

**Akademia Górniczo-Hutnicza w Krakowie**  
**Wydział Informatyki, Elektroniki i Telekomunikacji**



# **Hurtownie danych**

Konspekt zajęć laboratoryjnych  
prowadzonych w Katedrze Informatyki  
Studia Podyplomowe

*Marek Macura*

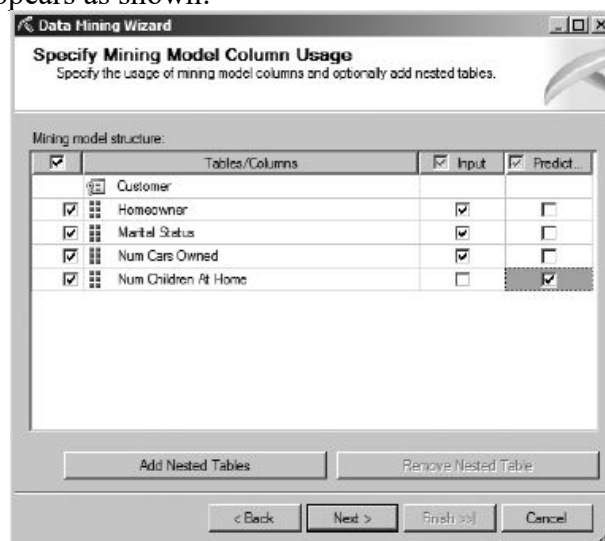
## I. Defining a Data Mining Model

**Business Need** The Maximum Miniatures Marketing Department wants to do a targeted mailing promoting the Mythic World line of figurines. Previous research has shown that the vast majority of buyers of the Mythic World product line do not have any children living at home. Unfortunately, the list of addresses and demographics purchased for the Mythic World mailing does not include a statistic on the number of children living at home. It does include the following facts about each household: Number of cars owned, Marital status and Whether the address is a rental property or occupied by the owner.

The marketing department would like to find a way, using the three facts included in the mailing list data and the information known about current customers, to predict which households have no children living at home. The mailing will then be sent only to those households likely to have no children living at home.

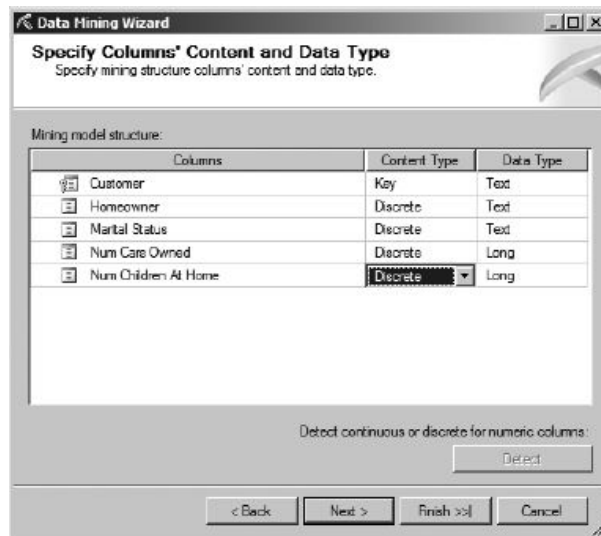
### Steps to Create a Mining Structure with One Mining Model

1. Right-click the Mining Structures folder in the Solution Explorer window for local **MaxMinSalesDM** cube and select New Mining Structure from the context menu. The Data Mining Wizard appears.
2. Click Next on the Welcome page. The Select the Definition Method page of the wizard appears.
3. Select the From existing cube radio button. Click Next.
4. Select Microsoft Decision Trees from the drop-down list. Click Next.
5. Select the Customer dimension. This is the main source of information for this mining operation. Click Next.
6. Leave the Customer attribute selected. This attribute serves as the primary key for the mining operation. Click Next. The Select Case Level Columns page of the wizard appears.
7. The columns selected here serve as the input and predictable columns in our model. Check the following attributes: Homeowner, Marital Status, Num Cars Owned and Num Children At Home.
8. Click Next. The Specify Mining Model Column Usage page of the wizard appears.
9. Uncheck the check box in the Input column for Num Children At Home. Check the check box in the Predictable column for Num Children At Home. The Specify Mining Model Column Usage page appears as shown:



10. Click Next. The Specify Columns' Content and Data Type page of the wizard appears.

11. Change the Content Type for Num Cars Owned to Discrete. Change the Content Type for Num Children At Home to Discrete. The Specify Columns' Content and Data Type page appears as shown:



12. Click Next. The Slice Source Cube page of the wizard appears.

13. We need to use only retail customers for our data mining (Account Num attribute of 5000 and above). Select Customer from the drop-down list in the first row of the Dimension column.

14. Select Account Num from the drop-down list in the first row of the Hierarchy column.

15. Select Range (Inclusive) from the drop-down list in the first row of the Operator column.

16. Click in the first row of the Filter Expression column. Two drop-down lists appear.

Activate the drop-down list on the left.

17. Expand the All entry, select 5000 from the list, and click OK.

18. Activate the drop-down list on the right. Scroll to the bottom of the list.

19. Click the Filter Members link. The Filter Members – Children of All dialog box appears.

20. Select Name in the drop-down list under Property.

21. Select Begins with in the drop-down list under Operator.

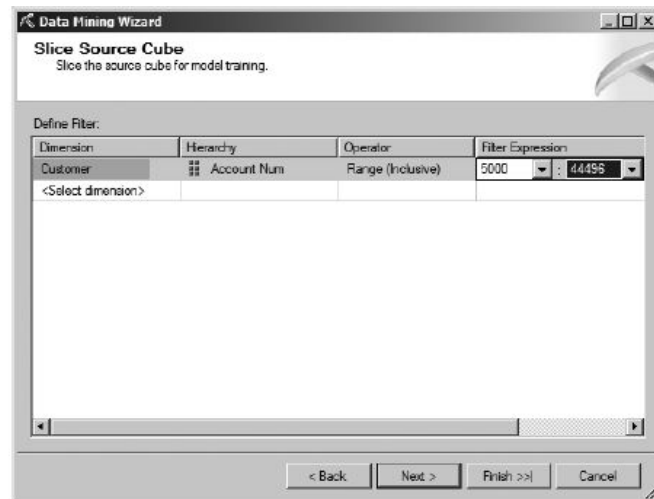
22. In the Value area, enter 444.

23. Click Test. The Filter Members dialog box appears, as shown:



24. Click OK to exit the Filter Members dialog box.

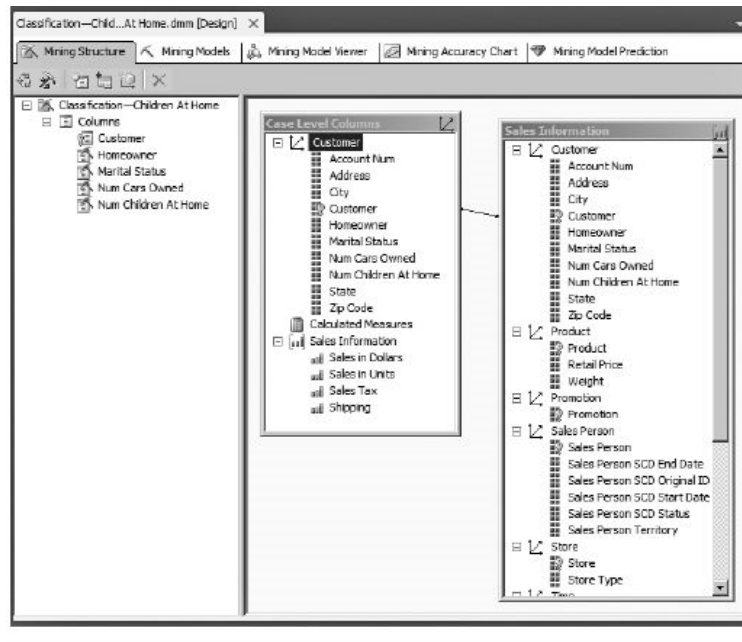
25. Select 44496 in the list and click OK. The Slice Source Cube page appears, as shown:



26. Click Next. The Create Testing Set page appears. We need to split our data into a training data set and a testing data set. We will split our data down the middle.
27. Change the Percentage of data from testing to 50%.
28. Click Next. The Completing the Wizard page appears.
29. Enter Classification—Children At Home for Mining structure name.
30. Enter Decision Trees—Children At Home for Mining model name. The Completing the Wizard page appears as shown:



31. Click Finish. The data mining structure containing our new data mining model is created, and the Data Mining Design tab appears as shown:



### Steps to Add Mining Models to the Mining Structure

1. Select the Mining Models tab on the Data Mining Design tab.
2. Click the Create a related mining model button on the Mining Models tab toolbar. The New Mining Model dialog box appears.
3. Enter Naive Bayes—Children At Home for Model name.
4. Select Microsoft Naive Bayes from the Algorithm name drop-down list.
5. Click OK.
6. Again, click the Create a related mining model button on the Mining Models tab toolbar. The New Mining Model dialog box appears.
7. Enter Clustering—Children At Home for Model name.
8. Select Microsoft Clustering from the Algorithm name drop-down list.
9. Click OK.
10. Once more, click the Create a related mining model button on the Mining Models tab toolbar. The New Mining Model dialog box appears.
11. Enter Neural Network—Children At Home for Model Name.
12. Select Microsoft Neural Network from the Algorithm name drop-down list.
13. Click OK.
14. Click the Save All button on the toolbar.
15. Click the Process the mining structure and all its related models button on the Mining Models tab toolbar. A dialog box informs you the server content appears to be out-of-date.
16. Click Yes to deploy the project. If prompted to overwrite the database, click Yes to continue.
17. When the deployment is complete, the Process Mining Structure dialog box appears. Click Run. The Process Progress dialog box appears.
18. When the models have all been trained and the process has been completed successfully, the Process Progress dialog box appears. Click Close to exit the Process Progress dialog box.
19. Click Close to exit the Process Mining Structure dialog box.

Now that we have several data mining models created and trained, let's look at how we begin to analyze the information. We do this by using the data mining model viewers, which are found on the **Mining Model Viewer** tab on the Data Mining Design tab.

