Workshop Reference Guide

Version 4.1

September 15, 2017

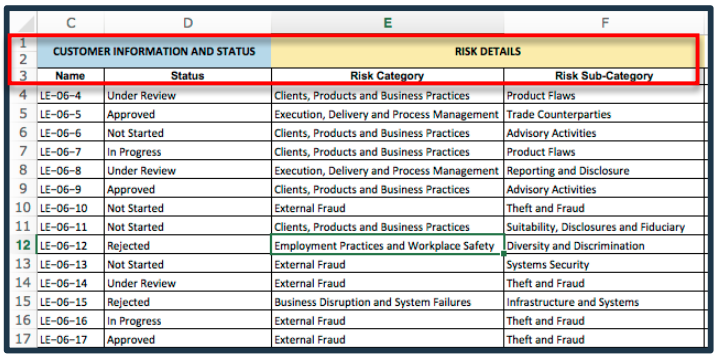
Table of Contents

Preparing Data

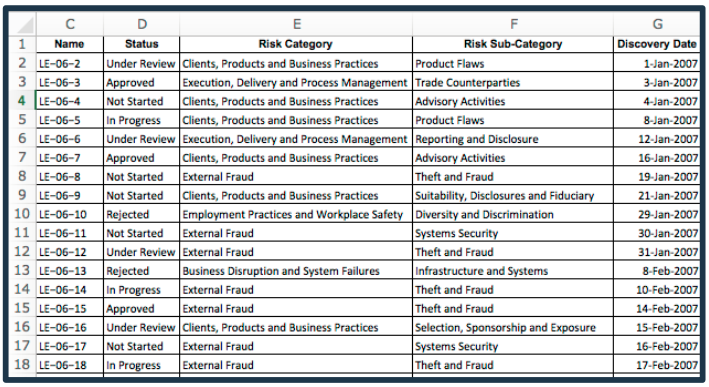
Remove formatting and aggregation on excel files

The ideal data set for Watson Analytics is rectangular piece of data - an excel sheet, a database table, etc. Database tables are by nature in a "raw" format, with consistent data types within columns and free of formatting. When using Excel files, you should format your sheet down to a raw form by:

- Removing header and footers. The first row of your excel sheet should be the row of column headers. There should not be anything below the last row of data
- Removing the row level subtotals, and aggregations, and calculations



	C	D	E	F
1	CUSTOMER INFORMATION AND STATUS		RISK DETAILS	
2				
3	Name	Status	Risk Category	Risk Sub-Category
4	LE-06-4	Under Review	Clients, Products and Business Practices	Product Flaws
5	LE-06-5	Approved	Execution, Delivery and Process Management	Trade Counterparties
6	LE-06-6	Not Started	Clients, Products and Business Practices	Advisory Activities
7	LE-06-7	In Progress	Clients, Products and Business Practices	Product Flaws
8	LE-06-8	Under Review	Execution, Delivery and Process Management	Reporting and Disclosure
9	LE-06-9	Approved	Clients, Products and Business Practices	Advisory Activities
10	LE-06-10	Not Started	External Fraud	Theft and Fraud
11	LE-06-11	Not Started	Clients, Products and Business Practices	Suitability, Disclosures and Fiduciary
12	LE-06-12	Rejected	Employment Practices and Workplace Safety	Diversity and Discrimination
13	LE-06-13	Not Started	External Fraud	Systems Security
14	LE-06-14	Under Review	External Fraud	Theft and Fraud
15	LE-06-15	Rejected	Business Disruption and System Failures	Infrastructure and Systems
16	LE-06-16	In Progress	External Fraud	Theft and Fraud
17	LE-06-17	Approved	External Fraud	Theft and Fraud



	C	D	E	F	G
1	Name	Status	Risk Category	Risk Sub-Category	Discovery Date
2	LE-06-2	Under Review	Clients, Products and Business Practices	Product Flaws	1-Jan-2007
3	LE-06-3	Approved	Execution, Delivery and Process Management	Trade Counterparties	3-Jan-2007
4	LE-06-4	Not Started	Clients, Products and Business Practices	Advisory Activities	4-Jan-2007
5	LE-06-5	In Progress	Clients, Products and Business Practices	Product Flaws	8-Jan-2007
6	LE-06-6	Under Review	Execution, Delivery and Process Management	Reporting and Disclosure	12-Jan-2007
7	LE-06-7	Approved	Clients, Products and Business Practices	Advisory Activities	16-Jan-2007
8	LE-06-8	Not Started	External Fraud	Theft and Fraud	19-Jan-2007
9	LE-06-9	Not Started	Clients, Products and Business Practices	Suitability, Disclosures and Fiduciary	21-Jan-2007
10	LE-06-10	Rejected	Employment Practices and Workplace Safety	Diversity and Discrimination	29-Jan-2007
11	LE-06-11	Not Started	External Fraud	Systems Security	30-Jan-2007
12	LE-06-12	Under Review	External Fraud	Theft and Fraud	31-Jan-2007
13	LE-06-13	Rejected	Business Disruption and System Failures	Infrastructure and Systems	8-Feb-2007
14	LE-06-14	In Progress	External Fraud	Theft and Fraud	10-Feb-2007
15	LE-06-15	Approved	External Fraud	Theft and Fraud	14-Feb-2007
16	LE-06-16	Under Review	Clients, Products and Business Practices	Selection, Sponsorship and Exposure	15-Feb-2007
17	LE-06-17	Not Started	External Fraud	Systems Security	16-Feb-2007
18	LE-06-18	In Progress	External Fraud	Theft and Fraud	17-Feb-2007

Other Excel File Specific Preparation

You can use Microsoft Excel 97 – 2003 workbook files (.xls) and Microsoft Excel 2007 and later workbook files (.xlsx).

.xlsx files that are saved in OpenOffice are currently not supported.

Password-protected Microsoft Excel files are not supported.

If more than one sheet you are prompted for which sheet to import

The .ods open source file format is not supported.

Files must be in list format - nested column or rows, i.e. crosstab, are not supported

You cannot have empty columns that are inserted before the data.

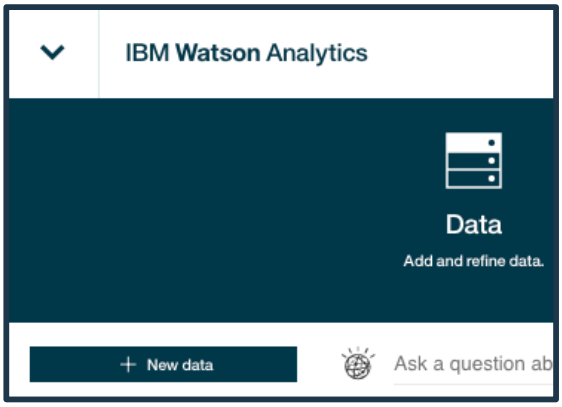
You must have a heading for every column. If a column doesn't have a heading, the column is ignored

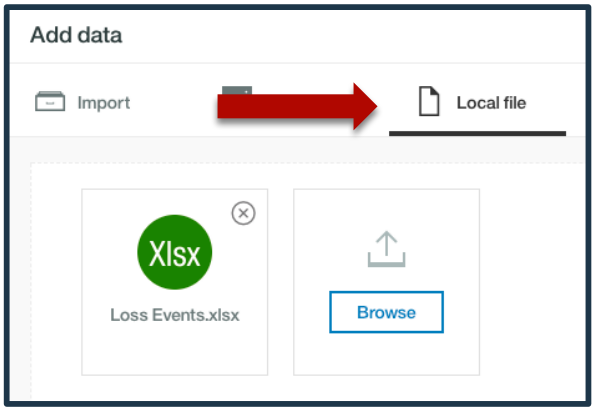
Uploading Data

Uploading an excel/CSV file

1. Navigate to the welcome page, using the page navigator

2. Click the "New Data'" button at the left of the screen, and choose "Local File"





3. Choose "Browse" and choose the "Loss Events" spreadsheet (provided by facilitator) from the resulting dialog box, and select "Import" from the bottom right.

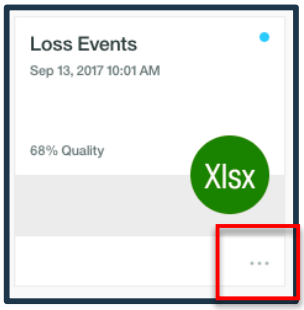
Refining Data

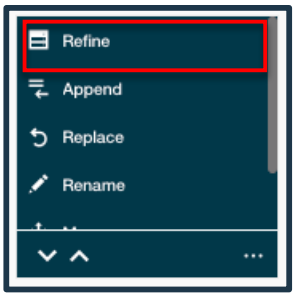
Refining Basics

The "Refine" capability allows you to modify and augment a loaded data asset. This could be as simple as filtering out unnecessary columns or rows, creating calculations, binning (data groups), or hierarchies for drill up/drill down. You can make changes to the original data asset, or save the changes to a new data asset.

1. Navigate back to the welcome page, using the page navigator

2. Click the ellipses on the data asset you want to refine, and choose the "Refine option.





Hiding/Filtering out unnecessary columns

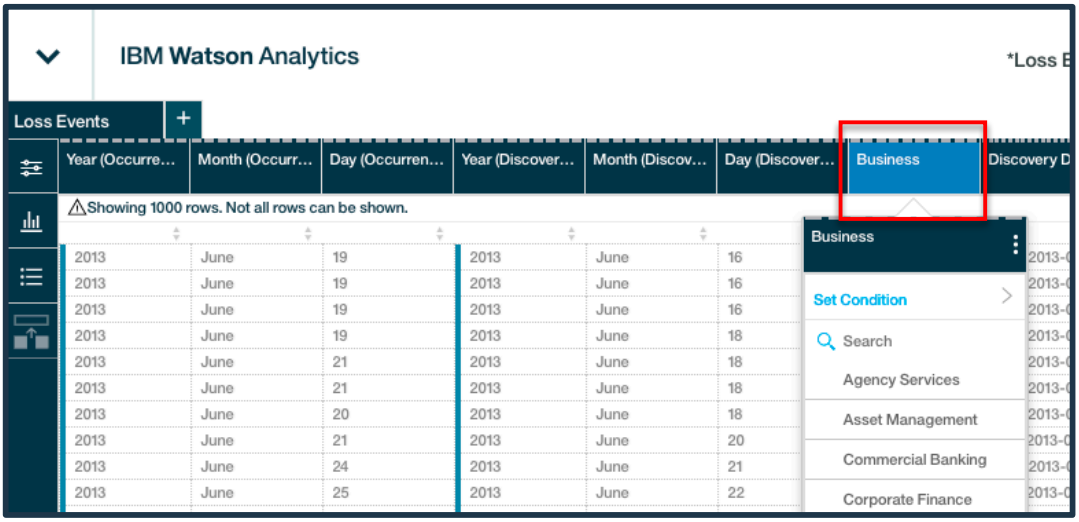
1. Click on the Actions button on the left-hand side. From here you have the option of showing/hiding columns of your choosing. Upon saving, this will remove the unchecked columns from the new refined data set. Here, leave all columns checked.

Note: there are certain columns that are hidden by default

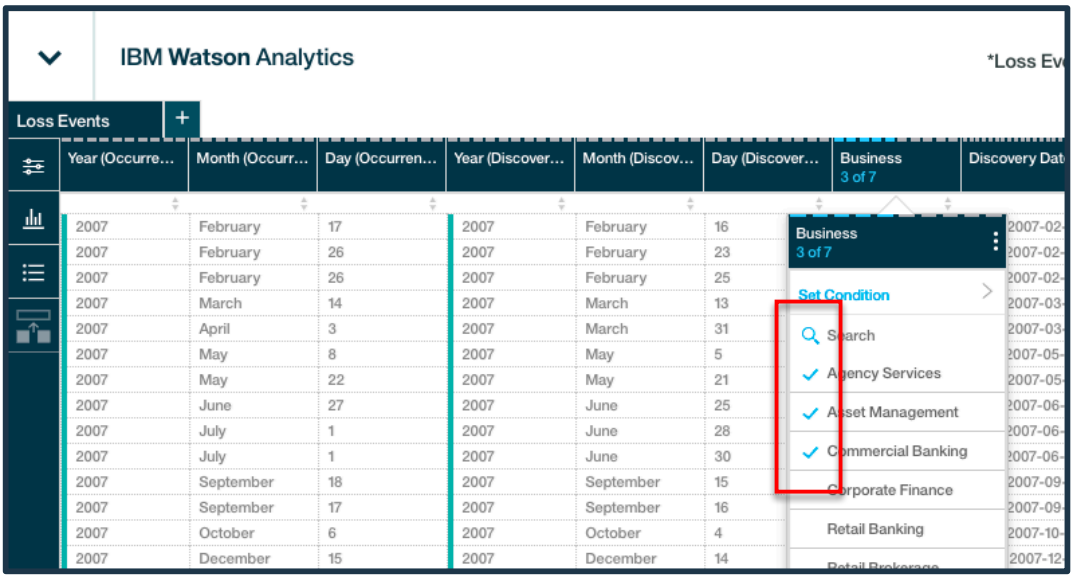
- "Rows" - This is a special column that contains a unique identifier for each row
- Any hierarchies that were automatically identified by Watson Analytics, with the exception of a time hierarchy, are hidden by default.

