

# Data modeling and table properties

DATA MODELING IN POWER BI

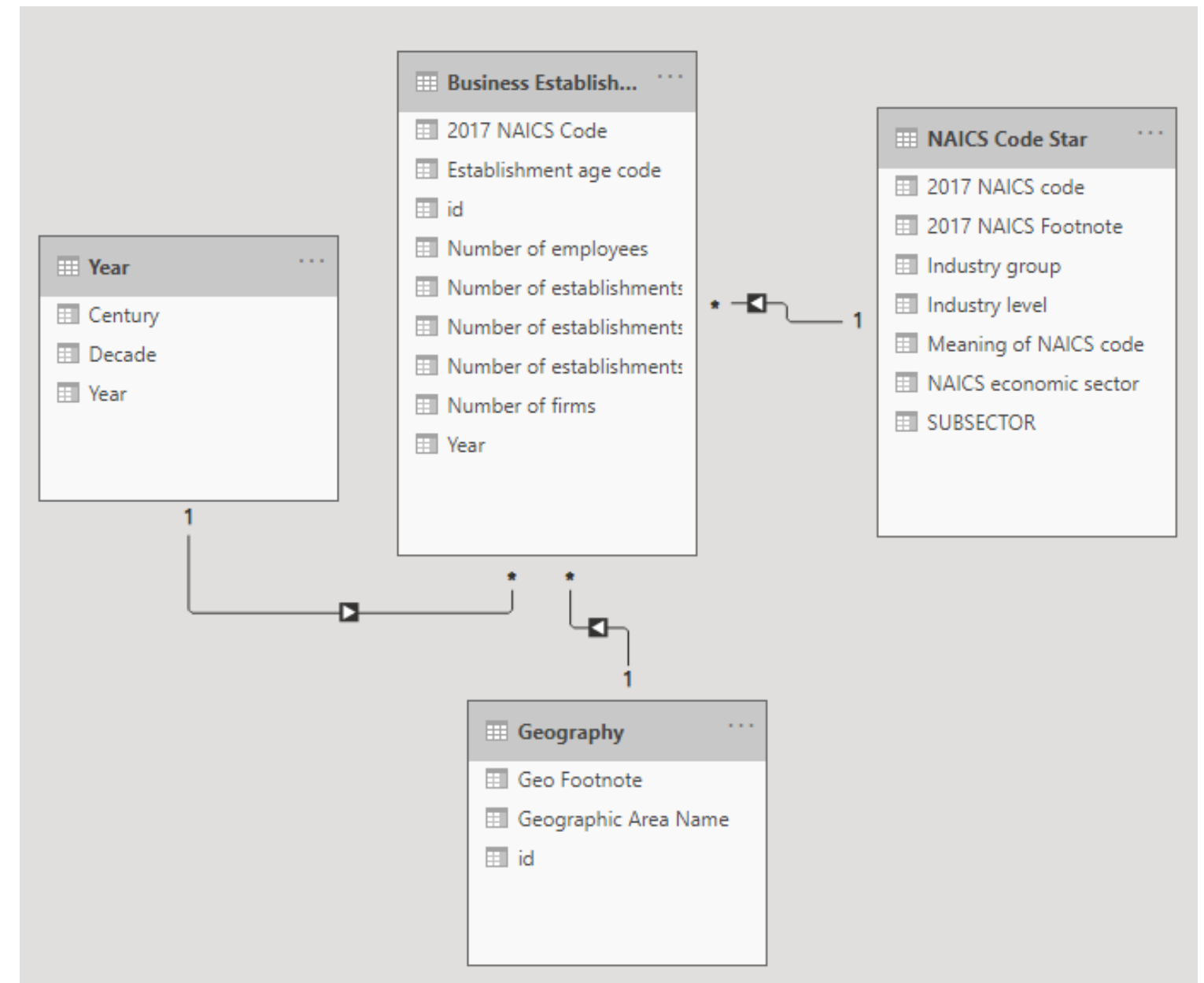


**Maarten Van den Broeck**

Content Developer at DataCamp

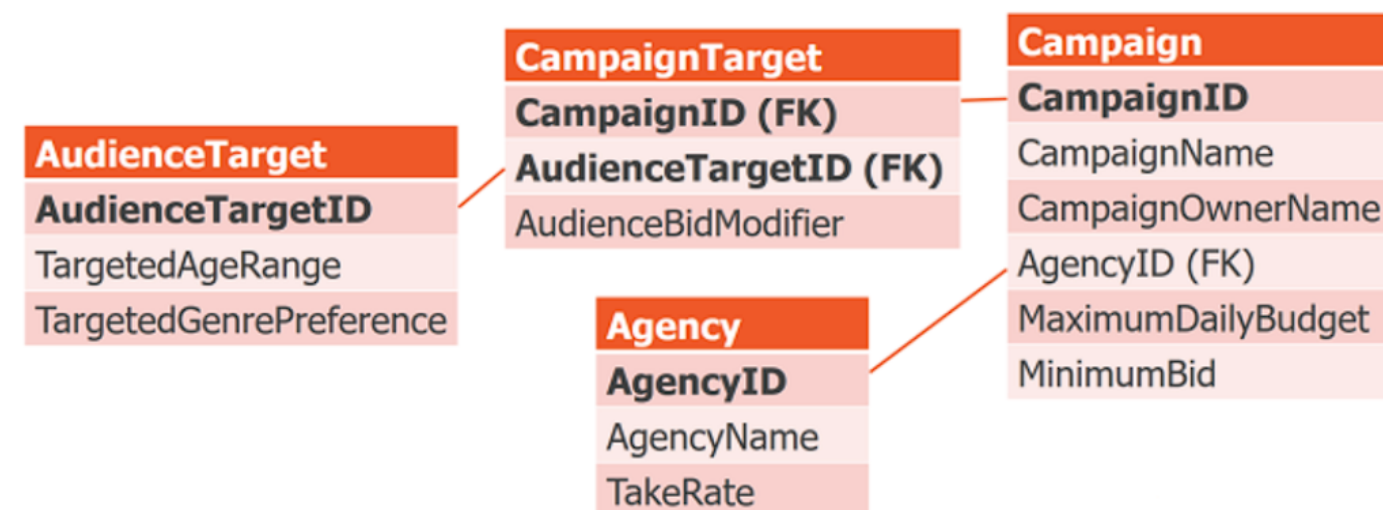
# What is a data model?

