Data modeling and table properties

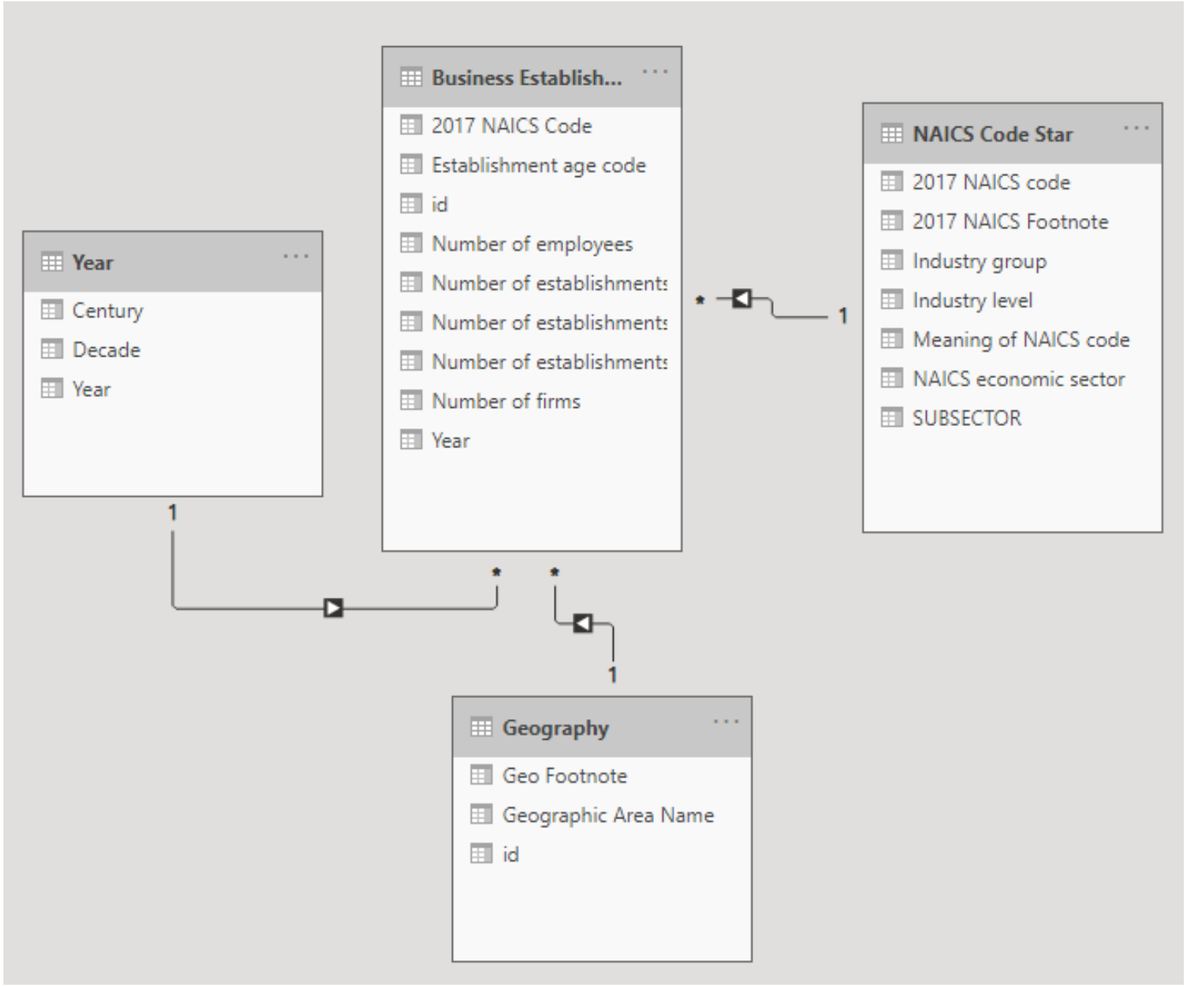
DATA MODELING IN POWER BI

Maarten Van den Broeck
Content Developer at DataCamp

datacamp

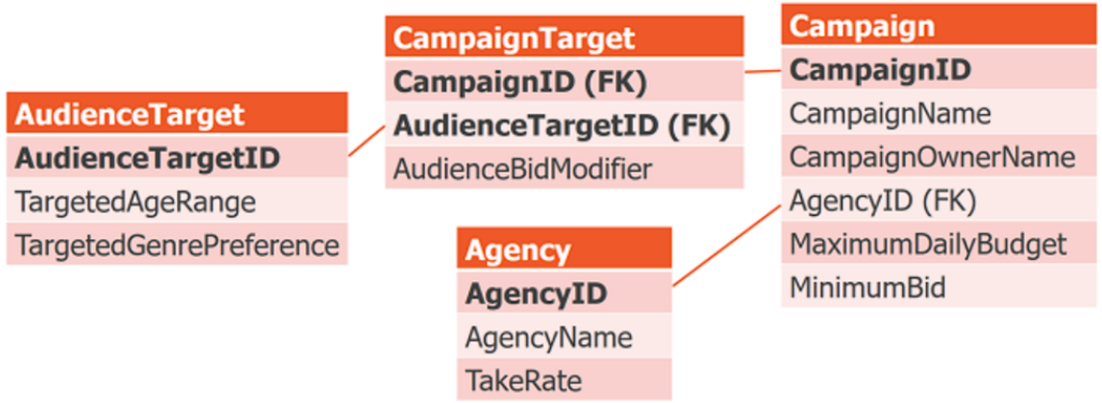
What is a data model?

