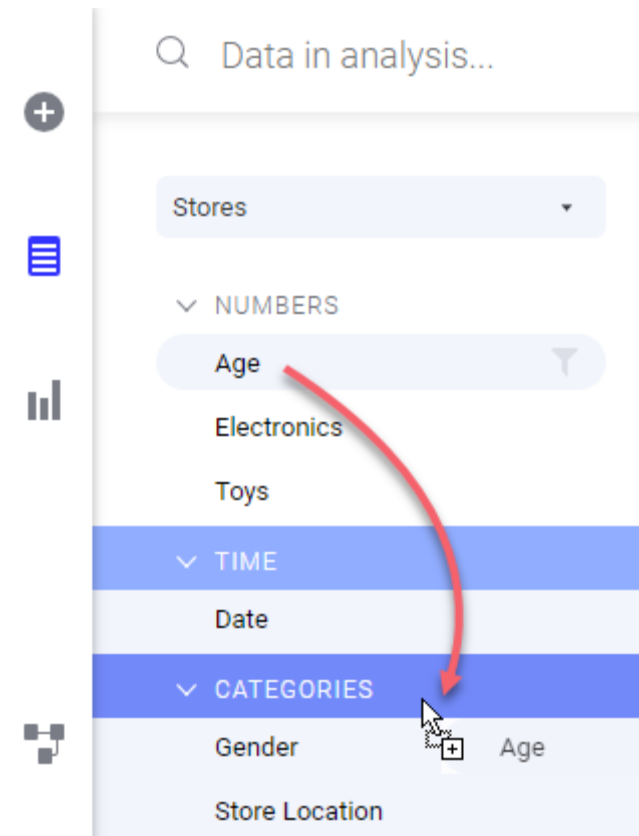


Handling Data

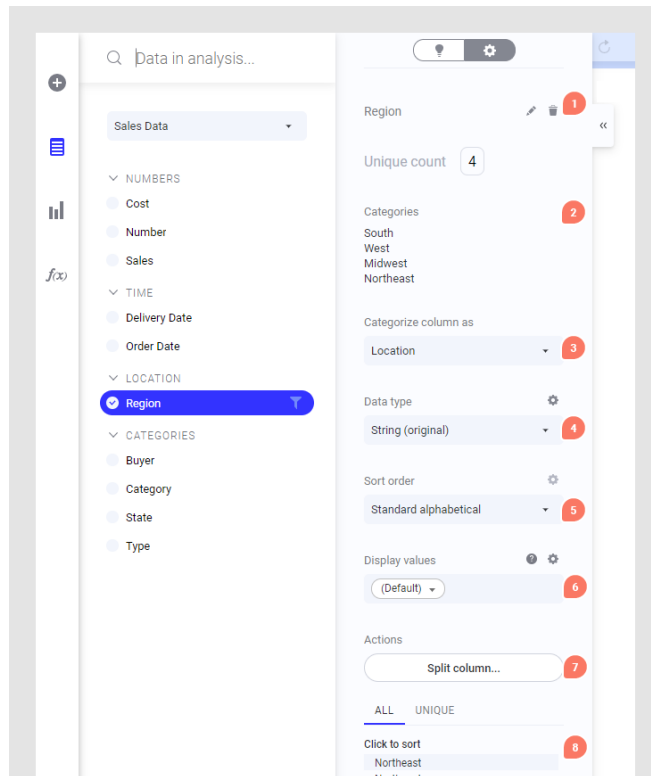
Data Types

- The category of a column affects what happens when you drag the column to a column selector or to a drop target in a visualization.
- It also affects which kinds of visualizations are advised in recommendations.
- If the suggested category for a column is not your preferred category, you can change the categorization.
- The different categories are Numbers, Currency, Time, Location, Categories, Identifiers, Text, Images, and Binary.

