MicroStrategy | 2023

Accessing Data Sources and Wrangling Datasets in Dossiers

Learning Path: Departmental Analyst

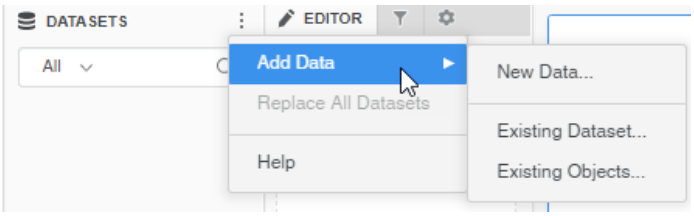
CONTENTS

1

IMPORTING DATA INTO DOSSIERS

Importing data from different data sources

You can use data from a variety of data sources in your dossier. From MicroStrategy datasets, personal spreadsheets, and cloud-based data sources like Salesforce to big data sources like Hadoop, you can import data from over 70 data sources into your dossier.



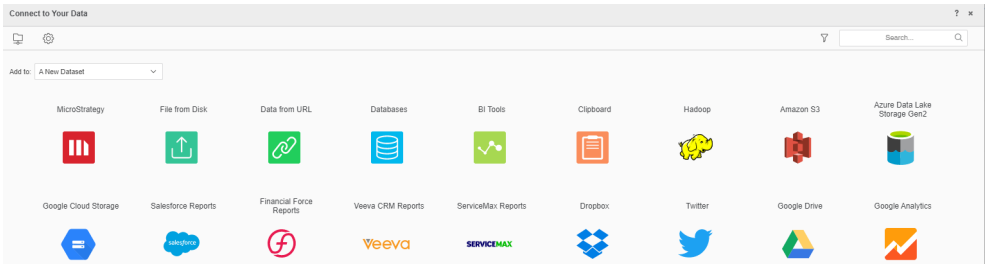
MicroStrategy can connect to enterprise resources that your business has already invested in, such as AWS and relational databases. You can use a single data source or combine different data sources in the same dossier by using multiple datasets or a single dataset. For example, you can import an Excel spreadsheet and a table from a relational database and combine them into a single dataset.

External data sources

You can import data from many of external data sources. Examples include:

- A file in a selected folder, URL, or file URI scheme
  - Excel files
  - Text files
  - Comma-separated values (CSV) files
  - Tab-separated values (TSV) files
  - Printer (PRN) files
  - JavaScript Object Notation (JSON) files
- A file stored on Dropbox or Google Drive
- A database
- A Hadoop file browser

To see a list of supported databases, scroll down on the Connect to Your Data page (displayed when you begin importing data).



You can select which data sources and databases are displayed on this page by clicking the **Customize Shortcuts** icon.

Best Practice

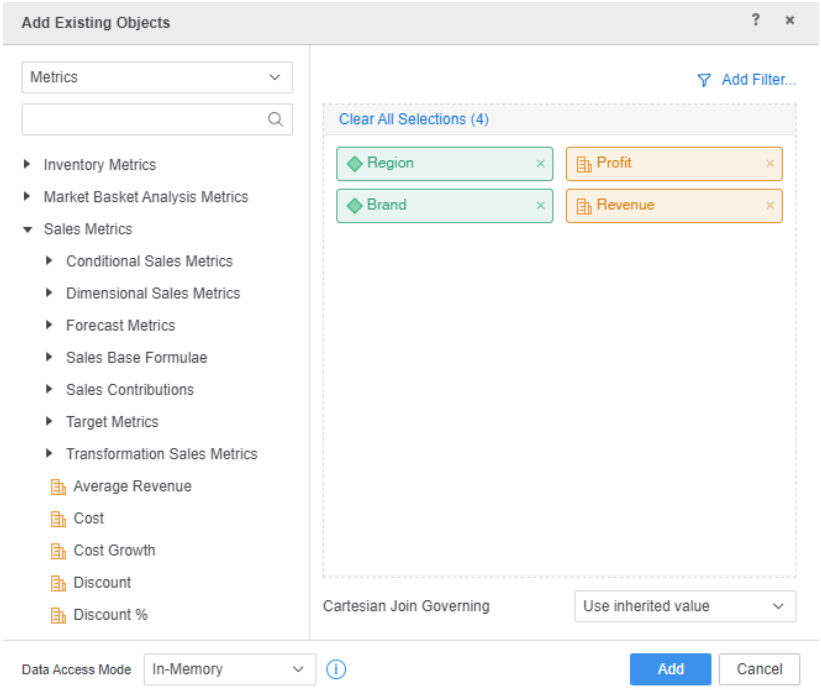
External datasets may not adhere to an organization's data governance standards. Accessing external datasets for utilization in dossiers is appropriate in an agile environment where users build and test their projects. Once they move a project to the departmental or enterprise environment, data should be subject to more rigorous data governance.

MicroStrategy datasets

Departmental and enterprise environments, due to their standards and guidelines, typically utilize MicroStrategy in-memory datasets that have been

certified for accuracy. Using MicroStrategy datasets in a dossier leverages the power of the enterprise platform to perform analysis on enterprise data. You can access data from MicroStrategy by adding any of the following to the dossier:

- An existing MicroStrategy report, including intelligent cube reports and MDX reports
- A report that you create within the dossier by selecting project objects such as attributes, metrics, and filters, as shown in the image below



- An existing intelligent cube, including imported data

An intelligent cube is a multi-dimensional set of data saved on a MicroStrategy server. You can share intelligent cubes as a single in-memory copy, to be used by many different documents and dossiers created by multiple users.

To learn more about in-memory and live datasets, see the Enterprise Analyst learning path.

Exercise: Access Library in the Education sandbox

The Education sandbox is a shared, cloud-based environment to help you learn about and experiment with MicroStrategy Web and Library.

Use the Education sandbox to complete a sequence of exercises related to importing data into a dossier and leveraging the data to create basic visualizations. Complete the steps below to access the sandbox.

Access Library in the Education sandbox

1. In your browser, navigate to **https://education.microstrategy.com/MicroStrategy/servlet/mstrWeb**.
2. Log in to the sandbox with your **MicroStrategy Account** credentials.

Open the Tutorial project

3. The MicroStrategy landing page opens. Click the **MicroStrategy Tutorial** project.

The MicroStrategy Tutorial home page opens.

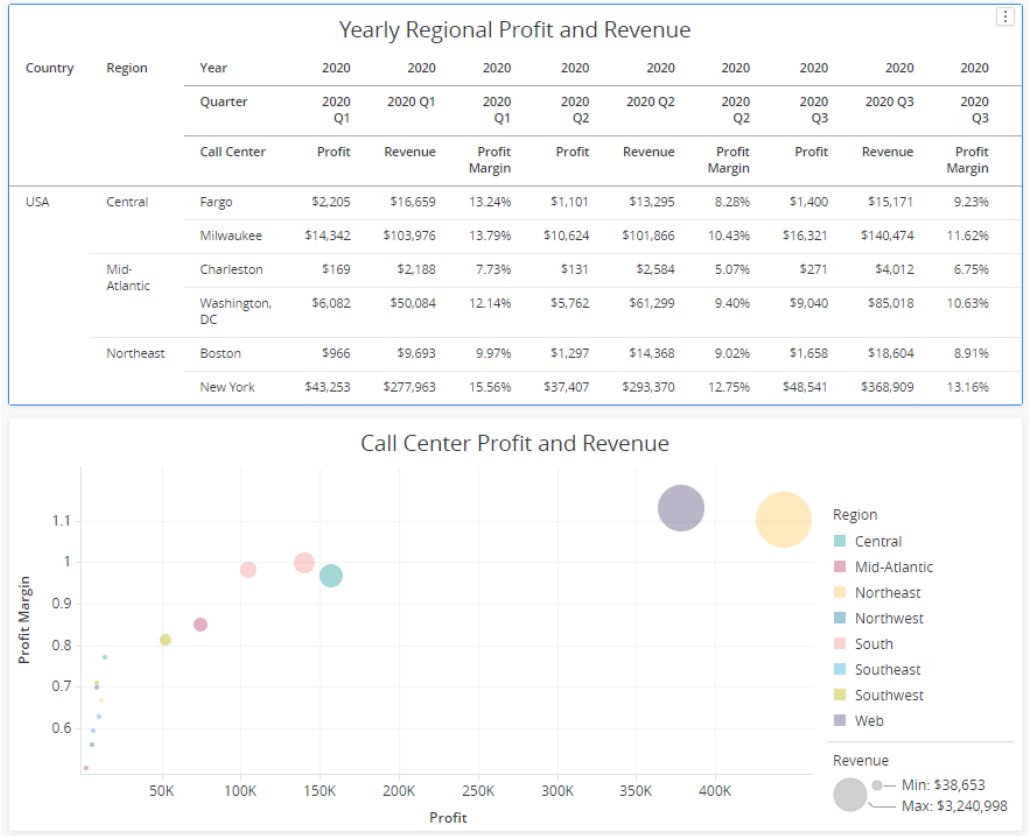
4. Click **Go to MicroStrategy Library**.
5. If prompted, enter your **MicroStrategy Account** credentials, and then click **Log in with Identity**.

Your Library home page displays.

Exercise: Import a MicroStrategy report and Excel data into a new dossier

You work for Radiant department stores as part of the analyst team. Your manager has provided you with the following mockup for the first page of a dossier:

Initial dossier mockup



In this exercise, you begin creating the dossier by importing data. Requirements for the first page include:

- A grid at the top, displaying detailed profit, revenue, and profit margin information for each call center for each quarter.
- A bubble chart at the bottom, displaying the minimum and maximum call center revenue values.

Your manager sent you additional analyses the dossier must include, which are:

- Employee contribution to total revenue and total profit. You know that a ring chart can be used to display contributions to a whole.

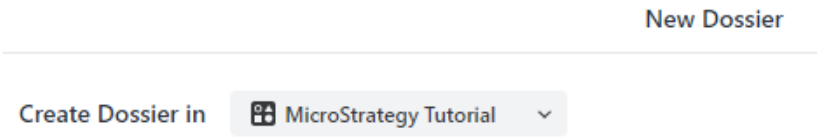
- The relationship between sales promotion types and purchase types. You know that a network visualization is an effective way to display relationships between items.
- Call center revenue and profit. You know that a map visualization is a good way to display data for geographic locations.

Many teams use MicroStrategy Workstation to build a dossier in an agile environment and deploy it to the enterprise or departmental environment in MicroStrategy Library once it is approved. However, for training purposes, the exercises in this lesson are completed in MicroStrategy Library.

Create a new dossier

You can create and edit dossiers in MicroStrategy Library.

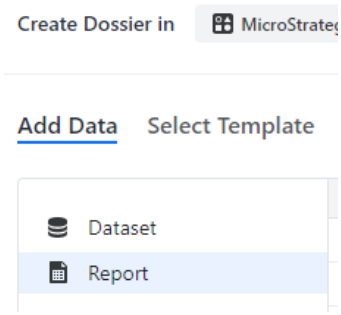
1. Access Library. For a reminder on how to do this, see *Exercise: Access Library in the Education sandbox, page 7*.
2. In the Library home page toolbar, click the **Create New** icon ⊕, and then select **Dossier**.
3. You want to create the dossier in the MicroStrategy Tutorial project. In the New Dossier window, in the **Create Dossier in** drop-down list, ensure **MicroStrategy Tutorial** is selected.



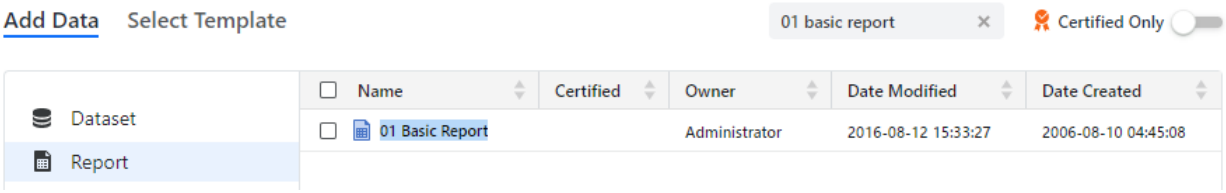
Import data from a MicroStrategy report

Add data you want to analyze to your dossier. Some of the data you need to create the ring chart and network visualization is available in a MicroStrategy report in the Tutorial project.

4 In the left pane of the New Dossier window, click **Report** to display a list of reports available to import into the dossier.



