



Chapter 2–Basic Operations

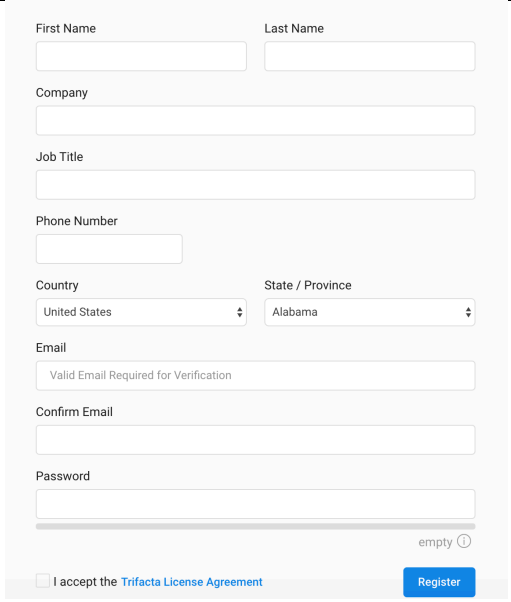
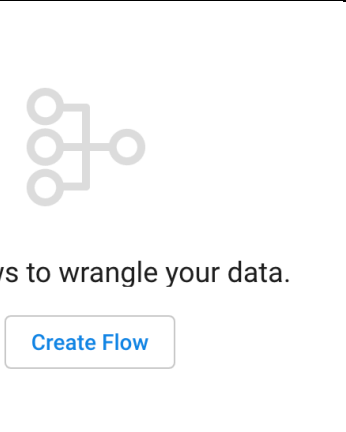
Lesson A | Uploading and Filtering Data | 5 Minutes

Lab Goals

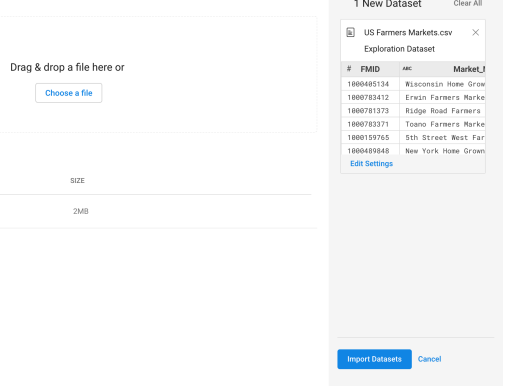
In this Chapter, you will:

- Register for Trifacta
- Create a Flow
- Upload a local file
- Add transforms to your Recipe
- Edit, undo/redo, add, and delete Recipe steps


Lesson Instructions

1.	Register for Trifacta If you haven't already, register for Trifacta. At the home page, click register, and fill in all of the required boxes with your information. Click on the box to accept the Trifacta licensing agreement and click register.	 The registration form includes fields for First Name, Last Name, Company, Job Title, Phone Number, Country (dropdown), State / Province (dropdown), Email, Confirm Email, and Password. It also has a checkbox for 'I accept the Trifacta License Agreement' and a 'Register' button.
2.	Create a New Flow In the Trifacta workspace, click the Create Flow button. Specify the name of the Flow: <ul style="list-style-type: none">• Flow Name: Training• Flow Description: Training Data	 The 'Create Flow' button is shown in a light blue box. Above it is a diagram of a flow with three input nodes and one output node.



3.	<p>Upload a Local File</p> <p>Click the Add Datasets button. In this interface, you can drag and drop or upload files from your desktop. You can also select data in HDFS or other connected sources. If you're using Wrangler, you can only import local files from your desktop.</p> <p>Select the US Farmers Market.csv dataset. We could select multiple files but for now we will stick with this one. Click the box to the right on the file name. Specify the dataset name and description as:</p> <ul style="list-style-type: none">• Dataset Name: US Farmers Markets• Dataset Description: Exploration Dataset <p>Click the Import Datasets button.</p> <p>You can now see that the dataset you just imported is now a part of your Training Flow.</p> <p>If you remember the terminology from Chapter 1B, you can see the Imported Dataset, which is the source data, US Farmers Markets.csv.</p> <p>In order to start Wrangling our data, we'll need to add a Recipe to our US Farmers Market dataset. Select the Imported Dataset icon and select Add New Recipe. Now we're ready to get Wrangling. Select Edit Recipe and you will be brought into the Transformer page.</p>	
4.	<p>Filtering Rows</p> <p>Let's start by introducing a few basic transformations and features in Trifacta. Scroll to the right to find the Spring_Day_A column. Click on the black portion of the data quality bar. As mentioned in last chapter, interacting with the various elements on the Transformer page will lead Trifacta to prompt suggestions based on your interactions with the product.</p> <p>Select the suggestion to delete the rows where Spring_Day_A is missing values. Select the Rows box in the bottom right if you would like to see only the rows affected. Add to your recipe. You should now see that you have a step in your recipe.</p> <p>You can also accomplish the same task by selecting the green portion of the data quality bar and choosing the keep suggestion.</p>	<p>delete row: ismissing([Spring_Day_A])</p>
5.	<p>Filtering Columns</p> <p>Open up the Columns View by clicking on the Columns button in the top left. Select the</p>	<p>drop col: Address_Zip,Latitude,Longitude</p>



	<p>Address_Zip column and then shift + click on the Longitude column. This should highlight three columns. Click on Action and then Drop. The columns view is the quickest way to perform multi-column transformations.</p>	
6.	<p>Undo & Redo</p> <p>If you have discovered, at any point throughout your Wrangling process, that you added a step you did not wish to add, you can easily Undo & Redo steps.</p> <p>Select the Undo  button and undo the last two steps you added.</p>	