# Canadian Disaster Database

# Power BI – ETL

# Task 1: Load web data with default format

1. Open Power BI Desktop
2. Click on “Get data”
   1. Type “web” in the search bar
   2. Click “Connect”
   3. In the pop-up window, type “<http://cdd.publicsafety.gc.ca/dl-eng.aspx?cultureCode=en-Ca&normalizedCostYear=1>” in the URL box
   4. Click “OK”, the click “Transform Data”
   5. In the pop-up window “Power Query Editor”, go to bottom left and change to “Column profiling based on entire data set”
   6. Under “View”, make sure to check all the box in “Data Preview” except for Monospaced

**How to install python in your computer for Digital Transformations**:

1. Download python 3.8.0 from <https://www.python.org/downloads/release/python-380/>
2. Install python - make sure to choose add to path in the first splash screen and disable path length limit at the end of the installation process
3. Open CMD with admin privileges (right click and click on run as admin)
4. Check python version with cmd: python --version
5. The command prompt should show python 3.8.0, if not, uninstall from control panel and make sure to choose path before installing again
6. Execute the following commands
7. python -m pip install --upgrade pip
8. pip install manager
9. pip install pandas
10. pip install regex
11. pip install matplotlib
12. Try running the python code in Power BI and check for any error codes and please let me know

(edited)

Python.org**Python.org**

[**Python Release Python 3.8.0**](https://www.python.org/downloads/release/python-380/)

The official home of the Python Programming Language

# Task 2: Develop Python script

1. Open “Digital\_Transformation\_CDD.ipynb” in Jupyter Notebook
2. Follow the notebook instruction to develop script to handle “carriage return” issue

# Task 3: Load web data with Python script

1. Open Power BI Desktop
2. Click on “Get data”
   1. Type “Python script” in the search bar
   2. Click “Connect”
   3. Copy and paste the code developed from Jupyter Notebook

import pandas as pd

import regex as re

import requests

import sys

if sys.version\_info[0] < 3:

from StringIO import StringIO

else:

from io import StringIO

cdd\_url = "http://cdd.publicsafety.gc.ca/dl-eng.aspx?cultureCode=en-Ca&normalizedCostYear=1%22

req = requests.get(cdd\_url)

