# Lab 3: Dimensional Modeling

## Part 1: Splitting data into facts and dimensions

Open the Power BI Desktop file named `3_1_split_dataset.pbix`, from the `Exercises` folder on the desktop, and go to Power Query.

Duplicate the Business "Establishment Survey" query and rename it to "Industry".

- Keep only the following columns in the "Industry" table: `NAICS code`, `NAICS Code Description`, `Industry group code`, `Industry group`, `Subsector Code`, `Subsector`, `Sector Code`, and `Sector`.
- Remove duplicate values from the dataset.

- Return to the "Establishment Survey" query and remove the columns you just added to "Industry": `NAICS Code Description`, `Industry group code`, `Industry group`, `Subsector Code`, `Subsector`, `Sector Code`, and `Sector`.
- Make sure to keep `NAICS code`! This column is the key we'll use to connect the dimension to the fact table. Exit *Power Query*.

Go to the *Model* view to verify that there is a relationship between "Industry" and "Establishment Survey".

*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.*

Go to the *Report* view and create a new clustered bar chart visual with the `Subsector` column from the dimension table as the *Axis* and **Average** `Number of employees` from the fact table as the *Values*.

**Which subsector counted the highest average number of employees?**

## Part 2: Load a new dimension

Import the file named `Time.txt` from the `Datasets` folder on the desktop.

Go to the *Model* view and verify that there is a relationship between "Time" and "Establishment Survey" tables, defined by the `Year` column.

*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.*

Return to the *Report* view and add a new slicer visual, slicing on `Decade` from the Time dimension.

Adjust the size of the slicer in the page until you see the options and change the slicer to be a list of values rather than a range.

**How many employees did the Food Manufacturing subsector count on average during the 90s?**

## Part 3: Create another dimension

Import the file named `EstablishmentAge.csv` from the `Datasets` folder on the desktop.

Go to Power Query and name the table "Age".

Remove duplicate values from the dataset and then close Power Query.

Navigate to the *Model* view and connect the Age dimension to the Establishment Survey fact by `Age code` if this is not the case yet.

*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.*

Return to the *Report* view and add a second slicer using `Establishment Age` from the new Age dimension.

**How many average employees did 3-year old firms in the Food Manufacturing subsector have during the 90s?**