Chapter 1. Getting started with Explorations

Explorations

Explore is a flexible workspace where you can discover and analyze data. You can also explore an existing visualization from a dashboard or story. Uncover hidden relationships and identify patterns that turn your data into insights. Correlated insights are represented by a green icon with a number on the x-axis, y-axis, or the title of a chart.

Starting explorations

Use one of several methods to start explorations.

Uploading data

Upload a data asset to your **My content** folder to use in your exploration.

Procedure

1. Click the **New** icon and then click **Upload files**.
   
2. Browse to where you saved your data asset and select it.
   The data asset appears in the **My content** folder.

Starting an exploration from an existing dashboard or story

When you are working on a dashboard or story, you can create or edit an exploration directly from a visualization.

About this task

Complete these steps to open a visualization in a new exploration or to add to an existing exploration:

Procedure

1. Open an existing dashboard or story.
2. Select a visualization.
3. Click the **Explorations** icon  in the toolbar.
4. Select **New exploration** or **Add to existing**.

Starting a new exploration from the New menu

From the **welcome** page, you can start a new exploration from the **New** menu.

Procedure

1. Click **New**  and then click **Exploration**.
2. Select a data source and click **Add**.
   A starting points page is generated from the data source you selected.

Starting a new exploration from a data asset on the welcome page

You can select the **Action menu** on a recently used data asset on the **welcome** page.

Procedure

1. On the **welcome** page, if the data asset you want to use is displayed as a tile in the **Recent** area, click the **Action menu**  icon.
2. Select **Create exploration**.

Adding a data source

Add a data source to your exploration to explore its data.

Procedure

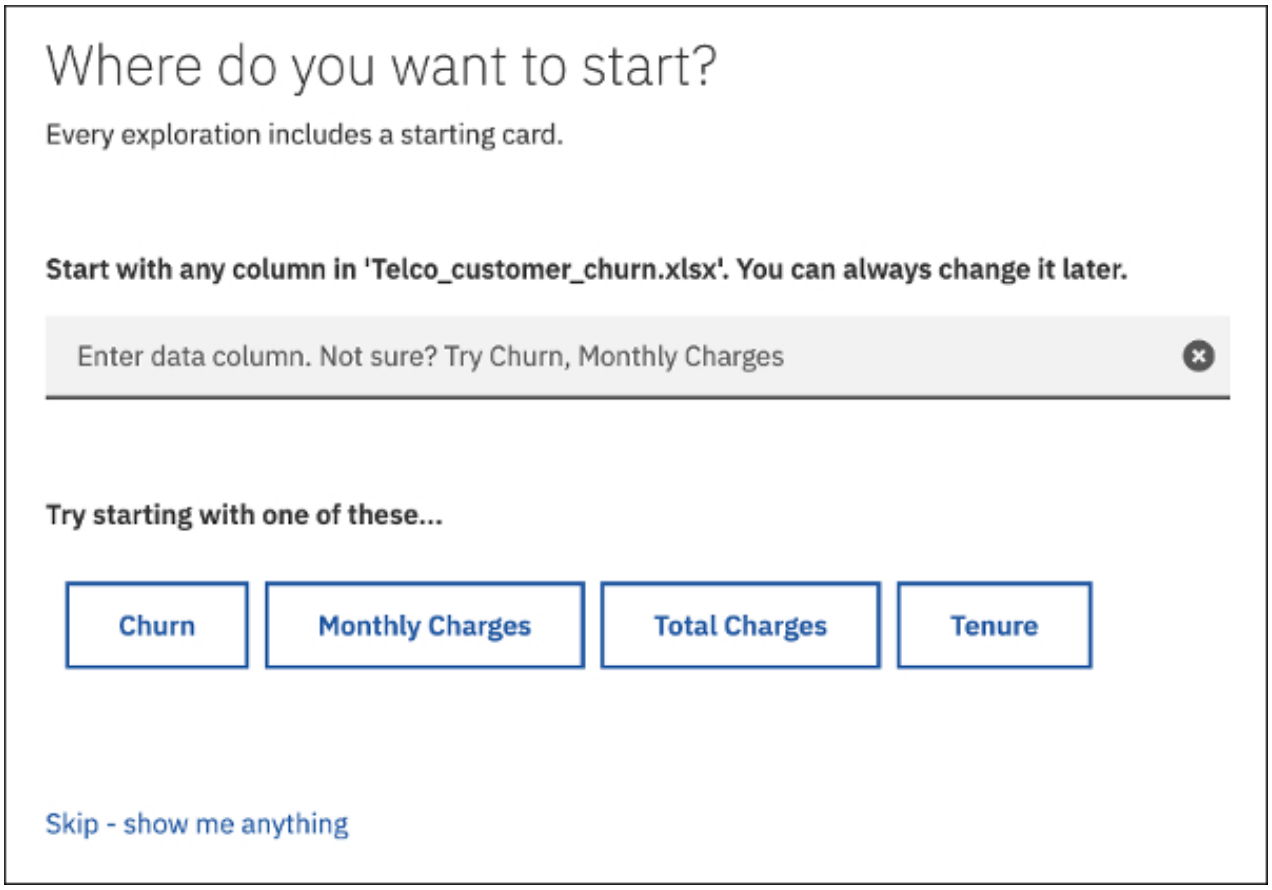
1. In the **Selected sources** pane, click the **Add a source** + icon.
2. Go to **My content** or the **Team content** folder and select the data source that you want to add. Click **Add**.
3. Expand the data source in the **Selected sources** pane to see what's available.
4. Use the starting points page to generate a relationship diagram from your data.

Chapter 2. Exploring relationships

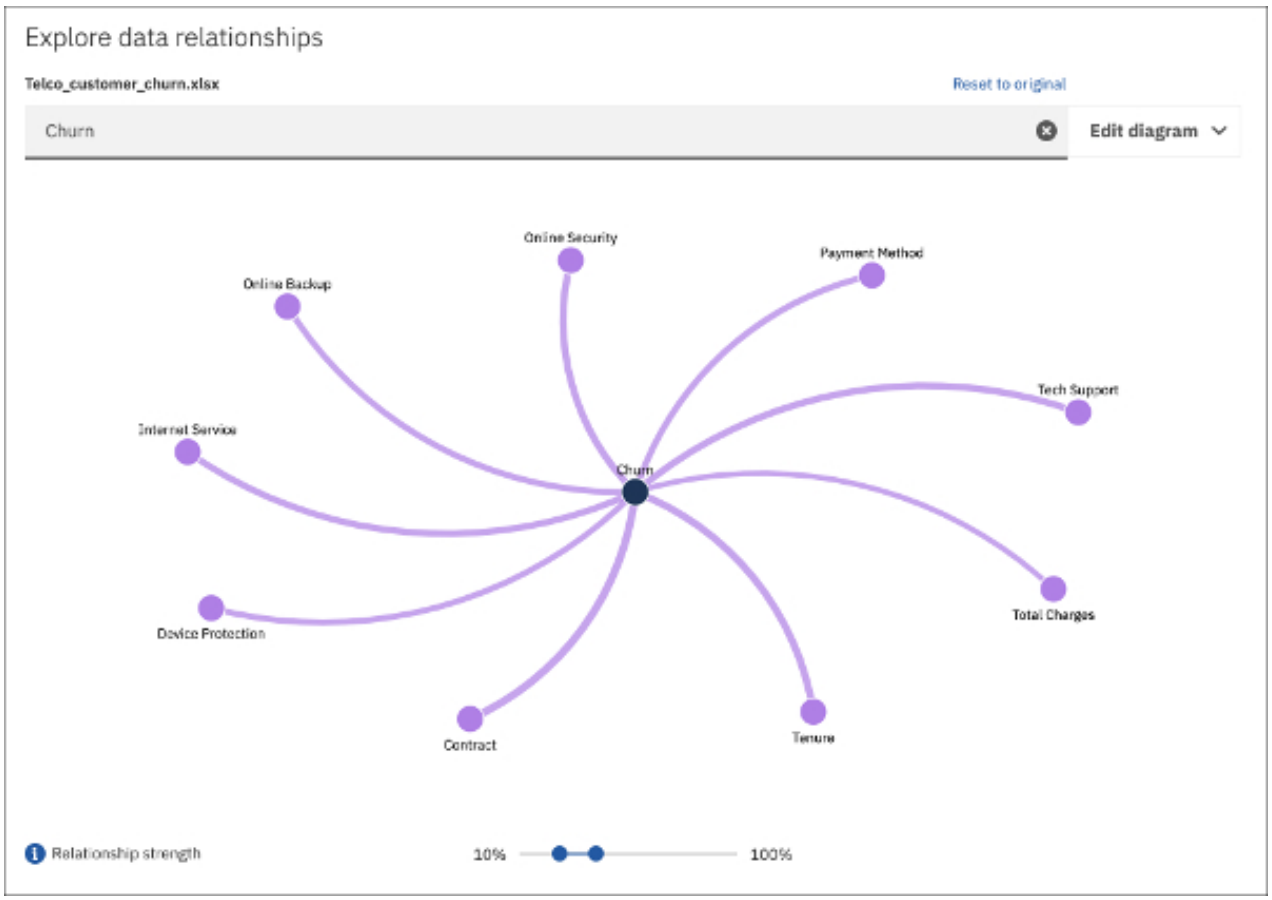
Explore relationships in your data

When you create an exploration, you can start from a data source. A starting points page is displayed with suggestions for how to get started.

You can type a column name that appears in your data source. Or, you can click one of the suggested columns that the system identifies as interesting. If you're not sure which column to start with, click **Skip - show me anything** to see a relationship diagram with some suggested starting point visualizations.

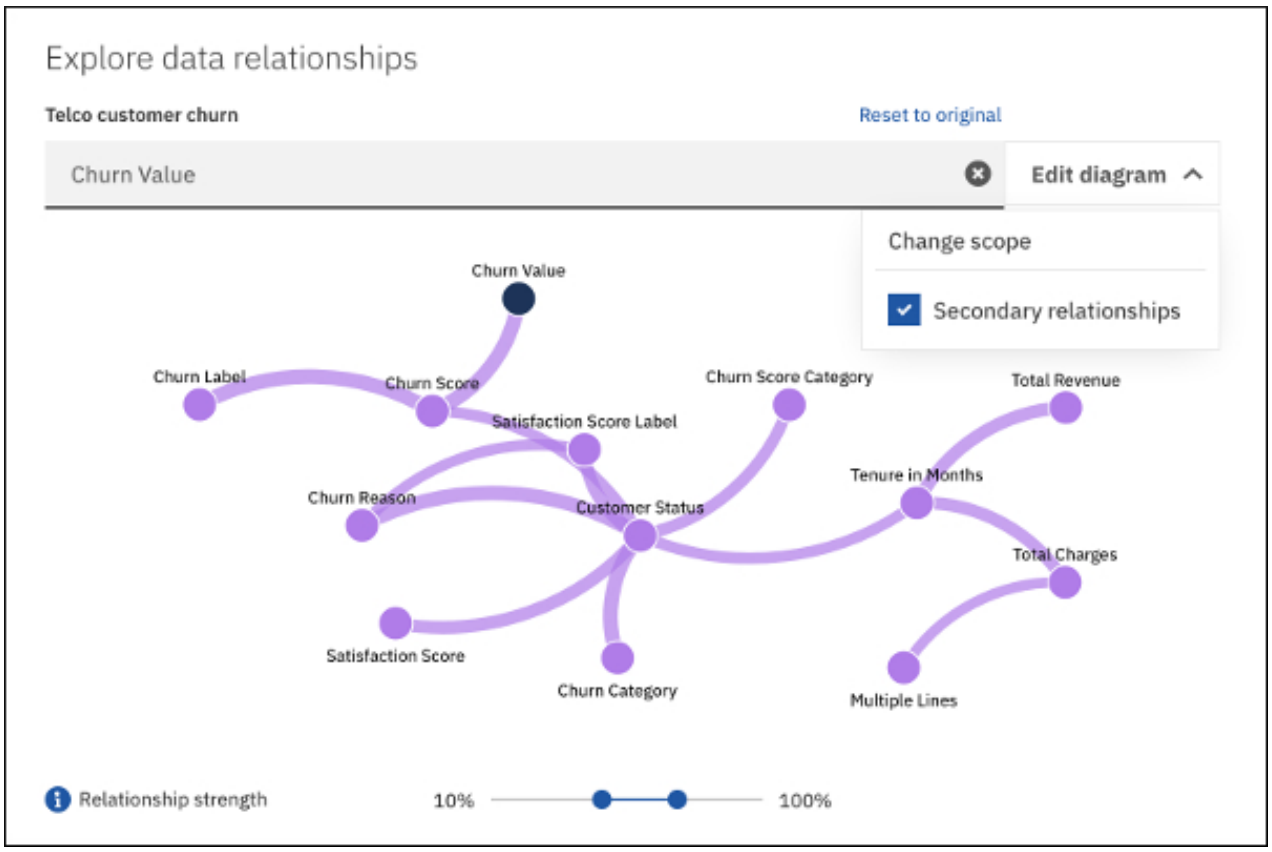


In the relationship diagram, the column that you start with is the prime focus and is represented by a dark blue node. Related fields are represented by purple nodes. Lines connect the nodes and represent relationships. The thickness of the line indicates the strength of the relationship.



The strongest primary relationships are displayed by default and are the direct relationships between the prime focus and the related fields. Secondary relationships are the relationships between other fields directly or indirectly related to the target.

To view both primary relationships and secondary relationships, select the **Secondary relationships** check box under **Edit diagram**.



The relationship diagram plots these fields based on a statistical evaluation of related items. The relationship diagram is not a picture of the data model. However, the model might be an influencing factor in the analysis. To improve performance when there are many rows in the data source, the analysis is based on a representative sample of the entire data.

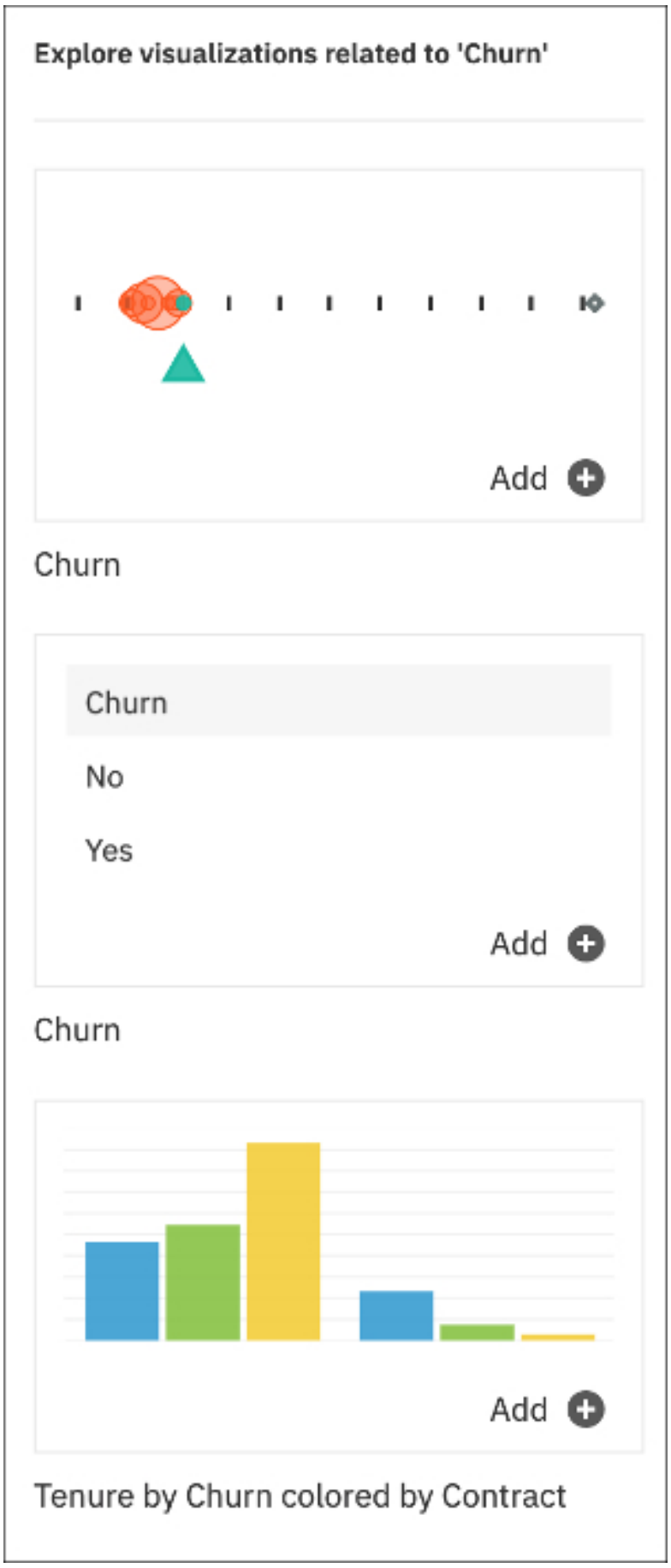
You can interact with the relationship diagram by selecting a node that you are interested in. As you do, the list of suggested starting point visualizations to the right of the diagram updates to include the nodes you selected. You can also use Ctrl+click to select multiple nodes.

Click **Reset to original** if you want to reset the scope and view of all the fields in the relationship diagram to the default setting.

Suggested starting points

Suggested starting point visualizations are displayed as thumbnails beside the relationship diagram. Select single nodes or multiple nodes in the relationship diagram to generate these visualizations.

Click a visualization if you want to add it to your exploration and view it at the same time. Click the plus icon  on the starting point visualization to add it to your exploration and maintain the current view.


Explore visualizations related to 'Churn'
Add
Churn
Churn
No
Yes
Add
Churn
Add
Tenure by Churn colored by Contract

Opening the relationship diagram

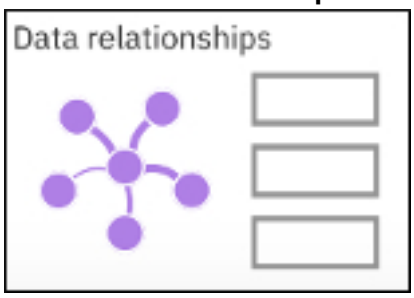
When you're viewing a visualization and you want to return to the relationship diagram, use the **Data relationships card** to return to the starting points view.

About this task

Complete the following steps to return to the starting points view to see a relationship diagram and the suggested starting points.

Procedure

1. Click **Explorations** icon in the side pane.

2. Click the **Data relationships card**.


Tip: The **Data relationships card** is also available from the **New card** menu on the toolbar.

Chapter 3. Visualizations

Visualizations

You can change the visualization type or change the columns that are used in the visualization.

Viewing cards in the navigation panel

View thumbnails of your visualizations, called cards, in the navigation panel to the left of the main view.

About this task

Cards are a collection of visualizations in your Exploration. Use the cards to open your visualizations to view details and modify them using the data slots.

Procedure

1. Click the **Explorations** icon in the side panel.
   
   Your cards are listed here.
2. Click the card that displays a visualization thumbnail that you want to open.
   The visualization opens in the main view.
3. View the generated text under the **Details** tab, or add more data items under the **Fields** tab.
   If you don't see the **Details** tab and **Fields** tab, click the **Show details and fields** icon.

