

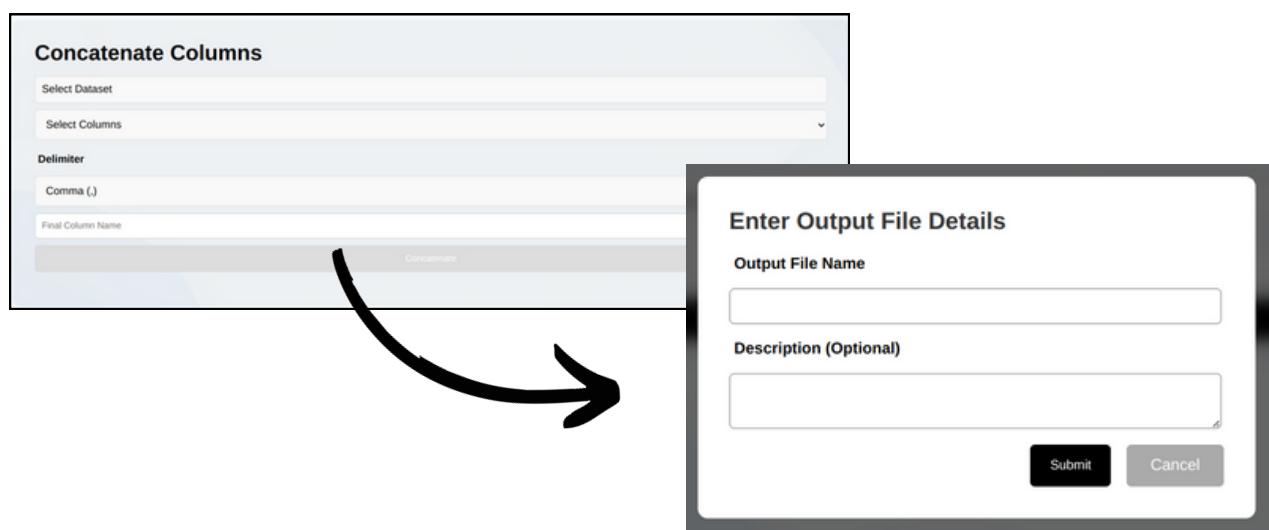
Choose a delimiter from the 'Delimiter' dropdown menu. Common delimiters include commas, semicolons, or spaces. This character will separate the values from each column in the new concatenated column.

4. Specify Final Column Name:

Enter a name for the final concatenated column in the 'Final Column Name' field. This name should be descriptive and reflect the contents or purpose of the new column.

5. Then Click on Concatenate:

When you click on Concatenate Button, a modal will appear displaying input fields for Name of your result dataset i.e the manipulated dataset and an optional description about it.



Standardize

This maps various text entries in a dataset to standardized values. making it invaluable for cleaning data and ensuring uniformity, particularly with variations in terminology or abbreviations.

≈

Standardize

Standardize the columns of a dataset based on a set of rules and mappings to ensure consistency

[Continue →](#)

