

Data modeling and table properties

DATA MODELING IN POWER BI

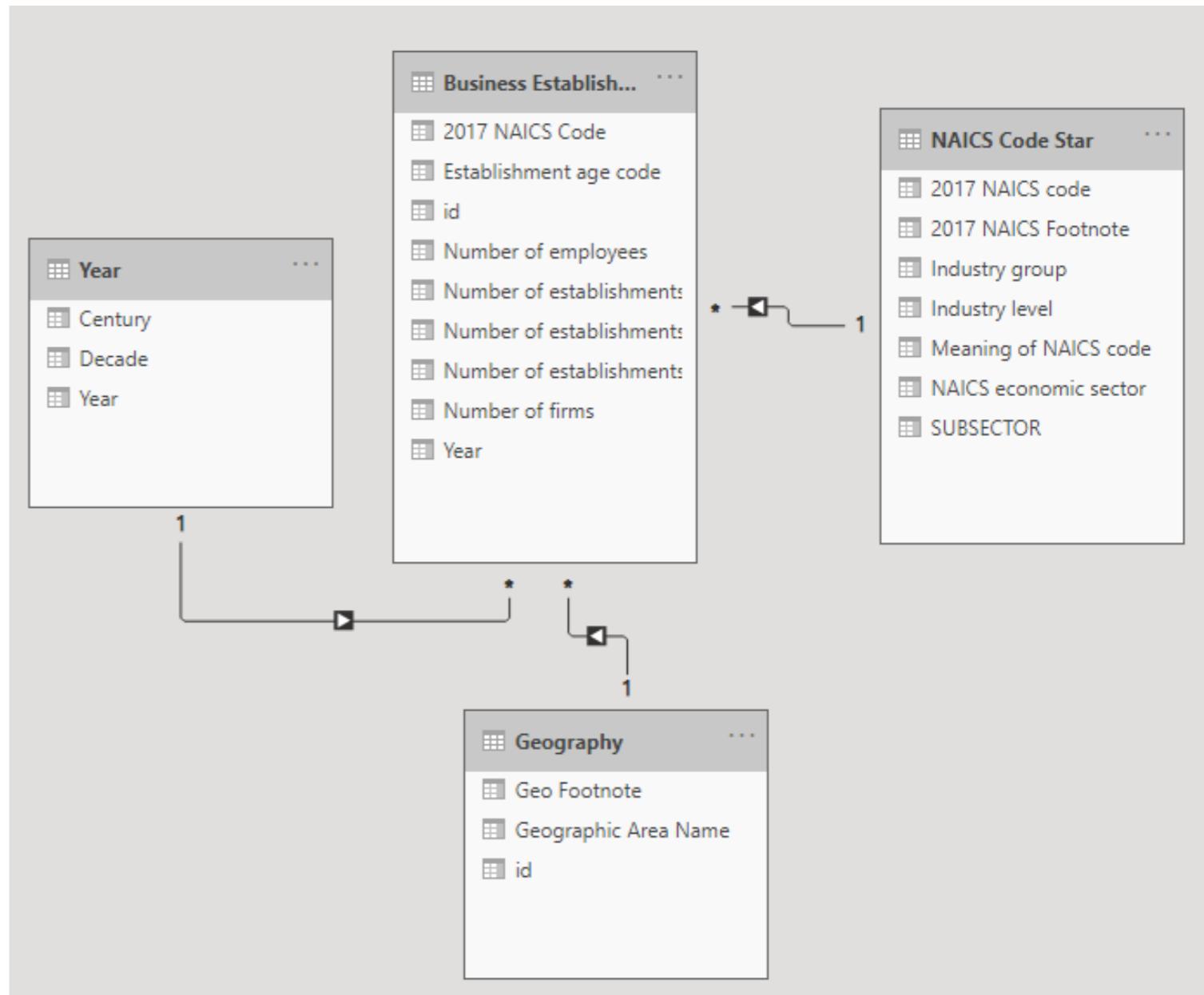


Maarten Van den Broeck

Content Developer at DataCamp

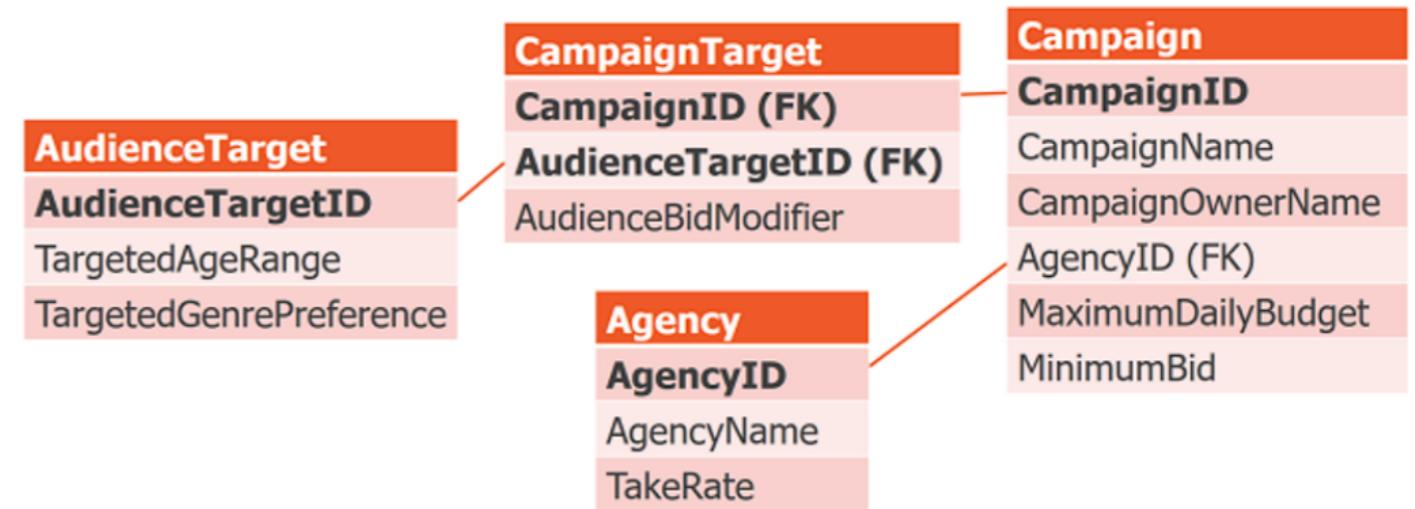
What is a data model?

- Conceptual view of data elements
- Typically a visual representation
- Data models include:
 - Tables
 - Columns
 - Relationships between tables
 - Data types
 - Keys



