


# Create a Statistical Model (Datasets)

Metric Insights allows Users to easily create a Statistical Model to find anomalies in data containing a large number of Dimension Values. This process can surface specific anomalies across thousand of records and Dimensions, allowing Reports and Metrics to be easily created with this reduced information.

For an example, we will create a Stats Model from an existing Dataset containing sales data from our test database. We want to find any values of 'Cost of Product', 'Sales Amount' or 'Gross Profit' that fall outside of 2 standard deviations from a 30-day moving average. We want to perform this evaluation across these dimensions: Country, Channel, and Product sub-category.

 You can only create Stats Models on Datasets that 'keep history', aka [Snapshot Datasets](#), and you need to use the 'All Data' View.

 To learn more about security requirements for Power Users, check [Understanding Power Users](#) and [Dataset/User Map Security Overview](#).

## 1. Creating Statistical Models from the Dataset Viewer

 Stats Models can only be created from the 'All Data' Views.

Dataset collected: Sunday 12/31/2017

Results

calendar_date	country	channel	product_subcategory	1 sum(total_cost)	2 sum(total_sales_amount)	3 sum(total_gross_profit)	4 avg(percent_gross_margin)
08/03/17	Australia	corporate sales	bottle opener	\$5,486	\$8,537	\$3,051	37.9
08/03/17	Australia	corporate sales	champagne	\$18,315	\$24,542	\$6,227	27.4
08/03/17	Australia	corporate sales	Chardonnay	\$9,735	\$16,613	\$6,878	41.6
08/03/17	Australia	corporate sales	Cylinder	\$7,607	\$11,550	\$3,943	34.1

In our example, the **All data View** of the Dataset provides the following:

### Measures :

1. Cost of product
2. Sales amount
3. Gross profit amount

### Dimensions:

- Product\_subcategory
- Country
- Channel

## 2. Defining Settings

Access **Actions > Build Stats Model** to open the Stats Model Editor

**1** Name: Daily Sales Data Statistical Model

**2** Date column: calendar\_date

**3** Moving Average Interval: 10 days

**4** Calculate UCL and LCL based on 3 standard deviations

**5** Only include rows where volatility is greater than 3 standard deviations

**6** Compute volatility for: Latest calendar\_date

**7** Dimensions: country

**8** Measure: sum(total\_cost)

Include totals for each Dimension: ☐

Show Dimensions with NULL values as: N/A

Include column to show particular measure(s) where the anomaly occurs: ☐

Save or cancel

calendar_date	country	channel	product_subcategory	sum(total_cost)	sum(total_sales_amount)	sum(total_gross_profit)	avg(percent_gross_margin)
08/03/17	Australia	corporate sales	bottle opener	\$5,486	\$8,537	\$3,051	37.9
08/03/17	Australia	corporate sales	champagne	\$18,315	\$24,542	\$6,227	27.4
08/03/17	Australia	corporate sales	Chardonnay	\$9,735	\$16,613	\$6,878	41.6

- 1. Name:** Default is 'Statistical Model' appended to your Dataset Name
- 2. Date column:** Choose a date from the Results field or use 'Snapshot Date' if your Dataset is snapshotted
- 3. Moving Average Interval:** Select a moving average range that will produce statistically relevant data
- 4. Calculate UCL and LCL based on .. standard deviations:** Upper (UCL) and Lower Control Limits (LCL) provide calculations used in I-MR statistics. Provide the multiplier used to generate upper and lower control boundaries
- 5. Only include rows where volatility is greater than .. standard deviations:** Set your Volatility limit to control the number of records returned in your Stats Model. The generated Stats Model returns only records that fall outside of a specified number of standard deviations. Lower numbers will include more significantly anomalous results
- 6. Compute volatility for:** You can compute volatility for the either the current (latest) Calendar date or for all Calendar values
- 7. Dimensions:** Dimension Value is defaulted, but can be changed or added to by clicking **[+Dimension]**
- 8. Measure:** Measure values is also defaulted but can be changed or added to by clicking **[+Field]**

### 3. Select Dimensions (Filters) and Measures for the Stats Model Results

Home Datasets / Daily Sales Data / All data ▲ 0 New... Content Admin Alex ?