Viewing visualization details

When you open a visualization, it is displayed in the main exploration area. The exploration pane displays the **Details** tab, the **Fields** tab, and the **Properties** tab.

If you don't see the exploration pane, click the **Show details and fields** icon.

Visualization details

The **Details** tab displays text that is generated to describe aspects of the data represented in the visualizations. These details are not obvious from looking at the visualization. For example, the details might reveal an average of the values over time.

Fields

The **Fields** tab is where you can add columns to build and modify visualizations. Add a column to each mandatory field.

Properties

The **Properties** tab is where you can modify properties that apply to your visualizations.

Creating a single visualization

While you work with your exploration, you might decide that you need another visualization.

About this task

Complete the following steps to create a single visualization.

Procedure

1. On the toolbar, select **New card**.
2. Select the blank **Single** card.
3. In the **Sources** window, expand the data asset that you want to use.

   If a different data asset is open, click **Go back** next to the name of the data asset that is open.

4. To create a new visualization, complete one of the following actions:
   - Drag and drop data items onto the **Create a visualization** area.

     IBM® Cognos Analytics creates a visualization to match the data items. For example, when you add Year or Department, a table is created. Drag in a measure, such as Revenue, and a bar visualization is created.
   - Click **Choose a type** and select a visualization type. Then, add a data item to each field.

Creating a visualization using search in data fields

You can quickly build a visualization that uses search in data fields.

Procedure

1. Click the **Explorations** icon in the side panel.

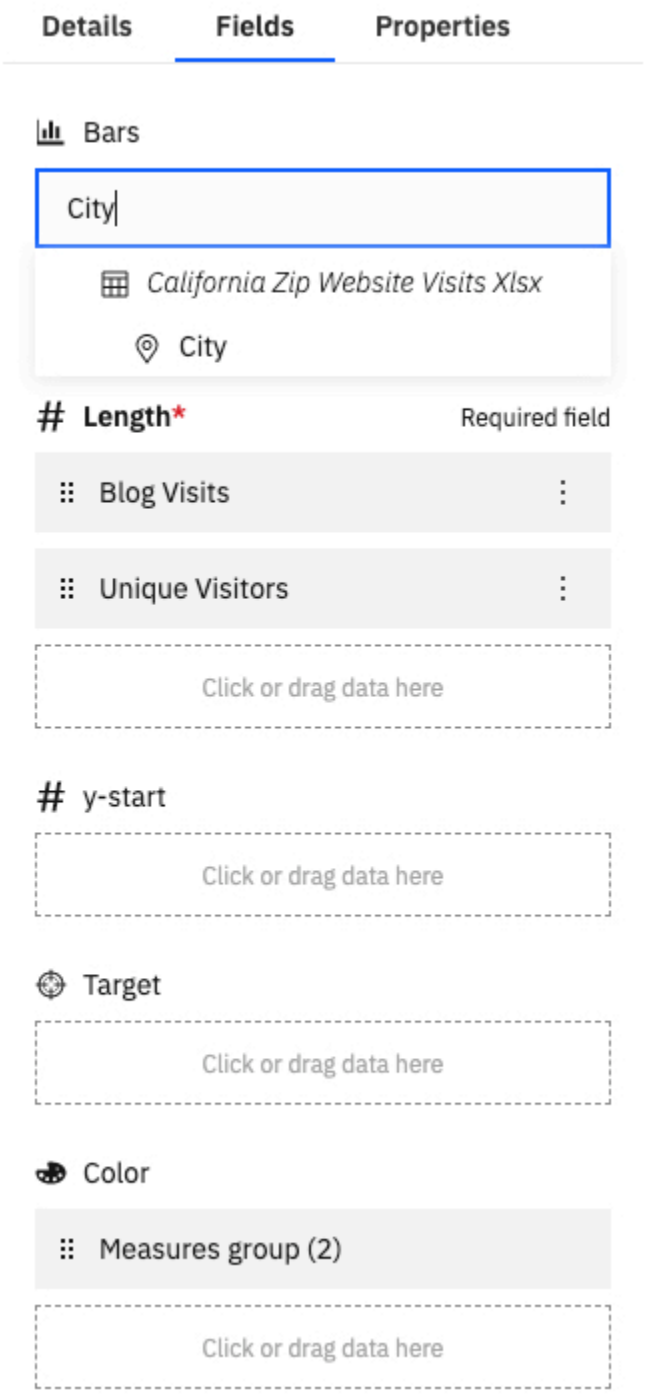
   

   Your cards are listed here.
2. Click the card that displays a visualization thumbnail that you want to open.

   The visualization opens in the main view.
3. Click the **Fields** tab.

   If you don't see the **Fields** tab, click the **Show details and fields** icon.
4. Search for the data you are looking for. Click the data to populate the data field.

Results



Comparing two visualizations

You can create your own comparison to analyze the data between two visualizations. Or, you can start with a recommended comparison. In either case, a summary of key information and differences between the two visualizations is generated.

About this task

Complete the following steps to create a comparison between two visualizations.

Note: When you create a new visualization, you can select a blank comparison card with two slots for visualizations.

Procedure

1. Click the **Explorations** icon in the side pane.
   The **Cards** pane opens.
2. Select a card to create a comparison.
   A visualization is displayed.
3. On the toolbar, click **Compare**.
   The **How do you want to compare?** page is displayed with guidance on how to create your own comparison or start with a recommendation.
4. Click **Add card** icon on a card thumbnail to add it to the list of cards in the navigation panel. Or, click the card thumbnail to add the new card and immediately view it.
5. Optionally, modify the data in one visualization to compare with the other visualization.
   a) Select one of the two visualizations.
   b) In the **Fields** tab, modify the visualization in some of the following ways, for example:
      - Remove filters.
      - Show top or bottom count.
      - Remove data items.
      - From the **Sources** pane, add new data items or filters. Or, use search in the data fields. For more information, see "Creating a visualization using search in data fields" on page 10.

Comparing two data points on a visualization

You can select two data points on an existing visualization and compare the data.

Procedure

1. Click the **Explorations** icon in the side pane.
   The **Cards** pane opens.
2. From the **Cards** pane, select the card that displays a visualization thumbnail that you want to open.
   The visualization opens in the main view.
3. Select two data points on the visualization.
4. Right-click and then click **Compare by**.
5. Type a column to compare the two data points.
   A table displays information about how the two data points compare to each other.

Advanced data analytics

IBM Cognos Analytics is a business intelligence tool for managing and analyzing data. It includes various self-service features that make it possible for users to prepare, explore, and share data. As part of this offering Cognos Analytics includes a number of predictive, descriptive, and exploratory techniques, also known as numeric intelligence. Cognos Analytics uses many statistical tests to analyze your data. It is important to understand the definitions of these tests as they apply to Cognos Analytics.

For more information, see the *IBM Cognos Analytics Dashboard and Stories Guide*.

Choosing a different visualization type

Visualizations communicate comparisons, relationships, and trends. They emphasize and clarify numbers. To choose a visualization type, consider what you want the visualization to illustrate and what appeals to the audience for the visualization.

Before you begin

For more information on visualization types, see the visualization documentation in the *IBM Cognos Analytics Dashboards and Stories User Guide*.

Procedure

1. From the **Cards** pane, select the card that represents the visualization you want to open.
2. Click the **Choose visualization type** icon  in the toolbar.
3. Click the visualization type that you want to use.

   Notice how each visualization type communicates data differently. For example, use a bar, column, or line visualization to compare a set of values. Use a line or area visualization to track relationships. Use a tree map or pie visualization to see the parts of a whole.

Area

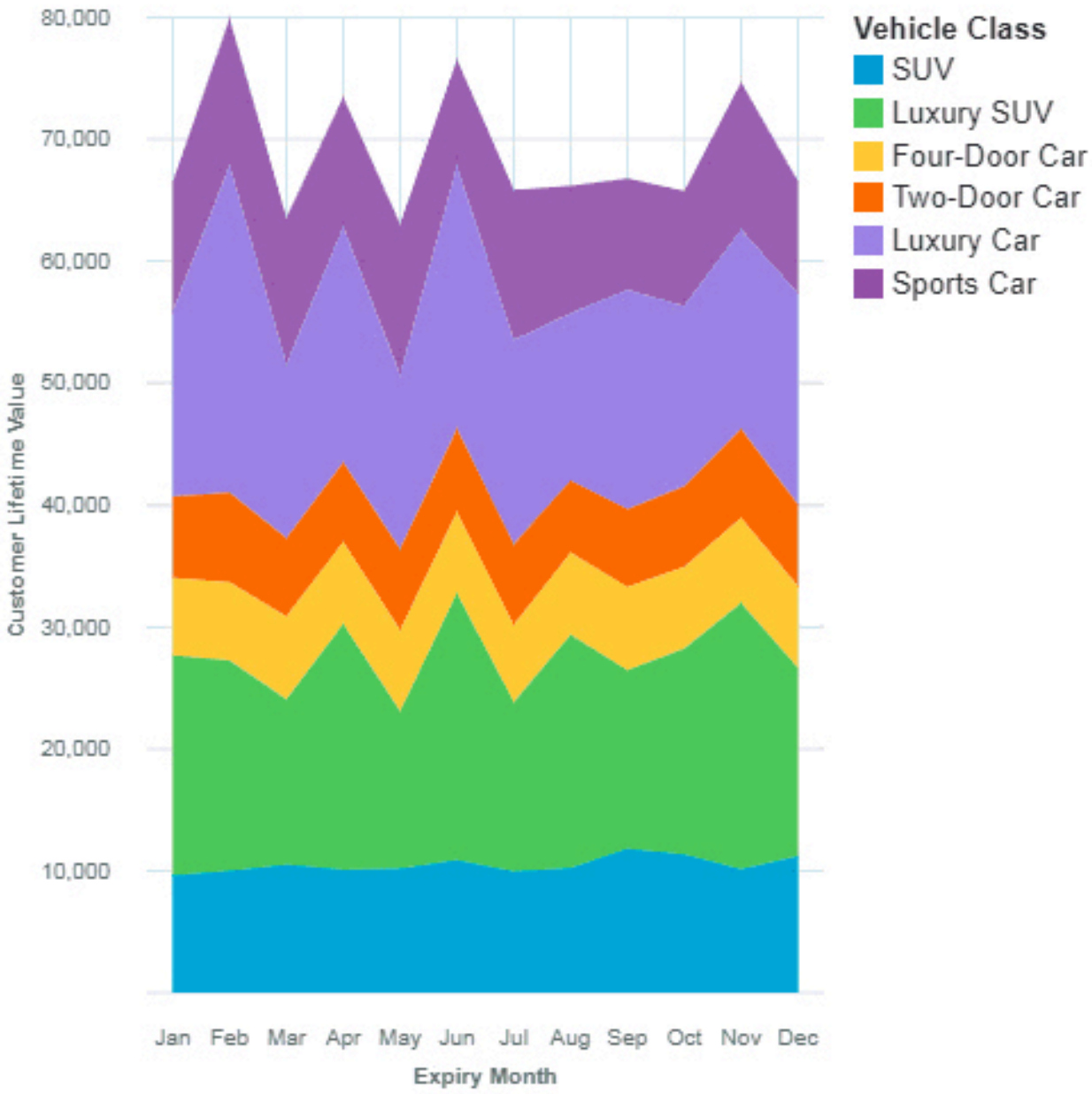
Use an area visualization to emphasize the magnitude of change over time.

Area charts are like line charts, but the areas below the lines are filled with colors or patterns. Stacked charts are useful for comparing proportional contributions in a category. They plot the relative value that each data series contributes to the total.

Because an area visualization stacks the results for each column or item, the total of all results is easily seen.

For example, an area visualization is excellent for looking at revenue over time across several products.

For example, this area visualization shows the customer lifetime value for each vehicle class per month. Because the area visualization stacks the results, you see the totals for each month.



The area visualization was created by dragging the following data items from the Sources panel:

- Drag **Expiry Month** type onto the **x-axis** field.
- Drag **Vehicle Class** onto the **Color** field.
- Drag **Customer Lifetime Value** onto the **y-axis** field

Samples

You can see an example of a word cloud visualization in the sample report **Customer lifetime value analysis**. You can find the sample here: **Team content** > **Samples** > **Reports** > **Customer lifetime value analysis**.

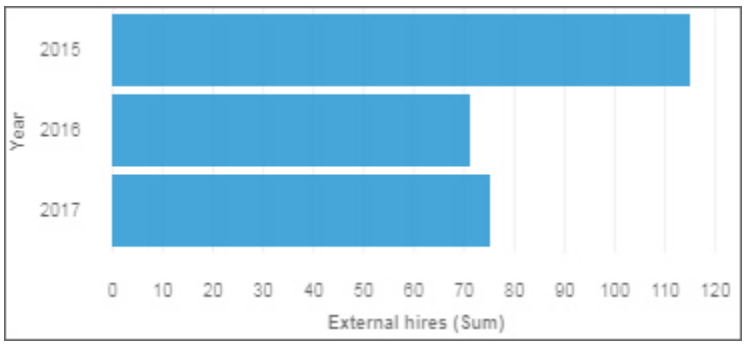
If any of the sample objects are missing, contact your administrator.

Bar

Use a bar visualization to compare values by one or more columns, such as sales for products or sales for products each month.

Bar visualizations use horizontal data markers that are arranged in groups to compare individual values. You can use bar visualizations to compare discrete data or to show trends over time.

A bar visualization can show change over a specific time period or can compare and contrast two or more columns in a time period or over time. If there are so many bars that the labels are impossible to read, filter the data to focus on a subset of the data or use a tree map.



Use the **Target** field to show measures that need to be compared against a target value.

Use the **y-start** field to define where the measure must start.

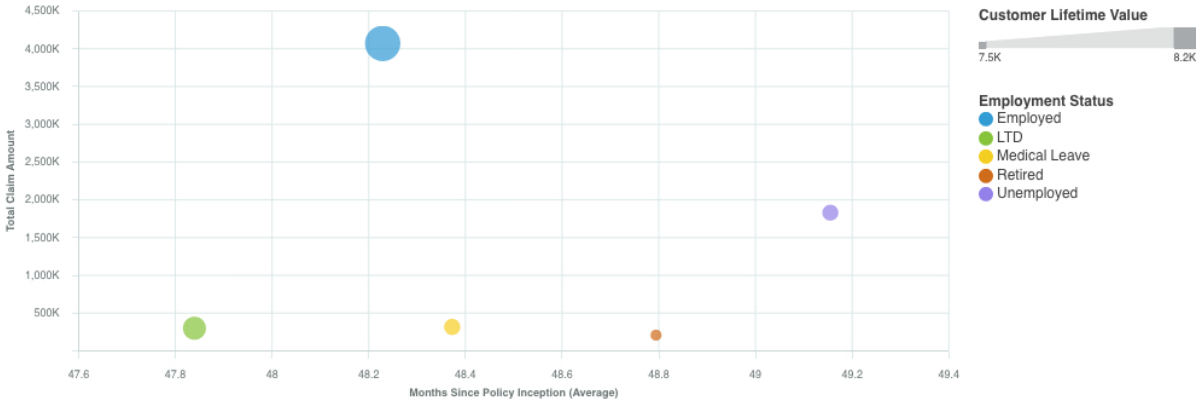
Bubble

Use a bubble visualization to show relationships among columns that contain numeric values, such as revenue and profit.

A bubble visualization uses data points and bubbles to plot measures anywhere along a scale. One measure is plotted along each axis. The size of the bubble represents a third measure. Use bubble visualizations to represent financial data or any data where measure values are related.

The bubbles are in different sizes and colors. The x-axis represents one measure. The y-axis represents another measure, and the size of the bubbles represents the third measure. In the example shown below, color is represented by an identifier.

The example that is shown represents the months since the policy inception.



Create the Bubble visualization by dragging the following data items from **Customer Analysis** in the **Sources** pane :

- Drag **Months Since Policy Inception** onto the **x-axis** field.

- Drag **Total Claim Amount** onto the **y-axis** field.
- Drag **Customer Lifetime Value** onto the **Size** field.
- Drag **Employment Status** onto **Color**

You can customize the bubble chart. For example, to make the x-axis of the bubble chart appear as it does in the sample, do the following steps:

1. Click the visualization, then in the **Data** pane, click the **<Total Claim Amount>** data item.
2. Click
3. Next to **Data format**, click and set the following options:
   - **Format type**: Currency
   - **Currency symbol**: K
   - **Currency symbol position**: End
   - **Number of decimal places**: 0
   - **Scale**: -3 (this presents values in thousands).
4. Click **OK**.

To change the size of the visualization, click the visualization, then set the following option in the properties pane:

- **Size** - **Width**: 700 px, **Height**: 300 px

Click to close the **Properties** pane.

Samples

You can see examples of visualizations in the sample report **Customer lifetime value analysis**. You can find the samples here: **Team content** > **Samples** > **Reports** > **Customer lifetime value analysis**.

If any of the sample objects are missing, contact your administrator.

Bullet

Use bullet charts to show measures that need to be compared against a target value.

In a call center, a bullet chart can be used to measure metrics like call volume, call answer speed, and percentage of abandoned calls.

In manufacturing, a bullet chart can be used to track metrics like number of defects and orders that are shipped.

In a fitness context, a bullet chart can be used to measure metrics like steps that are taken and calories that are burnt.

Bullet visualizations compare an actual measure (the bullet) to targeted measure (the target). Bullet visualizations also relate the compared measures against colored regions in the background that provide more qualitative measurements, such as good, satisfactory, and poor. Bullet visualizations can be shown at small sizes while still effectively conveying information.

A bullet visualization features a single, primary measure. For example, current year-to-date revenue. And compares that measure to one or more other measures to enrich its meaning. For example, compared to a target. The primary measure is displayed in the context of a qualitative range of performance, such as poor, satisfactory, and good.