5 In the search box, type **01 Basic Report**, and press **Enter**. The report displays in the search results.

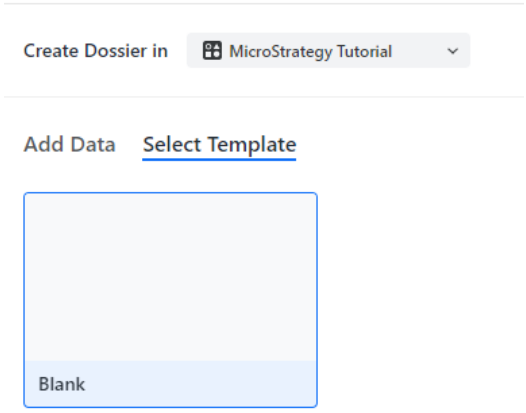


6 Select the check box for **01 Basic Report**.

Since the report objects exist in the project, they are already classified as attributes and metrics. This enterprise data has already been subjected to your organization's data governance reviews, so previewing and cleaning the data is not necessary or available.

7 Click **Select Template** to view your available templates. A dossier template is a structural foundation for building dossiers. A template can be as simple as displaying your company logo in the top right corner of every dossier. A more complex template can provide datasets to work with and various chapters

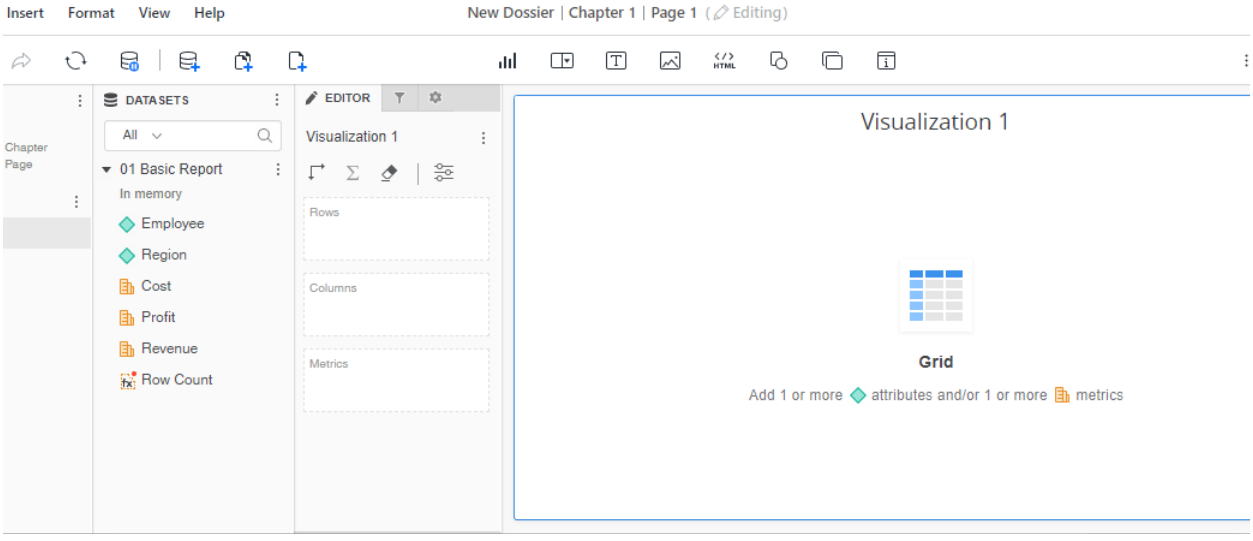
containing pages that already include blank visualizations. Templates can save you time when creating multiple similar dossiers.



Your administrator has not yet added any templates for you to use; the only available template is the Blank template, which provides a single blank grid.

8 Click **Create**.

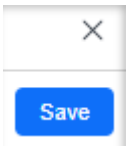
Your dossier is created and displayed, containing a single blank grid visualization. The dossier opens in Edit mode, as indicated by (Editing) after the dossier name above the toolbar. You use Edit mode to import data and create visualizations, filters, and more using that data.



01 Basic Report has been added to the dossier as a dataset. Its attributes and metrics are displayed in the Datasets panel. You can drag and drop any of these objects to create visualizations and to filter the visualizations based on the objects.

Save the dossier

9 Save your dossier by clicking **Save** in the far right of the toolbar.



10 In the **Name** box, type **Yearly Regional Performance**.

11 In the **Save In** drop-down list, select **My Reports**.

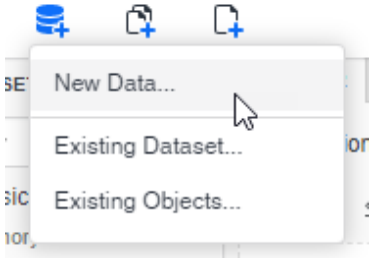
In the sandbox environment, you cannot save items in the Shared Reports folder.

12 Click **Save**.

Import data from an Excel file

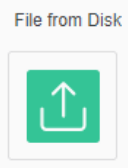
Your data can come from a variety of sources, including MicroStrategy data sources and external data sources. 01 Basic Report does not contain the data necessary to complete the grid and bubble chart, so you need to add another dataset to the dossier. Your manager provided you with an Excel file containing the data. Import the file into your dossier.

1 In the toolbar, click the **Add Data** icon , and then select **New Data**. Since the Excel file is saved in a file external to MicroStrategy, you are adding new data to MicroStrategy.

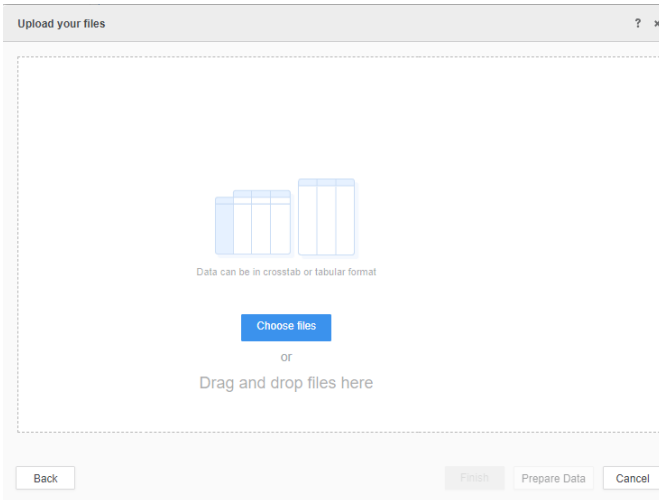


If you had not added 01 Basic Report when you created the dossier, you could add it here by selecting Existing Dataset.

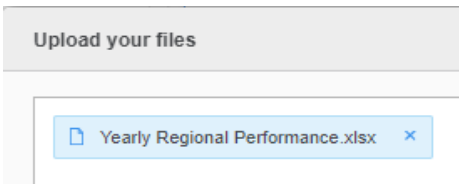
2 On the Connect to Your Data page, click **File From Disk**.



3 On the Upload Your Files page, click **Choose files**.



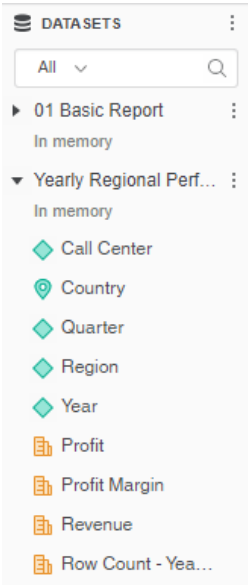
4 In the Open window, navigate to the folder where you saved the Exercise Files provided, select **Yearly Regional Performance.xlsx**, and click **Open**.



You can select additional files, which are combined into the same dataset. For this exercise, you are using only one file. To learn about combining datasets, see the Preparing Complex Data for Enhanced Storytelling chapter in the Enterprise Analyst learning path.

5 Click **Finish** to import the spreadsheet's data into the dossier.

You can preview the data and improve its structure and quality before importing it, if necessary. You learn more about previewing and improving data later in this lesson.



The file (referred to as a dataset once you add it to the dossier) displays in the Datasets panel in your dossier. The objects in the file display as attributes and metrics and are ready for you to use in visualizations and filters. The other dataset is collapsed in the image above so that you can focus on the newest dataset.

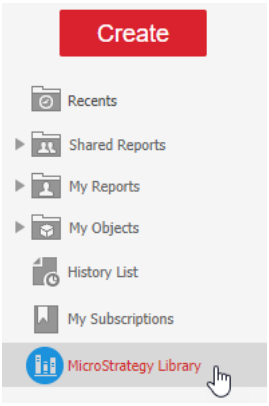
Country has a different icon, , which indicates that it has a geo (geographic) role. An attribute with a geo role contains latitude and longitude information. You learn more about using geo roles with map visualizations in *Adding geographical data for map visualizations, page 31*.

6 Save your dossier, by clicking **Save** in the right of toolbar.

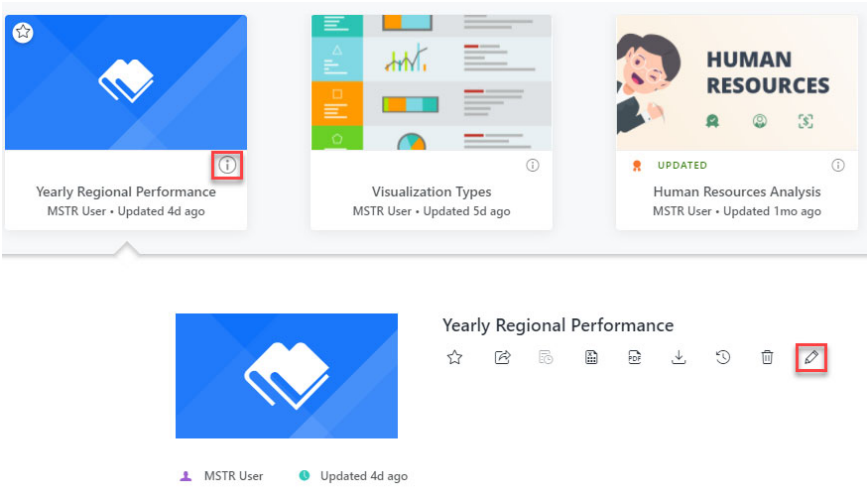
Viewing and editing a dossier in Library

The information below is provided as a reference only. Refer back to this section for high-level information on how to use and close Edit mode in Library.

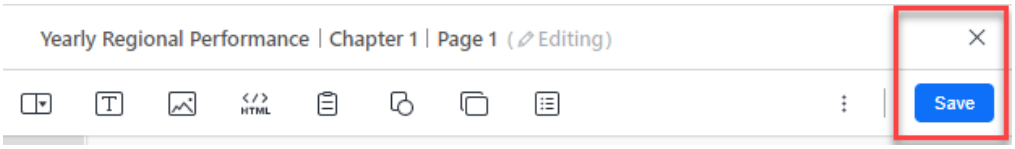
- To access Library from the Education sandbox, click **MicroStrategy Library** in the left pane.



- To open a dossier in Edit mode from the Library home page:
  - Click the **Information** icon ⓘ on the dossier and then click the **Edit** icon ✎.



- To exit Edit mode and return to viewing a dossier:
  - Click the **Save** button to save your changes and then click the **X**.



- To return to the Library home page to view the dossiers in your Library:
  - Click the **Library** icon in the upper left of the toolbar.

Exercise: Create a grid visualization

In this exercise, you use the data you imported in *Exercise: Import a MicroStrategy report and Excel data into a new dossier, page 8* to create the quickest and easiest visualization, a simple grid. A grid visualization displays data in rows and columns, and allows you to pivot, sort, move, drill, filter, and perform additional manipulations to analyze the data. Since a grid visualization displays the actual data, rather than graphic elements like bars or circles, you can use it to understand and prepare your data before displaying it on another type of visualization.

To learn more about using grid visualizations, see the Understanding Your Data Through Visualizations chapter of the Departmental Analyst learning path.

Following the guidelines from your manager's mockup, your grid visualization looks like the image below.

