**Exercise 1 Power Query to import public health data**

# Introduction

This is the document that I printed.

Pleas contact me at [Rloski@SQLMovers.com](mailto:Rloski@SQLMovers.com) if you have questions or are stuck.

# Market Saturation

C:\Power BI Data\Market Saturation & Utilization State-County\2024\Market\_Saturation\_and\_Utilization\_State\_County\_Dataset\_Release\_20\_20250122.csv

This dataset is available from <https://data.cms.gov/summary-statistics-on-use-and-payments/program-integrity-market-saturation-by-type-of-service/market-saturation-utilization-state-county>.

This next dataset comes from the Centers for Medicare & Medicaid Services (CMS).

The purpose of this dataset is to give us insight into the types of healthcare services that are being utilized across different counties. It helps us identify areas where there’s potentially high demand for services or, conversely, regions where there may be an oversupply of providers.

The dataset includes information covering roughly two years and is structured at the county level, giving us a localized view of healthcare needs and resource distribution.

One important note: some of the numerical data in this file is formatted in ways that might not translate cleanly into Power BI or other analysis tools. For example, certain values may use inconsistent formatting, which means we’ll need to check and possibly clean those numbers to ensure our analysis is accurate.

We’ll look at how to import this data and prepare it so we can analyze trends and uncover insights about healthcare service needs across various regions.

I need to summarize where I am going with this. It is helpful to note that this is going to require 3 final queries.

Step 1. Steps for importing and transforming the cms county-level dataset in power query

**1. Import the Data from CSV**

1. Go to **Home > New Source > Text/CSV**.
2. Browse to the file path where the CSV exists (the file Market\_Saturation\_and\_Utilization\_State\_County\_Dataset\_Release\_20\_20250122.csv in the “C:\Power BI Data\Market Saturation & Utilization State-County\2024” folder).
3. Step 2. Review the options
4. Step 3. Click transform data.

**2. Review Results**

Properties. Note the name.

Step 4. Note the applied steps. click on each step and view the function.

1. Source:
   1. File.Contents reads the file
   2. Csv.Document converts the file that is in comma delimited format into a table
2. Promote Headers. Takes the first row and makes values the column headers
3. Step 5. Change type: converts all of the columns to the appropriate data types

**2. General cleanup**

To organize the query, we need to make a couple changes. This query is not going to be shown directly to the report users. It will be used to create other tables. We will rename the query and put it in a special group for what I call “data source” queries.

1. Step 6. Rename the query to market\_saturation\_ds
2. Right click on query and Move to Group “Data Sources.” If the group doesn’t exist, then click New Group and name it “Data Sources.”
3. Right click on the Market\_Saturation\_DS query and uncheck the Enable load item   
   A screenshot of a computer

   AI-generated content may be incorrect.

Next, it is very possible that the source file will be in a different location and may have a different name. We are going to create a parameter.

1. Step 7. Click on the source applied step
2. Copy the full path for the csv file
3. Step 8. Create a new parameter
   1. Step 9. Click manage parameter
   2. Step 10. Click new parameter
   3. Step 11. Paste the full for the csv file into the current value text box (make sure to remove the leading and following quotes from the file name)
   4. Step 12. Name the parameter “market\_saturation\_path”
   5. Step 13. Change the type to “text”
   6. Step 14. Click ok
4. Right click the Market\_Saturation\_Path parameter and move to the “Parameters” group, creating if it does not exist
5. Go back to the Source for the “Market\_Saturation\_DS”
6. Delete the path to the csv file, including the double quotes
7. Step 15. Type market\_saturation\_path in place of the path
8. Source should look like the following and the rest of the steps should not change  
   

**3, Note the data issues**

There are a number of data issues to be aware of in the original data source. Some of them are handles. Some are not.

Step 16. Click on the applied step “promoted headers.”

* fips columns: numbers padded with zero
* number columns: commas in the number
* percentage columns have a percent sign
* dollar sign in payments
* “No” in moratorium column. This would be equivalent to false.

The problem is whether Power BI will convert these column values to the expected types

Step 17. Click the changed type step. view what changes were made. did power query convert the column values correctly.

Step 18. Note that the percentage divided the number by 100 and then changed the type to percentage.

“No” did not convert to a Boolean (true/fale)

**2. Correct data types**

Step 19. Review the data types for the columns. note the state\_fips data types as integer. this removes the leading zero from many of the states’ records. we need to get it back.

1. Step 20. Select the state\_fips column
2. Right click the column header
3. Step 21. Click the “change type” menu item, followed by text
4. Step 22. Click “replace current” on the dialog box

A screenshot of a computer

AI-generated content may be incorrect.

When you have a change type from text to number and change it back, you have the opportunity to change from number to text after first changing the type to number. This removes the formatting (like commas, zero padding) before changing to text. If you replace current, then the output is the string format up to that poin.

**2. Remove extra columns and rows**

There are a number of columns that I don’t see a use for. Rather than loading that data, we will remove those columns. At the same time, in order to keep the grain of the tables at the county/parish level, we will filter out rows that do not have a county fips

1. We are going to remove all of the columns to the right of the moratorium colum
   * Step 23. Select all of the columns in the market\_saturation\_ds to the right of the moratorium column
   * Right-click and click “Remove Columns”
2. Next we filter out all of the rows not related to COUNTY.
   * Step 24. Click on the down triangle on the “aggregation\_level” column
   * Unselect “(Select All)” and select “COUNTY”
3. We now need to limit the records to the neighboring states
   * Step 25. Create a new parameter with the name reportstates
   * Set the value to “LA,TX,MS,AR,TN”
   * The datatype is text
   * Move to Parameters Group
4. Use the parameter in the filter:
   * Edit the Filtered Rows step
   * After aggregation\_level] = "COUNTY"   
     add the following function  
     and List.Contains(Text.Split(ReportStates,","), [state])

At this point the data should filter to just the five states

**7. Convert moratorium to Logical**

* “moratorium” is a text field (with values “Yes”/“No”), we need to convert it to a True/False: The strings true and false (all lower case) will convert to Boolean (True/False)
  + Step 26. Select the column → transform > replace values
    - Step 27. Replace “yes” → true
    - Step 28. Replace “no” → false
  + Step 29. Change data type to true/false.

