Course Overview

Hi everyone, my name is Adam Crahen, and welcome to my course, Manipulating Data in Tableau Desktop. I am the head of data visualization engineering here at Pluralsight, and I'm also a 2018 Tableau Zen Master and the cofounder of a data visualization blog, thedataduo.com. Tableau is one of the world's most popular data visualization tools and is consistently ranked as a leader in Gartner's Magic Quadrant for BI and Analytics Platforms. Tableau helps people see and understand their data, but you need to know the why and the how behind data and analytics in order to be effective. In this course, you will learn how to transform your data before you drag and drop. Some of the major topics we will cover include organizing your data in the data pane, learning techniques for exploring your dataset, and taking a deep dive on Tableau's order of operations. By the end of this course, you will have all the skills and knowledge of manipulating your data to build the right analysis for the job. Before beginning this course, you should be familiar with Tableau Desktop's primary purpose, have a general knowledge of navigating through the tool, and know how to connect to your data. From here, you should feel comfortable diving into Tableau with courses on building common chart types, using calculated fields, and data visualization best practices. I hope you'll join me on this journey to learn about data visualization with the Manipulating data in Tableau Desktop course, at Pluralsight.

Organizing your Data

Overview

Hi everyone, and welcome to my course, Manipulating Data in Tableau Desktop. My name is Adam Crahen, and I am the head of data visualization engineering here at Pluralsight and a 2018 Tableau Zen Master. I am also the cofounder of a popular data visualization blog, thedataduo.com, which I cofounded with two-time Tableau Zen Master, Pooja Gandhi. You can also follow me on twitter @acrahen. In this module, you will learn how to organize your data in Tableau, which includes an introduction to the Data pane, understanding data field roles and types, how to organize and customize your fields, and finally, we'll cover tips on how to document your Tableau workbook. For this course, it would be helpful if you already have some familiarity with the Tableau Desktop interface and know how to connect to your data in Tableau. I will be using Tableau Desktop 2019.2 in this course. If you are on an older version of Tableau, the content of this course is still applicable. And finally, please download the exercise files so you have all the files used throughout this course at your fingertips. As we cover topics in this module, I want you to think about how you would feel about inheriting someone else's work. If their work is nicely organized, you probably feel pretty good about things. However, if upon opening someone else's workbook you encounter chaos, I would bet you wouldn't be feeling so good about things. This module will teach you some ways to organize your work for success.

Introduction to the Data Pane

In this clip, you will learn how data is organized in the Data pane within Tableau Desktop. When working with Tableau Desktop, all visualizations begin with the Data pane. The Data pane in Tableau organizes our data for us by displaying connected data sources, segmenting our fields into dimensions and measures, displaying any sets that we have in our data, or parameters that have been created in the workbook. Let's take a deeper look at each of these. Data sources are pretty simple. It is your data, and these are the files or databases you connected to that contain the underlying data you want to analyze. For more information on connecting to data sources, check out the Collecting and Preparing Data for Tableau Desktop course. Dimensions contain qualitative values like names and dates and are used to categorize or segment your data. Dimensions affect the view level of detail. In contrast, measures contain quantitative values, which are numeric and can be aggregated. A set is a subset of your data that is based on some condition. Sets can be created in Tableau, but they are also used by actions, set actions, and viz in tooltips. We'll cover these in more detail later in the course. And finally, parameters. These are dynamic values, or in other words, a manual input, which a user can change to replace constants within calculations, sets, filters, and reference lines. And with that, now you know how all your fields are organized in the Data pane.

Data Field Roles and Types

In this clip, you will learn about Tableau's data field roles and types. When we connect to any data source, Tableau analyzes that data source and assigns each data field to the Dimensions area or the Measures area of the Data pane, depending on the type of data the field contains. Tableau assigns a data type to each field and assigns a role of discrete or continuous. To take a closer look at data types, Tableau analyzes the contents of each field and assigns a data type. These data types could be numbers, decimals or whole; dates with or without timestamps; strings, which are text field; or Boolean fields, which are true false. These data types can be changed later. In my opinion, the concept of data roles in Tableau is one of the most important foundational concepts to understand. This concept defines what Tableau will draw on the canvas as we drag and drop. There are two roles in Tableau, discrete and continuous. These are mathematical terms. Discrete means individually separate and distinct while continuous means forming an unbroken whole, without interruption. Let's dive into each of these in more detail. In Tableau, discrete fields are represented as blue pills. Tableau treats discrete values as finite. We can drag a discrete pill to our Column or Row shelf, and Tableau will add headers to our view. Here is an example showing a discrete dimension with values 1 through 10 on the Columns shelf. Each member of this dimension has its own header. We can draw borders between all the values because headers create panes for each member. We can see Tableau drew our marks within the pane of each header. In Tableau, continuous fields are represented as green pills. Tableau treats continuous values as having an infinite range. When we drag a continuous pill to our Column or Row shelf, Tableau will add an axis or axes to the view. Here is an example showing a continuous dimension with values of 1 through 10 on the Columns shelf. Notice Tableau drew an axis and plotted our marks along that axis. Tableau automatically adds padding on either side as we can see by the presence of 0 and 11 on the axis. This view results in a single pane, and we cannot draw borders between the values. I want to reiterate something here. I frequently encounter people who have learned that blue pills are dimensions and green pills are measures. This is completely false and needs to be unlearned. Blue pills do not equal dimensions, and green pills do not equal measures. On the last two slides, I showed how the same field could be either discrete or continuous. If your data contains location fields, Tableau may assign a geographic role to your data. When a geographic role is assigned, Tableau will geocode your data, or in other words, it will assign latitude and longitude coordinates for the field so it can be mapped. These geographic roles can also be changed later. Tableau will assign geographic roles automatically if your field names match a level. For example, if you have a field named Country, Tableau will identify that and assign the geographic role automatically. If your fields do not get automatically assigned a role, you can click on the field type icon next to the field in the Data pane and pick a geographic role from the list. Once a field is assigned a geographic role, you will notice that the field type icon has changed to a globe, and Tableau has geocoded your data and added generated latitude and longitude fields in your Measures pane. For more detail on geographic roles and mapping, check out the Mapping Spatial Data course. Now that we know what data types and roles are, in the next clip, we will learn how to change data roles and types and how this affects what Tableau draws on the canvas.

