ICT394 Business Intelligence Application Development

# Lab 01: Introduction to BI Application Development

The focus of this lab is to give you a basic introduction to the process through which BI applications are developed.

Previous Edit: 16th February 2021

Last Edit: 1st September 2022

## Aims:

At the completion of this lab you should be able to

* Connect to an **external data source** in Power BI Desktop (PBID)
* Demonstrate a basic understanding of the “shaping and combining” processes in Power BI Desktop
* Save Power BI files to use at the next workshop

## For this lab, you will need:

* Power BI Desktop (PBID) installed
  + It is a free download, so my suggestion is that all students should try to download and install it on their machines themselves.
  + NB: there is no MAC or Unix version – Please run PBID in a Windows 10 VM in Virtual Box.
  + There are a number of installation guides for Power BI Desktop, see: <https://aka.ms/pbiSingleInstaller>
* This lab is based largely on “Getting started with Power BI Desktop”.

<https://docs.microsoft.com/en-us/power-bi/desktop-getting-started>

## Introduction

**For the first three labs**, we will be using PBID, a free, standalone Microsoft application that can assist us with understanding the BI application development process. During the semester, we will be using a variety of other tools as well, **primarily Power BI**.

In this exercise, we are going to **build a dashboard using several data sources** that will help us to **decide the best US states for sales of sunglasses** (not very exciting, but it will help us to become more familiar with PBID so we can do some more exciting stuff later).

The basic steps to create the dashboard are:

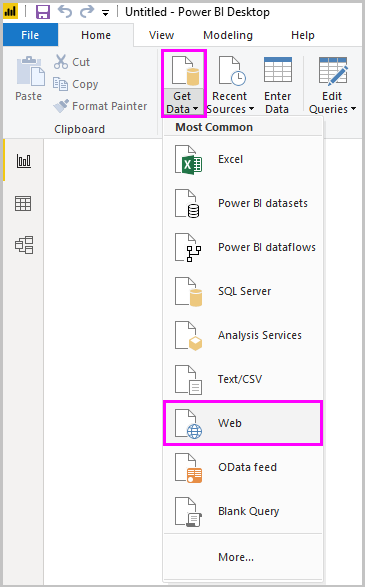
1. Connect to the data
2. Shape the data
3. Create visualizations and reports.

## Connect to Data

There is a very large range of different types of data that can be loaded into PBID; this list continues to grow. To what types of data you can connect to, select Get Data from the External Data ribbon.

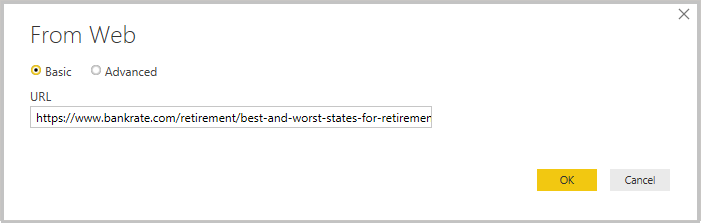
Select More and scroll down through the list.

For now though, we are going to connect to a web page, so select Web from the list



Imagine you're a data analyst working for a sunglasses retailer. You want to help your client **target sunglasses sales where the sun shines most frequently**.

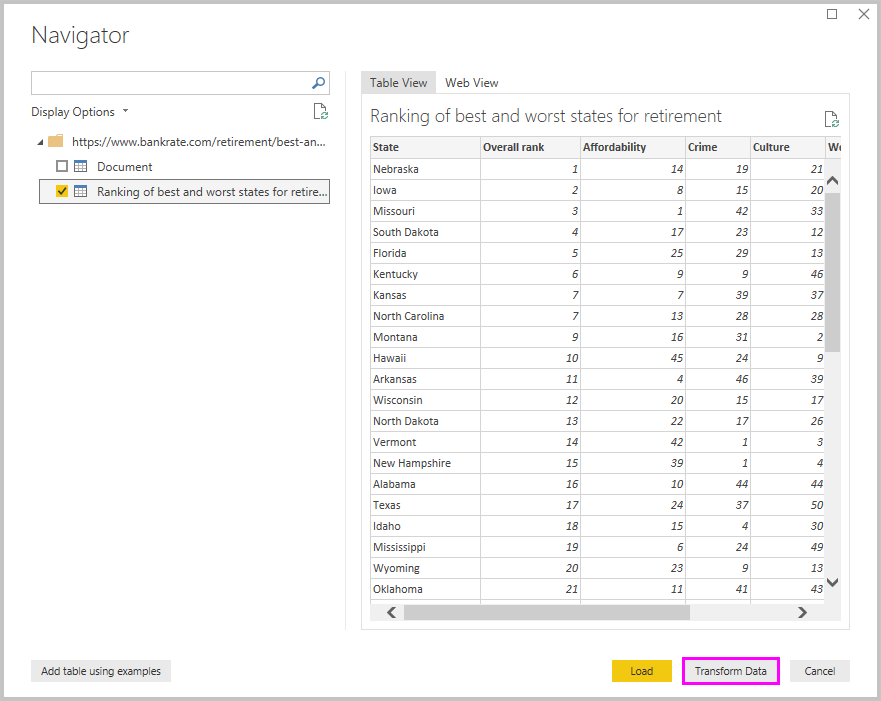
In the **From Web** dialog box, paste the address <https://www.bankrate.com/retirement/best-and-worst-states-for-retirement/>  into the **URL** field, and select **OK**.



If prompted, on the **Access Web Content** screen, select **Connect** to use anonymous access.

The query functionality of Power BI Desktop goes to work and contacts the web resource. The **Navigator** window returns what it found on the web page, in this case a table called **Ranking of best and worst states for retirement**, and a document. You're interested in the table, so select it to see a preview.

At this point you can select **Load** to load the table, or **Transform data** to make changes in the table before you load it.



* If you encounter any issues with the web link, download the **Ranking of best and worst states for retirement** file (Sheet 1.csv) from <https://www.bankrate.com/retirement/best-and-worst-states-for-retirement/>
* Also, the table data used in this guide might change over time.