Hiding/Filtering out unnecessary rows

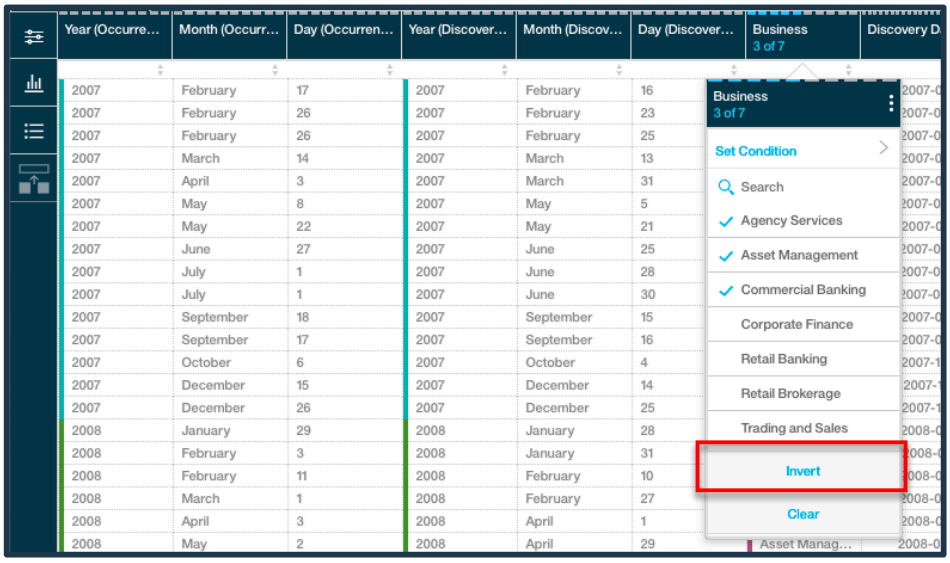
1. Click on a column header in the data refinement screen



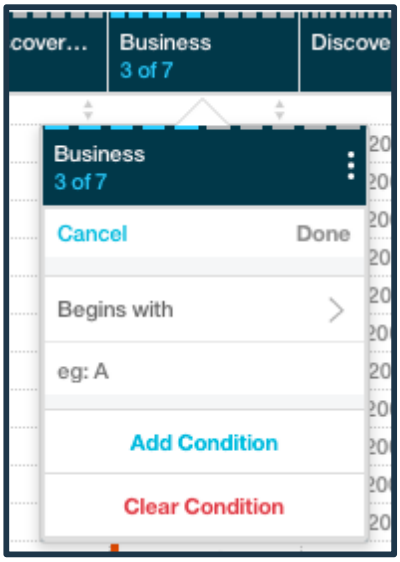
2. Here you have the option of filtering only a subset of rows of your choosing.



3. Alternatively, select the items you don't want to keep, and select Invert at the bottom. (Useful when you want to keep many items, with the exception of a few)



4. To create a more advanced filter, use the Set Condition option to search for strings (equals, contains, begins with, etc)

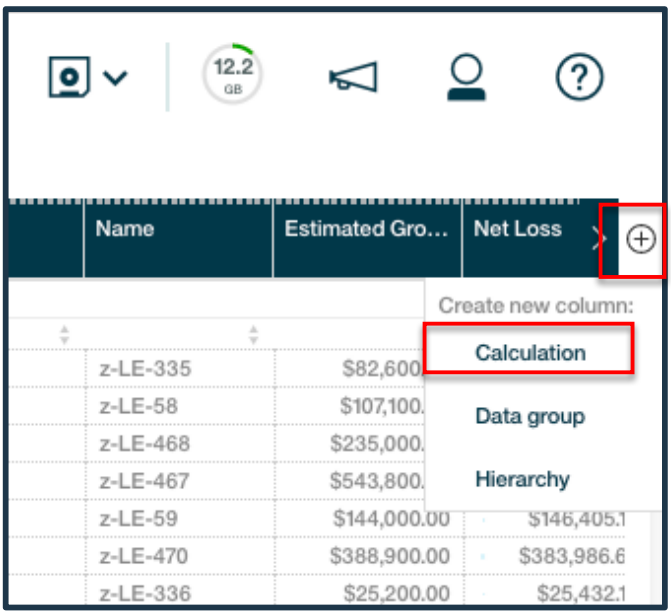


In our case, leave all rows **selected** as per default

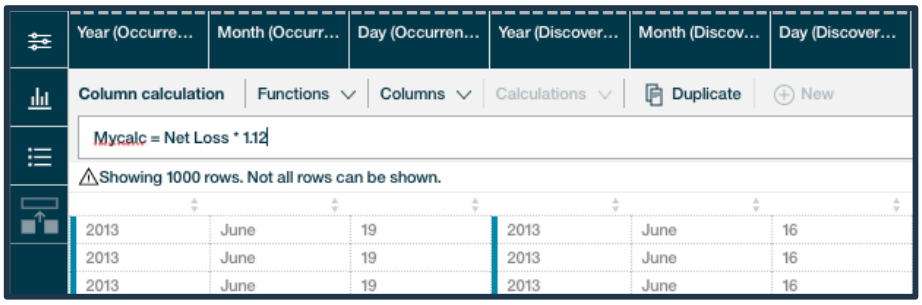
Creating Calculations

Oftentimes, you may have a need to create calculated items. For example, you have the fields Quantity sold and Price, and you wish to determine Revenue.

1. Click on the **plus** sign in the data strip and select **Calculation**



2. Enter the name of your new calculation on the left, then begin to build your calculation by starting with either a **Column**, a **Function** or **Number**

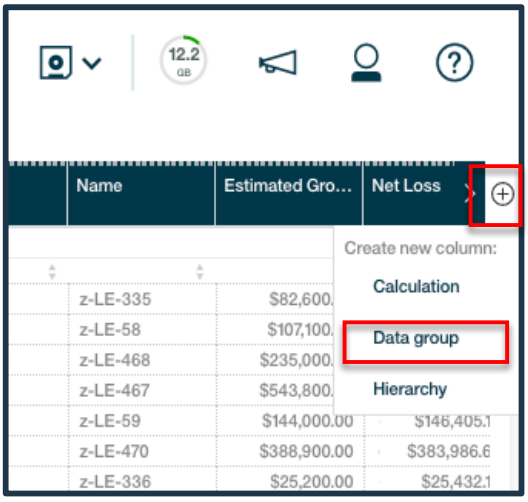


3. Click Done to complete

Creating Groupings with Categorical Data or Numerical Data

The purpose for grouping arises when you have a field with many distinct values, and you'd like group them for analysis. For example, you have a variety of ages, and you'd like to group them into Young, Middle Aged, and Seniors. Or perhaps you have a range of Amount Spent and you'd like to group them into groups of spenders.

1. Click on the **plus** sign in the data strip and select **Data group**

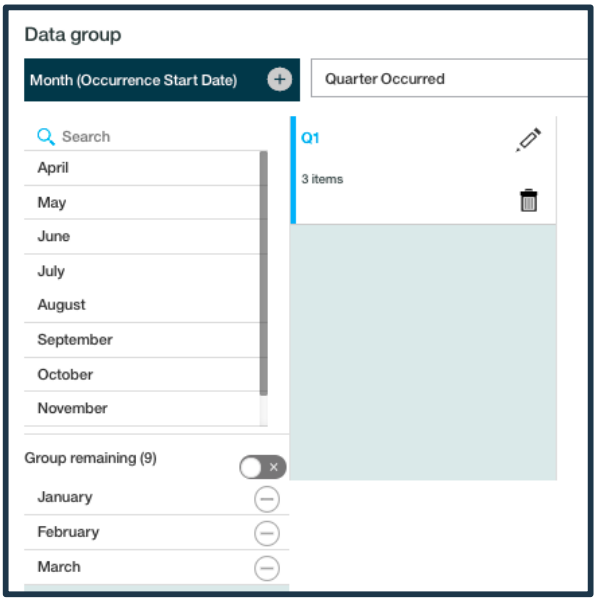


2. This will create a new column in your dataset, so enter "**Quarter Occurred"** in the text box

3. Select the "Month (Occurrence Start Date)" by clicking the plus button.

4. For **Categorical** Data

    a. Select the items on the left that you want to combine as part of a grouping, then select **New Group** on the right and call it "Q1" then click OK.



    b. Repeat step 4 three more times to create Q2, Q3 and Q4, keeping in mind that when you have only group left to create you can choose **Group Remaining** at the bottom left.

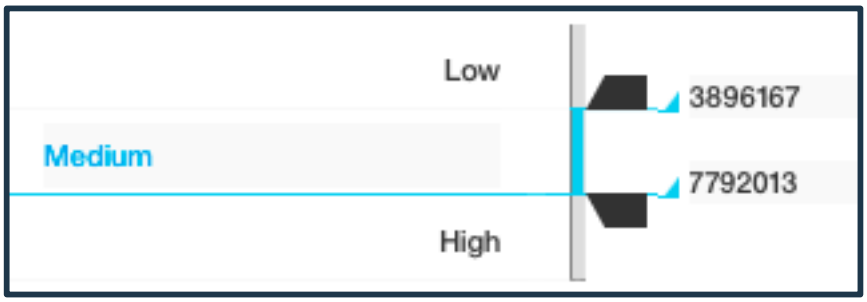
    c. Click the Done Button

5. For **Numerical** Data

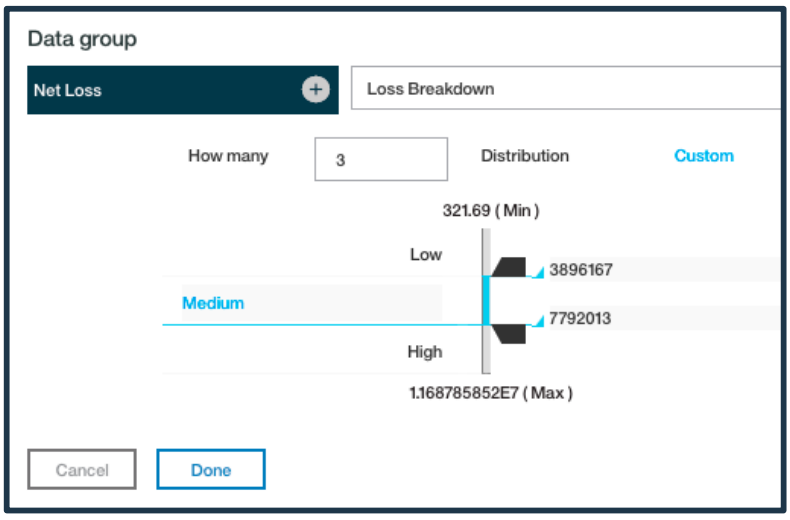
   a. Repeat step 1 to create a new data group, and this time select "Net Loss" as the field to group. Set the number of groups you wish to separate your data into by specifying "how many" to **3.**

   b. By default, your groupings will be distributed at an equal rate. However, you can choose to have your groupings distributed non-equally or in a custom range by sliding the scale to fit your needs.

   c. You have the option to change the names of the grouping by selecting the default name and creating your own custom title



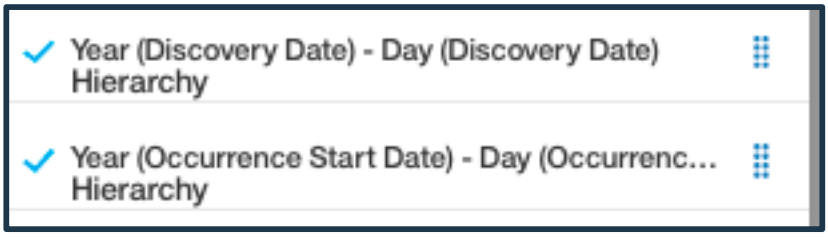
   d. Call your group "**Loss Breakdown**" and click Done.



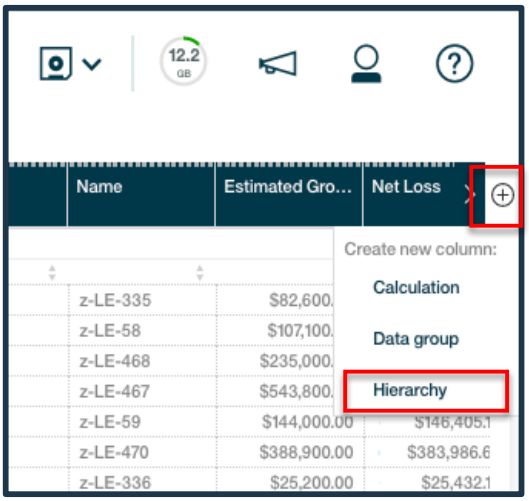
Creating Hierarchies

The creation of hierarchies will enable the drill up and drill down functionality on a data item. For example, if you have Product Type, you may wish to drill down to the Products that make up this product type.

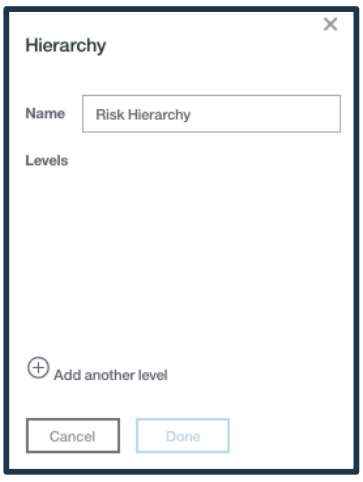
1. First check to see if the hierarchy you wish to create was already created automatically by Watson Analytics. In the Actions menu, scroll to the bottom of the field list to see the automatically detected hierarchies - they will be in the form: "Data item 1 - Data item 2". If so, you may just need to unhide the column.



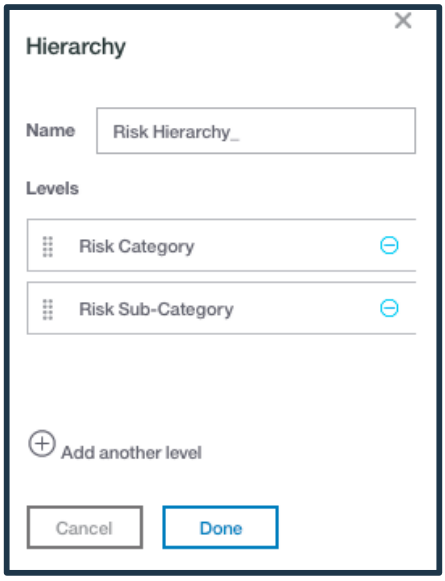
2. To create a new hierarchy, click on the **plus** sign in the data strip and select **Hierarchy**



3. As this will create a new data item, call your hierarchy **Risk Hierarchy**



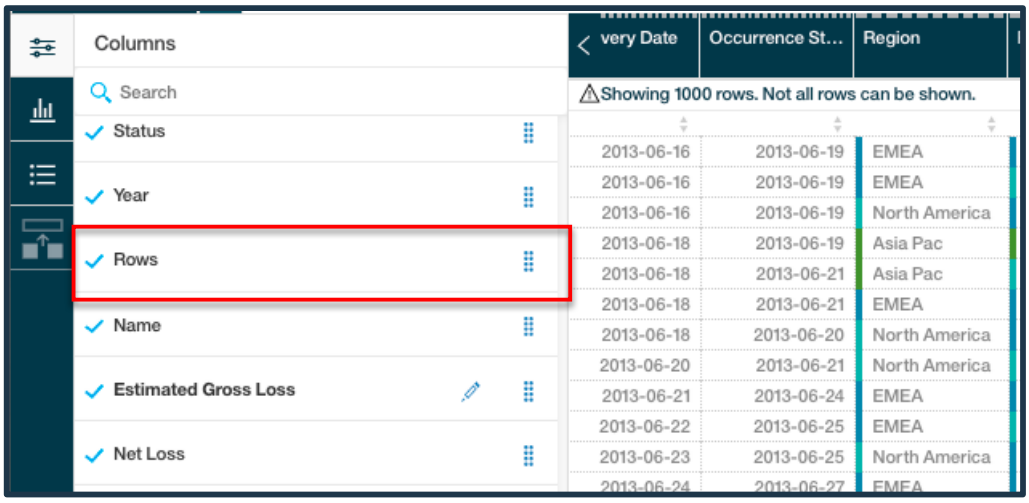
4. Then click Add Another Level to put the data items that make up this hierarchy, starting with the parent which in our case is the **Risk Category** field, and then the first child which in our case is the **Risk Sub-Category** field.