**8. Create LocationID Column**

To create a unique identifier for location:

1. Step 30. Select state\_fips and county\_fips columns
2. Go to **Add Column > Column From Examples**.
3. Step 31. Rename the column locationid
4. In the new column, type a few examples concatenating the value from state\_fips with the value from county\_fips
5. Power Query will auto-generate a formula like:

Text.Combine({[state\_fips],[county\_fips]})

**9. Split the data source into three queries**

Now we need to create three queries from this data source (Market\_Saturation\_DS):

* Market\_Saturations\_Counties (information about counties)
* Market\_Saturation\_Measures (The type of service)
* Market\_Saturation\_Facts (the numbers we are reporting)

Right click on the Martet\_Saturation\_DS query and click Reference. This creates a new query that references Market\_Saturation\_DS.

Do the same thing two other times

Move all three of these new queries to the Market Saturation group

Step 32. Rename each of these market saturation queries to the names of the thre tables above:

* Market\_Saturation\_DS (2) -> Market\_Saturation\_Counties
* Market\_Saturation\_DS (3) -> Market\_Saturation\_Measures
* Market\_Saturation\_DS (4) -> Market\_Saturation\_Facts

**9.** Market\_Saturation\_Counties

This table will contain the state, county, state\_fips, county\_fips and LocationID columns. The key for this will be the LocationID, which will also be in the Market\_Saturation\_Facts table.

1. Remove all of the columns except the ones related to the county
   1. Step 33. Select the state, county, state\_fips, county\_fips and locationid columns.
   2. Right click on one of the selected columns and click “Remove Other Columns”
2. We need to remove the duplicates
   1. Step 34. Select the locationid column
   2. Right click on the LocationID column and select “Remove Duplicates”
3. Finally, we move the LocationID so that it in the left. Drage the column to the left

**9A.Market\_Saturation\_Measures**

For this table we are going to have one record for each type\_of\_service. We will add an integer index column which we will use in the main Market\_Saturation\_Facts table.

1. Step 35. Select the type\_of\_service column and remove other columns
2. Step 36. Select the type\_of\_service column and remove duplicats
3. Step 37. Click the index column (in the add column tab) and select from 1)
4. Before we go forward, we need to rename the Index column to Type\_Of\_Service\_Key
   1. In the formula bar, find the word “Index” in double quotes.
   2. Step 38. Replace that text with type\_of\_service\_key
5. Move the Type\_Of\_Service\_Key to the left

**9B. Market\_Saturation\_Facts**

This is the table that has all of our numbers. Before we can use this table, we need to make a couple changes. We need to remove the county columns. We need to replace the type of service column with the Type\_Of\_Service\_Key. Finally, the reference period has two dates in a string. We need to put those two dates in their own columns.

1. Remove columns
   1. Step 39. Select the columns aggregation\_level to county\_fips
   2. Right click and select Remove Columns
   3. **Don’t remove the Location\_ID**
2. We need to get the Type\_Of\_Service\_Key from the Market\_Saturation\_Measures
   1. Step 40. Click merge queries in the home tab
   2. Step 41. Select market\_saturation\_measures in the dropdown menu

Step 42. Select the type-of\_service column in both tables

Step 43. Note the comment at the bottom that all of the records in the first table have a match in the second table

* 1. Step 44. Click ok
  2. Market\_Saturation\_Measures is now a new column, with Table as the value

1. We need to expand the table
   1. Step 45. Click the two arror ins the market\_saturation\_measures table
   2. Check only the Type\_Of\_Service\_Key column
   3. Uncheck the “Use original column name as prefix”
   4. Step 46. Click ok

A screenshot of a computer

AI-generated content may be incorrect.

1. Split the reference period into Period\_StartDate and Period\_EndDate
   1. Step 47. Select reference\_period column
   2. Right click and select Add Column From Examples
   3. Step 48. Rename column period\_startdate
   4. Step 49. Type in the first date
   5. The values suggested are correct
   6. The expression looks good
   7. Step 50. Click ok
   8. Do the same of the second date (Period\_EndDate)
   9. Remove reference\_period and type\_of\_service columns
   10. Step 51. Change type for period\_startdate and period\_enddate to date
2. Move the LocationID, Type\_Of\_Service\_Key, Period\_StartDate and Period\_EndDate to the front of the table (left)

Finally, Click Close and apply

# Louisiana Hospital Administrative Regions

Before we dive into our next exercise, I want to introduce you to the Excel workbook we’ll be working with. This isn’t just any clean, perfectly structured spreadsheet. Instead, it’s a realistic example of the kind of Excel file you might get from a business user or another system.

Take a look at this Excel file ("C:\Power BI Data\Census\Louisiana DOH Regions.xlsx"). In this workbook, I’ve created a report-style layout that maps Louisiana hospital regions to parishes. But to make it more interesting—and challenging—it includes some typical ‘Excel formatting traps’ that can cause problems when importing data into Power BI.

For example:

* The region names are shown in a stair-step format using merged cells. Visually, this looks nice in Excel, but in Power BI, it creates missing values and makes the data hard to work with. We’ll need to find a way to fill those missing region values so each row has the correct data.
* There’s a decorative map image included to show the regions visually. While helpful in a report, it’s irrelevant for analysis and something we’ll ignore or remove in Power BI.
* There’s extra header information at the top of the sheet that’s not part of the data table itself. This is another common real-world scenario, and we’ll have to clean it up before loading the data.