Grid visualization sample

Yearly Regional Profit and Revenue

Country	Region	Year	2020	2020	2020	2020	2020	2020	2020
		Quarter	2020 Q1	2020 Q1	2020 Q1	2020 Q2	2020 Q2	2020 Q2	2020 Q3
		Call Center	Profit	Revenue	Profit Margin	Profit	Revenue	Profit Margin	Profit
USA	Central	Fargo	$2,205	$16,659	13.24%	$1,101	$13,295	8.28%	$1,400
		Milwaukee	$14,342	$103,976	13.79%	$10,624	$101,866	10.43%	$16,321
	Mid-Atlantic	Charleston	$169	$2,188	7.73%	$131	$2,584	5.07%	$271
		Washington, DC	$6,082	$50,084	12.14%	$5,762	$61,299	9.40%	$9,040
	Northeast	Boston	$966	$9,693	9.97%	$1,297	$14,368	9.02%	$1,658
		New York	$43,253	$277,963	15.56%	$37,407	$293,370	12.75%	$48,541
	Northwest	San Francisco	$524	$5,878	8.92%	$324	$5,310	6.10%	$582
		Seattle	$819	$7,643	10.71%	$786	$8,735	9.00%	$739
	South	Memphis	$11,855	$77,168	15.36%	$12,826	$97,595	13.14%	$13,693
		New Orleans	$12,059	$88,757	13.59%	$11,815	$103,323	11.43%	$16,338
	Southeast	Atlanta	$577	$5,739	10.06%	$718	$8,344	8.60%	$694
		Miami	$804	$8,424	9.54%	$1,275	$13,954	9.14%	$1,429
	Southwest	Salt Lake City	$937	$9,133	10.26%	$607	$8,731	6.96%	$1,624
		San Diego	$5,729	$45,717	12.53%	$4,854	$48,347	10.04%	$5,486
Web	Web	Web	$12,109	$81,620	14.84%	$14,147	$108,158	13.08%	$20,294

Not all of the columns are shown in the sample; use the scroll bar at the bottom of the visualization to display more columns.

The last row appears to be blank but does contain data in the columns all the way to the right. This is an error in your data that you fix when you wrangle your data in *Exercise: Replace a dataset with improved data, page 40*.

Prerequisite

This exercise uses a dossier you created in a different exercise. Before you begin, ensure that you have completed *Exercise: Import a MicroStrategy report and Excel data into a new dossier, page 8*.

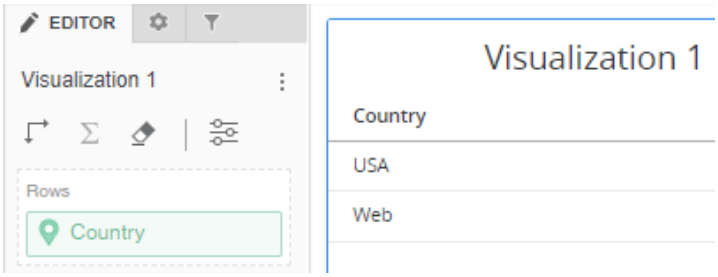
Add attributes and metrics to the grid

1 In the Education sandbox Library, open the **Yearly Regional Performance** dossier in Edit mode, if not already there.

For a reminder on how to do this, see ***Exercise: Import a MicroStrategy report and Excel data into a new dossier, page 8***.

2 The data you need to display in the grid visualization is in the Yearly Regional Performance dataset. To focus only on this dataset, in the Datasets panel, click the arrow icon ▼ next to **01 Basic Report** to collapse it.

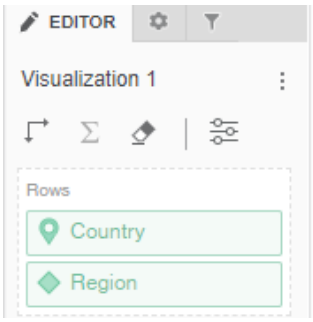
3 Double-click the **Country** attribute. Country is added to the Rows drop zone in the Editor panel. The various countries (which are attribute elements) are displayed as rows in the grid. The row header is Country.



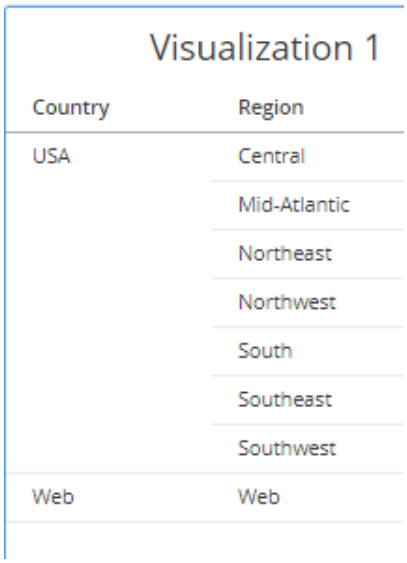
To add data to the visualization on the selected page, you can double-click dataset objects, drag and drop them into the drop zones in the Editor panel, or drag and drop them directly into the visualization container.

Rows, columns, and metrics are the most common drop zones in the Editor panel; however, different visualizations sometimes have different requirements. For example, a Heat Map visualization includes a Color By drop zone to color each tile by its relative value.

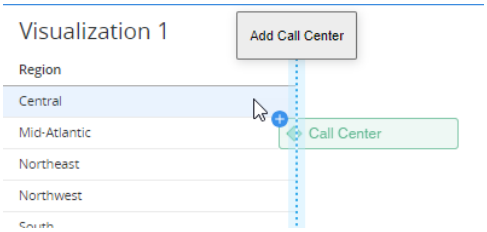
4 Double-click the **Region** attribute to add it to the **Rows** drop zone, below Country, in the Editor panel.



Region is displayed to the right of Country in the grid:



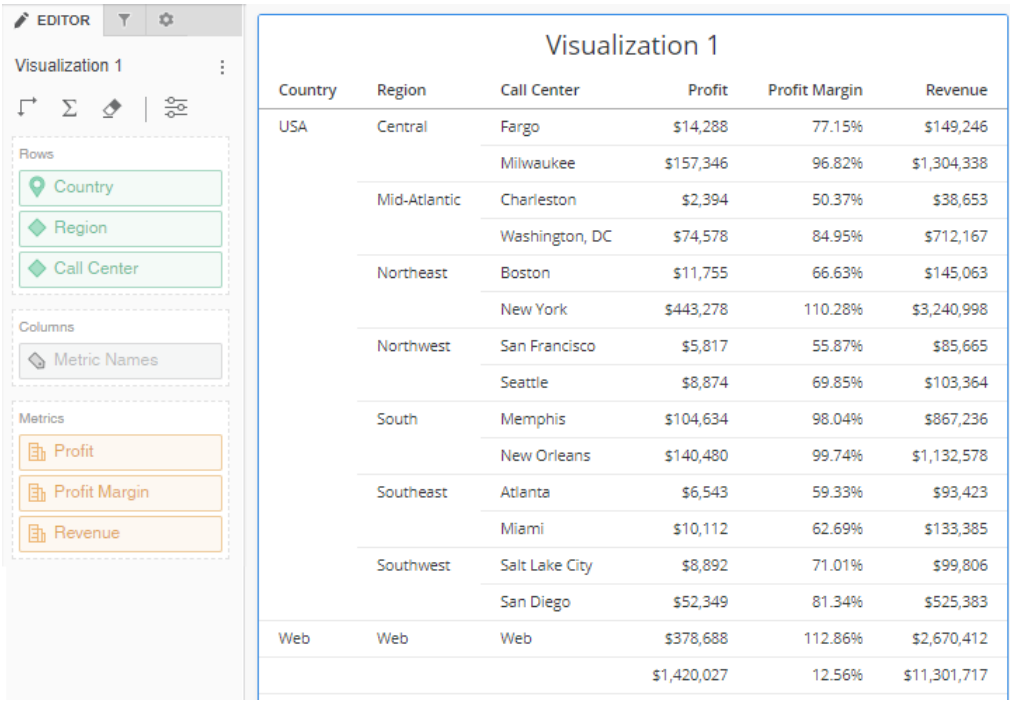
5 Drag **Call Center** from the Datasets panel to the right of Region in the grid.



You can add, remove, and move objects in either the visualization itself or in the Editor panel.

6 SHIFT+click **Profit**, **Profit Margin**, and **Revenue** to select all three metrics. Drag and drop them in the **Metrics** drop zone in the Editor panel.

The values for the metrics are displayed in columns in the grid with the metric names in the column headers.



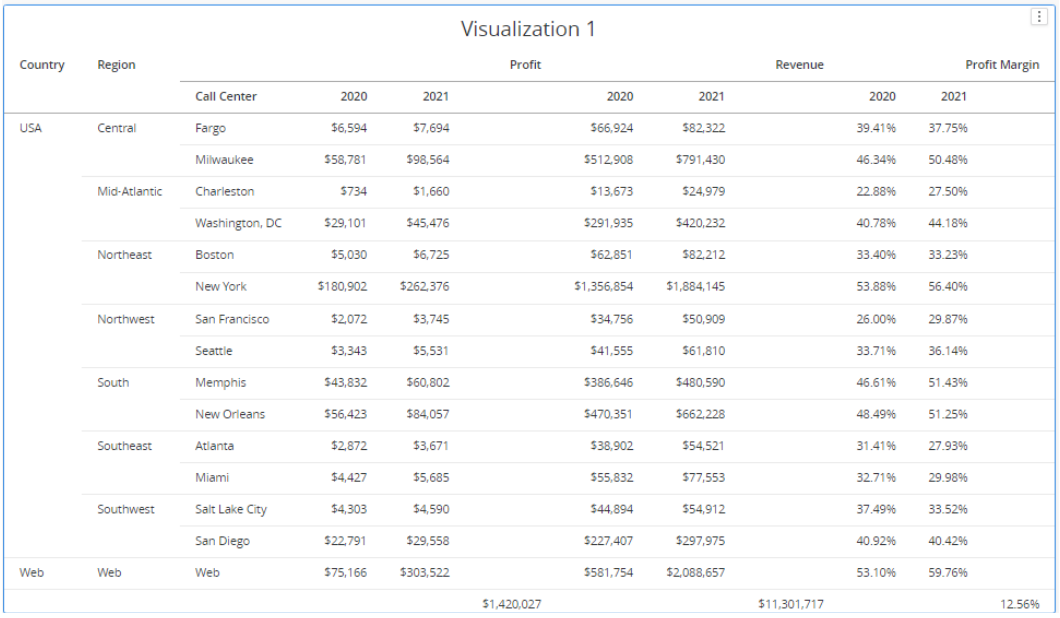
Country	Region	Call Center	Profit	Profit Margin	Revenue
USA	Central	Fargo	$14,288	77.15%	$149,246
		Milwaukee	$157,346	96.82%	$1,304,338
	Mid-Atlantic	Charleston	$2,394	50.37%	$38,653
		Washington, DC	$74,578	84.95%	$712,167
	Northeast	Boston	$11,755	66.63%	$145,063
		New York	$443,278	110.28%	$3,240,998
	Northwest	San Francisco	$5,817	55.87%	$85,665
		Seattle	$8,874	69.85%	$103,364
	South	Memphis	$104,634	98.04%	$867,236
		New Orleans	$140,480	99.74%	$1,132,578
	Southeast	Atlanta	$6,543	59.33%	$93,423
		Miami	$10,112	62.69%	$133,385
	Southwest	Salt Lake City	$8,892	71.01%	$99,806
		San Diego	$52,349	81.34%	$525,383
Web	Web	Web	$378,688	112.86%	$2,670,412
			$1,420,027	12.56%	$11,301,717

An object called Metric Names has been automatically added to the Columns drop zone. This object contains the names of the metrics and determines where to display the metric names. The metric names display in the column headers, since Metric Names is placed in the columns. In the following steps, you move the Metric Names object to explore the different locations in which the visualization can display metric names.

Change the location of metrics and metric names

1 Change the order that the metrics are displayed on the grid. In the Metrics drop zone, drag **Profit Margin** below Revenue.

2 Drag **Year** from the Datasets panel to the **Columns** drop zone, below Metric Names. On the grid, the years are displayed below the metric names.



Visualization 1

Country	Region		Profit			Revenue			Profit Margin	
		Call Center	2020	2021		2020	2021		2020	2021
USA	Central	Fargo	$6,594	$7,694		$66,924	$82,322		39.41%	37.75%
		Milwaukee	$58,781	$98,564		$512,908	$791,430		46.34%	50.48%
	Mid-Atlantic	Charleston	$734	$1,660		$13,673	$24,979		22.88%	27.50%
		Washington, DC	$29,101	$45,476		$291,935	$420,232		40.78%	44.18%
	Northeast	Boston	$5,030	$6,725		$62,851	$82,212		33.40%	33.23%
		New York	$180,902	$262,376		$1,356,854	$1,884,145		53.88%	56.40%
	Northwest	San Francisco	$2,072	$3,745		$34,756	$50,909		26.00%	29.87%
		Seattle	$3,343	$5,531		$41,555	$61,810		33.71%	36.14%
	South	Memphis	$43,832	$60,802		$386,646	$480,590		46.61%	51.43%
		New Orleans	$56,423	$84,057		$470,351	$662,228		48.49%	51.25%
	Southeast	Atlanta	$2,872	$3,671		$38,902	$54,521		31.41%	27.93%
		Miami	$4,427	$5,685		$55,832	$77,553		32.71%	29.98%
	Southwest	Salt Lake City	$4,303	$4,590		$44,894	$54,912		37.49%	33.52%
		San Diego	$22,791	$29,558		$227,407	$297,975		40.92%	40.42%
Web	Web	Web	$75,166	$303,522		$581,754	$2,088,657		53.10%	59.76%
					$1,420,027			$11,301,717		12.56%

The blank column name after the years for each metric occurs because of the value at the bottom of the column. These values, which are totals for each metric, were not labeled in the Excel spreadsheet, and therefore are not labeled in the dataset in the dossier. You fix the issue when you wrangle data in ***Exercise: Replace a dataset with improved data, page 40****.*

3 To display the years above the metric names, in the Columns drop zone, drag **Year** above Metric Names. On the grid, the years are now displayed above the metric names.