Dataset Collected: Monday 01/01/2018

Back All data

Name: Daily Sales Data Statistical Model

Date column: calendar\_date

Moving Average Interval: 30 days

Calculate UCL and LCL based on 3 standard deviations

Only include rows where volatility is greater than 2 standard deviations

Compute volatility for: Latest calendar\_date

**1 Dimensions**

- country
- channel
- product\_subcategory

☒ Include totals for each Dimension

Show Dimensions with NULL values as N/A

**2 Measure**

- sum(total\_cost)
- sum(total\_sales\_amount)
- sum(total\_gross\_profit)

**3** + Derived Field

**4** ☒ Include column to show particular measure(s) where the anomaly occurs

**5** ☒ List one measure per row

**6** Save or cancel

**Results**

Show: All of 62588 rows

calendar_date	country	channel	product_subcategory	sum(total_cost)	sum(total_sales_amount)	sum(total_gross_profit)
2017-08-03 00:00:00	Australia	e-mail marketing	white wine	8,069	12,863	4,795
2017-08-03 00:00:00	Australia	e-mail marketing	wine glasses	10,890	16,133	5,243
2017-08-03 00:00:00	Australia	e-mail marketing	wine rack	3,762	5,174	1,412
2017-08-03 00:00:00	Australia	store visit	bottle opener	7,812	12,084	4,272

- Dimensions:** There are 3 filters selected. The **Include Totals for each Dimension** checkbox is also selected to generate stats for every unique combination of filter values against all other filter values including aggregated totals. In this example, a Total Value will be calculated for All Countries, All Channels, All Product-subcategories, as well as all channels and products for each country, all countries and products for each channel, etc
- Measures:** Select those measures you are interested in
- [+Derived Field]** to add any number of computed fields
  - For more details about Derived Fields, check [Understanding Derived Fields](#) article
- Include column to show particular measure(s) where the anomaly occurs:** Check to include a generated column that concatenates all Measure Names greater than your specified volatility limit (in this example, all measures with values greater than 2 standard deviations from 30-day moving average)
- List one measure per row:** Select to alternately display one row for each separate measure with an anomaly instead of concatenating
- [Save]** to create your Stat Dataset

**NOTE:** Stat Dataset creation process may take some time and you can exit this page while processing is still going

## 4. Review Results of One Row Per Measure Setting

Results

country	channel	product_subcategory	Measure	Value	Average	SD	NUM SD	UCL	LCL
Australia	corporate sales	full cabinet	sum(total_cost)	82,863	35,823	19,367	2.43	93,925	-22,279.4
Australia	corporate sales	wine rack	sum(total_cost)	18,380	9,781	4,258	2.02	22,556	-2,993.7
Australia	e-mail marketing	Chardonnay	sum(total_cost)	35,831	20,990	6,442	2.30	40,317	1,663
Australia	e-mail marketing	half cabinet	sum(total_cost)	5,365	2,081	1,483	2.23	6,530	-2,387.5
Australia	e-mail marketing	red wine	sum(total_cost)	142,418	79,644	26,636	2.36	159,557	-269.9
Australia	e-mail marketing	white wine	sum(total_cost)	72,581	39,161	14,739	2.27	83,377	-5,055.2
Germany	corporate sales	champagne	sum(total_cost)	18,769	46,743	13,885	2.01	68,397	5,089
Australia	corporate sales	full cabinet	sum(total_sales_amount)	117,405	52,293	27,273	2.39	134,113	-29,525.8
Australia	corporate sales	wine rack	sum(total_sales_amount)	25,012	13,392	5,800	2.00	30,792	-4,008.3
Australia	e-mail marketing	Chardonnay	sum(total_sales_amount)	61,316	35,945	10,980	2.31	68,885	3,006
Australia	e-mail marketing	half cabinet	sum(total_sales_amount)	10,618	4,030	2,925	2.25	12,805	-4,745.2
Australia	e-mail marketing	red wine	sum(total_sales_amount)	216,856	122,494	40,952	2.30	245,348	-360.6
Australia	e-mail marketing	white wine	sum(total_sales_amount)	121,667	63,049	26,067	2.22	142,048	-14,351.2
Germany	corporate sales	champagne	sum(total_sales_amount)	17,882	44,556	13,211	2.01	84,190	4,923
Australia	corporate sales	full cabinet	sum(total_gross_profit)	34,542	16,471	7,952	2.27	40,326	-7,384.2