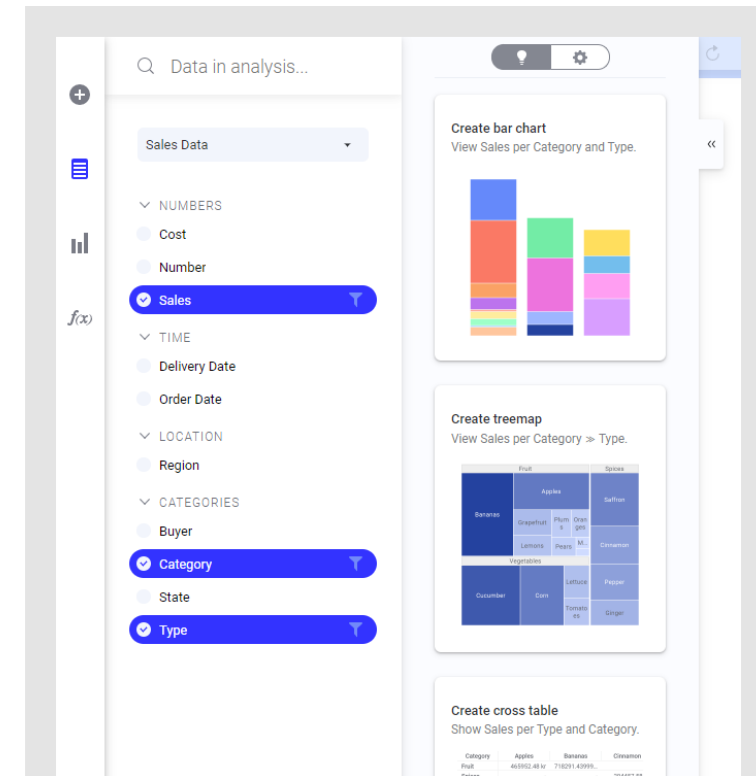


Data in Analysis Flyout

Details on Selected Column



Recommendations from selected columns



Data in Analysis Flyout – Splitting Columns

Split Column

Split using this separator
@

Number of new columns
2

Number of characters to remove from original string before splitting

From beginning
0

From end
0

☒ Split from beginning of string
☐ Split from end of string

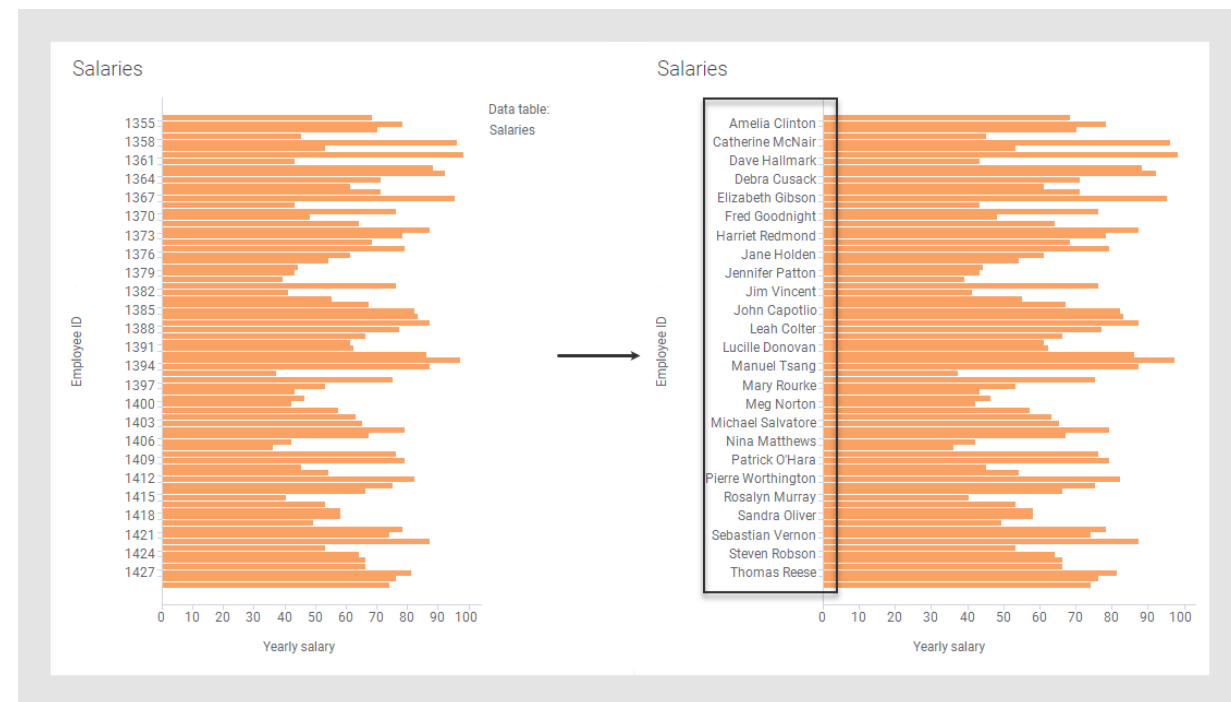
Preview

email	email (Split 1)	email (Split 2)
String	String	String
john.smith@mycompany.se	john.smith	mycompany.se
barbara.no@guest.fi	barbara.no	guest.fi

Data in Analysis Flyout – Custom Values

```
Concatenate([Employees].[First name], " ", [Employees].[Last name])
```

Employee ID	First name	Last name
1354	Adrian	Jones
1355	Amelia	Clinton
1356	Anne	Chang
1357	Bonnie	Foster
1358	Catherine	McNair
1359	Charlotte	Gordon
1360	Daniel	Osterman
1361	Dave	Hallmark
1362	David	Roswell
1363	David	Williams
1364	Debra	Cusack
1365	Diana	Ericsson
1366	Douglas	Moore
1367	Elizabeth	Gibson
1368	Elizabeth	Turnell
1369	Eric	Ballard
1370	Fred	Goodnight
1371	Grace	Milburn
1372	Harold	Kelly
1373	Harriet	Redmond



Data in Analysis Pop-up Menu

Data Tables

Option	Description
Rename...	Allows you to change the name of the selected data table.
Sort order	Changes the display of columns in the flyout. Choose from: Categorized, No sorting, Ascending or Descending views (and Hierarchical, for cube data).
Set as default data table	Sets the selected data table to be the default data table in the analysis.
Data table properties	Opens the Data Table Properties dialog where you can edit the properties of the data tables in the analysis, including relations across tables (internal and external).

Columns

Option	Description