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



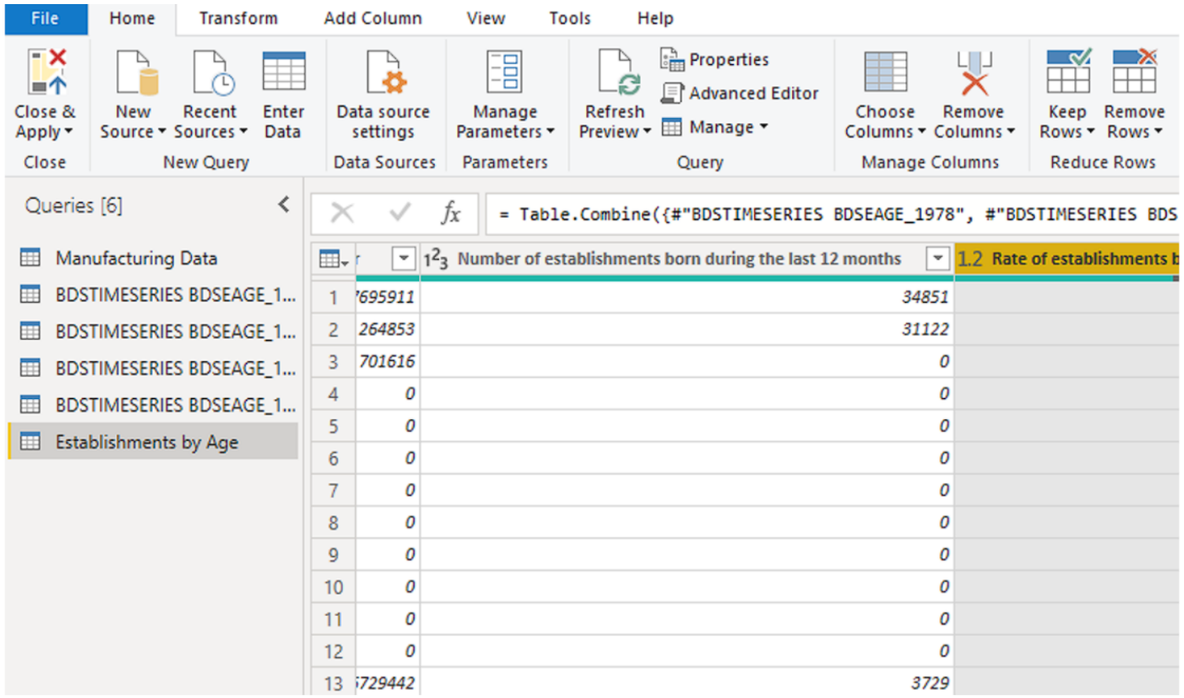
Data modeling

- 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



Columns and row management

Operation

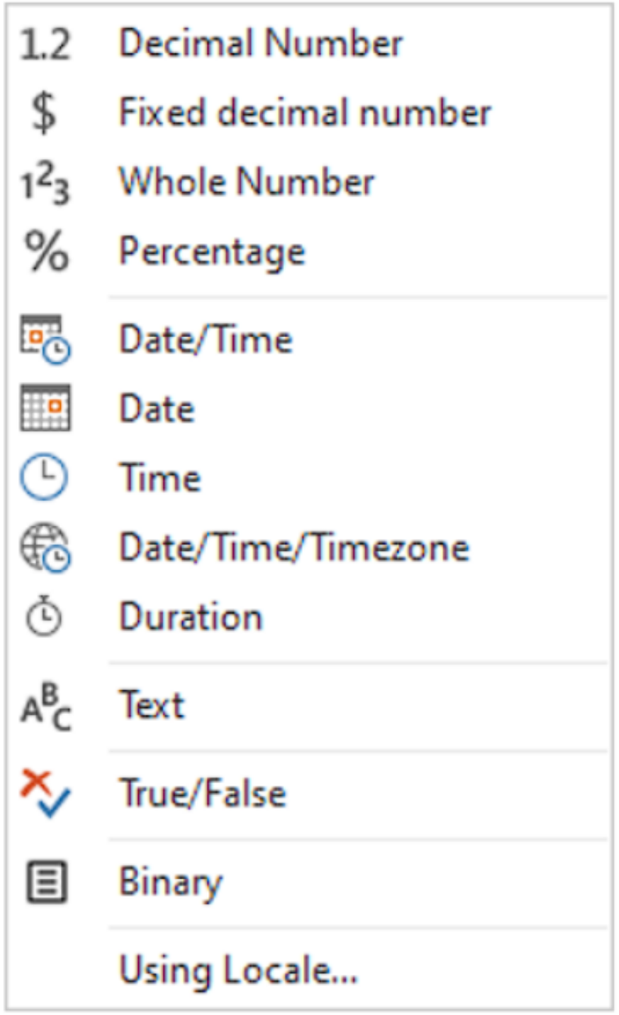
- Keep or remove specific columns
- Keep or remove specific rows
- Split a single column in multiple columns
- Summarize/group rows in a table by the contents of a column

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

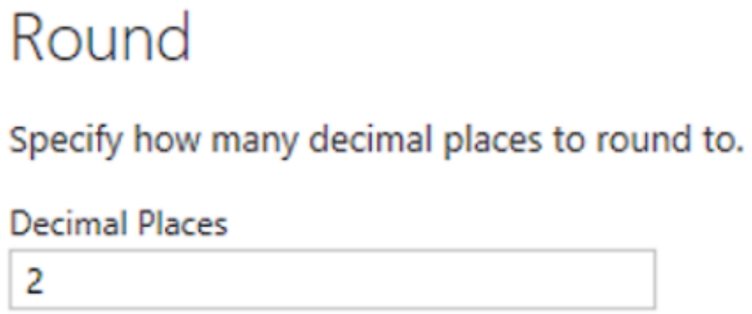
- Choosing the right data type is essential:
  - Constrain data to a specific shape
  - Optimize storage
  - Enable specific functionality
- Power Query infers data type on first few hundred rows



Rounding