These kinds of formatting issues are exactly what you’ll run into in real projects. So the goal of this exercise is to show you how to handle imperfect Excel files and get them ready for analysis in Power BI.

Close the Excel Workbook before going on to the next step. Excel will prevent you from importing if the Workbook is open in Excel.

## Import Excel

Step 52. Class exercise: importing and cleaning excel data in power bi

Step 53. Scenario

You have an Excel workbook that contains Louisiana parish data organized for reporting purposes. However, it includes merged cells, extra header rows, and formatting designed for human reading rather than data analysis. Your goal is to import this workbook into Power BI and clean the data for analysis.

Step 54. Steps

Step 55. Import the excel workbook

Step 56. Start a new report

* Open Power BI Desktop.
* Step 57. Click file > new (or simply start from the launch screen).

Step 58. Get data from excel

Step 59. Click home > get data > excel workbook

Step 60. Connect to the file

* Browse to your Excel file (C:\Power BI Data\Census\Louisiana DOH Regions.xlsx).
* Step 61. Click open.

Step 62. Explore available data

Step 63. View sheets and suggested tables

* In the Navigator window:
  + You’ll see a list of worksheets from the Excel file.
  + You may also see **Suggested Tables** at the bottom.
    - Power BI tries to detect where the table of data starts and ignores decorative or header rows.
* Step 64. Click on different items to preview the data.

Step 65. Choose the suggested table

* Step 66. Select the suggested table that looks like your parish data.
* Notice how it:
  + Excludes top rows with extra info (like maps or notes).
  + Starts directly with the data you want.
* Step 67. Click transform data instead of load to review and clean the data.

Step 68. Examine the code using ai

At this point I want you to examine the code with AI. Be very careful doing this with your own queries. Don’t use AI to examine the code unless you are sure that the information is not being used to train other models. You could find the structure of you super secret database table shared for everyone.,

* Open the query using the Advanced Editor
* Open your favorite AI tool.
* Step 69. Paste the following prompt: “please explain the following m code (power query”.
* Then copy the query from the Advanced editor at the previous prompt
* If you are unable to access an AI agent, view the result from my chat in ChatGPT Demo.docx

Step 70. Clean and transform data

Step 71. Review the data preview

* Look at your columns:
  + Region
  + Region Name
  + Location ID
  + Abbreviation
  + Longitude / Latitude
  + Parish Name

Step 72. Check and change data types

* Right-click **Location ID** column.
* Step 73. Click change type > text.

🔹 **Important:**

* When prompted:
  + Choose **Replace Current**.
* Why? To avoid converting the ID to a number and back to text, which could drop leading zeros.

A quick side note. I have decided to keep the Location ID as a string. The location id is a string that combines the state FIPS code (the first two characters) with the County/Parish FIPS code. The normal representation if to use string, with states like Alabama having 01 as its state FIPS code. One could argue that a number would perform better.

Whichever data type you choose, be consistent. The location id should text wherever it occurs or number wherever.

Step 74. Fill down region information

Because the Excel sheet uses merged cells, some rows are blank under Region and Region Name.

* Step 75. Select the region and region name columns (one at a time or both together).
* Right-click one of the selected columns.
* Choose **Fill > Down**.

This fills in the missing region values into all rows.

Step 76. Move the query to the geography group

* Step 77. Rename the query “louisiana hospital administrative regions”
* Right click the query name in the query list
* Step 78. Click move to group
* Step 79. Click geography if it is in the list or create a new group called geography

Step 80. Replace file name with parameter

* Step 81. Click on source applied step
* Copy the path to the Excel file without the double quotes (C:\Power BI Data\Census\Louisiana DOH Regions.xlsx)
* Step 82. Create a new parameter
  + Manage Parameters in Home tab
  + New Parameter
  + Step 83. Paste the path into current value
  + Step 84. Change name to “louisianadohregionspath”
  + Step 85. Change type to text
  + Move the parameter query you just created to the Parameters folder
* Edit Source step for the Louisiana Hospital Administrative Regions
  + Step 86. Replace the path to the excel file with “louisianadohregionspath”

Step 87. Review applied steps

* In the Query Settings pane, review the **Applied Steps** list.
* Notice:
  + Navigation
  + Promoted Headers
  + Removed Top Rows
  + Filtered Rows
  + Step 88. Changed type
  + Filled Down

These show how Power BI transformed the data.

Step 89. Close and apply

* Step 90. Click close & apply in the home tab.
* Power BI will load your cleaned data into the data model.

Step 91. Confirm your data

* In Data view, confirm:
  + All rows have region names.
  + Leading zeros remain in Location IDs.
  + The dataset looks ready for reporting.

## CDC Places

The CDC Places dataset ([PLACES Data Portal | PLACES | CDC: https://www.cdc.gov/places/tools/data-portal.html](https://www.cdc.gov/places/tools/data-portal.html)) provides some interesting information about health conditions in different geographical areas. There are several datasets. For this exercise, we will be using the PLACES: Local Data for Better Health, County Data 2024 release dataset (<https://data.cdc.gov/500-Cities-Places/PLACES-Local-Data-for-Better-Health-County-Data-20/swc5-untb/about_data>) as well as some other releases.

We are going to load several files as a part of this load. The first is a file (PLACES\_and\_500\_Cities\_\_Data\_Dictionary\_20250701.csv) that describes the measures used in the dataset. The second group of files contain the actual data (PLACES\_\_Local\_Data\_for\_Better\_Health\_\_County\_Data\*.csv).

This will involve several steps including getting a list of all of the files in the folder, identify the file or files needed for each step, parsing and in the case of groups of files, joining the output.