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



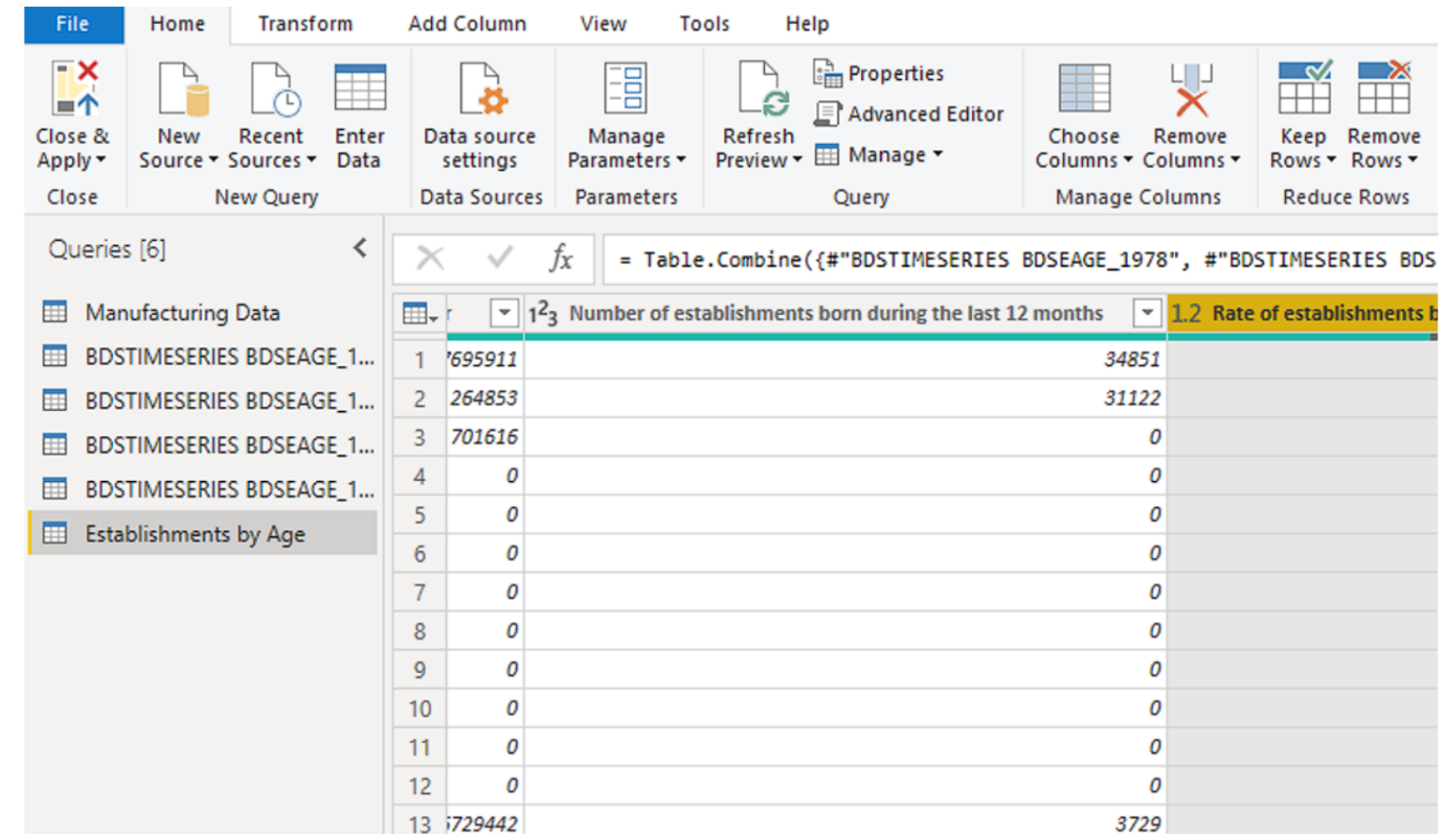
# Data modeling

- 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 displays the Power Query Editor interface. The ribbon at the top includes tabs for File, Home, Transform, Add Column, View, Tools, and Help. The Transform tab is active, showing options like Data source settings, Manage Parameters, Refresh Preview, and Query. The left pane shows a list of queries, with 'Establishments by Age' selected. The main area shows a data table with columns for establishment count and rate.

	1.2.3	Number of establishments born during the last 12 months	1.2	Rate of establishments b
1	695911	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	729442	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 <sup>2</sup> <sub>3</sub>	Whole Number
%	Percentage
	Date/Time
	Date
	Time
	Date/Time/Timezone
	Duration
A <sup>B</sup> <sub>C</sub>	Text