Graphical user interface, text, application

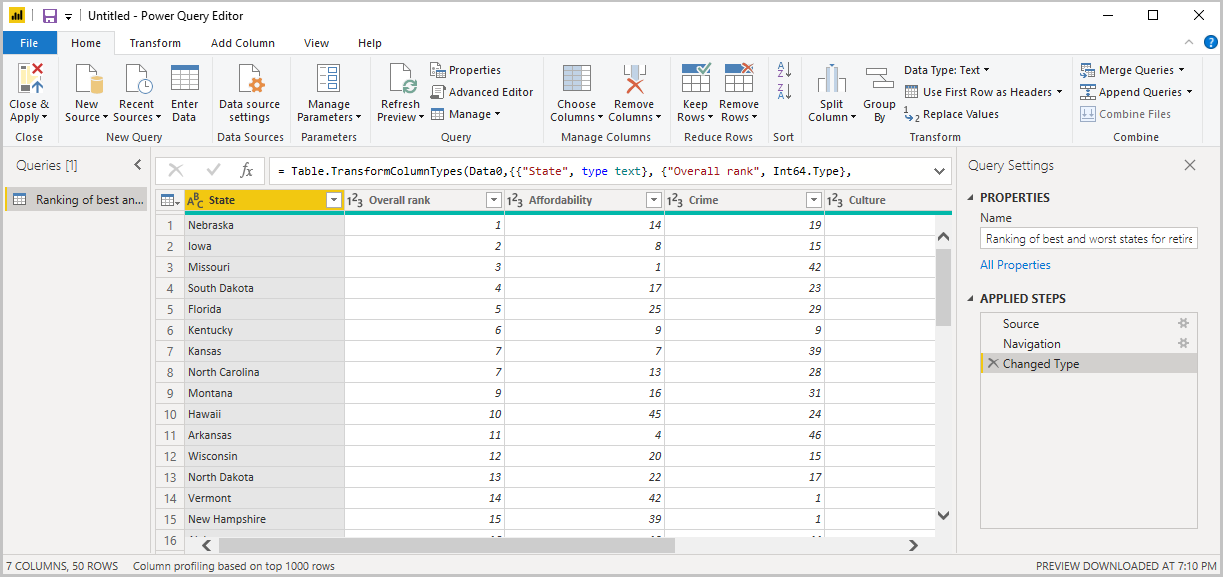
Description automatically generated

**Then** select Get Data from the External Data ribbon, and select Excel as the source data. At this point you can select **Load** to load the table, or **Transform data** to make changes in the table before you load it.

Table

Description automatically generated

**Transform:** When you select **Transform data**, Power Query Editor launches, with a representative view of the table. The **Query Settings** pane is on the right, or you can always show it by selecting **Query Settings** on the **View** tab of Power Query Editor.

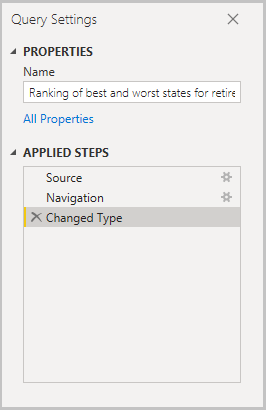


## Shape data

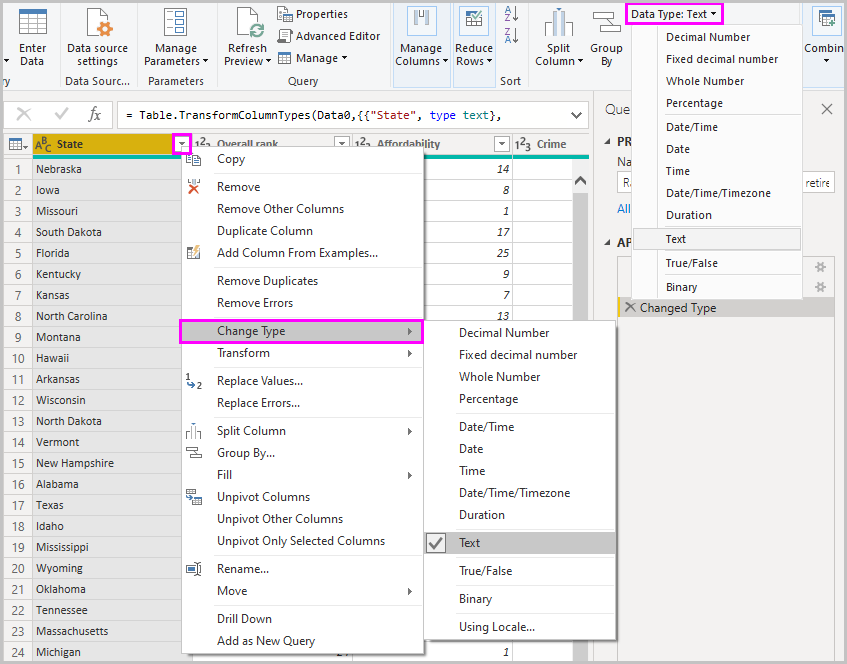
Now that you're connected to a data source, you can adjust the data to meet your needs. To shape data, you provide Power Query Editor with step-by-step instructions for adjusting the data while loading and presenting it. **Shaping doesn't affect the original data source**, only this view of the data.

**Shaping can mean transforming the data,** such as renaming columns or tables, removing rows or columns, or changing data types. Power Query Editor captures these steps sequentially under **Applied Steps** in the **Query Settings** pane. Each time this query connects to the data source, those steps are carried out, so the data is always shaped the way you specify. This process occurs when you use the query in Power BI Desktop, or when anyone uses your shared query, such as in the Power BI service.

Notice that the **Applied Steps** in **Query Settings** already contain a few steps. You can select each step to see its effect in the Power Query Editor. First, you specified a web source, and then you previewed the table in the **Navigator** window. In the third step, **Changed type**, Power BI recognized whole number data when importing it, and automatically changed the original web **Text** data type to **Whole numbers**.

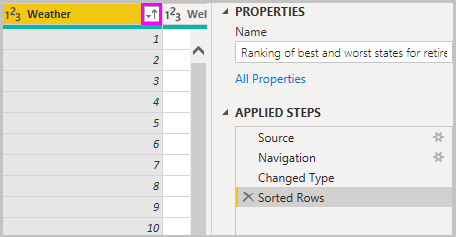


If you need to change a data type, select the column or columns to change. Hold down the **Shift** key to select several adjacent columns, or **Ctrl** to select non-adjacent columns. Either right-click a column header, select **Change Type**, and choose a new data type from the menu, or drop down the list next to **Data Type** in the **Transform** group of the **Home** tab, and select a new data type.

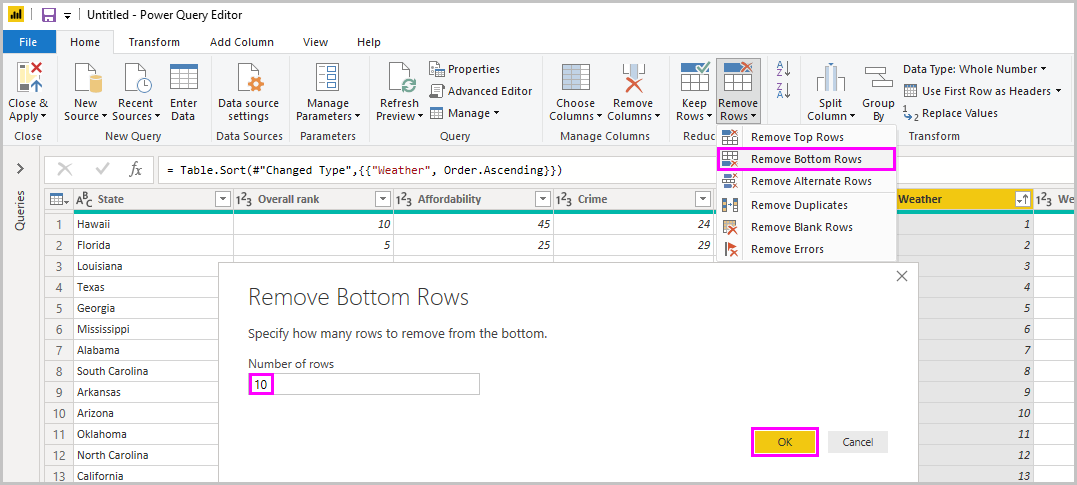


You can now apply your own changes and transformations to the data and see them in **Applied Steps**.