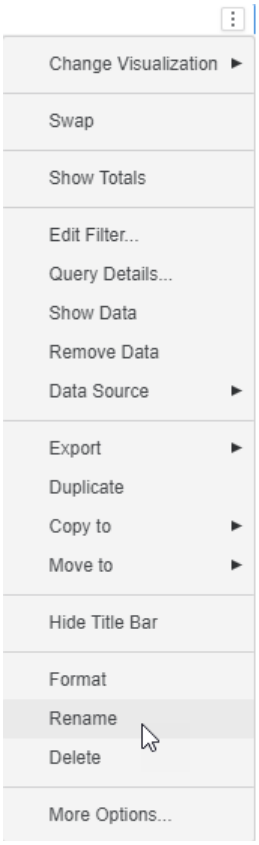
Country	Region	Year	2020	2020	2020	2021	2021	2021
		Call Center	Profit	Revenue	Profit Margin	Profit	Revenue	Profit Margin

4 Drag **Quarter** from the Datasets panel to the **Columns** drop zone in the Editor panel, between Year and Metric Names. On the grid, the years are displayed at the top, followed by quarters, and finally the metric names.

Country	Region	Year	2020	2020	2020	2020	2020	2020	2020
		Quarter	2020 Q1	2020 Q1	2020 Q1	2020 Q2	2020 Q2	2020 Q2	2020 Q3
		Call Center	Profit	Revenue	Profit Margin	Profit	Revenue	Profit Margin	Profit

Rename the visualization

1 Rename the visualization by clicking the menu icon in the visualization's title bar and selecting **Rename**. Type **Yearly Regional Profit and Revenue**.



The grid now looks like the sample provided at the beginning of this exercise, *Grid visualization sample, page 16*.

2 Save your dossier by clicking **Save** in the right of the dossier toolbar.

Best Practice

It is helpful to save your dossier throughout the lesson to ensure that all your changes have been saved.

Exercise: Create a bubble chart visualization

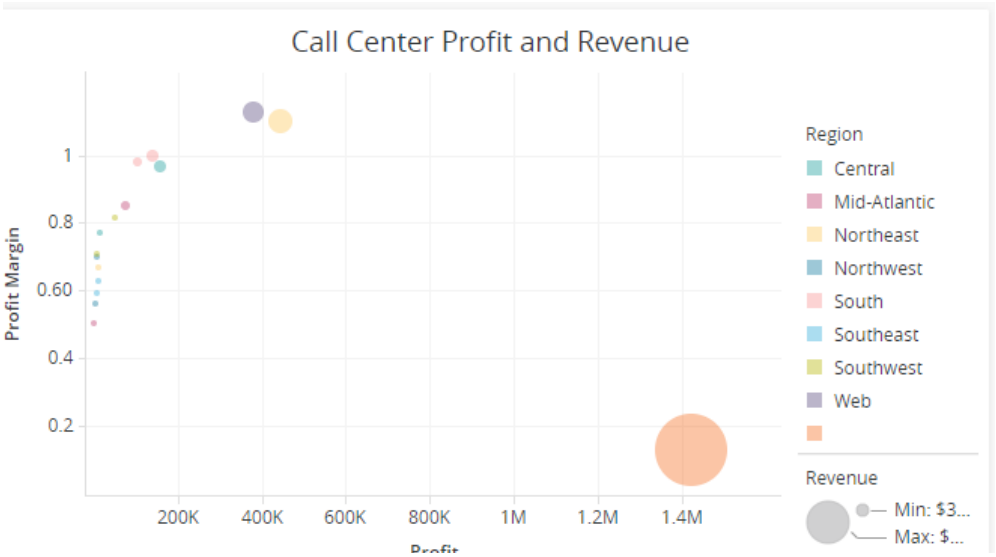
Continuing to follow the guidelines from your manager's mockup, you now create a bubble chart visualization in your dossier. A bubble chart visualization highlights the trends for the values of two or more metrics. Each attribute element (such as Central and South) is represented by a bubble. You graph the bubbles on a scatter plot and can size and color them by metric values and attributes.

To learn more about when to use bubble charts, see the Understanding Your Data Through Visualizations chapter of the Departmental Analyst learning path.

While the bubble chart displays much of the same information as the grid visualization, you can more easily see the differences in revenue and profit between the different call centers.

Your bubble chart visualization looks like the image below.

Bubble chart sample



Prerequisites

This exercise builds on progress you've made with the dossier so far. Before you begin:

1 Access Library. For a reminder on how to do this, see *Exercise: Access Library in the Education sandbox, page 7*.

2 In Library, open the **Yearly Regional Performance** dossier in Edit mode. For a reminder on how to do this, see *Exercise: Import a MicroStrategy report and Excel data into a new dossier, page 8*.

3 Ensure that you have:

- Imported data into the dossier in *Exercise: Import a MicroStrategy report and Excel data into a new dossier, page 8*.
- Created the grid visualization in *Exercise: Create a grid visualization, page 16*.

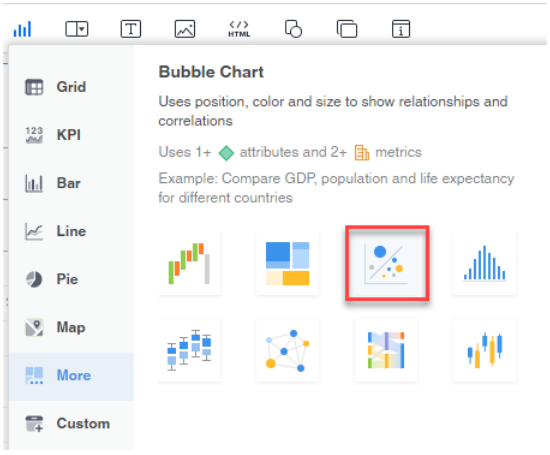
Create a bubble chart visualization

1 The toolbar above the dossier canvas area gives you several options to add items to the page, such as visualizations, filters, text, and shapes.

Click the **Visualization** icon to display the Visualization Gallery, which is used to create new visualizations.

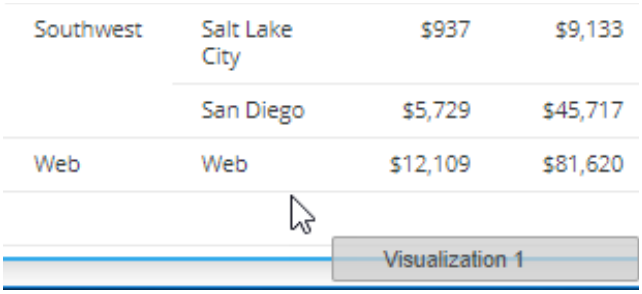
The Visualization Gallery contains all the available types of visualizations, grouped by type. If you hover your cursor over a visualization icon in the gallery, you can quickly view the description of a visualization type and its data requirements. If you click a visualization icon, a blank visualization of that type is added to your dossier. A blank visualization on the dossier also displays the data requirements for its visualization type.

2 To add a bubble chart, in the Visualization Gallery, point to **More** and then click the **Bubble Chart** icon.



A new, blank bubble chart visualization is added next to the grid. You need to add at least two metrics and one attribute to this bubble chart visualization. Notice that the drop zones in the Editor panel have changed because you are using a different type of visualization (a grid visualization contains Rows, Columns, and Metrics drop zones).

3 Drag and drop the new visualization (automatically labeled Visualization 1) below the grid. When the blue line is below the grid instead of on the left or right, you have the blank visualization in the correct location.



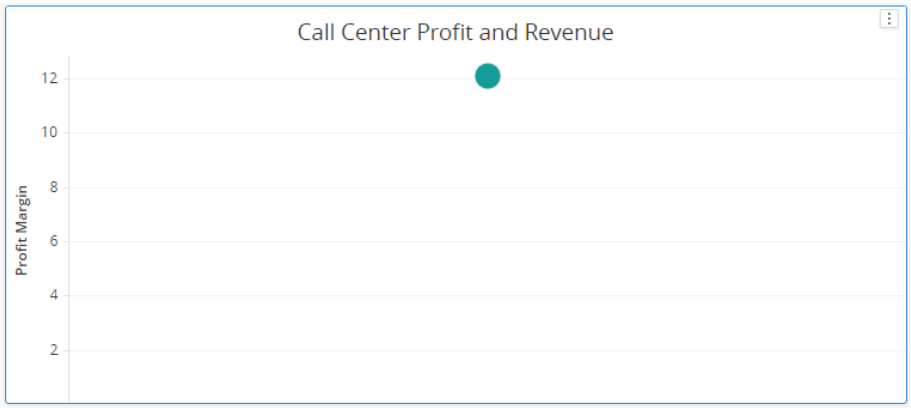
4 Rename the visualization **Call Center Profit and Revenue** using the menu icon ⋮ in the visualization's title bar.

5 The data you need to display in the bubble chart is in the Yearly Regional Performance dataset. If you did not already collapse the 01 Basic Report in a previous exercise, click the arrow icon ▼ next to **01 Basic Report** to collapse it. This helps you focus only on the dataset you need.

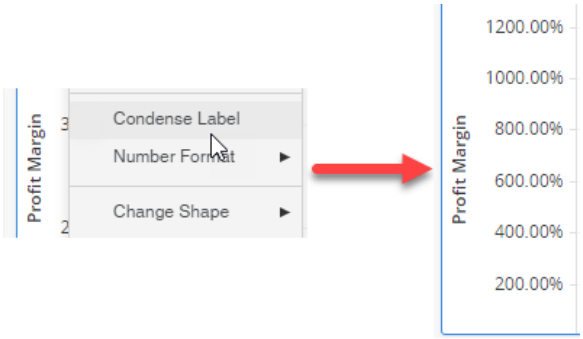
6 Double-click the **Profit Margin** metric from the Yearly Regional Performance dataset to display it in the bubble chart. The metric is added to the Vertical drop zone in the Editor panel.

The vertical axis displays a scale of Profit Margin values, and the position of the bubble on the graph's vertical axis is calculated from the Profit Margin values. A single bubble is displayed, representing the profit margin for all countries

and all years in the dataset. Notice that the vertical axis displays values, while the horizontal axis does not.



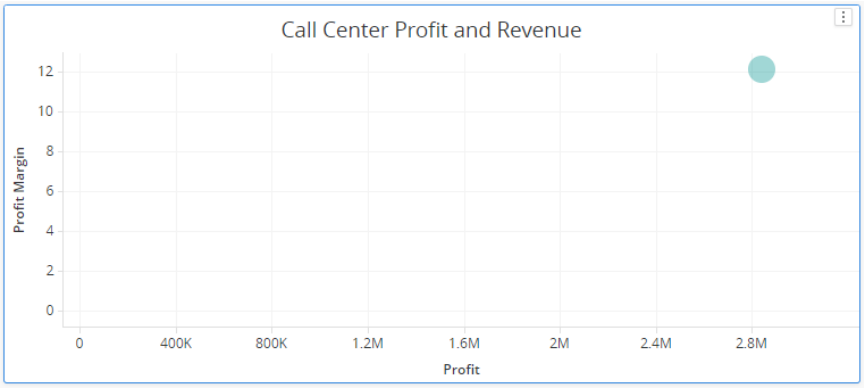
The vertical axis labels are automatically condensed to save space and simplify the analysis. For example, the combined Profit Margin value is 1209.51%, and the vertical axis grid line that it is closest to is labeled as 12. To adjust this, you could right-click one of the vertical axis labels, and then click ***Condense Label*** *to clear the check mark indicating that the labels are condensed. The labels then display as 200.00%, 400.00%, and so on, up to 1200.00%.*



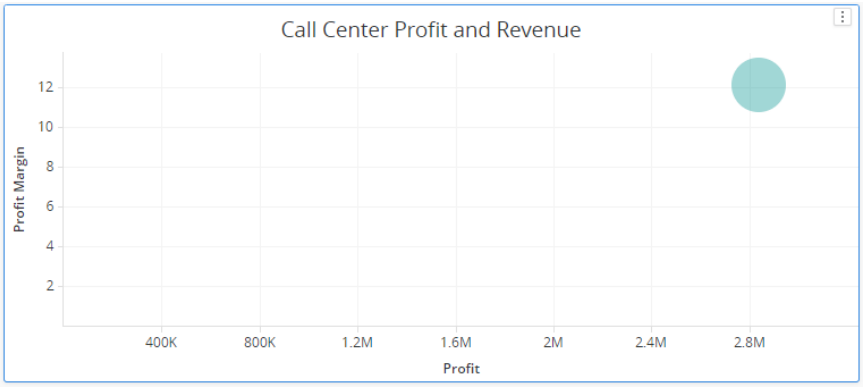
7 Double-click the **Profit** metric to add it to the bubble chart. The metric is added to the Horizontal drop zone in the Editor panel.