Changing Data Field Roles and Types

In this clip, you will learn how changing data field roles and types influences what Tableau draws on the canvas. Now let's jump over to Tableau and take a closer look at data field roles and types. Okay, here we are back in Tableau using the Tableau workbook from the exercise files. Open the Clean Data Types tab. Here we are connected to a data source that contains six fields, each with a different data type and two rows of data. I put all six fields on the Row shelf so we can see the headers for each field and the values. Let's take a look at how Tableau assigned our fields in the Data pane. Look at the dimensions in the Data pane. Notice the icons next to each of our field names. Tableau analyzed our data and determined the data type for each field. Our Boolean field contains true/false values and was assigned the Boolean data type as indicated by the true/false icon. Our Date field was assigned a date data type as indicated by the calendar icon. The Date Time field was assigned the date time data type as indicated by the calendar with a clock icon. And our String field was assigned the string data type as indicated by the Abc icon. Not only did Tableau assigned the data types, but all of these fields were assigned to the Dimensions pane, and the default role for each of them is discrete as indicated by the blue color of the icons. Now look at our Measures pane. Tableau assigned our number fields, both the decimal and whole number fields, to the Measures pane because numbers can be aggregated. Notice the icon next to each of these is the pound sign, but if we click on the icon next to the decimal field, we can see Tableau assigned this field the Number (decimal) data type. If we click on the icon next to the Number (whole) field, we can see Tableau assigned it the Number (whole) data type. In this menu, we can see that the Default data type is what Tableau assigned, but we can select another data type if Tableau got it wrong. Now, not only did Tableau assign these field's data types, but they were assigned to the Measures pane, and the default role for each of them is continuous as indicated by the green color. Now let's switch gears. Click on the Messy Data Types tab, click on the Messy Data Types data source, and then click on the View Data icon. Here, we can see that we have 10 rows of data. We have a record ID, a year, country, and sales fields in our data source. Number of records is automatically generated by Tableau. Let's close the window. Tableau has assigned all of our fields to the wrong place. We're going to fix it. From the Measures pane, double-click on the ID field. Notice Tableau has assigned this field a numerical data type and a default role as a continuous measure. When we double-clicked, Tableau drew an axis and is trying to aggregate our data by summing it on the Rows shelf. The record ID field is a discrete value and shouldn't be summed. Click on the back button. Now right-click on the ID field and select Convert to Dimension. Tableau has now assigned this field with a default role as a discrete dimension. If we double-click on it again from the Dimensions pane, we can now see that Tableau is correctly drawing headers for this field. Okay, hit the back button. From the Dimensions pane, drag Name to the Marks card. As we hover over the mark, we can see that Name is actually a country. Tableau has assigned this field as a string data type in a default role as a discrete dimension. That is actually correct, but we also want to assign a geographic role. Tableau did not recognize this field as country automatically due to the name. Hit the back button. Now click on the icon next to Name, hover over Geographic Role, and select Country Region. Notice the icon next to Name has changed to a globe indicating this value has an assigned geographic role. Now drag the Name field to the Marks card. Now we can see that Tableau has automatically drawn a map, and it has generated latitude and longitude coordinates for us because that did not exist in our data source. Okay, once again, hit the back button. If we remember when we looked at our raw data, we had a numerical sales value, but Tableau has assigned the Sales field as a string data type and a discrete dimension. Double-click on Sales. Tableau has drawn headers and displayed our sales values because it is discrete. However, we really want to aggregate our sales values. Hit the back button, click on the icon next to Sales, and change it from String to Number (whole). Now you can either right-click on the field and select Convert to Measure, or you can simply drag the field from the Dimensions pane to the Measures pane. Now our Sales field is assigned a numerical data type, and the default role is a continuous measure. If we double-click on Sales again, we can see that Tableau is now drawing an axis and has aggregated our sales values. When we hover over the mark, we can see the total of $147, 000 in sales. Okay, let's hit the back button again. From the Dimensions pane, double-click on Year. Tableau has assigned the field as a string data type with the default role of discrete dimension, so Tableau has drawn headers with our individual values. If we click the back button again, we can change our year value to be numerical. Click on the icon next to Year, and change it to the Number (whole) data type. Right-click on the field and select Convert to Continuous. Notice we have a continuous field in our Dimensions pane as indicated by the green pound sign next to the field name. Double-click on this field, and notice that Tableau has now put this field on the Columns shelf and not the Row shelf. It has drawn an axis for our years and plotted marks for each year in our data source. To demonstrate the concept of discrete versus continuous one step further, double-click on the Sales field to add it to the Row shelf. We now have two continuous fields drawing two axes and showing sales totals over time. I can click on the axis to highlight the pane of data. In this view, we have a single pane of data. In the bottom left corner, we can see that we have one row by one column. Now, right-click on the Year field on the Columns shelf and select Discrete. Notice Tableau has switched the mark type to bars. We now have a header for every year. I can click on 2019 to highlight the pane of data. Notice only the 2019 bar was highlighted. If we look at the bottom-left corner again, we now have 1 row by 10 columns of data. In other words, we now have 10 panes of data in our view. We can change the mark type back to a line, but we still have discrete years in this view with headers. I encourage you to explore your data in a similar way until you have a solid grasp on how the difference between discrete and continuous affects what Tableau will draw on your canvas. One last note. Your data source, like multidimensional cubes, may limit your ability to change data rules. If you can extract your data, you should be able to regain all of Tableau's functionality.

Organize and Customize Fields

In this clip, you will learn how to organize and customize your fields in the Data pane. Now, let's jump into Tableau and learn how to organize and customize our fields. In this demo, you will learn how to group fields, sort fields, find fields, rename fields, combine fields, hide fields, and create calculated fields. So here we are back in Tableau. Go ahead and click on the Organize and Customize tab. Here we are connected to the Sample - Superstore data source that ships with Tableau. I have joined the Orders and Returns tables on Order ID and the Orders and People tables on region. First, we are going to talk about grouping fields. If we click on the caret of the Data pane, we can see that we have two options for grouping, Group by Folder or Group by Data Source Table. When you connect to a data source with a single table, grouping by folder is enabled by default. When you connect to a data source with multiple tables, grouping by data source table is enabled by default. Let's create a new folder for customer. From the Dimensions pane, click on Customer ID, and while holding down Command if you're on a Mac or Ctrl if you're on a PC, click on Customer Name to select both fields. Right-click on one of the pills, and hover over Folders, and select Create Folder. Enter Customer for the name of the folder, and click OK. Notice Tableau has created a folder in the Dimensions pane that we can expand or collapse by clicking on a caret next to the folder. Now, let's say we want to add segment into the Customer folder. You can right-click on the field and hover over Folders, hover over Add to Folder, and select Customer, or you can simply drag the field into the folder box. Notice how Tableau draws a light box around the fields within the folder as you hover over them. Let go of the field, and it has now been added to that folder. Let's add another folder for Order. In the Dimensions pane, right-click on the white space and select Create Folder. Enter Order for the folder name, and click OK. Notice we have an empty folder in our view. We can click on Order ID and drag it directly into the Order folder, or we could right-click on Order Date, hover over Folders, hover over Add to Folder, and select Order. Again, we can collapse this folder by clicking the caret next to the Order's folder. Now let's click on the caret at the top of the Dimensions pane and change our data source to Group by Data Source Table. Now our fields are organized by the data source table name. For example, all the fields that came from the Orders table are grouped together unless they have been organized in a custom group or hierarchy, which we will cover in a later clip. We can swap between the Group by Folder and Group by Data Source Table options at any time. Now let's talk about sorting fields. If you have a data source with many fields, you may want to sort the fields so they can be found easier. Again, click on the caret in the Dimensions pane. Currently, we are set to sort by field name. This means the fields are sorted alphabetically within our current grouping either by folder or by data source table name. Notice how Category is at the top of the list of the Orders table. Now click on the caret, and change the sort to Sort by Data Source Order. Notice how the fields have moved around, and Row ID is now at the top of the orders table. This sort is the order of the columns within our data source. Now let's talk about finding fields. If we change our data source back to Group by Folder and collapse our folders, it isn't as easy to see which fields are in our view. If you have a large data source, you may not want to keep scrolling or expanding folders. You can simply click on the magnifying glass icon at the top of the Dimensions pane. If we start typing customer in the window, Tableau will filter the Dimension pane to just the fields that contain the text. For example, if we type name in the search box, Tableau will show us Customer Name and Product Name. Simply click the X on the search box to return to the full-field list. This can really come in handy when you're working with a really large data source. Now let's talk about renaming fields. Sometimes you may have to rename fields within your data source. This could happen for a variety of reasons. For example, maybe the names come from a database, and everything has underscores in the name, or they are very lengthy. You may just want to change them for presentation purposes. Another reason might be just to change it for ease of use. A field name may be recognized by another name within your organization, or maybe you want to order the fields in a certain order and use prefixes to sort them alphabetically like I have done with my data sources. Let's rename the field. From the Dimensions pane, double-click on category to add the field to the Rows shelf. Notice Tableau automatically places the field name as a label above our headers. I usually turn these off, but sometimes it can come in handy when designing a text table. Let's rename Category to Department. You can either right-click on the field and select Rename, or you can click on the field and hold down for a few seconds until the pill changes so that you can rename the field. Type in Department. Notice how Tableau has updated our field name and the field label within our view. To revert to the default name, right-click on the field and select Rename. Click on the little arrow on the right of the pill, and then press Return on the keyboard, and the name will revert to the original field name. Renaming the field will not change the field name in your data source. Tableau stores this name just for use within Tableau. Now let's talk about combining fields. Combining fields is used to create a cross product of members from different dimensions. Let's demonstrate this by selecting both Category and Sub-Category. Right-click on one of the pills, and hover over Create, and select Combined Field. Tableau will create a new field in our Dimensions pane called Category & Sub-Category. Double-click on this field to add it to the Rows shelf. Here we can see that the two-dimension values were concatenated together using a comma delimiter. We can change the appearance of this combined field. Right-click on this field and select Edit Combined Field. A new window will appear. We can rename the field to something more friendly like Product, and we can change the delimiter from a comma to something else like space, bar, space. We can click apply and see the results behind the window. There is a 10 character limit for that delimiter. There are some limitations to using a combined field. For example, you cannot count distinct members of a combined field; however, you could do this using a calculated field. Now let's talk about hiding fields. There are a few reasons to want to hide a field from the Dimensions pane. For example, if you are working with a large data source, you may want to reduce the fields that you have to scroll through. Another reason might be that you want to reduce the size of the extract you are working with. When you extract a data source, hidden fields are automatically removed from the extract. Let's look at the field Order ID. We have the original Order ID from our Orders table, but we also have Order ID from the Returns table if the order had a return associated with it. We don't really need the Order ID from the Returns table because we have it from the original order. Right-click on Order ID (Returns) and select Hide. Notice the field is grayed out now in the Dimensions pane. If we right-click in the white space of our Data pane we can see that we have the option to show hidden fields checked. Click on that option, and now our hidden field has been removed from the Data pane. You could click this option again to show the field. However, if you extract the data source, the field will not be included in your extract. Right-click on the data source and select Extract Data. A new window will appear. Click Extract at the bottom to rerun the extract. Now right-click in the white space of the Data pane, and notice the option to show hidden fields is grayed out. Our hidden field has been removed from the extract. To get it back, we need to go back to our live data source. Right-click on the data source, and click Use Extract. We have now reverted to the live data source as indicated by the single database icon in our Data source window. Now right-click in the white space again. Select Show Hidden Fields. Notice our Order ID (Returns) field is back, but it is grayed out. Right-click on the field and select Unhide. Now we can re-extract our data source to get the field back in our extract. Let's try one more thing here. Double-click on Category and Sales from the Data pane. Here we have a simple visualization. Right-click in the white space of our data pane and select Hide All Unused Fields. Notice all of the fields except for the ones we use in our view have been grayed out. We can select the option to show hidden fields to cut our list down to just what we used. If we now run the extract once again, this would be much smaller than if we included all of the fields from the entire data source. Finally, let's talk about creating calculated fields. Tableau has an extensive ability to create custom calculations. There are many ways to create a calculated field. One way is to click on the caret at the top of the Dimensions pane and select Create a Calculated Field. Tableau will open a new window. You can name your calculated field here. There is a caret on the right side of this dialog box that opens a context menu of all of the functions that can be used within your calculations. I'm going to name my calculation Pi and enter the function pi. Now close the window. I am going to Option+drag this field to Text on the Marks card and select MINIMUM from the Drop Field window. This will display the value of pi on the canvas. Calculated fields can be renamed and organized just like any other field from our original data source. One last note. Your data source, like multidimensional cubes, may limit your ability to organize and customize your fields. If you can extract your data, you should be able to regain all of Tableau's functionality.