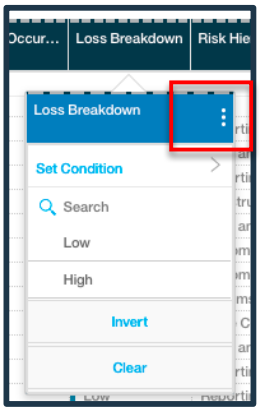
5. Click Done to complete

Adding the Rows column

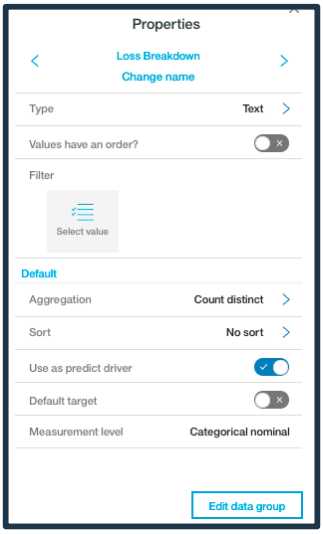
To simplify the creation of visualizations that require the count of the number of rows in your data asset, add the Rows column. Click on the Action button then ensure there's a checkmark beside the **Rows** field.



Changing the Name, Type, or default Aggregation for a field



1. Click on the column header of the **Loss Breakdown**, and open the menu for that field by clicking on the 3 ellipses icon



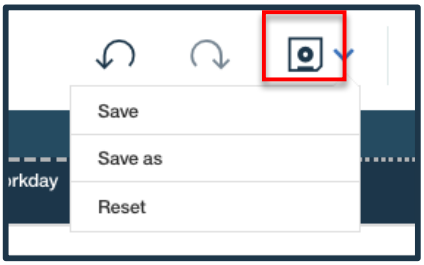
2. Choose Properties

   - Click on Change Name to rename the field
   - Click on the existing Type to change the data type of the field
   - Click on the existing Aggregation to change how this field is aggregated by default
   - You can specify what columns you want to use when you create a prediction
   - For each column that you choose, you can specify whether it is a prediction target.
   - You can also see what its measurement level is.

Note: Prediction targets are fields that are influenced by the other fields in the data set.

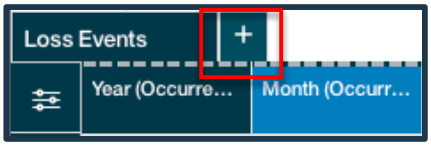
3. Toggle the "use as a predict driver" box to unchecked for **Loss Breakdown**, then repeat steps 1-3 for the **Recovery Amount (percent)** field

The measurement level of a column classifies the information in a field. Having the correct measurement type ensures that you get valid insights from a prediction. This value is automatically derived from the settings of the Type and Aggregation properties of a column.



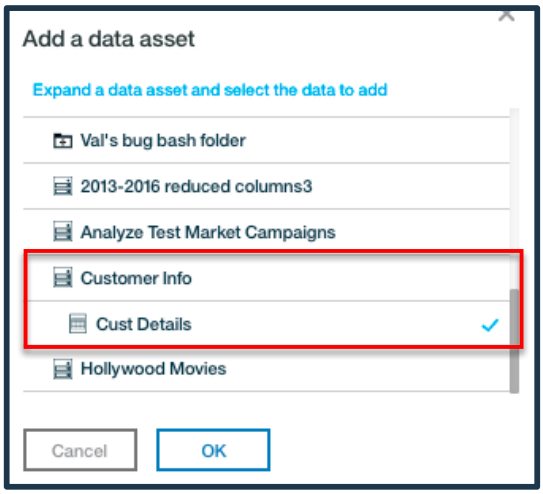
4. Save your dataset by clicking on the save button in the toolbar and select **Save As** and name your dataset "Loss Events Refined"

Optional: Joining multiple data sources



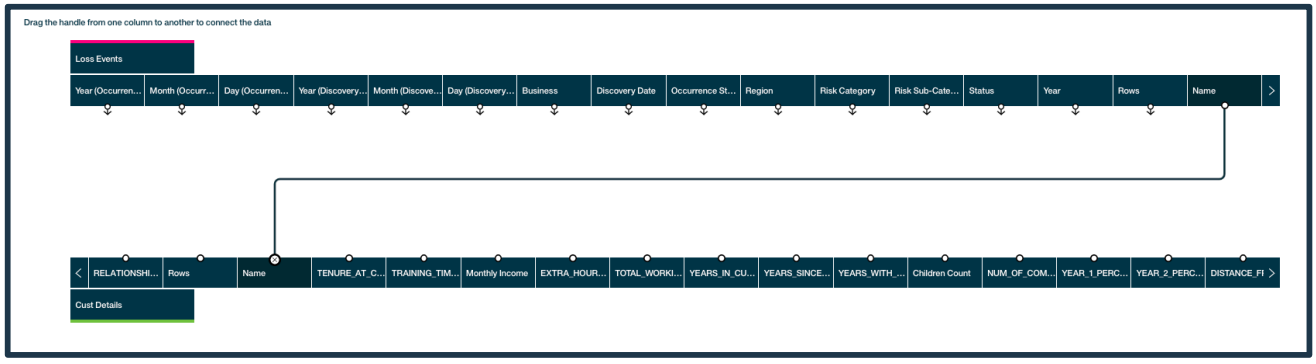
1. Click on the + sign to the right of your open worksheet

2. Navigate to the Shared folder, then locate the file called **Customer Info**. Select "Cust Details" then click Ok.

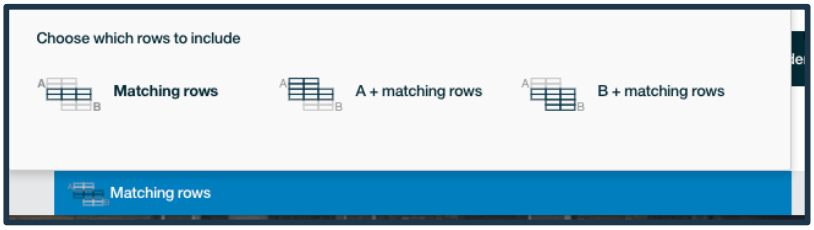


3. Click on the **join** button along the left side panel 

4. Now you can link your two datasets using common fields. From the Loss Events datast, select the "Name" field, then from the Cust Details dataset also locate the Name field (you may need to scroll to the right).



Note: You can also choose the type of join which is most appropriate:



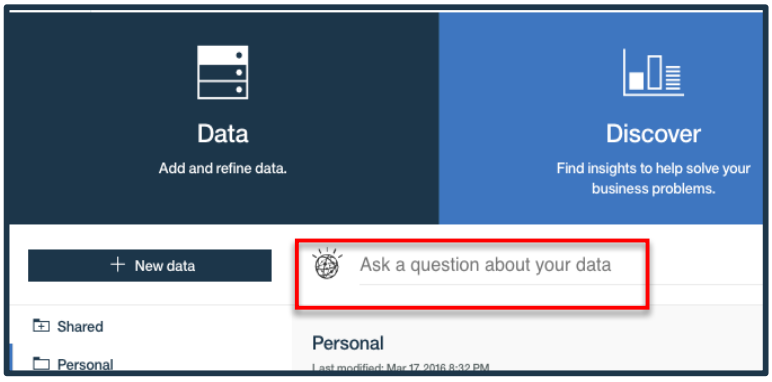
Discover

Discover Basics

Use the Discover capability in Watson Analytics against a data asset to explore and visualize data, identify relationships, and explain outcomes - all towards the goal of answering business questions.

1. Navigate back to the welcome page, using the page navigator

2. Interacting with a data asset is done through a Discovery Set, which is a workbook that will store all of your visualizations (ie, discoveries). If you don't know which data asset to use, complete the following actions:



   - Enter a question on the Data landing page. IBM® Watson Analytics™ analyzes all of the data assets you have access to and creates a list of starting points that might answer your question.
   - Select one of the starting points - this will create a Discovery Set on a specific data asset.

3. If you know which data asset to use, click on the data asset that you want to use - this will create a Discovery Set. Then do one of the following:

    - Ask a question and then select a suggested starting point
    - Select one of the suggested starting points
    - Choose the specific visualization you want to start with from the bottom. Note: If you choose a specific visualization, Watson Analytics pre-populates it with the first few columns of data in the data asset.

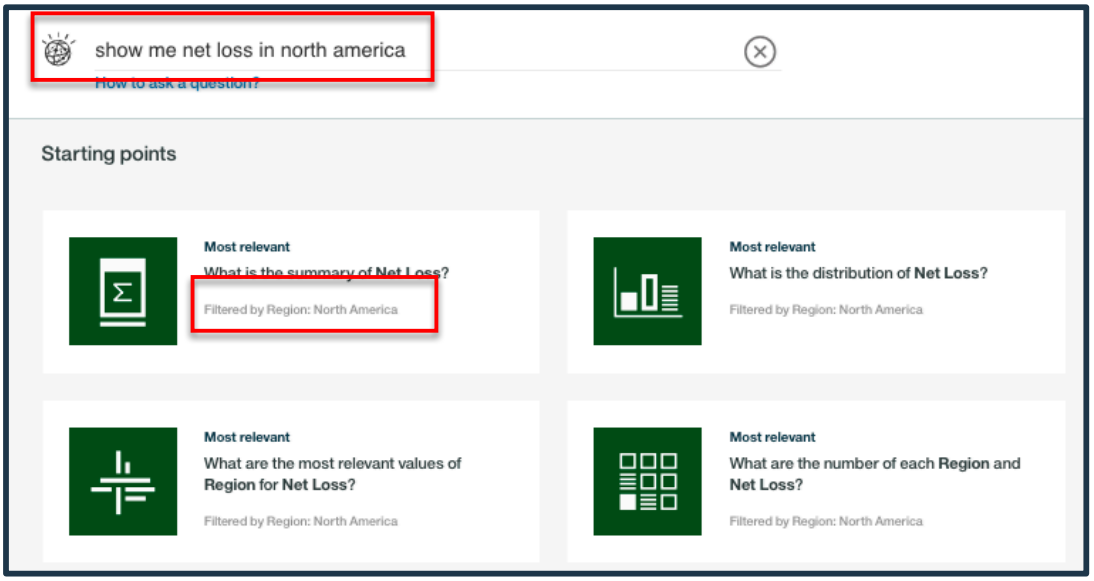
Asking a Question with the Natural Language Dialogue box

When you first choose a Data Asset to Discover, you're presented with suggestions (starting points) that were based on the analysis of your data during upload. You can click on any of those suggestions, and it will display the visualization.

There is also a Natural Language Dialogue box to create your own visualizations, which makes use of column headers (field names) and or specific data values, and keywords.
The simplest way to get started with the Natural Language Dialogue box is to use column headers.

1. Enter the name of a column header(s), and Watson Analytics will update the suggestions to show you visualizations based on the column header(s) you entered
- Choosing one of these suggestions will display the visualization

2. Using specific data values will act as a filter. For example, if you have a column called State, which contains various state names, using "New York" in the Natural Language Dialogue box will filter all the suggestions related to New York:



TIP: Use the “How to Ask a question” link to access a question wizard.
The use of keywords will provide more specific results

Compare
How do the values of <Column heading> compare by <Column heading>?
How do the values of <Column heading> compare by <Column heading> and <Column heading>?

Trend
What is the trend of <Column heading> over <Column heading>?
What is the trend of <Column heading> over <Column heading> by <Column heading>?

Contribution
What is the contribution of <Column heading> and <Column heading>?

Correlation, Relationship
What is the relationship between <Column heading> and <Column heading>?

Breakdown
What is the breakdown of <Column heading> by <Column heading>?
What is the breakdown of <Column heading> by <Column heading> and <Column heading>?

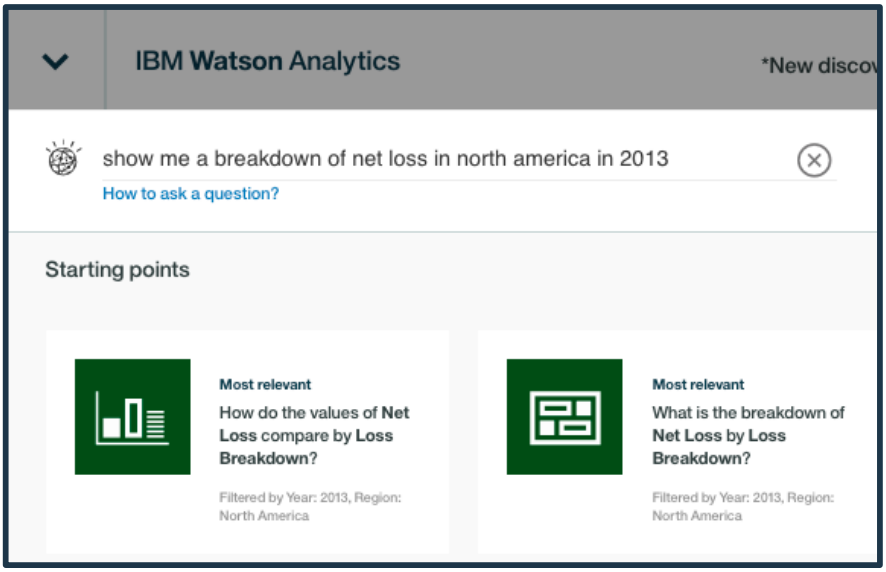
Grouping
What is the grouping of <Column heading> and <Column heading>?

Where
Use for map-based visualizations or when you include a location-based column in your question.

When
Use with columns that include times or dates.

How long
Use with columns that include time durations.

Example:



3. Keywords to sum, sort, or count values - use the following keywords in a question to sum, sort, or count values in the resulting visualization.

Tip: Place this type of keyword in front of the column heading that you want it to act upon.