The horizontal axis displays a scale of Profit values. The single bubble's position has moved to indicate the intersection of the Profit value and the

Profit Margin value for all countries and all years. Now both axes have values displayed.



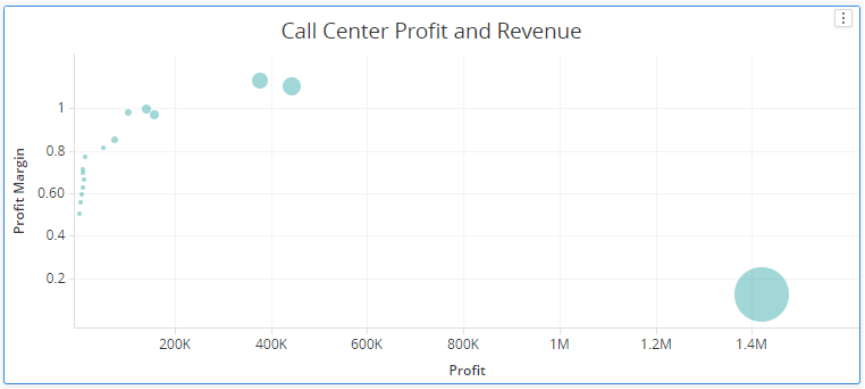
8 Drag the **Revenue** metric to the **Size By** drop zone. The size of the bubble increases, representing the single Revenue value (for all countries and all years).



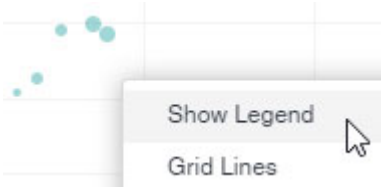
9 Drag the **Call Center** attribute to the **Break By** drop zone.

A bubble is displayed for each call center. The bubbles vary in size, depending on the revenue for that call center. Each bubble is positioned horizontally

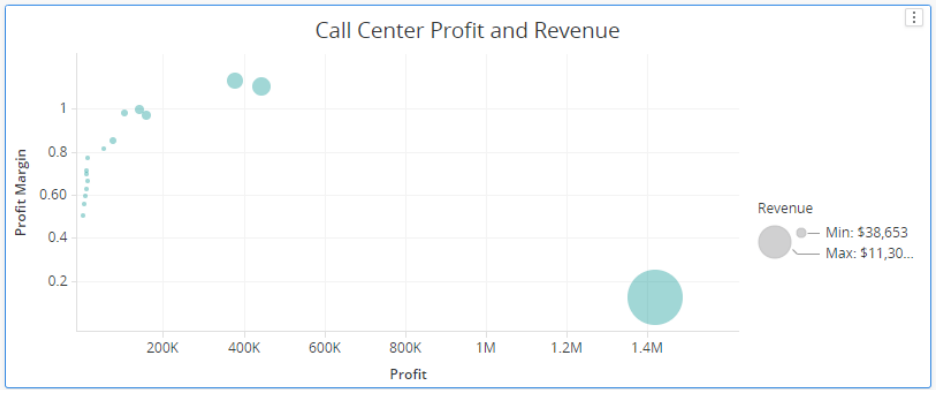
according to the call center's Profit value and vertically according to the call center's Profit Margin value.



10 Add a legend, so that analysts can quickly understand that the bubbles are sized by Revenue. Right-click inside the bubble chart visualization, and select **Show Legend**.



A legend displays in the bubble chart, showing the minimum and maximum revenue values.

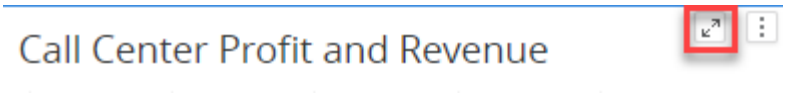


11 Drag **Region** to the **Color By** drop zone.

Each region is assigned a color, as shown in the legend. Each bubble is colored according to the region that it is located in.

The bubble chart looks like the sample provided in the beginning of the exercise, *Bubble chart sample, page 22*. To view it in more detail, click the

Maximize icon in the top right corner of the visualization's title bar. To return the visualization to its previous size, click the **Restore** icon in the visualization's title bar.

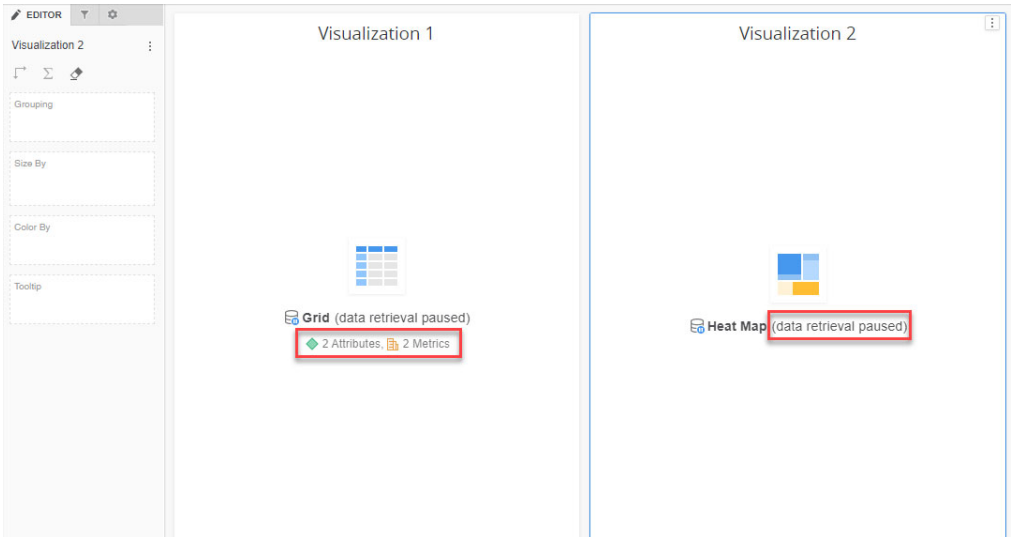


12 **Save** your dossier.

Pausing data retrieval during dossier creation

As you add visualizations to your dossiers, you can pause the data query. If the dossier accesses large datasets, this allows you to quickly edit and modify your dossier without delaying the retrieval of data from the data source. After clicking the **Pause Data Retrieval** icon the toolbar, you can modify, add, and remove objects and filters on the dossier. After you complete editing in Pause mode, click the **Resume Data Retrieval** icon to retrieve relevant data from the data source.

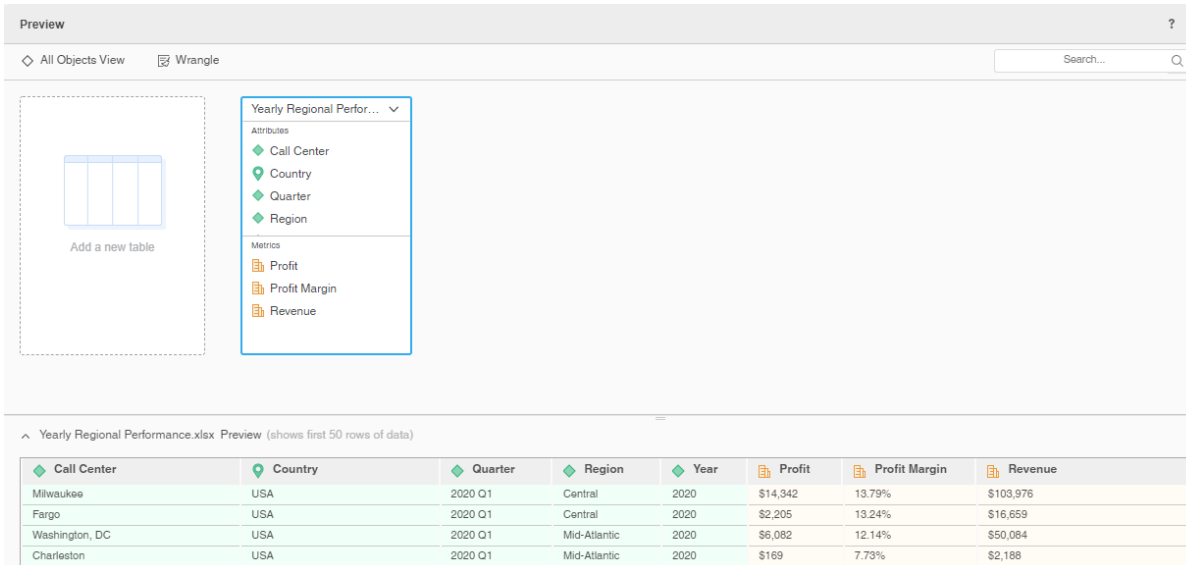
For example, the dossier shown below was paused and then two attributes and two metrics were added to Visualization 1, a grid. A second visualization, a heat map, was added to the dossier, but no objects have been added to it yet.



You can also open a dossier in paused data mode from the Library home page by right-clicking the dossier and selecting **Edit Without Data**.

Improving the quality of your data's structure

Your data is imported into MicroStrategy as tables with columns and rows. In *Exercise: Import a MicroStrategy report and Excel data into a new dossier, page 8*, as you imported the Yearly Regional Performance.xlsx file, the following table was created:



You improve the quality of your data structure during the import process. For example, you might need to rename a column or table to match your organization's naming conventions or remove a data column that contains personally identifiable information that the company must keep confidential. After you select the data to import, you can prepare the data using the Preview page. The Preview page also allows you to view the first 50 rows of data.

You can improve the quality of your data's structure by changing the table. For example, you can:

- Select which columns to include or exclude in the import.
- Combine columns to define a multi-form attribute if your data contains different attribute forms of the same attribute, saved in separate columns.

  An attribute can contain multiple attribute forms. An attribute form is an additional category for any data that your organization saves about any of its business attributes. For example, a Customer attribute can have the forms ID, Last Name, First Name, Address, and Email.
- Delete the table completely, so that it is not imported into your dossier.

- Rename the table. The table's name is displayed in the dossier's Datasets panel.

You can also improve the quality of your data's structure by changing individual data columns. For example, you can:

- Select whether the column is an attribute or a metric.
- Rename the column.
- Change the column's data type (text, date, or number).
- Define geographical information to display on map visualizations.
- Map the column to an attribute that already exists in your project.

  When you import data, MicroStrategy automatically attempts to link imported data to attributes that already exist in the dossier. You can change the linking by manually linking a column to group or display imported data based on a project attribute.

Adding geographical data for map visualizations

To create a map visualization, you need geographical location data. When you import data, MicroStrategy automatically identifies data columns that contain geographical information, such as city names or locations using latitude and longitude. These columns are assigned geo roles, which add latitude and longitude information to the attribute. Geo roles include:

- Area Code
- City
- Country
- County
- Latitude
- Longitude
- State
- Zip Code

Map visualizations use geo roles to display geographic data. The geo role determines the visualization's default base map. For example, if the geo role is country, a world map displays; if the geo role is state, a U.S. state names map displays.

If a geographic data column is not automatically identified during data import, you can assign it a geo role yourself. You can change the geo role of an automatically assigned column.

You can also create attributes using geographic information at a higher level than the selected data column. For example, you can create a state attribute based on city information, even if there is not a state column in your dataset.

Exercise: Improve the data structure during data import

You need to import data that supports creating the requested network visualization, which will display the relationship between sales promotion types and purchase types. This data is saved as the Sheet2 worksheet in an Excel spreadsheet. In this exercise, you improve the quality of the data's structure during import by:

- Combining columns to define multi-form attributes

  The worksheet contains two columns with information about the type of sales promotion with which customers buy items: ID and description (DESC). Each column is imported as a separate attribute.