Each data mining algorithm has its own set of viewers, enabling us to examine the trained data mining algorithm. These viewers present the mining information graphically for ease of understanding. If desired, we can also use the Microsoft Generic Content Tree Viewer to look at the raw data underlying the graphical presentations.

## II. Connecting to an Existing Analysis Services Model with Power BI

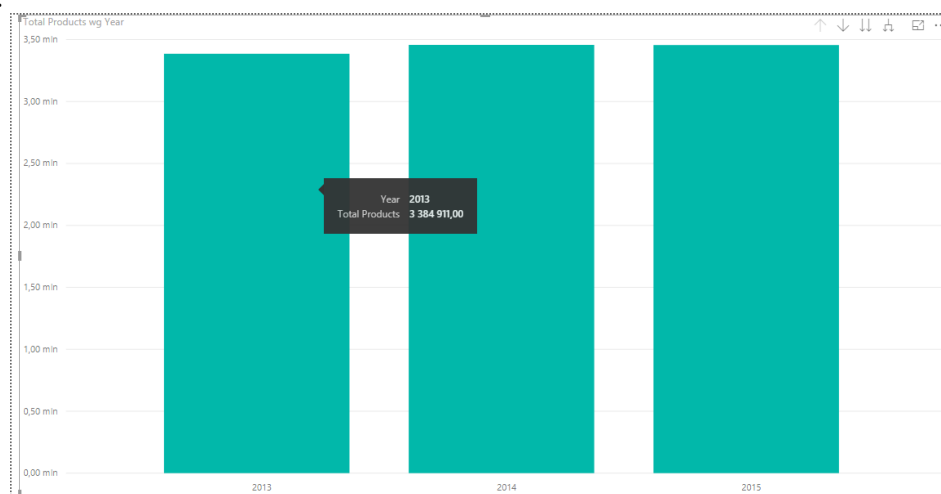
**Business Need** The VP of manufacturing for Max Min, Inc., would like to use Power BI for self-service data visualizations. He wants to use Power BI reports to explore the manufacturing data, but he does not want to have to create his own data models.

### Steps to Create Live Connection to the MaxMinManufacturing Tabular Model

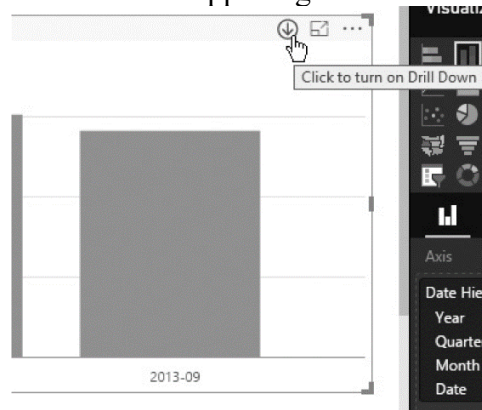
1. Click the File tab of the menu and select New from the menu. The Start dialog box appears.
2. Click Get Data. The Get Data dialog box appears.
3. Select SQL Server Analysis Services Database and click Connect. The SQL Server Analysis Services Database connection dialog box appears.
4. For Server, type (local) for the MaxMinManufacturing SSAS model.
5. For Database, type MaxMinManufacturingDM.
6. Select the Connect live option.
7. Click OK. The Navigator dialog box appears.
8. Any perspectives defined within the data model as well as an entry for the data model as a whole are displayed in the Navigator dialog box when operating in this mode. Select Max Min Manufacturing DM.
9. Click OK. You immediately see the Report page with the tables in the selected model visible in the Fields area.

### Steps to Explore the Data Using Drilldown

1. Expand the Measures table in the Fields area.
2. Click the check box next to Total Products in the Fields area. A column chart showing the total products produced across all of the data in the model is displayed.
3. Scroll down in the fields area and expand the Dim Time table.
4. Click the check box next to Date Hierarchy. The column chart shows the total produced for each of the three years in the data model.
5. Hover over the 2013 column. The detail information for this column appears in a tooltip, as shown:



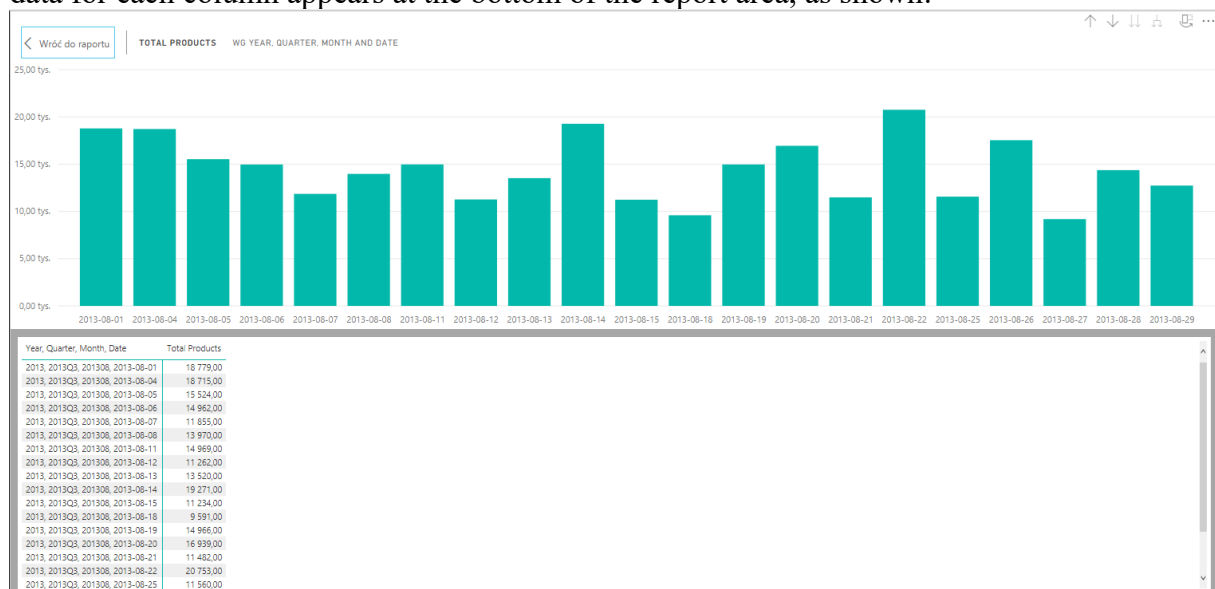
6. Because we chose to add the entire Date Hierarchy to the visualization, our column chart has drilldown enabled. Right-click the 2013 column and select Drill Down from the context menu. You see the Total Products for the four quarters of 2013.
7. Right-click the 2013-Q3 column and select Drill Down from the context menu. You see the Total Products for each month in the third quarter of 2013.
8. Click the 2013-08 column. The selected column retains its color while the color fades on the other two columns. This draws special attention to the selected column.
9. Click the 2013-08 column again to have all the columns return to the same color intensity.
10. We can change the chart so a click causes a drilldown action rather than a highlight action. Click the single down arrow in the upper-right corner of the chart, as shown:



The arrow changes from dark on light background to light on dark background. This indicates single-click drilldown has been activated.

11. Once again, click the 2013-08 column. This time you see the Total Products for each day in August 2013.

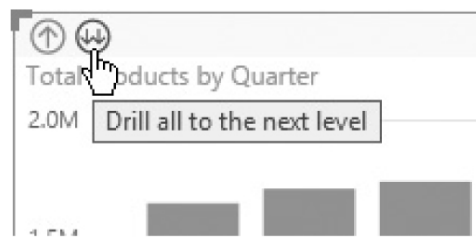
12. Right-click one of the columns and select See Data from the context menu. The detail data for each column appears at the bottom of the report area, as shown:



13. Click the button in the upper-right corner of the report area to pivot the windows. The chart and the detail data window will appear side by side.

14. Right-click one of the columns and select Drill Up from the context menu. You see the Total Products for each month in the third quarter of 2013.

15. Click the button to pivot the windows again and they appear top to bottom again.
16. Click Back to Report in the upper-left corner of the report area to remove the detail data window.
17. Click the Drill Up arrow in the upper-left corner of the chart. You see the Total Products for each quarter of 2013.
18. Click the Drill Up arrow again. You see the Total Products for each year.
19. In addition to drilling down to the next level for a single column, we can drill down to the next level for all columns at once. Right-click one of the columns and select Expand All from the context menu. You see the Total Products for all quarters of all years in the data model.
20. Click the Drill all to the next level double down arrow in the upper-left corner of the chart, as shown:



You see the Total Products for all months of all years in the data model.

21. Click Save in the upper-left corner of the Power BI window. The Save As dialog box appears.
22. Enter Max Min Manufacturing Live Connection for File name.
23. Click Save.

### III. Creating a Paginated Report with a Matrix

**Business Need** The Maximum Miniatures Marketing Department would like to analyze wholesale sales for each store over time. The marketing department wants to be able to select a year and view the quarters within the selected year. They would like to be able to drill down to the month level within a quarter. They would also like to be able to drill down to specific products within a store. In addition to the sales dollar amount, they would like to see the sales as a percent of total sales for the quarter. We use the Max Min Sales DM model as the data source for this report.

#### Steps to Create a Report Server Project

1. Open SQL Server Data Tools.
2. Click the New Project button on the toolbar. The New Project dialog box appears.
3. In the Installed Templates area, select Reporting Services under Business Intelligence.
4. Select Report Server Project in the center of the dialog box.
5. Enter **MaxMinReports** for Name and set the Location to the appropriate folder. This will be the only project in this solution, so leave Create directory for solution unchecked.
6. Click OK to create the project.

#### Steps to Create a Shared Data Source

1. Right-click the Shared Data Sources folder in the Solution Explorer window and select Add New Data Source from the context menu. The Shared Data Source Properties dialog box appears.
2. Enter **MaxMinSales** for Name.
3. Select Microsoft SQL Server Analysis Services from the Type drop-down list.
4. Click Edit. The Connection Properties dialog box appears.



5. Enter the name of the Analysis Services server hosting the Max Min Sales model for Server name.
6. From the Select or enter a database name drop-down list, select MaxMinSalesDM. The Connection Properties dialog box appears.
7. Click OK to exit the Connection Properties dialog box.
8. Click OK to exit the Shared Data Source Properties dialog box. The shared data source is created.