Rename...	Allows you to change the name of the selected column. Note: The columns use the same name as their corresponding filters, so renaming a column will also change the filter name, if a filter exists for the column.
Delete	[Only available for in-memory data tables.] Deletes the selected columns from the data table.
Create filter	[Only available for columns which currently do not have a corresponding filter.] Creates filters for the selected columns.
Change categorization	Allows you to change the category for the selected column.
Geocoding	Allows you to find a geocoding hierarchy in the library and manually specify that a column should be used for geocoding.

Data in Analysis Flyout – Filter Transformation

- Filtering out data using filters, or deleting marked rows, might sometimes not be permanent enough to remove unwanted data from your analysis, because reloading the data or resetting the filters might bring the undesired data back.
- If you want to make sure that values gets permanently removed, you can instead add a **Filter rows** transformation.
 1. In the **Filters** panel, or in the **Data in analysis** flyout, filter your data to include only what you want to keep in the analysis.
 2. When you are done, right-click the filter and select **Create Filter Transformation** from the pop-up menu.

Data Canvas

The screenshot displays the Data Canvas interface, which is used for building data pipelines. The main workspace shows a flow diagram with the following components and steps:

- AccountFeed (Salesforce)**: The initial data source, marked with a red circle 3.
- Added rows**: A transformation step that adds data from **Additional Sales Rows.txt** (marked with a red circle 4) to the main flow (marked with a red circle 6).
- Added columns**: A transformation step that adds data from **Sales 2016** (marked with a red circle 7) to the main flow (marked with a red circle 2).
- AccountFeed**: The final output data table, marked with a red circle 2.