## 5. Editing Stats Model

To change options access **Edit > Edit Statis Model**

Remove the option to List one measure per row - and Save

Save or cancel

## 6. Reading the Results

country	channel	product_sa...	Measures over 2 Standard Deviations	calendar_d...	sum(total_...	sum(total_...	sum(total_...	sum(total_...	sum(total_...	sum(total_...	sum(total_...	sum(total_...	sum(total_...	sum(total_...	sum(t...	sum(t...
Australia	corporate sa...	full cabinet	sum(total_cost), sum(total_sales_amount), sum(total_gross_profit)	2017-12-31 ...	82,863	117,405	34,542	2.43	19,367	35,823	93,825	-22,279.4	2.38	27,273	52,293	134,...
Australia	corporate sa...	wine rack	sum(total_cost), sum(total_sales_amount), sum(total_gross_profit)	2017-12-31 ...	18,380	25,012	6,632	2.02	4,258	9,781	22,556	-2,993.7	2.00	5,800	13,392	30.7
Australia	e-mail mark...	Chardonnay	sum(total_cost), sum(total_sales_amount), sum(total_gross_profit)	2017-12-31 ...	35,831	61,316	25,484	2.30	6,442	20,990	40,317	1,663	2.31	10,980	35,945	68.0
Australia	e-mail mark...	half cabinet	sum(total_cost), sum(total_sales_amount), sum(total_gross_profit)	2017-12-31 ...	5,385	10,618	5,232	2.23	1,483	2,681	6,530	-2,367.5	2.25	2,925	4,830	12.8
Australia	e-mail mark...	red wine	sum(total_cost), sum(total_sales_amount), sum(total_gross_profit)	2017-12-31 ...	142,418	216,856	74,439	2.36	26,638	79,644	159,557	-269.9	2.30	40,952	122,494	245.
Australia	e-mail mark...	white wine	sum(total_cost), sum(total_sales_amount), sum(total_gross_profit)	2017-12-31 ...	72,581	121,667	49,086	2.27	14,739	39,161	83,377	-5,055.2	2.22	26,067	63,849	142.
Australia	website visit	gift set	sum(total_gross_profit)	2017-12-31 ...	25,554	32,579	7,025	1.90	8,873	42,397	69,015	15,780	1.96	11,424	54,954	89.2
Germany	corporate sa...	champagne	sum(total_cost), sum(total_sales_amount), sum(total_gross_profit)	2017-12-31 ...	18,769	17,862	-787.8	2.01	13,885	46,743	88,397	5,089	2.01	13,211	44,556	84.1
Germany	website visit	champagne	sum(total_gross_profit)	2017-12-31 ...	60,291	56,549	-3,741.6	0.68	14,130	50,663	93,052	8,274	0.62	13,800	48,039	89.4

Review **Results** of this Model will quickly highlight those elements with Anomalies:

1. For email Marketing in Australia, Chardonnay has anomalies in all 3 Measures - *Total Cost*, *Total Sales Amount*, *Total Gross Profit*
2. For Website Visits in Australia, the Gift Set has only one measure showing an anomaly - *Total Gross Profit*
3. And Germany is showing anomalies for Champagne in two channels

## 7. Save Stat Model as a New View

You can create Rules for the Stats Dataset and save these as Views.

1. **[Save as View]**
2. **Name:** Add meaningful name for your View

3. **Visibility:** Select "Public" if you want others to be able to View and Use this Stat Model, otherwise it this view will only be available to you
4. **[Save]**
5. Statistical Models are saved as Views

💡 All editing of Statistical Models must be done from the Dataset Viewer as Stat Datasets will not appear in Dataset List (**Content > Datasets**)

## 8. Creating Reports and Metrics from the View

111

Selecting **Action > Build Report** will take you directly to a defaulted report that you can edit, or simply Publish and Enable to display on your Homepage.

💡 For details on creating elements from Statistical Datasets refer to this article for more information: [Sourcing Reports / Metrics / Dimensions from "Existing Datasets - SQL"](#)