Average
Average sales per region

Total
Total rainfall per month

Maximum
Maximum balance by month

Minimum
Minimum golf score by game

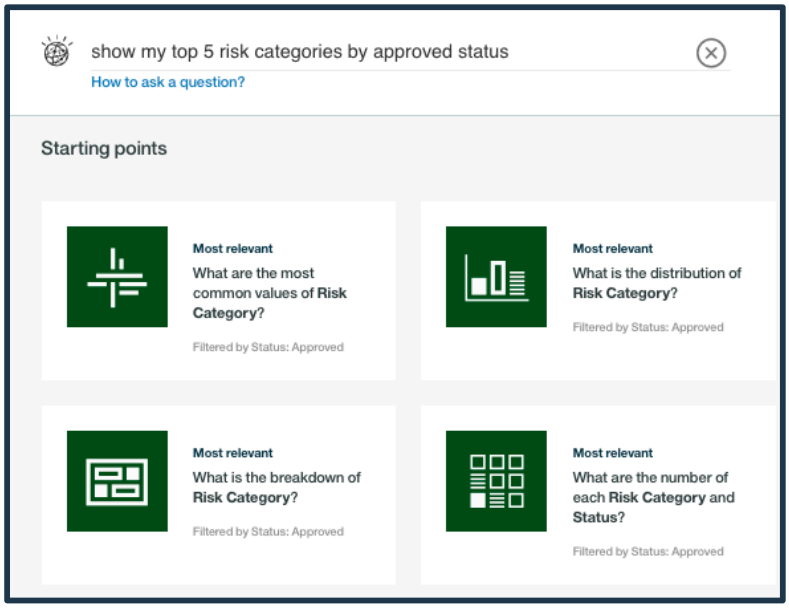
Top, bottom, best, worst, highest, lowest, most, least
Sorts the results.

Top products by sales
Include a specific value to filter a number of records in the sort.

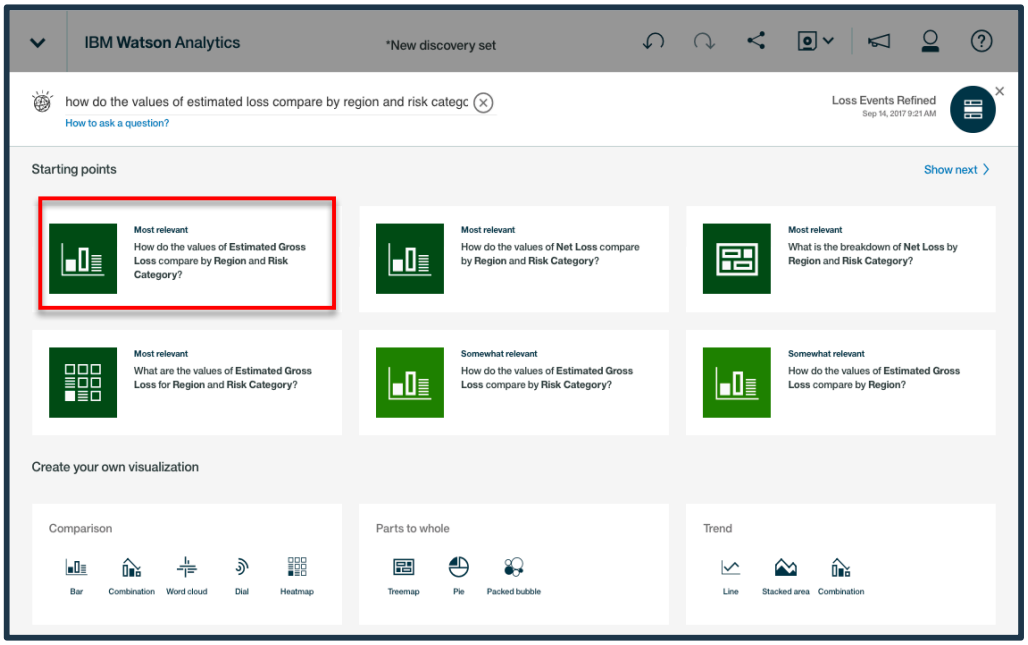
Rows
A count of sales by sales person.

How many, Count
Includes a count in the visualization.

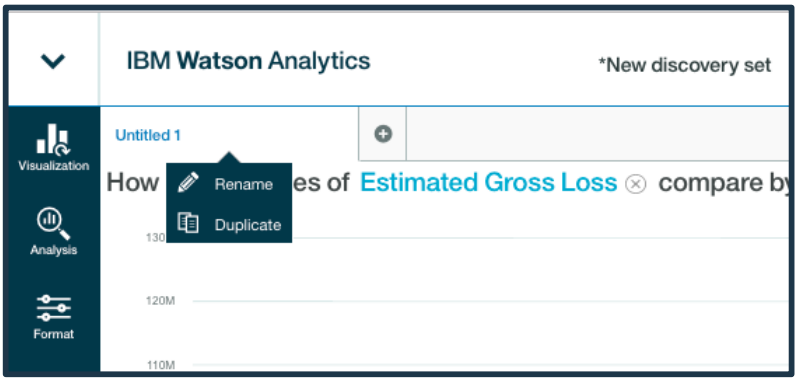
Example:



In our case, ask the question ***How do the values of estimated loss compare by region and risk category?*** then select the most relevant starting point.

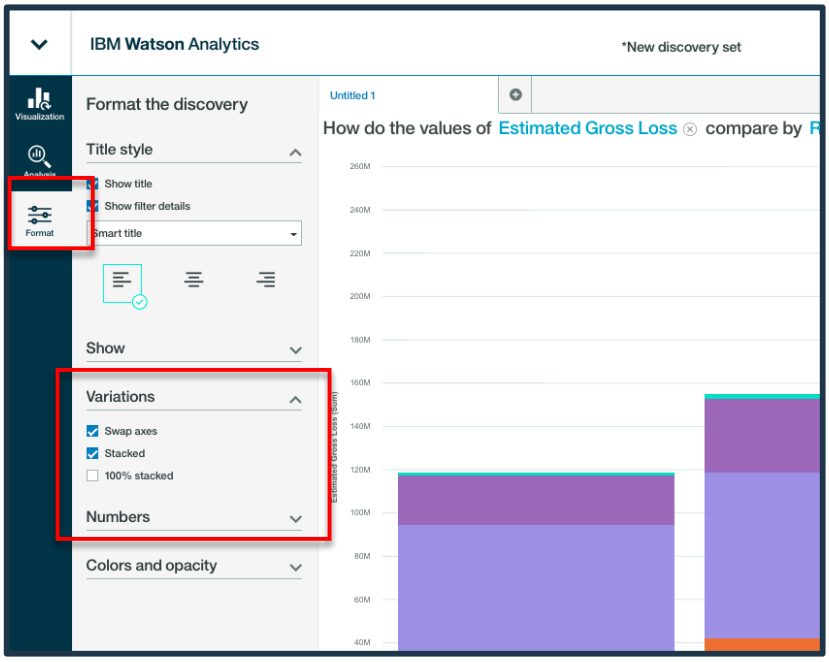


4. A Discovery Set is composed of tabs. Every time you generate a visualization, it will create a new tab. **NOTE**: Click on the tab to rename it - this is recommended to allow for easier dashboard assembly in the Display capability.



Formatting

1. Click on the Format button on the left toolbar to expand the formatting pane, then expand the **Variations** drop-down and check the “stacked” checkbox



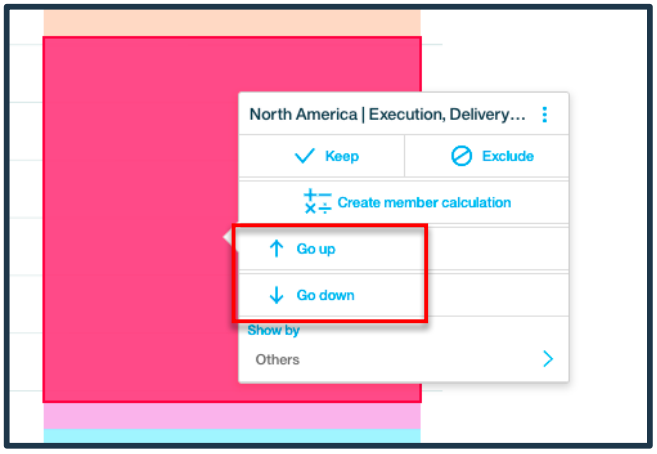
2. Expand the **Colors and opacity** drop-down and select a palette of your choosing.
3. Click the Format button again to close the pane.

Drill Up/Down

1. Right click on a data element inside a visualization, and you're be presented with an option to Go Up or Go Down, representing drill up and drill down, respectively.

Because we created a hierarchy in a previous step we can drill-down on Risk Category. Here we can see how Reporting and Disclosure is a risk in North America, and how that differs from APAC with a Vendor and Supplier risk.

Note: If the data element you click on is not part of a hierarchy, your ability to drill up and down will be limited.



Manipulating Data in a Visualization

For each visualization, the data elements that drive the visualization can be updated in a number of ways.

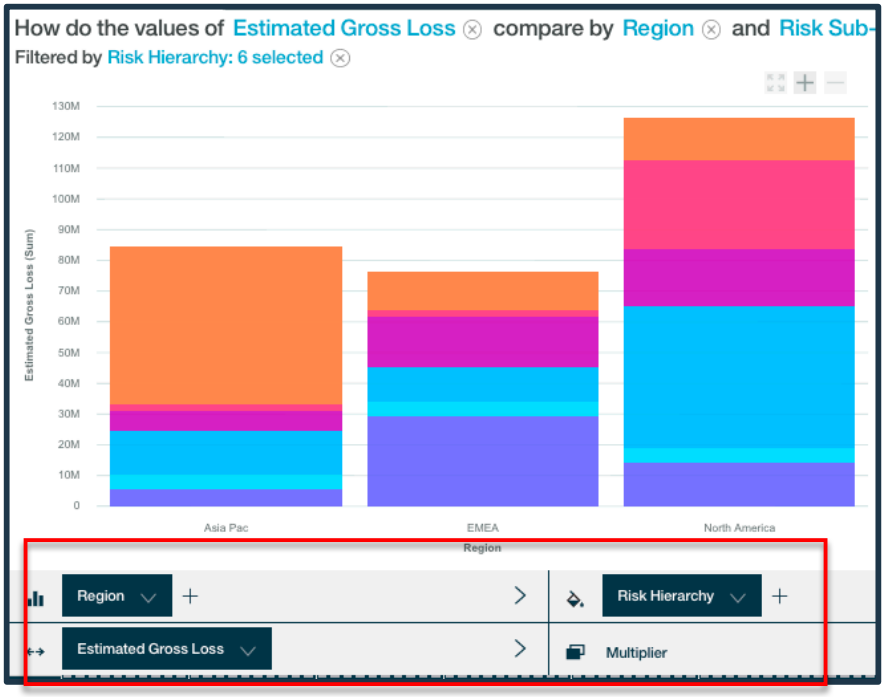
The Title Bar

The blue highlighted items in the title are the data elements (columns) that are present in the visualization. If you click on any of those items, you are presented with a list of other suggested data elements that you could use instead.

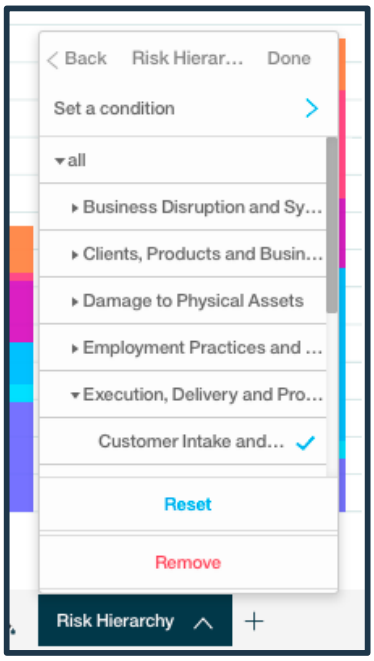
Manually changing the elements of the visualization

The aspects of the current visualization are visible at the bottom of the screen.

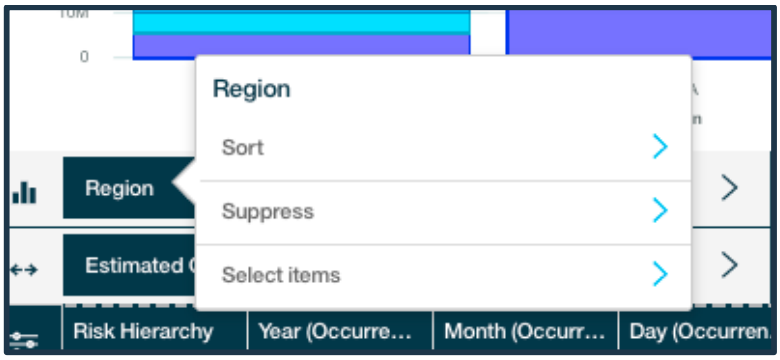
- Drag and drop items from the data tray into drop zones, to replace the existing data item or to add a new aspect to the visualization



- Left clicking on a data element will provide a menu to filter the elements for this visualization (Local filter)