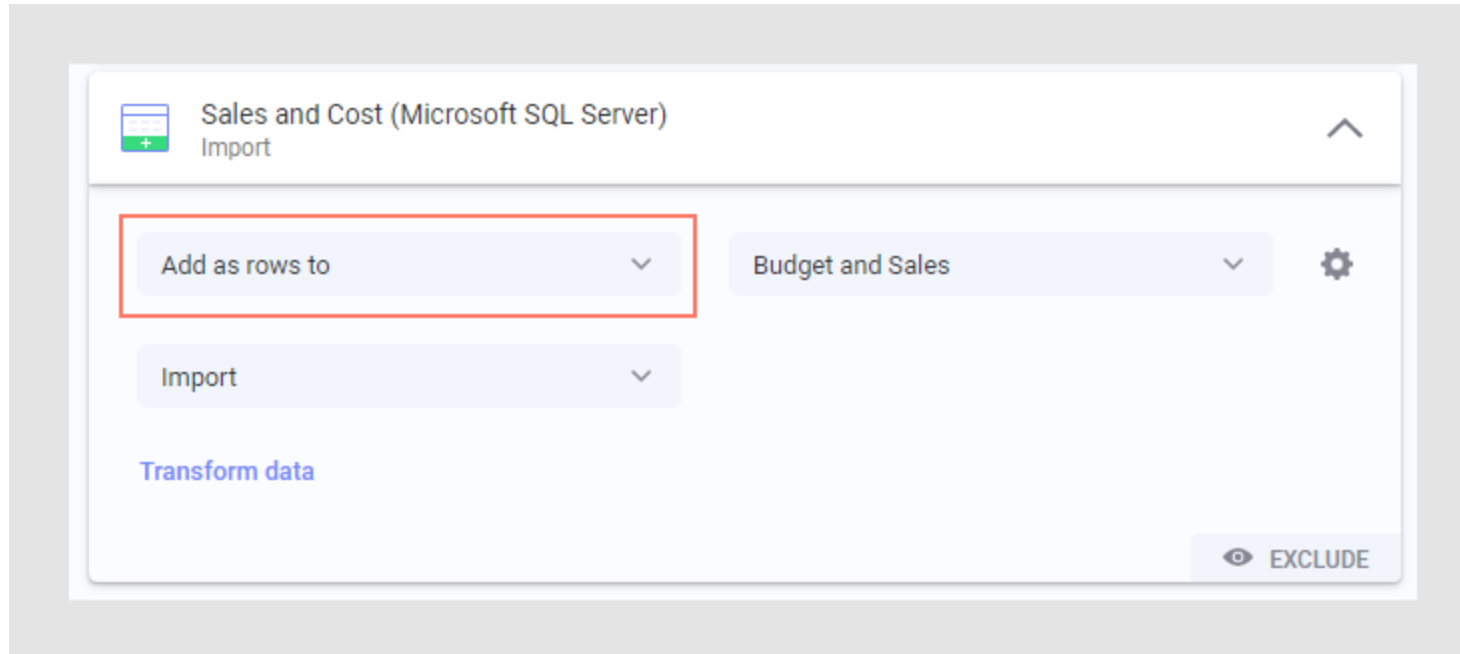
The bottom panel provides details for the **AccountFeed (Salesforce)** data source:

- AccountFeed (Salesforce)**: Loaded data from AccountFeed (Salesforce) (marked with a red circle 8).
- DATA**: A table showing the data loaded from the source (marked with a red circle 9).