Promo Type DESC	Promo Type ID
No Promo	1
No Promo	1
No Promo	1
No Promo	1

  You combine the columns to create a single attribute, called Promo Type, that contains both forms. In a network visualization, you use one attribute to display the nodes. Having multiple forms in a single attribute allows multiple descriptive categories to be displayed in each node. You can display the number (the ID) as a quick reference or to save space in a visualization, the description to provide more information, or both at the same time.

Promo Type	
Promo Type ID@Promo Type ID	Promo Type DESC@Promo Type DESC
1	No Promo
1	No Promo
1	No Promo
1	No Promo

  You also create a Purchase Type attribute, with an ID form and a description form, for the same reason.

- Converting an attribute to a metric

  The Item Qty column is identified as an attribute, but it should be a metric, because it is a business measure.

- Converting a metric to an attribute

  The Order Number column is identified as a metric, but it should be an attribute, because it provides context for business measures.

The Excel file with the necessary data is provided in the Exercise Files.

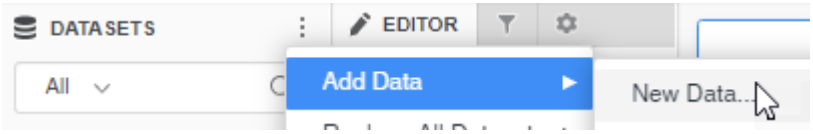
Prerequisites

This exercise builds on progress you've made with the dossier so far. Before you begin:

1. Access Library. For a reminder on how to do this, see *Exercise: Access Library in the Education sandbox, page 7*.
2. In Library, open the **Yearly Regional Performance** dossier in Edit mode. For a reminder on how to do this, see *Exercise: Import a MicroStrategy report and Excel data into a new dossier, page 8*.
3. Ensure that you have:
   - Imported data into the dossier in *Exercise: Import a MicroStrategy report and Excel data into a new dossier, page 8*.
   - Created the grid visualization in *Exercise: Create a grid visualization, page 16*.
   - Created the bubble chart visualization in *Exercise: Create a bubble chart visualization, page 22*.

Improve the data structure during data import

1. In the Datasets panel, click the menu icon ⋮, point to **Add Data**, and then select **New Data**.



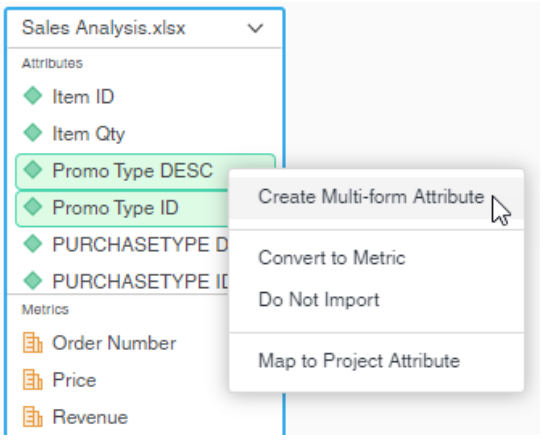
2. On the Connect to Your Data window, click **File From Disk**.
3. On the Upload Your Files window, click **Choose files**.
4. In the Open window, navigate to the folder where you saved the Exercise Files provided, select **Sales Analysis.xlsx**, and click **Open**.
5. Click **Prepare Data**.

6 If your Excel file has multiple worksheets, you can select which ones to import. On the Select Worksheets page, clear the **All Worksheets** check box, then select only the **Sheet2** check box.

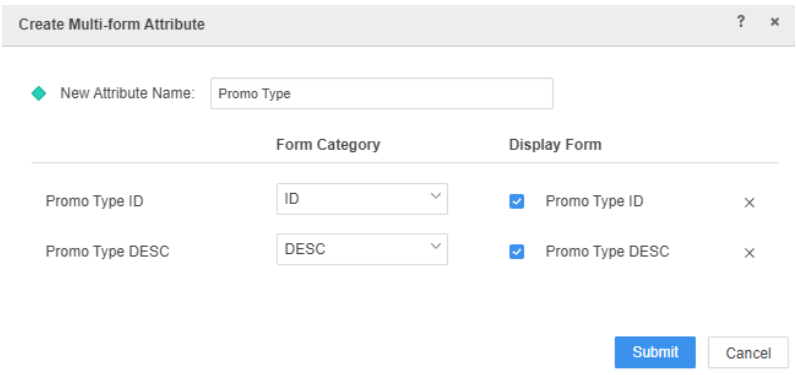
7 Click **Select**.

Create a multi-form attribute

8 In the table at the top of the Preview page, SHIFT+click **Promo Type DESC** and **Promo Type ID** to select both columns. Right-click the selected columns and click **Create Multi-form Attribute**.



9 On the Create Multi-form Attribute page, in the **New Attribute Name** box, rename the attribute to **Promo Type.** The ID and description forms have automatically been identified with the correct form category. You want each attribute form to be available to display, so leave the Display Form check boxes selected.



10 Click **Submit**.

11 Create the **Purchase Type** attribute by combining the **PURCHASETYPE DESC** and **PURCHASETEYPE ID** columns the same way you created the Promo Type

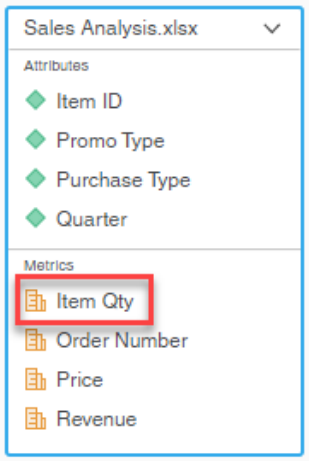
attribute. The resulting attribute looks like the following at the bottom of the Preview window:

Purchase Type	
PURCHASETYPE ID@PURCHASETYPE ID	PURCHASETYPE DESC@PURCHASETYPE DESC
2	Phone
2	Phone
2	Phone

Convert an attribute to a metric

12 Item Qty should be a metric, instead of an attribute. In the table at the top of the Preview page, right-click **Item Qty**, and select **Convert to Metric**.

The column is now displayed as a metric:



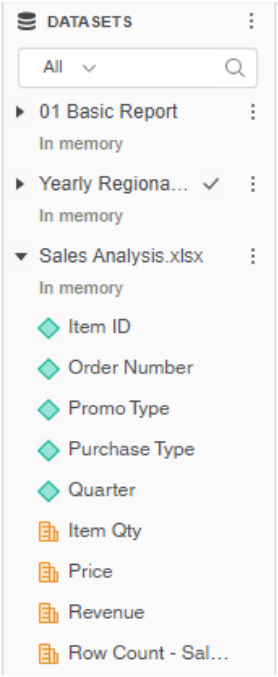
Convert a metric to an attribute

13 Order Number should be an attribute, instead of metric. In the table at the top of the Preview page, right-click **Order Number**, and select **Convert to Attribute**. The column is now displayed as an attribute.

Import the data

14 Click **Finish** to import the worksheet's data into the dossier.

The file is displayed in the Datasets panel in your dossier.



The other datasets are collapsed in this image so that you can focus on the objects from the newly imported dataset.

15 **Save** your dossier.

The datasets you imported into your dossier do not have defined relationships linking them. If you link them, you can display information from multiple datasets in a single visualization. To learn more about linking datasets, see the Enterprise Analyst learning path.

Improving the quality of your data values: Data wrangling

In addition to improving the structure of your data, you can refine the quality of your data values, which are the individual cells in the table. When you refine your data (referred to as data wrangling), you view a sample of the actual values that will be imported into the dossier. Viewing sample values allows you to assess its quality and usability. For example, when you choose to wrangle the data imported from the Yearly Regional Performance.xlsx file, the following data sample displays:

Year	Quarter	Country	Region	Call Center	Profit	Revenue	Profit Margin
2020	2020 Q1	USA	Central	Milwaukee	14342.1682000001	103976	0.137937295145034
2020	2020 Q1	USA	Central	Fargo	2204.875	16659	0.132353382555976
2020	2020 Q1	USA	Mid-Atlantic	Washington, DC	6081.64060000003	50084	0.121428811596518
2020	2020 Q1	USA	Mid-Atlantic	Charleston	169.0242	2188	0.0772505484460694
2020	2020 Q1	USA	Northeast	Boston	966.4432	9693	0.0997052718456618
2020	2020 Q1	USA	Northeast	New York	43252.6285	277963	0.155605704716095
2020	2020 Q1	USA	Northwest	San Francisco	524.2604	5878	0.0891902687989111
2020	2020 Q1	USA	Northwest	Seattle	818.7432	7643	0.107123276200445
2020	2020 Q1	USA	South	New Orleans	12059.1125000001	88757	0.135866607704182
2020	2020 Q1	USA	South	Memphis	11854.9245	77168	0.153624876891976
2020	2020 Q1	USA	Southeast	Atlanta	577.4073	5739	0.100611134343962
2020	2020 Q1	USA	Southeast	Miami	803.8388	8424	0.0954224596391264
2020	2020 Q1	USA	Southwest	San Diego	5729.01790000004	45717	0.125314825994707
2020	2020 Q1	USA	Southwest	Salt Lake City	936.7489	9133	0.102567491514289
2020	2020 Q1	Web	Web	Web	12109.0356	81620	0.148358681695663
2020	2020 Q2	USA	Central	Milwaukee	10624.1295	101866.4	0.104294738009785
2020	2020 Q2	USA	Central	Fargo	1101.3861	13295.2	0.0828408824237318
2020	2020 Q2	USA	Mid-Atlantic	Washington, DC	5762.44510000004	61299.3	0.0940050718360575
2020	2020 Q2	USA	Mid-Atlantic	Charleston	130.8854	2583.9	0.0506542048840899
2020	2020 Q2	USA	Northeast	Boston	1296.5443	14368.4	0.0902358160964338
2020	2020 Q2	USA	Northeast	New York	37407.1927000002	293369.9	0.12750862545885
2020	2020 Q2	USA	Northwest	San Francisco	323.7593	5309.6	0.0609762128973935
2020	2020 Q2	USA	Northwest	Seattle	786.484900000001	8734.6	0.0900424633068487
2020	2020 Q2	USA	South	New Orleans	11814.8323	103322.8	0.114348742968638
2020	2020 Q2	USA	South	Memphis	12826.3542	97595.2	0.131424027001328
2020	2020 Q2	USA	Southeast	Atlanta	717.6971	8343.8	0.0860156163858195
2020	2020 Q2	USA	Southeast	Miami	1275.2045	13954.4	0.0913826954325517

1 - 50 / 121 Rows

A total of 1000 rows of data are available by default, although this spreadsheet contains only 121 rows. For larger datasets, you can increase the sample size to 10,000 rows, since 1000 rows are not always representative of the data as a whole in a larger dataset. You can choose whether the sample rows are selected randomly or from the first rows of the data source.

Data wrangling is an optional step in the data import process.

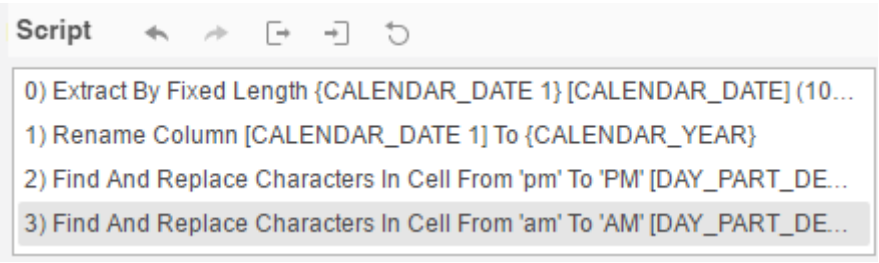
Improve the quality of your data values by reviewing them to assess quality and usability. If you discover a problem, you can, for example:

- Remove white space

- Copy data into white space
- Delete duplicate rows or cells
- Find and replace specific data
- Join columns
- Create a new column by extracting data from an existing column
- Split a column or cell

After you define the action, it is applied to the data sample, so that you can view the results. At this stage, it is not applied to the actual data. You are instead building a script of actions. This script is applied when you finish importing the data into the dossier.