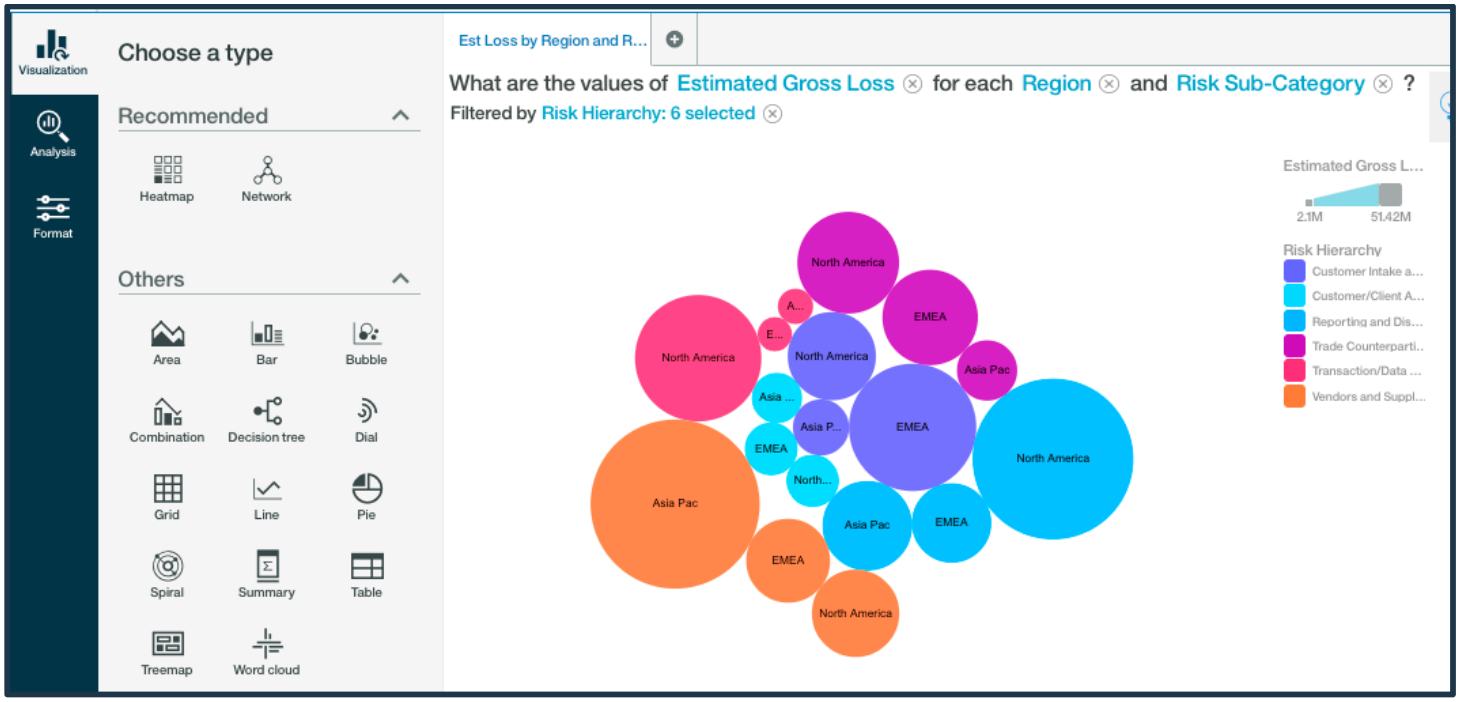
- Right clicking on a data element will provide sorting and suppression options, or summarization options if you click on a measure



Changing Visualizations

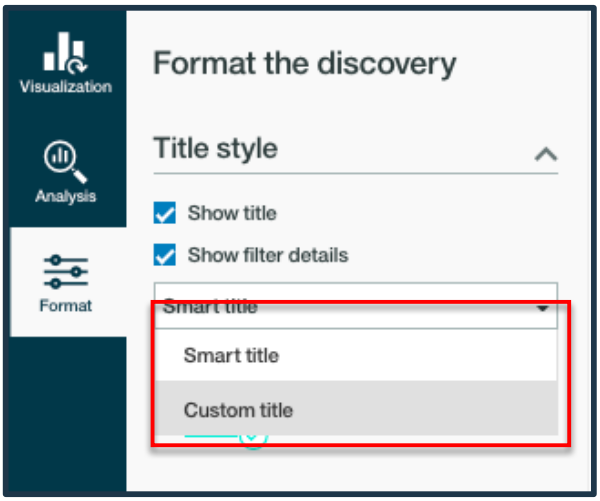
To change the visualization, select the visualization icon at the top left. You will be presented with recommended visualizations, which are based on the data elements currently being used.

1. In our case, let's change our stacked column chart to a packed bubble chart



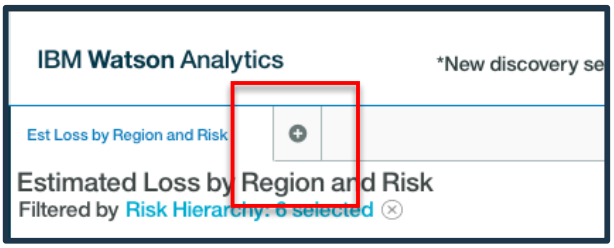
2. Click on the Format button, then expand the **Title Style** and change the "smart title" drop-down to "custom title". Now give your discovery a name such as "Estimated Loss by Region and Risk"

3. Click format button to close the pane.



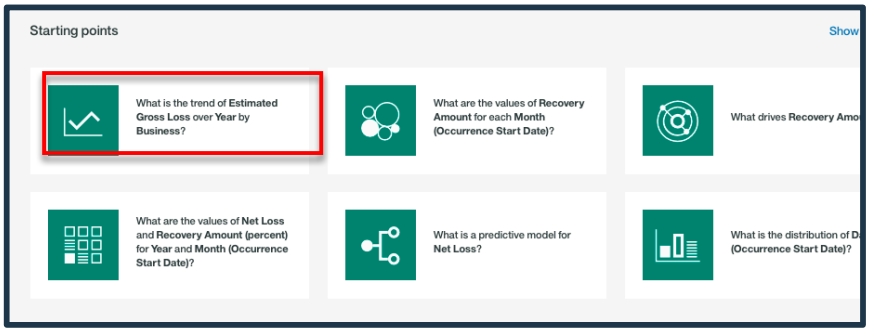
Starting Points

1. Click on the New Tab “plus” button to create a new discovery.

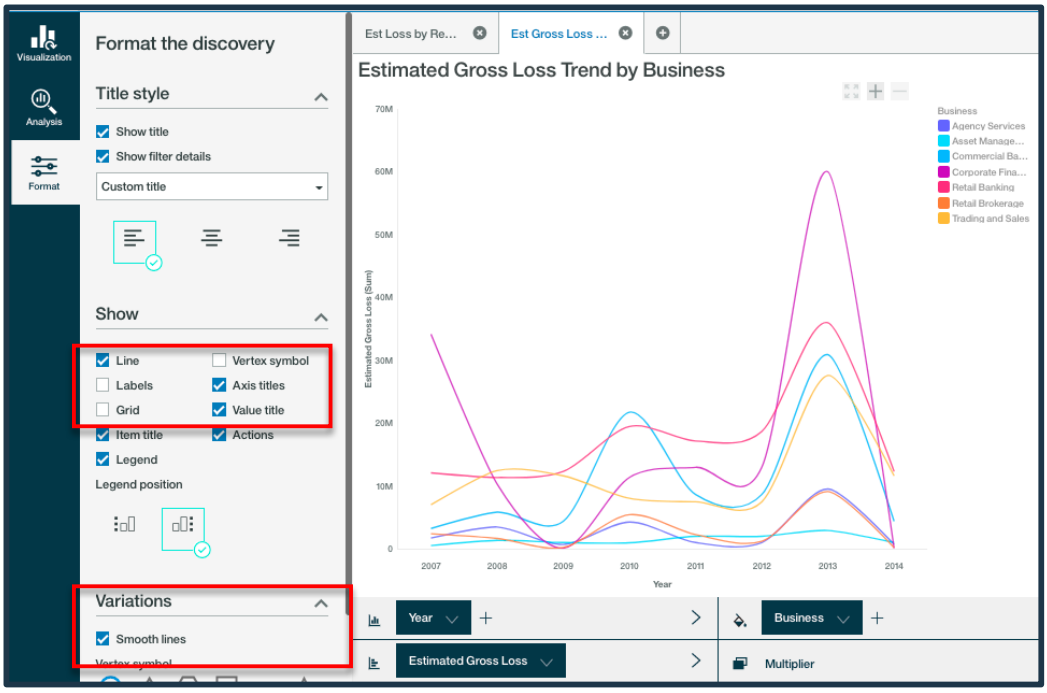


When a dataset is uploaded to Watson Analytics, the data and metadata is analyzed to determine underlying patterns and relationships. When you click on a dataset, those resulting analyses and their resulting visualizations are presented as smart starting points based on relevance.

2. Here, the suggested visualization with the highest relevancy is the trend of estimated gross loss over year by business, so click on that



3. I can see that there was a large spike in estimated gross loss for corporate banking. Change the color palette to match the colors chosen in a previous step, rename the tab in the discovery set and change the title to a custom title.
4. Expand the **Show** drop-down and uncheck Grid and Vertex Symbol. Expand **Variations** and check “smooth lines”.



Predictive Visualizations

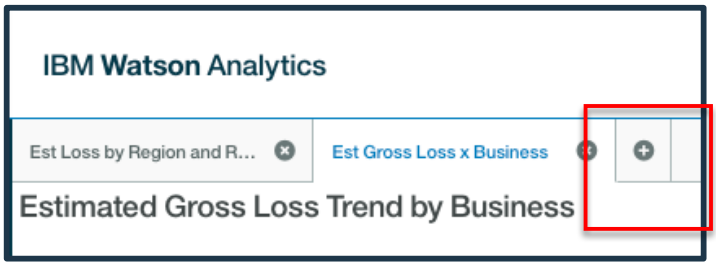
With the Discover section of Watson Analytics, there are some advanced, or predictive, capabilities. The main goal of the predictive capabilities in Watson Analytics provides an explanation of an outcome by identifying the patterns that exist in a historical data set. For example, say you have a historical data set where each row represents the characteristics of a single sales opportunity, and for each opportunity, the outcome was a win or a loss. Using the predict capability would be able to tell you which combination of characteristics were more important, and develop the profiles of opportunities you are likely to win, and likely to lose.

Keep in mind that not every dataset will be ideal for this capability. If your data does not have an outcome to predict, or if it does not have enough columns or the relevant columns, or the data is of poor quality, then your results may vary.

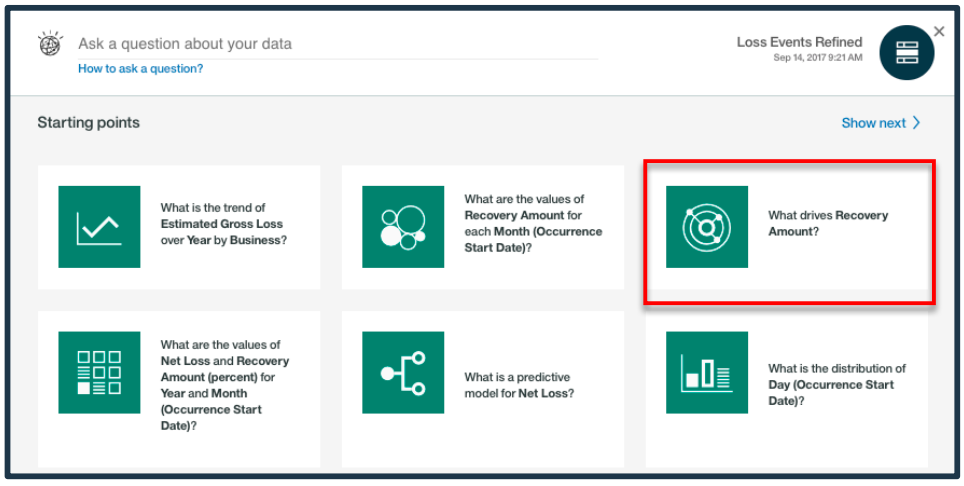
Drivers (Spiral Chart)

This visualization shows the single-factor or two-factor drivers (influencers) of a given target.

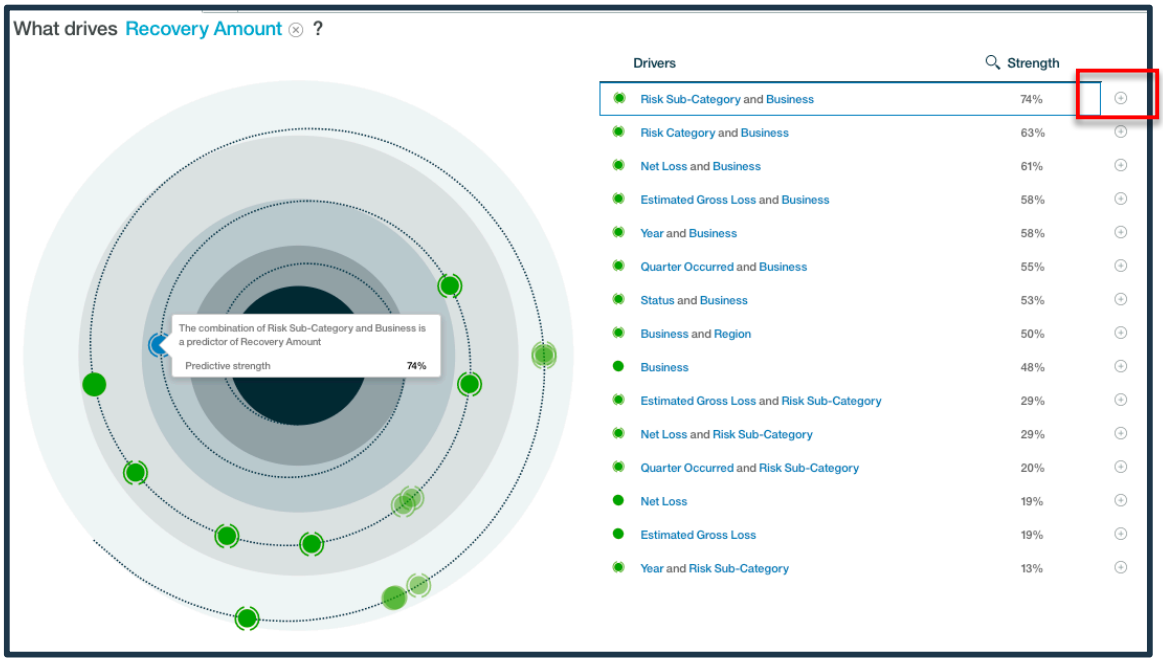
1. Click on the New Tab "plus" button to create a new discovery



2. Here Watson Analytics has already determined that there's a relevant statistical relationship that we should review. I could also use the natural language query to say What Drives <field 1> or show me a predictive model for <field 1>. In our case, let's examine the recommended visualization.