How to Standardize a Dataset

Click on the Standardize on the dashboard, it will open a form like below

Standardize Data

Select Dataset
standardize.csv

Select Column
course

Mappings

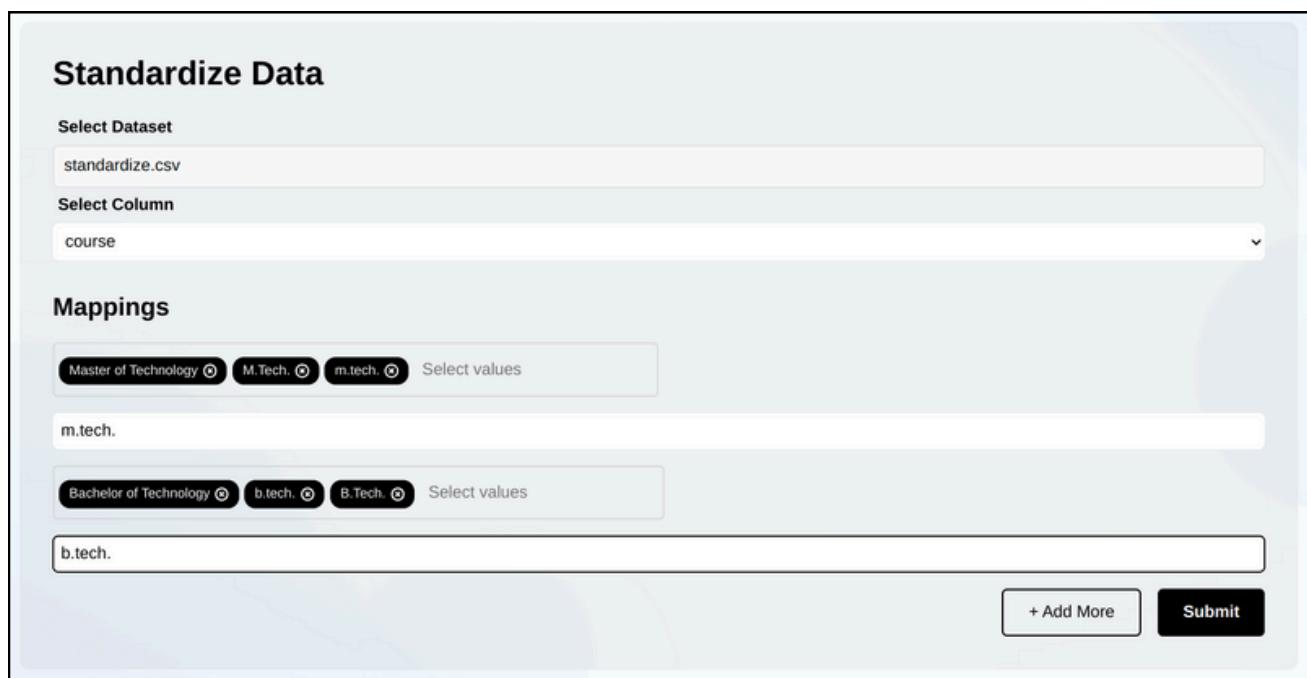
Master of Technology M.Tech. m.tech. Select values

m.tech.

Bachelor of Technology b.tech. B.Tech. Select values

b.tech.

[+ Add More](#) [Submit](#)



1. Select Datasets:

Begin by selecting the dataset you want to standardize using the 'Select Dataset' dropdown.

2. Select Column:

Choose the column that contains the data you wish to standardize from the 'Select Column' dropdown. This column will contain the values that need to be mapped to a standard format.

3. Define Mappings:

For each variant of a term in the selected column, enter the standardized term in the 'Mappings' section.

- **Add Variant:** Click on 'Select values' to add all different existing variants of a term that need to be mapped to a single standard value.
- **Standard Value:** In the adjacent text box, specify the standard term that all selected variants should map to.
- **Example:** For "Master of Technology", add variants like "M.Tech." and "m.tech." to map to "Master of Technology".

4. Add more Mappings:

Click '+ Add More' to include additional mappings for other terms as needed.

5. Submit Changes:

Once all mappings are correctly set, click the 'Submit' button

Standardize Data

Select Dataset
standardize.csv

Select Column
course

Mappings

Master of Technology	M.Tech.	m.tech.	Select values
m.tech.			
Bachelor of Technology	b.tech.	B.Tech.	Select values
b.tech.			

+ Add More Submit



Enter Output File Details

Output File Name

Description (Optional)

Submit Cancel

a modal will open similarly in the case of merge, enter the result dataset name and optional description, and click submit to apply standardisation.

Split

There are two functionalities for splitting columns: General Split and Address Split. The General Split allows you to divide a column into multiple new columns based on a specified delimiter. The Address split uses

 **Split**
Split the columns of a dataset based on a delimiter

Continue →

an API to break down an address column into components like Street Address, Post Office, District, State, Pincode, and Country, based on the PIN code.

General Split

Click on the Standardize on the dashboard and select 'General Split' tab, it will open a form like below

The screenshot shows a user interface for splitting columns. At the top, there's a title 'Split Columns' and a navigation bar with tabs: 'General Split' (which is active and highlighted in dark grey) and 'Address Split' (which is white). Below the tabs, there are two main sections: 'Select Dataset' and 'Select Column'. In the 'Select Dataset' section, a dropdown menu is open, showing 'split(id,full name).csv' as the selected item. In the 'Select Column' section, there are three dropdown menus: one for 'Select Column' containing 'Select Column', one for 'Select Delimiter' containing ',', and a text input field for 'Enter Column Names' containing '1'. Below these sections is a large text input field labeled 'Column 1 Name'. At the bottom right of the form is a prominent 'Convert' button.