### Steps to Create a New Report with a Dataset

1. Right-click the Reports folder in the Solution Explorer window and select Add | New Item. (Selecting Add New Report launches the Report Wizard, which we are not using for this report.) The Add New Item dialog box appears.
2. Select Report in the center area. Enter **Wholesale Sales By Month Report** for Name.
3. Click Add. The Report Design tab appears.
4. In the Report Data window, select New | Data Source from the toolbar. The Data Source Properties dialog box appears.
5. Enter **MaxMinSales** for Name.
6. Select the Use shared data source reference option.
7. Select the **MaxMinSales** shared data source from the drop-down list.
8. Click OK to exit the Data Source Properties dialog box.
9. Right-click the MaxMinSales data source you just created in the Report Data window and select Add Dataset from the context menu. The Dataset Properties dialog box appears.
10. Enter **WholesaleSales** for Name.
11. Click Query Designer. The MDX Query Designer window appears.
12. Enter the following MDX query:  

```
SELECT NON EMPTY { [Measures].[Sales in Dollars] } ON COLUMNS, NON EMPTY { ([Time].[Year - Quarter - Month - Date].[Month].ALLMEMBERS * [Store].[Store].[Store].ALLMEMBERS * [Product].[Product].[Product].ALLMEMBERS ) } DIMENSION PROPERTIES MEMBER_CAPTION, MEMBER_UNIQUE_NAME ON ROWS FROM ( SELECT ( STRTOSET(@TimeDateHierarchy, CONSTRAINED) ) ON COLUMNS FROM ( SELECT ( [Customer].[Account Num].&[4500] : [Customer].[Account Num].&[4996] ) ON COLUMNS FROM [MaxMinSales])) CELL PROPERTIES VALUE, BACK_COLOR, FORE_COLOR, FORMATTED_VALUE, FORMAT_STRING, FONT_NAME, FONT_SIZE, FONT_FLAGS
```
13. Click Query Parameters icon (6th from the left on top) and fill table as shown:

Parameter	Dimension	Hierarchy	Multiple values	Default
TimeDateHierarchy	Time	Year - Quarter - Month - Date	<input type="checkbox"/>	Calendar 2015
<Enter Parameter>				

Klik OK and then klik ! icon to execute query and then OK to close designer and next OK to close Dataset window. The results should look like:

Year	Quarter	Month	Store	Product	Sales in Dollars
Cale...	Quarter...	Januar...	Online	Americ...	51265,32
Cale...	Quarter...	Januar...	Online	Beer al...	55641,9
Cale...	Quarter...	Januar...	Online	Beer al...	60278,4
Cale...	Quarter...	Januar...	Online	Scotlon...	28019
Cale...	Quarter...	Januar...	Online	British ...	58612,4
Cale...	Quarter...	Januar...	Online	British ...	57596
Cale...	Quarter...	Januar...	Online	British L...	39522,56
Cale...	Quarter...	Januar...	Online	British ...	45664,08
Cale...	Quarter...	Januar...	Online	Bull Mo...	76885,2
Cale...	Quarter...	Januar...	Online	Cape H...	30693,6
Cale...	Quarter...	Januar...	Online	Chicag...	19530
Cale...	Quarter...	Januar...	Online	Dragon ...	52855,48
Cale...	Quarter...	Januar...	Online	Disumme...	33079,04
Cale...	Quarter...	Januar...	Online	Eagle Pa...	78132,6
Cale...	Quarter...	Januar...	Online	Eagle P...	40147,2
Cale...	Quarter...	Januar...	Online	Eddisto...	27606,6
Cale...	Quarter...	Januar...	Online	Elvin Ac...	26666,4
Cale...	Quarter...	Januar...	Online	Elvin De...	130754,4
Cale...	Quarter...	Januar...	Online	Female ...	108469,2
Cale...	Quarter...	Januar...	Online	Flying ...	41569,08
Cale...	Quarter...	Januar...	Online	French ...	50881,6
Cale...	Quarter...	Januar...	Online	French ...	41995,8

14. Expand the Parameters folder in the Report Data window.
15. The TimeDateHierarchy parameter was created. Double-click the entry for the TimeDateHierarchy parameter. The Report Parameter Properties dialog box appears.
16. Uncheck the Allow multiple values check box.
17. Fill available values section as shown:

Właściwości parametru raportu

Ogólne

Dostępne wartości

Wartości domyślne

Zaawansowane

Wybierz wartości, które będą dostępne dla tego parametru.

Wybierz jedną z następujących opcji:

☐ Brak

☐ Określ wartości

☒ Pobierz wartości z zapytania

Zestaw danych: (Ostrzeżenie: możliwy negatywny wpływ na wydajność)

TimeDateHierarchy

Pole wartości:

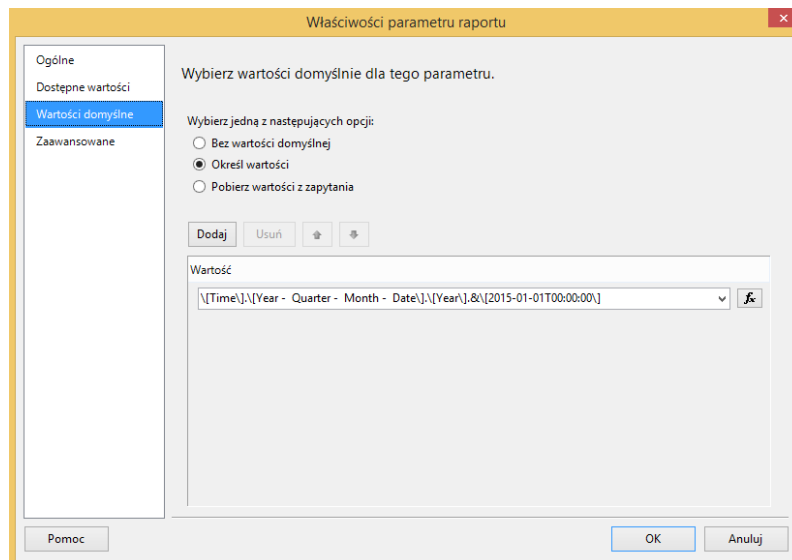
ParameterValue

Pole etykiety:

ParameterCaptionIndented

Pomoc OK Anuluj

18. Fill default values section:



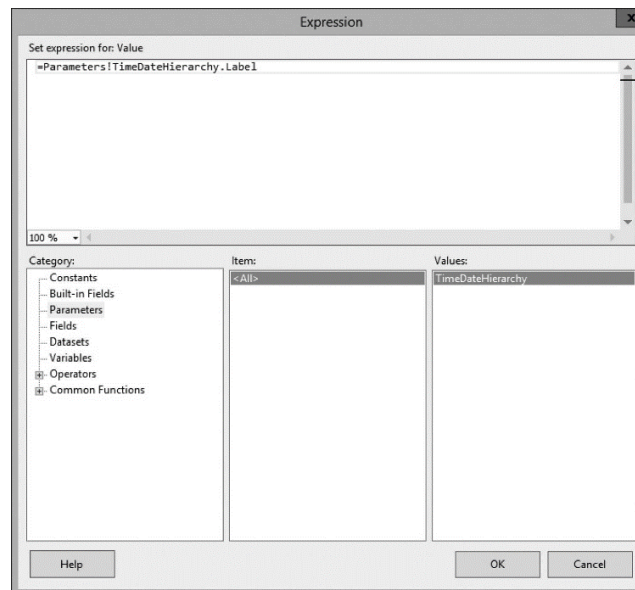
19. Click OK to exit the Report Parameter Properties dialog box.
20. Right click WholesaleSales dataset and fill parameters section as follows:  
parameter name: **TimeDateHierarchy** parameter value: **[@TimeDateHierarchy]** and click OK.

### Steps to Create the Report Layout

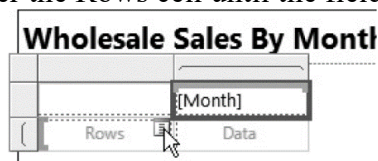
1. Click the report design area to activate the Report menu, and then select Report | Report Properties from the main menu. The Report Properties dialog box appears.
2. Select Landscape for Orientation.
3. Click OK.
4. Drag a text box from the Toolbox and drop it on the report body layout area.
5. Using the toolbar or the Properties window, modify the properties of the text box as follows:

Property	Value
Font: FontSize	16pt
Font: FontWeight	Bold
Location: Left	0in
Location: Top	0in
Size: Width	6in
Size: Height	0.375in

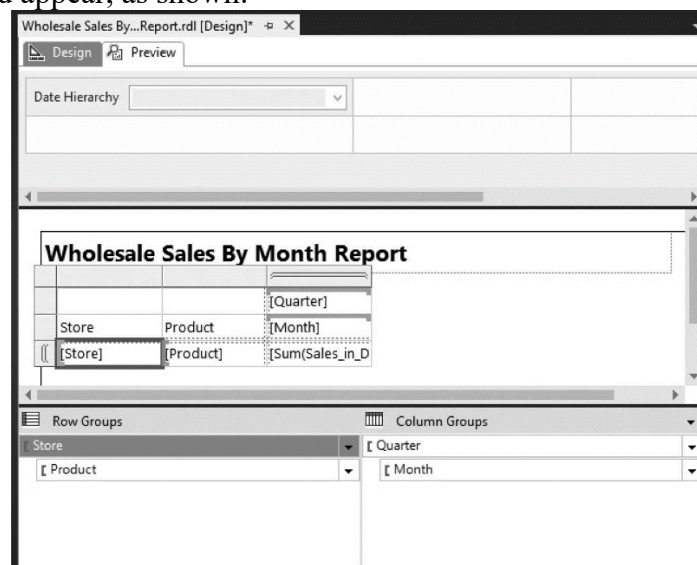
6. In the text box, enter **Wholesale Sales By Month Report - .**
7. With the text edit cursor still blinking at the end of the string you just entered, right-click the text edit cursor and select Create Placeholder from the context menu. The Placeholder Properties dialog box appears.
8. Click the *fx* button next to the Value drop-down list. The Expression dialog box appears.
9. Select Parameters in the Category pane.
10. Double-click the TimeDateHierarchy entry in the Values pane. This will create an expression representing this parameter.
11. In the Set expression for: Value area, replace “Value” with Label. The Expression dialog box appears, as shown:



12. Click OK to exit the Expression dialog box. Click OK to exit the Placeholder Properties dialog box.
13. Drag a matrix from the Toolbox and drop it on the report body layout area below the text box. This will create a tablix data region using the matrix template.
14. From the Report Data window, drag the Month field and drop it in the Columns cell.
15. In addition to dragging fields from the Report Data window, we can use the pop-up field picker to select fields. Hover over the Rows cell until the field picker icon appears, as shown:



16. Click the field picker icon to display the pop-up field list. Select the Product field from the field list.
17. Select the Sales\_in\_Dollars field in the Data cell.
18. Drag the Quarter field and drop it on the line forming the top of the Month cell.
19. Drag the Store field and drop it on the line forming the left side of the Product cell. The report layout should appear, as shown:



## Steps to Enable Drilldown

1. As requested in the business needs, we are going to enable drilldown in this report. Click the drop-down arrow for the Month group in the Column Groups area at the bottom of the screen. Select Group Properties from the drop-down menu. The Group Properties dialog box appears.
2. Select the Visibility page.
3. Select the Hide option under the When the report is initially run prompt.
4. Check the Display can be toggled by this report item check box.
5. Select Quarter from the drop-down list.
7. Click the drop-down arrow for the Product group in the Row Groups area. Select Group Properties from the menu. The Group Properties dialog box appears.
8. Select the Visibility page.
9. Select the Hide option under the When the report is initially run prompt.
10. Check the Display can be toggled by this report item check box.
11. Select Store from the drop-down list.
12. Click OK to exit the Group Properties dialog box.
13. Click the Preview tab. The report appears, as shown:

Wholesale Sales By...Report.rdl [Design] X

Design Preview

Date Hierarchy 2015 View Report

1 of 1 100% Find

**Wholesale Sales By Month Report 2015**

		2015-Q1	2015-Q2	2015-Q3	2015-Q4
Store	Product				
Online		8841975.28	8823874.84	8978806.4	9051652.32
Retail-Store 1		347636.16	348292.2	358511.56	314317.56
Retail-Store 2		337734.04	328844.36	291433	265555.72
Retail-Store 3		301872.8	285150.84	244707.56	202801.12
Retail-Store 5		609228.72	461444	400568.2	360686.04
Wholesale		29391877.12	29950211.08	29380094.6	27109714.16

### Steps to Accomplish Additional Report Formatting

1. Click the Design tab. Let's make a few changes to improve the looks and readability of the matrix.
2. Select the cell containing the Sum of the Sales\_in\_Dollars field (the detail cell for the matrix). Set the following properties in the Properties window: "Format" as C and "Size:Width" as 1.5 in  
The "C" value provides currency formatting for the number in this cell.
3. Select the cell containing the Quarter field. Set the following properties in the Properties window: „BackgroundColor" as CornflowerBlue „Font:FontWeight" as Bold and „TextAlign" as Right.
4. Select the cell containing the Month field. Use the toolbar buttons to set the font weight to bold, the background color to PowderBlue, and the alignment to align right. (Hint: Click the More colors link to select the color by name.)
5. Select the cell containing the Store field. (Not the Store heading.) Use the toolbar buttons to set the font weight to bold and the background color to CornflowerBlue.
6. Select the cell containing the Product field. Use the toolbar buttons to set the font weight to bold and the background color to PowderBlue.

### Steps to Add a Calculated Column

1. Right-click the detail cell. Select Tablix: Insert Row | Inside Group - Below. A new detail cell is created below the existing detail cell.
2. Right-click the new detail cell and select Textbox: Expression from the context menu. The Expression dialog box appears.
3. Enter the following expression:  

$$=Sum(Fields!Sales\_in\_Dollars.Value) / Sum(Fields!Sales\_in\_Dollars.Value, „Quarter”)$$
4. Click OK to exit the Expression dialog box. You will see “<<Expr>>” as a placeholder for the expression.
5. Click to the right of the <<Expr>> placeholder in the new detail cell until you see the blinking edit cursor in the cell.
6. Enter a space followed by **of quarter** in the cell after the placeholder.
7. Click the <<Expr>> placeholder until it is highlighted.
8. Right-click the <<Expr>> placeholder and select Placeholder Properties from the context menu. The Placeholder Properties dialog box appears.
9. Select the Number page.
10. Select Percentage from the Category list.
11. Select the Font page.
12. Check the Bold check box.
13. Click OK to exit the Placeholder Properties dialog box.
14. Click elsewhere to unselect this cell, and then click this cell again. Now the entire cell is selected, rather than just one portion of the content.
15. Use the toolbar buttons to set the alignment to align right.
16. Drag the right side of the report layout area (report body) as far left as possible.
17. Select the Preview tab. Your report appears, as shown:

		2015-Q1	2015-Q2	2015-Q3	2015-Q4
Store	Product				
Online		\$8,841,975.28	\$8,823,874.84	\$8,978,806.40	\$9,051,652.32
		22.20% of quarter	21.95% of quarter	22.64% of quarter	24.26% of quarter
Retail-Store 1		\$347,636.16	\$348,292.20	\$358,511.56	\$314,317.56
		0.87% of quarter	0.87% of quarter	0.90% of quarter	0.84% of quarter
Retail-Store 2		\$337,734.04	\$328,844.36	\$291,433.00	\$265,555.72
		0.85% of quarter	0.82% of quarter	0.73% of quarter	0.71% of quarter
Retail-Store 3		\$301,872.80	\$285,150.84	\$244,707.56	\$202,801.12
		0.76% of quarter	0.71% of quarter	0.62% of quarter	0.54% of quarter
Retail-Store 5		\$609,228.72	\$461,444.00	\$400,568.20	\$360,686.04
		1.53% of quarter	1.15% of quarter	1.01% of quarter	0.97% of quarter
Wholesale		\$29,391,877.12	\$29,950,211.08	\$29,380,094.60	\$27,109,714.16
		73.79% of quarter	74.51% of quarter	74.09% of quarter	72.67% of quarter

18. Try expanding the quarter columns and the store rows. Click the Save All button on the toolbar when you are done exploring the report.

## IV. Creating the Max Min Sales Data Model in Power BI

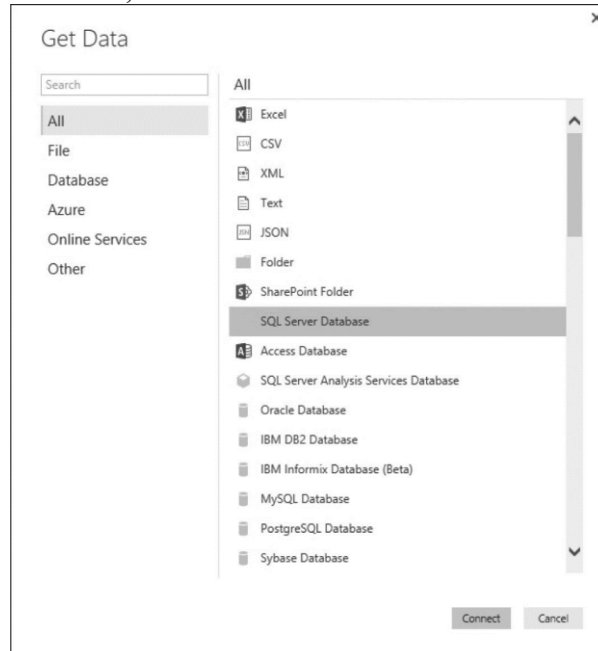
**Business Need** The VP of sales for Max Min, Inc., would like to use Power BI for self-service analytics. Analysis will be based on sales in U.S. dollars and sales in number of units. He wants to create his own data model so he can add data and measures on his own in the



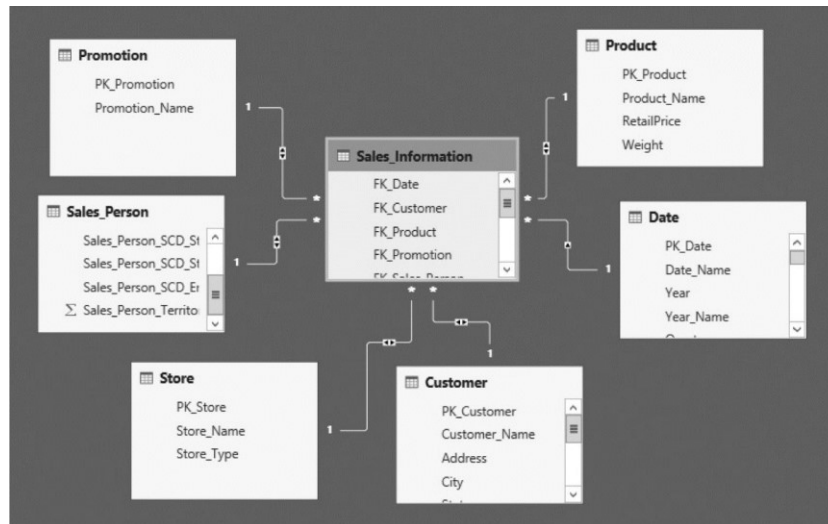
future. However, he wants to get the data into the model in the most straightforward way possible.

### Steps to Bring the MaxMinSalesDM Data into a Power BI Model

1. Open Power BI Desktop. The Start dialog box appears.
2. Click Get Data. The Get Data dialog box appears.
3. Select SQL Server Database, as shown:



4. Click Connect. The SQL Server Database connection dialog box appears.
5. For Server, type the name of the SQL Server instance hosting the MaxMinSalesDM database. You can type **(local)**.
6. For Database, type **MaxMinSalesDM**. The SQL Server Database Connection dialog box should appear.
7. Ensure the Import option is selected and click OK. The Access a SQL Server Database dialog box appears.
8. If the Access a SQL Server Database dialog box appears, make sure the Use my current credentials option is selected and click Connect.
9. If an Encryption Support dialog box appears, click OK.
10. The Navigator dialog box appears. Check the box for each of the following tables: Customer, Date, Production, Sales\_Information, Sales\_Person and Store.
11. Click Load. The data from the selected tables is loaded into the Power BI data model. Once the data load is complete, you will be taken to the Report page of Power BI Desktop.
12. Click the button for the Relationships page on the left side of the window. With some rearranging, the Relationships page appears as:



The relationships in the model were created from relationships (foreign key constraints) that exist in the MaxMinSalesDM database. Each of the lookup tables (Customer, Date, Product, Promotion, Sales\_Person, and Store) has a relationship to the main data table (Sales\_Information).