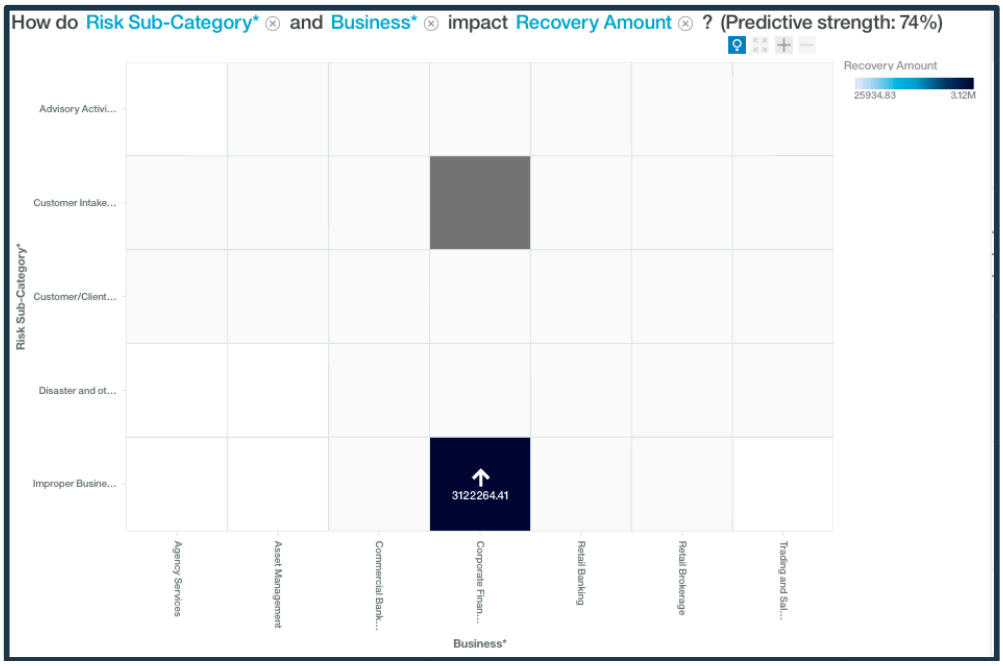
3. The closer a pin is to the center of the spiral chart, the more influence it has (ie, the better it can explain the outcome). This is also shown on the right hand side with a strength indicator.



Note: Here **strength** is a patented combination of factors which is analogous to R^2

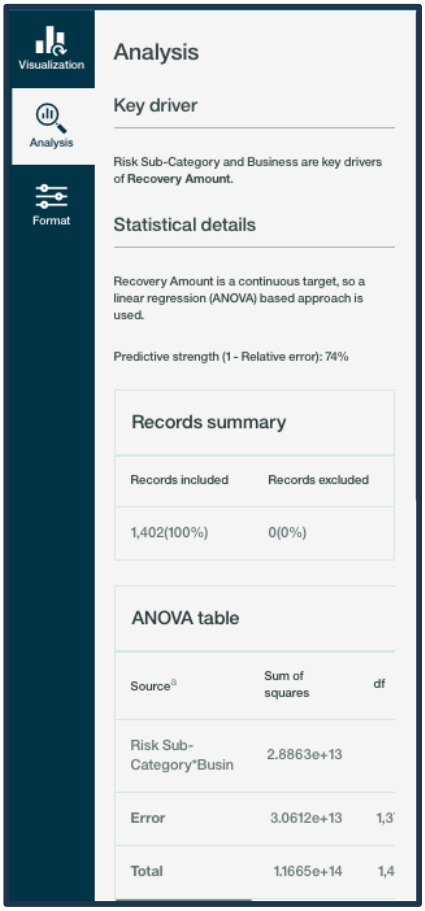
In our case the combination of **Risk Sub-Category** along with **Business** unit is the leading driver of understanding recovery amount.

4. To view more details about any of the drivers, click on the + icon and it will show you additional details on a new tab.



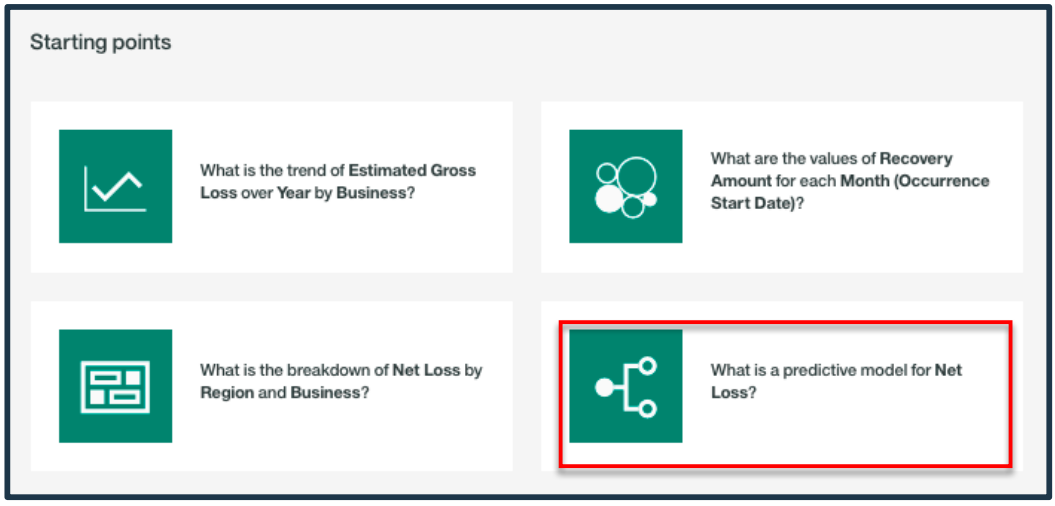
Interpretation: Recovery amount is unusually high for corporate business units with a risk sub-category of improper business or market practices.

5. Clicking on the Analysis button on the left toolbar will provide the statistical details surrounding how the model was built



Predictive Model (Decision Tree)

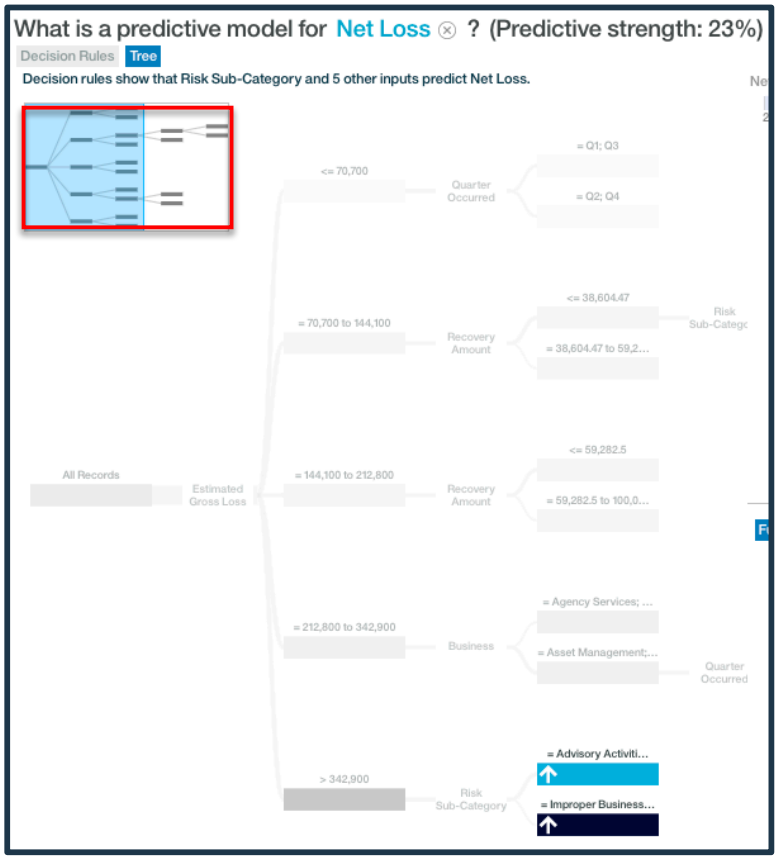
1. If you want to see a multi-factor explanation of an outcome, add a new tab to your discovery set and select the recommendation for a predictive model, but this time Watson Analytics is suggesting using **Net Loss** as the target.



The Decision Tree

The decision tree shows you patterns of characteristics that lead to a certain outcome, which you could also think of as profiles. Reading from left to right, each branch in the tree is a unique pattern, that leads to the likelihood of an outcome occurring in the past. In the screenshot below, the lighter boxes represent lower satisfaction, and the darker boxes represent higher satisfaction.

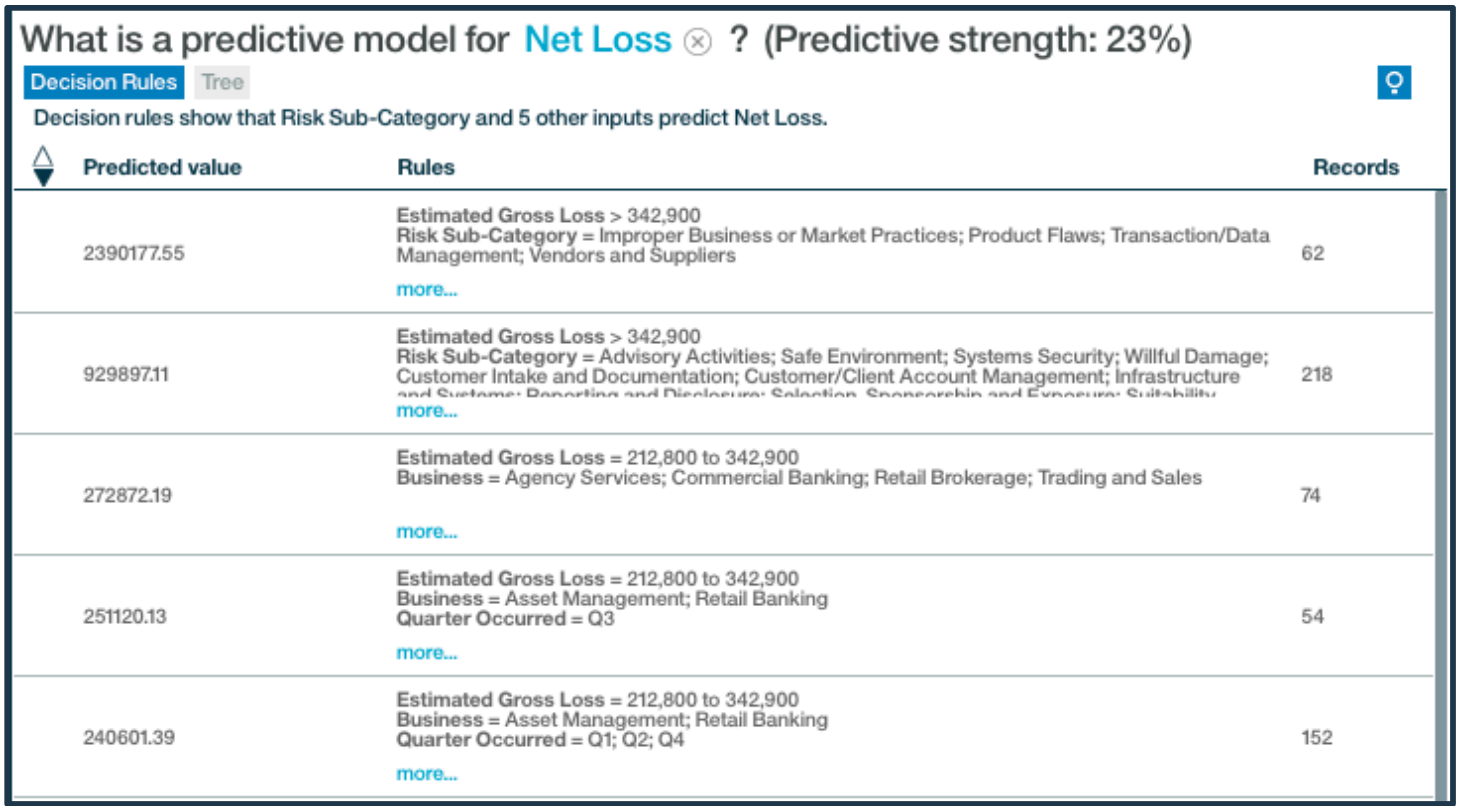
You can focus on specific paths by collapsing/expanding branches of this tree by clicking on individual nodes.



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.

In our example, the leading predictor of **Net Loss** is a combination of a number of factors: When Estimated Gross Loss is greater than 342K, when the Risk Sub-Category contains a specific set of categories, etc.



What is a predictive model for Net Loss ? (Predictive strength: 23%)

Decision Rules | Tree

Decision rules show that Risk Sub-Category and 5 other inputs predict Net Loss.

Predicted value	Rules	Records
2390177.55	Estimated Gross Loss > 342,900 Risk Sub-Category = Improper Business or Market Practices; Product Flaws; Transaction/Data Management; Vendors and Suppliers more...	62
929897.11	Estimated Gross Loss > 342,900 Risk Sub-Category = Advisory Activities; Safe Environment; Systems Security; Willful Damage; Customer Intake and Documentation; Customer/Client Account Management; Infrastructure and Systems; Reporting and Disclosure; Selection, Sponsorship and Exposure; Suitability more...	218
272872.19	Estimated Gross Loss = 212,800 to 342,900 Business = Agency Services; Commercial Banking; Retail Brokerage; Trading and Sales more...	74
251120.13	Estimated Gross Loss = 212,800 to 342,900 Business = Asset Management; Retail Banking Quarter Occurred = Q3 more...	54
240601.39	Estimated Gross Loss = 212,800 to 342,900 Business = Asset Management; Retail Banking Quarter Occurred = Q1; Q2; Q4 more...	152

Save your discovery set using the save button in the upper right toolbar.

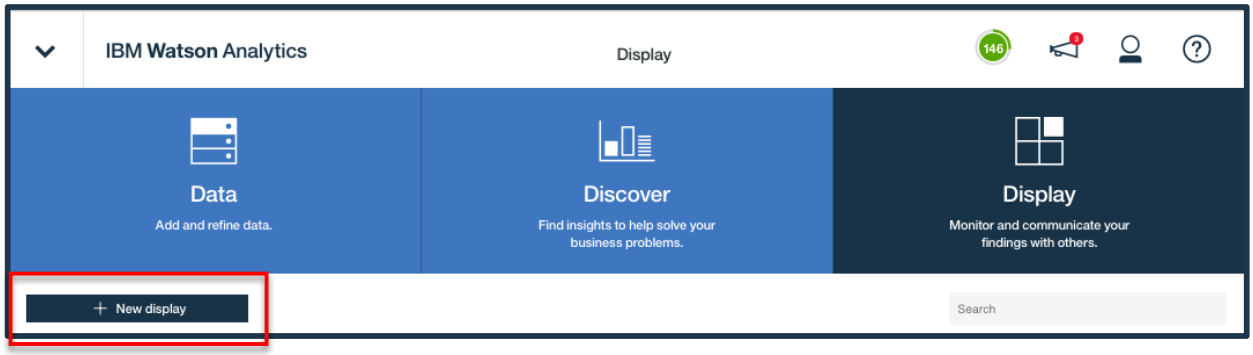
Display

Display Basics

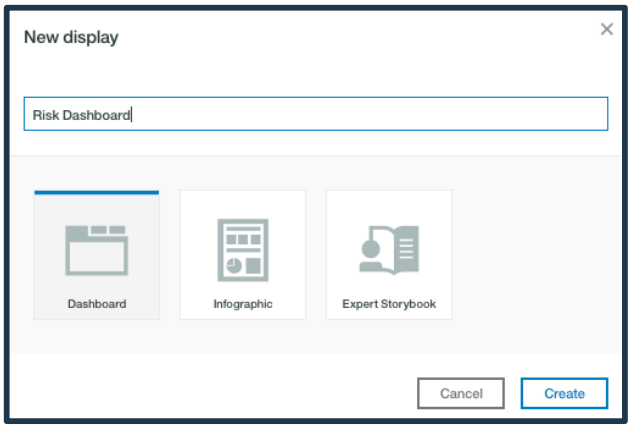
The Display capability allows you to assemble multi-tab, interactive dashboards that span one or more Data Assets. Displays are built using saved discoveries from one or more Discovery Sets.