### Get list of files from folder

#### Load folder

1. Get Data
2. Folder
3. Browse to “C:\Power BI Data\CDC Places”
4. Step 92. Note options:
   1. Combine will create one dataset from all of the files
   2. Load will Load the data
   3. Transform Data will bring over first 1000 file names for you to work on
5. Step 93. Click transform data

#### Organize

1. Put the new query (CDC Places) into Data Sources group
2. Step 94. Create a new text parameter (cdcplacesfolder) and set the value to c:\power bi data\cdc places. move new parameter to the parameters group
3. In CDC Places query, replace the folder path ("C:\Power BI Data\CDC Places") with the name of the new parameter (CDCPlacesFolder)
4. Uncheck the Enable Load property of the CDC Places query

#### Step 95. Review

The output is a list of all of the files in the folder. The list is recursive, so if you have multiple subfolders, they will show as well. You can see many of the attributes of the file.

The Content column contains the actual data that is in the files.

### Measure file

#### Get the file

1. Step 96. Create a reference to the cdc places query
   1. Right click CDC Place query
   2. Step 97. Click reference
2. Step 98. Rename to “cdc places measures”
3. Move to group “CDC Places.” Create if not present

#### Parse the file

Step 99. Click the “binary” link in the content column for the row with the file name “places\_and\_500\_cities\_\_data\_dictionary\_20250701.csv”

Step 100. Note the steps created. the process gets the contents from the row. then it converts it to a table by parsing it as a csv file. finally, it changes the data types.

#### Correct the column names

One problem with the output. It doesn’t get the column names from the first row

1. Step 101. Click x to the left of the changed type step of the query. this deletes that step
2. Step 102. Click “use first row in headers” (home tab, toward right side of menu)
3. Step 103. Note that the process created two steps: the first to make the first row the column headers; the second to change types
4. Step 104. Review the data types
   1. Step 105. Note, we could do additional data cleansing: change x to null and change type on the “places release” columns to whole number
5. Step 106. Review whether measureid is unique
   1. You can select in the View tab “Column Distribution” to display uniqueness  
      A screenshot of a computer

      AI-generated content may be incorrect.

### Data Files

To get the data files, we need to filter down to the files that we want and exclude the files that don’t work.

#### Get files

1. Step 107. Create new reference to cdc places
2. Step 108. Rename to cdc places ds
3. Filter to just the files we need
   1. Step 109. Click the down arrow on the extension column and select “.csv”
   2. Right click any of the “PLACES\_\_Local\_Data\_for\_Better\_Health\_\_County\_Data\_” files
   3. Step 110. Click text filters -> begins with
4. Edit the text in the formula bar. Remove all of the text in the file name from the year on. What should remain is “PLACES\_\_Local\_Data\_for\_Better\_Health\_\_County\_Data\_”
5. We are going to merge the two filters
   1. In the Filtered Rows 1 step, copy the formula starting from “Text.StartsWith” until the second to last parenthesis  
      Text.StartsWith([Name], "PLACES\_\_Local\_Data\_for\_Better\_Health\_\_County\_Data\_")
   2. Edit the Filtered Rows step. Add after “[Extension] = ".csv")” the word “and” and then paste the Text.StartsWith text
   3. The Filtered Rows and the Filtered Rows 1 steps should look identical
   4. Delete the Filtered Rows 1 step
6. This is going to be the source for the rest of the queries
   1. Leave in the Data Sources group
   2. Uncheck the Enable Load

#### Parse files

This step will create a dataset out of the files. What this involves is parsing the Content column from each of the rows into tables and then expanding the tables. We don’t click the Binary link in the Content column because that will only expand one row. We click the expand button on the Content column (two down arrows pointing to a line).

1. Step 111. Click the expand button
2. Step 112. Select the 2024 release as the sample file
3. Step 113. Click ok

#### Step 114. Review the output

The important steps in this are the Invoke Custom Function 1 and the Expanded Table Column 1.

Here is a short version of what this query does. See ChatGPT description in “PowerBI\_M\_Query\_Explanation\_CDC\_Places.docx”

Invoke Custom Function1 creates a new Table type column which is build by apply the “Transform File” function to the Content column. Power Query created the function based on the 2024 release file that we selected when expanding the column.

The Expanded Table Column 1, simply expands each of the tables into rows in the table. It also gets the column names that the function found for the sample file.

#### Filter the query to those in ReportStates

1. Edit the CDC Places DS query
2. Step 115. Click on down arrow on the location name column header
3. Unlselect “(blank)”
4. Edit the Filtered Rows 1 stepmin the formula bar
   1. Add “and List.Contains(Text.Split(ReportStates,","), [StateAbbr])” after “[LocationName] <> ""”
5. Step 116. Change types by detect
   1. Step 117. Select the whole table (ctl-a)
   2. In Transform tab, click Detect Datatype
6. Fix datatypes
   1. Step 118. Change datatype on totalpop18plus to whole number
   2. Step 119. Change datatype on locationid to text

## County file

We need to fill the LocationID for all of the records. The sample file should have all of the county/parish names with the locationid.

### Build basic query

* 1. Step 120. Create reference to the cdc places ds
  2. Step 121. Rename to cdc places county
  3. Move to Geography group
  4. Filter LocationName column, removing blanks
  5. Edit the formula adding “and [LocationID]<> null”
  6. Keep only geography related columns (Select and “Remove other columns”)
     + StateAbbr
     + StateDesc
     + LocationName
     + LocationID
     + Geolocation
  7. Remove duplicates. Select LocationID column and choose remove duplicates

## Fact table

Finally, we pull it all together to create a Fact table where only the references to the dimensions are present and the numbers.