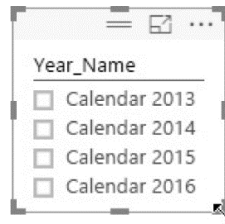
13. Click the button for the Data page on the left side of the window. (The Data button is right above the Relationships button.) You see the data in the Customer table.

14. Click on any of the table names in the **Fields** on the right side of the window. You will see the data loaded into the model for that table.

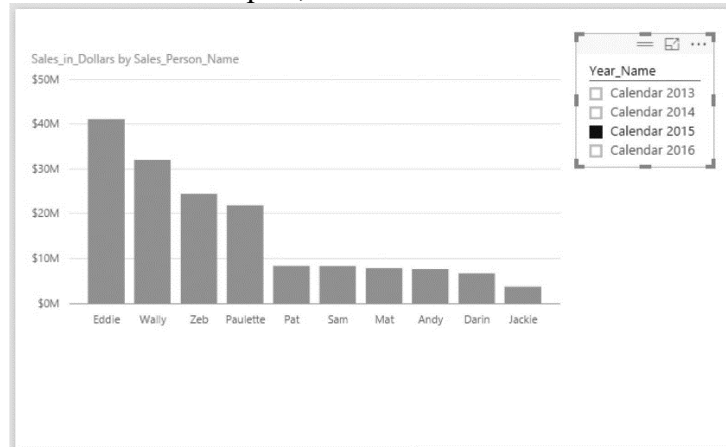
### Steps to Visualize Data in a Power BI Model

1. Click the Report button on the left side of the window (above the Data button). You see the blank report layout area.
2. In the Fields area on the right, expand the Sales\_Information table and scroll so you can see all of the fields in the Sales\_Information table. Note the Greek letter sigma next to several of the fields in the Sales\_Information table. This signifies the model knows how to add up the values for these fields across multiple records in the table. For example, we can get the total Sales\_in\_Dollars for a customer or get the total Sales\_in\_Units for a particular month.
3. Click the check box next to Sales\_in\_Dollars in the Fields area. A bar chart showing the total sales in dollars across all of the data in the model is displayed.
4. Click the Page View button on the ribbon and select Actual Size from the drop-down menu. The bar chart is now bigger and easier to read.
5. In the Fields area, expand the Sales\_Person table.
6. Click the check box next to Sales\_Person\_Name. You may need to make the Fields area wider or hover over the item to see the tooltip to determine which one to select.
7. Click somewhere in the report layout area that is outside of the chart so the chart is no longer selected.
8. Scroll up in the Fields area and expand the Date table.
9. Click the check box next to Year\_Name. A table of the data in the Year\_Name field is created.
10. Drag this table so it is on the right side of the report layout area near the top.
11. In the Visualizations area, click the Slicer button. The Year\_Name table is now a slicer. A slicer filters the values on the report page by what is selected. Note that having no items selected in the slicer is the same as having all items selected in the slicer.
12. Use the sizing handles to remove the whitespace surrounding the year names in the slicer as shown:





13. Arrange the two items on the report, as shown:



14. Click the Calendar 2015 entry in the slicer. The chart shows sales for each salesperson in calendar year 2015.

15. Click the Calendar 2014 entry in the slicer. The chart shows sales for 2014. Note that by default the chart is sorted by Sales\_in\_Dollars descending. When we use the slicer to switch the data from one year to

another, the sort order changes according to who were the top sellers in that year.

16. Hold down CTRL and click Calendar 2015. The chart now shows the combined sales results for calendar year 2014 and calendar year 2015.

17. Click the Save button at the top of the window. The Save As dialog box appears. Navigate to the location where you would like to save the content created.

18. Create a new folder called **PowerBI** . Double-click the PowerBI folder to navigate into it.

19. Enter **Max Min Sales Information** for File name.

20. Click Save.

Congratulations. You've built your first interactive Power BI report.

## V. Adding Hierarchies to the Max Min Sales Information Data Model

**Business Need** We are still working to complete the Max Min Sales Information Power BI data model in order to begin using it to create visualizations.

### Steps to Add Hierarchies

1. On the left side of the window, select Report. The Report view of the model appears.
2. In the Fields area, expand the entry for the Customer table.
3. Right-click the entry for the State field. (You can also click the ellipsis (...) button to see the same context menu.)
4. Select New hierarchy from the context menu. A new hierarchy called State Hierarchy is created. The State field is set as the highest level in the hierarchy.
5. Right-click the State Hierarchy entry and select Rename from the context menu.
6. Type **Geographic Hierarchy** and press ENTER . The hierarchy moves so it is in the correct sort order among the fields in the Customer table.

7. Select the City field and drop it on top of the Geographic Hierarchy entry. City is added as a lower level of the hierarchy.
8. Select the Customer field and drop it on top of the Geographic Hierarchy entry. Customer is added as a lower level of the hierarchy. The Fields area should appear as shown
9. Collapse the entry for the Customer table.
10. Expand the entry for the Date table.
11. Select the Quarter field and drop it on top of the Year field. A new hierarchy is created called Year Hierarchy with Year as the highest level and Quarter as the lowest level.
12. Right-click the Year Hierarchy and select Rename from the context menu.
13. Type **Date Hierarchy** and press ENTER .
14. Select the Month field and drop it on top of the Date Hierarchy entry. Month is added as a lower level of the hierarchy.
15. Select the Date field and drop it on top of the Date Hierarchy entry. Date is added as a lower level of the hierarchy.
16. Collapse the Date Hierarchy entry. We are doing this to make it easier to correctly complete the following steps.
17. Right-click the Year field entry (the Year field entry not in the Date Hierarchy) and select Hide from the context menu.
18. Repeat Step 17 for the Quarter, Month, and Date fields.
19. Now, if we want to use Year, Quarter, Month, or Date in our report, we must access them through the Date Hierarchy.
20. Collapse the entry for the Date table.
21. Expand the entry for the Store table.
22. Select the Store field and drop it on top of the Store Type field. A new hierarchy called Store Type Hierarchy is created.
23. Right-click the Store field entry that is not part of the hierarchy and select Hide from the context menu.
24. Right-click the Store Type field entry that is not part of the hierarchy and select Hide from the context menu.
25. Click Save.

## **VI. Completing the Max Min Sales Information Data Model**

**Business Need** We need to complete the Max Min Sales Information Power BI data model in order to begin using it to create visualizations.

### **Steps to Clean Up the Customer Table**

1. Start up Power BI Desktop. The Start dialog appears.
2. Click Max Min Sales Information.pbix on the left side of the dialog box. The Max Min Sales Information Power BI file opens displaying the Sales\_in\_Dollars by Sales\_Person\_Name report.
3. On the left side of the window, select Data. The Data view of the model appears.
4. Select the Customer table in the Fields area, if it is not selected by default.
5. The PK\_Customer field is used for creating relationships in the model, but does not provide any value as we analyze the data. Right-click the PK\_Customer heading and select Hide in Report View from the context menu.
6. Right-click the Customer\_Name heading and select Rename from the context menu.
7. Type **Customer** and press ENTER .
8. Click the Address heading.
9. On the Modeling tab of the ribbon, select Address from the Data Category drop-down list.

10. Click the City heading.
11. On the Modeling tab of the ribbon, select City from the Data Category drop-down list.
12. Click the State heading.
13. On the Modeling tab of the ribbon, select State or Province from the Data Category drop-down list.
14. Click the ZipCode heading.
15. On the Modeling tab of the ribbon, select Postal Code from the Data Category drop-down list.
16. Right-click the ZipCode heading and select Rename from the context menu.
17. Type **Zip Code** and press ENTER .
18. Right-click the Homeowner heading and select Rename from the context menu.
19. Type **Home Owner?** and press ENTER .
20. Right-click the MaritalStatus heading and select Rename from the context menu.
21. Type **Married?** and press ENTER .
22. Right-click the NumCarsOwned heading and select Rename from the context menu.
23. Type **Number of Cars Owned** and press ENTER .
24. On the Modeling tab of the ribbon, select Average from the Default Summarization drop-down list.
25. Right-click the NumChildrenAtHome heading and select Rename from the context menu.
26. Type **Number of Children at Home** and press ENTER .
27. On the Modeling tab of the ribbon, select Average from the Default Summarization drop-down list.

**Steps to Clean Up the Date Table** In the Date table, there is a human-readable version of each field and a sortable version of each field. We will hide the sortable version of the field while renaming and setting the Sort By property for the human-readable version of the field.

1. Select the Date table in the Fields area.
2. Right-click the PK\_Date heading and select Rename from the context menu.
3. Type **Date** and press ENTER .
4. In the Format drop-down list on the Modeling tab of the ribbon, select Date Time | 3/14/2001 (M/d/yyyy).
5. Right-click the Date\_Name heading and select Rename from the context menu.
6. Type **Long Date Name** and press ENTER .
7. Right-click the Year heading and select Rename from the context menu.
8. Type **YearSort** and press ENTER .
9. Right-click the YearSort heading and select Hide in Report View from the context menu.
10. Right-click the Year\_Name heading and select Rename from the context menu.
11. Type **Year** and press ENTER .
12. Click the Sort By Column button on the Modeling tab of the ribbon.
13. Select YearSort from the menu.
14. Right-click the Quarter heading and select Rename from the context menu. (Scroll right when needed throughout these instructions.)
15. Type **QtrSort** and press ENTER .
16. Right-click the QtrSort heading and select Hide in Report View from the context menu.
17. Right-click the Quarter\_Name heading and select Rename from the context menu.
18. Type **Quarter** and press ENTER .
19. Click the Sort By Column button.
20. Select QtrSort from the menu.
21. Right-click the Month heading and select Rename from the context menu.