Documentation

In this clip, you will learn how to document your Tableau workbook. Documentation is really important in a business setting. This enables a clean handoff of your project to other resources, but it is also important for your own personal reference. You may forget exactly how you calculated a figure months down the road. All of this is to provide metadata about your data source for yourself and others. If you don't know, metadata is data about other data. Here are some methods for capturing metadata within your data source. Group your data by folders. It is much easier to work with a large, organized data source than just hunting around for data. While grouping data in folders, be consistent with naming conventions of your fields. We already learned how to do this in the previous clip. Comment your fields. This is an opportunity to inform users about where the data comes from and any important context they might need when using your data source. Comment your calculations. This is your opportunity to embed important business logic used to derive a calculation in the data source. These are good practices for your own working habits, but they become very important when you're the one curating a data source that will be published for others to consume. Embedding this metadata will result in you answering less questions. Now let's jump into Tableau and learn some tips about how to document your Tableau workbook. In this demo, you will learn how to comment fields, comment calculations, and use the describe fields option. Okay, here we are back in Tableau. This time, click on the Documentation tab. Here we are connected to a very organized version of the Sample - Superstore data from the Orders, Returns, and People tables. Notice right off the bat that all of my fields are organized by folder. My dimensions are grouped into folders for Customer, Location, Order, Product, Sales Territory, and Shipping. My measures have been grouped into Data Source Measures and Calculated Fields. This makes it very easy to look for fields if I am unfamiliar with the data source. Let's expand the Calculated Fields folder in the Measures pane. In here, we can see that we have a calculated field as indicated by the equals sign next to the data type icon. If I right-click on Profit Ratio and select Edit calculated field, the calculation editor will appear. Within the calculation, you can enter comments for your users to better understand what is being calculated. You can do this by prefixing the text with two forward slashes. Notice I have more than one line of comments within this calculation. When properly commented, the text will appear gray. If you only have one slash, for example, the text will be displayed as black, and the calculation will become invalid. Let's add the slash back in here and click OK. Another way to add metadata to our data source is by commenting the field. Right-click on Profit Ratio, hover over Default Properties, and select Comment. A dialog box will appear where you can enter some text about the calculation. I have already entered the same comment I had in my calculation window, along with the formula. Click OK and close the box. Now hover your mouse over Profit Ratio. Notice the comment will appear. This is a great way to provide information, and it does not even require opening the calculation window. Let's explore another use case for comments. Here at Pluralsight, we have several certified data sources. We have a Data Source Information folder on these data sources. Let's go ahead and expand it. Here we have a calculated field. Let's right-click on the field and select Edit calculated field. Notice the calculated field is a blank calculation with two single quotes. This adds very little overhead to the data, but go ahead and close this window. If you hover over the field, you can see some important information like who owns the table, the refresh cadence, the source of the underlying data, and how the data was curated and joined. This is just another way of providing information to your users. All of this metadata will be published along with the data source if you publish it to Tableau Server. One more thing to know about the Data pane is how to describe a field. Right-click on Profit Ratio and select Describe. I'm going to double-click on the title bar to expand the window to a full screen. Here, Tableau will tell us the default role of the field, the data type. It shows the text that we added as a comment, and the Status tells us whether or not the calculation is valid. Below this, we can see the formula of the calculation. Notice it says the domain for this field has not been loaded. Click Load at the bottom to retrieve. Let's click on Load. Tableau is now telling us the formula results in a single value of 12%. This value is computed across the entire data source. Let's close this box. Now drag Profit Ratio to Text on the Marks card. Here we can see the 12%. Let's hit the back button. Now let's see how this window differs for a field from our original data source. Click on the search icon, and start typing category. Right-click on Category and select Describe. Notice how this window differs. The Describe function now tells us what table this database column comes from, the data type, and some other information about the field. Notice the Domain at the bottom tells us that the field has three members and shows us those distinct values. Let's close this dialog box. Now double-click on Category, and we can see the same three members on our rows. It does not take a long time to document your Tableau workbook. If you build this into your workflow while you are designing a view, it will become a good habit and one that someone else will praise you for if they ever inherit your work.

Summary

To wrap up this module on organizing your data, you should now have a solid understanding of using the Data pane, including how Tableau assigns data fields and roles to your data source. You should be able to change data roles and types if they are in the wrong place and know if Tableau is going to draw an axis or a header based on the role of the field you add to your view. You should now know how to organize and customize fields through grouping, sorting, combining, et cetera. And finally, you should know how to document and inspect your data through the use of comments and the Describe option. Here are some things to remember from this module. Discrete fields, or blue pills, draw headers, while continuous fields, or green pills, draw axes. Data types and roles determine what Tableau draws on the canvas. These can be changed if the default assignment is incorrect. And make sure to always document your workbooks so you can refer to your previous work or ensure your work can be handed off to another resource smoothly. I hope you join me in the next module where we'll discuss exploring your data.

Exploring your Data

Overview

Hi, this is Adam Crahen, and thanks for continuing along this journey of manipulating data in Tableau Desktop. In this module, you will learn how to explore your data, which includes an introduction to Tableau's generated fields, methods for exploring your dimensions and measures, how to create bins, and finally, how to explore distributions in your data source. As an analyst or someone who is communicating through data, we play the role of a translator. We need to communicate what happened, what is happening, and what will happen. We are often the last line of safety between being accurate and miscommunicating results. We need to explore our data to look for data pitfalls. There is no clean and perfect data. In this clip, you will be exposed to some techniques for exploring your dataset. This will help you get a better understanding of the underlying data before working on your final product.

Generated Fields

In this clip, you will learn about the fields that Tableau generates in your workbooks. So what are generated fields? These are the fields that Tableau automatically creates, and these fields are not from your original data source. In Tableau, there are five fields Tableau generates for you. These are Measure Values, Measure Names, Number of Records, Latitude (generated), and Longitude (generated). Let's take a closer look at each of these fields. First, we have Measure Values, which always appears at the bottom of the Measures section of the Data pane. This field contains all of the measures in your data source combined into a single field with continuous values. You can add and remove measures from the Measure Values shelf once it's added to your view. Next, we have Measure Names, which always appears at the bottom of the Dimensions section of the Data pane. This field contains the names of all the measures in your data source combined into a single field with discrete values. This field goes hand in hand with Measure Values. Once added to the view, you can add/remove Measure Names from the view by adding or removing a field from the Measure Values shelf, or you can simply filter by measure name. Next, there is Number of Records, which appears in the Measures section of the Data pane. This is a calculated field that adds the number 1 to every row of your data source. This field can be added to your view to aggregate the values of the field. This allows you to count the number of records in the data source and slice it by any dimension in the workbook. Finally, we have generated Latitude and Longitude fields. These fields appear when fields in your data source have been assigned a geographic role or when you are connected to a spatial file. Tableau automatically geocodes your data for use on a map. Now that we know what generated fields are, in the next clip, we will learn how to use them to explore your dataset.