1. Step 122. Create a reference to the cdc places ds
2. Step 123. Rename to cdc places fact
3. Move to the CDC Places group

### Get the location id

Some of the older records don’t have location ids. We will get them by joining the CDC Places Fact to the CDC Places County

1. Merge CDC Places Fact to CDC Places Group.
   1. In Home tab, click on Merge Queries
   2. Step 124. Select cdc places county in the drop down menu
   3. Verify that the Join Kind is left outer
   4. Step 125. Select stateabbr and locationnmae in both sets
   5. Step 126. Click ok
2. Expand CDC Places County
   1. Step 127. Click expansion button on the column header
   2. Step 128. Select locationid
   3. Make sure that “Use original column name as prefix” is checked
3. Step 129. Replace locatinid with the cdc places count.locationid, if it is null
   1. Step 130. Rename locationid to locationidsource
   2. Add a Conditional column.
   3. Step 131. Name it locationid
   4. Step 132. Select column name locationidsource equals null then cdc places count.locationid  
      else locationidsource
   5. Edit formula adding: “, type text” right before final closing parenthesis

### Remove columns that are covered by ids

Remove the following columns

* StateAbbr
* StateDesc
* LocationName
* Category
* Measure
* Data\_Value\_Footnote\_Symbol
* Data\_Value\_Footnote
* CategoryID
* LocationIDSource
* Geolocation
* CDCPlacesCounty.LocationID

### Convert values to proper percentages

Percentages that are displayed represent about 100 times the value that they have. If you have 33% of something, that represents .33 or that item. I prefer to store the percents as their decimal equivalent. The data values in this set are what is normally displayed for percents, not the value

* 1. Step 133. Click on data\_value colum
  2. On the Transform tab (the same button on Add Column adds a new column)
     + Step 134. Click standard -> divide
     + Enter 100
     + Step 135. Replace type number with percentage.type
  3. Do the same for the following two columns
     + High\_Confidence\_Limit
     + Low\_Confidence\_Limit

Step 136. Note that in the sample i put together, i used chatgpt to merge the three division statements into one statement.

## County file additional changes

The CDC Places county table can be joined to the Louisiana Hospital Administrative Regions table to provide more context. In addition the GeoLocation can be parsed into Latitude and Longitude.

### Enrich with data from Lousian Hospital Administrative regions

1. Step 137. Click merge queries
2. Step 138. Select “louisiana hospital administrative regions” in the drop down
3. Step 139. Select the locationid in both tables
4. Verify that the Left Outer is the join type
5. Only 64 of the 570 Rows match
6. Expand the column “Louisiana Hospital Administrative Regions”
   1. Step 140. Select only the region name column
   2. Uncheck the “Use original column name as prefix”
7. Handle the missing Region Name
   1. Right click Region Name
   2. Step 141. Replace values
   3. Step 142. Type “null” in the value to find (important: all lower case)
   4. Step 143. Type “n/a” in the replace with box

### Parse the GeoLocation

1. Step 144. Select the geolocation column
2. Right click and select Add Column from Examples
3. Step 145. Rename column longitude
4. Step 146. Select the first cell in the geolocation column. select the first number and paste into the longitude column
5. Repeat for Latitude
6. Step 147. Select longitude and latitude columns. change type to decimal number

## Step 148. Create year table

Step 149. Create new query (blank query)

Go to Advanced Editor. Paste the following M code and click OK

let

StartYear = 2010,

EndYear = Date.Year(DateTime.LocalNow()),

YearList = List.Numbers(StartYear, EndYear - StartYear + 1),

YearTable = Table.FromList(YearList, each [Year = \_], {"Year"})

in

YearTable

Where does the error start?

Look for documentation for Table.FromList. What might be wrong?

Step 150. Replace the formula in “converted to table” with the following:

= Table.FromList(YearList, Splitter.SplitByNothing(), type table [Year=Int64.Type])

Move the Year query to a Date group

## CDC Places Fact – Copy population totals

The total population for each row seems to be dependent on the location and file that it came from. First, we are going to parse out the release year and then we will copy the population information to a new table.

In addition a number of measures change their values. So, some of the measures are duplicated for a year. For example, in the release year file for 2022, thee are COLON\_SCREEN records for the year 2020. In the release year 2023, there are also COLON\_SCREEN records for 2020. I assume that the release 2023 records represent a correction to the release year 2022 values (an assumption worth checking). We will correct this so that there are no duplicates.

1. Extract the release year from the Source.Name
   1. In the CDC Places Fact table, select the Source.Name column
   2. Step 151. Click “add column from examples”
   3. Step 152. Rename column1 to release year
   4. Step 153. Type the year from the source.name
   5. Check the function. It should look like it is getting the text between two delimiters (both “\_”)
   6. Step 154. Click ok
   7. Step 155. Change type for column to whole number
   8. Remove the Source.Name column
2. Remove the duplicate Release year values
   1. Step 156. Select the year column header and click on sort descending
   2. Step 157. Select the release year column header and click on descending
   3. Notice that the second sort is added to the existing Sort
   4. Next, we are going to remove all of the duplicates. But we need to first load all of the data into memory.
   5. At the front of the Table.Sort formula in the formula bar put your cursor one space before the words “Table.Sort.” Otherwise, the next step will delete the Table.Sort function.
   6. Step 158. Type table.buffer and type open parenthesis. delete the close parenthesis that power query gives you
   7. Put a close parenthesis at the end of the line
   8. Hit enter and take a break
   9. Step 159. Select columns
      1. Year
      2. Data\_value\_type
      3. MeasureID
      4. DataValueTypeID
      5. LocationID
   10. Remove duplicates
3. Step 160. Create new table from cdc places fact
   1. Reference CDC Places Fact
   2. Step 161. Rename “cdc places population”
   3. Step 162. Select columns
      1. LocationID
      2. Release Year
      3. Total Population
      4. TotalPop18Plus
   4. Remove Other Columns
   5. Step 163. Select table (ctl a)
   6. Remove duplicates

# Add Texas Regions

The public health regions for Texas are on the following page:

<https://www.dshs.texas.gov/center-health-statistics/texas-county-numbers-public-health-regions>

The counties of Texas are assigned to Public Health Regions and Health Service Regions. We will treat the Public Health Regions to be the equivalent of the Region Name that we assigned from the Louisiana spreadsheet.