1. Select Dataset:

Choose the dataset you want to work with from the 'Select Dataset' dropdown.

2. Select Column:

Pick the column you intend to split from the 'Select Column' dropdown.

3. Select Delimiter:

Choose a delimiter that separates the data in the selected column. Common delimiters include commas (,), semicolons (;), or spaces.

4. Enter New Column Names:

Specify the names for the new columns that will be created post-split in the 'Enter Column Names' fields.

5. Split the Column:

Click on the 'Convert' button and enter Output file name and optional description in modal to execute the splitting of the column based on the chosen delimiter.

Address Split

Click on the Standardize on the dashboard and select 'Address Split' tab, it will open a form like below

The screenshot shows a 'Split Columns' interface. At the top, there are two tabs: 'General Split' (unselected) and 'Address Split' (selected, indicated by a dark background). Below the tabs, there are two dropdown menus. The first dropdown is labeled 'Select Dataset' and contains the option 'split(id,full name).csv'. The second dropdown is labeled 'Select Address Column' and contains the option 'Select Column'. At the bottom of the interface is a large, dark rectangular button labeled 'Split Address'.

1. Select Dataset:

Use the 'Select Dataset' dropdown to pick the dataset containing the address data.

2. Select Address Column:

Choose the address column to be split from the 'Select Column' dropdown.

3. Split Address:

Click on the 'Split Address' button enter Output file name and optional description in modal. The system will segregate the address into specified components: Street Address, Post Office, District, State, Pincode, and Country.

Conversion Operations



Convert To CSV

Easily transform datasets from formats like XML and JSON into CSV

Continue →



Convert to XML or JSON

Easily transform datasets from CSV to XML or JSON

Continue →

Convert to CSV

Easily transform datasets from XML and JSON formats into CSV, which is widely used for its simplicity and compatibility with numerous data processing applications.

1. Select Dataset:

Use the interface to drag and drop a file, or click to select an XML or JSON file from your computer.

2. Convert:

Click the 'Convert' button to initiate the transformation of the dataset into a CSV file.

3. Download/Save:

Once converted, the new CSV file will be available for download or further processing within Datasync.

Convert to XML/JSON

Convert datasets from CSV to either XML or JSON formats, which are excellent for hierarchical data structures and are extensively used in web data and API interactions.

1. Select Dataset:

Drag and drop a CSV file into the designated area, or use the click option to upload from your system.

2. Choose Format:

Select the desired output format (XML or JSON) from the 'Convert To:' dropdown menu.

3. Convert and Download:

Hit the 'Convert' button to start the conversion process. After conversion, download the new file or utilize it as needed within the platform.