22. Type **MnSort** and press ENTER .
23. Right-click the MnSort heading and select Hide in Report View from the context menu.
24. Right-click the Month\_Name heading and select Rename from the context menu.
25. Type **Month** and press ENTER .
26. Click the Sort By Column button.
27. Select MnSort from the menu.
28. Right-click the Day\_Of\_Year heading and select Hide in Report View from the context menu.
29. Right-click the Day\_Of\_Year\_Name heading and select Rename from the context menu.
30. Type **Day of Year** and press ENTER .
31. Click the Sort By Column button.
32. Select Day\_Of\_Year from the menu.
33. Right-click the Day\_Of\_Quarter heading and select Hide in Report View from the context menu.
34. Right-click the Day\_Of\_Quarter\_Name heading and select Rename from the context menu.
35. Type **Day of Quarter** and press ENTER .
36. Click the Sort By Column button.
37. Select Day\_Of\_Quarter from the menu.
38. Right-click the Day\_Of\_Month heading and select Hide in Report View from the context menu.
39. Right-click the Day\_Of\_Month\_Name heading and select Rename from the context menu.
40. Type **Day of Month** and press ENTER .
41. Click the Sort By Column button.
42. Select Day\_Of\_Month from the menu.
43. Right-click the Month\_Of\_Year heading and select Hide in Report View from the context menu.
44. Right-click the Month\_Of\_Year\_Name heading and select Rename from the context menu.
45. Type **Month of Year** and press ENTER .
46. Click the Sort By Column button.
47. Select Month\_Of\_Year from the menu.
48. Right-click the Month\_Of\_Quarter heading and select Hide in Report View from the context menu.
49. Right-click the Month\_Of\_Quarter\_Name heading and select Rename from the context menu.
50. Type **Month of Quarter** and press ENTER .
51. Click the Sort By Column button.
52. Select Month\_Of\_Quarter from the menu.
53. Right-click the Quarter\_Of\_Year heading and select Hide in Report View from the context menu.
54. Right-click the Quarter\_Of\_Year\_Name heading and select Rename from the context menu.
55. Type **Quarter of Year** and press ENTER .
56. Click the Sort By Column button.
57. Select Quarter\_Of\_Year from the menu.

### **Steps to Clean Up the Product and Promotion Tables**

1. Select the Product table in the Fields area.

2. The PK\_Product field is used for creating relationships in the model, but does not provide any value as we analyze the data. Right-click the PK\_Product heading and select Hide in Report View from the context menu.
3. Right-click the Product\_Name heading and select Rename from the context menu.
4. Type **Product** and press ENTER .
5. Click the RetailPrice heading and select Rename from the context menu.
6. Type **Retail Price** and press ENTER .
7. In the Format drop-down list on the Modeling tab of the ribbon, select Currency | \$ English (United States).
8. In the Default Summarization drop-down list on the Modeling tab of the ribbon, select Average.
9. Click the Weight heading.
10. In the Default Summarization drop-down list on the Modeling tab of the ribbon, select Average.
11. Select the Promotion table in the Fields area.
12. The PK\_Promotion field is used for creating relationships in the model, but does not provide any value as we analyze the data. Right-click the PK\_Promotion heading and select Hide in Report View from the context menu.
13. Right-click the Promotion\_Name heading and select Rename from the context menu.
14. Type **Promotion** and press ENTER .

**Steps to Clean Up the Sales\_Information Table** We will rename the Sales\_Information table itself to get rid of the underscore character. We will also hide all of the foreign key fields used to link this table to the other tables in the model. Finally, we will delete the Sales\_Tax and Shipping columns, which do not contain any data.

1. In the Fields area, right-click the Sales\_Information table and select Rename from the context menu.
2. Type **Sales Information** and press ENTER .
3. Right-click the FK\_Date heading and select Hide in Report View from the context menu.
4. Repeat Step 3 to hide the following fields:  
FK\_Customer, FK\_Product, FK\_Promotion, FK\_Sales\_Person and FK\_Store.
5. Right-click the Sales\_in\_Dollars heading and select Rename from the context menu.
6. Type **Sales in US Dollars** and press ENTER .
7. In the Format drop-down list on the Modeling tab of the ribbon, select Currency | \$ English (United States).
8. Right-click the Sales\_in\_Units heading and select Rename from the context menu.
9. Type **Sales in Units** and press ENTER .
10. Right-click the Sales\_Tax heading and select Delete from the context menu. The Delete Column dialog box appears.
11. Click Delete.
12. Right-click the Shipping heading and select Delete from the context menu. The Delete Column dialog box appears.
13. Click Delete.

### **Steps to Clean Up the Sales\_Person and Store Tables**

1. In the Fields area, right-click the Sales\_Person table and select Rename from the context menu.
2. Type **Sales Person** and press ENTER .
3. Select the Sales Person table in the Fields area.
4. Right-click PK\_Sales\_Person and select Hide in Report View from the context menu.

5. Right-click Sales\_Person\_Name and select Rename from the context menu.
6. Type **Sales Person** and press ENTER .
7. Right-click Sales\_Person\_SCD\_Original\_ID and select Hide in Report View from the context menu.
8. Right-click Sales\_Person\_SCD\_Status and select Rename from the context menu.
9. Type **Status** and press ENTER .
10. Right-click Sales\_Person\_SCD\_Start\_Date and select Rename from the context menu.
11. Type **Sales Person Start Date** and press ENTER .
12. In the Format drop-down list on the Modeling tab of the ribbon, select Date Time | 3/14/2001 (M/d/yyyy).
13. Right-click Sales\_Person\_SCD\_End\_Date and select Rename from the context menu.
14. Type **Sales Person End Date** and press ENTER .
15. In the Format drop-down list on the Modeling tab of the ribbon, select Date Time | 3/14/2001 (M/d/yyyy).
16. Right-click Sales\_Person\_Territory and select Rename from the context menu.
17. Type **Sales Territory** and press ENTER .
18. In the Default Summarization drop-down list in the Modeling tab of the ribbon, select Do Not Summarize.
19. Select the Store table in the Fields area.
20. Right-click PK\_Store and select Hide in Report View from the context menu.
21. Right-click Store\_Name and select Rename from the context menu.
22. Type **Store** and press ENTER .
23. Right-click Store\_Type and select Rename from the context menu.
24. Type **Store Type** and press ENTER .
25. Click Save.

## VII. Loading World Population Data from the Internet

**Business Need** We will be analyzing world population, Gross Domestic Product (GDP), and world CO<sub>2</sub> emissions from the late 20<sup>th</sup> century to the present using Power BI. We will begin by obtaining population information by country for that time period from the Internet.

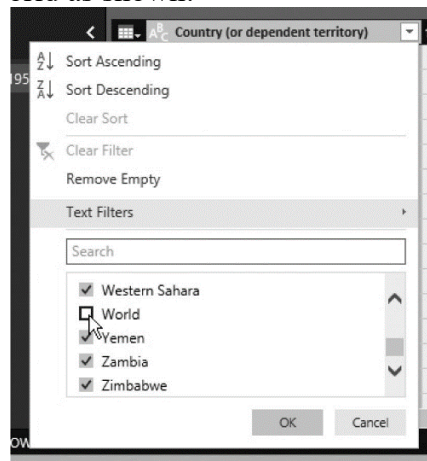
### Steps to Create a Connection to Web Data Source

1. Click the File tab of the menu and select New from the menu. The Start dialog box appears.
2. Click Get Data. The Get Data dialog box appears.
3. Select Other from the category list on the left, and then select Web.
4. Click Connect. The From Web connection dialog box appears.
5. Enter the following URL:  
[http://en.wikipedia.org/wiki/List\\_of\\_countries\\_and\\_dependencies\\_by\\_population](http://en.wikipedia.org/wiki/List_of_countries_and_dependencies_by_population)
6. Click OK. Power BI will connect to this web page and look for any tables on the page. You will see the Navigator dialog box with entries for the entire web page document and four tables within that document.
7. Check the box next to Estimates between the years 1950 and 1980 (In Thousands) [edit]. You see a sample of the data in this table.
8. Click Edit. (The Load button is very tempting with its bright yellow color. Be sure you click Edit instead.) You see the Query Editor with the information from the population table loaded. In addition to connecting to the web page and navigating to the desired table, you will see Power BI automatically added a step to change the data type of some of the fields as they were imported.

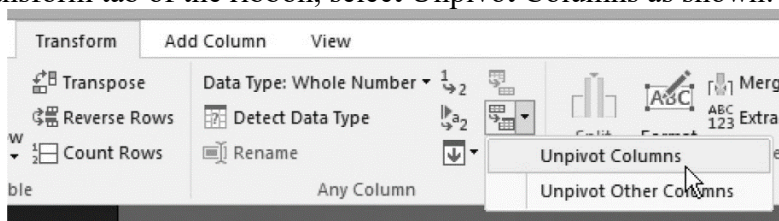


## Steps to Transform the Data in the Query Editor

1. The N° column should be selected, but click the N° column heading to make sure.
2. On the Home tab of the ribbon, click Remove Columns. The N° column is removed.
3. Click the % column heading to select that column. Don't be confused by the data type indicator in the column heading. This particular column heading looks a bit strange because it includes the "1.2" that indicates it is a decimal data type and the "%" that is the actual name of the column.
4. Scroll the data view to the right.
5. Hold down CTRL and select the %2 column. Both the % and %2 columns are selected.
6. Continue to hold down CTRL and select %3, %4, %5, and %6. Scroll right as needed.
7. On the Home tab of the ribbon, click Remove Columns. All of the selected columns are removed.
8. Click the drop-down arrow in the Country (or dependent territory) column. The sorting and filtering drop-down appears.
9. Scroll down to find "World" in the list of countries and dependent territories.
10. Uncheck the entry for World as shown:



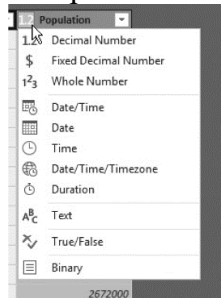
11. Click OK to exit the sorting and filtering drop-down. The row with the world totals is removed from the data.
12. Click the 1950 column heading.
13. Scroll right, hold down SHIFT, and click the 1980 column heading. The columns from 1950 through 1980 are selected.
14. On the Transform tab of the ribbon, select Unpivot Columns as shown:



The columns are unpivoted to create rows with new columns called Attribute and Value. The items in the Attribute column were the column names before the pivot. The figures in the Value column were the contents of those columns before the pivot.