We are going to merge this table from the web with the one from the spreadsheet.

1. Step 164. Create a parameter called texasregionsweburl and set the value to
2. New Source, Web
3. Step 165. Change url source from parameter and select texasregionsweburl.
4. Step 166. Note the html table at the top
5. Step 167. Select the html table (texas counties, public health regions, and administrative regions)
6. Step 168. Rename to texas public health regions
7. Move to Geography, uncheck Enable Load
8. Clean up to join to Louisianan Hospital Administrative Regions
9. First clean up the Louisiana table. There are a number of columns that we don’t use. Keep the following columns:
   1. Region
   2. Region Name
   3. LocationID
10. In the Texas Public Health Regions cleanup the data type
    1. FIPS Code -> change type to text (replace current)
11. Keep only the FIPSCode and Public HealthRegion columns
12. Step 169. Rename columns
    1. FIPSCode -> LocationID
    2. Public HealthRegion -> Region
13. Step 170. Create region name
    1. Step 171. Select region column and add column from examples
    2. Step 172. Name column region name
    3. Step 173. Type in the new column “tx region “ followed by the region number
    4. The result should be something like combining.
    5. What does Text.Form do?
    6. Step 174. Click ok
14. Join the Louisiana and Texas queries together
    1. Edit the Louisiana Hospital Administrative Regions
    2. Step 175. Click append queries, leave two tables selected
    3. Step 176. Select the texas public health regions table
    4. You will get a warning. Take it seriously  
       A screenshot of a computer error

       AI-generated content may be incorrect.
    5. Set the value to Public.
    6. The meaning of this is that you have two queries running that come from different places. And there is potential that you could send sensitive data out somewhere you don’t want it to go.
    7. Step 177. Rename the louisiana hospital administrative regions to hospital administrative regions

# Power BI Exercise: Data Modeling and Cleanup

## Step 1. Goal

Clean up and prepare the Power BI data model to support effective analysis by:

* - Managing relationships
* - Hiding or removing unnecessary tables and columns
* - Building hierarchies
* - Marking data categories
* - Formatting columns and creating useful metadata

## Step 2. Refresh data

Open the completed model from previous exercises and perform a Refresh to ensure all data is current.

## Step 3. Explore the data

Switch to Table View and scan the available tables and values. Look at each table. You can filter the columns

Then switch to Model View to understand how the tables are related. Examine different relationships. Arrange the tables so that the relationships between the tables is clear.

## Step 4. Remove or hide unused tables

We are going to report on these tables:

* - CDC Places Fact
* - CDC Places Measures
* - CDC Places Population
* - CDC Places County
* - Market\_Saturation\_Facts
* - Market Saturation Measures
* - Year table

We will hide the unused tables Hospital Administrative Regions and Market\_Saturation\_Counties. But before we do let’s examine all of the relationships to/from these tables.

Hospital Administrative Regions has relationships to 2 tables: Market\_Saturation\_Counties; CDC Places County

Market\_Saturation\_Counties has relationships with 3 tables: Hospital Administrative Regions, Market\_Saturation\_Facts; CDC Places County

Delete these relationships (4 in all).

Hide the Hospital Administrative Regions and Market\_Saturation\_Counties tables.

Note that if you delete the tables, the queries will normally be deleted. However, since the Hospital Administrative Regions table is used in other queries it will remain

## Step 5. Create clean model layouts

Use Model View to create layouts named Market Saturation, CDC Places, and CDC Population.

Add the Market\_Saturation\_Facts to the Market Saturation layout and add related tables and verify relationships.

Add the CDC Places Fact to the CDC Places layout and add related tables and verify relationships.

Add the CDC Places Population to the CDC Population layout and add related tables and verify relationships.

## Review all the relationships between table

Right click some of the lines between table and select “Properties.” Or select the line and look at the Properties tab on the right.

Note the properties.

One of the lines is dotted. That means that it is inactive. The reason it was inactive was that there was an active relationship to one of the tables we hid. You can make it active.

Create a second relationship between the Year table and the CDC Places Fact table, on the Release Year column. Note that it is inactive.

## Step 6. Add a date table

Disable Power BI's Auto date/time feature. This is in Options in PBI Desktop.

A screenshot of a computer

AI-generated content may be incorrect.

Create a calculated Date table using DAX. We will use a template from SQLBI. <https://www.sqlbi.com/tools/dax-date-template/> Download, then open up the report and copy the date table.

Open the DAX for the date table and set the FirstYear variable: VAR FirstYear = 2015

Mark it as a Date Table.

## Step 7. Define relationships with date table

Link Date[Date] to two columns in the Market\_Saturation\_Facts table.

* Period\_EndDate
* Period\_StartDate

Note that only one active relationship is active between these two tables.

## Step 8. Build hierarchies

* Examples:
* Geography Hierarchy: State Abbreviation > Region Name > Location Name
* Calendar Hierarchy: Calendar Year > Calendar Month > Day of Month
* Measure Hierarchy: Category Name > Measure Short Name

## Hide or put columns into display folders

In the table CDC Places Measures select all of the columns EXCEPT:

* Category Name
* Measure Full Name
* Measure Hierarchy
* Measure Short Name

Set the Display folder for all of those columns to “Metadata”

In the CDC Places Fact, we will hide the columns that we don’t need to see. Some of them are needed in the model but don’t need to be visible. Others we are hiding for now until we can determine if we need them.

Hide the following columns:

* Data\_Value\_Unit
* DataSource
* DataValueTypeID
* LocationID
* MeasureId
* Year

## Step 9. Column formatting and metadata

Format and categorize columns. Hide IDs and internal fields. Set proper summarization (Sum or Average).