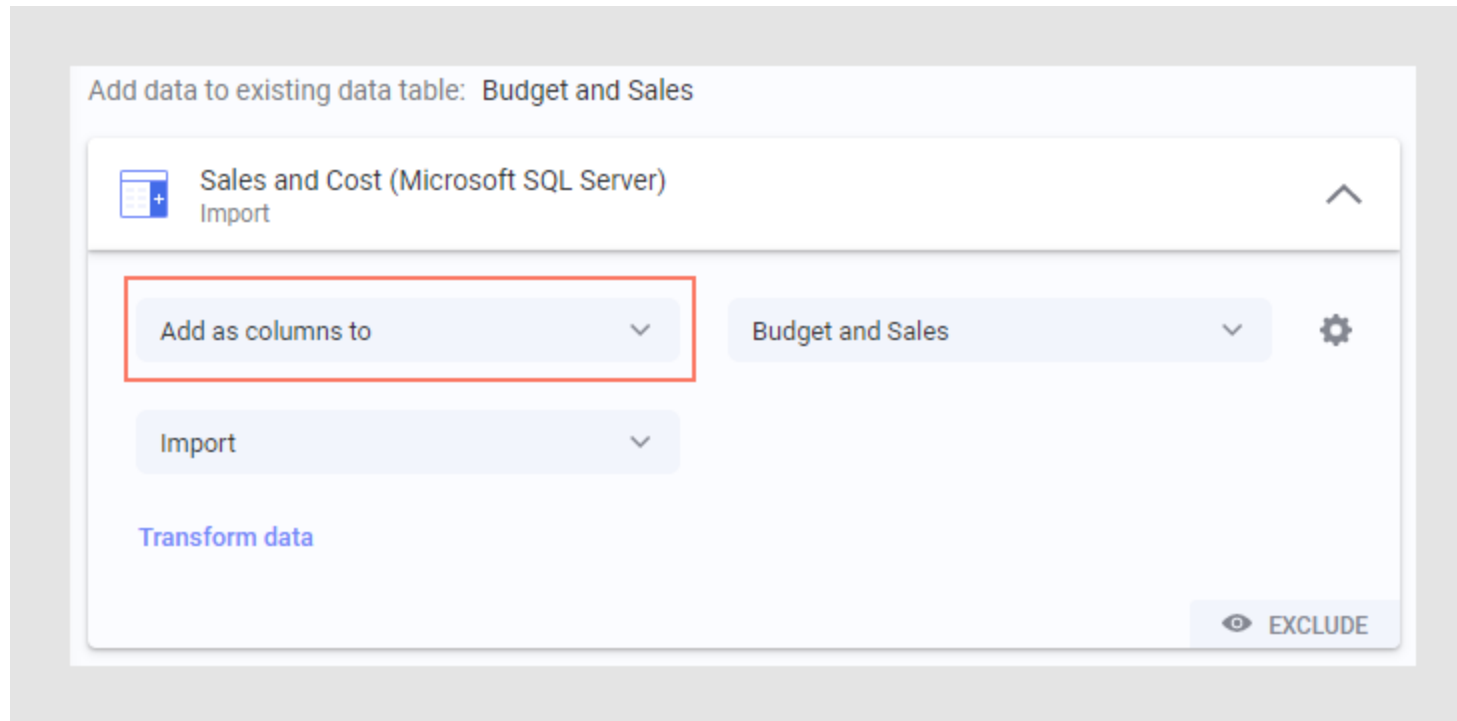
Id	ParentId	Type	CreatedBy	CreatedDate	IsDeleted	LastModifiedDate	SystemModsta...	Comr
String	String	String	String	DateTime	Boolean	DateTime	DateTime	
0D51100002Fdz...	0011100000yd9...	TrackedChange	00511000000n6...	8/21/2019 7:54...	False	8/21/2019 7:54...	8/21/2019 7:54...	
0D51100002Fdz...	0011100000yd9...	TrackedChange	00511000000n6...	8/21/2019 7:56...	False	8/21/2019 7:56...	8/21/2019 7:56...	