Exploring Dimensions and Measures

In this clip, you will learn how to explore dimensions and measures in your dataset. I want to encourage you to build throwaway views, many, many throwaway views. Tableau is a fantastic tool for data exploration and getting quick insights into record counts, distributions, and granularity. When I get a new dataset, even if I curated the data myself, I spend some time to explore the data to make sure I understand what I am working with, so explore and discard the views. Now let's jump into Tableau and learn how to explore your dimensions and measures. Okay, here we are back in Tableau. Click on the Exploring Your Data tab. Here we are connected to the same Sample - Super source dataset that is using the Orders, Returns, and People tables. When you connect to a dataset, any dataset, the first thing you should do is double-click on Number of Records. If you remember, Number of Records is a generated field that Tableau creates to allow us to count the number of records. When we hover over the bar, we can see that there are 9, 994 marks in our view. Another way to keep track of what you are visualizing is to peek in the bottom-left corner to see the row and column counts, or panes of data. Tableau will also display a measure from your view in the bottom-left corner. Now 10, 000 rows isn't a lot of data. Let's create some throwaway exploratory views to get familiar with this dataset. Drag Number of Records to Label on the Marks card. Double-click on Category. Here we can see our three-dimension members. It looks like we have a lot of office supply data. Click on the Swap Rows and Columns icon at the top of the screen. I prefer to look at bar charts with long labels horizontally so I don't have to turn my head sideways, and the analyst in me is going to want to sort these bars. Click on the Sort Descending icon at the top of the screen. From the Dimensions pane, drag Sub-Category on top of Category on the Rows shelf to replace the field. Click on the Sort Descending icon at the top of the screen again. Here we can see that Binders has the most number of records, followed by Paper. Now drag Category onto Color. I would expect to see Office Supplies near the top of this sub-category view. By simply adding color to our view, we can see that 6 out of the top 10 sub-categories come from the Office Supplies category. Another thing we should notice here is that Category has a one-to-one relationship with Sub-Category, or Tableau would have created a stacked bar with multiple colors. For example, from the Dimensions pane, drag Region onto the Sub-Category pill on the Rows shelf to replace it. Here we can see how the number of records breaks down across each region and category. Because there are records from each category in every region, Tableau has created a stacked bar chart. Click on the Sort Descending icon again at the top of the screen. Now we can quickly see that West leads the way, and South is bringing up the rear in terms of record count. Let's remove Category from the Marks card by dragging it off the screen. Since we have Region on the Rows shelf, let's start looking at location. From the Dimensions pane, drag State to the first position on the Rows shelf. Here I notice two things right away. First, the most records are coming from California. I guess that's to be expected since the West region had the highest record count. Secondly, we have a one-to-one relationship between state and region. In other words, we do not have any custom regions that cross state borders. On the Rows shelf, let's move State to the second position. Click on the Sort Descending icon at the top of the screen. We can now see how the states rank within each region. Let's put these on a map and check out some of our other generated fields. Click on the Clear Sheet icon at the top of the screen. Drag Country to the Marks card. Notice Tableau has automatically added generated Latitude and Longitude to the Columns and Rows shelves. This geocoded data came from Tableau's assigned geographic roles. So, okay, we have just one country, and it's the United States. From the Measures pane, add Number of Records to Size on the Marks card. Click on Size, and drag the slider to the second hash mark along the slider. These hash marks are Tableau's recommended sizes. If you go higher than the second one, Tableau may have some overlapping marks at some point in your view. Now drag State onto the Marks card. Drag Region onto Color on the Marks card, and here we can see the same breakdown as our state bar chart, but we can also see how the marks are dispersed across the map. If we change the mark type on the Marks card, we could change this map to be a field map, but then we'd also need to move Number of Records from Size on the Marks card to Color; otherwise, we are just looking at a region map. Let's hit the Undo button twice. From here, drag Postal Code onto the Marks card. Now we are breaking apart the state counts into a more granular level of detail. We can also take a look at the filled map or a choropleth map by changing the mark type, but I'm not going to bother with this view because the points on our map show that not every postal code in the country has records in this dataset. The data is widely dispersed across the country. Let's clear the sheet by clicking the Clear Sheet icon at the top of the screen. We have some dates in here, so let's take a look at the granularity of this data. Again, double-click on the Number of Records to add it to the Rows shelf. If you were on a Mac, Option+drag Order Date to the Columns shelf. Right-click and drag if you're on a PC, and this will open a drop field window. Select the green, or continuous, YEAR of Order Date option. Now we know that there are four years of data in this dataset. Click on the plus on the Columns shelf to drill down to the QUARTER of Order Date. We can see that we have data at this level, and there is a repeating pattern for each year. Click the plus on the pill on the Columns shelf to drill down to the MONTH. Okay, we still have data at this date level, and we still have a repeating pattern at the monthly level. It looks like September, November, and December are strong months, but for some reason every October is horrible. Click on the plus sign again, and we can see that we still have data at a weekly level. Click the plus once more, and we can see that there is still plenty of data at the daily level. If we had wider gaps between the points of our lines, we would know that the data is more dispersed at a certain date level. Finally, let's explore our measures a little bit. Click on the Clear Sheet icon at the top of the screen. Double click on Measure Names, and add it to the Columns shelf. Notice Tableau has already added the Measure Values card to this sheet and added all of our measures. We can see the total discounts, number of records, total profit, total quantity, and sales amount for the entire dataset. We can't see profit ratio because Measure Values is currently formatted as a whole number, but we already know it's 12% from the last module. We can add measures to this list, like a distinct count of customer IDs. I'm going to Option+drag Customer ID to the Measure Values shelf and select the CNTD, or a distinct count, from the Drop Field window. Here I can see that we have 793 customers in this dataset. So, as you can see, we built many charts in this short clip. The total is somewhere in the double digits. I don't know exactly. I lost count of all the variations. One thing to note is that none of these views will make it onto any of my final dashboards. Everything here was ad hoc and exploratory and is a throwaway view. But as you can see, we learned quite a bit about this dataset just by exploring our data for a few minutes. Now, a good analyst wouldn't stop here, but I think you get the point. Build those throwaway views to learn as much as you can about your data before you visualize anything.

Creating Bins

In this clip, you will learn how to create bins in Tableau Desktop. But first, what is a bin? The process of creating a bin in Tableau creates a dimension from a measure. This takes a second to wrap your head around. We take our continuous measure with an unlimited number of possible values, and we chunk it into discrete bins that have a limited set of values by size. So why do we do this? To analyze a distribution. When we visualize data using bins, we can see the shape and spread of the data. We can look for missing values or gaps in our data, and we can look for outliers. Here is an example. Let's say we have 1, 000 orders that total X dollars. The highest sales amount in those orders was $1, 000. The lowest sales amount in those orders was $10. I want to create 10 bins of $100 and count how many of the 1, 000 orders are in each bin. Here is the number of orders that have a total sales amount between $0 and $99. Here is the number of orders that have a total sales amount between $100 and $199. And here is the rest of the distribution. This is called a histogram. It shows the shape and size of the distribution. In Tableau, we can build histograms using bins. Now let's jump into Tableau and learn how to create discrete bins from a continuous measure. Okay, here we are back in Tableau. Once again, click on the Exploring Your Data tab. In this clip, we are going to create some bins for our Sales and Profit measures. In the Measures pane, right-click on the Sales field. Hover over Create and select Bins. A new window will appear. First, we can name our field. The name is defaulted to be the measures' name with bin in parentheses. Here we can see the minimum value of sales is $0 and the max value of sales is $22, 638. The difference between these two values is $22, 638 again, and we can see there are 3, 590 distinct sales values in the dataset. Using these values, Tableau has suggested a bin size of $503. Let's update this to just be $500 so we're using a round figure. Now click OK, and notice Tableau has created a new dimension for us. The data type icon next to the pill is a histogram indicating that this field is a bin. Drag the Sales (bin) dimension to the Columns shelf. Here we can see these discrete bins in the $500 increments. Okay, now click on the Undo button, and let's create one more bin from our Profit field. But this time, let's parametrize the bin size. In the Measures pane, right-click on the Profit field, hover over Create, and select Bins. Here we can see the minimum profit value is minus $6, 600, indicating we lost money on one of our sales, the maximum profit value is $8, 400, and the difference between these two values is $15, 000. There are 4, 415 distinct profit values in the dataset. However, rather than using this adjusted bin size, I want to be able to dynamically change the bin size. Click on the drop-down next to Size of bins, and notice here we can create a new parameter, or we can select an existing parameter. Let's click on Create a New Parameter. Tableau has opened a new window allowing us to configure the parameter. Change the name to Bin Size. For the allowable values, the bin is already set to allow a range of values. Let's make the Minimum 0, set the Max to 200, and update the Step size to $25. This means users can select values in $25 increments. I am going to update the Current value to 100, and this is a dollar amount, so let's change the Automatic Display format to Currency (Custom) with no decimal places. Click OK to close the Parameter window. Now click OK to close the Bin window, and now Tableau has created our new discrete dimension of Profit (bin) in the Dimensions pane. Drag the Profit (bin) dimension to the Columns shelf, and here we can see that we have bins for every $100. Click in the box, and update the value to $50. Notice how the bin sizes on the Columns shelf have changed. Now let's type $85 into the parameter box and hit Return on the keyboard. Tableau has rounded the value to the nearest $25 increment since we set up a step size in our parameter and changed the value to $75. In the next clip, we will use these bins to explore distributions in our dataset.