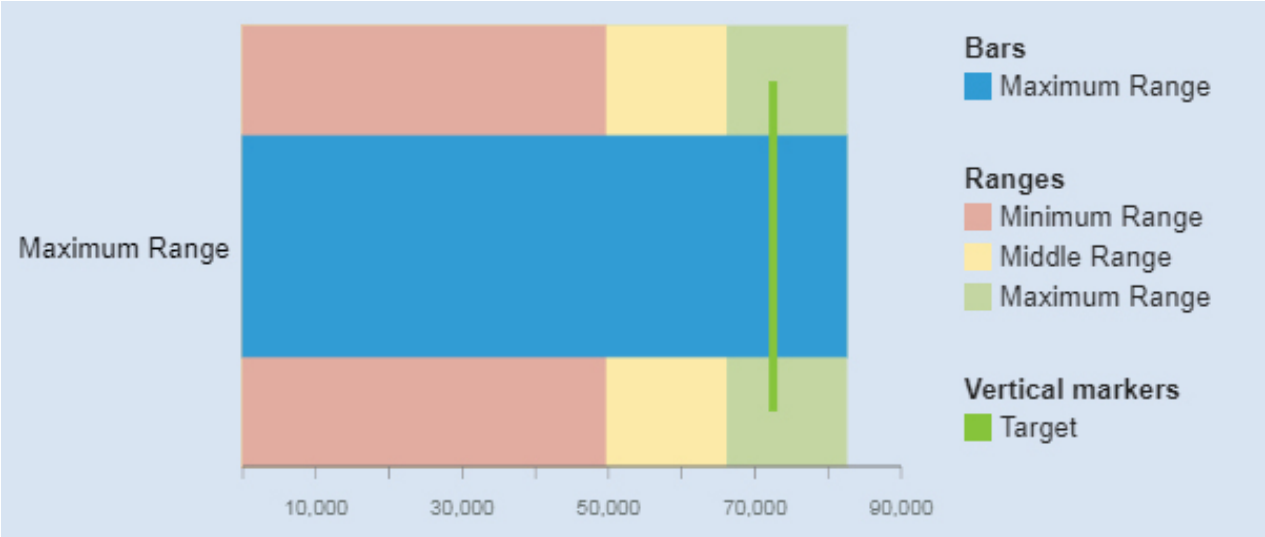
If you select a bullet visualization, then specify the following fields:

- The **Actual bar** field specifies the actual measure.
- The **Target** field specifies the target measure.
- The **Minimum range** field specifies the minimum qualitative range.

- The **Medium range** field specifies the middle qualitative range.
- The **Maximum range** field specifies the higher qualitative range.

Note: Drill-through is not available for a bullet visualization.

Make sure the minimum, medium, and maximum ranges relate to the actual and target measure.



The bullet visualization was created by dragging the following data items from the Sources panel:

- Drag **Minimum Range** onto the **Minimum Range** field.
- Drag **Minimum Range** onto the **Minimum Range** field.
- Drag **Maximum Range** onto the **Maximum Range** field
- Drag **Maximum Range** onto the **Actual Bar** field
- Drag **Target** onto the **Target** field
- Drag **Vehicle class** onto the **Extra Data** field

Samples

You can see an example of a bullet visualization in the sample report **Customer lifetime value analysis**. You can find the sample here: **Team content** > **Samples** > **Reports** > **Customer lifetime value analysis**.

If any of the sample objects are missing, contact your administrator.

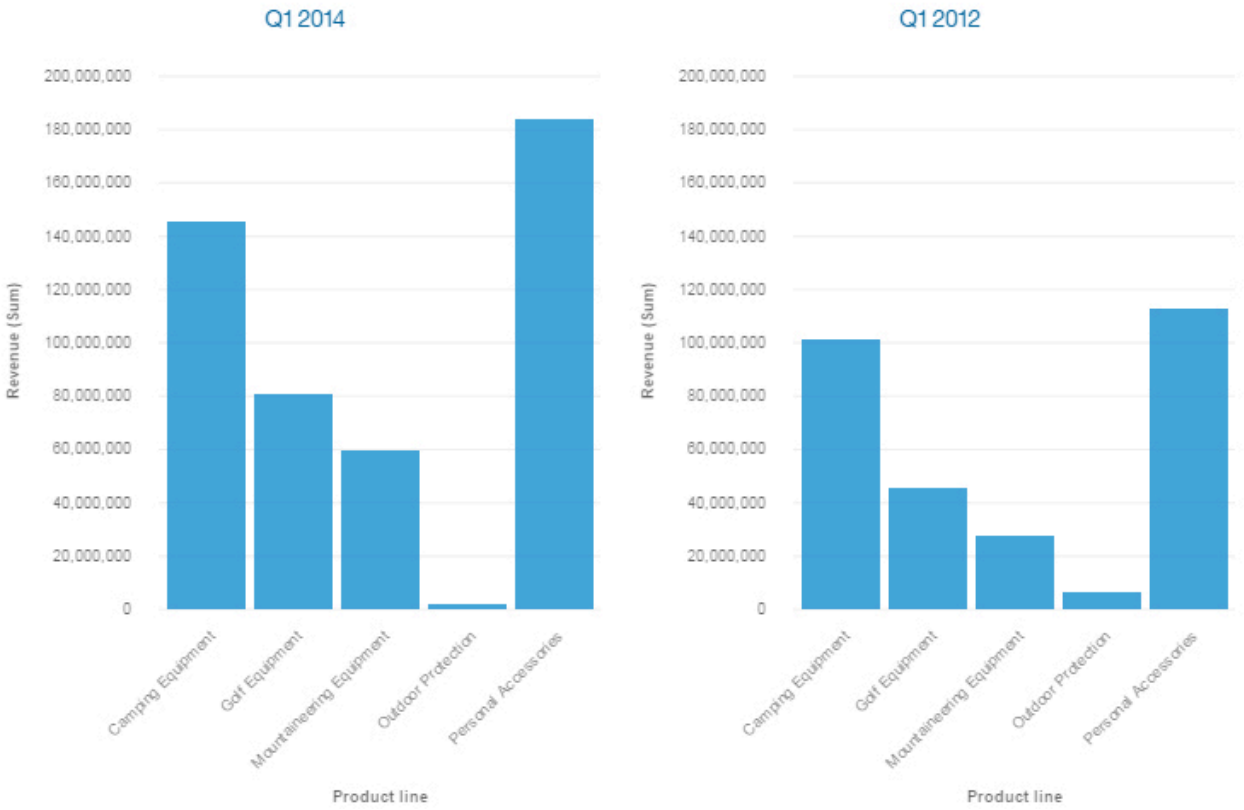
Column

Use a column visualization to compare values by one or more columns, such as sales for products or sales for products each month.

Column visualizations use vertical data markers that are arranged in groups to compare individual values. Use column visualizations to compare discrete data or show trends over time.

A column visualization shows change over a specific time period or can compare and contrast two or more columns in a time period or over time. If there are so many bars that the labels are impossible to read, filter the data to focus on a subset of the data or use a tree map.

For example, revenue for each product line is grouped by quarter, which emphasizes performance in each quarter.



Use the **Target** field to show measures that need to be compared against a target value.

Use the **y-start** field to define where the measure must start.

Crosstab

Use a crosstab when you want to show the relationships between three or more columns. Crosstabs show data in rows and columns with information summarized at the intersection points.

For example, this crosstab shows the course costs for each department by organization.

Course cost	GO Asia Pacific corporate						
	Executive Offices	Finance	Sales (Corporate)	Procurement	Human Resources	Marketing	Summary
2014	10,500	29,750	10,500	18,250	11,000	15,500	95,500
2015	11,500	46,250	14,750	7,750	11,500	30,750	122,500
2016	12,500	43,000	33,250	5,000	14,500	22,500	130,750
2017	13,500	24,000	36,000	7,000	15,750	21,750	118,000
Summary	48,000	143,000	94,500	38,000	52,750	90,500	466,750

Starting from Cognos Analytics version 11.1.4, you can drag data from the **Selected sources** pane and insert data in a column/row or drop the data on top of existing data to replace it.

Data player

Use a data player to see an animation of the impact of a column on the other visualizations.

Decision tree

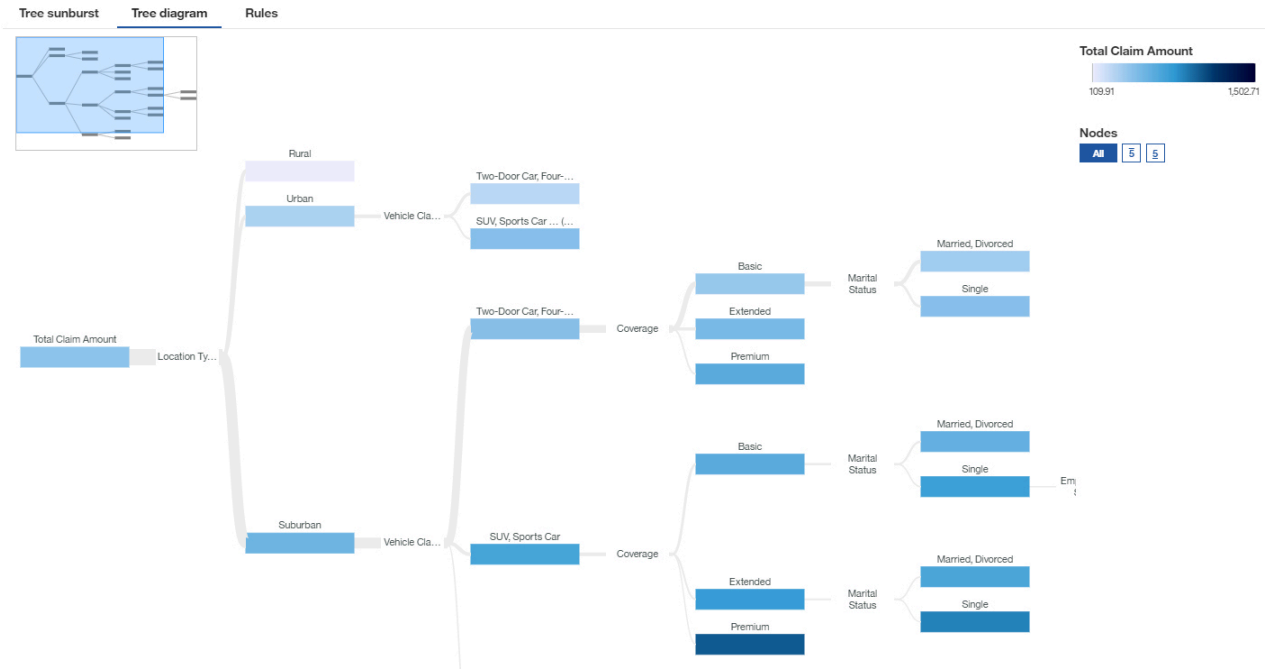
A decision tree shows a connected hierarchy of boxes to represent the values of records.

Records are segmented into groups, which are called nodes. Each node contains records that are statistically similar to each other with respect to the target field. For example, a node might contain the records for males who have more than 18 years of education. Nodes can then be used to predict a target's field value. For example, the node about males and education might be used to predict salary.

Each branch in a decision tree corresponds to a decision rule. For more info about decision rules, see "Viewing decision rules" on page 22

To improve performance, due to number of rows in the data source, the analysis is based on a representative sample of the entire data.

For example, a decision tree visualization can look like this:



Note: Filters are not supported for decision tree visualizations.

For more information, see "Exploring a decision tree visualization" on page 19.

Exploring a decision tree visualization

A decision tree visualization is used to illustrate how underlying data predicts a chosen target and highlights key insights about the decision tree.

About this task

The predictive strength of a decision tree determines the degree to which the decisions represented by each branch that is shown in the tree, predicts the value of the target.

Decision trees have a single target. If the target field of the decision tree is continuous, then the key insight indicators highlight unusually high or low groups. If the target field of the decision tree is categorical, then the key insight is the mode of the node. The mode of the node is the most frequently occurring category or categories of the target field within the group.

To improve performance, due to number of rows in the data source, the analysis is based on a representative sample of the entire data.

When you review a decision tree:

- If you want to see all the drivers, use either the **Tree diagram** tab or the **Rules** tab.
- If you want to focus on key drivers, use the **Tree sunburst** tab.

To edit or add key drivers, click the  on the target field.

Insights are different depending on the type of your target. If you are predicting a continuous measure, for example income, age, or profit, then the decision tree shows within the node the average value of the target given the conditions so far within the group that is represented by the node. For example, if you have a tree that is predicting income and you have a branch that has gender and then city. If you follow the path from male to Chicago, then the value that is in the Chicago node, is the average income of males in Chicago.

Procedure

1. If you have a continuous measure, the following example illustrates the decision tree.

   The color shows whether the value of the node is associated with high, medium, or low values of the target. The color of the node is based on the average of the target for the measure. The higher the average value of the target for a node, the darker the color.

   For example, shown next is the detailed visualization for Restaurant Spending. The international terminal is a strong predictor for high restaurant spending for business travelers. Poor airport cleanliness is a predictor of low spending for conference/convention travelers.

   The minimap helps you move around the areas of the tree. The minimap is especially useful if there are many nodes.

   In this example, the top five highest target values are indicated with a number. You can choose between the following options:

   - Full tree. No highest, or lowest values are indicated specifically.

   - Top five highest target values. The top five highest target values are shown.

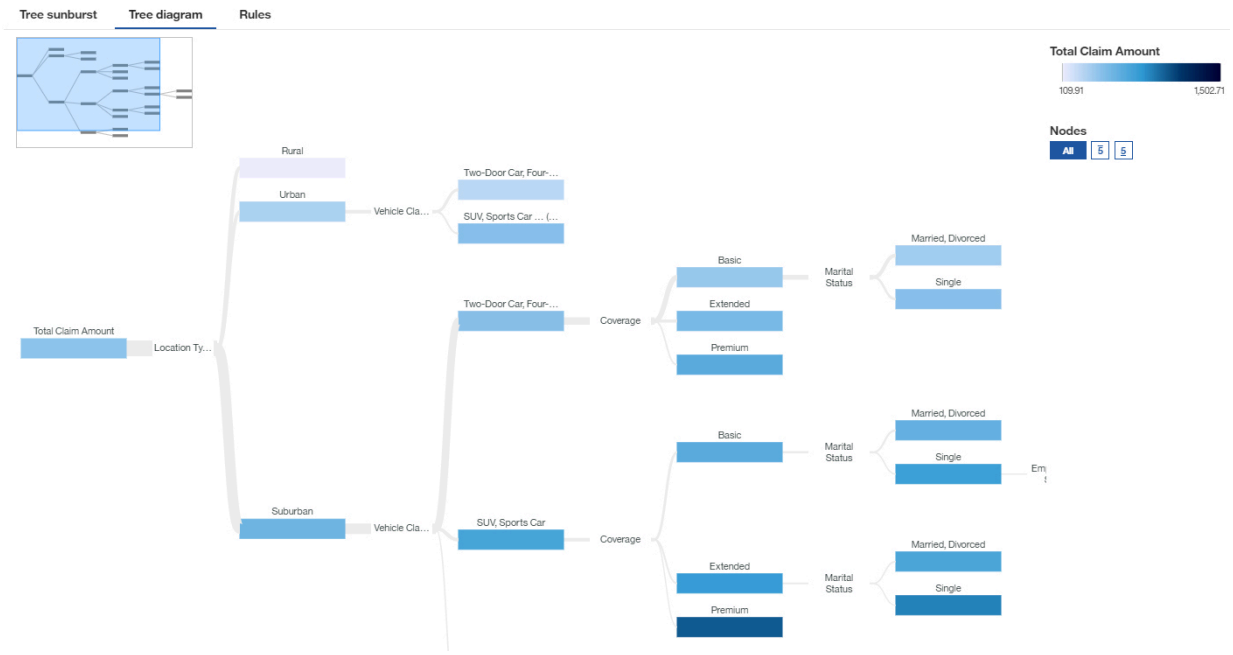
     

   - Top five lowest target values. The five lowest target values are shown.

   If you have a categorical measure, select the category for which you want to see the top five or lowest five targets from the **Top 5 nodes for:** menu or from the **Bottom 5 nodes for:** menu.

   In case you zoomed in too far, the top five or bottom five nodes are not visible.

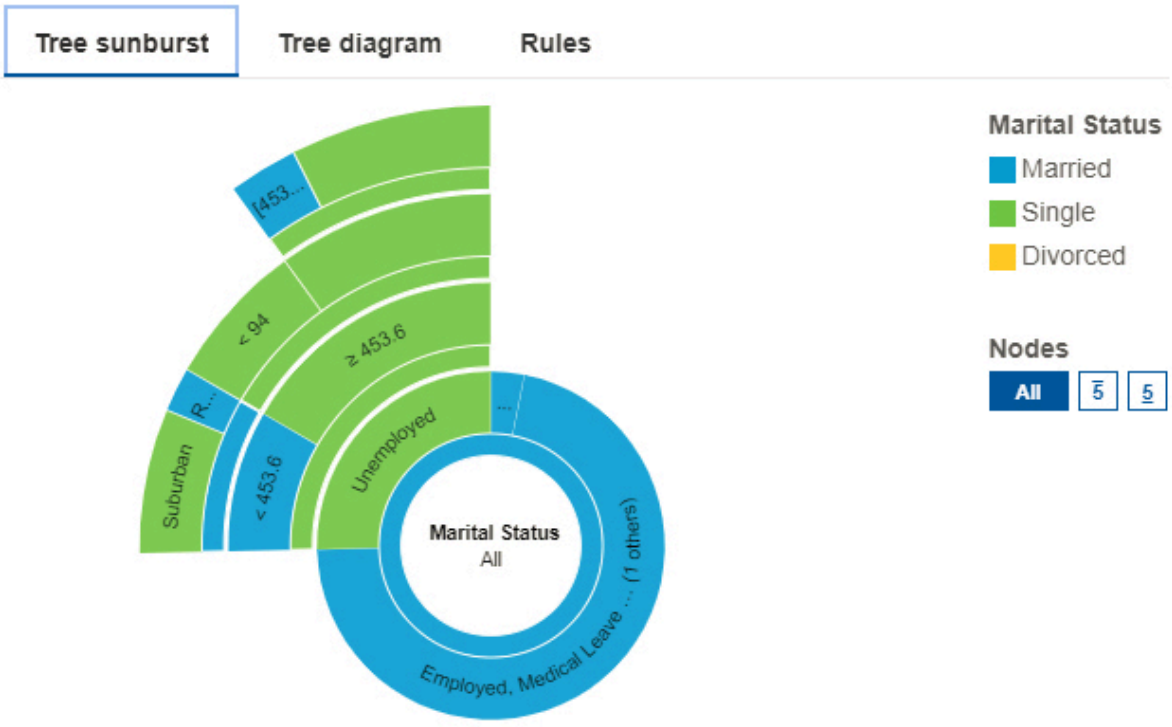


2. If you have a categorical measure, the following example illustrates the decision tree.

   The color shows which field value or values are represented the most.

   In the **Tree sunburst** tab, you can see that if the measures within the decision tree are strong predictors for a target value or target values, then the colors prevail in that node. The non-significant values are left out.

   For example, shown next is the detailed visualization of the marital status in the **Tree sunburst** tab. It shows that being employed is a strong predictor for being married.



   In the **Tree diagram** tab, the nodes visually show the distribution of the people by marital status.

Viewing decision rules

A decision rule predicts an outcome in the target field. Viewing the decision rules helps you determine which conditions are likely to result in a specific outcome.

For example, consider some hypothetical decision rules that might predict churn. These rules might identify classifications based on the ranges for customer age and number of previous claims. From these rules, you might observe that customers who have no or 1 claim and are older than 50 are more likely to churn.

Each branch in a decision tree corresponds to a decision rule.

Procedure

1. In a decision tree, tap **Rules**.
2. Review the decision rules.
3. To return to the visualization, tap **Tree diagram**.

Driver analysis

A driver analysis visualization shows you the key drivers, or predictors, for a target. The closer the driver is to the right, the stronger that driver is.

IBM Cognos Analytics uses sophisticated algorithms to deliver highly interpretable insights that are based on complex modeling. You don't have to know which statistical tests to run on your data. Cognos Analytics picks the right tests for the data.