For example, **for sunglasses sales you're most interested in the weather ranking**, so you decide to sort the table by the **Weather** column instead of by **Overall rank**. Drop down the arrow next to the **Weather** header and select **Sort ascending**. The data now appears sorted by weather ranking, and the step **Sorted Rows** appears in **Applied Steps**.



You're not very interested in selling sunglasses to the worst weather states, so you decide to remove them from the table. From the **Reduce Rows** group of the **Home** tab, select **Remove Rows** > **Remove Bottom Rows**. In the **Remove Bottom Rows** dialog box, enter 10, and then select **OK**.



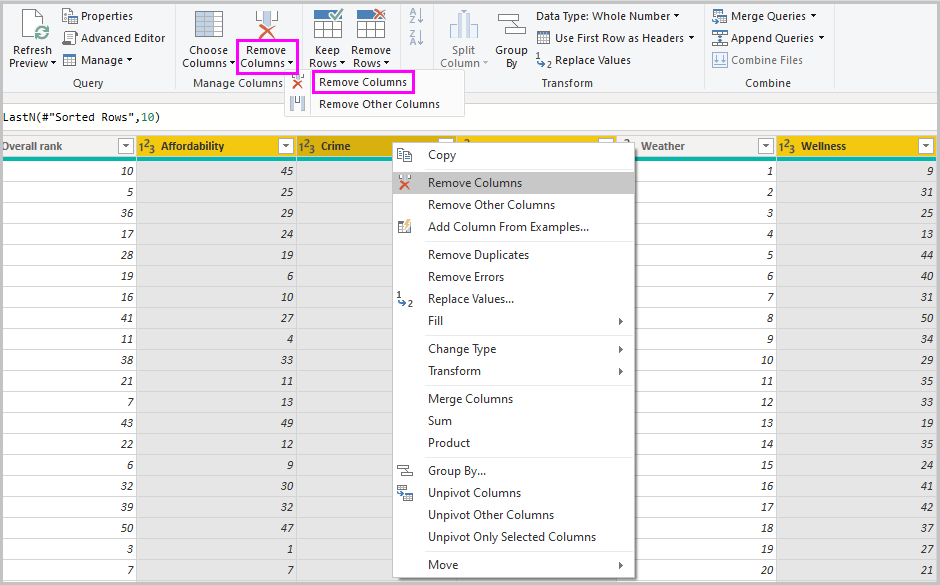
The bottom 10 worst weather rows are removed from the table and the step **Removed Bottom Rows** appears in **Applied Steps**.

A screenshot of a computer

Description automatically generated

You decide the table has too much extra information for your needs and to remove the **Affordability**, **Crime**, **Culture**, and **Wellness** columns. Select the header of each column that you want to remove. Hold down the **Shift** key to select several adjacent columns, or **Ctrl** to select non-adjacent columns.

Then, from the **Manage Columns** group of the **Home** tab, select **Remove Columns**. You can also right-click one of the selected column headers and select **Remove Columns** from the menu. The selected columns are removed, and the step **Removed Columns** appears in **Applied Steps**.

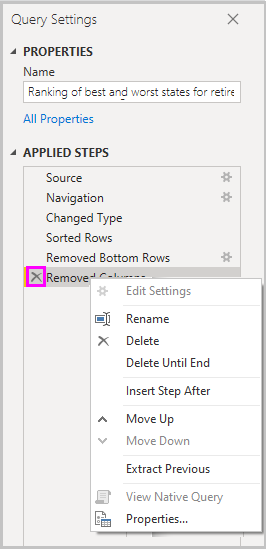


On second thought, **Affordability** might be relevant to sunglasses sales after all. You'd like to get that column back. You can easily undo the last step in the **Applied Steps** pane by selecting the **X** delete icon next to the step. Now redo the step, selecting only the columns you want to delete. **For more flexibility, you could delete each column as a separate step**.

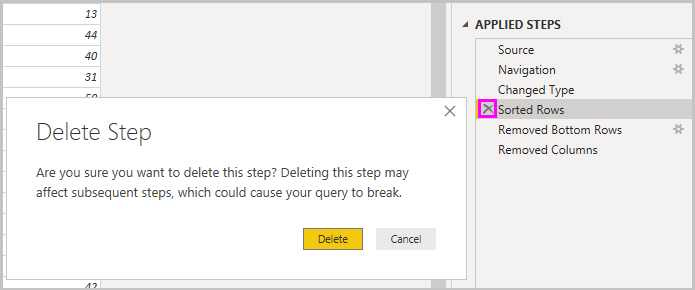
You can right-click any step in the **Applied Steps** pane and choose to delete it, rename it, move it up or down in the sequence, or add or delete steps after it. For intermediate steps, Power BI Desktop will warn you if the change could affect later steps and break your query.