15. Click the Attribute column heading so the Attribute column is selected on its own.
16. Right-click the Attribute column heading and select Rename from the context menu.
17. Type **Year** and press ENTER.
18. Right-click the Value column heading and select Rename from the context menu.
19. Type **Population** and press ENTER.

20. On the Transform tab of the ribbon, select Standard | Multiply. The Multiply dialog box appears.
21. Type **1000** for Value.
22. Click OK. The Population numbers are multiplied by 1000 to show their true magnitude.
23. Click the data type indicator in the Population column, as shown:

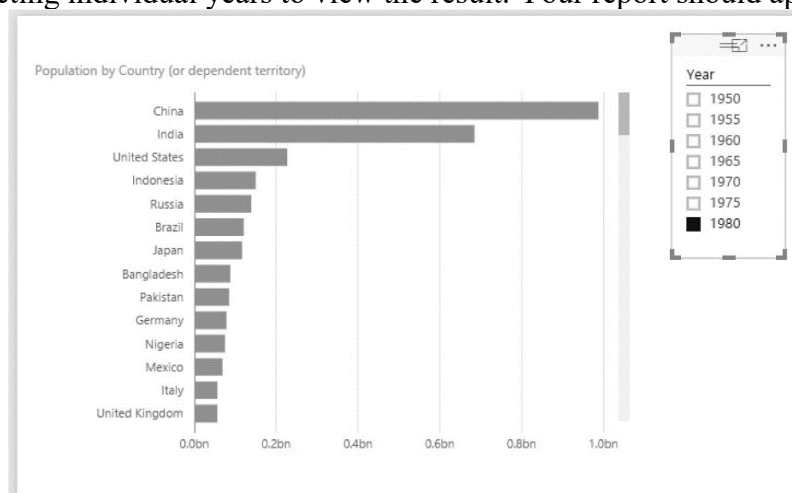


The data type drop-down list appears.

24. Select Whole Number. The data type of the Population column is changed to whole number.
25. On the right, under Properties, replace the content of Name with **Population 1950-1980**.
26. On the Home tab of the ribbon, click Close & Apply. The data is pulled from the web page, the transformations are applied, and the transformed data is loaded into the data model.

### Steps to Verify the Data

1. On the Home tab of the ribbon, click Page View | Actual Size.
2. In the Fields area on the Report page of Power BI Desktop, check the box for Population. A column chart is created.
3. In the Visualizations area, click the Stacked bar chart icon. The chart is changed to a bar chart.
4. In the Fields area, check the box for Country. At the moment, this is a meaningless number because it is totaling the population for each country across all of the years in the data model.
5. Click anywhere in the report layout area outside of the chart. The chart is unselected.
6. In the Fields area, check the box for Year. A table listing the Years is created.
7. In the Visualizations area, click Slicer. The Year table is converted to a slicer.
8. Drag the Year slicer to the upper-right corner of the layout area and size it to remove the whitespace.
9. Select the bar chart and size it to fill most of the rest of the layout area.
10. Try selecting individual years to view the result. Your report should appear similar to:





11. Click Save.
12. If necessary, navigate to the PowerBI folder you created previously.
13. Enter **Population-GDP-CO2 by Country** for File name.
14. Click Save.

## VIII. Completing the Population, GDP, and CO<sub>2</sub> Data Model

**Business Need** We need to complete the data import for our World Population, GDP, and CO<sub>2</sub> model.

### Steps to Create a Connection to Web Data Source

1. On the Home tab of the ribbon, click Recent Sources. The Recent Sources drop-down appears.
2. Select the entry that begins “[http://en.wikipedia.org/wiki/List\\_of\\_countries\\_by\\_...](http://en.wikipedia.org/wiki/List_of_countries_by_...)” The Navigator dialog box appears with entries for the entire web page document and the four tables in that document.
3. Check the box next to Estimates between the years 1985 and 2015(In Thousands) [edit]. You see a sample of the data in this table.
4. Click Edit. (Again, the Load button is very tempting with its bright yellow color. Be sure you click Edit instead.) You see the Query Editor with the information from the population table loaded. Power BI automatically created steps to connect to the web page, load the data from the table, and change the data type of some of the fields as they were imported.

### Steps to Transform the Data in the Query Editor

1. The N<sup>o</sup> column should be selected, but click the N<sup>o</sup> column heading to make sure.
2. Hold down CTRL and select %, %2, %3, %4, %5, %6, and %7. Scroll right as needed.
3. On the Home tab of the ribbon, click Remove Columns. All of the selected columns are removed.
4. Click the drop-down arrow in the Country (or dependent territory) column. The sorting and filtering drop-down appears.
5. Scroll down to find “World” in the list of countries and dependent territories.
6. Uncheck the entry for World.
7. Click OK to exit the sorting and filtering drop-down. The row with the world totals is removed from the data.
8. Click the 1985 column heading.
9. Scroll right, hold down SHIFT, and click the 2015 column heading. The columns from 1985 through 2015 are selected.
10. On the Transform tab of the ribbon, select Unpivot Columns. The columns are unpivoted to create rows with new columns called Attribute and Value. The items in the Attribute column were the column names before the pivot. The figures in the Value column were the contents of those columns before the pivot.
11. Click the Attribute column heading so the Attribute column is selected on its own.
12. Right-click the Attribute column heading and select Rename from the context menu.
13. Type **Year** and press ENTER.
14. Right-click the Value column heading and select Rename from the context menu.
15. Type **Population** and press ENTER.
16. On the Transform tab of the ribbon, select Standard | Multiply. The Multiply dialog box appears.
17. Type **1000** for Value.
18. Click OK. The Population numbers are multiplied by 1000 to show their true magnitude.

19. Click the data type indicator in the Population column. The data type drop-down list appears.
20. Select Whole Number. The data type of the Population column is changed to whole number.
21. On the right under Properties, replace the content of Name with **Population** and press ENTER .
22. On the Home tab of the ribbon, click Combine. If you do not see Combine on the ribbon, click Append Queries.
23. If you clicked Combine, select Append Queries from the Combine drop-down list. The Append dialog box appears.
24. From the Table to append drop-down list, select Population 1950–1980.
25. Click OK. The data we imported into the Population 1950–1980 table is appended to our current Population query.
26. On the Home tab of the ribbon, click Close & Apply. The data is pulled from the web page, the transformations are applied, the 1950–1980 data is appended, and the result is loaded into the data model. Note, the data visualization is not impacted by the changes made to the data model, so it will be unchanged.

### Steps to Load the GDP from the Excel File

1. Click Get Data. The Get Data dialog box appears.
2. Select Excel and click Connect. The Open dialog box appears.
3. Navigate to the location where you saved the “GDP in Current US Dollars.xlsx” file.
4. Select the “GDP in Current US Dollars.xlsx” file and click Open. The Navigator appears showing entries for the two spreadsheet tabs in the Excel workbook.
5. Check the entry for the Data tab and click Edit. (Again, be sure you don’t click Load.)
6. On the Home tab of the ribbon, click Use First Row As Headers. The first row of data becomes the headings for the columns.
7. On the Home tab of the ribbon, click Choose Columns. The Choose Columns dialog box appears with a list of the columns.
8. Click (Select All Columns) to uncheck all of the items in the list.
9. We only want GDP figures for those years that are present in our Population table. Check the following columns: Country Name, 1960, 1965, 1970, 1975, 1980, 1985, 1990, 1995, 2000, 2005 and 2010.
10. Click OK. Only the selected columns remain.
11. We need to replace the “. .” entries in the table with a null value. Click the 1960 column heading.
12. Hold down SHIFT and click the 2010 column heading to select all of the columns from 1960 to 2010.
13. On the Home tab of the ribbon, click Replace Values. The Replace Values dialog box appears.
14. Enter .. for Value To Find. We want to leave Replace With blank.
15. Click OK.
16. On the Transform tab of the ribbon, select Unpivot Columns.
17. Click the Attribute column to select that column by itself.
18. Right-click the Attribute column and select Rename from the context menu.
19. Type **Year** and press ENTER .
20. Right-click the Value column and select Rename from the context menu.
21. Type **GDP** and press ENTER .
22. Click the data type indicator in the GDP column. The data type drop-down list appears.
23. Click Whole Number.

24. On the right under Properties, replace the content of Name with **GDP Data** and press ENTER .
25. On the Home tab of the ribbon, click Close & Apply. The data is pulled from the Excel file, the transformations are applied, and the result is loaded into the data model.

### **Steps to Load the CO<sub>2</sub> Data from the Excel File**

1. Click Get Data. The Get Data dialog box appears.
2. Select CSV and click Connect. The Open dialog box appears.
3. Navigate to the location where you saved the “CO<sub>2</sub> Emissions in kt.csv” file.
4. Select the “CO<sub>2</sub> Emissions in kt.csv” file and click Open. The CO<sub>2</sub> Emissions in kt.csv preview screen appears.
5. Click Edit.
6. On the Home tab of the ribbon, click Use First Row As Headers. The first row of data becomes the headings for the columns.
7. On the Home tab of the ribbon, click Choose Columns. The Choose Columns dialog box appears with a list of the columns.
8. Click (Select All Columns) to uncheck all of the items in the list.
9. We only want GDP figures for those years that are present in our Population table. Check the following columns: Country Name, 1960, 1965, 1970, 1975, 1980, 1985, 1990, 1995 ,2000 ,2005 and 2010.
10. Click OK. Only the selected columns remain.
11. We need to replace the “..” entries in the table with a null value. Click the 1960 column heading.
12. Hold down SHIFT and click the 2010 column heading to select all of the columns from 1960 to 2010.
13. On the Home tab of the ribbon, click Replace Values. The Replace Values dialog box appears.
14. Enter .. for Value To Find. We want to leave Replace With blank.
15. Click OK.
16. On the Transform tab of the ribbon, select Unpivot Columns.
17. Click the Attribute column to select that column by itself.
18. Right-click the Attribute column and select Rename from the context menu.
19. Type **Year** and press ENTER .
20. Right-click the Value column and select Rename from the context menu.
21. Type **CO<sub>2</sub> Level** and press ENTER .
22. Click the data type indicator in the GDP column. The data type drop-down list appears.
23. Click Whole Number.
24. On the Home tab of the ribbon, click Close & Apply. The data is pulled from the Excel file, the transformations are applied, and the result is loaded into the data model.
25. Click Save.