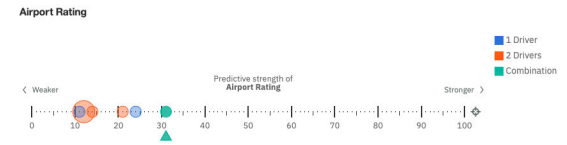
Key drivers for both continuous and categorical targets are available in the driver analysis visualization in dashboards and explorations.

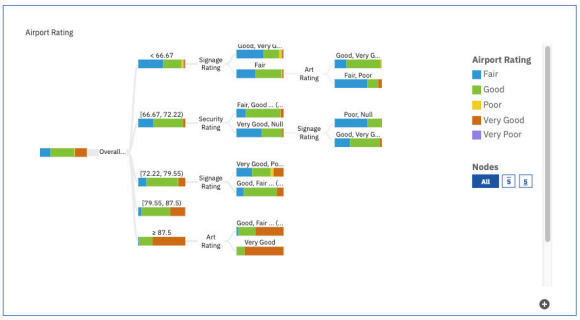
For more info, see *Statistical tests* documentation in the *IBM Cognos Analytics Dashboards and Stories User Guide*.

For example, this driver analysis visualization shows that the combination of overall satisfaction, signage rating, security rating, and art rating are the strongest drivers of the target airport rating.

To edit or add key drivers, click the  on the target data slot.

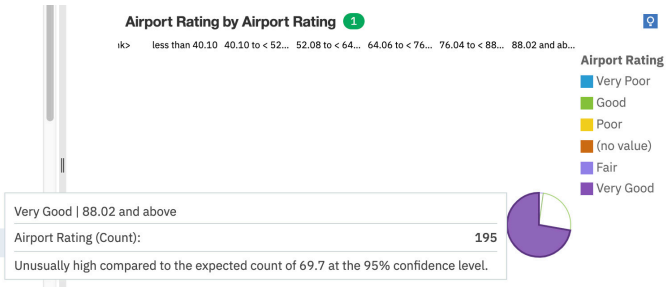
To improve performance, due to number of rows in the data source, the analysis is based on a representative sample of the entire data.





If you hover over a data point, then the driver analysis visualization shows what drives the overall airport rating.

If you click a data point in the tree, other recommended visualizations are shown.



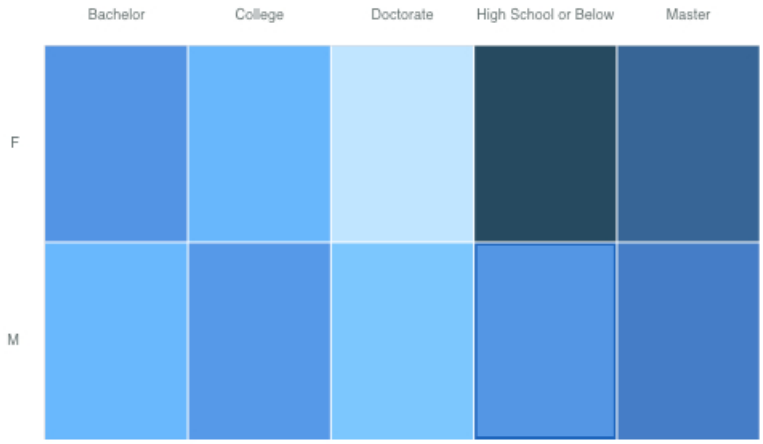
Note: Filters are not supported for driver analysis visualizations.

Heat map

Use a heat map visualization to visualize the relationship between columns, represented in a matrix type view.

A heat map visualization uses color and intensity of the color to show the relationship between two columns.

For example, this heat map visualization shows the average customer lifetime value by gender and education.



Create the heatmap visualization by dragging the following data items from the **Sources** panel :

- Drag **Gender** onto the **Rows** field.
- Drag **Education** onto the **Columns** field.
- Drag **Customer Lifetime Value** onto the **Heat** field.

Samples

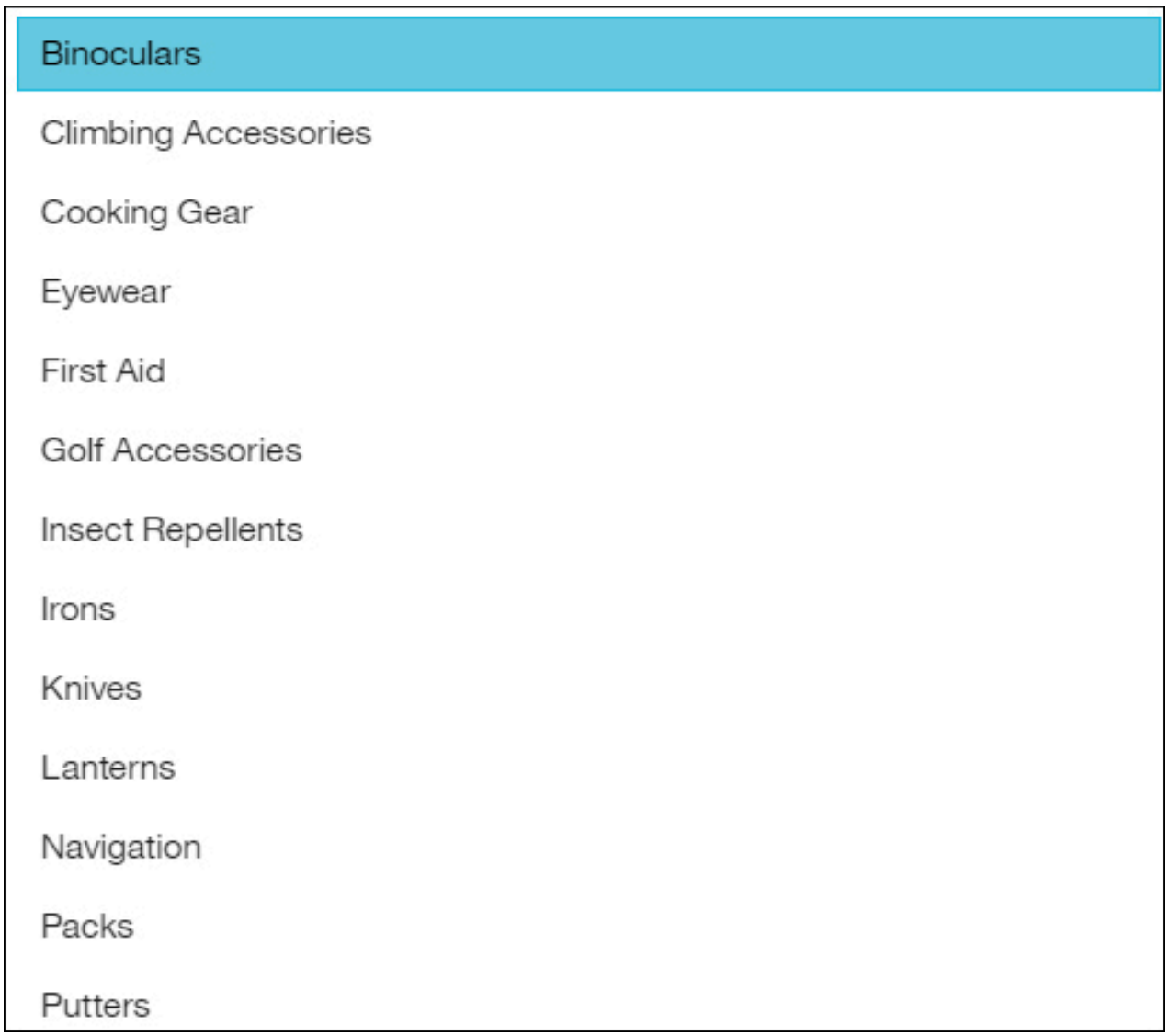
You can see examples of visualizations in the sample report **Customer lifetime value analysis**. You can find the samples here: **Team content** > **Samples** > **Reports** > **Customer lifetime value analysis**.

If any of the sample objects are missing, contact your administrator.

Hierarchy

Use a hierarchy when you want to see the data in rows and columns.

For example, this hierarchy shows product types.



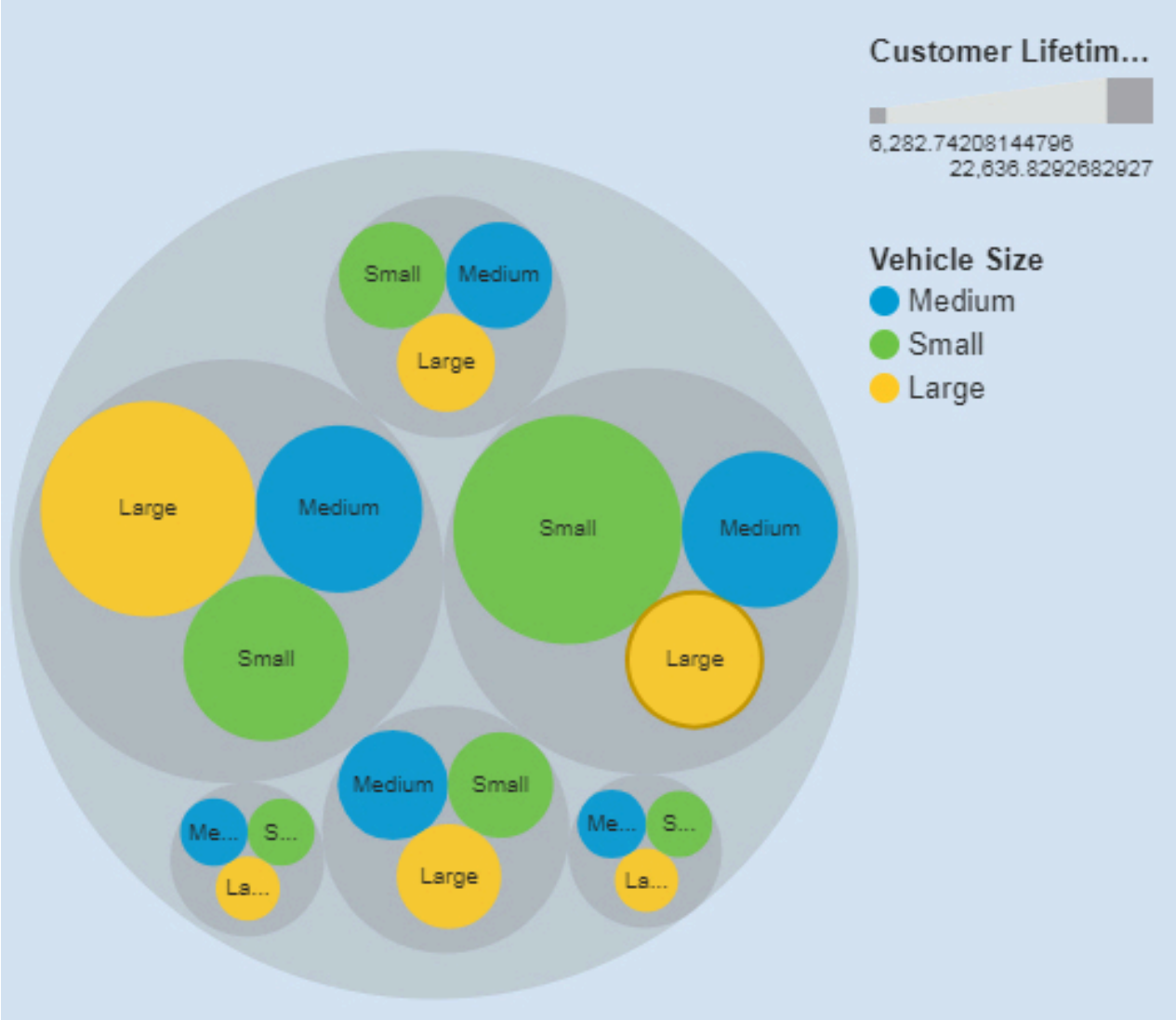
Hierarchy bubble

Use a hierarchy bubble visualization when you want to show relationships among columns that contain values, such as net loss. It is similar to the bubble visualization but the bubbles are tightly packed instead of spread over a grid. The bubbles use nesting to represent the hierarchy. A hierarchy bubble visualization shows a large amount of data in a small space.

The size of each bubble shows a quantitative dimension of each data point. It shows many levels within a hierarchy and relationships between groups based on assigned attributes. It uses bubble size and color to convey comparative information about categories.

The bubbles are in different sizes and colors.

For example, this hierarchy bubble visualization shows customer lifetime value by vehicle class per vehicle size. Each bubble is a different vehicle class in one of the three vehicle size. The size of each bubble is determined by the customer lifetime value of that vehicle class. The colors of the bubbles are determined by the vehicle size.



The hierarchical packed bubble visualization was created by dragging the following data items from the Sources panel:

- Drag **Vehicle Class** and **Vehicle Class** onto the **Bubbles** field.
- Drag **Customer Lifetime Value** onto the **Size** field.
- Drag **Vehicle Size** onto the **Color** field

Samples

You can see an example of a word cloud visualization in the sample report **Customer lifetime value analysis**. You can find the sample here: **Team content** > **Samples** > **Reports** > **Customer lifetime value analysis**.

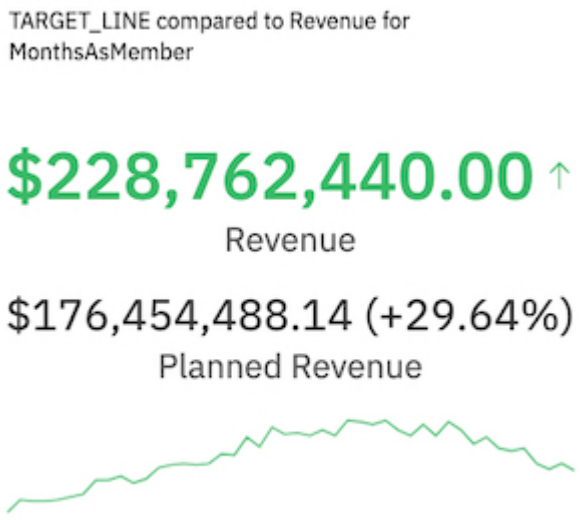
If any of the sample objects are missing, contact your administrator.

KPI

Use a KPI visualization to display a key performance indicator (KPI) that contains two related measures, such as revenue and planned revenue. Optionally, you can display a sparkline and a meaningful shape in your KPI visualizations.

A KPI visualization compares a base value to a target value and shows the variance between the two measures.

For example, this KPI visualization shows the actual revenue in green with an up arrow to indicate that revenue is up compared to the target. In this case, the target value is planned revenue. A sparkline displays the shape of the variation over time and is the same color as the base value.



Create a similar KPI visualization by dragging measures from your own data source to the fields in an empty KPI visualization:

1. On the toolbar, select **New card**.
2. Select the blank **Single** card.
3. Click **Choose a type** and select the **KPI** visualization type. Then, add a data item to each field.
4. Drag a measure onto the **Base value** field. This value is the actual target.
5. Drag a measure onto the **Target value** field.
6. Drag another measure onto the **Time** field. This value creates a sparkline for your KPI visualization. You can add multiple measures, for example Years and Months, to the **Time** field.

Use the properties to customize a KPI visualization. For example, the properties are set by default to display a green conditional color when the target is met and a red conditional color if the target is not met. To display the actual target in another color, under properties, expand the rule and then select a different **Text color**.

Complete the following steps to edit a conditional color rule and select a custom colors:

1. Select the KPI visualization on your exploration.
2. Click the **Properties** tab.
3. Under **Rules**, expand the rule that you want to edit.
4. From **Text color**, select a color.

The following information describes the KPI properties under **Rule style**:

- **Text color**

  Set the color for the value, sparkline, and indicator shape.

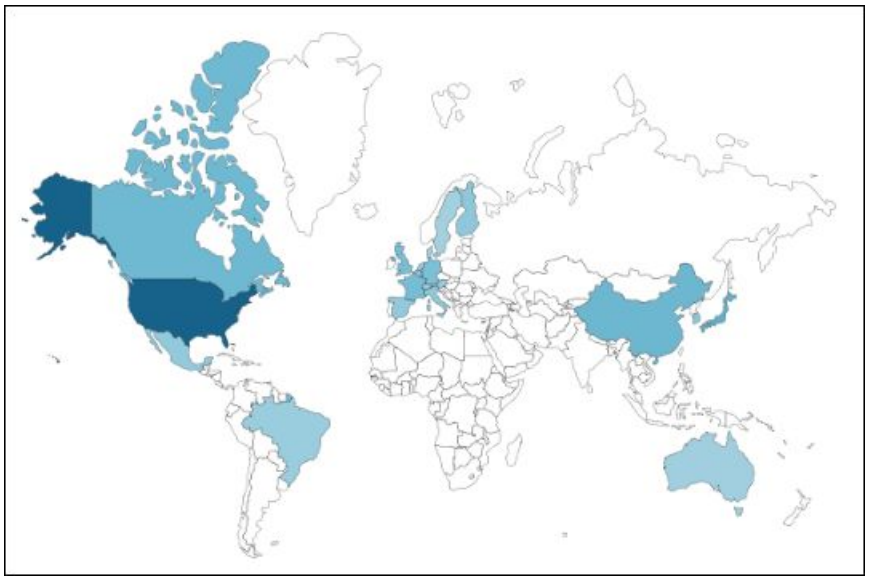
- **Indicator**

  Select a shape to display on the KPI visualization when the when the rule is met. For example, you might want to display a down arrow when your base value falls below a certain threshold compared to the target value.

Legacy map

Use a legacy map when you want to see patterns in your data by geography. You can use a legacy map when you are not connected to the internet.

For example, this legacy map visualization shows revenue by retailer country with the darker color indicating higher revenue.



For more information, see https://www.ibm.com/support/knowledgecenter/SSEP7J_11.1.0/com.ibm.swg.ba.cognos.ug_ca_legacymaps.doc/ug_ca_legacymaps.pdf.