A screenshot of a computer

Description automatically generated

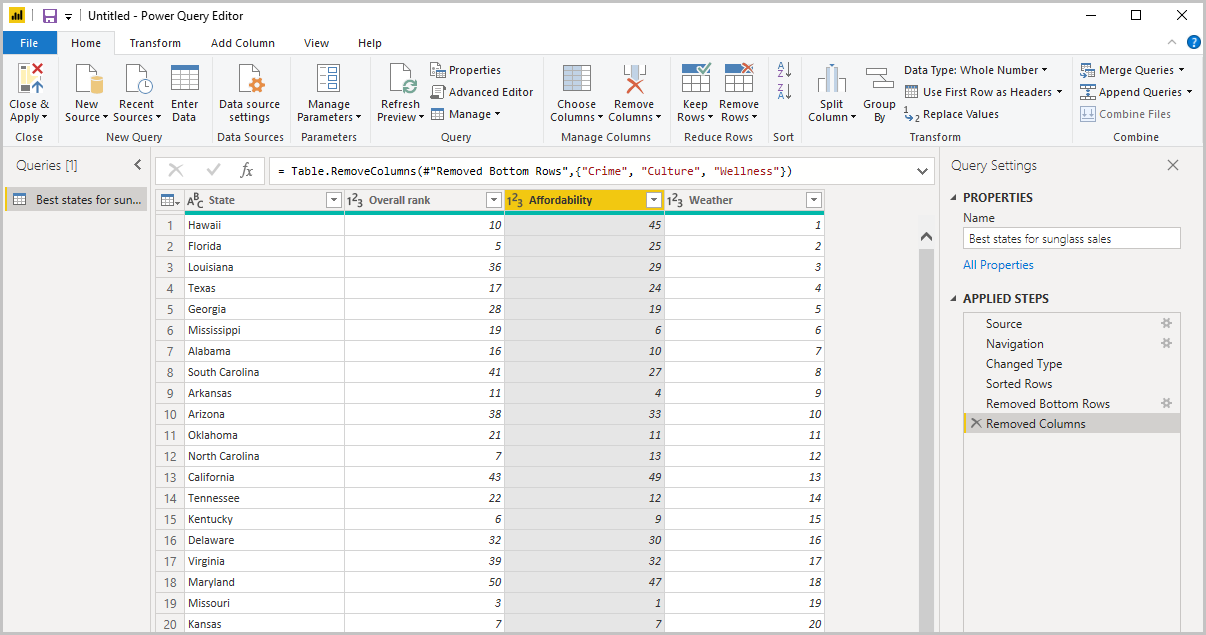


For example, if you no longer wanted to sort the table by **Weather**, you might try to delete the **Sorted Rows** step. Power BI Desktop warns you that deleting this step could cause your query to break. **You removed the bottom 10 rows after you sorted by weather, so if you remove the sort, different rows will be removed**. You also get a warning if you select the **Sorted Rows** step and try to add a new intermediate step at that point.



Finally, you change the table title to be about sunglass sales instead of retirement. Under **Properties** in the **Query Settings** pane, replace the old title with Best states for sunglass sales.

The finished query for your shaped data looks like this:



## Combine data

The data about various states is interesting and will be useful for building additional analysis efforts and queries. But there's one problem, most data out there use two-letter abbreviations for state codes, not the full names of the states. To use that data, you need some way to associate your state names with their abbreviations.

You're in luck. **Another public data source** does just that. However, the data will need a fair amount of shaping before you can combine it with your sunglass table.

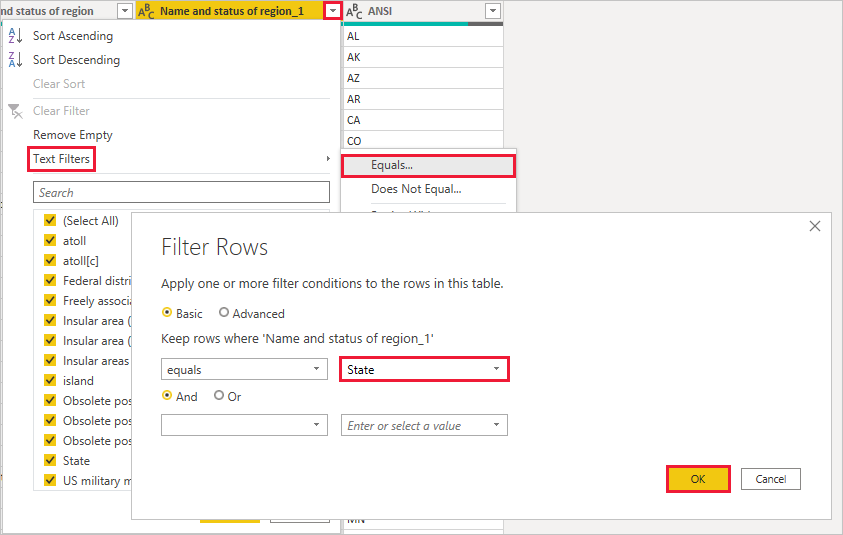
To import the state abbreviations data into Power Query Editor, select **New Source** > **Web** from the **New Query** group on the **Home** tab of the ribbon.

In the **From Web** dialog box, enter the URL for the state abbreviations site <https://en.wikipedia.org/wiki/List_of_U.S._state_abbreviations>

In the **Navigator** window, select the table **Codes and abbreviations for U.S. states, federal district, territories and other regions**, then select **OK**. The table opens in Power Query Editor.

Remove all columns except for **Name and status of region**, **Name and status of region2**, and **ANSI**. To keep only these columns, hold down **Ctrl** and select the columns. Then, either right-click one of the column headers and select **Remove Other Columns**, or, from the **Manage Columns** group of the **Home** tab, select **Remove Other Columns**.

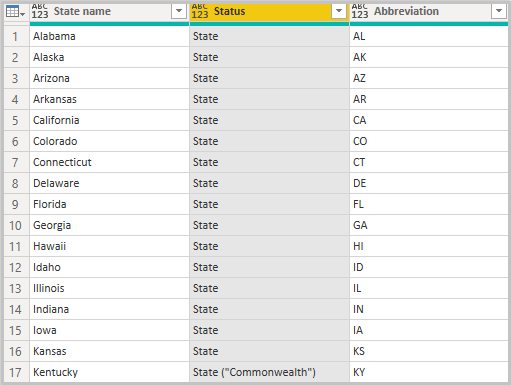
Drop down the arrow next to the **Name and status of region2** column header and select **Filters** > **Equals**. In the **Filter Rows** dialog box, drop down the **Enter or select a value** field next to **equals** and select **State**.



With extra values like **Federal district** and **island** removed, **you now have a list of the 50 states and** their official **two-letter abbreviations**. You can rename the columns to make more sense, for example **State name**, **Status**, and **Abbreviation**, by right-clicking the column headers and selecting **Rename**.

Note that all of these steps are recorded under **Applied Steps** in the **Query Settings** pane.

Your shaped table now looks like this:

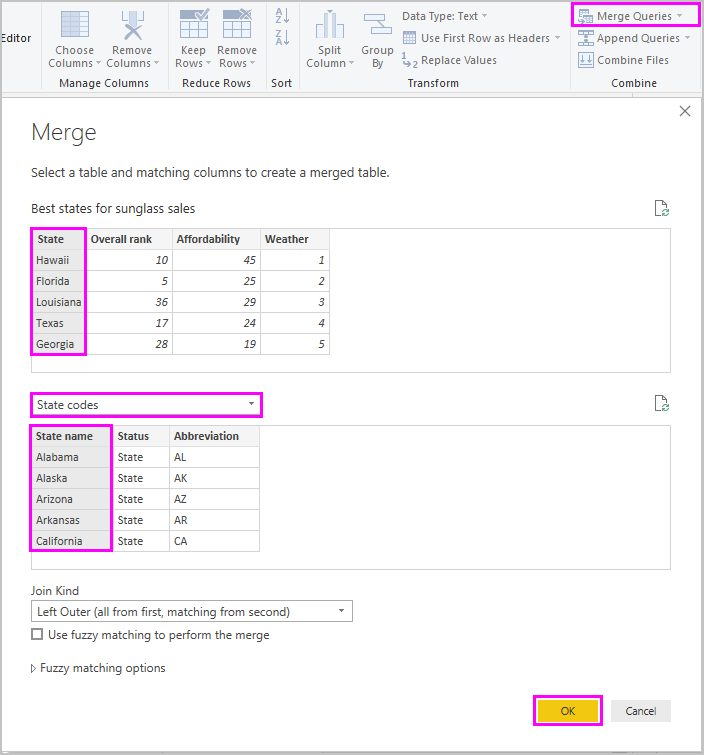


Retitle the table to State codes in the **Properties** field of **Query Settings**. With the **State codes** table shaped, you can **combine these two tables into one**. Since the tables you now have are a result of queries you applied to the data, they're also called queries.