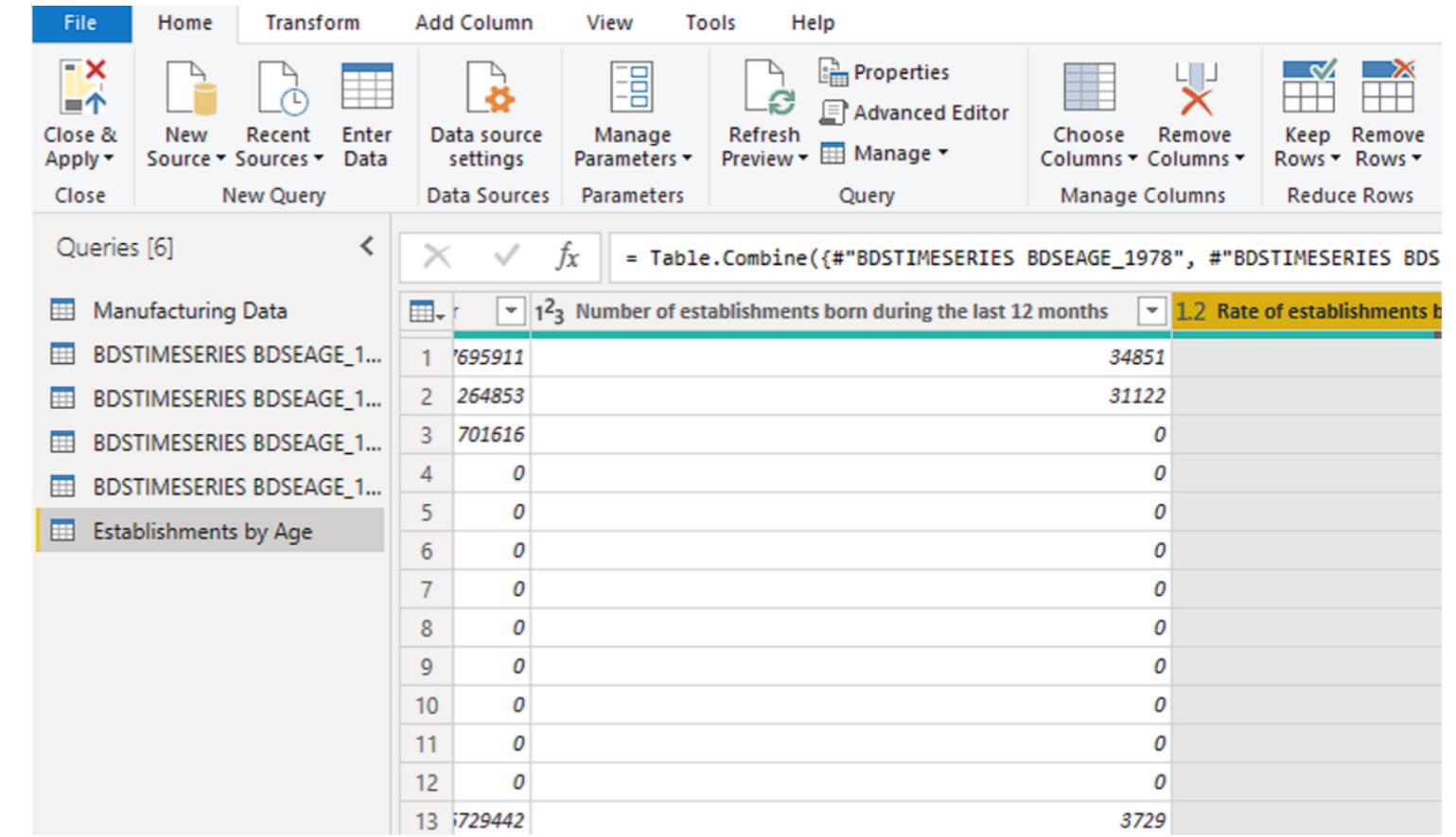
Data modeling

- The process of creating a data model
- Why model data?
 - Data \neq perfect
 - Reshape data for analysis
 - Compress data usage
 - Easier to understand model



Data modeling in Power BI & Power Query

- Power Query is the data preparation tool of different Microsoft products, including Power BI
- Main goals:
 - Manage queries
 - Data modeling
- Data modeling: 80% in Power Query, 20% in Power BI



The screenshot shows the Microsoft Power Query Editor window. The ribbon at the top includes File, Home, Transform, Add Column, View, Tools, and Help. The Home tab is selected, displaying icons for Close & Apply, New Source, Recent Sources, Enter Data, Data source settings, Manage Parameters, Refresh Preview, Properties, Advanced Editor, and Manage. Below the ribbon is a 'Queries [6]' pane listing six items: Manufacturing Data, BDSTIMESERIES BDSEAGE_1..., BDSTIMESERIES BDSEAGE_1..., BDSTIMESERIES BDSEAGE_1..., BDSTIMESERIES BDSEAGE_1..., and Establishments by Age. The 'Establishments by Age' query is currently selected. To its right is a preview grid showing 13 rows of data. The first column contains integers from 1 to 13. The second column contains values: 1695911, 264853, 701616, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, and 1729442. The third column contains values: 34851, 31122, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, and 3729. A formula bar at the top of the preview area shows the formula: = Table.Combine({#"BDSTIMESERIES BDSEAGE_1978", #"BDSTIMESERIES BDS}).