We are going format the values in the CDC Places Fact table.

Select the Data\_Value column in the table. Change the name to CDC Places Value

Change the Format to Percentage. In Advanced properties, Set the Summarize By to Average.

Select both the High\_Confidence\_Limit and Low\_Confidence\_Limit and set the Format to Percentage and the Summarize By to Average.

Next select the TotalPopulation and TotalPop19plus and set the format to Whole Number and Thousands separator to “Yes”. Change Summarize By to None.

## Set the data category for columns

Select the following columns one by one in the CDC Places County table

* Latitude
* Longitude
* LocationName
* StateAbbr
* StateDesc

For each of them, select the Data Category (in advanced properities).

Also, change the Summarize by for Latitude and Longitude to None

## 11. Create Calculated Columns

Example: Create “Population 2024 Release”' as a new calculated column. Pm the CDC Places County table

The expression is

CALCULATE(MAXX(RELATEDTABLE('CDC Places Population'),'CDC Places Population'[TotalPopulation]),'CDC Places Population'[Release Year] =2024)

Format as Whole Number and set the comma for thousands separator.

This expression gets the records from the CDC Places Population table that is related to the current CDC Places County record. It then filters to the Release Year 2024.

## 12. Create Bins/Groups

Use New Group feature to bin population data into 5 bins. This breaks up the data into bins based on a number

Select the new Population 2024 Release column CDC Places County.

New Group, using binning. Set the bin count to 5, which will create 5 groups based on Population.

Select LocationName,, new group. Select East and West Baton Rouge, Call the group Baton Rouge and name the data group as a whole Baton Rouge. Check Include Other Group

## 13. Create Manual Groups

Manually group selected parishes like East and West Baton Rouge into one group.

## 13. Create Manual Groups

Manually group selected parishes like East and West Baton Rouge into one group.

Another way to create a group is to add a column.

Create a new column on the CDC Places County called Rural/Urban. Set the expression:

Rural/Urban = SWITCH(TRUE

    , 'CDC Places County'[Population 2024 Releaase] < 100, "Rural tiny"

    , 'CDC Places County'[Population 2024 Releaase] < 5000, "Rural"

    , 'CDC Places County'[Population 2024 Releaase] < 100000, "Suburban"

    , "Urban"

)

Next create a new column Rural/Urban Number with expression

Rural/Urban Number = SWITCH(TRUE

    , 'CDC Places County'[Population 2024 Releaase] < 100, 1

    , 'CDC Places County'[Population 2024 Releaase] < 5000, 10

    , 'CDC Places County'[Population 2024 Releaase] < 100000, 50

    , 100

)

Hide the Rural/Urban Number.

Select the Rural/Urban column. Under Advanced in propertis, set the sort by column to Rural/Urban Number. This will sort the Rural/Urban column in population order

## 13. Change the order by on Date Calendar Month

The Calendar Month is text and like text it will sort alphabetically. So far there no internal mechanism to order it in month order.

But you can change the Sort By property on the column.

Select the Calendar Month and change the Sort By column. In the Report view and Table view, the menu will change to show properties of the column. In the Model view, the Properties menu will show. You change Sort By in Advanced section. Select Calencar MonthNumber as the sort by column

## 13. Format dates in the Date table

I prefer a simple Short Date for the display of a date. But PBI prefers a long date format.

Select Date column. In Format drop down select Short date (near top of list)

Change the “StarOf” dates to shor date. In Data tab (right of screen), type start in the search bar. Check formatting on each column. They should all be dates. If so, select all of the columns displaying in the Date table. Change format to Short Date. Then do the same for other dates showing.

Do the same with EndOf as the search.

## 13. Clean up Market\_Saturation\_Facts

We will do many of the same things with the Market\_Saturation\_Facts

Hide: LocationID; Type\_Of\_Service\_Key

Select the percentage columns and change the Format to Percentage. Change the Summarize By to Average

Select the number\_of colums, set the Thousands separator to Yest

Select the average columns, set the Summarize by to Average

Select total\_payment and verify that the Format is currency

Exercise 3 Power BI Report Creation

This detailed guide walks through the process of building reports in Power BI. It includes precise actions, user interface guidance, and suggested spots for screenshots to enhance clarity.

# 1. Launching Power BI and Opening the Report View

* Open Power BI Desktop.
* Open the following reports:
* Exercise 3 Reports.Start.pbix
* Exercise 3 Reports.Final.pbix

The first file is the one that you will be editing. The second file represents what I intended the results to be.

# 2. Adding a Basic Visualization

* From the Data pane, drag the field 'StateAbbr' onto the canvas.
* Next, drag a the CDC Places Value from CDC Places Fact to the same
* Power BI may attempt to use a map visualization automatically based on geographic fields.
* If an error appears related to mapping, ignore it for now—we'll switch visual types shortly.

# 3. Exploring and Changing Visual Types

* With the visual selected, go to the Visualizations pane and click on different visual icons:
* - Try Bar chart (Stacked/Clustered), Pie chart, Funnel, etc.
* - Notice how the layout and interpretation of the data changes.
* Avoid visuals that seem nonsensical for your data (e.g., area chart for categorical state data).
* The Narrative visual gave this message
* A screenshot of a computer

  AI-generated content may be incorrect.If you get this message, don’t click on the button. Delete the visual and start over

# 4. Using Slicers for Interactive Filtering

* Remove the CDC Places Value column from the visual (alternatively Click a blank spot on the canvas to deselect visuals. Drag 'StateAbbr' again to the canvas)
* In the Visualizations pane, click the 'Slicer' icon (filter icon).
* Change slicer formatting:
* - Go to Format menu your Visual → Slicer Settings.  
  A screenshot of a computer

  AI-generated content may be incorrect.
* - Change style to 'Dropdown' and set 'Show Select All'
* - View the slicer
* - Try selecting multiple states using Ctrl-click.
* - Click the 3 dot ellipsis to allow for Search  
  A screenshot of a computer

  AI-generated content may be incorrect.