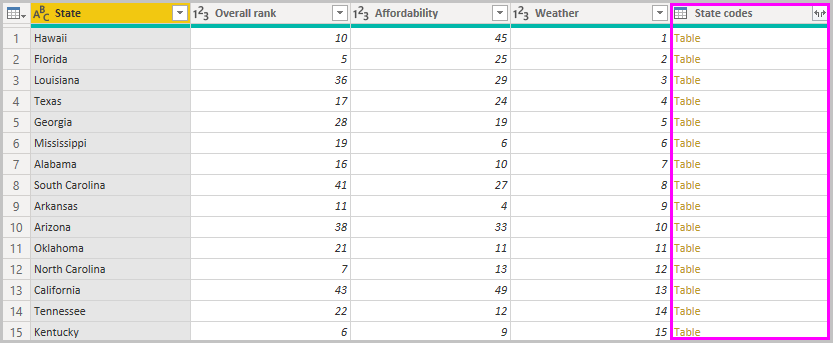
**To merge or to append?** There are two primary ways of combining queries: merge and append. When you have one or more columns you'd like to add to another query, you **merge the queries**. When you have additional rows of data you'd like to add to an existing query, you append the query.

**In this case, you want to merge** the **State codes** query into the **Best states for sunglasses** query. To merge the queries, switch to the **Best states for sunglasses** query by selecting it from the **Queries** pane on the left side of Power Query Editor. Then select **Merge Queries** from the **Combine** group in the **Home** tab of the ribbon.

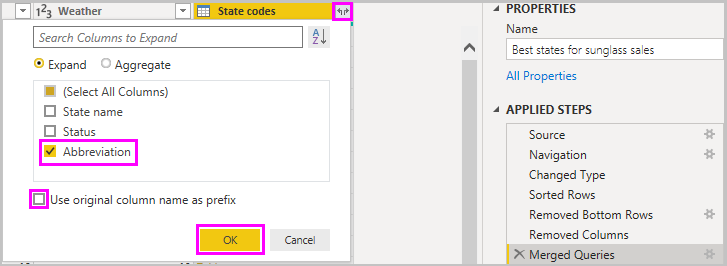
In the **Merge** window, drop down the field to select **State codes** from the other queries available. Select the column to match from each table, in this case **State** from the **Best states for sunglasses** query and **State name** from the **State codes** query. If you get a **Privacy levels** dialog, select **Ignore privacy levels checks for this file** and then select **Save**. Select **OK**.



A new column called **State codes** appears on the right of the **Best states for sunglass sales** table. It contains the state code query that you merged with the best states for sunglass sales query. All the columns from the merged table are condensed into the **State codes** column. You can expand the merged table and include only the columns you want.

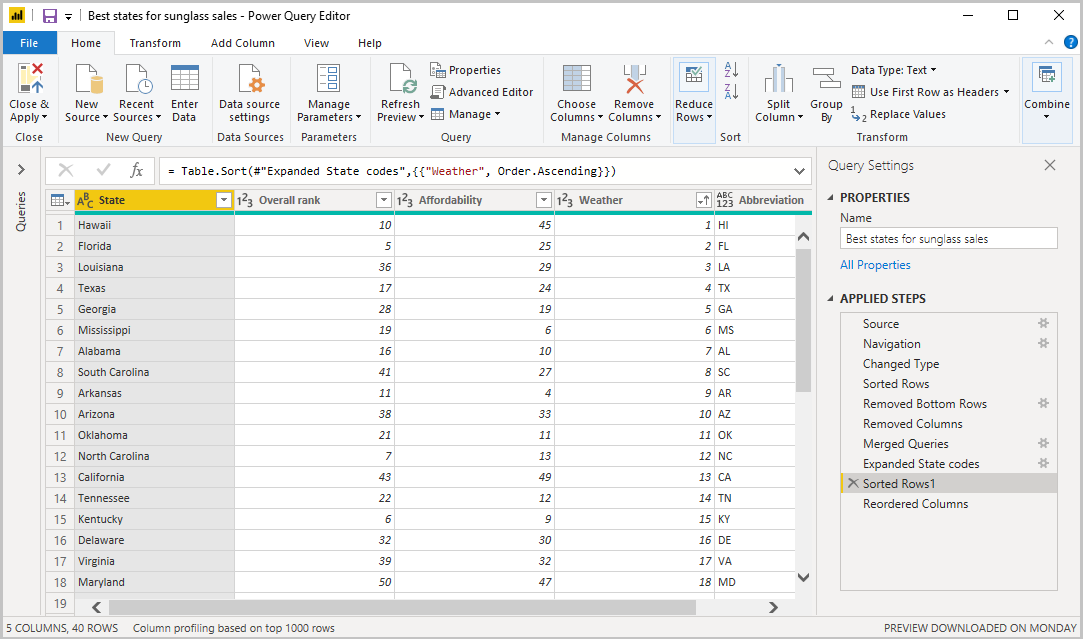


To expand the merged table and select which columns to include, select the **Expand** icon in the column header. In the **Expand** dialog box, select only the **Abbreviation** column. Deselect **Use original column name as prefix**, and then select **OK**.

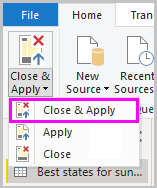


**Note:** You can play around with how to bring in the **State codes** table. Experiment a bit, and if you don’t like the results, just delete that step from the **Applied Steps** list in the **Query Settings** pane. It's a free do-over, which you can do as many times as you like until the expand process looks the way you want it.

You now have a single query table that combines two data sources, each of which has been shaped to meet your needs. This query can serve as a basis for lots of additional, interesting data connections, such as demographics, wealth levels, or recreational opportunities in the states.