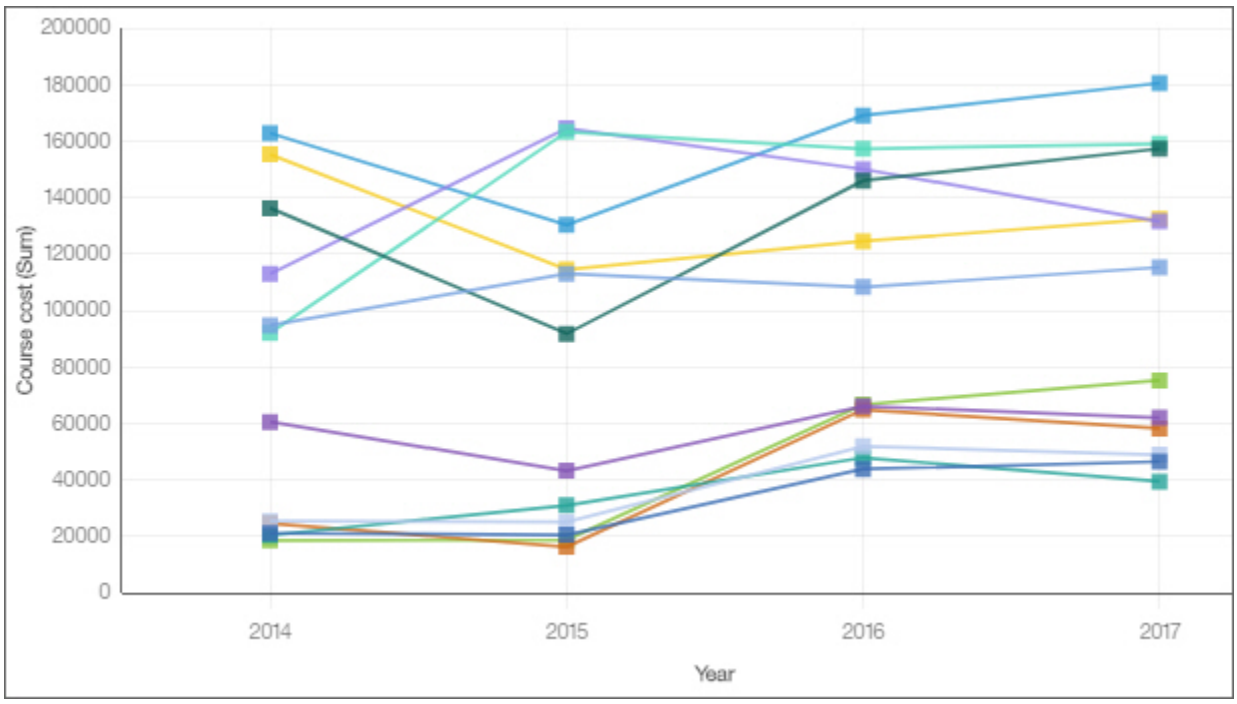
Line

Use a line visualization to show trends over time.

A line visualization can compare trends and cycles, infer relationships between variables, or show how a single variable is performing over time.

For an effective line visualization, use a time column in the x-axis, such as years, quarters, months, or days. If the x-axis shows something else, such as Canada, Netherlands, UK, and US, use a bar or column visualization.

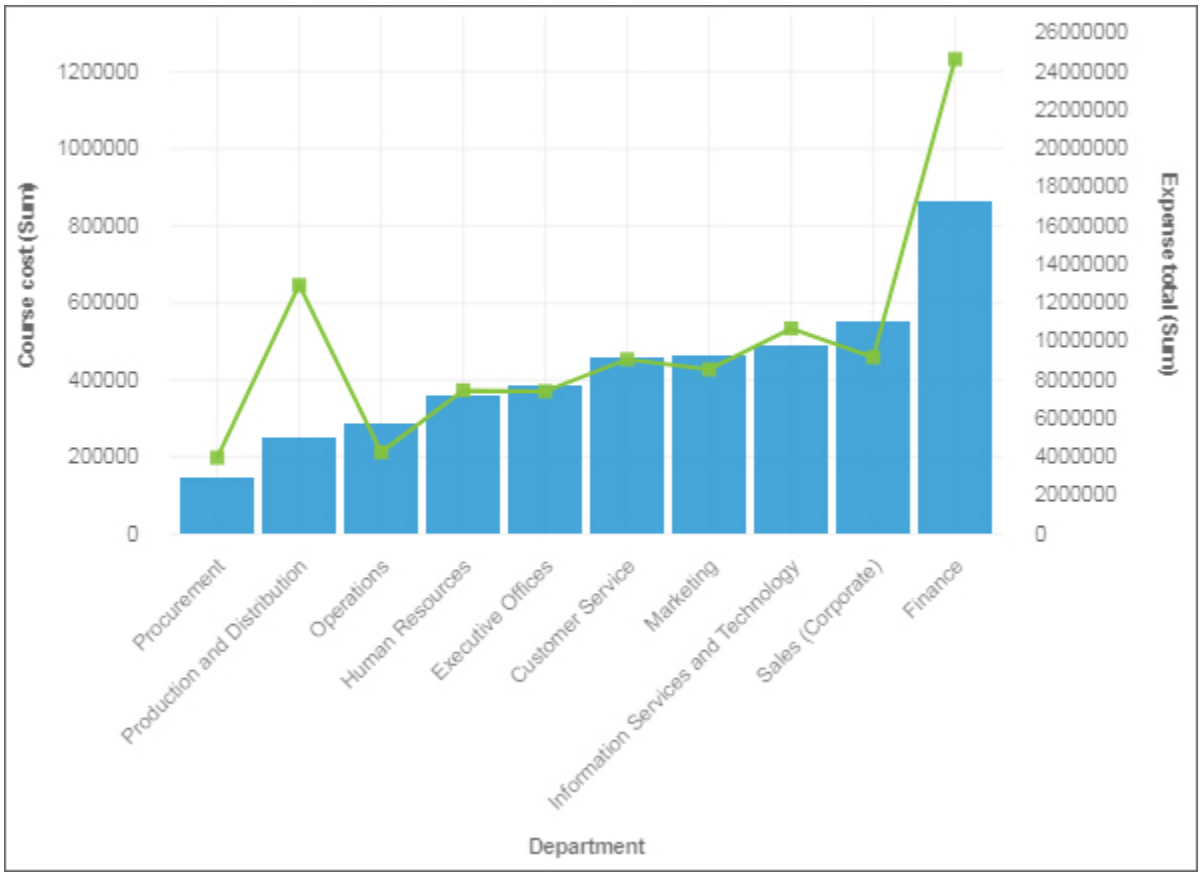
For example, this line visualization shows the trend in course costs by department over year.



Line and column

Use a line and column visualization to highlight relationships between multiple data series by combining bars and lines with one visualization.

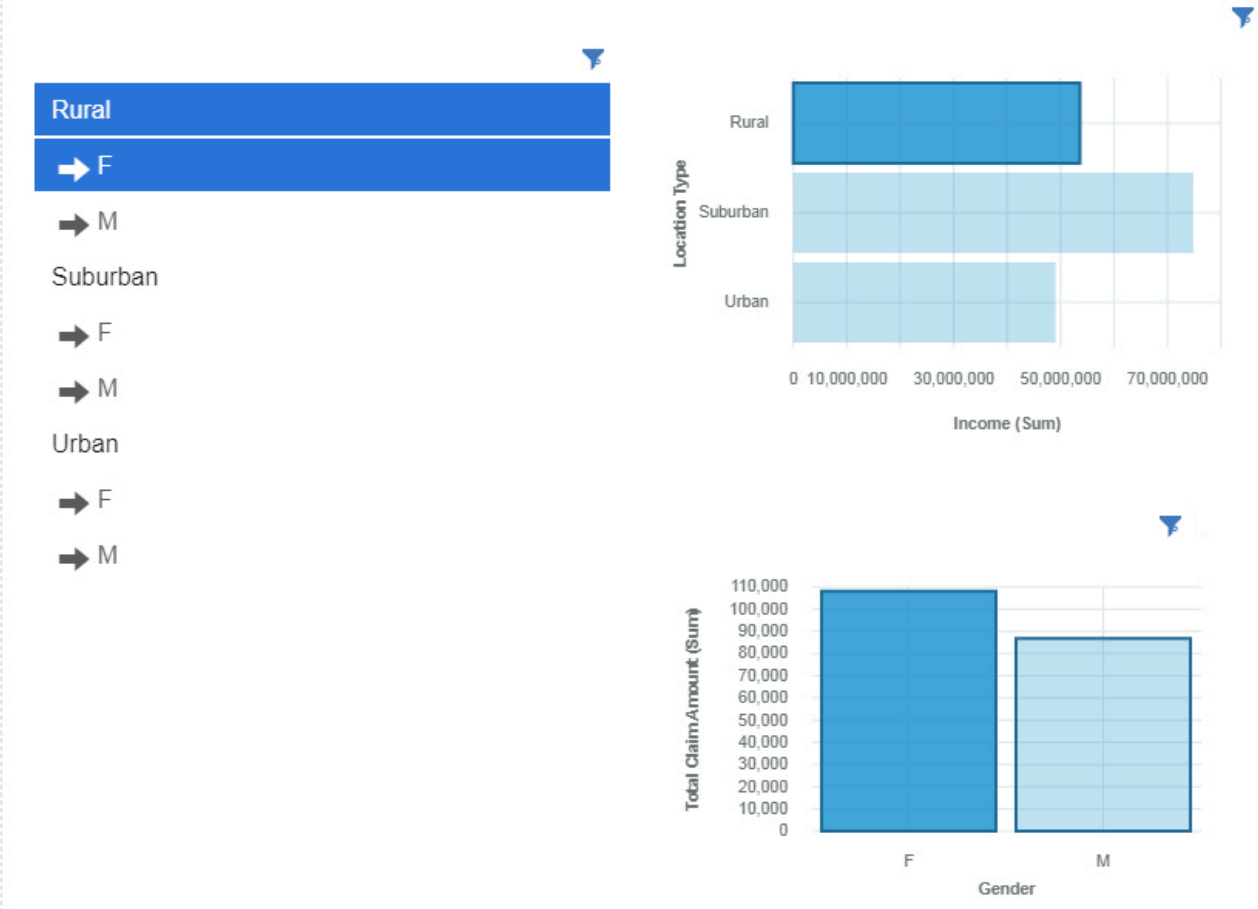
For example, this line and column visualization shows the relationship between course cost and expense totals by department.



List

Use a list visualization to create an overview the data in a hierarchical way.

Another use of the list visualization is to create filter widget. The next example show how you can use the list visualization as a filter widget.



Map

Use a map when you want to see patterns in your data by geography.

Your data asset must contain geographical data, such as countries, states, provinces, or continents.

Note: Maps do not show animations if you set your ease of access system settings to not display animations.

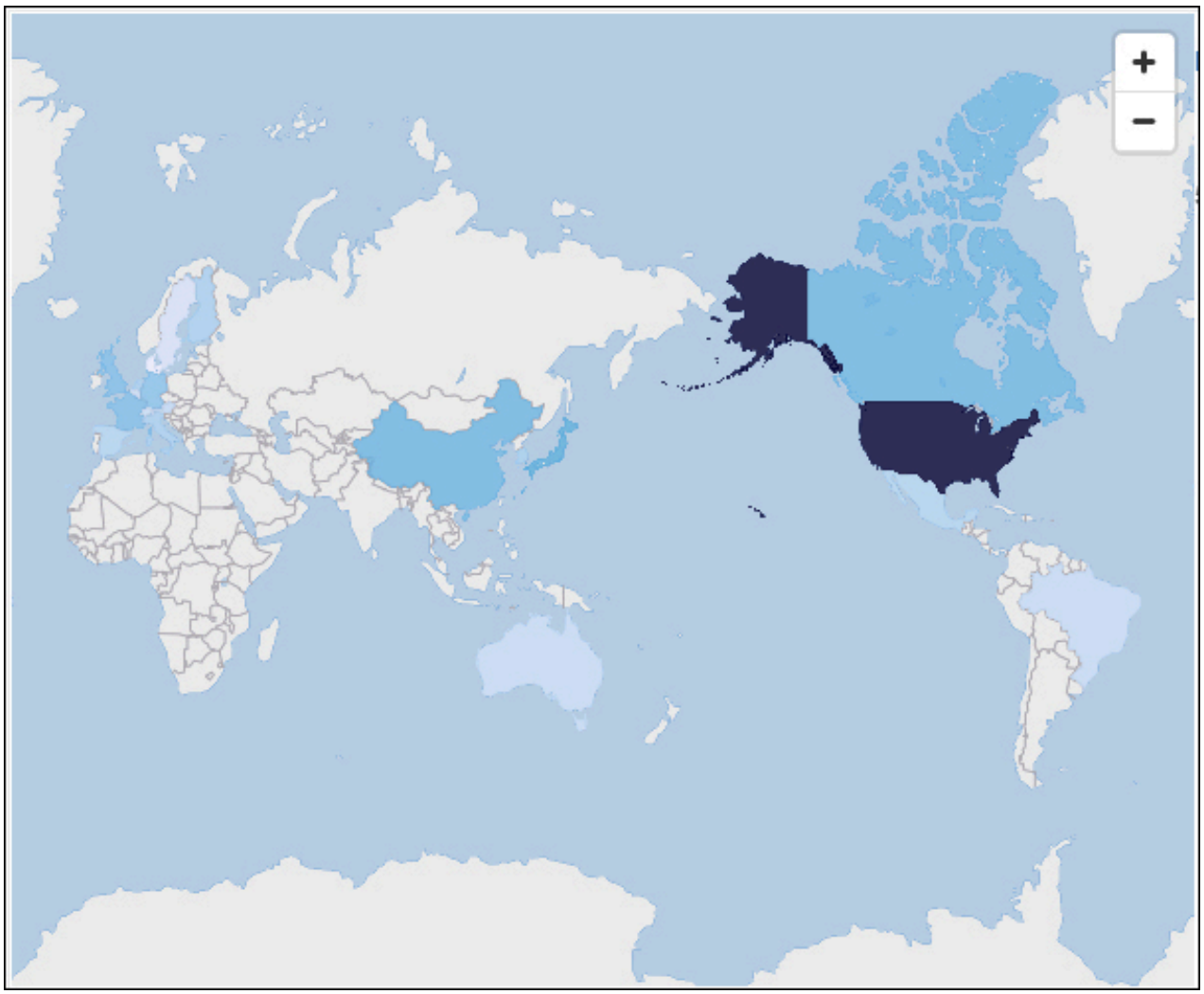
Maps in Cognos Analytics support the following continents:

- North America
- South America
- Africa
- Asia
- Europe
- Antarctica
- Oceania

To determine whether a column is mappable, Cognos Analytics analyzes a sample of 2000 values in the location column, looking for recognizable place names. If 80% or more are recognized as map values, Cognos Analytics produces a map.

For example, you have four countries in your location column: Brazil, China, India, and Russia. The typographical error for India means that only 75% of the values are recognizable place names and you will not see a map as a starting point. But if you have five countries and one has a typographical error in it, you see a map.

For example, this map visualization shows revenue by retailer country with the darker color indicating higher revenue.



Marimekko

A marimekko visualization is similar to a stacked column visualization. It shows data through varying heights and includes an added dimension of data through varying column widths. The width of the columns is based on the value that is assigned to the width field. Individual segment height is a percentage of the respective column total value.

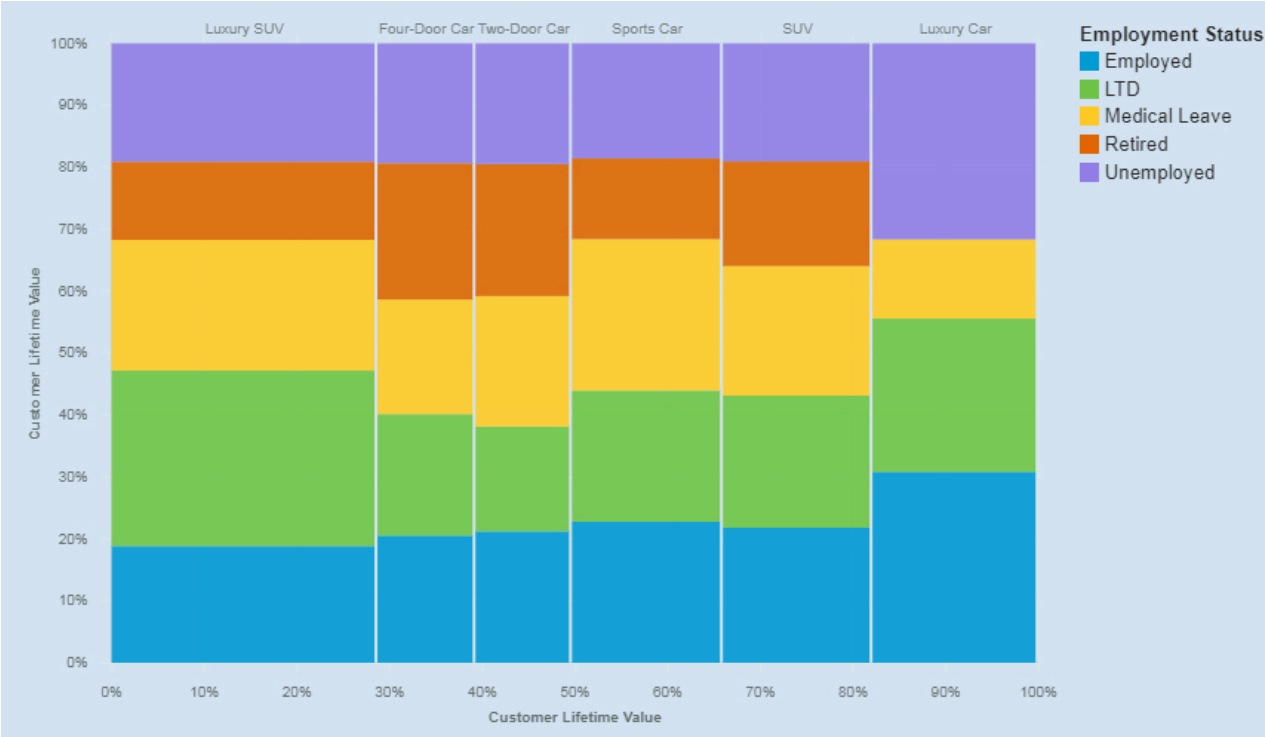
You can quickly spot large segments, such as a specific vertical that has a large share of a region. You can also identify white space such as an under-represented vertical in a specific region.

The marimekko visualization is useful for part-to-whole comparisons, where you need to show an extra measure/variable.

The marimekko visualization allows data to be depicted along two dimensions simultaneously. For example, market segments are often arrayed along the x-axis, with the width of each column corresponding to the financial value of a segment. You use marimekko visualizations in cases, for example, where you want to show the revenue contribution per product line. Or the gross domestic product per country.

The marimekko visualization can display total or partial number. If you want to use stacked percentages instead of number, then use the **Display as stacked percentage chart** option.

The following example shows the contribution of customer lifetime value and employment status in different vehicle classes with the option **Display as stacked percentage chart** enabled.



The marimekko visualization was created by dragging the following data items from the Sources panel:

- Drag **Vehicle Class** type onto the **Bars** field.
- Drag **Customer Lifetime Value** onto the **Length** field.
- Drag **Employment Status** onto the **Color** field

Samples

You can see an example of a word cloud visualization in the sample report **Customer lifetime value analysis**. You can find the sample here: **Team content** > **Samples** > **Reports** > **Customer lifetime value analysis**.

If any of the sample objects are missing, contact your administrator.

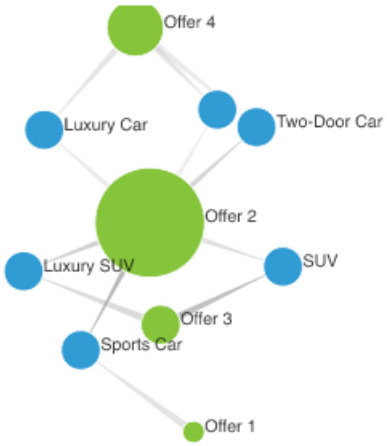
Network

Use a network visualization when you want to see the connections among columns in your data asset. A network visualization is a good choice to show connections, networks, and points of intersection.