# 5. Adding a Year Slicer

* Drag the 'Year' field onto the canvas.
* Convert the visual to a Slicer.
* By default, Power BI may show it as a slider (range).
* Explore Slicer Settings → Options: change to Dropdown, Between, Less Than, etc.
* Choose Dropdown, Show Select all

# 6. Filtering Pages Using the Filters Pane

* The data only occurs from 2018 to 2022, but the Year slicer has more years
* Open the Filters pane on the right side.
* Drag the 'Year' field into the 'Filters on all pages' section.
* Set an advanced filter: Year is between 2018 and 2022.
* Click 'Apply filter' to activate the filter.

# 7. Creating a Table with Geographic Hierarchy

* Click a blank space to deselect existing visuals.
* Drag a 'Geography Hierarchy' field (e.g., State Abbreviation, Region Name, Location Name) to the canvas.
* Convert it to a Table visual.
* Format columns for visibility (widen if necessary).

# 8. Adding Measures to the Table

* Drag values such as 'Low Confidence,' ‘CDC Places Value, ‘and 'High Confidence' into the table.
* You can rename columns by clicking their headers in the Visualizations pane.
* Align and resize visuals for consistency.

# 9. Converting the Table to a Matrix

* Create a new page
* Copy the two slicer to that page and allow Sync
* Copy the table from the first page and copy to the second page
* Go to the second page (I forget this regularly)
* Select the table and convert it to a Matrix visual.
* Drag 'Year' to the Columns section of the Matrix.

# 10. Experiment with drill up and down

* Use the '+' icon in the visual to drill down through hierarchy levels (e.g., State → Region → County).
* Also, you can drill down and up using the arrows in the upper right corner  
    
  The two arrows pointing down will drill down on the whole table and remove the StateAbbr  
  The trident pointing down will drill down all the way, leaving the StateAbbr.  
  The Up arrow will drill up  
  Try it!
* The display is stairstep. If you want to have tabular
* Go to Format → Layout and Style Presets and select 'Tabular' in the drop down.
* Enable 'Repeat Row Headers' to repeat higher-level rows.

# 11. Using Column Charts (Line and Clustered Column)

* Duplicate the page with the matrix
* Select the matrix
* Select the 'Line and Clustered Column Chart' visual type.
* In the X-Axis, delete the LocationName
* Mover the Low and High Confidence values to the Line y-axis
* The CDC Places Value will remain in the Column y-axis
* Move the Year column to the Column legend
* Assign one value to the 'Column Y-Axis' and the other to the 'Line Y-Axis'.

# 12. Creating a Measure Slicer

* Drag the 'Measure Hierarchy' field to the canvas.
* Convert it to a slicer and format it as a dropdown.
* Enable single-selection mode to allow focus on one metric at a time.

# 13. Using the scatter chart, set up the data

A scatter chart is one way to find correlations in data. We will see if there is a correlation between the percentage of people in a county getting regular checkups and high blood pressure. We need to create two measures:

Annual Checkup Average = CALCULATE(AVERAGE('CDC Places Fact'[CDC Places Value]), 'CDC Places Measures'[Measure short name] = "Annual Checkup")

High Blood Pressure Average = CALCULATE(AVERAGE('CDC Places Fact'[CDC Places Value]), 'CDC Places Measures'[Measure short name] = "High Blood Pressure")

* The easiest way to handle this is to right click on the CDC Places Fact table, New Measure, then paste the each of the two paragraphs in as separate measures. Hit green check
* Set the formatting for both new measures to percentage
* Next create a new page. Copy and paste the two slicers (StateAbbr and Year) at the top
* Drag the Geography hierarchy. Change the visual to a Matrix.
* Remove the LocationName column from the hierarchy in the Rows
* Drag Year into columns
* Then drag the two measures that we just created into the Values
* Change the Format of the Matrix to Tabular Layout and Repeat row headers

# 14. Create the scatter chart

* Copy the matrix and past it below
* Select the second matrix and click on Scatter Chart. There will be an error that we will correct
* Drag the High Blood pressure variable from X Axis to the Y Axis.
* Drag the Year to the Play Axis
* Click in the space above each year to see what the values are for that year
* Delete Region Name from the Values
* Then drag Region Name under Geography

# 14. Test the interaction of the visuals

* Click above 2021 in the Year play bar so that the scatter chart is showing 2021
* Click 2019 in the Matrix at the top. Notice that the Scatter chart changes

# 15. Displaying the CDC Measures in a tree map

A tree map is intended to give an overview of how much is in a category. The boxes to the left and toward the top are bigger than the ones to the left and on the bottom. It organizes hierarchically.

* Create a new page and copy the StateAbbr and Year slicers from another page
* Drage the Measure Hierarchy from CDC Places Measure.
* Drag the CDC Places Value into the same visual
* Change to Treemap
* Drill down and up to see different views.

# 16. Comparing States Side-by-Side

* Duplicate a report page and remove the slicer for 'State Abbreviation'.
* Add two Tree Maps side by side.
* Use Filters on each visual to limit one to LA and one to TX.
* Rename visuals and adjust titles accordingly.

# 17. Create drill through report

* Create a new page
* Drag Measure Hierarchy to canvas and change to Matrix
* Drag Year to the Columns
* Drag CDC Places Value, High Confidence, Low Confidence to values
* Select Page
* Drag StateAbbr and LocationName to Drill through in the visualizations tab
* Notice the arrow in the top
* Uncheck Keep all filters
* Add two Card visuals, setting one to StateAbbr and the other LocationName
* Go to a page with the StateAbbr and LocationName in a matrix.
* Right click row and click drill through to the new page you just created