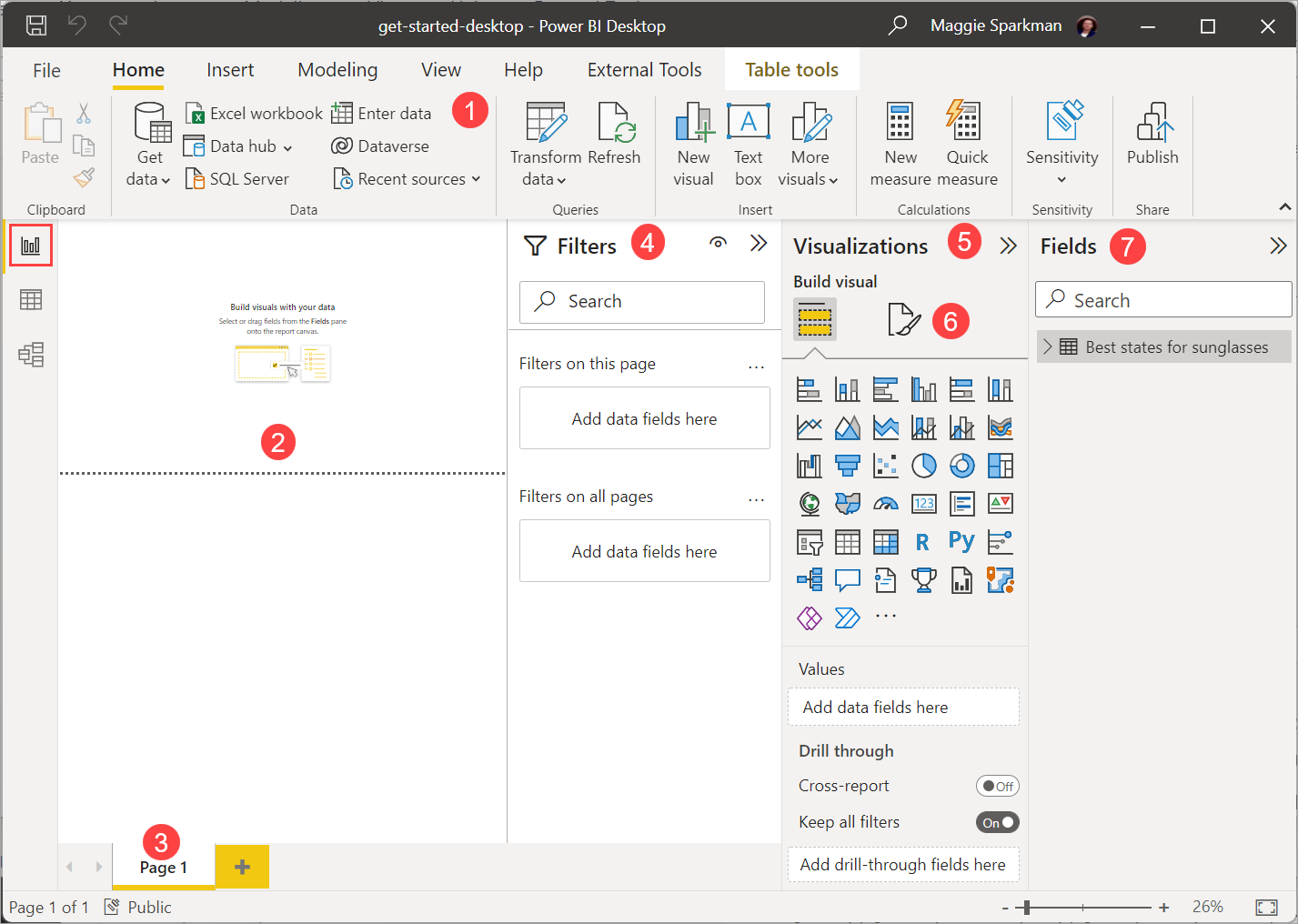
For now, you have enough data to create an interesting report in Power BI Desktop. Since this is a milestone, apply the changes in **Power Query Editor** and load them into Power BI Desktop by selecting **Close & Apply** from the **Home** tab of the ribbon. You can also select just **Apply** to keep the query open in Power Query Editor while you work in Power BI Desktop.



You can make more changes to a table after it is loaded into Power BI Desktop and reload the model to apply any changes you make. To reopen **Power Query Editor** from Power BI Desktop, select **Edit Queries** (or **Transform data from the Queries group**) on the **Home** tab of the Power BI Desktop ribbon.

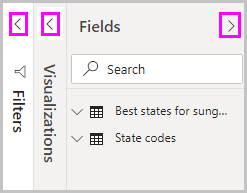
## Build reports

In Power BI Desktop **Report** view, you can build visualizations and reports. The **Report** view has **seven** main areas.

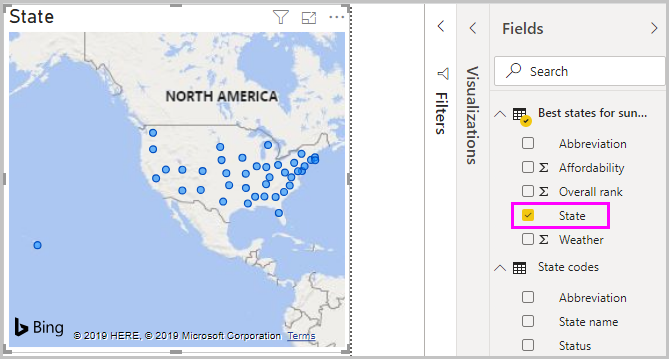


1. The ribbon at the top, which displays common tasks associated with reports and visualizations.
2. The canvas area in the middle, where you create and arrange visualizations.
3. The pages tab area at the bottom, which lets you select or add report pages.
4. The **Filters** pane, where you can filter data visualizations.
5. The **Visualizations** pane, where you can add, change, or customize visualizations, and apply drill through.
6. The **Format** pane, where you design the report and visualizations.
7. The **Fields** pane, which shows the available fields in your queries. You can drag these fields onto the canvas, the **Filters** pane, or the **Visualizations** pane to create or modify visualizations.

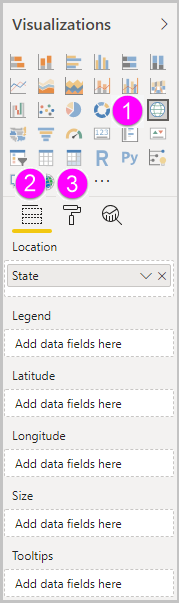
You can expand and collapse the **Filters**, **Visualizations**, and **Fields** panes by selecting the arrows at the tops of the panes. Collapsing the panes provides more space on the canvas to build cool visualizations.



To create a simple visualization, just select any field in the fields list or drag the field from the **Fields** list onto the canvas. For example, drag the **State** field from **Best states for sunglass sales** onto the canvas, and see what happens.



Look at that! Power BI Desktop recognized that the **State** field contained geolocation data and automatically created a map-based visualization. The visualization shows data points for the 40 states from your data model. The **Visualizations** pane shows information about the visualization and lets you modify it.