Network visualizations display a set of nodes, represented by symbols, and links, represented by paths, to show the relationship between entities or items.

Use the **From** and **To** fields to define the relationship that you want to investigate.

For example, a network visualization can show offer acceptance by Vehicle Class.



Create the Network visualization by dragging the following data items from the **Offers** section in the **Sources** pane  :

- Drag **Offer** onto the **From** field.
- Drag **Vehicle Class** onto the **To** field.
- Drag **Accepted** onto the **Line width** field.

Next, set the size and node properties.

1. Click the visualization, then click . Set set the following options in the **Properties** pane:
   - **Size - Width**: 500 px, **Height**: 300 px
   - **Nodes minimum size**: 20
   - **Nodes maximum size**: 100
2. Click to close the **Properties** pane.

Samples

You can see examples of visualizations in the sample report **Customer lifetime value analysis**. You can find the samples here: **Team content** > **Samples** > **Reports** > **Customer lifetime value analysis**.

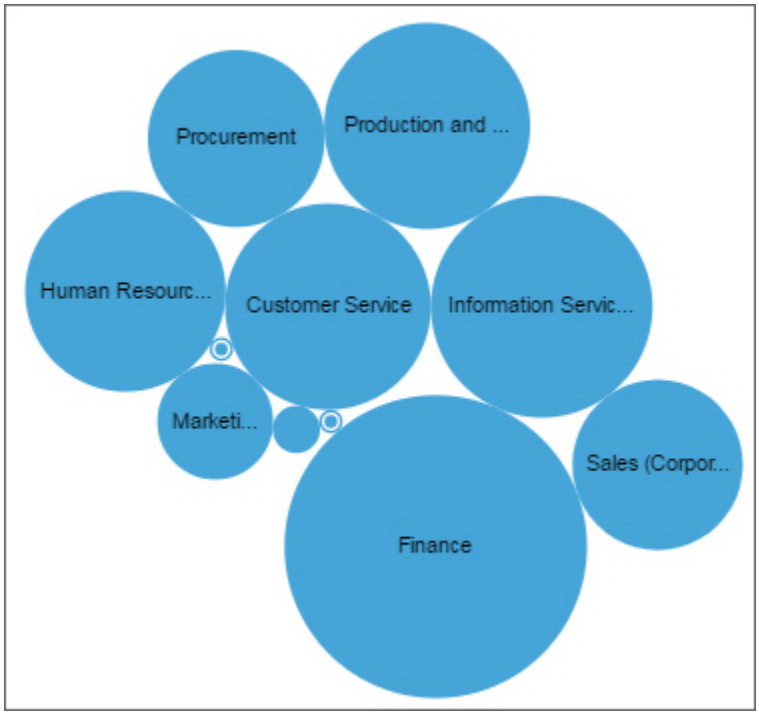
If any of the sample objects are missing, contact your administrator.

Packed bubble

Use a packed bubble visualization when you want to show relationships among columns that contain numeric values, such as revenue. It is similar to the bubble visualization but the bubbles are tightly packed instead of spread over a grid. A packed bubble visualization shows a large amount of data in a small space.

The bubbles are in different sizes and colors.

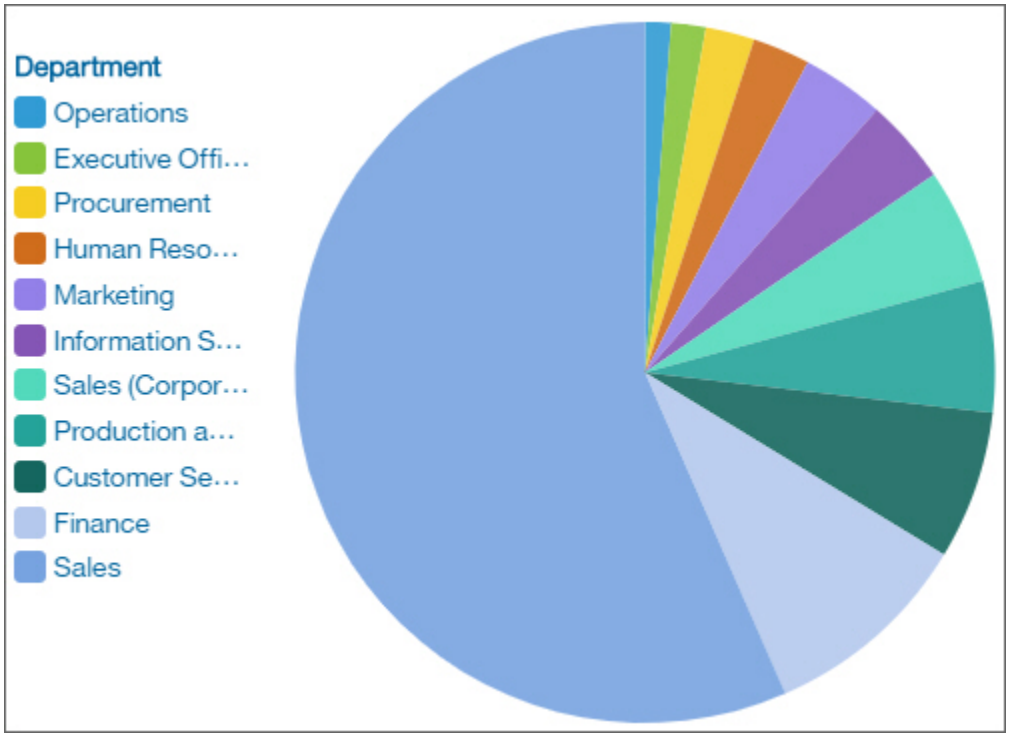
For example, this packed bubble visualization shows external hires by department. Each bubble is a different department. The size of each bubble is determined by the number of external hires for that department.



Pie

Use a pie visualization to highlight proportions. Each slice shows the relative relationship of each part to the whole.

For example, this pie visualization shows the number of course days for each department.



Point

Use a point visualization to show trends over time.

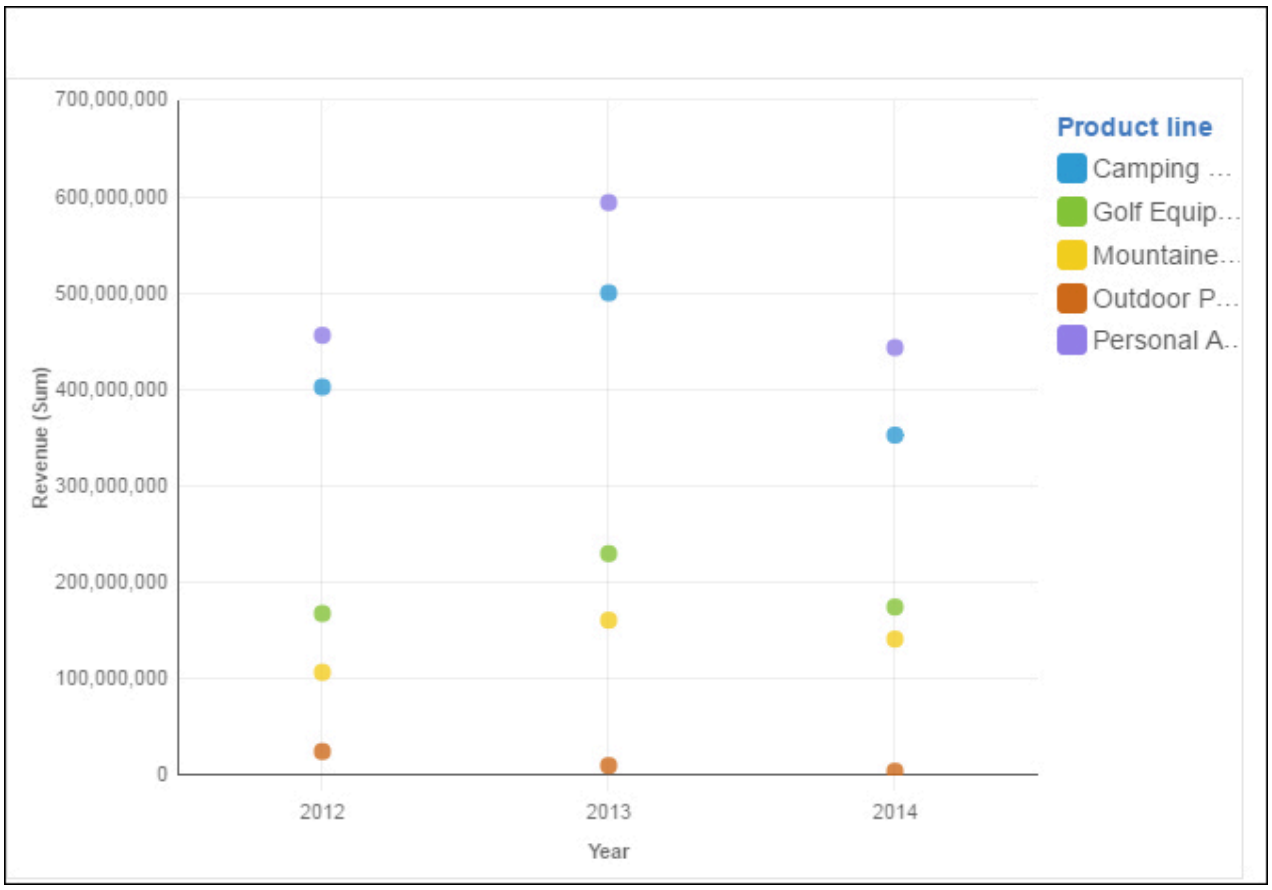
A point visualization can compare trends and cycles, infer relationships between variables, or show how a single variable is performing over time.

A point visualization is like a line chart without the connecting lines.

For an effective line visualization, the x-axis should show time, such as years, quarters, months, or days. If the x-axis shows something else, such as Canada, Netherlands, UK, and US, use a bar visualization.

Data values are plotted vertically.

For example, this line visualization shows revenue over quarter by order method type. Web orders have grown dramatically over this period.

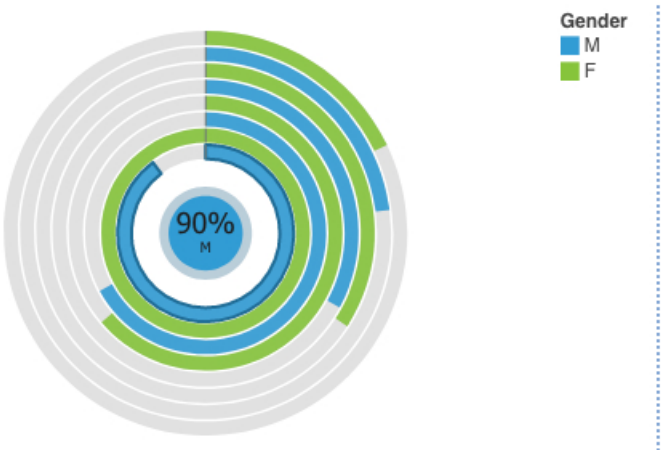


Radial

In a radial visualization, each bar appears in a circle with longer bars that represent larger values. Hover over a bar to see the details about it, such as the exact value represented by the bar. Each bar starts at 12 noon and goes in a clockwise direction for positive values and counterclockwise for negative values.

Radial visualizations, also known as dial charts or speedometer charts, show information as reading on a dial. The radial visualization is valid only with one category.

For example, this visualization shows renewals by offer type and gender.



Create the Radial visualization by dragging the following data items from the **Customer Analysis** section in the **Sources** pane :

- Drag **Renew Offer Type** onto the **Bars** field.
- Drag **Number of Policies** onto the **Length** field.
- Drag **Gender** onto the **Color** field.

The next step is to set the sort properties for **Renew Offer Type** and **Gender**.

1. Click the visualization, then in the **Data** pane, click the **<Renew Offer Type>** data item.
2. Click
3. In the **Properties** pane, for **Sort order**, select **Ascending**.
4. In the **Data** pane, click the **<Gender>** data item.
5. In the **Properties** pane, for **Sort order**, select **Descending**.
6. Click to close the **Properties** pane.

Samples

You can see examples of visualizations in the sample report **Customer lifetime value analysis**. You can find the samples here: **Team content** > **Samples** > **Reports** > **Customer lifetime value analysis**.

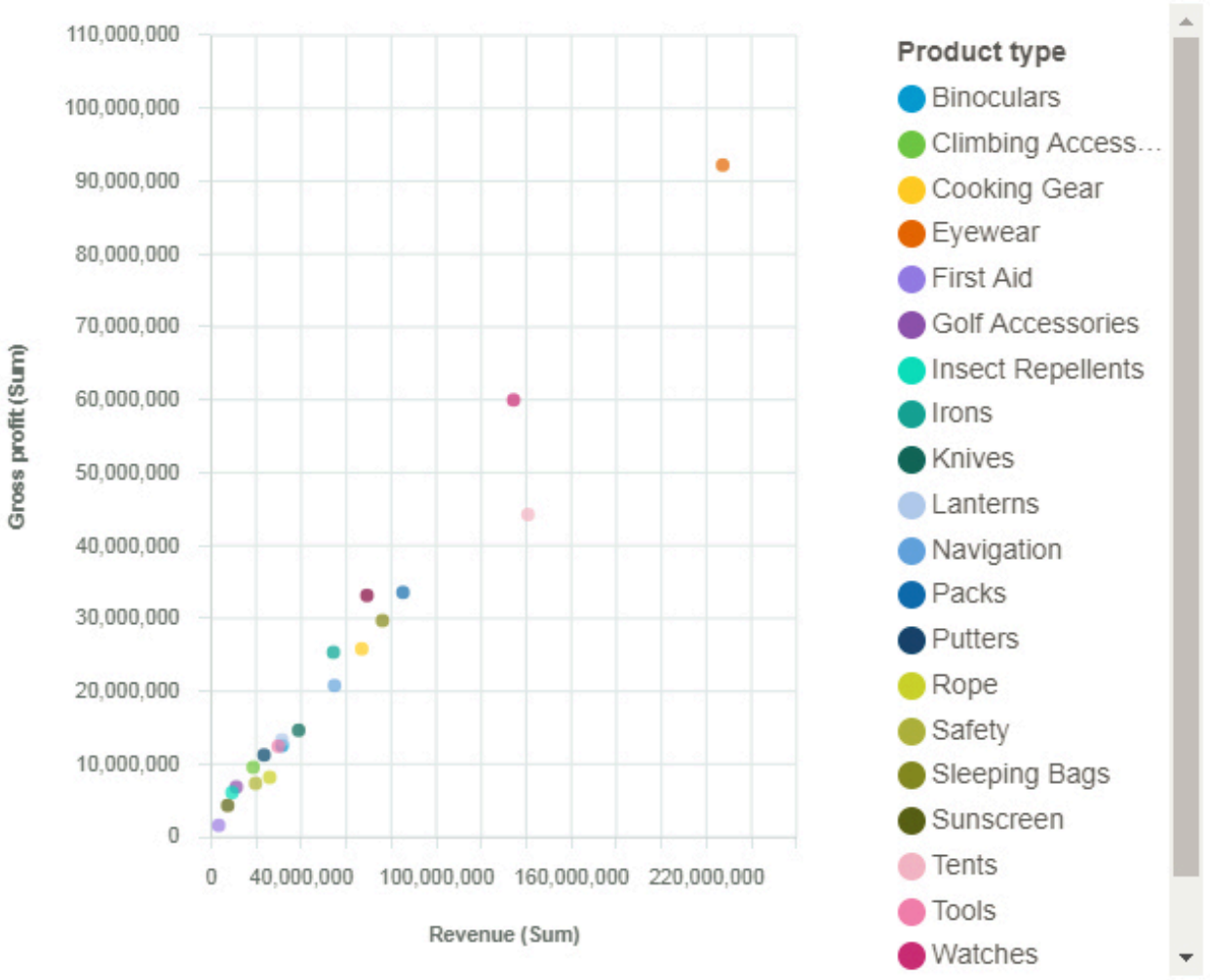
If any of the sample objects are missing, contact your administrator.

Scatter

Scatter visualizations use data points to plot two measures anywhere along a scale, not only at regular tick marks.

Scatter visualizations are useful for exploring correlations between different sets of data.

The following example shows the correlation between revenue and gross profit for each product type.



Spiral

A spiral visualization shows you the key drivers, or predictors, for a given target. The closer the driver is to the center, the stronger that driver is.

IBM Cognos Analytics uses sophisticated algorithms to deliver highly interpretable insights that are based on complex modeling. You don't have to know which statistical tests to run on your data. Cognos Analytics picks the right tests for the data.

Key drivers for both continuous and categorical targets are available in the spiral visualization in dashboards and explorations.

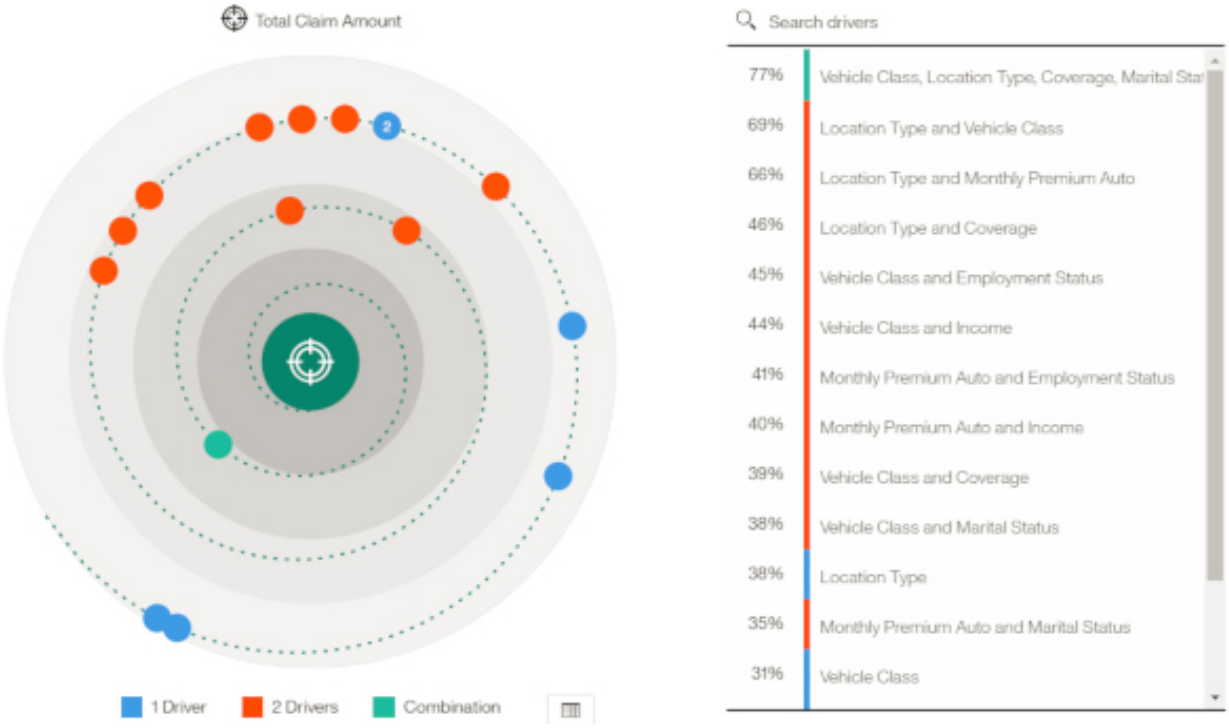
For more information, see *Statistical tests* documentation in the *IBM Cognos Analytics Dashboards and Stories User Guide*.

