Lab 4: Star and Snowflake schemas

Part 1: Star schemas

Create a new page and name it "Industry insights" and navigate to the *Data* view to get familiar with the columns in the "Industry" dimension.

In the *Report* view, add a table visualization with the following columns from the "Industry" dimension: `Industry group`, `NAICS Code description`, `Sector`, and `Subsector`.

Note that the columns in the Industry table make up a hierarchy. Starting from the highest level, each category gives a more granular view of what industry the establishments operate in.

Put the columns in the right order, going from less to more detail.

How is the Industry hierarchy structured?

- ○ Industry group > NAICS Code description > Sector > Subsector
- ○ Sector > Subsector > Industry group > NAICS Code description
- ○ Sector > Subsector > NAICS Code description > Industry group

Part 2: Snowflake schemas

- Go to Power Query and duplicate the Industry queries four times.
- Rename the duplicated queries to "NAICS code", "Industry group", "Subsector", and "Sector".

Knowing the hierarchy as: `Sector` > `Subsector` > `Industry group` > `NAICS Code description`.

- For each table, keep only the code and text description columns corresponding to that level of the hierarchy, and also keep the code of the next level in the hierarchy,if there is one.
- Make sure to remove duplicate values.

- Close Power Query and go to the *Model* view.
- Delete the automatically created relationships between the "Industry" star dimension and the four tables you just created.

If for some reason a new table is not showing up in the Model view, you can manually add a relationship using the Manage relationships icon in the Home menu. Click "New..." and select the tables and columns where you want to define the relationship.

- Create a relationship between "NAICS code" and "Establishment Survey" tables, using `NAICS code`.
- You can zoom in and out on the data model by using the `+` and `-` in the bottom right corner of the Power BI window.

If for some reason a new table is not showing up in the Model view, you can manually add a relationship using the Manage relationships icon in the Home menu. Click "New..." and select the tables and columns where you want to define the relationship.

Check that the other snowflake dimensions are connected to each other in the right order. If there are dotted-line relationships, double-click on the relationship to go into the *Properties* and select "Make this relationship active."

Find the Hierarchical level table that contains "Food Manufacturing" and tell how many rows would you have to change if the name of the subsector changes.

Part 3: The performance analyzer

- Go to the *Report* view and create a new page named "Comparison".
- Create a line chart with the decade on the x-axis, the **average** number of employees on the y-axis, and the subsector (from the *star* dimension "Industry") on the legend.

Create another line chart with the decade on the x-axis, the **average** number of employees on the y-axis, and the subsector (from the *snowflake* dimension "Subsector") on the legend.

- Change the title of the first line chart to "Star schema".
- Change the title of the second line chart to "Snowflake schema".

- Open the *Performance analyzer* and start recording to begin performance analysis.
- Refresh all visuals on the page. Do this several times to see if there is a consensus on whether the star or snowflake schema is faster in this case.

- Looks like there isn't a big difference in speed between both approaches for our relatively small dataset.
- Dig a level deeper and go take a look at what the drivers of performance are for each visual.

What takes the most amount of time for both visuals?

- ○ DAX query
- ○ Visual display
- ○ Other