## **IX. Completing the Population-GDP-CO<sub>2</sub> by Country Data Model**

**Business Need** We need to complete the Population-GDP-CO<sub>2</sub> by Country Power BI data model in order to begin using it to create visualizations.

### **Steps to Create the Year Intermediate Table**

1. Open Power BI Desktop. The Start dialog box appears.
2. Click “Population-GDP-CO<sub>2</sub> by Country.pbix” on the left side of the dialog box. The Population-GDP-CO<sub>2</sub> by Country Power BI file opens.
3. Click the Data button on the left side of the window.

4. Click Enter Data on the Home tab of the ribbon. The Create Table dialog box appears.
5. In row 1 under Column1, type **1950** and press ENTER .
6. In row 2 under Column1, type **1955** and press ENTER.
7. In row 3 under Column1, type **1960** and press ENTER .
8. Continue this process until you have rows for every five years from 1950 to 2015.
9. Replace “Table1” with **Year** for Name at the bottom of the dialog box. The Create Table dialog box appears as shown:

Create Table

Create a table by typing or pasting content.

	Column1
1	1950
2	1955
3	1960
4	1965
5	1970
6	1975
7	1980
8	1985
9	1990
10	1995
11	2000
12	2005
13	2010
14	2015

Name:

Load Edit Cancel

10. Click Load. The Year table is created in the data model.
11. Select the Year table in the Fields area.
12. Right-click the Column1 column heading and select Rename from the context menu.
13. Type **Year** and press ENTER .
14. In the Default Summarization drop-down list in the Modeling tab of the ribbon, select Do Not Summarize.

**Steps to Create the Country Intermediate Table.** Each of the tables we imported has slightly different names for some countries and territories. We certainly could have taken the time to add more Replace Values operations in the import queries to bring each set of countries into a consistent naming convention. Rather than taking the time to do that in these exercises, we will be using multiple fields in the Countries table. Each field is adapted to the country naming used in a particular data table. Also, country and territory names tend to be a bit fluid. You may need to update the content of the PopulationTableCntry/Terr field to synchronize it with the current naming in the Wikipedia table before bringing the Countries table into the model.

1. Click Get Data on the Home tab of the ribbon. The Get Data dialog box appears.
2. Select SQL Server Database and click Connect. The SQL Server Database dialog box appears.
3. Enter (local) as server name.
4. Enter **CountryInfo** for Database.
5. Click OK. The Navigator appears.
6. Check Countries and click Load. The Countries table is added to the model.
7. Right-click the Countries entry in the Fields area and select Rename from the context menu.
8. Type **Country** and press ENTER .

## Steps to Create Relationships in the Model

1. On the left side of the window, select Relationships. The Relationships view of the model appears.
2. In the Population table, click Country (or dependent territory) and drop it on PopulationTableCntry/Terr in the Country table. An active, one-to-many relationship that filters in both directions is created. (Note the double arrow in the middle of the line created to represent the relationship.)
3. In the GDP Data table, click Country Name and drop it on GDPCntry/Terr in the Country table. An active, one-to-many relationship that filters in both directions is created.
4. In the CO2 Emissions in kt table, click Country Name and drop it on CO2TableCntry/Terr in the Country table. An active, one-to-many relationship that filters in both directions is created.
5. In the Population table, click Year and drop it on Year in the Year table. An active, one-to-many relationship that filters in both directions is created.
6. In the GDP Data table, click Year and drop it on Year in the Year table. An inactive, one-to-many relationship that filters from the Year table to the GDP Data table is created. This new link is inactive because there is already an active path from Population to GDP Data through Country. If this path were active, there would be a second active path from Population to GDP Data. The Power BI model will not support this. We will fix this issue in a moment.
7. In the CO2 Emissions in kt table, click Year and drop it on Year in the Year table. Another inactive, one-to-many relationship that filters in from the Year table to the GDP Data table is created. Again, this is inactive to avoid having two active paths from Population to CO2 Emissions in kt. The problem is our first four relationships filter in both directions. We don't need the bidirectional filtering for these relationships.
8. Let's change these relationships to single-direction filtering. Double-click the relationship line from the Population table to the Country table. The Edit Relationship dialog box appears.
9. Change the Cross filter direction drop-down list to Single. The Edit Relationship dialog box appears as shown:

Edit Relationship

Select tables and columns that relate to one another.

Population

Country (or dependent territory)	Year	Population
Ecuador	1950	3370000
Tajikistan	1950	1531000
Mauritius	1950	482000

Country

Country/Territory	Capital	Continent	PopulationTableCntry/Terr	CO2TableCntry/Terr
Abkhazia	Sukhumi, Abkhazia	Asia	Abkhazia	Abkhazia
Afghanistan	Kabul, Afghanistan	Asia	Afghanistan	Afghanistan
Akrotiri and Dhekelia	Episkopi Cantonment, Akrotiri and Dhekelia	Asia	Akrotiri and Dhekelia	Akrotiri and D

Cardinality: Many to One (\*:1)

Cross filter direction: Single

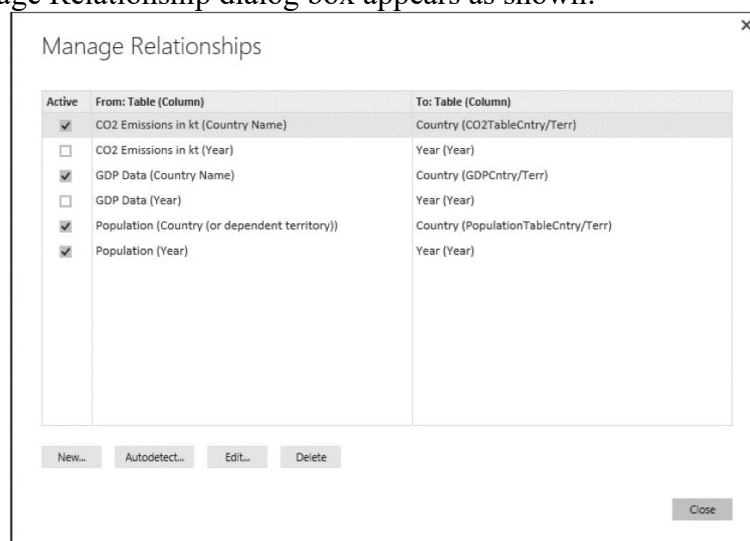
☒ Make this relationship active

☐ Assume Referential Integrity

OK Cancel

10. Click OK to exit the Edit Relationship dialog box.

11. As an alternative to double-clicking each link, we can use the Manage Relationships dialog to make these same changes. Click Manage Relationships on the Home tab of the ribbon. The Manage Relationship dialog box appears as shown:



12. Select the relationship from CO2 Emissions in kt (Country Name) to Country (CO2TableCntry/Terr).

13. Click Edit. The Edit Relationship dialog box appears.

14. Change the Cross filter direction drop-down list to Single.

15. Click OK to exit the Edit Relationship dialog box.

16. Repeat Steps 12–15 to change the following relationships to single:

From: Table (Column)	To: Table (Column)
GDP Data (Country Name)	Country (GDPCntry/Terr)
Population (Year)	Year (Year)

17. Check the box next to the relationship from CO2 Emissions in kt (Year) to Year (Year) to make this relationship active.

18. Check the box next to the relationship from GDP Data (Year) to Year (Year) to make this relationship active. These relationships were created with single-direction cross-filtering, so that property does not need to be changed.

19. Click Close to exit the Manage Relationships dialog box.

Notice the small arrows on the relationship connections all go from the Country table to the other tables or from the Year table to the other tables. Therefore, when we select a country in the Country table, it will filter the other tables to display just the data for that country. When we select a year in the Year table, it will filter the other tables to display just the data from that year. With the single-direction cross-filtering, there is no problem having all of the relationships active.

With the filtering working in this manner, we don't want to select countries and years in the tables containing the data. They should only be selected in the intermediate tables. Therefore, let's hide the Year- and Country-related fields in the data tables. Likewise, when filtering, we don't want the special linking fields in the Country table selected (PopulationTableCntry/Terr, GDPCntry/Terr, CO2TableCntry/Terr). In addition, the data from the Population 1950-1980 table was appended to the Population table. Therefore, we do not want to use the Population 1950-1980 table in our reporting. To simplify our model, we will hide that table, too. We will also add a bit of formatting to the data.

## Steps to Do the Final Model Clean-Up

1. Right-click the heading of the Population 1950-1980 table and select Hide in Report View from the context menu. The table is grayed out.

2. Right-click the Country (or dependent territory) field in the Population table.
3. Select Hide in Report View from the context menu.
4. Right-click the Year field in the Population table.
5. Select Hide in Report View from the context menu.
6. Right-click the Country Name field in the CO2 Emissions in kt table.
7. Select Hide in Report View from the context menu.
8. Right-click the Year field in the CO2 Emissions in kt table.
9. Select Hide in Report View from the context menu.
10. We can also hide fields while viewing the model through the Data view. Click the Data button on the left side of the window. Power BI Desktop switches to Data view.
11. Select the GDP Data table in the Fields area.
12. Right-click the Country Name column heading and select Hide in Report View from the context menu.
13. Right-click the Year column heading and select Hide in Report View from the context menu.
14. Select the Country table in the Fields area on the right.
15. Right-click the PopulationTableCntry/Terr column heading and select Hide in Report View from the context menu.
16. Right-click the CO2TableCntry/Terr column heading and select Hide in Report View from the context menu.
17. Right-click the GDPCntry/Terr column heading and select Hide in Report View from the context menu.
18. Click Save.
19. Close Power BI Desktop.