For example, this spiral visualization shows that the combination of vehicle class, location type, coverage, marital status, and employment status are the strongest drivers of the target, total claim amount.

You can exclude drivers from the analysis. Right-click on a driver and click the **Edit drivers** icon . Select the drivers that you want to include in the analysis.

To edit or add key drivers, click the on the target data slot.

To improve performance, due to number of rows in the data source, the analysis is based on a representative sample of the entire data.

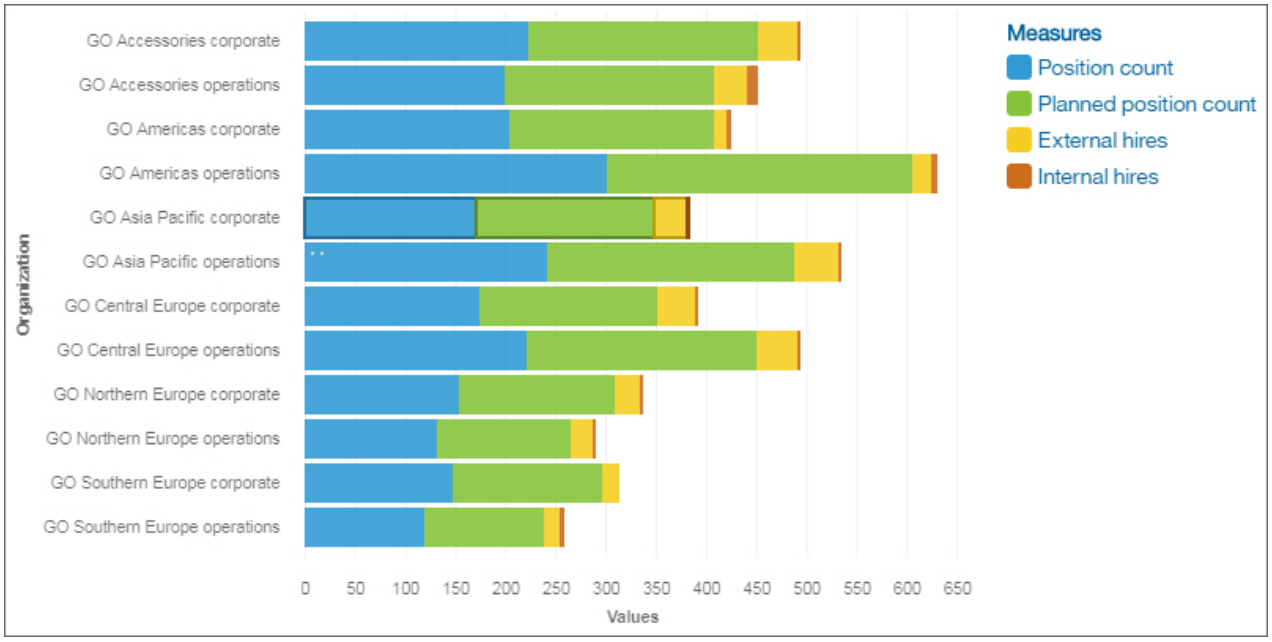


Note: Filters are not supported for spiral visualizations.

Stacked bar

Use a stacked bar visualization to compare the proportional contributions for each item to the total, such as sales for products and sales for products each month.

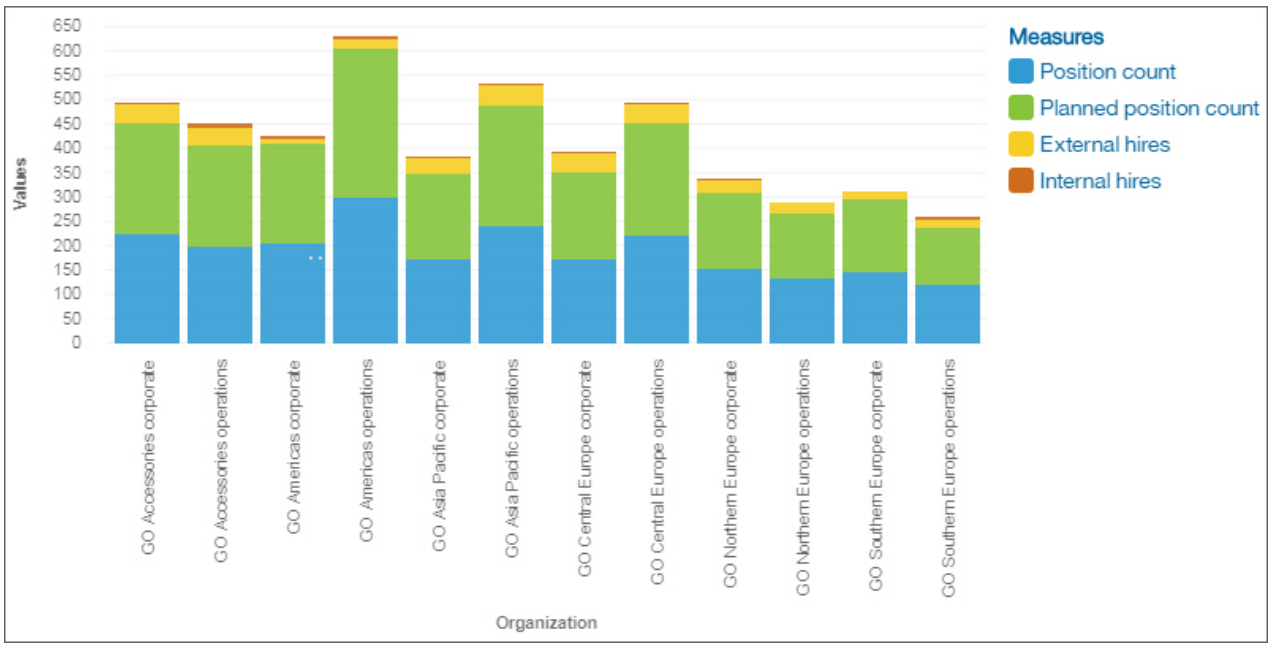
A stacked bar visualization can show change over a specific time period or compare the proportional contributions for each item to the total. If there are so many bars that the labels are impossible to read, filter the data to focus on a subset of the data or use a tree map.



Stacked column

Use a stacked column visualization to compare the proportional contributions for each item to the total, such as sales for products and sales for products each month.

A stacked column visualization can show change over a specific time period or can compare the proportional contributions for each item to the total. If there are so many bars that the labels are impossible to read, filter the data to focus on a subset of the data or use a tree map.



Summary

Use a summary visualization when you want to see the total for a measure or the count for a categorical column.

For example, this summary visualization shows total revenue for all product types.

For example, this summary visualization shows the number of departments in your organization.

Department	Course cost
Customer Service	459,250
Executive Offices	384,000
Finance	863,750
Human Resources	361,250
Information Services and Technology	491,750
Marketing	465,750
Operations	288,000
Procurement	146,250
Production and Distribution	249,450
Sales	4,056,000
Sales (Corporate)	549,550
Summary	8,315,000

Adding more columns to a table

You can focus on points that are of interest to you by adding more data to the visualization.

1. Drag another column to the field where you want additional data.
2. Drop the column beside the existing column.

Starting from Cognos Analytics version 11.1.4, you can drag data from the **Selected sources** pane and insert data in a column/row or drop the data on top of existing data to replace it.

Treemap

Use a tree map visualization to identify patterns and exceptions in a large, complex data asset.

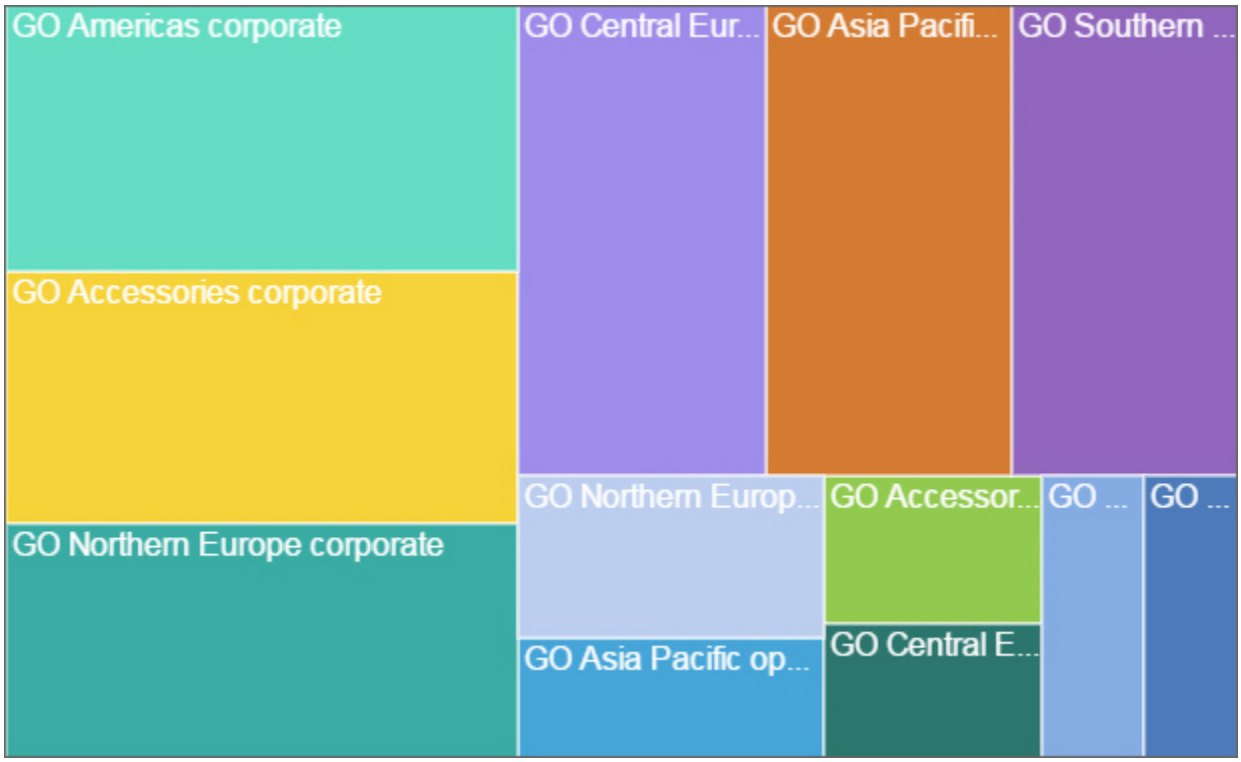
Treemaps show relationships among large numbers of components by using size and color coding in a set of nested rectangles.

A treemap that is colored by category identifies the level 1 category by color. The sizes of the rectangles represent the values. In a treemap that is colored by value, the sizes of the rectangles represent one of the values and the color represents a second set of values. Do not use data that includes negative numbers. A treemap ignores negative numbers.

Many data assets have a hierarchical structure. For example, you have data about the profit margin of food items in a grocery store. Under the general category of fruit, there is a category for citrus fruit. Various

citrus fruits are listed, such as grapefruit, orange, and lemon. A treemap tells you how each citrus fruit is performing when compared to each other and to other types of food.

For example, this treemap visualization shows course cost by organization.



To deselect a box that you've selected, Ctrl+click the selected box.

Waterfall

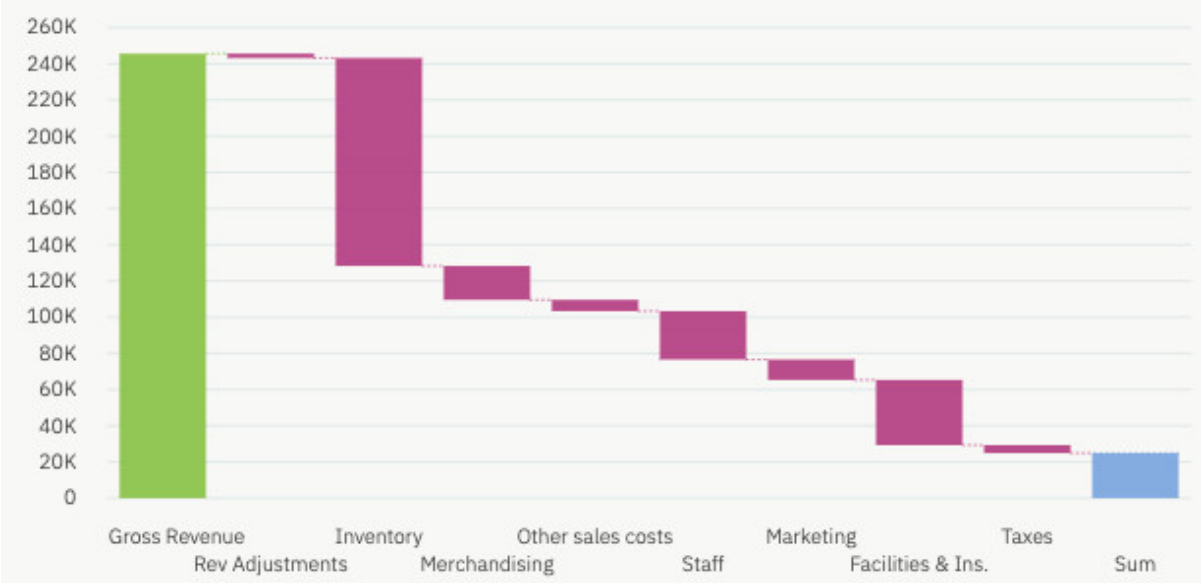
Use a waterfall visualization to understand the cumulative effect a series of positive and negative values have on an initial value. The bars in a waterfall visualization are not totals.

A waterfall visualization shows how an initial value is increased and decreased by a series of intermediate values, leading to a final cumulative value shown in the far right column. The intermediate values can either be time-based or category-based.

Some examples of waterfall visualizations are as follows:

- Viewing the net income after you add the increases and decreases of revenue and costs for an enterprise over a quarter.
- Cumulative sales for products across a year with an annual total.

This waterfall visualization shows the policy holder delta by month.



Creating a waterfall visualization

1. Create a new exploration. For more information, see “Starting a new exploration from the New menu” on page 1.
2. Open the sample data module: **Select a source** > **Team content** > **Samples** > **Data** > **Customer analysis**.
3. Click **Visualizations** and click **Waterfall** to add the waterfall visualization to the exploration.
4. Click **Sources**

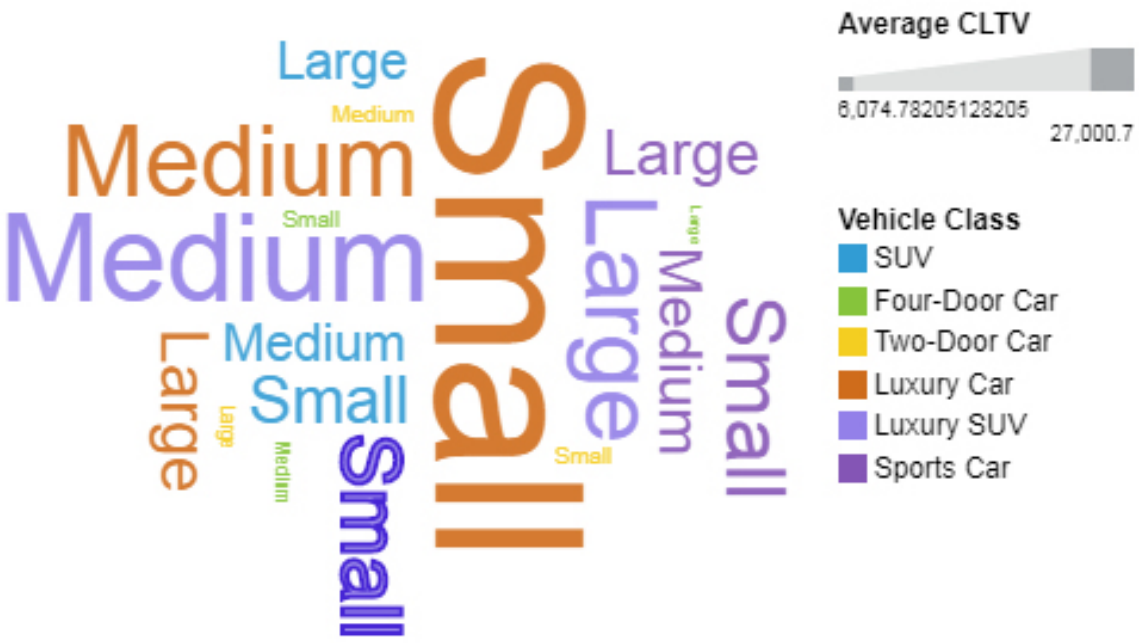
   
5. Drag the following data items from the Policy Holders section:
   - Drag **Month** onto the **x-axis**.
   - Drag **Delta** onto the **y-axis**.

Word cloud

Use a word cloud visualization when you want to see a text-based visualization of a column. The text height represents the scale. The name itself is the different members of the column.

Tip: The data asset should contain at least 15 columns and at least 100 rows to create an effective word cloud.

For example, this word cloud visualization shows the customer life time value by vehicle size and class.



The word cloud was created by dragging the following data items from the Sources panel:

- Drag **Vehicle size** type onto the **Words** field.
- Drag **Average CLTV** onto the **Size** field.
- Drag **Vehicle class** onto the **Color** field

Samples

You can see an example of a word cloud visualization in the sample report **Customer lifetime value analysis**. You can find the sample here: **Team content** > **Samples** > **Reports** > **Customer lifetime value analysis**.

If any of the sample objects are missing, contact your administrator.

Insights in visualizations

IBM Cognos Analytics provides analytic insights that help you to detect and validate important relationships and meaningful differences based on the data that is presented by the visualization.

Insights are available by clicking the **Insights** icon of eligible visualizations. When you turn on insights, the summary appears in the **Insights** box and related visualization elements are highlighted and details are provided in the corresponding tooltip message. You can control each available insight separately.

Procedure

1. In a visualization that supports insights, click the **Insights** icon.
2. Depending on the visualization, the following insights are shown:
   - **Average** Provides the mean of the displayed target value.
   - **Predictive strength** Shows the predictive strength of the relationship between the target and explanatory fields.
   - **Fit line** Shows when either a linear or quadratic relationship exists between the target and explanatory fields.

- **Meaningful differences** Shows values that are most significantly higher or less than the average or trend.
- **Most frequent** Shows values that are most frequently reported.

Choosing correlated insights

Based on your visualization you are presented with statistically based, correlated, insights.

About this task

If correlated insights are available that are related to the main visualization, a green icon with a number 

is shown on either the x-axis or y-axis. The number indicates the available correlated insights.