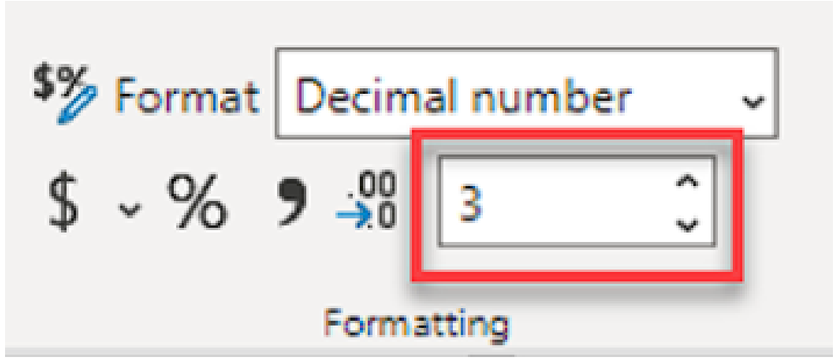
Power Query

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



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)

United States Census Bureau

ASMAREA2017.AM1831BASIC01

ALL | TABLES | MAPS | PAGES

1 Results FILTER | DOWNLOAD

Annual Survey of Manufactures: Summary Statistics for Industry Groups and Industries in the U.S.: 2019 and 2018

Survey/Program: Annual Economic Surveys

Table: AM1831BASIC01

Annual Survey of Manufactures: Summary Statistics for Industry Groups and Industries in the U.S.: 2019 and 2018

Survey/Program: Annual Economic Surveys **TableID:** AM1831BASIC01

Dataset: ASMAREA2017

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

datacamp

Let's practice!

DATA MODELING IN POWER BI