Exploring Distributions

In this clip, you will learn how to explore distributions. Now that we know how to build bins, let's put them to use. Let's jump into Tableau and learn how to explore distributions using bins and histograms. So here we are back in Tableau, and we are once again on the Exploring Your Data tab. In the last clip, we created bins for Sales and for Profit. In this clip, we are going to explore distributions using these bins. First, let's take a look at the Sales (bin). From the Dimensions pane, drag the Sales (bin) field to the Columns shelf. What I'm interested in here is how our sales amounts are distributed across our data source. Drag Number of Records to the Rows shelf, and let's set our worksheet to Entire View. Now let's take a second to absorb what's in our view. Our original sales bins were hard coded to use $500 increments. The bin field on our Columns shelf is discrete, meaning that we have individual headers for each bin. If we look in the bottom-left corner of the screen, we can see that we have one row by 46 columns, so we have 46 bins visualized here. We can see that most of the records in this dataset are between $0 and $499. I wish we could break this down a little more, so let's try to break it down. In the Dimensions pane, right-click on the Sales (bin) and select Edit. Update the bin size to use our Bin Size parameter that we created in the last clip. Now our view has updated to look at sales in bins of $200 because that is the current value in our parameter. If we hover over the first bar, we can see that 7, 415 records have a total sales amount between $0 and $199. Let's confirm this. Right-click on the first bar and select View Data, and click on the Full Data tab. The default in Tableau is to show all underlying fields. There is a little-known checkbox at the top of this window that we can uncheck to only see the fields in our current view. Uncheck Show all fields so we can see just the Sales (bin) and the sales amounts. Here we can see that the sales values on all of these records is below $200, which is what we wanted to confirm. We can close this window. Drag Number of Records to Detail on the Marks card. Right-click on the field, hover over Quick Table Calculation, and select Percent of Total. If we hover over the first bar again, we can see that 74% of our records fall into this bucket. A couple of notes here. While we have some gaps in our bins, Tableau is showing the empty bins, so we get a sense of the spread of our data. If you right-click on the Sales (bin) field on the Columns shelf, we can see that the Show Missing Values option is checked. If we uncheck it, Tableau will show us just the bins that have data. This is handy if we want to inspect values. Change our Bin Size to $25, and now we can see how many and what percentage of our orders were between $0 and $24 or $25 and $49. Just be very aware that we have removed the gaps. This option should only be used for analysis. You should never publish anything or present any data with the Show Missing Values option unchecked. Let's change the Bin Size back to $200, and right-click on the Sales (bin), and recheck the option to show missing values. Another thing we can do here is to right-click on the Sales (bin) and make it continuous. This option will help us visualize a more traditional histogram where the bars are touching. Let's take a quick look at the Profit (bin) now. Remove the Sales (bin) from the Columns shelf. From the Dimensions pane, drag the Profit (bin) onto the Columns shelf. If we peek in the bottom-left corner, we can see that we have 75 bins of $200 showing the distribution of our profit. We did lose money on some products, so we have values that are negative. Right-click on the Profit (bin) and select Continuous. If we remove the Number of Records field from the Rows shelf, Tableau will automatically change our Mark type to Gantt bars, and now we can see the spread of our individual profit bins. This view is called a barcode chart or a strip plot. Now these are just a couple of ways to put the power of bins to use in exploring distributions in your dataset.

Summary

To wrap up this module on exploring your data, you should now know what fields Tableau generates in your data source, how to explore dimensions and measures in your data source, how to create bins, and how to explore distributions in your data using bins and histograms. Here are some things to remember from this module. Generated fields come in handy when exploring your dataset granularity. Explore your data before you build a finished product. Bins are used to chunk the values of a continuous measure into discrete buckets. Know and understand the distributions of measures in your dataset. I hope you join me in the next module, Transforming Your Data.

Transforming your Data

Overview

Hi, this is Adam Crahen, and thanks for continuing along on this journey of manipulating data in Tableau Desktop. In this module, you will learn how to transform your data, which includes sorting fields, creating groups, creating sets, and creating hierarchies. As I mentioned in the last module, there is no perfectly clean data, and sometimes we don't have access to correct the issues further back in the data pipeline where they should be corrected. We often need to adjust the data for presentation purposes. In this module, you will learn some techniques for transforming your data.