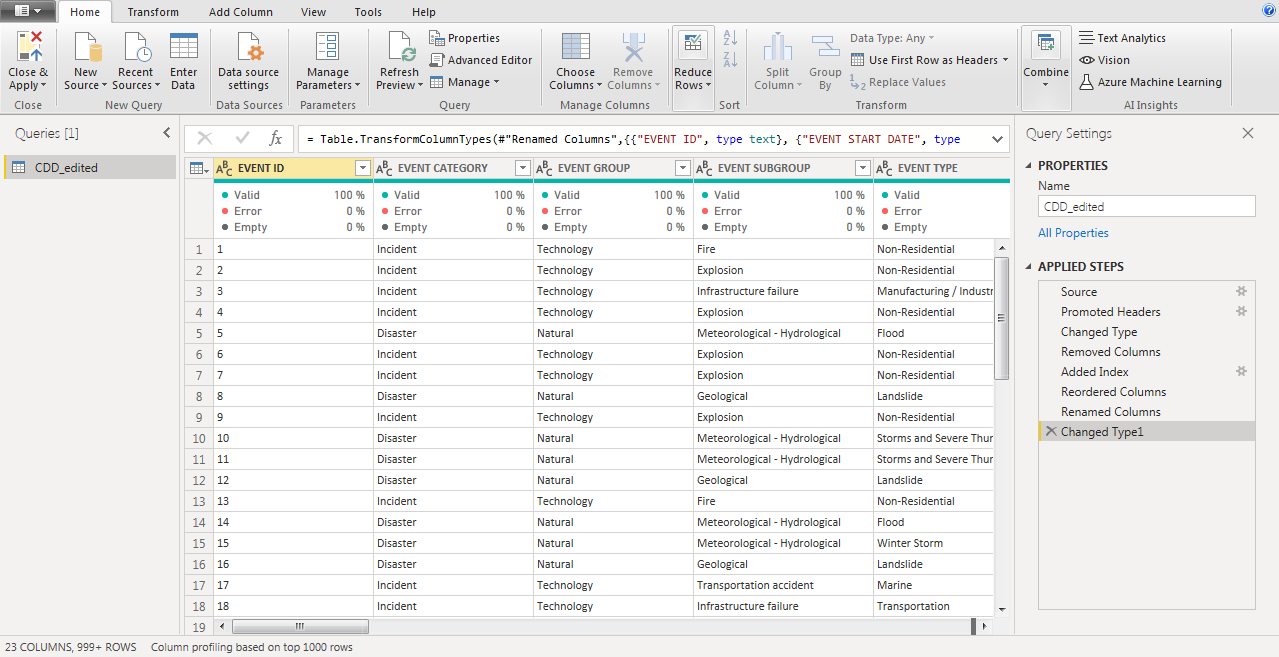
data = req.text

data\_new = re.sub(r"\.\s\*\n+", ". ", data)

CDD\_edited = pd.read\_csv(StringIO(data\_new), sep="\t")

* 1. In the pop-up Navigator window, select “CDD\_edited”, then click “Transform data”
  2. In the pop-up window “Power Query Editor”, go to bottom left and change to “Column profiling based on entire data set”

1. Perform basic data quality check:
   1. Right click the last column, then click “Remove”
   2. Right click table icon, then click “Add Index Column” “From 1”
   3. Drag the Index column to the left (trick: move just one column and edit column sequence in the formula bar)
   4. Double click on “Index” column and rename it to “EVENT ID
   5. Change column type by clicking the icon before each column:
      1. IDs -> text
      2. DATEs -> date
      3. Numbers -> Whole or Decimal Number



# Task 4: Data Transformation – Province

1. Under “Home”, click “New Source”, then click “Blank Query”
2. In the formula bar, type “= CDD\_edited”
3. Double click on “Query1” on the left Queries window, type “Fact\_table”
4. Control +click event related columns, right click column name, then click “Remove Columns”
   1. "EVENT CATEGORY"
   2. "EVENT GROUP"
   3. "EVENT SUBGROUP"
   4. "EVENT TYPE"
   5. "COMMENTS"
   6. "MAGNITUDE"
5. Add a column to indicate if the event affected Alberta
   1. Click on the table icon, the click “Add Conditional Column”
   2. In the pop-up window:
      1. “New column name”: type “AB”
      2. IF “PLACE” “contains” “AB” Then “1” Else “0”
      3. Click “OK”
   3. Check if any other keyword identifies Alberta
      1. Click the triangle icon besides the “AB” column
      2. Only select “0” then click “OK” to filter all the records without the keyword “AB”
      3. Click on “PLACE” and check “Value distribution”
         1. For example, “AB” column should be 1 when “PLACE” column is “Prairie Provinces”
   4. Add other keywords that identify Alberta
      1. In the right side “Query Setting” window, remove “Filtered Rows”
      2. Double click on “Add Conditional Column” then click “Add Clause”
         1. Else if “PLACE” “contains” “Prairie Provinces” Then “1”
         2. Trick: Instead of adding each clause separately, we can copy the code in the formula bar, paste then change the keyword
            1. Copy the following code to the formula bar:

*= Table.AddColumn(#"Removed Columns", "AB", each if Text.Contains([PLACE], "AB") then 1 else if Text.Contains([PLACE], "Across Canada") then 1 else if Text.Contains([PLACE], "Alberta") then 1 else if Text.Contains([PLACE], "Prairie") then 1 else if Text.Contains([PLACE], "Western Canada") then 1 else if Text.Contains([PLACE], "Edmonton") then 1 else 0)*

1. Repeat step 5 to add column for each province
   1. Trick: Instead of adding each column separately, we can copy the code in the formula bar, paste then change the keyword
   2. under “Home”, double click on “Advanced Editor”
   3. Copy the following code to the Advanced Editor

