Lab 4: Identifying Performance Problems

Part 1: Data types and calculations

Open the Power BI Desktop file named `4_1_data_types.pbix` , from the Exercises folder on the desktop, and go to Power Query.

- Open *Power Query* and go to the "Business Establishments" table.
- Duplicate the `Number of employees` column and rename it to `Number of employees (text)` .

Change the data type of `Number of employees (text)` to "Text" and close the *Power Query* editor.

Create a new measure in "Business Establishments" named `Sum Employees (text)` that converts the text to a numeric value and then sums those values. The DAX formula to do that is:
```
SUMX('Business Establishments', VALUE('Business
Establishments'[Number of employees (text)]))
```

Create two table visuals:

- The first table visual should include `Rowguid` and `Number of employees` from "Business Establishments".
- The second visual should include `Rowguid` and `Sum Employees (text)` from "Business Establishments".

Open the *Performance analyzer* and refresh the visuals two or three times, focusing on the DAX query component.

Which version of the "Number of employees" column is faster to process?

○ Numeric

○ Text

Part 2: Calculated versus computed columns

Open the Power BI Desktop file named `4_2_custom_columns.pbix` , from the Exercises folder on the desktop, and go to Power Query.

- Rename the `Summary Statistics for Manufacturing BIG` table to `Summary Statistics (computed column)` .
- Make a duplicate of the previous table and rename it as `Summary Statistics (calculated column)` .

Go to `Summary Statistics (computed column)` and merge `GeographySummary` , `GeographyVariant` , and `GeographyNation` together into a column called `id` . Exit *Power Query.*

Add a new **id column** in `Summary Statistics (calculated column)` concatenating `GeographySummary` , `GeographyVariant` , and `GeographyNation` columns.

Navigate to the *Model* view and create a relationship from the `Geography` table to each of the Summary Statistics fact tables, using `id` .

If for some reason the 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 verify if the aggregation method of `Year` in both fact tables is set to "Don't summarize". Change if necessary.

Create two table visuals:

- The first table should include `id` , `Year` , and `Number of employees` from `Summary Statistics (computed column)` .
- The second table should include `id` , `Year` , and `Number of employees` from `Summary Statistics (calculated column)` .

Open the Performance Analyzer and refresh visuals two or three times.

Did the table with a reference to a calculated column load faster, slower, or the same speed as the table whose geography ID was created in Power Query?

○ Faster

○ Slower

○ Same

Part 3: An alternative to bi-directional filtering

Open the Power BI Desktop file named `4_3_bidirectional.pbix` , from the Exercises folder on the desktop.

Navigate to the *Model* view and note that there is a relationship between the `Summary Statistics` fact table and the `Year` and `Geography` dimensions.

- Return to the *Report* view and add a new measure in `Summary Statistics` and call it `Slicer Summary Stats` .
- Your should use `INT()` wrapped around `NOT ISEMPTY()` to return `1` if there is a value present in the "Summary Statistics" fact table and `0` if there isn't.

- Create a slicer on `Year` from the `Year` dimension. Change the slicer to display a list of values.
- Create another slicer on `Geographic Area Name` from the `Geography` dimension.

Add a *Card* visual with the `Number of employees` from the fact table.

Filter each slicer by the rows where `Slicer Summary Stats` is equal to 1 (meaning that there are rows present in the fact table).

How many years of data exists in the "Summary
Statistics" fact?