1	1695911	34851
2	264853	31122
3	701616	0
4	0	0
5	0	0
6	0	0
7	0	0
8	0	0
9	0	0
10	0	0
11	0	0
12	0	0
13	1729442	3729

Columns and row management

Operation

- Keep or remove specific columns
- Keep or remove specific rows
- Split a single column in multiple columns
- Summarize/group rows in a table by the contents of a column

Example

- Remove empty column
- Keep top row as header
- DD/MM/YYYY column split in DD, MM, YYYY columns
- Sum or median of all rows

Data types

- Choosing the right data type is essential:
 - Constrain data to a specific shape
 - Optimize storage
 - Enable specific functionality
- Power Query infers data type on first few hundred rows

1.2	Decimal Number
\$	Fixed decimal number
1 ² ₃	Whole Number
%	Percentage
 	Date/Time
 	Date
 	Time
 	Date/Time/Timezone
 	Duration
A ^B _C	Text
  	True/False
 	Binary
Using Locale...	

Rounding

Power Query

- Actually *changes* the data, not just formatting
- Typically not the right answer

Round

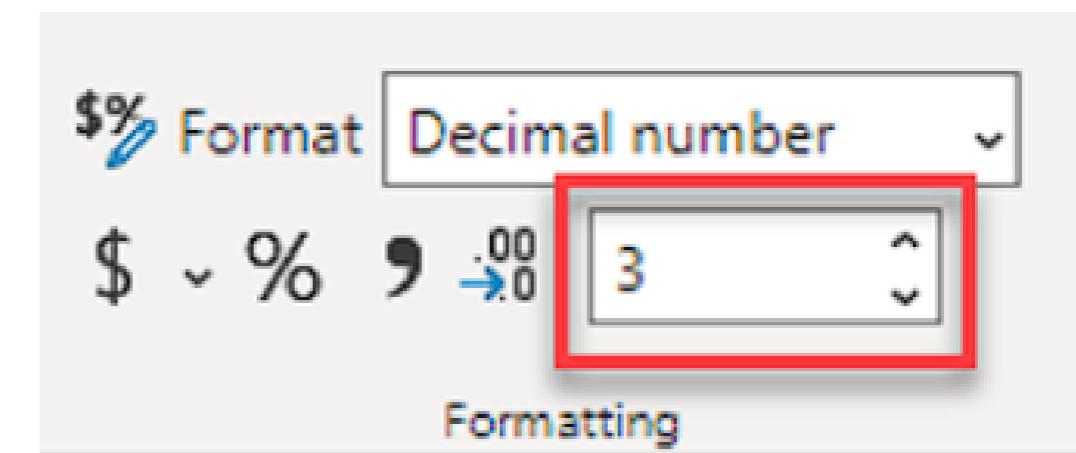
Specify how many decimal places to round to.

Decimal Places

2

Power BI

- Changes how the data *appears*, not how it's stored
- Generally a better answer than rounding in Power Query



The dataset

- United States Census Bureau survey data of manufacturers
- Summary statistics for manufacturing firms
- North American Industry Classification System (NAICS)

The screenshot shows the United States Census Bureau website interface. The top navigation bar includes links for 'ALL', **TABLES**, 'MAPS', and 'PAGES'. A search bar contains the query 'ASMAREA2017.AM1831BASIC01'. Below the search bar, the title 'Annual Survey of Manufactures: Summary Statistics for Industry Groups and Industries in the U.S.: 2019 and 2018' is displayed, along with the 'Survey/Program: Annual Economic Surveys' and 'Dataset: ASMAREA2017' information. The 'TableID: AM1831BASIC01' is also mentioned. The main content area features a table with the following data:

Geographic Area Name	2017 NAICS code	Meaning of NAICS code	Year	Sales, value of shipments, ...	Relative sta
United States	31-33	Manufacturing	2018	5,890,662,939	
United States	31-33	Manufacturing	2019	5,731,187,355	
United States	311	Food manufacturing	2018	784,628,928	
United States	311	Food manufacturing	2019	794,130,354	
United States	3111	Animal food manufacturing	2018	57,501,604	

Let's practice!

DATA MODELING IN POWER BI

Load and transform data

DATA MODELING IN POWER BI



Maarten Van den Broeck
Content Developer at DataCamp

Let's practice!

DATA MODELING IN POWER BI