*let*

*Source = CDD\_edited,*

*#"Removed Columns" = Table.RemoveColumns(Source,{"EVENT CATEGORY", "EVENT GROUP", "EVENT SUBGROUP", "EVENT TYPE", "COMMENTS", "MAGNITUDE"}),*

*#"Added Conditional Column" = Table.AddColumn(#"Removed Columns", "AB", each if Text.Contains([PLACE], "AB") then 1 else if Text.Contains([PLACE], "Across Canada") then 1 else if Text.Contains([PLACE], "Alberta") then 1 else if Text.Contains([PLACE], "Prairie") then 1 else if Text.Contains([PLACE], "Western Canada") then 1 else if Text.Contains([PLACE], "Edmonton") then 1 else 0),*

*#"Added Conditional Column1" = Table.AddColumn(#"Added Conditional Column", "BC", each if Text.Contains([PLACE], "BC") then 1 else if Text.Contains([PLACE], "Across Canada") then 1 else if Text.Contains([PLACE], "British Columbia") then 1 else if Text.Contains([PLACE], "Western Canada") then 1 else 0),*

*#"Added Conditional Column2" = Table.AddColumn(#"Added Conditional Column1", "MB", each if Text.Contains([PLACE], "MB") then 1 else if Text.Contains([PLACE], "Across Canada") then 1 else if Text.Contains([PLACE], "Manitoba") then 1 else if Text.Contains([PLACE], "Prairie") then 1 else if Text.Contains([PLACE], "Western Canada") then 1 else 0),*

*#"Added Conditional Column3" = Table.AddColumn(#"Added Conditional Column2", "NB", each if Text.Contains([PLACE], "NB") then 1 else if Text.Contains([PLACE], "Across Canada") then 1 else if Text.Contains([PLACE], "New Brunswick") then 1 else if Text.Contains([PLACE], "Eastern Canada") then 1 else if Text.Contains([PLACE], "Maritime") then 1 else if Text.Contains([PLACE], "Martime") then 1 else 0),*

*#"Added Conditional Column4" = Table.AddColumn(#"Added Conditional Column3", "NL", each if Text.Contains([PLACE], "NL") then 1 else if Text.Contains([PLACE], "Across Canada") then 1 else if Text.Contains([PLACE], "Newfoundland") then 1 else if Text.Contains([PLACE], "Eastern Canada") then 1 else if Text.Contains([PLACE], "Labrador") then 1 else if Text.Contains([PLACE], "Saint-Pierre-et-Miquelon") then 1 else 0),*

*#"Added Conditional Column5" = Table.AddColumn(#"Added Conditional Column4", "NT", each if Text.Contains([PLACE], "NT") then 1 else if Text.Contains([PLACE], "Across Canada") then 1 else if Text.Contains([PLACE], "Northwest Territories") then 1 else 0),*

*#"Added Conditional Column6" = Table.AddColumn(#"Added Conditional Column5", "NS", each if Text.Contains([PLACE], "NS") then 1 else if Text.Contains([PLACE], "Across Canada") then 1 else if Text.Contains([PLACE], "Nova Scotia") then 1 else if Text.Contains([PLACE], "Eastern Canada") then 1 else if Text.Contains([PLACE], "Maritime") then 1 else if Text.Contains([PLACE], "Martime") then 1 else 0),*

*#"Added Conditional Column7" = Table.AddColumn(#"Added Conditional Column6", "NU", each if Text.Contains([PLACE], "NU") then 1 else if Text.Contains([PLACE], "Across Canada") then 1 else if Text.Contains([PLACE], "Nunavut") then 1 else 0),*

*#"Added Conditional Column8" = Table.AddColumn(#"Added Conditional Column7", "ON", each if Text.Contains([PLACE], "ON") then 1 else if Text.Contains([PLACE], "Across Canada") then 1 else if Text.Contains([PLACE], "Ontario") then 1 else if Text.Contains([PLACE], "Eastern Canada") then 1 else if Text.Contains([PLACE], "Toronto") then 1 else if Text.Contains([PLACE], "Fort Albany") then 1 else if Text.Contains([PLACE], "National Capital Region") then 1 else 0),*