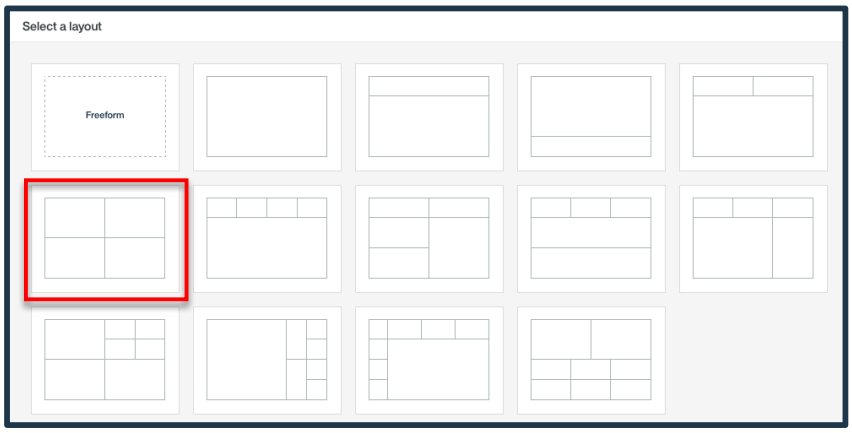
1. Navigate back to the welcome page, using the page navigator

2. Click on Display and click “New Display”



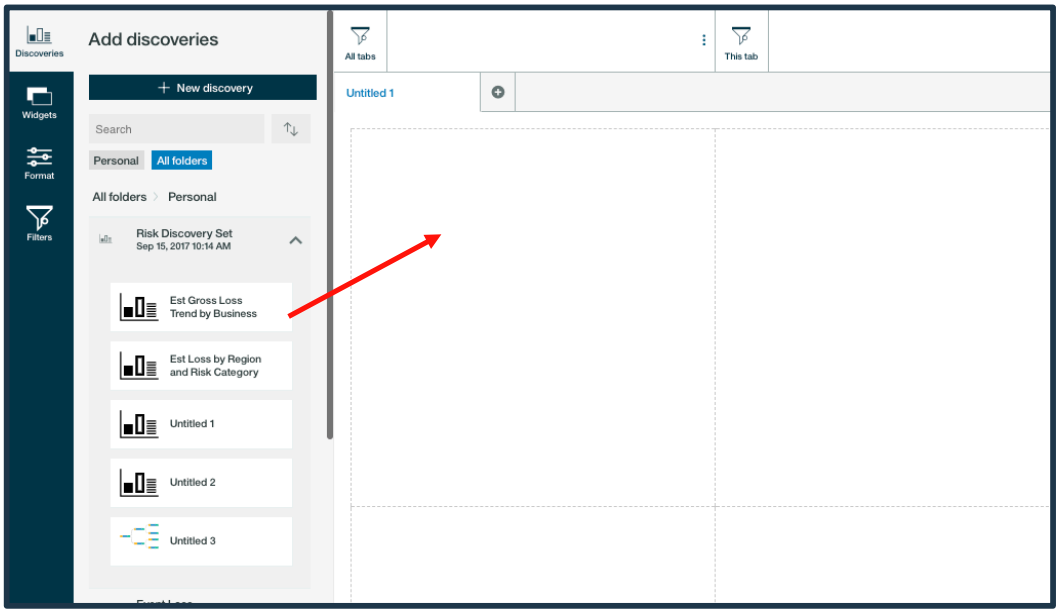
3. Modify the name of the dashboard if desired, and choose a dashboard template, then select a basic 4x4 dashboard template.



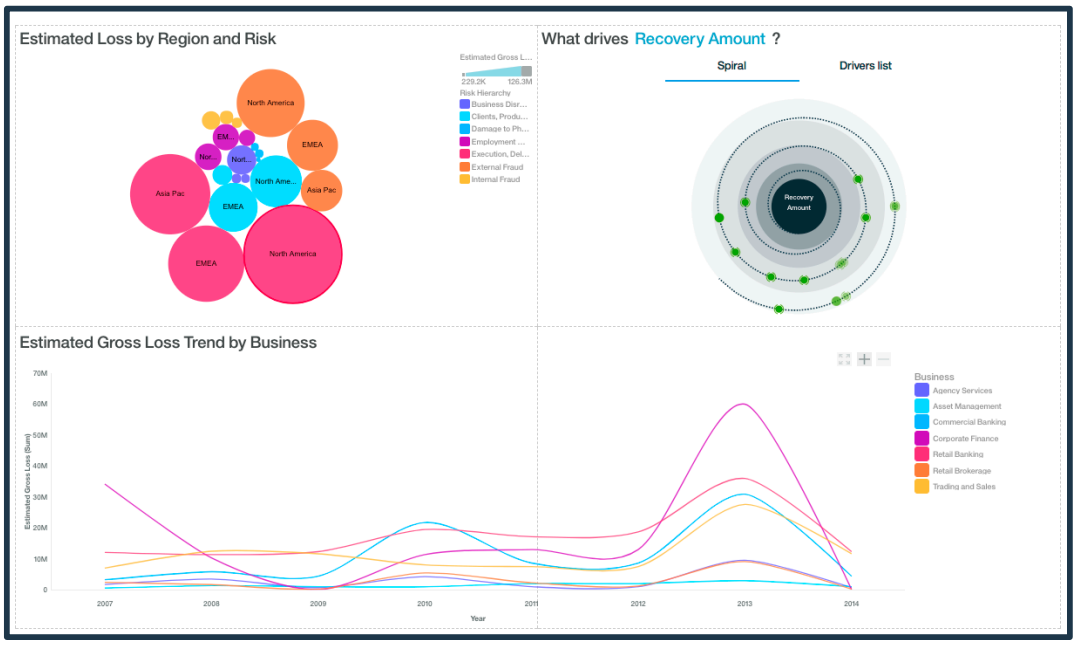


- A starting template with gridlines facilitates the creation of a dashboard. Objects can automatically snap into and fill grid boxes. To do this, drop your object in the center of a grid box - when it turns blue, release the object and it will snap into place

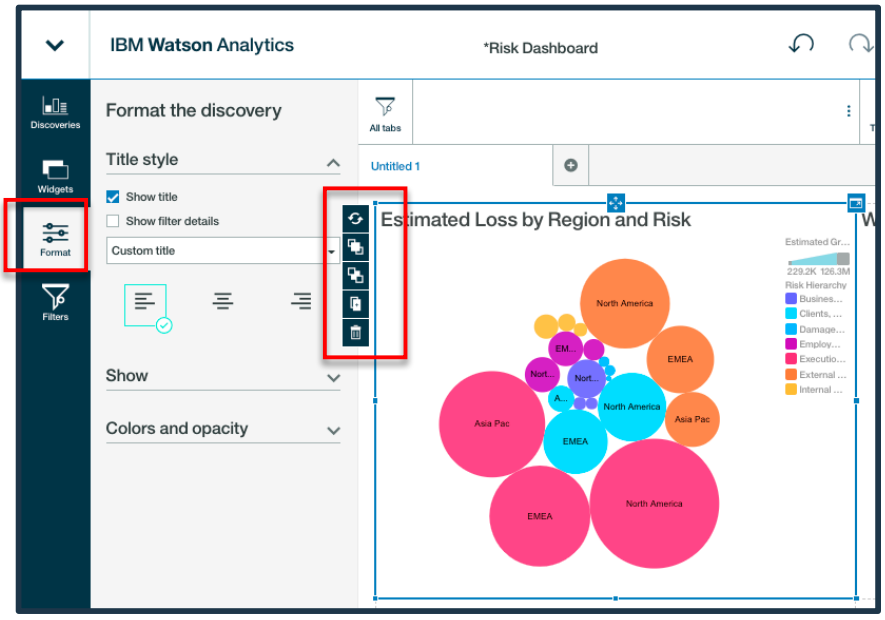
4. Navigate on the left to the location of the discovery set you saved in the previous step and drag the Estimated Loss by Region and Risk Category (bubble chart) into the upper-left box.
   Note: As you're dropping the discovery into the drop zone



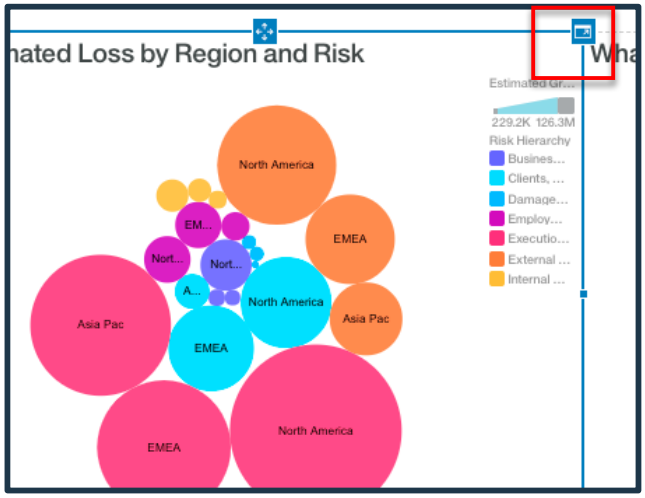
5. Repeat step 4 to drop the remaining discoveries into boxes. Drop the spiral discovery into the upper right drop-zone and the Estimated Loss Trend by Business (line chart). You can re-size the line chart to fit the lower half of the chart.



6. To duplicate or delete a discovery after its been dragged in, click in the whitespace. This displays a popup menu. To see the properties of an option, click on the “Format” icon on the left hand side.



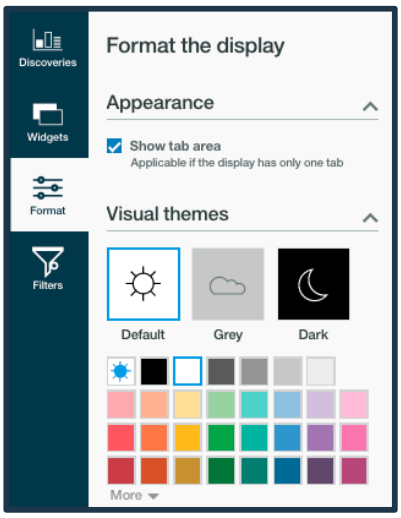
- When selected, an object:
  i. can be resized using the grab handles
  ii. can be moved by clicking and dragging using the arrows button
  iii. the top right corner arrow maximizes the visualization and shows you the underlying data items for that visualization



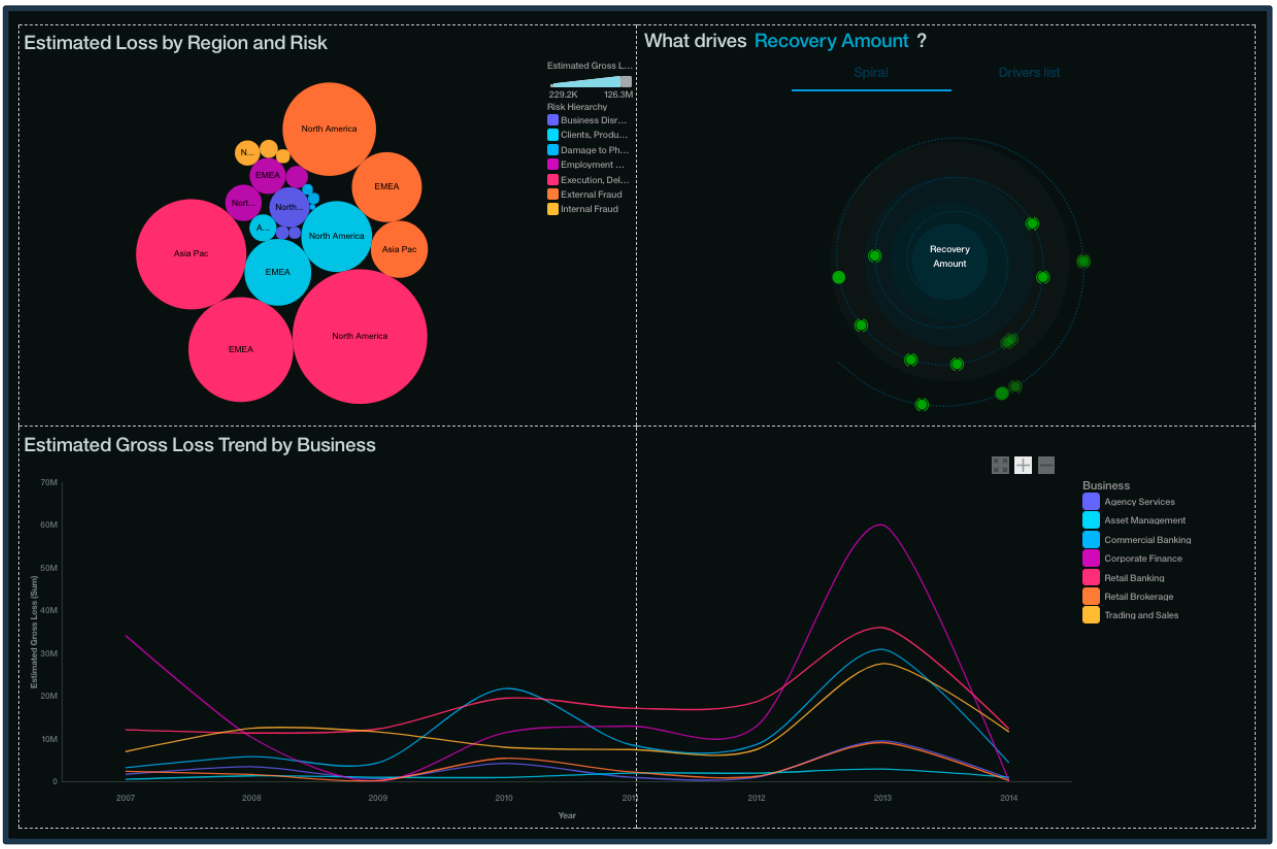
Click on the visualization icon on the left to change the visualization
Use the drop zones at the bottom to modify the visualization