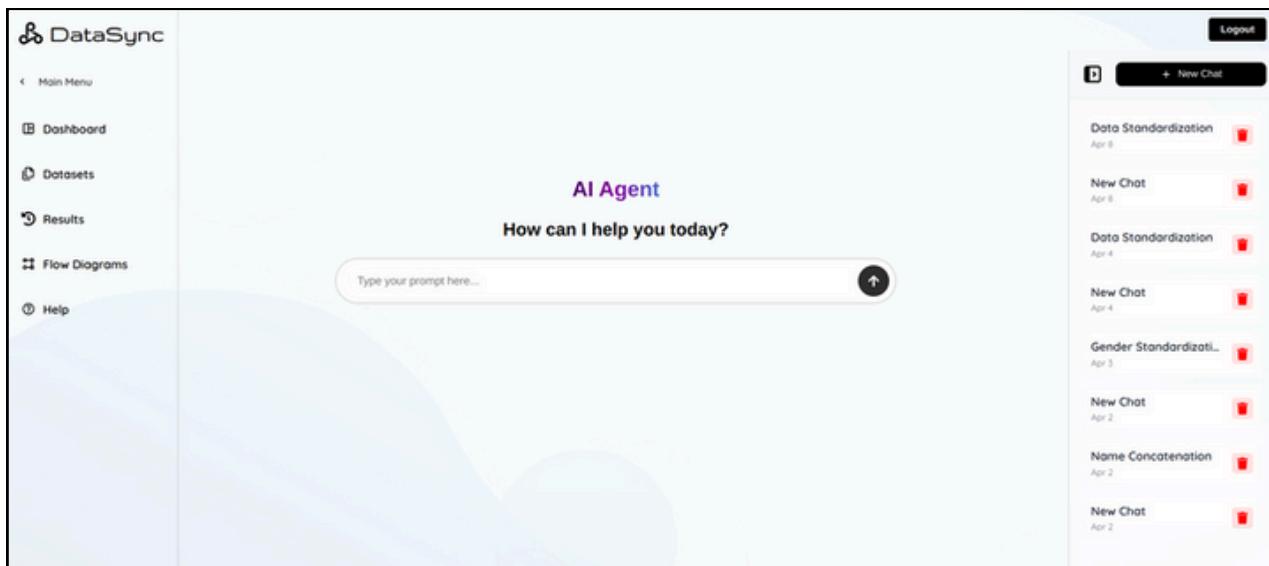


Agentic AI

The Agentic AI feature in DataSync provides a powerful interface where users can interact directly with AI to perform operations on datasets. This guide will walk you through how to utilize the AI chat feature effectively.

Starting a New Chat

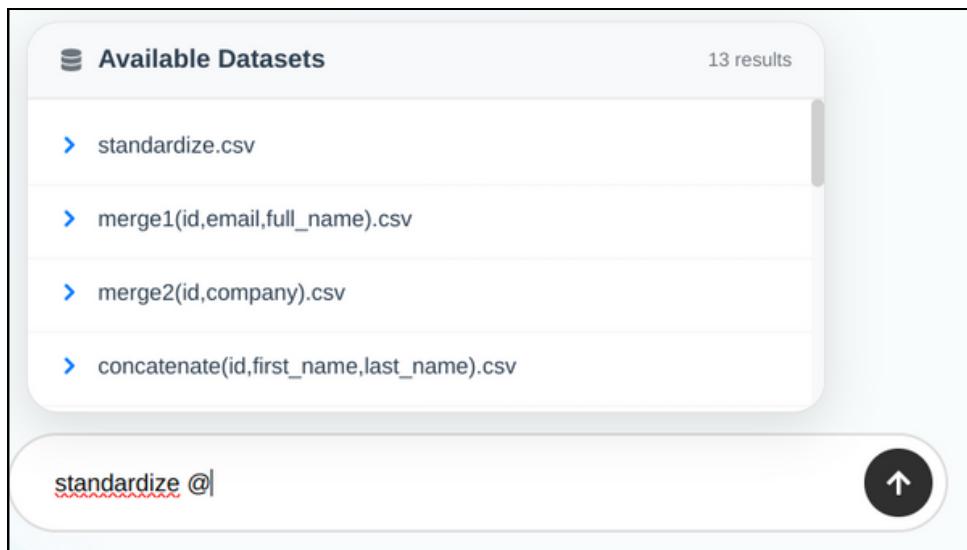
To start a new conversation, click the New Chat button located at the top-right of the chat panel.



Using the AI Agent

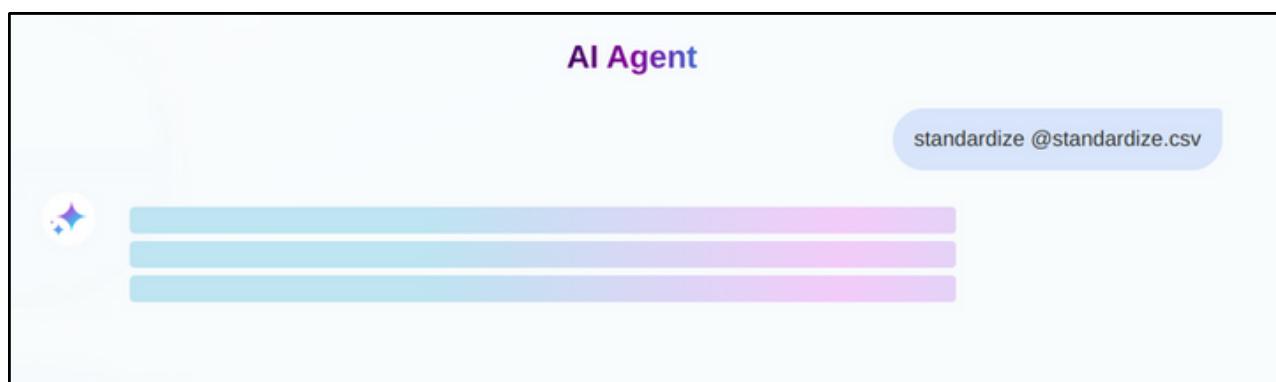
- **Typing Commands:** Simply type your query or command in the chat textbox and press Enter or click the send button (arrow icon).