*#"Added Conditional Column9" = Table.AddColumn(#"Added Conditional Column8", "PE", each if Text.Contains([PLACE], "PE") then 1 else if Text.Contains([PLACE], "Across Canada") then 1 else if Text.Contains([PLACE], "Prince Edward Island") then 1 else if Text.Contains([PLACE], "Eastern Canada") then 1 else if Text.Contains([PLACE], "Maritime") then 1 else if Text.Contains([PLACE], "Martime") then 1 else 0),*

*#"Added Conditional Column10" = Table.AddColumn(#"Added Conditional Column9", "QC", each if Text.Contains([PLACE], "QC") then 1 else if Text.Contains([PLACE], "Across Canada") then 1 else if Text.Contains([PLACE], "Quebec") then 1 else if Text.Contains([PLACE], "Qu?bec") then 1 else if Text.Contains([PLACE], "Eastern Canada") then 1 else if Text.Contains([PLACE], "St. Lawrence River") then 1 else 0),*

*#"Added Conditional Column11" = Table.AddColumn(#"Added Conditional Column10", "SK", each if Text.Contains([PLACE], "SK") then 1 else if Text.Contains([PLACE], "Across Canada") then 1 else if Text.Contains([PLACE], "Saskatchewan") then 1 else if Text.Contains([PLACE], "Prairie") then 1 else if Text.Contains([PLACE], "Western Canada") then 1 else if Text.Contains([PLACE], "Saskatoon") then 1 else 0),*

*#"Added Conditional Column12" = Table.AddColumn(#"Added Conditional Column11", "YT", each if Text.Contains([PLACE], "YT") then 1 else if Text.Contains([PLACE], "Across Canada") then 1 else if Text.Contains([PLACE], "Yukon") then 1 else 0)*

*in*

*#"Added Conditional Column12"*

* 1. Click “Done”

1. Create a new column indicates how many provinces affected by the event (More on this later)
   1. Click the table icon and click “Add Custom Column”
      1. “New column name”: “Affected\_Province\_Count”
      2. “Custom column formula”: “=[AB]+[BC]+[MB]+[NB]+[NL]+[NT]+[NS]+[NU]+[ON]+[PE]+[QC]+[SK]+[YT]”
      3. Click “OK”
2. Remove records which do not affect Canadian provinces
   1. Click the triangle icon in the “Affected\_Province\_Count” column
   2. Uncheck “0” then click “OK”
3. Convert from wide table to long table
   1. Control click all province columns
   2. Right click any selected column and click “Unpivot columns”
   3. Remove unrelated rows:
      1. Click the triangle icon in the “Value” column
      2. Uncheck “0” then click “OK”
   4. Remove unrelated columns:
      1. Control click “PLACE” and “Value” columns
      2. Right click any select column the select “Remove columns”
   5. Double click “Attribute” and type “Affected\_Province”
4. Distribute fatality, cost, and payment to each affected provinces
   1. Since there is no provincial level information, let’s assume the impact is evenly distributed across affected provinces
   2. For example, if one event’s fatality rate is 10 people and the event affected ON and QC, we assume that there are 5 people from ON and 5 from QC
   3. Click the table icon and click “Add Custom Column”
      1. “New column name”: “Fatalities\_by\_Province”
      2. “Custom column formula”: “=[FATALITIES]/[Affected\_Province\_Count]”
      3. Click “OK”
5. Repeat step 10 to add column for each attribute
   1. Trick: Instead of adding each column separately, we can copy the code in the formula bar, paste then change the keyword
   2. under “Home”, double click on “Advanced Editor”
   3. Append the following code to the Advanced Editor

*#"Added Custom1" = Table.AddColumn(#"Removed Columns1", "Fatalities\_by\_Province", each [FATALITIES]/[Affected\_Province\_Count]),*

*#"Added Custom2" = Table.AddColumn(#"Added Custom1", "Injured\_Infected\_by\_Province", each [#"INJURED / INFECTED"]/[Affected\_Province\_Count]),*

*#"Added Custom3" = Table.AddColumn(#"Added Custom2", "Evacuated\_by\_Province", each [EVACUATED]/[Affected\_Province\_Count]),*

