



Lesson 5 – Wrangling Tools

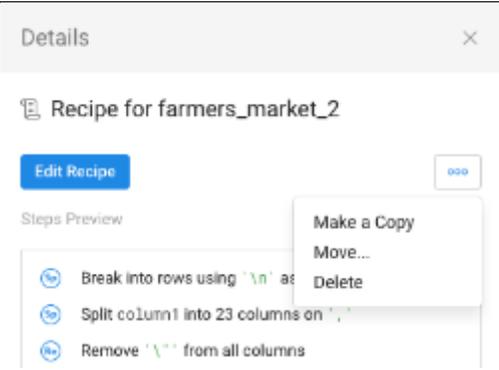
Chapter 5B | Aggregate | 5 Minutes

Chapter Goals

In this Chapter, you will:

- Use the Aggregate function to create data summarization

Chapter Instructions

YOUR GOAL: Compute the average number of market days for each state.	
<p>1. Create a copy of the Wrangled Farmers Market Dataset.</p> <p>You now want to build an aggregation on top of the current US Farmers Market dataset. However, an aggregation will change the granularity of the dataset, and you do not want to lose any of the individual market-level data. Consequently, you need to branch the dataset. You can do this by copying the current farmers market dataset.</p> <p>Return to the Training Flow details page. Click on the recipe icon for US Farmers Markets dataset, and click the  icon.</p> <p>Select Make a Copy. Trifacta will create a new dataset: Copy of US Farmers Markets. You will build the aggregation inside this new dataset. You can now see that you have two different recipes, and two different Wrangled datasets branching off from the source dataset for US Farmers Markets.</p> <p>Click on the Copy of US Farmers Markets dataset.</p> <ul style="list-style-type: none"> • Notice that the new dataset contains the same recipe and data as the original dataset. 	
<p>2. Aggregate data by state.</p> <p>In the Choose a Transformation box select the Aggregate transform.</p> <p>The Functions section is where you can aggregate your data with mathematical functions such as sum, average, max, etc. In this section let's find the average number of farmers' market occurrences for the Spring by state.</p> <p>In the Functions box choose Average from the aggregate functions. Then choose the Spring_Occur_Total, which we created in the</p>	<pre>aggregate value: average(Spring_Occur_Total) group: Address_State</pre>



	<p>previous lesson, from the columns list.</p> <p>In the Group by box choose the Address_State column.</p> <p>Notice the preview in yellow above the Builder. You can scroll through to preview the results.</p> <p>Add the Aggregation to your recipe.</p>	
<p>3.</p>	<p>Aggregate data while maintaining the details of your dataset</p> <p>If you want to maintain the details of your dataset, you can aggregate using the Derive transform. Click in the recipe above the Aggregate step we just created and insert a new step. Choose a Transformation, select Derive, and like aggregate, we are going to choose the AVERAGE function on the Spring_Occur_Total column, then group by Address_State. You'll see in the preview we get the same values as the Aggregate, but in a new column, while maintaining the details of the dataset. Click Cancel in the Transform Builder</p>	<p>Derive value: average(Spring_Occur_Total) group: Address_State</p>