Sorting Fields

In this clip, you will learn all the different ways to sort your data in Tableau. So what is sorting? It means arranging your data in a particular order. In Tableau, sorting is a little bit different than other applications. We sort dimensions by data source order, alphabetically, by a field, usually a measure, or manually. I will show you some ways to sort that include options for users to sort data in a view after the view has been published to Tableau Server. I personally do not like these options, and I sometimes disable them so users do not break the view that I designed. I think as designers of visual communications we should be thinking about presenting our data in the most meaningful way and one that doesn't require our users to click on a bunch of icons and possibly distort that communication. Additionally, I will show you some ways to sort data from within Tableau Desktop as you're building a view. I will also show you how to dynamically sort a view multiple different ways using calculated fields and parameters. This will create a more dynamic and intuitive user experience for your users. Let's jump into Tableau and learn all the different ways we can sort our data to transform a view. We will cover user options for sorting, designer options, the difference between nested and non-nested sorting, and dynamic sorting via a parameterized calculation. Okay, here we are back in Tableau. Click on the User Sorting tab. Here we are connected to our Sample - Superstore dataset again using the Orders, Returns, and People tables. First, we are going to look at some options that your users would see if they were interacting with your views on Tableau Server. In this view, we are looking at our sales data. On the Rows shelf, we have discrete pills for Category and customer Segment, which are creating row headers. On the Columns shelf, we have a discrete pill for Region, which is creating column headers for each region. This pill is followed by a continuous measure, Sales, which is creating an axis within each region. The bars are encoding our sales amount and are colored by customer segment. For example, if we hover over the first bar, we can see the total amount of furniture sales in the Consumer segment for the Central region. Since we have dimensions on both the column and rows, a user can sort this view in multiple ways. First, let's start with field labels. If you hover over the Category field label, notice a hidden icon appears, and we now have the ability to sort this field alphabetically. One click on this icon will sort Category in ascending order alphabetically. Two clicks will sort it in descending order, and three clicks will clear the sort. If we hover over the Segment field label, we have the same sorting options available. However, if we hover over the Region field label, notice we do not get any options to sort this field. A user still has the ability to sort the Region field by hovering over the individual rows for segment, or in other words, the lowest grain of data on our Rows shelf. Notice instead of the alphabetical icon, the icon here is an ordered bar chart. If we click it once, we are sorting Region by the descending order of sales by the Category and Segment dimensions. If we hover over the Sort icon, Tableau will explain the sort as a nested sort, meaning that in each pane of data in this view the bars will be ordered in terms of largest to smallest values. The order of regions, however, will be based on the row that you picked. For example, because we clicked on Consumer in the Furniture category, the East region is first on the column shelf. If we click on the Sort icon on the Technology Consumer row, the West region is now first in the list. Notice that the sort icon was automatically cleared from the Furniture Consumer row. Now I want to explain the difference between the sort behavior of the fields on rows versus columns one step further. Remove Region from the Columns shelf. If you are at all familiar with SQL, the values on our rows act as a GROUP BY clause. For example, if I click on Worksheet, Show Caption, you can see the query would read something like select Category, Segment, and calculate the sum of sales from the Superstore dataset, and group the sum of sales by Category and Segment. Essentially, what this is saying is that the sum of sales bar is the total amount for each category and segment combination. Now, when we add the dimension to the Columns shelf, we are slicing this grouped value by region. Instead, drag Region to the Rows shelf. Now hover over the Region field label, and notice we can now sort the region alphabetically. Additionally, when we hover over the rows, we have no sort icons because we have no dimensions on columns. This just demonstrates that the way you build your view will impact how users can sort the fields from these icons. Finally, hover over the axis at the bottom of the screen. Notice we have a bar sorting icon. This indicates that we can sort by the measure. One click here is a nested descending sort, two clicks is an ascending nested sort, and three clicks clears the sort. Let's switch gears and focus on sorting from a designer perspective. Click on the Dynamic Sorting tab. I think it is confusing to expect a user to know that if they hover over various parts of a workbook they will get different options to sort. I would encourage you, as a designer of visualizations, to build good views that remove that need. I often hide field labels or float empty text boxes over headers and axes to disable some of these options. In this view, we have a simple horizontal bar chart because I don't want to turn my head sideways to read labels. These bars are not sorted, and it's starting to hurt my eyes. There are icons at the top of the screen that allow a designer to sort the view by ascending or descending order. Notice as we click these buttons Tableau has added a sort icon to the Sub-Category pill on the Rows shelf. This icon lets us know that a sort has been applied to this dimension. Now right-click on the pill on the Rows shelf and select the Clear Sort option to remove the sort. Instead of making your users hunt and peck for hidden sort icons, build better views from the start. Notice I have created two parameters, one to specify the measure in my view and another to control the sorting of the view. Click on a caret of the Measure's parameter, and select Edit Parameter so we can see how this is set up. Here I am using a float data type with a list of values. Notice that I'm using values of 1 through 4, but I'm displaying text strings for these values. I almost always do this for strings as it makes writing calculations easier. I have four options for Sales, Profit, Quantity, and Customers. Okay, go ahead and close this window. From the Measures pane, right-click on Measure, and select Edit. Here you can see the case statement I created. It basically says when the value of the parameter is 1, then I want you to sum the sales and so on. Notice that I have specified my aggregation in the calculation. The reason for this is because my fourth measure is Customers. I want to do a distinct count on Customer ID so I can show sub-category by unique customers. Now close this window. The Measure field has already been added to the Columns shelf. Double-click on the title. Notice here that I have also injected my parameter into my title text. You can do this by clicking on Insert and picking the parameter from the list. Tableau will show the display as text here and not the value. Okay, go ahead and close this window. Now, as I change my measure in the parameter, the title and value being visualized also changes. Click on the caret on the Sort parameter, and select Edit Parameter so we can see how this is set up. The Parameter window will now appear. Here you can see I have the same float data type and list set up, but my Display As values are to sort in descending, ascending, or alphabetically. Click OK and close the window. From the Measures pane, right-click on the Sort field and select Edit. In this calculation, I am using my Sort parameter. When the value is 1, or descending order, I want to sort the negative value of the Measure calculation we created. When the value is 2, or the ascending order, sort by the Measure calculation. The trick here is number 3, to sort Sub-Category alphabetically. In a case statement, all of the results need to be the same data type. We really need this field to be numeric. However, Sub-Category is a string, so I have wrapped our string in the ASCII expression which converts the first character of the string into an integer. Since our Measure calculation is aggregated, we also need to wrap this calculation in MIN. Close the window. Right-click on the Sub-Category pill on the Rows shelf and select Sort. In this window you have options for sorting your dimension, by data source order, alphabetically, by a field, manually, or nested. We are going to use the Field sort because our view is using multiple measures, and it is very dynamic. Leave the Sort Order on Ascending, and change the Field Name to the Sort calculation we have created. Notice Aggregation shows as Custom because we have already specified the different aggregations in our calculation. I also want to point out that the Sort field isn't even on our canvas, but we are still able to use it in our sort configuration. Now close the window. Click on Worksheet and Show Caption. Double-click on the caption at the bottom of the screen. Notice I have injected the Sort parameter into the text here as well. Close the window. Let's change our Sort to Ascending. Now let's sort Sub-Category alphabetically. Let's update our view to Profit Descending and Profit Ascending, and Customers Ascending and Customers Descending. As you can see, we can now dynamically control the measure that's being visualized, the text that's being displayed, and the order of the marks on our view, all with just a few clicks. To me, this type of user experience is much more intuitive than asking users to hunt and peck for hidden icons all over the screen. As a designer of visualizations, always take time to consider the user experience of your work, even if it means adding hidden complexity to your view. What I mean by hidden complexity is that all of these calculations are invisible to the user, but provide a better intuitive experience in transforming their view.

Creating Groups

In this clip, you will learn how to create groups and what they are used for. So what are groups? Groups are used to combine members in a field. This could be used to correct data errors. For example, you could group misspellings of a member into one field. Or it may be used for presentation purposes. For example, let's say you have 10 dimension members, but 2 of them are major. Maybe you want to show those two members individually and then aggregate the data for the remaining member into an Other member. Groups can be used in calculations. Groups are also used to create custom territories on maps. For more information about this process, watch the Mapping Spatial Data in Tableau Desktop course. Let's just into Tableau and learn how to create groups. Okay, here we are back in Tableau. Click on the Groups tab. Here we are connected to the same Sample - Superstore dataset that is connected to the Orders, Returns, and People tabs. In this clip, we are going to create groups for a few different reasons. First, in this view, we are looking at the number of distinct orders by our four different shipping modes. Let's say we really just want to look at counts by express and standard shipping. There are a few different ways to create groups. We could select the First, Second, and Standard shipping class modes directly from our view. A popup will appear, and we can select the paperclip icon to create a group. Notice our three members have been grouped into a single row. Tableau has created a new field in our Dimensions pane called Ship Mode (group), and this field has also replaced the Ship Mode field on the Rows shelf. Let's right-click on the new field in the Dimensions pane and select Edit Group. We can click on the caret next to the first group to expand it so we can see the members that the group contains. Click on that long, ugly group name. Click Rename at the bottom of this window, and call this group Standard. If we click Apply, we can see behind our window that we now have Same Day and Standard shipping members for this group. But if you remember, the requirement was to have two groups for express and standard shipping methods. Now, back in the Group window, click on Same Day and select Group. Name this new group Express and click OK. Now Tableau is showing the data for our two groups. If we drag the original Ship Mode field to the Rows shelf, we can see the individual members on our rows. Okay, let's remove the two fields from the Rows shelf. Now let's compare the number of orders for everybody in our dataset named Adam to the rest of the dataset. From the Dimensions pane, right-click on Customer Name, hover over Create, and select Group. Manually select the three customers with the name Adam and click on Group. Name this group Adams. Now, rather than selecting all the other members in this table and grouping them, we can simply click on Other at the bottom of this window. Now click OK. Notice we have a new Customer Name (group) in the Dimensions pane. Double-click on this field to add it to our Rows shelf, and now we can see that the Adams had 27 orders versus 4, 982 of the rest of our customers. You can use groups to correct for some data errors. However, groups are not dynamic, meaning that new Adams will not automatically be added to this group.