	True/False
	Binary
Using Locale...	

# Rounding

## Power Query

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

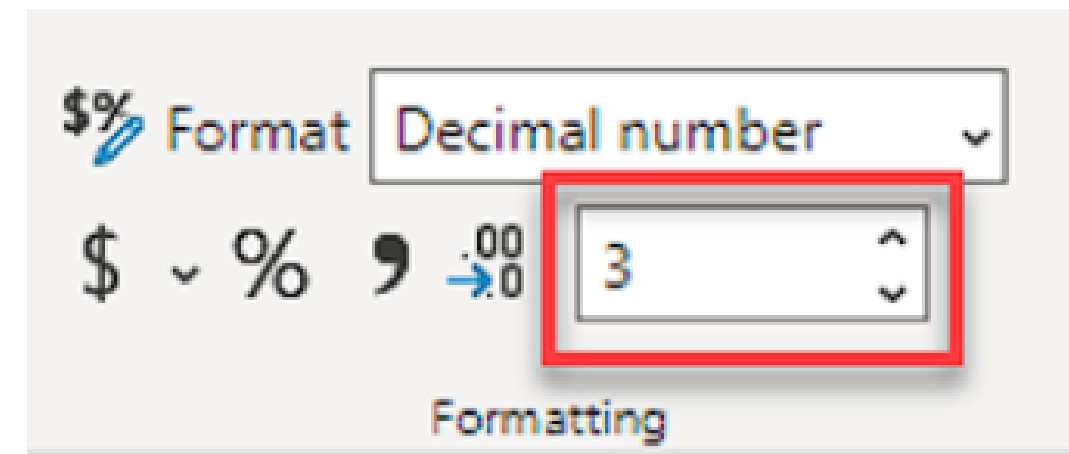
### Round

Specify how many decimal places to round to.

Decimal Places


## 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)

		Q ASMAREA2017.AM1831BASIC01				
ALL TABLES MAPS PAGES		<b>Annual Survey of Manufactures: Summary Statistics for Industry Groups and Industries in the U.S.: 2019 and 2018</b>				
1 Results FILTER   DOWNLOAD		Survey/Program: Annual Economic Surveys TableID: AM1831BASIC01 Dataset: ASMAREA2017				
<b>Annual Survey of Manufactures: Summary Statistics for Industry Groups and Industries in the U.S.: 2019 and 2018</b> Survey/Program: Annual Economic Surveys Table: AM1831BASIC01		Geographic Area Name	2017 NAICS code	Meaning of NAICS code	Year	Sales, value of shipments, ...
		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