To access correlated visualizations, complete the following actions:

Procedure

1. From the visualization, click the green icon 

.
2. Click any of the statistically based insights that are presented from the menu.
   A new card is created.

Choosing recommended visualizations

Recommended visualizations are thumbnails that display visualizations that might be appropriate for your data.

Procedure

1. From the **Cards** pane, select the card that represents the visualization you want to open.
2. Click **Choose visualization type**  in the toolbar.
   The recommended visualizations are displayed.
3. Click the thumbnail for the recommended visualization that you want to work with.

Choosing related visualizations

When a visualization is in focus in your exploration, the system recommends some related visualizations that are not specifically what you requested. Based on the data analysis, these related visualizations might be of interest to you.

About this task

Related visualizations replace one of the data elements in the visualization or add another data element to create a new visualization. Related visualizations use a combination of learned user interactions, statistics and interestingness to suggest useful next steps.

To access related visualizations, complete the following actions:

Procedure

1. From the **Cards** pane, select the card that represents the visualization you want to open.
2. Click **Related**  in the toolbar.

Chapter 5. Principles of advanced data analytics

Principles of advanced data analytics

IBM Cognos Analytics is a business intelligence tool for managing and analyzing data. It includes self-service features for users to prepare, explore, and share data. Cognos Analytics includes predictive, descriptive, and exploratory techniques, also known as numeric intelligence. Cognos Analytics uses many statistical tests to analyze your data.

It is important to understand the definitions of these tests as they apply to Cognos Analytics.

Numeric algorithms are used as part of the workflow to provide features to the user that communicate information about the numeric properties and relationships in their data.

Business oriented

Unlike traditional statistical software, where the target audience is an experienced data analyst, the algorithms of Cognos Analytics are aimed at users who are familiar with, but not an expert in data analysis. This means that when Cognos Analytics considers tradeoffs, usefulness is chosen over complexity.

Trustworthy

Business data is lot more complicated than text-book examples that are used in statistical courses or in web searches and examples. Cognos Analytics uses algorithms that are robust and that are able to cope with a range of varieties of unusual data. Cognos Analytics does this because although the algorithms that are more brittle are able to get slightly better results than robust algorithms, they require you to make sure that they are applicable and build correct data transformations for the results to be meaningful. A minor drop in accuracy is worth the safety that is provided by an algorithm, which does not give wrong results when the data is not as it is supposed to be.

Intelligent

Almost all algorithms require decisions to be made about them; levels of confidence, which combinations of fields to explore, data transformations. The details of these decisions can be found in the descriptions.

Cognos Analytics chooses appropriate values automatically, by examining properties of the data. As a user you might not discover all the decisions that are made.

Summary

In Cognos Analytics, the numeric algorithms and procedures are designed to produce trustworthy results automatically. To get the best possible prediction, classification or analysis, a professional statistician analyzes the data by using IBM SPSS Statistics or IBM SPSS Modeler. The goal of Cognos Analytics is to provide qualitative insights that help you to understand your data and its relationships, and to do so automatically for a wide variety of types of data. Cognos Analytics aims at providing results similar to a professional statistician without getting in the way of the business user.

Data preparation

Data preparation is a pre-analysis step that is used by most data analytic algorithms to ensure that the data is suitable for analytic use.

Overview

Data preparation is critical in IBM Cognos Analytics. Only prepared data is entered into analysis for key drivers, decision trees, and relationships that are displayed in the advanced analytics visualizations: Spiral, Driver analysis, Decision tree, Sunburst, and Explore relationships. Data is not automatically prepared for other visualizations and their corresponding insights.

Algorithms

All applied algorithms are based on values of a single field at a time. Missing values are removed or handled for each field, all numeric predictor driver fields are binned. All categorical fields are adjusted for large number of categories and outliers are handled in the target field. While all data preparation influences the analysis results, corresponding data preparation summaries are not currently reported to you.

Details

Data preparation and subsequent key drivers, decision trees and relationships are based on a data sample with approximately 10,000 rows when the original data is larger. Bernoulli random sampling, equal probability without replacement random sampling, is applied to uploaded data and any connected data sources that are supporting random sampling. Otherwise, systematic sampling is used.

Data preparation for numeric fields

A field is treated as numeric whenever it contains numeric information and its usage property is set to measure.

Overview

Because numeric data can be varied in their distribution, IBM Cognos Analytics transforms non-target numeric fields into ordinal bins, reducing the dependence of analytic algorithms on the format of numeric data.

Algorithms

The basic algorithm that is used is equal frequency binning. Numeric data is divided into a fixed number of bins that are attempting to put an equal number of rows of data into each bin. Missing values are placed in their own bin. Cognos Analytics attempts to use knowledge about missing values in predictor fields to build a better model. For example, if a field of data represents when an item was tested, Cognos Analytics uses missing values (which might represent that an item was never tested) to help predict the values of other fields.

Details

Certain field exclusion criteria apply to numeric fields. A numeric field is excluded from further analysis if it has only a single value, including the missing value. Otherwise, the numeric field is binned and the default number of bins is 5. If a field has no more than 10 unique numeric values, then binning is not attempted, and each unique value is given its own category. If zero occurs in more than 40% of rows, it is always given a separate category. Missing values are placed in their own bin and do not affect the binning procedure.

Data preparation for categorical fields

A field is treated as categorical whenever its usage property is set to attribute or identifier.

Overview

The main information that is extracted from categorical fields is observed frequency for each unique category value. Appropriate analytic methods are applied to categorical fields, but their accuracy and performance can be adversely affected when the number of different categories becomes large. The main data preparation step is to start merging categories when their number becomes large.

Algorithms

The basic algorithm that is used is merging categories. Categories are sorted by their frequency in descending order and the categories beyond default number are merged in a single category. Missing values are treated as a single separate category. In other words, IBM Cognos Analytics uses missing values in a similar way as for the numeric fields. Categorical fields are treated as nominal. Intrinsic order is not assumed among categories.

Details

Certain field exclusion criteria apply to categorical fields. A categorical field is excluded from further analysis if it has only a single value or the number of unique, non-merged categories exceeds 50% of the number of valid data rows.

Otherwise, the categorical field is merged and the default number of non-merged categories is 49. The rest of the categories are merged into a single extra category. All categories with row count smaller than 3 also get merged. A categorical field is also excluded if the percentage of valid data rows corresponding to the merged category exceeds 25%.

Missing values are treated as a separate category and considered in the merging step as such.

Data preparation for target fields

Specification of the target field is required for key drivers and decision tree visualizations.

Overview

Always specify the target field and at least one extra field. Models are trained by using supplied target values and are used to detect predictive relationships and eventually to predict target values given the input field values. Data preparation for the target field differs from the data preparation for the rest of the fields. Missing values in the target are not used for building models, but the rest of the information is preserved and sometimes adjusted to obtain unbiased models.

Algorithms

The main data preparation step related to target fields is removal of all data rows with missing target value. This happens before any other data preparation steps. While it ensures that only reliable information is used for model building, the number of removed rows can be substantial. The resulting model might have a limited scope in such instances. Numeric target fields are not binned, but the extreme outliers are handled to not adversely affect the later created models. Categorical target fields are treated much like other categorical fields. The only difference is that missing values have been removed for the categorical targets.

Details

Extreme outliers are detected based on lower and upper boundaries. The upper boundary is constructed by using an upper percentile such that only 2.5% percent of target values are found to have a greater value. The difference between the upper percentile and the median is multiplied by 2.5 and added to the

median to obtain the upper boundary. Similar steps are applied to obtain the lower boundary. The target values that are found beyond the computed boundaries are replaced by the corresponding boundary value in all subsequent analysis.

One-way key drivers

One-way drivers are a model-based exploratory tool.

Overview

Given a target field, the tool uses a statistical model to analyze any other available data field and estimates its strength in predicting the target values. Such data fields are called target predictors or drivers. Each potentially relevant data field is analyzed and only the top drivers regarding their predictive strength are displayed. You obtain insights regarding available drivers and their ranking according to their predictive strength for the specified target in the data. One-way driver analysis results are available both in the driver analysis and spiral visualizations. Visual drill-down into each separate driver is enabled for driver analysis visualization in Explore only.

Algorithms

Analysis for each one-way driver is based on a statistical model that includes the target and a single categorical predictor. The model is applied after the data preparation step for the target field and all potential predictor fields. For example, all numeric predictor fields are binned during the data preparation step and treated as categorical in the analysis. One-way ANOVA is applied for numeric targets and Chi-square test of independence is applied for categorical targets with the chi-square adjustment for sparse data.

For each field in the list of potential drivers, a hypothesis test on whether the field has a significant impact on the target is performed. Only those fields which pass the test and have sufficiently high predictive strength are selected as possible one way key drivers.

Details

Preliminary analysis based on smarts capabilities reduces the number of potential drivers in some cases. The goal is to remove irrelevant or redundant fields. The list of used drivers is available in the UI and you can add any initially excluded drivers to the analysis. The top 20 resulting drivers with predictive strength higher than 10% are available for display.

Some restrictions are enforced on the size of the data to improve performance and speed. If the data contains more than 250 fields, the least relevant fields are excluded before driver analysis. You can add the excluded fields back into the analysis through the UI as described above. If specified data contains more than 10,000 rows, it might be sampled down to approximately 10,000 rows for purpose of driver analysis. A warning is displayed in such instances: *To improve performance, due to the number of rows in the data source, the analysis is based on a representative sample of the entire data.* The results are expected to closely approximate results that would be obtained by using all the rows in the original data.

Two-way key drivers

Two-way drivers rely on modeling and ranking pairs of categorical predictors at one time.

Overview

Given a target field, IBM Cognos Analytics uses a statistical model for analysis of a pair of other data fields and estimates its strength in predicting the target values. Search over different predictor pairs is usually not exhaustive and also some high-ranking pairs can be filtered out from the final results. The goal is to provide an overview and variety of predictor pairs that improve upon predictive strength of a single predictor models that are displayed as one-way drivers. Therefore, the insights obtained from one-way drivers are expanded and the user obtains relevant information on the pairs of fields in the data. Both one way driver and two-way driver analysis results are available in the driver analysis and spiral charts. They

can be viewed separately by selecting a corresponding chart viewing option. Each displayed one-way or two-way driver can be expanded into a new visualization directly from the Driver analysis visualization in Explore.

Algorithms

Analysis for each two-way driver is based on a statistical model that includes the target and a pair of categorical predictors. The model is applied after data preparation and building all the one-way drivers. The first predictor in the pair is selected from the top 50 one-way drivers and the second is selected from the top 25 one-way drivers. This search strategy ensures that most of the top-ranking predictor pairs would be considered for modeling. The two-way ANOVA (analysis of variance) analysis is applied for numeric targets and Chi-square test of independence is applied for categorical targets with the chi-square adjustment for sparse data.

For each considered pair of fields, a hypothesis test on whether the pair has a significant impact on the target is performed. Only those pairs which pass the test and have sufficiently high predictive strength are selected as possible two-way drivers.

Details

The restriction of selection of data fields and data rows for one-way drivers apply to the two-way drivers as well. This is expected as potential predictor fields for two-way drivers are selected from the top one-way drivers based on their respective predictive strength. However, the model significance of one-way driver and the minimum predictive strength is not required for their entry into a two-way model. A resulting two-way driver must have its predictive strength higher than 10% and provide more than 10% relative improvement over the predictive strength for each of the contained one-way drivers. Relative improvement is computed as the percentage of the difference between 100% and predictive strength of the nested one-way driver. Resulting two-way drivers that satisfy these criteria are ranked by their predictive strength and the top 20 are made available for display.

Decision tree

Decision trees are more complex models than the one-way and two-way drivers. They extend the sequence as the combination models. The main difference is that decision trees support discovery of interaction among multiple predictors and therefore deeper insights than the drivers.

Overview

Given the target field, the algorithm searches across all other data fields and adds them to the model to improve its strength in predicting the target values. The search across different predictors is iterative; after the search adds one predictor, the search continuous to add the next predictor that improves the model the most. The goal is to find the best set of predictors and an optimal way of combining them so that an optimal model is computed. The insights that are obtained from decision trees are presented in the form of decision rules where combination of predictors and corresponding values provide a single prediction for the target value. Decision rules are ranked by strength so that you can easily find the rules that are the most relevant and interesting. Decision rules that are generated by the decision tree are mutually exclusive. The decision rules also provide a complete rule set such that a corresponding rule exists for any combination of the predictor values in the data. Also, available is the overall decision tree predictive strength that provides relative improvement over the basic model. The results are available through three different visualizations: sunburst, tree, and decision rules. They each have certain advantages by displaying the decision tree structure and the corresponding decision rules content. Overall decision tree predictive strength is also available in the driver analysis visualization.

Algorithms

The decision tree model is computed after data preparation and building all the one-way drivers. The first tree predictor is selected as the top one-way driver. Categories of the predictor are merged when the adverse impact on the predictive strength is smaller than a certain threshold. The next step is to find the

best predictor to split each tree node that consists of the merged categories. The process is continued until a stopping rule applies to a tree node. Possible options for stopping are that all categories for every candidate predictor are merged into a single node or that the number of nodes exceeds maximum number of nodes. Categories with fewer than a minimum number of rows are always merged with another category. This means that none of the nodes in the tree can contain fewer than the minimum number of rows. The same procedure is used for continuous and categorical targets, only the impurity function is different.

Details

Impurity functions

Impurity function values are used as the main criterion for splitting and merging potential tree nodes. Impurity function total for continuous trees is the sum of squares per node, while Gini impurity measure is used for the categorical targets. Gini impurity total is computed as a sum of squares of count proportions across all target categories per node that is subtracted from one and the results that are multiplied by the number of rows. Improvement in impurity function value is information gain.

When splitting each node IBM Cognos Analytics looks for a predictor field with a largest information gain computed as total impurity across all potential children nodes subtracted from the parent node impurity. Before Cognos Analytics selects the predictor, Cognos Analytics attempts to merge some of the potential children nodes that initially correspond to each predictor category. Information loss is computed by subtracting impurity of non-merged nodes from the impurity of merged nodes. As long as information loss is smaller than a threshold, the nodes are merged. This process helps to create relatively small trees that are easy to visualize and comprehend while still preserving the overall strength of the tree.

Stopping rules

Candidate nodes are always merged if they are based on fewer than 25 rows. If all categories of a predictor are merged, it cannot be used for splitting a certain node. When none of the predictors can split the specific node, the process stops for the node. The overall process of generating the tree stops when none of the nodes can be split or when the number of generated nodes exceeds 36.

Variable importance

Variable importance corresponds to a relative tree error reduction when the corresponding predictor is included in the tree. It is computed by comparing the errors of an initial tree and a restricted tree that is generated by the rest of the predictors in the initial tree. The error of the initial tree is subtracted from the error of the restricted tree and the result is divided by the error of the restricted tree. Variables with zero or negative importance are removed from the tree. The tree error is computed as the sum of squares for continuous targets and as classification error for categorical targets.

Predictive Strength

Predictive strength for tree with continuous target is computed similarly to key drivers. The contents of leaf nodes are considered. Variance contribution for each leaf node is added and divided by the overall variance for the data. This is relative error for the tree. It is subtracted from one to obtain predictive strength that is compatible with the R-squared measure that is used for key drivers.

For categorical targets, Cognos Analytics computes classification accuracy based on the classification error that is added from all the leaf nodes. Relative classification accuracy improvement over the basic model, also known as adjusted count R-square, is reported as the tree predictive strength. It is computed by subtracting the tree error from the basic model error and dividing the result by the basic model error. For example, the classification accuracy of the model can be as high as 95%, but if the majority class appears for 90% of the rows in the data, then the predictive strength of the tree is reported as 50% only. This is parallel to the continuous target case where the basic model is represented by the overall mean value. Predictive strength that is measured by R-squared is based on the tree relative improvement in reducing the overall variance.

Cognos Analytics displays only the trees that have predictive strength larger than 10%. A tree for continuous target is displayed in a driver analysis or spiral visualization if its predictive strength is

show source <source-name>

Displays a list of relevant fields and details for the specified data source. The toolbar displays the <source-name> that is in context. By default, Cognos Analytics uses the active source in your dashboard or exploration panel. From the results, you can click field names to obtain more information for those particular fields. Clicking field names is equivalent to asking show column <column-name>.

When results exceed the number of displayed items, click **More** to view up to 100 data sources. Scroll to the bottom of the list and click **Less** to collapse the list.

show column <column-name>
For the specified column, information and related fields are displayed. Clicking the related fields is equivalent to asking show chart <column1> and <column2>. If the specified column is determined to have influencers, you can enter what influences <column-name> to see its list of influencers.

what influences column <column-name>
Displays a list of fields that influences the results in the specified column.

show chart <column1> and <column2>
Displays visualizations that show the relationship between <column1> and <column2>. Scroll through the visualizations by clicking the left and right arrows. Each visualization includes an information icon in the upper-right corner. Hover over the icon to see descriptions about the underlying data. You can optionally enter more columns, but excessive columns can result in less effective visualizations.

Clicking Show related visualizations returns visualizations based on influential and related fields.

Clicking Create dashboard from the charts creates a new dashboard based on the most recently generated charts. Typing Create related dashboard produces the same resulting dashboard. If the charts contain top or bottom aggregations, these modifiers are applied to the generated dashboard.

Applying aggregations and filters can help to add focus and create more compelling visualizations. Common aggregations include total, average, count, maximum/minimum, top/bottom, best/worst, and so on. Here are a few aggregation examples:

show top <num> <column1> by <column2>
Displays the top values from <column1> based on the context of <column2>. For example, show top 5 Sales by Region. If <num> is not specified, a default value of 10 is used.

<column1> is an aggregated or non-aggregated measure, while <column2> is a categorical column.

show average <column-name>
Displays the average for all values found in <column-name>.

how many <column-name>
If <column-name> is a category, the number of distinct items is returned. If <column-name> is a measure, the sum total is returned.

show maximum <column-name>
Displays the highest value found in <column-name>.

show minimum <column-name>
Displays the lowest value found in <column-name>.

show total <column-name>
Displays the sum total for all values found in <column-name>.

You can add filters for geographical strings (such as Country or State) or temporal strings (such as Month or Year). Filtering by date is not supported, at this time. Aggregations and filters can be combined to produce more granular results. Here are some examples, based on sample data:

- show Education by Income where Income is less than 1000
- show Education by Income where Income > 100K

Optionally use K (to denote thousands) or M (to denote millions).

- show Revenue in 2017 and 2018
- show Income by Month for New York City
- what are the top 5 States by average Inventory, excluding California

Filtered visualizations include a filter icon (), located in the upper-right corner of the chart. Hover over the icon to display the applied filtering.

create dashboard

Generates a new dashboard based on the currently selected data source. You can modify the visualizations, tabs, order, etc. and save your new dashboard. By default, the generated dashboard will include advanced analytics and predictive charts.

Automatically generating dashboards for larger data sources may result in performance issues. To circumvent this, you can enter `create simple dashboard` to generate a basic dashboard. Then you can modify the dashboard by replacing charts with more sophisticated visualizations, such as driver analysis or spiral charts.