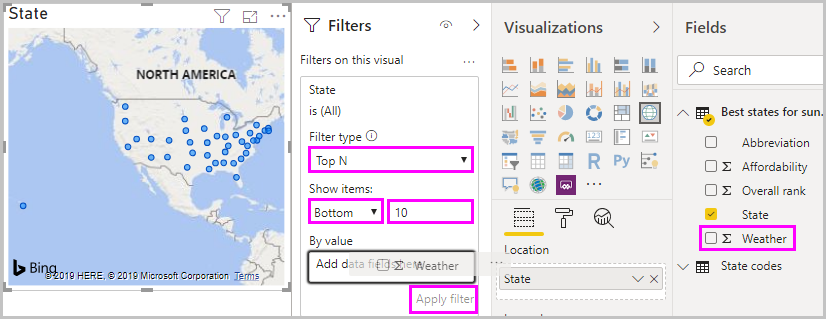


1. The icons show the type of visualization created. You can change the type of a selected visualization by selecting a different icon, or create a new visualization by selecting an icon with no existing visualization selected.
2. The **Fields** option in the **Visualization** pane lets you drag data fields to **Legend** and other field wells in the pane.
3. The **Format** option lets you apply formatting and other controls to visualizations.

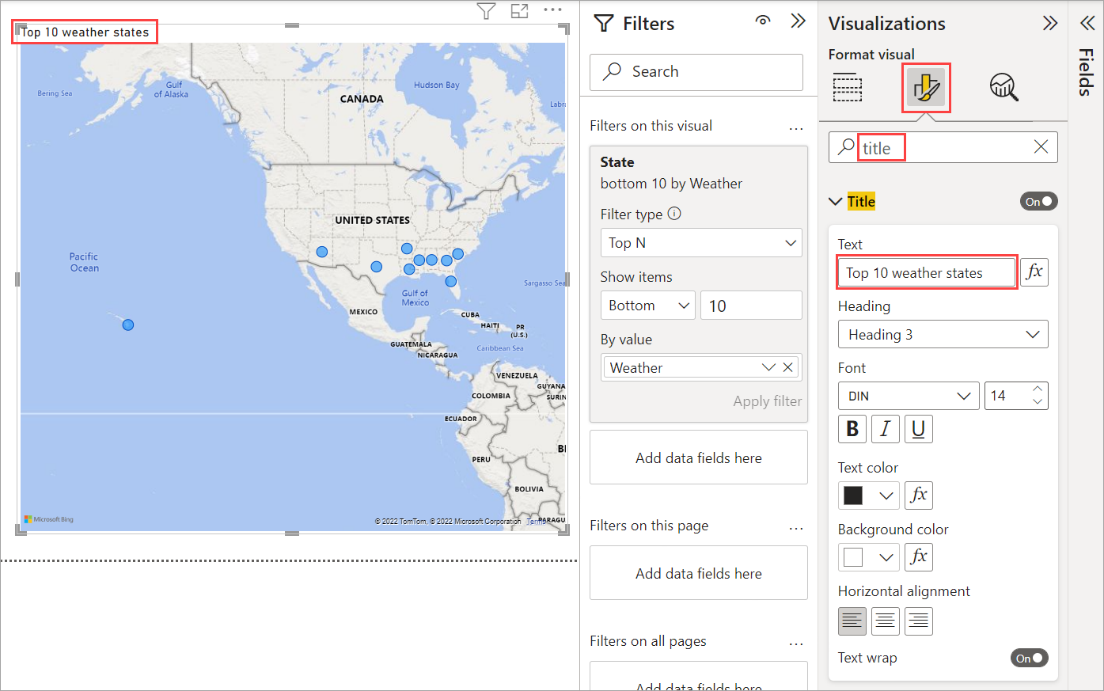
The options available in the **Fields** and **Format** areas depend on the type of visualization and data you have.

You want your map visualization to **show only the top 10 weather states**. To show only the top 10 states, in the **Filters** pane, hover over **State is (All)** and expand the arrow that appears. Under **Filter type**, drop down and select **Top N**. Under **Show items**, select **Bottom**, because you want to show the items with the lowest numerical ranks, and enter 10 in the next field.

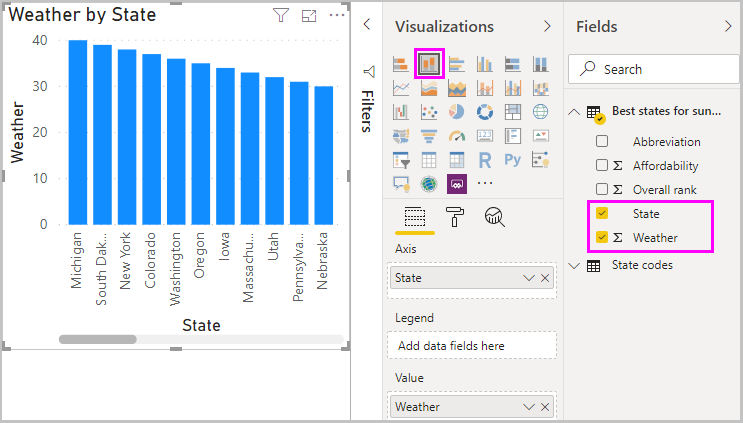
Drag the **Weather** field from the **Fields** pane into the **By value** field, and then select **Apply filter**. You now see only the top 10 weather states in the map visualization.



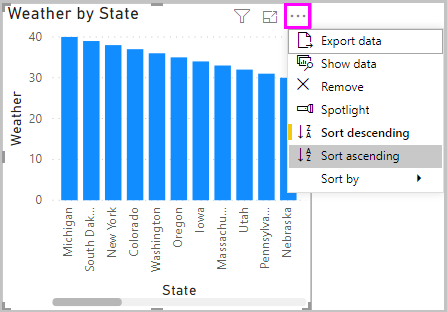
Retitle your visualization by selecting the **Format** icon in the **Visualization** pane, selecting **Title**, and typing Top 10 weather states under **Title text**.



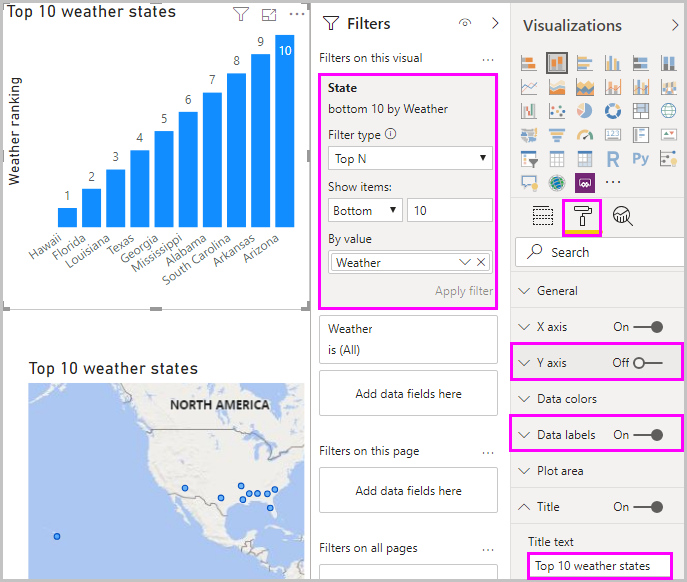
To add a visualization that **shows the names of the top 10 weather states** and their ranks from 1 to 10, select a blank area of the canvas and then select the **Column chart** icon from the **Visualization** pane. In the **Fields** pane, select **State** and **Weather**. A column chart shows the 40 states in your query, ranked from highest to lowest numerical rank, or worst to best weather.