Creating Sets

In this clip, you will learn how to create sets and what they are used for. So what are sets? Sets are subsets of your data that are based on some condition. Sets are used to compare or ask questions about the members that are inside or outside of the set. When we talk about the types of sets, there are dynamic sets, which is a condition to define rules, or top to define limits, like top 10 or top 25, and there are fixed sets that are based on a manual selection of members from one or more dimensions. Tableau also uses sets behind the scenes when you set up dashboard actions or utilize Viz in Tooltips. We aren't going to cover how to use these in this course, but set actions extend Tableau's set functionality by allowing users to change or update the values in a set. This is a really powerful feature that will be worth your time to explore. Now let's jump into Tableau and learn how to create sets. Okay, here we are back in Tableau. Click on the Dynamic Sets tab. We are connected to a copy of the Sample - Superstore dataset, which is connected to the Orders, Returns, and People tables. In this view, we can see that we have Customer Name on Rows and Profit on Columns. We are going to create a subset of these customers based on profit. So our customer name is sorted by profit in descending order. In the Dimensions pane, right-click on the Customer Name field, hover over Create, and select Set. Let's name the set Top Customers by Profit. Now we have a few options here. We are on the General tab. We could manually select some customers from this list which would create a fixed set. A more common use case would be to create a dynamic set that will update as your data changes. Let's click on Condition. Here we could set a condition based on a field. For example, we could say create a set using the sum of profit for each customer that's greater than $500. If we hit OK here, as our data changes, users could be added or removed to the set based on their future order profitability. Another option is to just write a calculation. For example, we could create a set with the formula CONTAINS Customer Name Adam. This set would automatically add new customers that have Adam in their first or last name. Let's click on None at the top of this window. Finally, click on Top. Here you can set a limit using a field or by writing a formula. Let's click on By field. The current configuration is to show the top 10 customers by the sum of profit. Now this is great, but if you show someone a top 10 list, they will say, what does the top 15 look like, or 20, and so on. A good tip for you is that anytime someone gives you a top N number, create a parameter so it can be changed easily or even by the user themselves. So instead of using number 10, click on the drop-down and select the Top Customers parameter, which I already set up to allow any integer. Now click OK. This Top Customers parameter is currently set at 5 and has already been added to our sheet. Notice we now have a new section in our Data pane that contains our sets. The icon next to our set is two rings indicating a subset. If we drag our set to the first position on the Rows shelf, we can see that Tableau is telling us which customers are in the set and which customers are out. Update the value in our parameter to 10 and now 15. Notice how our set dynamically updates. There are a lot of ways you can use sets. We could drag the set from the Rows shelf to the Filters shelf and simply filter on In to keep only a certain number of rows in our data. Click on the Undo button at the top of the screen. Or we could get creative with calculations. In the Dimensions pane, click on the search icon and look for top. I have created a calculation that will result in labeling the top N customers with their actual name, and then we'll simply group all others into an Other bucket. Drag this field to the second position on the Rows shelf, between our set and Customer Name. Notice how everything in the Out section of our set has Other for the Top Customer Label. Right-click on the Top Customer Label calculated field in the Dimensions pane so we can see how this works. Sets are similar to Boolean fields when used in calculation. The first line of this calculation simply says IF Top Customers by Profit. This really means if top customers by profit equals N. If you were trying to reference the members that are out of the set, you would need to add the word not in front of the set. Let me click Apply, and behind the calculation window we can see that our top customers are now labeled with Other, and the out customers are labeled by their name. Let's remove not from the calculation and click on OK to close the window. So what can we do with this new view? Drag Customer Name off the Rows shelf, and notice we have now lost our sort because that was the dimension that was sorted in descending order. Right-click on the Top Customer field on the Rows shelf and select Sort. Sort by Field, sort Descending, and close the window. Now you have two options. If you drag the set off Rows, Other will move to the position since it has the highest sum of profit. Click on the Undo button. Another option is to simply right-click on the set field on the Rows shelf and select Show Header to hide the field. This field is still on rows partitioning our data, but now we don't need to look at the in and out text, and Other remains at the bottom of our view. To demonstrate one more thing, let's replace our label calculation with the Customer Name field on the Rows shelf. Maybe we also want to include a few Adams in our set. Click on the three headers in the Out section that contain Adam. Hover over the marks, and a popup will appear. Click on the Set icon. From here, we could create a new set or add or remove these members from a current set. Let's add them to the Top Customers set. After all, Adams are pretty good people. Notice how the Adams moved up and are now in the set. However, in the Data pane, right-click on our set and select Edit Set. Notice on the General tab we have individuals selected from the list. If we click on Condition, we have no conditions set up. And if we click on Top, our top N configuration is gone. Manually adding members to our set converted it from a dynamic set to a fixed set. Sets are really powerful. You should explore these more as they can come in handy as you work through multiple situations in Tableau.

Creating Hierarchies

In this clip, you will learn how to create hierarchies and what they are used for. So what are hierarchies? It is a way to organize dimensions that allows for a drill-down capability. Here is an example looking at a date hierarchy. We could start off by rolling our dates up to the year level, and then we could drill down to the quarter, and then we could drill down to the month. From there, we could drill down to the week, and finally, we could drill down to the day level. Let's jump into Tableau and learn how to create hierarchies. Okay, here we are back in Tableau. This time, click on the Hierarchies tab. We are connected to a copy of the Sample - Superstore dataset, which is connected to the Orders, Returns, and People tables. Remember the group we created for Ship Mode to compare express versus standard shipping modes? We have now been asked to create a hierarchy that allows us to drill from our main level into our row level shipping detail. Click on the magnifying glass, and type mode to find our two fields. Click on Ship Mode (group), and while holding down Command on a Mac or Ctrl on a PC, click on Ship Mode. Right-click on either pill, hover over Hierarchy, and select Create Hierarchy. Name the hierarchy Shipping and click OK. Notice how these fields are now organized in the Dimensions pane. There is a new field with a hierarchy icon that can be collapsed or expanded to show the members. Also, notice after we created the hierarchy a minus appears on the Ship Mode group pill on the Rows shelf. If we click on that, Ship Mode will disappear from the Rows shelf. If we now click on the plus on that field, Ship Mode will return. One thing I really don't like about hierarchies is similar to sorting in that your users are going to have to hunt and peck for little hidden icons. Hover over the field headers for Ship Mode (group), and a little minus icon will appear on a header. Your users will have to click this to collapse ship modes. However, hover again, and click on the plus to expand the hierarchy. Okay, let's create another hierarchy. Remove the two fields from the Rows shelf. We want to create a hierarchy called Location. From the Dimensions pane, select City and Country. Right-click on either pill, hover over Hierarchy, and select Create Hierarchy. Name the new hierarchy Location. Scroll down, and we can see our hierarchy is out of order. We can simply drag Country above City in the hierarchy order on the Dimensions pane. Now let's add our other locations. Right-click on State, hover over Hierarchy, Add to Hierarchy, and select Location. Let's repeat this process for Postal Code. Now, scrolling down to our Location hierarchy, we can see that our fields are not in the right order. Simply drag State above City in the hierarchy. Now drag the Location hierarchy to the Marks card. Notice that Tableau has automatically created a map because these fields have geographic roles assigned to them using our generated fields Latitude and Longitude. We can click on the plus of the Country pill to drill down to the State level, or we can click on the minus to roll back up to the Country level. If you wanted to remove the hierarchy, you could simply right-click on the hierarchy from the Dimensions pane and select Remove Hierarchy. Let's clear the sheet using the icon at the top of the screen and try one more. From the Measures pane, double-click on Sales to add it to the Rows shelf. From the Dimensions pane, Option+drag Order Date to the Columns shelf. Right-click and drag if you're on a PC. This will open a Drop Field window. Select the continuous YEAR of Order Date from the Drop Field window. Tableau has a built-in hierarchy for dates. Notice there is a plus on the field on our Columns shelf. Your users can hover over the continuous axis for Order Date and click on the plus that appears to drill down to quarter, or month, or week, or day. We can click on the minus to roll back up to the week level. These can be handy if your uses know these hidden icons exist and your data has the granularity to support the view. However, I've rarely used these in a business setting as this is a huge training issue, and it can also cause some issues if you're using table calculations. I would rather use a parameter to inject a calculation to change my date levels in a view or swap a dashboard view with different geo levels. These do allow for a drill down, but they aren't always as dynamic as they appear.

Summary

To wrap up this module on transforming your data, you should now know how to sort fields from both a user and a designer point of view, including ways to sort your views dynamically using parameterized calculations. You should know how to create groups to transform messy data into a nicer presentation, how to create sets to ask questions about subsets of data based on a condition, and how to create hierarchies so you can perform drilldowns to different levels of detail. Here are some things to remember from this module. Always consider the user experience when thinking how users may sort a visualization. Groups can be used in calculations like any other dimension. Sets are dynamic if set up with a condition rule or a top limit. And hierarchies allow users to drill down to different levels of detail. I hope you join me in the next module, Order of Operations.

Order of Operations

Overview

Hi, this is Adam Crahen. In this module, we will cover Tableau's order of operations. In this module, you will gain an understanding of the order of operations, and then we will walk through several examples, including converting a dimension filter to a context filter, converting a table calculation to a fixed level of detail expression, and finally, converting a dimension filter to a table calculation filter. Understanding the order of operations is like looking under the hood of a car. It seems really intimidating, but once someone shows us around, it all starts to make sense. Let's get under the hood of Tableau.

Understanding Order of Operations

In this clip, you will be exposed to Tableau's order of operations. So what is the order of operation? It is also known as Tableau's query pipeline, which dictates of the order of filters, calculations, and data blending in Tableau. You're probably still wondering what this all means, so let's take a deep dive into what we are talking about. This is Tableau's order of operations. The gray boxes indicate a filtering operation, the blue boxes indicate a calculation operation, and the plum box indicates data blending. It is really important to know what is on this chart, what these items mean, and how this plays out in Tableau. For example, in the last module, we created a set. That set is calculated in the fourth position in this chart. If we want to filter that set by a dimension in our data source, the dimension filter would happen in the fifth position in this chart, or in other words, after the set is calculated. This means that the filter is not impacting our set like we want or expect. We need to apply our filter before the set is calculated, which means we need to add it to context. By adding the filter to context, it is promoted, or moved up this chart to the third position before the set is calculated. Understanding the order of operations is the key to demystifying Tableau. There are so many ways this comes into play while working with Tableau. Having a good handle on this will really change the way you think about solving problems. We will walk through several examples in this module where order of operations comes into play.