*#"Added Custom4" = Table.AddColumn(#"Added Custom3", "Estimated\_Cost\_by\_Province", each [ESTIMATED TOTAL COST]/[Affected\_Province\_Count]),*

*#"Added Custom5" = Table.AddColumn(#"Added Custom4", "Normalized\_Cost\_by\_Province", each [NORMALIZED TOTAL COST]/[Affected\_Province\_Count]),*

*#"Added Custom6" = Table.AddColumn(#"Added Custom5", "Federal\_DFAA\_Payments\_by\_Province", each [FEDERAL DFAA PAYMENTS]/[Affected\_Province\_Count]),*

*#"Added Custom7" = Table.AddColumn(#"Added Custom6", "Provincial\_DFAA\_Payments\_by\_Province", each [PROVINCIAL DFAA PAYMENTS]/[Affected\_Province\_Count]),*

*#"Added Custom8" = Table.AddColumn(#"Added Custom7", "Provincial\_Department\_Payments\_by\_Province", each [PROVINCIAL DEPARTMENT PAYMENTS]/[Affected\_Province\_Count]),*

*#"Added Custom9" = Table.AddColumn(#"Added Custom8", "Municipal\_Cost\_by\_Province", each [MUNICIPAL COSTS]/[Affected\_Province\_Count]),*

*#"Added Custom10" = Table.AddColumn(#"Added Custom9", "OGD\_Cost\_by\_Province", each [OGD COSTS]/[Affected\_Province\_Count]),*

*#"Added Custom11" = Table.AddColumn(#"Added Custom10", "Insurance\_Payments\_by\_Province", each [INSURANCE PAYMENTS]/[Affected\_Province\_Count]),*

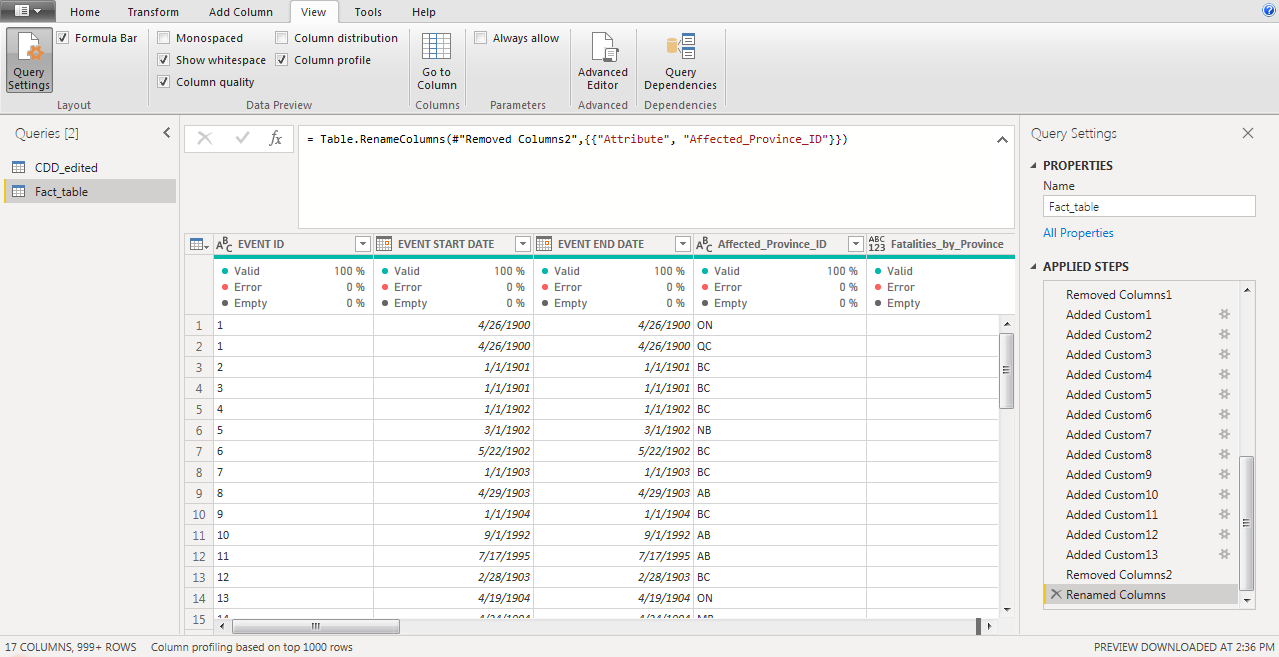
*#"Added Custom12" = Table.AddColumn(#"Added Custom11", "NGO\_Payments\_by\_Province", each [NGO PAYMENTS]/[Affected\_Province\_Count]),*