- **Using '@' for Datasets:** If your operation involves datasets:
 - Type @ to trigger a dropdown menu.
 - Select the desired dataset from the dropdown list that appears.



Understanding AI Responses

- **Thoughts and Process Descriptions:** The AI may provide thoughts or describe its process in italicized text with a gray background. This helps you understand how the AI is processing your request.



- **Final Outputs:** Once the AI completes an operation, it will display the final output directly in the chat. This could be a data table or confirmation of a successful operation.

The screenshot shows the DataSync AI Agent interface. On the left is a sidebar with options: Main Menu, Dashboard, Datasets, Results, Flow Diagrams, and Help. The main area is titled "AI Agent" and contains a conversation log. The first message is from the AI: "I need to standardize the specified file. First, I'll get the headers to identify the columns. Then, I'll retrieve distinct values from a column to create mappings for standardization. Finally, I'll perform the standardization operation...". The second message is from the user: "standardize @standardize.csv". Below this, the AI continues: "The headers are obtained. Now I'll choose a column for standardization and get its distinct values. I'll select the "course" column for this example...". A CSV data preview table is shown with columns Id, First_name, Last_name, and Course. The data rows are: 1 Luce Powlett M.Tech and 2 Herminia Say M.Tech. At the bottom, there's a text input field "Type your prompt here..." and a send button. To the right, a sidebar shows a list of saved chats with timestamps: Apr 8, Apr 8, Apr 4, Apr 4, Apr 3, Apr 2, Apr 2, and Apr 2. Each entry has a delete icon next to it.

Example Operations

Standardizing Data:

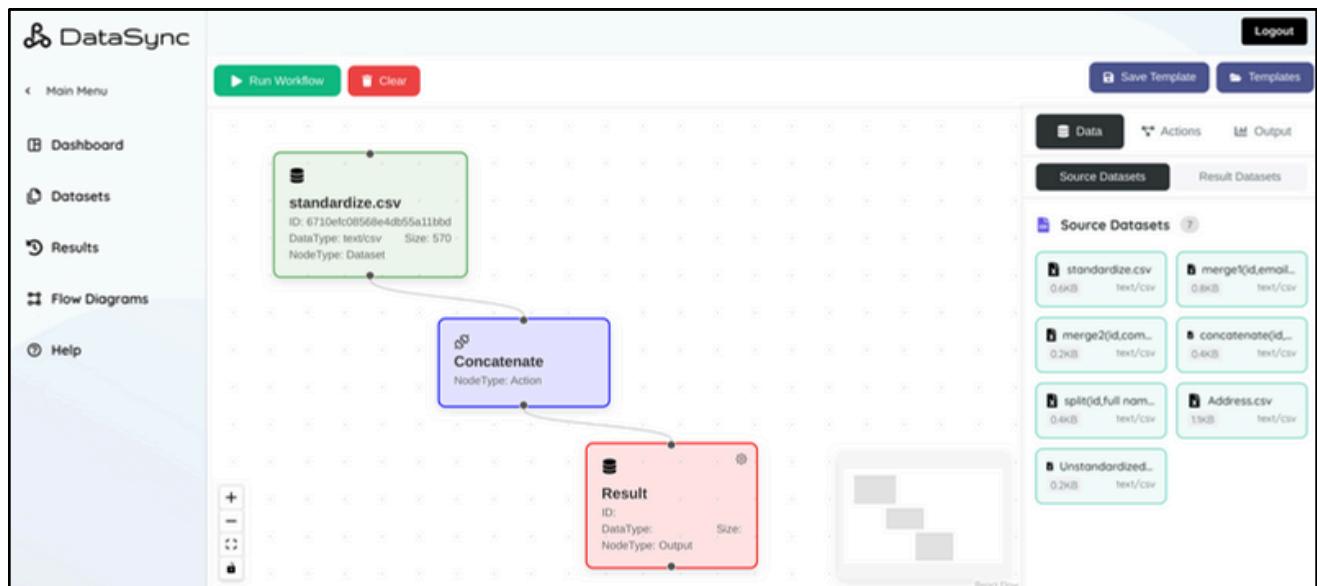
To standardize data, you might start by asking the AI to "Show headers of the dataset". Following the headers, you can instruct the AI to "Standardize the course column in dataset xyz.csv".