Convert a Dimension Filter to a Context Filter

In this clip, you will learn how to convert a dimension filter to a context filter. As a reminder, here is the order of operations. I have grayed out several of the levels. In this clip, we are going to focus on converting a dimension filter to a context filter so we can use filter sets and fixed level of detail expressions. By converting our dimension filter to a context filter, we are promoting it up the order of operations in Tableau. Let's jump into Tableau and learn how to convert a dimension filter to a context filter. Okay, here we are back in Tableau. This time click on the Order of Ops Example 1 tab. We are connected to a copy of the Sample - Superstore dataset, which is using the Orders, Returns, and People tables. In this clip, we are going to look at filtering sets and fixed LOD expressions. First, let's take a look at sets. In this view, we are looking at the top five customers by the sum of profit. We have Customer Name on the Rows shelf and the set, Customer Profit, is filtered to only show the N, or in other words, we're only including the top five customers in our view. Now, here it is important to note that the sets are calculated before dimension filters in the order of operations. To prove this, let me filter by year. Right now we are looking at every year in the dataset. Notice we have five names, and Seth is the most profitable customer. If I filter to 2015, our view drops to only three records. This means that two of our top customers for our entire dataset did not have an order in 2015. Click on 2016. Now we only have four names. One of our customers didn't have an order in 2016, and notice Seth, who is our most profitable customer overall, has a negative $192 profit. So not only did we lose money on him, he was far from the most profitable customer in 2016. So why is this working like this? We told Tableau to include the top five customers in our view. It's because of the order of operations. To remedy this situation, we need to add our filter to context. Right-click on YEAR on the Filters shelf and select Add to Context. Notice the pill has the turned gray. This is your indication that it is now in context. Context filters occur before sets are calculated in the order of operations. We now have the top five customers for 2016. Natalie Webber was our most profitable customer. Click on 2015. Now we have the top five most profitable customers in 2015. Notice that neither Natalie or Seth are even in this list. Now click on All. Notice Seth is back to the top of the list, and Natalie doesn't make the top five when we look at all four years. Now let's take a look at filtering LOD expressions. Right-click on the YEAR filter and select Remove from Context. Notice our filter has moved back to blue, which lets you know that it is back to a dimension filter. Hold down Command on a Mac and Ctrl on a PC, and drag the Order Date from the Filters shelf to the Rows shelf to make a copy. Drag the discrete SUM of LOD Profit field from the Marks card to the second position on the Rows shelf. From the Measures pane, right-click on the LOD Profit and select Edit. This calculation returns the total profit at a customer level. If we look back at our view, Seth has $21, 633 of total profit, and now we can see that most of it came in 2018. If we filter by 2018 only, notice that our fixed LOD expression didn't change. Fixed LOD expressions are calculated at the same level as sets. Now right-click on YEAR again and select Add to Context. Notice that the LOD profit total is now the same as the sum of profit. The context filter now happens before the fixed LOD expression is calculated.

Convert a Table Calculation to a Fixed Level of Detail Expression

In this clip, you will learn how to convert a table calculation to a fixed level of detail expression. As a reminder, here is the order of operations, and I have grayed out several levels. In this clip, we are going to focus on converting a table calculation to a fixed level of detail expression so we can promote the calculation to move up the order of operations. This will allow us to filter dimensions without impacting figures displayed in the view. Now let's jump into Tableau and learn how to convert a table calculation to a fixed level of detail expression. Okay, here we are back in Tableau. Click on the Order of Ops Example 2 tab. We are connected a copy of the Sample - Superstore dataset, which is still connected to the Orders, Returns, and People tables. In this clip, we are going to look at the order of how fixed LOD expressions and table calculations are calculated. Notice in this view we are looking at the percent of total profit by subcategory. We have two columns of numbers that are exactly the same. The first column is calculated via table calculation, and the second column is calculated using an LOD expression. Right-click on the LOD Sub-Category % of Total, and take a look at the calculation. The first line, or in other words, the numerator, is summing profit across all of the rows that are in our view. The second row, or the denominator, is a fixed LOD expression that returns the total profit across our entire dataset. This value will not change unless a dimension filter is added to context, like we showed in the last clip. Now close the window. Our Sub-Category filter is not in context. The order of operations comes into play here because we want to filter by our Sub-Category dimension. In the order of operations, dimension filters come after LOD expressions; however, they come before table calculations are calculated. Let's focus on the first row of data. Copiers make up 21.2 % of our total profit. Let's uncheck the filter for Phones. Notice we now have a different value in our Table Calculation column. According to our table calculation, copiers makes up 25.1 % of profit while our LOD still says 21.2. Due to the order of operations, our LOD value remained unchanged because it was calculated before our dimension filter. Our table calculation was computed after the filter was applied. That calculation isn't wrong. It's just saying that copiers make up 25.1 % of the total profit in the view, not across the entire dataset. You can see how knowing the order of operations can come into play as you're attempting to answer more complex questions about your data.

Convert a Dimension Filter to a Table Calculation Filter

In this clip, you will learn how to convert a dimension filter to a table calculation filter. As a reminder, here are the order of operations, and I have grayed out several levels. In this clip, we are going to focus on converting a dimension filter to a table calculation filter. This process demotes the filter to move it down the order of operations. This will allow us to simulate a single select discrete dimension filter by using a table calculation. Let's jump into Tableau and learn how to convert a dimension filter to a table calculation filter. Okay, here we are back in Tableau. Click on the Order of Ops - Example 3 tab. We are connected to the same Sample - Superstore dataset, which is connected to the Orders, Returns, and People tables. In this clip, we are going to hack our way around using a single select discrete dimension filter and convert it to using a table calculation. In this view, we are looking at total sales by month in the year 2018. We are also computing the percent difference from the previous month. For example, in February 2018, our percent difference is - 19.6. This number is calculated by taking $36, 352 from February, subtracting $45, 234 from January, and dividing that total by $45, 234 from January. This calculation is a table calculation computing data across rows. We know from the last clip that table calculations are computed after dimension filters. Notice that MONTH of Order Date is on the Filters shelf. The filter is shown on the very right side of our screen. Let's filter by December 2018. Notice that our table calculation value is null. This is because we now only have December 2018 in our view, and the table calculation requires two months of data to calculate the percent difference. Let's click on All to show all of our data again. From the Measures pane, right-click on the Table Calc Filter and select Edit. This hack in calculation was created by Pooja Gandhi. If we take a look at the comment, the MONTH of Order Date field is calculated by truncating the order date to the very first day of the month. This field was then shown on Rows as a discrete field. Pooja's calculation uses a LOOKUP function, which returns a row from a table based on the offset specified. In this case, the offset is 0, which means it's a current row. This function requires the field to be aggregated, so it was also wrapped in a MAX aggregation. So what this means is that we're using a table calculation to look up the value from the current row. Now close the window. This field has been added to our Filters shelf, formatted to look the same way as our MONTH of Order Date, and it is currently set to show all views. Okay, take note that our December 2018 percent change is - 25.4 %. Click on December 2018 in our Table Calc Filter, and notice our percentage is still correct, and we only have one row of data in our view even though this particular row is computed across two rows of data. I am going to right-click on my canvas and select show the caption. In the caption, I have visualized what the final view of this table could look like. We are often asked to place big numbers on dashboards for key performance indicators. Now, this type of filter could be especially important when multiple views are placed on the dashboard and a single select filter option is required. We can't use a dimension filter because the table calculation for the month-over-month change happens afterwards. In this case, by using a table calculation as a filter, we have demoted the filter operation down the order of operations, and the filter now happens at the same time that the table calculation in our view is calculated. Notice, if we select December on our normal dimension filter, the KPI view nulls out. This would not be a good experience on a dashboard, and this was a really creative solution to solve this specific problem that could only be solved by using the order of operations.

Summary

To wrap up this module, you should now have a good understanding of Tableau's order of operations. We covered examples like converting a dimension filter to a context filter to move up the order of operations. We converted a table calculation to a fixed level of detail expression to move up the order of operations, and we converted a dimension filter to a table calculation filter to move down the order of operations. These examples should help you understand how these processes can help you achieve your desired results when working in Tableau. Here are some things to remember from this module. Context filters are required to filter sets or fixed level of detail expressions. Fixed level of detail expressions are not impacted by dimension filters. Table calculations are computed near the end of the order of operations. And you can move up or down the order of operations depending on your view configuration. To wrap up this course, we now have covered how to organize, explore, and transform your data, and we peeked under the hood of Tableau's order of operations. I want to thank you for coming along this journey with me. I hope this course on manipulating data in Tableau desktop gives you wings as you continue along your Tableau learning path.