*#"Added Custom13" = Table.AddColumn(#"Added Custom12", "Utility\_People\_Affected\_by\_Province", each [#"UTILITY - PEOPLE AFFECTED"]/[Affected\_Province\_Count])*

*in*

*#"Added Custom13"*

1. Remove unrelated columns
   1. "FATALITIES"
   2. "INJURED / INFECTED"
   3. "EVACUATED"
   4. "ESTIMATED TOTAL COST"
   5. "NORMALIZED TOTAL COST"
   6. "FEDERAL DFAA PAYMENTS"
   7. "PROVINCIAL DFAA PAYMENTS"
   8. "PROVINCIAL DEPARTMENT PAYMENTS"
   9. "MUNICIPAL COSTS"
   10. "OGD COSTS"
   11. "INSURANCE PAYMENTS"
   12. "NGO PAYMENTS”
   13. "UTILITY - PEOPLE AFFECTED"
   14. "Affected\_Province\_Count"
2. Double click on “Attribute” and type “Affected\_Province\_ID”



# Task 5: Create Dimension Tables

1. Create Lookup Table for Disaster
   1. Under “Home”, click “New Source”, then click “Blank Query”
   2. In the formula bar, type “= CDD\_edited”
   3. Double click on “Query1” on the left Queries window, type “Lookup\_disaster”
   4. Control +click event related columns, right click column name, then click “Remove Columns”
      1. "PLACE"
      2. "FATALITIES"
      3. "INJURED / INFECTED"
      4. "EVACUATED"
      5. "ESTIMATED TOTAL COST"
      6. "NORMALIZED TOTAL COST”
      7. "FEDERAL DFAA PAYMENTS"
      8. "PROVINCIAL DFAA PAYMENTS"
      9. "PROVINCIAL DEPARTMENT PAYMENTS"
      10. "MUNICIPAL COSTS"
      11. "OGD COSTS"
      12. "INSURANCE PAYMENTS"
      13. "NGO PAYMENTS"
      14. "UTILITY - PEOPLE AFFECTED"
      15. "EVENT START DATE"
      16. "EVENT END DATE"
2. Create Lookup table for Province
   1. Under “Home”, click “New Source”, then click “Web”
   2. In the URL bar, type <https://www.ncbi.nlm.nih.gov/books/NBK7254/>
      1. In the pop-up window, check “table 1” and click “OK”
   3. Right click table icon, click “Use First Row as Header”
   4. Double click on “Table1” on the left Queries window, type “Lookup\_province”
   5. Double click on column name to rename as “Province Name” and “Province ID”
   6. Reorder the column so that “Province ID” is on the left

# Task 6: Create dimension table for date

1. Create Lookup table for Date
   1. First click “Close & Apply” to save all the changes so far and exit Power Query Editor
   2. In the Power BI Desktop interface, under “Modeling” tab, click “New table”
      1. In the formula bar, copy and paste the following code, then click check mark