Sample data wrangling script

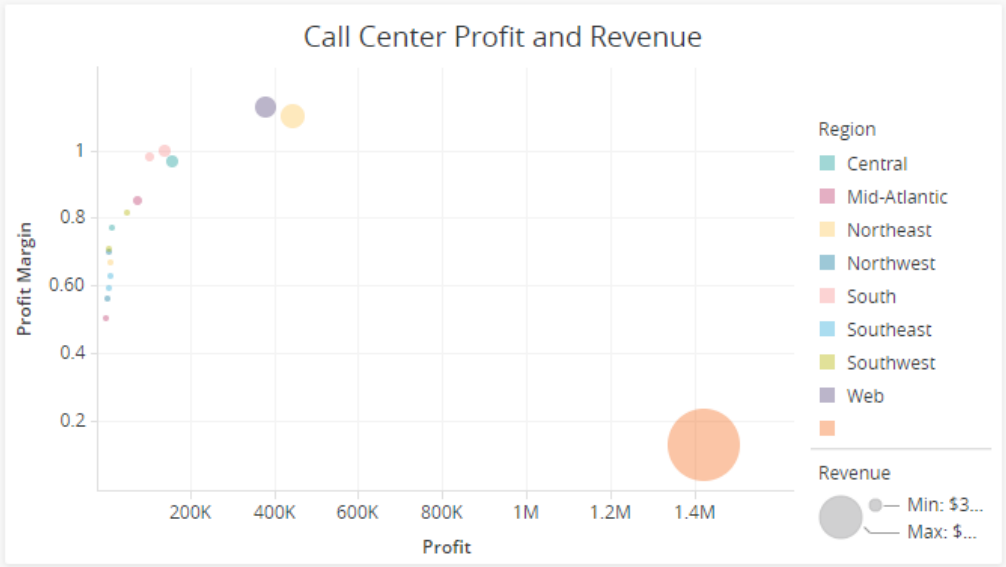


You can edit the script by undoing and redoing actions. You can also export the completed script to apply to another dataset.

Exercise: Replace a dataset with improved data

You started creating the requested map visualization that displays call center revenue and profit but realized you need to improve your data to meet the requirement. When you imported the Yearly Regional Performance.xlsx file to create your dossier, one attribute, Country, was automatically assigned a geo role, but you need a second attribute with geographical information for the call centers. You can use data preview options to assign the city geo role to the Call Center attribute, since call centers are cities.

Additionally, when displaying all regions in the bubble chart you created in *Exercise: Create a bubble chart visualization, page 22*, you noticed the large bubble in the lower right does not have a region label in the legend. You can wrangle the data to discover what is causing the problem and fix it.



Since data wrangling is part of the data import process, in this exercise you:

- Re-import the file and add a new geo role
- Wrangle the data to find and resolve the problem with the missing region label
- Replace the old dataset objects with the updated ones

Best Practice

Replacing or editing the dataset allows you to retain the visualizations that use objects from the old dataset. If you instead deleted the old dataset, the objects on the visualizations would also be deleted.

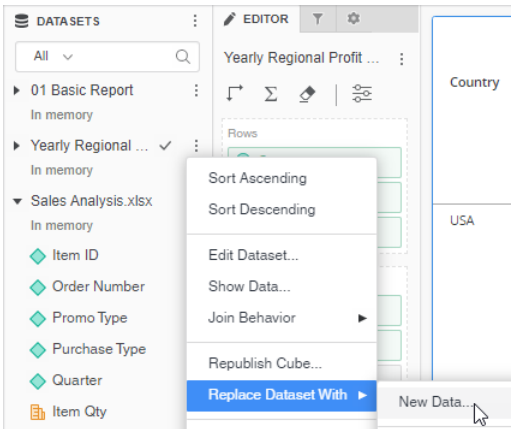
Prerequisites

This exercise builds on progress you've made with the dossier so far. Before you begin:

1. Access Library. For a reminder on how to do this, see *Exercise: Access Library in the Education sandbox, page 7*.
2. In Library, open the **Yearly Regional Performance** dossier in Edit mode. For a reminder on how to do this, see *Exercise: Import a MicroStrategy report and Excel data into a new dossier, page 8*.
3. Ensure that you have:
   - Imported data into the dossier in *Exercise: Import a MicroStrategy report and Excel data into a new dossier, page 8*.
   - Created the grid visualization in *Exercise: Create a grid visualization, page 16*.
   - Created the bubble chart visualization in *Exercise: Create a bubble chart visualization, page 22*.

Replace the dataset with improved data

1. In the Datasets panel, click the menu icon ⋮ next to the Yearly Regional Performance dataset, point to **Replace Dataset With**, and select **New Data**.



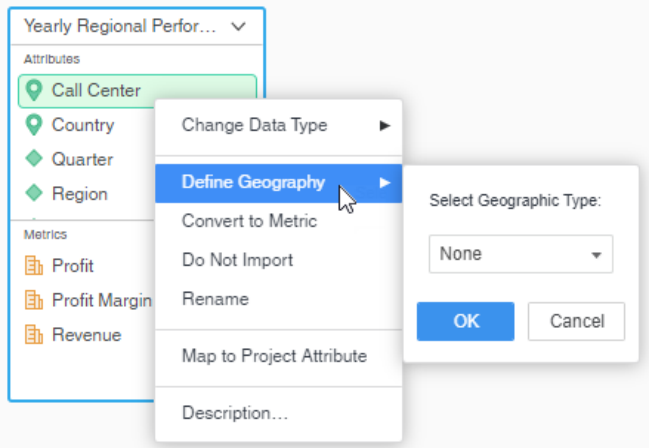
2. On the Connect to Your Data page, select **File from Disk**.
3. On the Upload Your Files page, click **Choose Files**.

4 In the Open window, navigate to the folder where you saved the Exercise Files provided, select **Yearly Regional Performace.xlsx**, and click **Open**.

5 On the Connect to Your Data page, click **Prepare Data**.

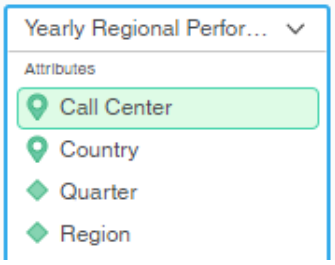
Add a geo role to the Call Center attribute

6 Right-click **Call Center** in the table at the top of the Preview page, and point to **Define Geography**.



7 From the drop-down list under **Select Geographic Type**, select **City**, since Call Center is a city. Click **OK**.

The icon for the Call Center attribute now indicates that it has a geo role, so you can use it to plot locations on a map.

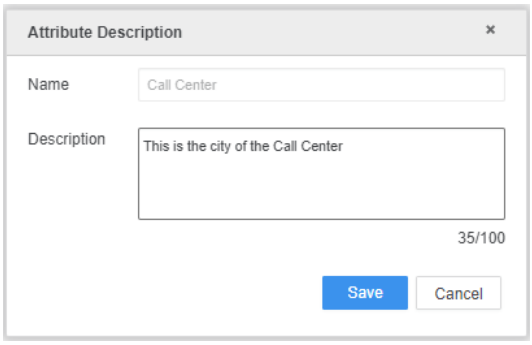


Add a tooltip description

You can add descriptions to attributes and metrics that display as tooltips when users hover over the attribute or metric in a grid in a dossier. Adding descriptions helps users understand the data they are analyzing.

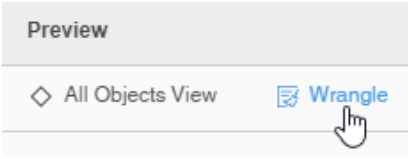
8 Right-click **Call Center** in the table at the top of the Preview page, and select **Description.**

9 In the Description box, type **This is the city of the Call Center** and click **Save**.

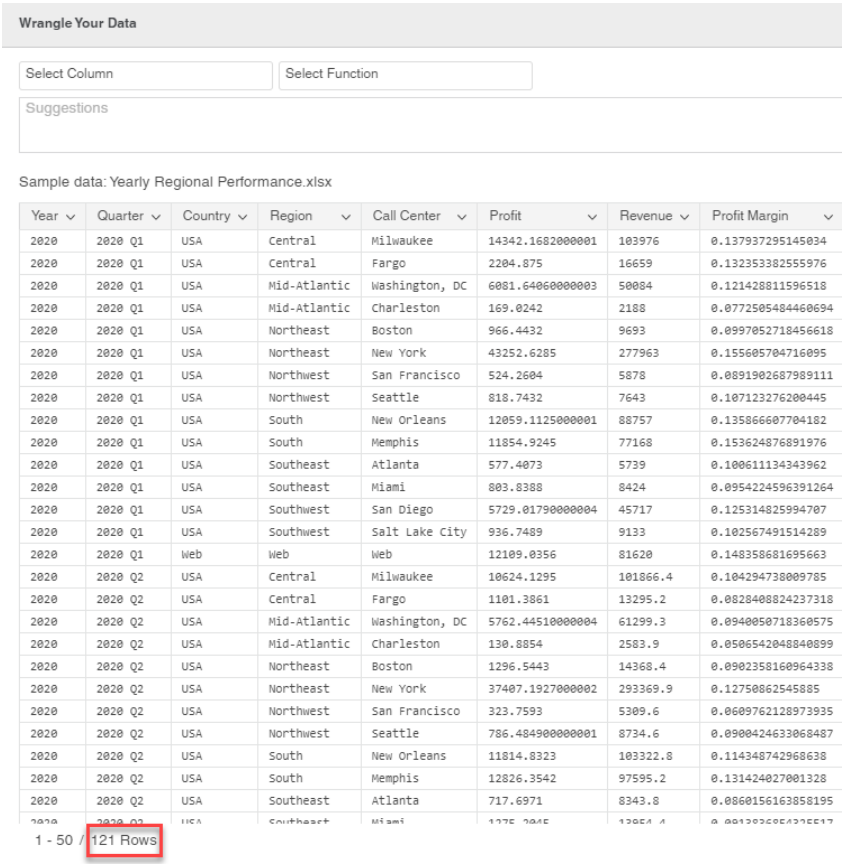


Discover the cause of the extra bubble in the visualization

10 Click **Wrangle**.



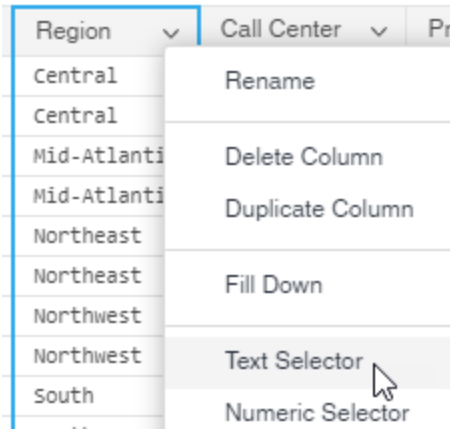
The dataset contains 121 rows of data.



Year	Quarter	Country	Region	Call Center	Profit	Revenue	Profit Margin
2020	2020 Q1	USA	Central	Milwaukee	14342.1682000001	103976	0.137937295145034
2020	2020 Q1	USA	Central	Fargo	2204.875	16659	0.132353382555976
2020	2020 Q1	USA	Mid-Atlantic	Washington, DC	6081.64060000003	50084	0.121428811596518
2020	2020 Q1	USA	Mid-Atlantic	Charleston	169.0242	2188	0.0772505484460694
2020	2020 Q1	USA	Northeast	Boston	966.4432	9693	0.0997052718456618
2020	2020 Q1	USA	Northeast	New York	43252.6285	277963	0.155605704716095
2020	2020 Q1	USA	Northwest	San Francisco	524.2604	5878	0.0891902687989111
2020	2020 Q1	USA	Northwest	Seattle	818.7432	7643	0.107123276200445
2020	2020 Q1	USA	South	New Orleans	12059.1125000001	88757	0.135866607704182
2020	2020 Q1	USA	South	Memphis	11854.9245	77168	0.153624876891976
2020	2020 Q1	USA	Southeast	Atlanta	577.4073	5739	0.100611134343962
2020	2020 Q1	USA	Southeast	Miami	803.8388	8424	0.0954224596391264
2020	2020 Q1	USA	Southwest	San Diego	5729.01790000004	45717	0.125314825994707
2020	2020 Q1	USA	Southwest	Salt Lake City	936.7489	9133	0.102567491514289
2020	2020 Q1	Web	Web	Web	12109.0356	81620	0.148358681695663
2020	2020 Q2	USA	Central	Milwaukee	10624.1295	101866.4	0.104294738009785
2020	2020 Q2	USA	Central	Fargo	1101.3861	13295.2	0.0828408824237318
2020	2020 Q2	USA	Mid-Atlantic	Washington, DC	5762.44510000004	61299.3	0.0940050718360575
2020	2020 Q2	USA	Mid-Atlantic	Charleston	130.8854	2583.9	0.0506542048840899
2020	2020 Q2	USA	Northeast	Boston	1296.5443	14368.4	0.0902358160964338
2020	2020 Q2	USA	Northeast	New York	37407.1927000002	293369.9	0.12750862545885
2020	2020 Q2	USA	Northwest	San Francisco	323.7593	5309.6	0.0609762128973935
2020	2020 Q2	USA	Northwest	Seattle	786.484900000001	8734.6	0.0900424633068487
2020	2020 Q2	USA	South	New Orleans	11814.8323	103322.8	0.114348742968638
2020	2020 Q2	USA	South	Memphis	12826.3542	97595.2	0.131424027001328
2020	2020 Q2	USA	Southeast	Atlanta	717.6971	8343.8	0.0860156163858195

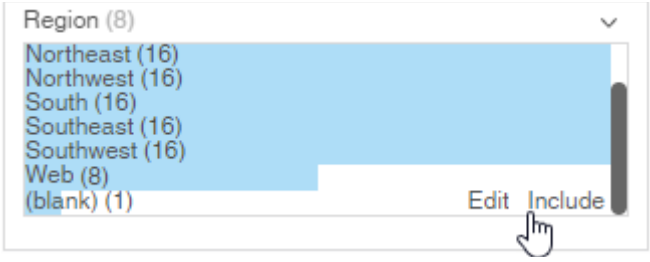
11 Since the bubble colors in the visualization were assigned by region, you need to find the region cell that is blank.

a Click the arrow icon in the Region column header, and select **Text Selector** (selector is another word for filter).