This screenshot shows the AI Agent interface with a focus on the chat history. The history includes three entries: "Data Standardization" (Apr 8), "New Chat" (Apr 8), and "Data Standardization" (Apr 4). Each entry is accompanied by a red trash can icon for deletion.

Chats are automatically saved, and you can review them anytime by scrolling through the chat history or by accessing the Saved Chats section from the dashboard.

Flow Diagrams

The Flow Diagrams tab in DataSync allows you to create, visualize, and run data transformation workflows. You can integrate various datasets, apply actions such as merge, split, or standardize, and generate outputs all within a graphical interface.



Creating a Workflow

1. Starting a New Workflow:

- Click on the Data tab to drag and drop dataset nodes into the workspace.
- Use the Actions tab to add transformation actions like Concatenate, Split, Merge, or Standardize between your dataset nodes.

2. Configuring Nodes:

- Connect dataset nodes to action nodes by dragging a line from the dataset to the action.
- Each action node needs to be configured by connecting the datasets involved in the action.

The screenshot shows the 'Data' sidebar with tabs for 'Actions' and 'Output'. Under 'Source Datasets', there are seven items listed:

- standardize.csv (0.6KB, text/csv)
- merge1(id,email... (0.8KB, text/csv)
- merge2(id,com... (0.2KB, text/csv)
- concatenate(id,... (0.4KB, text/csv)
- split(id,full nam... (0.4KB, text/csv)
- Address.csv (1.1KB, text/csv)
- Unstandardized... (0.2KB, text/csv)

3. Setting Up Actions:

- Once an action node is placed and connected, select it to configure its parameters.
- A form will appear in the right sidebar where you can specify the necessary parameters and conditions for the action

The screenshot shows the 'Actions' sidebar with tabs for 'Data' and 'Output'. Under 'Data Transformations', there are four actions listed:

- Concatenate**: Combine multiple datasets vertically.
- Split**: Divide dataset by columns or conditions.
- Merge**: Join datasets using common keys.
- Standardize**: Normalize data formats and values.

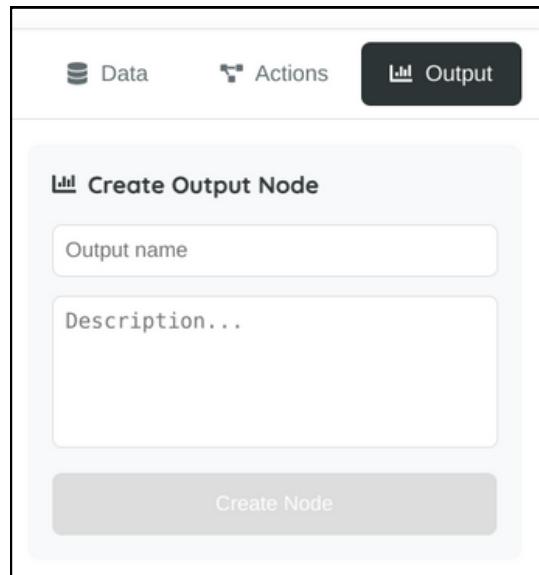
click on the Action, and set parameters in the form

The screenshot shows the configuration form for the 'Concatenate' action:

- Dataset**: Select Dataset (dropdown).
- Select Columns**: Select Columns (dropdown).
- Delimiter**: Comma (.) (dropdown).
- Final Column Name**: (text input field).
- Set Parameters**: (button).

4. Defining Output:

- Use the Output tab to create an output node.
- Connect the action node to the output node to designate where the result of the workflow should be saved.



Running the workflow

Execute the Workflow:

- Once your workflow is fully configured with datasets, actions, and outputs linked, click the Run Workflow button at the top left of the page to process the workflow.
- The system will execute all steps and save the final results to the designated output node.

Viewing Results

After running the workflow, the result will be available in the Results tab under the specified output node. You can view, export, or delete the result dataset from there.



Contact Us

✉ 21je0430@iitism.ac.in

✉ 21je0508@iitism.ac.in