Changing Themes

1. To change the theme (built-in color palettes), click on any white space outside of the dashboard and click the format button on the left toolbar



2. You can toggle any of the built-in themes and keep a preference of your choosing, for instance the dark theme:

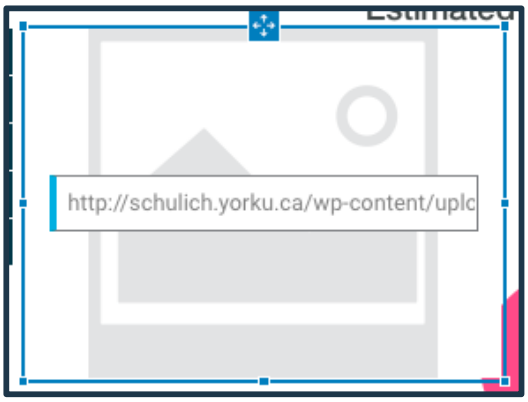


Note: You can also change the background to any of the colors you prefer

Adding Other Objects

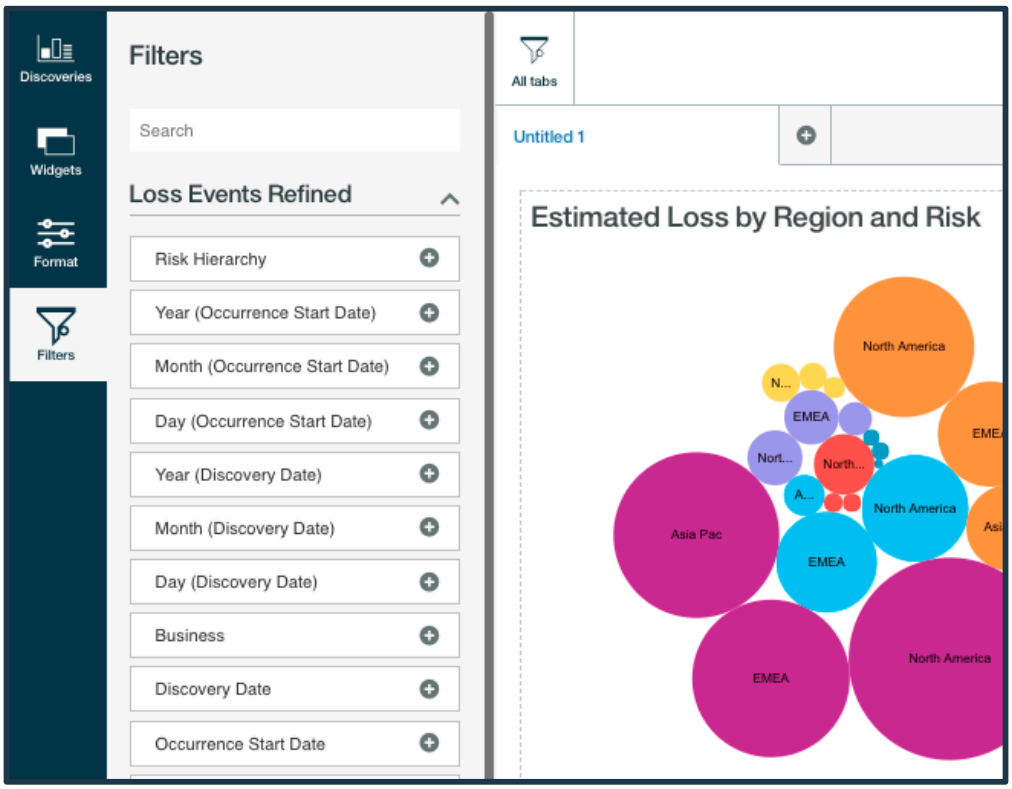
Click on Widgets on the left hand side of the Display screen to add text, images, videos, webpages.

Note: You can do a google search for an image of your choosing and save the image location, then paste into the widget



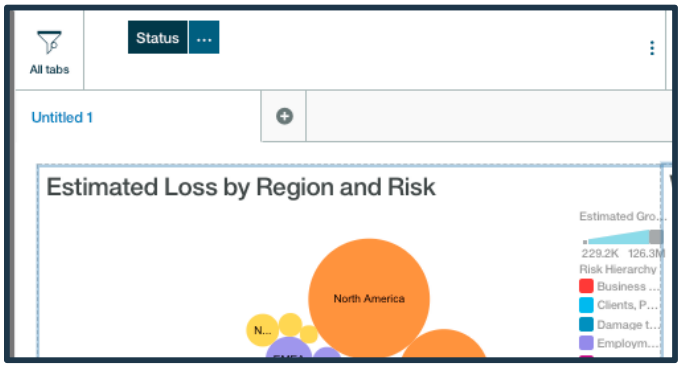
Filters

Filters are columns from a data asset that can be used to filter all tabs in a dashboard (global filter) or for a specific tab (local filter)

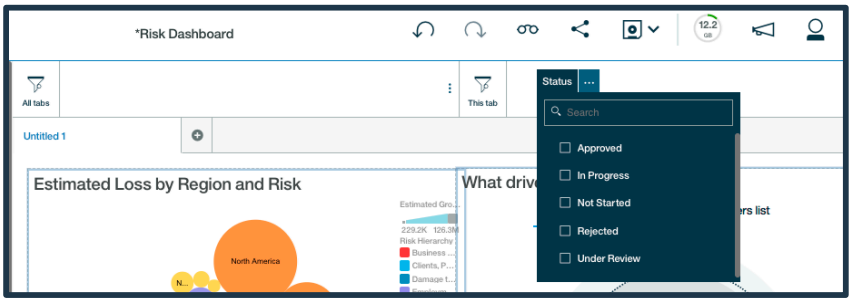


Display

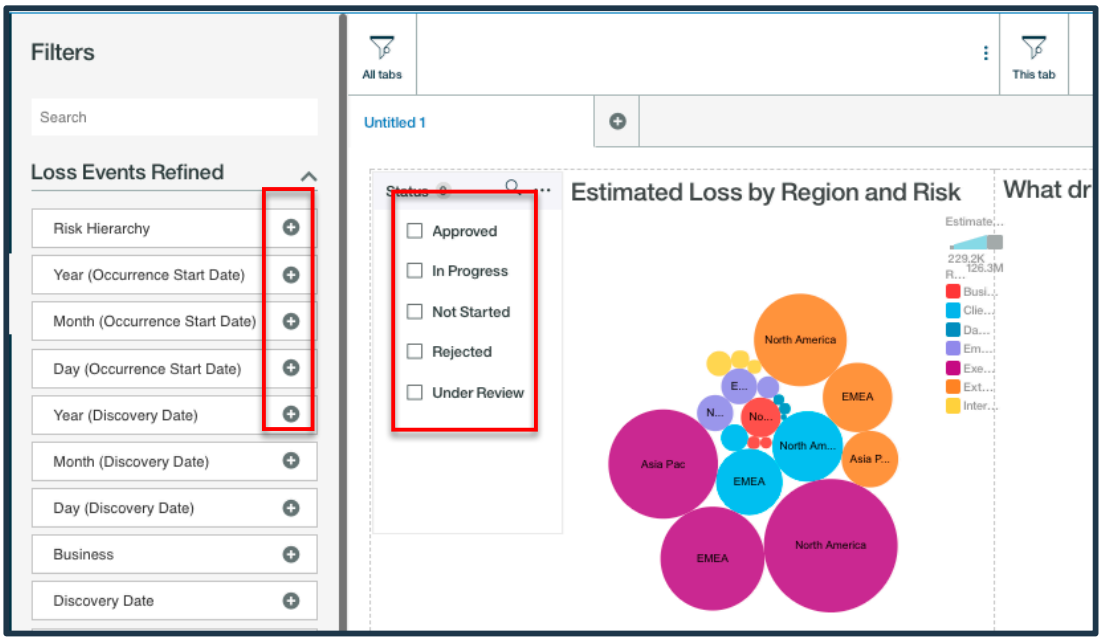
For a global filter, drag a column into the “All tabs” area. Clicking on the filter will then give you the option to select the items you want.



For a local filter, drag a column into the “This tab” area. Alternatively, you can also drag a column right onto the dashboard - this will act as a local filter as well.



For a local filter, click the “plus” sign to the right of the object to add it to your dashboard

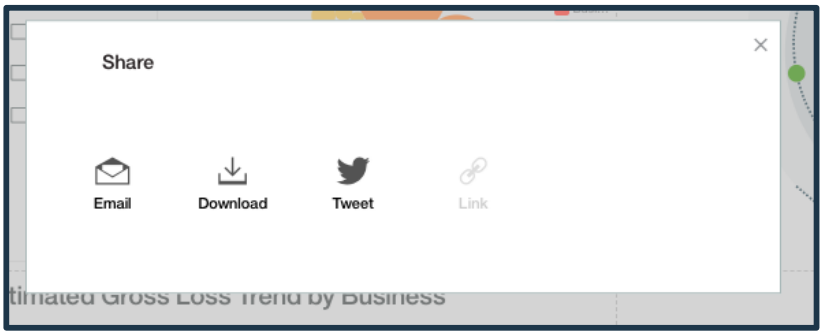


Sharing

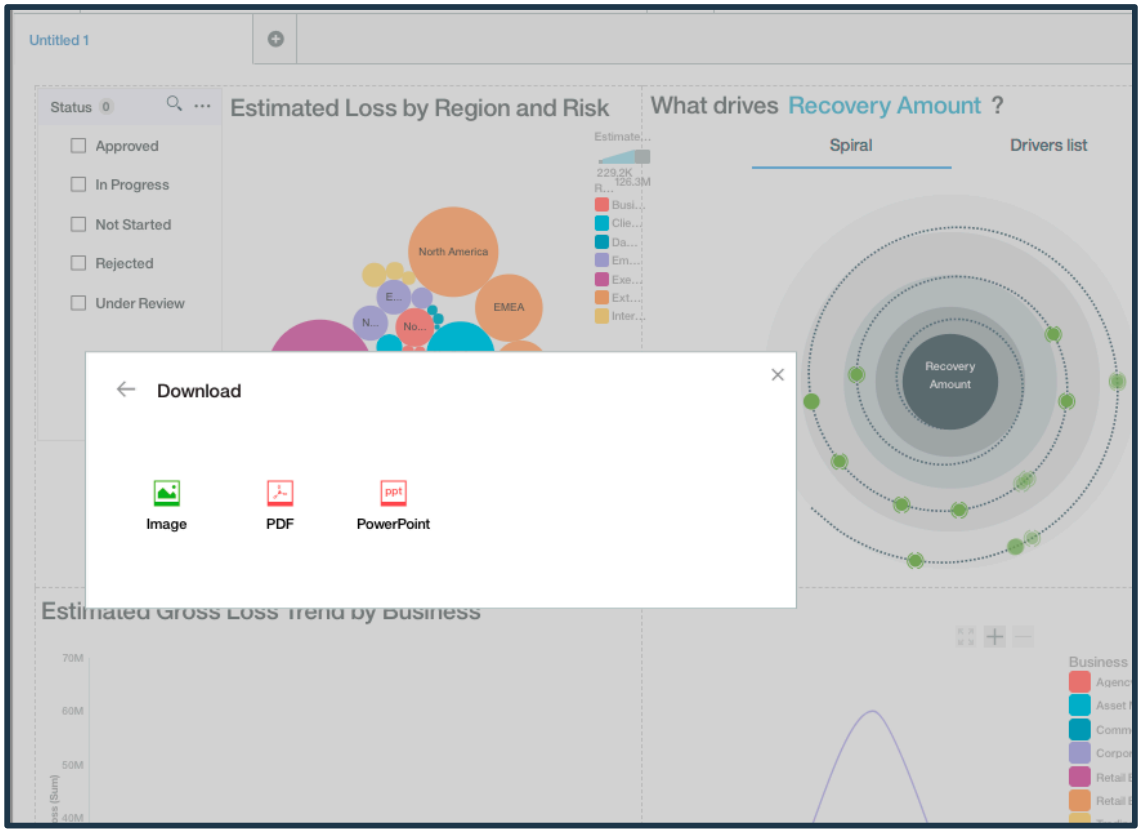
After you have populated your dashboard with your insights you can share it with others in your organization in a number of ways:

1. Selecting the share button from the tool bar

- Now you can choose to share via email, you can tweet out your dashboard or you can download it



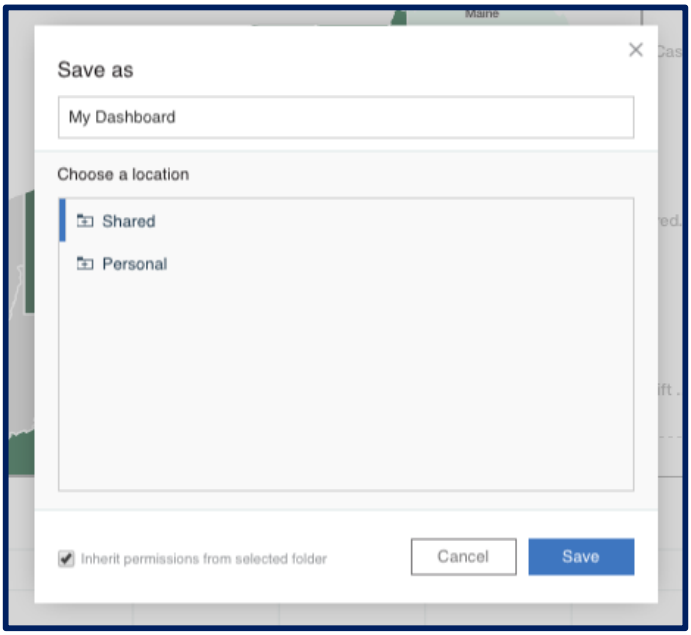
- If you choose to **download** your dashboard, you have the option of downloading as an image, a PDF or a PPT



2. You can share your display asset with coworkers in your Watson Analytics tenant:

   - Click on 'Save As' in the toolbar to save your dashboard:

   - Here you can choose to save in the **shared** area:



Now your asset will appear in the shared folder in the display area of Watson Analytics