To switch the order of the ranking so that number 1 appears first, select the **More options** ellipsis at the upper right of the visualization, and select **Sort ascending** from the menu.



To limit the table to the top 10 states, apply the same bottom 10 filter as you did for the map visualization. Retitle the visualization the same way as for the map visualization. Also, in the **Format** section of the **Visualization** pane, change **Y axis** > **Axis title** from **Weather** to Weather ranking to make it more understandable. Then, turn the **Y axis** selector to **Off**, and turn **Data labels** to **On**. Now, the top 10 weather states appear in ranked order along with their numerical rankings.

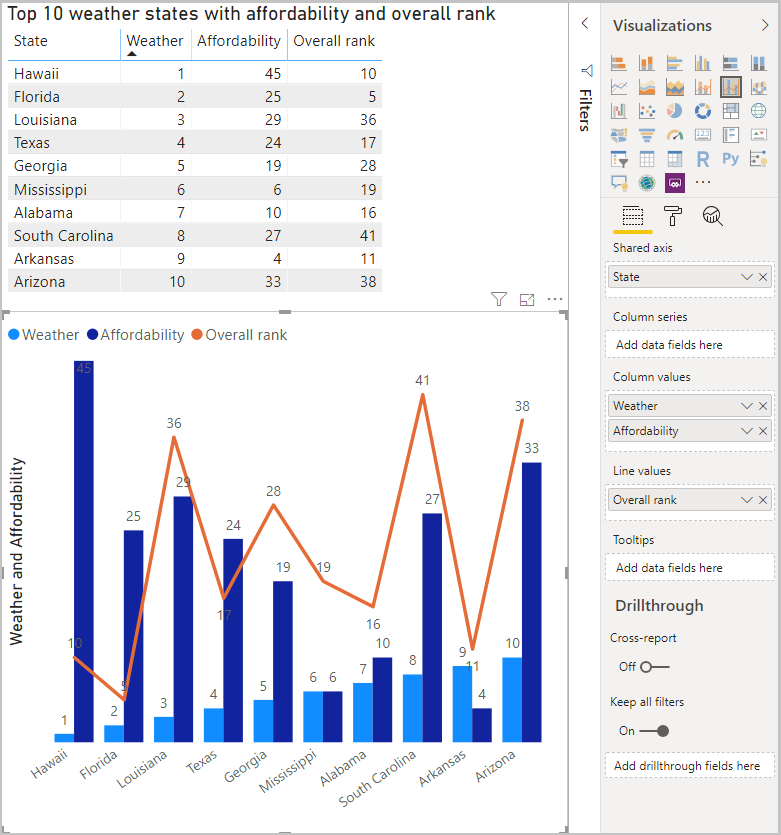


Or the window below – depending on the version of Power BI.

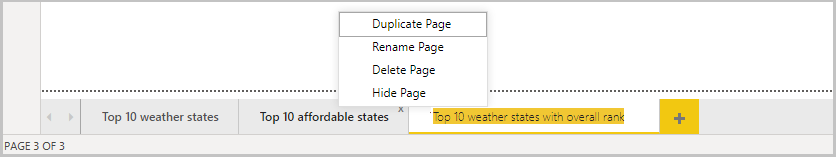
A screenshot of a computer

Description automatically generated with medium confidence

You can make similar or other visualizations for the **Affordability** and **Overall ranking** fields or combine several fields into one visualization. There are all sorts of interesting reports and visualizations you can create. These **Table** and **Line and clustered column chart** visualizations shows the top 10 weather states along with their affordability and overall rankings:

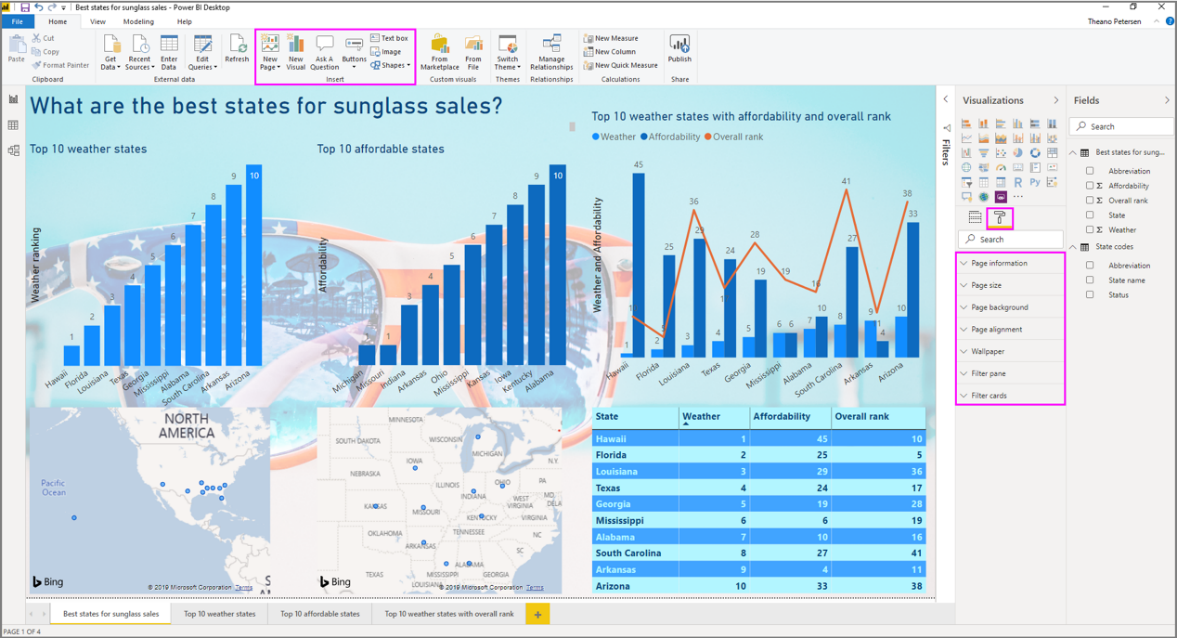


You can show different visualizations on different report pages. To add a new page, select the **+** symbol next to the existing pages on the pages bar, or select **Insert** > **New Page** in the **Home** tab of the ribbon. To rename a page, double-click the page name in the pages bar, or right-click it and select **Rename Page**, and then type the new name. To go to a different page of the report, select the page from the pages bar.



You can add text boxes, images, and buttons to your report pages from the **Insert** group of the **Home** tab. To set formatting options for visualizations, select a visualization and then select the **Format** icon in the **Visualizations** pane. To configure page sizes, backgrounds and other page information, select the **Format** icon with no visualization selected.

When you finish creating your pages and visualizations, select **File** > **Save** and save your report.



For more information about reports, see [Report View in Power BI Desktop](https://docs.microsoft.com/en-us/power-bi/desktop-report-view).