2 rows 16 columns

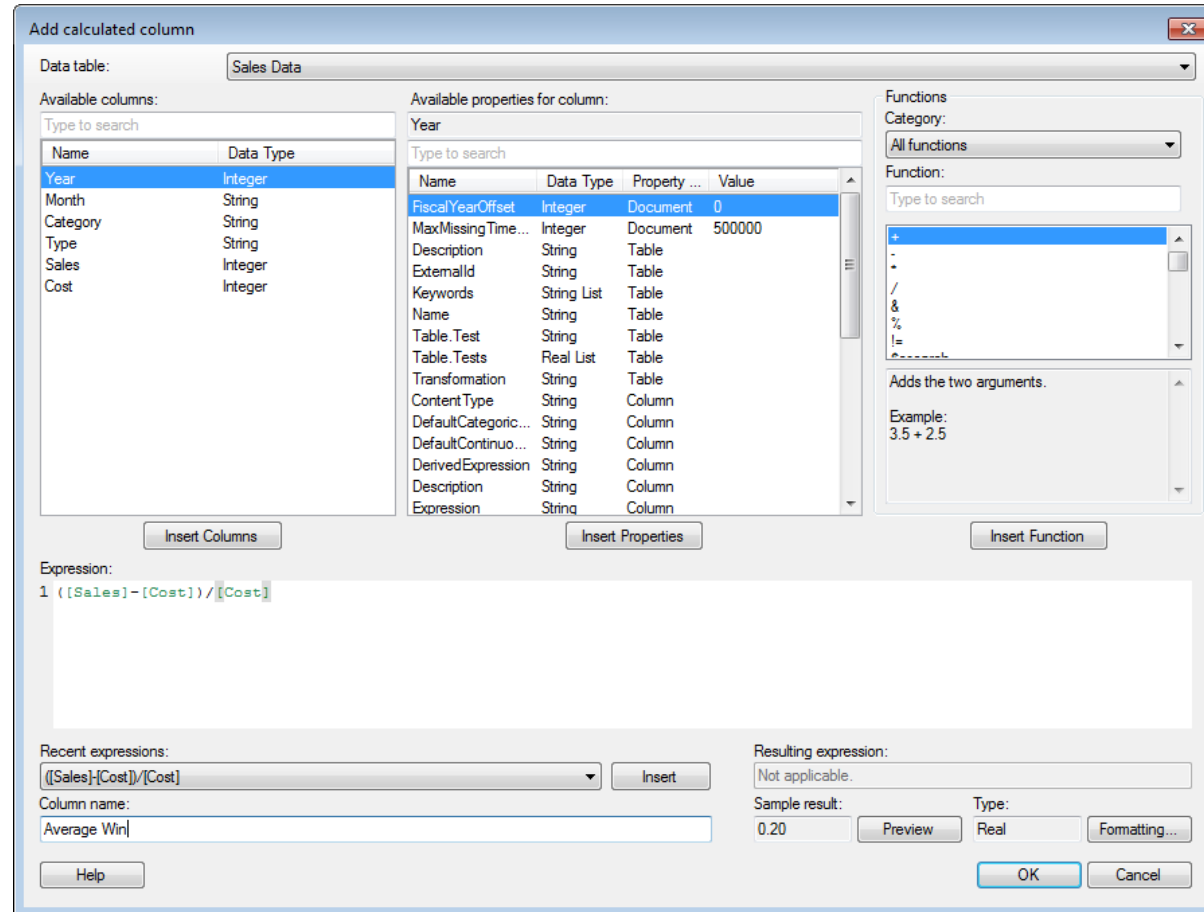
Data Canvas – Add Rows



Data Canvas – Add Columns



Data Canvas – Calculated Columns



Exercise

- Using the data from NEISS (injuries.csv and products.csv) find:
 - a) What 3 demographic groups are more prone to injuries?
 - b) For each of the three groups above, which product category do they get hurt with?
 - c) Which 3 product categories cause more injuries in the population?