*DimDate =*

*ADDCOLUMNS (*

*CALENDAR (*

*DATE ( YEAR ( MIN ( CDD\_edited[EVENT START DATE] ) ), 1, 1 ),*

*DATE ( YEAR ( MAX ( CDD\_edited[EVENT START DATE] ) ), 12, 31 )*

*),*

*"DateAsInteger", FORMAT ( [Date], "YYYYMMDD" ),*

*"Year", YEAR ( [Date] ),*

*"MonthNo", FORMAT ( [Date], "MM" ),*

*"YearMonthNo", FORMAT ( [Date], "YYYY/MM" ),*

*"YearMonth", FORMAT ( [Date], "YYYY/mmm" ),*

*"MonthShort", FORMAT ( [Date], "mmm" ),*

*"MonthLong", FORMAT ( [Date], "mmmm" ),*

*// "WeekNo", WEEKDAY ( [Date] ),*

*// "WeekDay", FORMAT ( [Date], "dddd" ),*

*// "WeekDayShort", FORMAT ( [Date], "ddd" ),*

*"Quarter", "Q" & FORMAT ( [Date], "Q" ),*

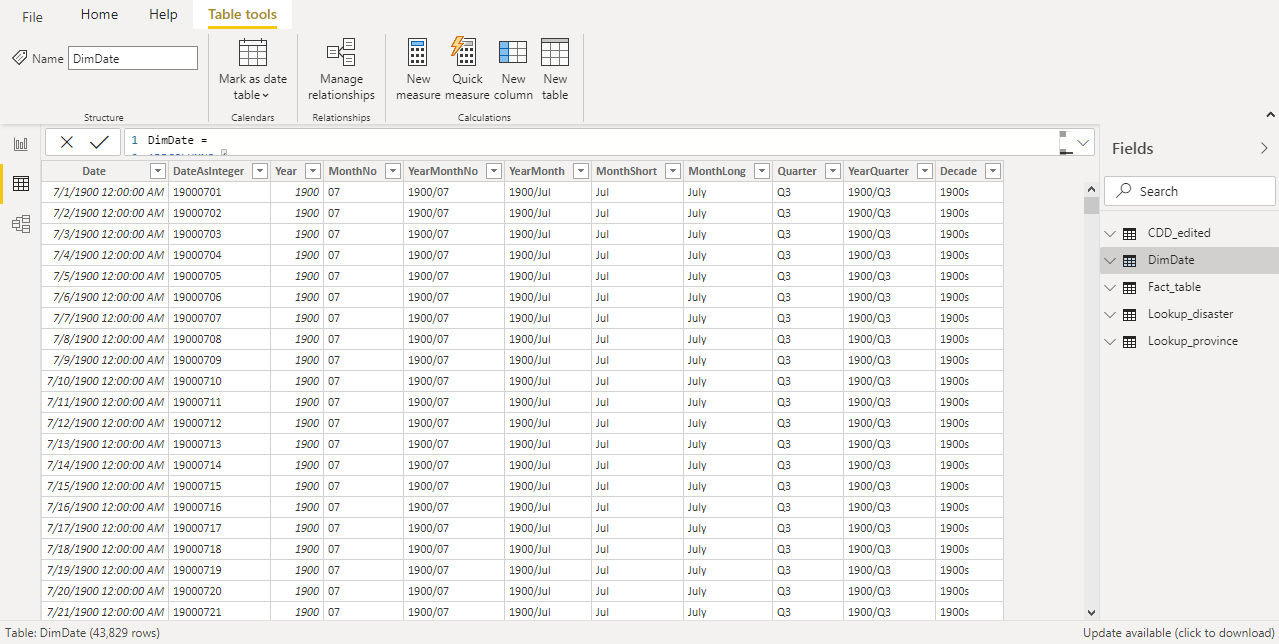
*"YearQuarter",*

*FORMAT ( [Date], "YYYY" ) & "/Q"*

*& FORMAT ( [Date], "Q" ),*

*"Decade", ROUNDDOWN( YEAR( [Date] )/10, 0 ) & "0s"*

*)*



# Task 7: Edit Table Relationship and Create Data Modeling

1. Go to the 3rd icon (Model“) on the left bar
2. Remove all the existing linkages among tables by right click on the link, then click “Delete”
3. Move all the Fact/Dimension/Lookup table above the fact table, while keep CDD\_edited table on the side
4. Create table relationship by dragging the common column from one table to another (direction does not matter)
   1. EVENT ID in Lookup\_disaster – EVENT ID in Fact\_table
   2. Province ID in Lookup\_province – Affected\_Province\_ID in Fact\_table
   3. Date in DimDate – EVENT START DATE in Fact\_table