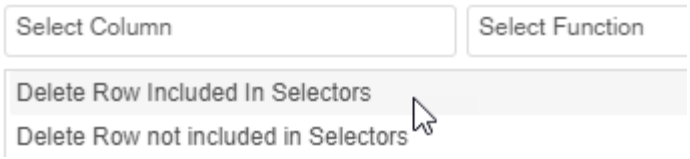


b The list of values in the Region column is displayed at the bottom of the screen. Scroll down, and discover that one cell is blank. This is causing the blank region in the bubble chart.

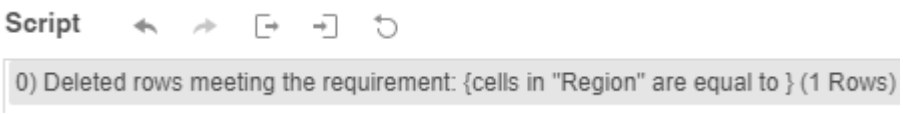
c Hover over the **(blank)** entry, and click **Include** to include the cell in the filter.



d A list of selected functions is displayed at the top of the page. Click **Delete Row Included In Selectors**.



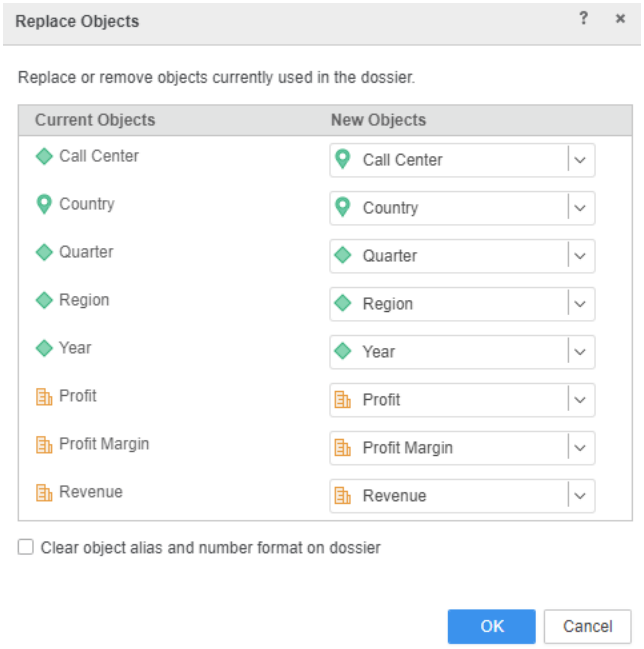
The row with the blank region is deleted from the sample; notice that there are now 120 rows instead of 121. The script, displayed in the upper right corner, contains one entry:



12 Click **Apply** to exit the Wrangle Your Data page.

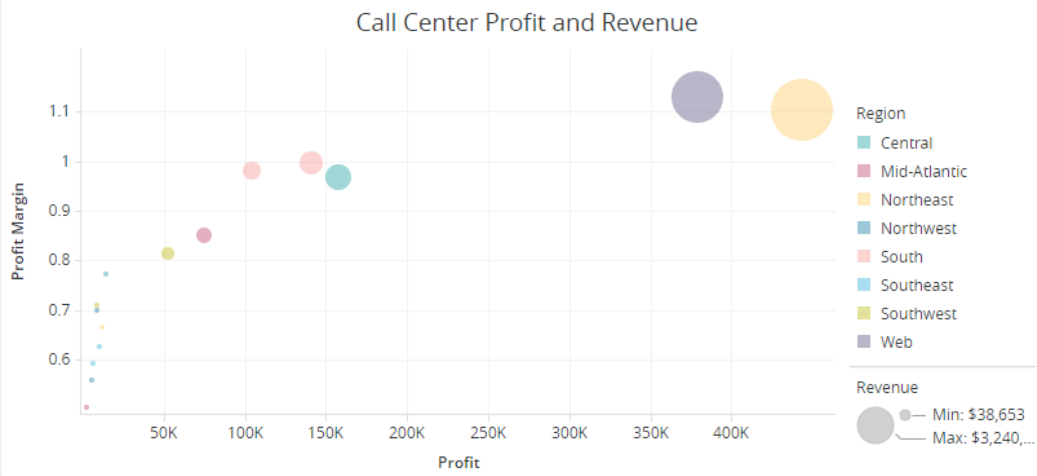
13 Click **Finish** to apply the data wrangling script to the data and import the updated data into the dossier.

14 The Replace Objects window is displayed. Notice that Call Center automatically maps to the new geo-located Call Center. Because you have not removed any of the attributes or metrics, keep the default selections in the New Objects drop-down lists.

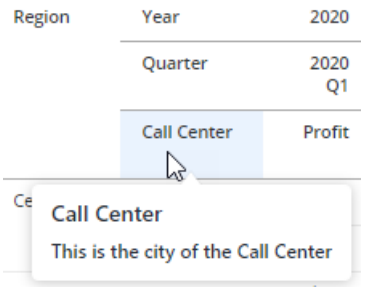


15 Click **OK**.

The extra, blank bubble is no longer displayed on the bubble chart.



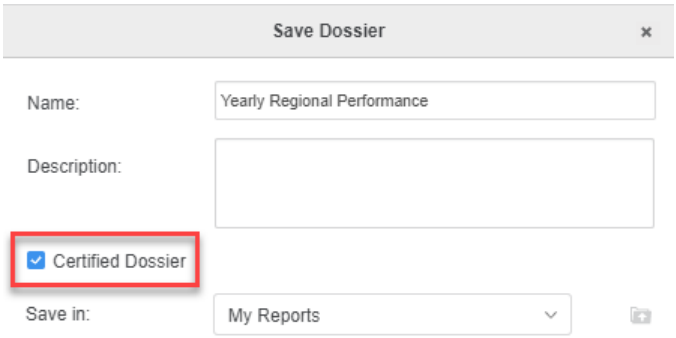
16 Hover over the **Call Center** column header in the grid visualization. Notice that the description you added displays in a tooltip.



17 **Save** your dossier.

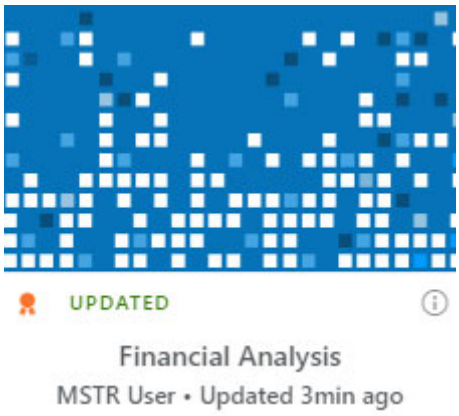
Certifying data to ensure accuracy

Because you make important business decisions based on your data, ensuring that the data is accurate and complete is critical. MicroStrategy provides data governance, the overall management and integrity of data across the enterprise, with dossier certification. Users with a certifier role can ensure data accuracy by certifying a dossier.



Certified dossiers help to maintain a single version of truth while providing flexibility for self-service analytics. Documents, reports, datasets, and HyperIntelligence cards can also be certified, so that you can ensure that these objects comply with your company's data governance policies.

For example, a financial services business analyst verifies that the data in his company's dossier is clean and accurate. He tests a dossier's data, and when he finds the data is complete and accurate, he certifies the dossier. In MicroStrategy Library, which is used for accessing and collaborating on dossiers, the certified dossier displays a badge to indicate that the dossier has been verified.



Exercise: Certify a dossier

In your organization, Departmental Analysts are responsible for verifying data accuracy in the dossiers they create and then certifying them. In this exercise, you certify the dossier that you created in the previous exercises. Then, you change its cover image to help you quickly differentiate it from other dossiers.

Prerequisites

This exercise uses the dossier that you worked on in other exercises. Before you begin:

1. Access Library. For a reminder on how to do this, see *Exercise: Access Library in the Education sandbox, page 7*.

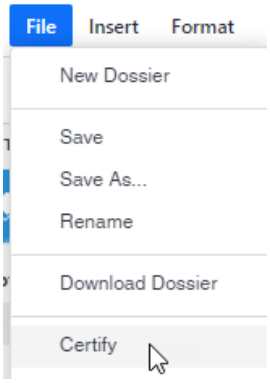
2. In Library, open the **Yearly Regional Performance** dossier in Edit mode. For a reminder on how to do this, see *Exercise: Import a MicroStrategy report and Excel data into a new dossier, page 8*.

3. Ensure that you have:

   - Imported data into the dossier in *Exercise: Import a MicroStrategy report and Excel data into a new dossier, page 8*.
   - Created the grid visualization in *Exercise: Create a grid visualization, page 16*.
   - Created the bubble chart visualization in *Exercise: Create a bubble chart visualization, page 22*.
   - Improved the quality of data structure and values in *Exercise: Improve the data structure during data import, page 33* and *Exercise: Replace a dataset with improved data, page 40*.

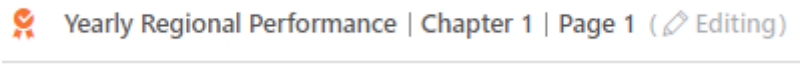
Mark the dossier as certified

1 From the **File** menu, select **Certify**.



You can also certify a dossier by selecting the Certified Dossier check box in the Save As window.

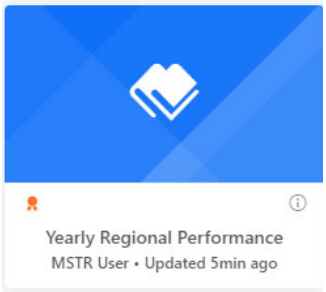
A Certification badge displays next to the dossier title, indicating that the dossier is certified.



View the dossier in your Library home page

2 Click the **Library** icon in the toolbar. Your Library home page displays.

The Yearly Regional Performance dossier displays a certified badge in the bottom left corner.



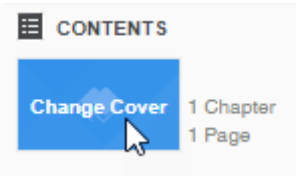
For more information on using Library to access dossiers and collaborate with other users, see the Business User learning path.

Change the dossier's cover image

To help you quickly identify different dossiers in your Library, you can give each dossier a unique cover image. You can choose a MicroStrategy stock image or use a web URL to an image.

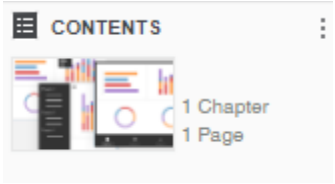
3 On your Library home page, right-click the **Yearly Regional Performance** dossier and select **Edit**.

4 In the Contents panel, hover your cursor over the book icon so that Change Cover is displayed. Click **Change Cover**.



5 In the Change Cover window, select the multiple visualization image (at the far right of the third row, as shown below), and click **Save**.

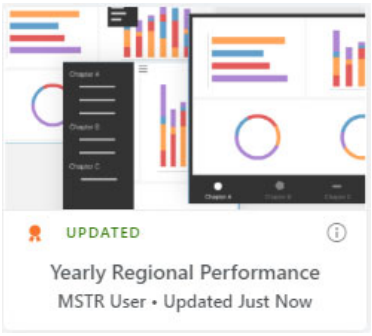
The image displays in the Contents panel and is the thumbnail image in Library.



6 **Save** your dossier.

7 Click the Library icon in the upper left to go to your Library home page.

The dossier displays on your Library home page with the updated cover image.



Copyright Information

Trademark Information