

This article aims at walking you through creating a Topic Modeling study and showing you how to explore the results.

### Agenda:

- What is Topic Modeling?
- Topic Modeling Approach
  - Three Broad Benefits of Topic Modeling
  - Bottom-up Approach
- Topic Modeling Study Example
  - Defining the set of mentions
  - Process
  - Launch a Topic Modeling study
  - Results of the study
  - Deep-dive into the cluster themes
  - Settings
  - Manage the clusters
    - Hide a cluster
    - Group and detach clusters
  - Convert a cluster into a Topic or a Classifier
    - Track cluster theme over time
    - Convert as a Topic or Classifier
    - Edit and enrich your Classifier or Topic
    - Incorporate your Classifier or Topic into your reporting

Should you want to learn more about Topic Modeling:

- [Watch our webinar](#) to see Topic Modeling in action
- [Read our Best Practice Guide](#)

---

## What is Topic Modeling?

Topic modeling is an unsupervised machine learning technique that's capable of scanning a set of documents, detecting word and phrase patterns within them, and

**automatically clustering word groups and similar expressions that best characterize a set of documents.**

Thanks to advances in computing power and data science, Synthesio has developed AI-powered text mining that scans millions of online data-points and surfaces hidden, consumer-defined conversation themes. By productizing over 4 years of Ipsos' methodology and expertise, Topic Modeling provides brands with a fast, scalable approach to turning unstructured social data into intelligence.

---

## **Topic Modeling Approach**

Topic analysis models enable you to sift through large sets of data and identify the most significant themes in a very simple, fast, and scalable way.

### **Three Broad Benefits of Topic Modeling**

- **Topic Detection**
  - Uncover the unintuitive or underlooked topics that matter most to your consumers.
  - Identify these topics quickly and early-on as they emerge
- **Dashboard Creation**
  - No matter our best intentions, topic taxonomy may not reflect how consumers naturally discuss topics
  - Topic Modeling helps us validate and consider all the topics that are important to our consumers
- **Data Analysis**
  - Eliminate the need to manually analyze millions of mentions.

### **Bottom-up Topic Modeling Approach**

This approach enables to create clusters of mentions talking about related themes.

The bottom-up approach is the ability to infer organic patterns and cluster similar expressions without needing to define topic tags or train data beforehand.

- **Advantages**
  - Personalized insights as the model adapt to your dataset
  - Provides unique insights, i.e., the unknown unknowns

---

## Topic Modeling Study Example

### Define the set of mentions

The first thing to do is define the set of mentions you want to launch your Topic Modeling study on. Here are some reminders and advice:

- Topic Modeling studies are limited to 600K mentions.
- If the total number of mentions in your dashboard is higher than 600K, you need to filter your data and narrow down your scope to be able to launch a Topic Modeling study.
- Remember the adage "Garbage in - garbage out": should you want to have satisfying results after running your Topic Modeling study, try to choose a set of mentions as homogeneous as possible. Indeed, if the mentions you picked for your study are too diverse or talking about topics that have nothing in common, it's likely that your results will not be insightful.

We would recommend that you browse through your mentions stream in the Listen Tab to create favorite filters with an appropriate number of mentions to launch Topic Modeling Studies. You can create multiple favorite filters and launch multiple Topic Modeling studies.

### Process

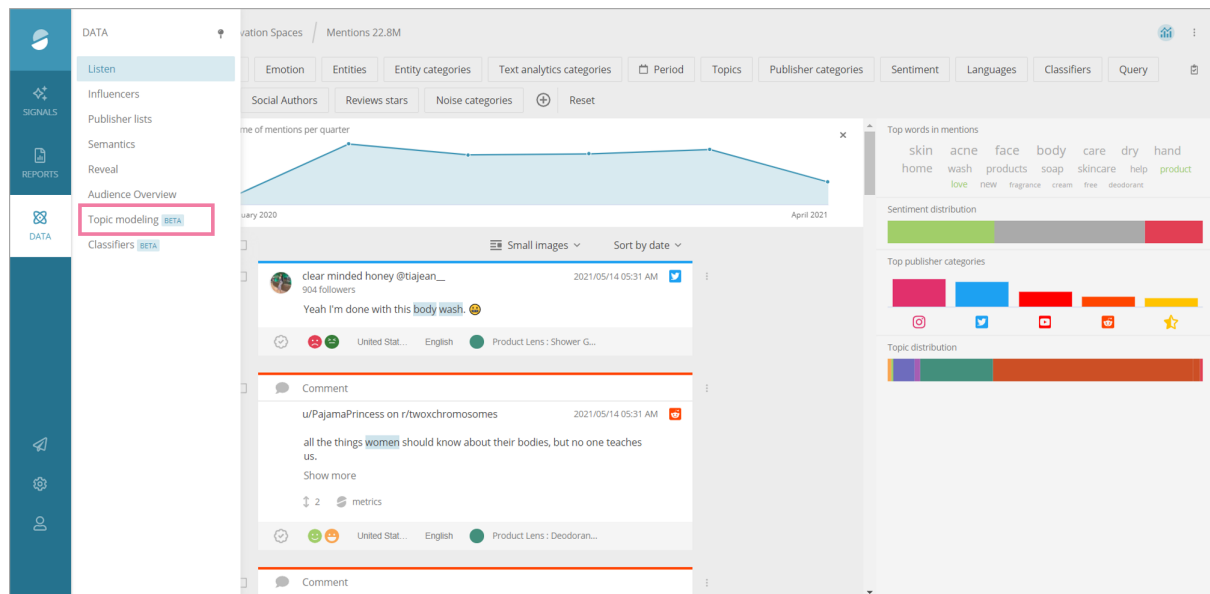
Here are some explanations on how the model is working once you chose your set of mentions for your Topic Modeling study.

- 3-step process:
  - Pre-processing: Clean the data and remove noise
  - Clustering: Our multilingual AI model interprets each mention and regroups similar discussions
  - Post-processing: Visualize the most representative posts and ease insight discovery

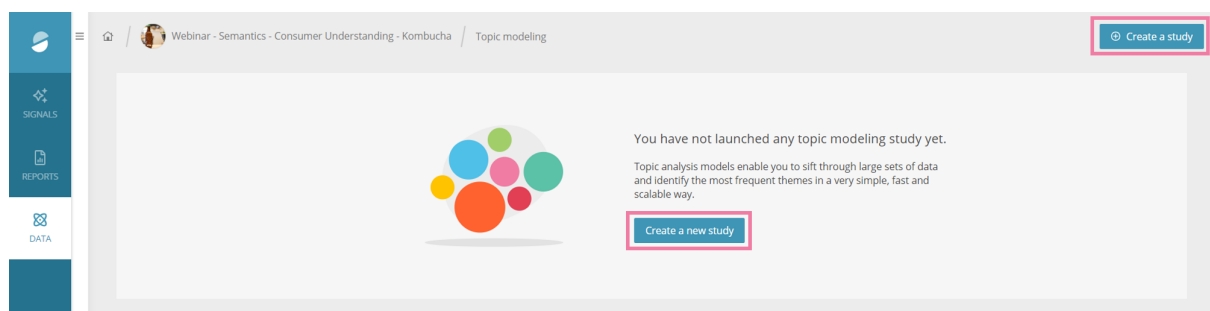
- After the cleaning, your study is run on a sample of 100K mentions within the chosen set of mentions
- You can then see the results and the different clusters that have been detected

## Launching a Topic Modeling Study

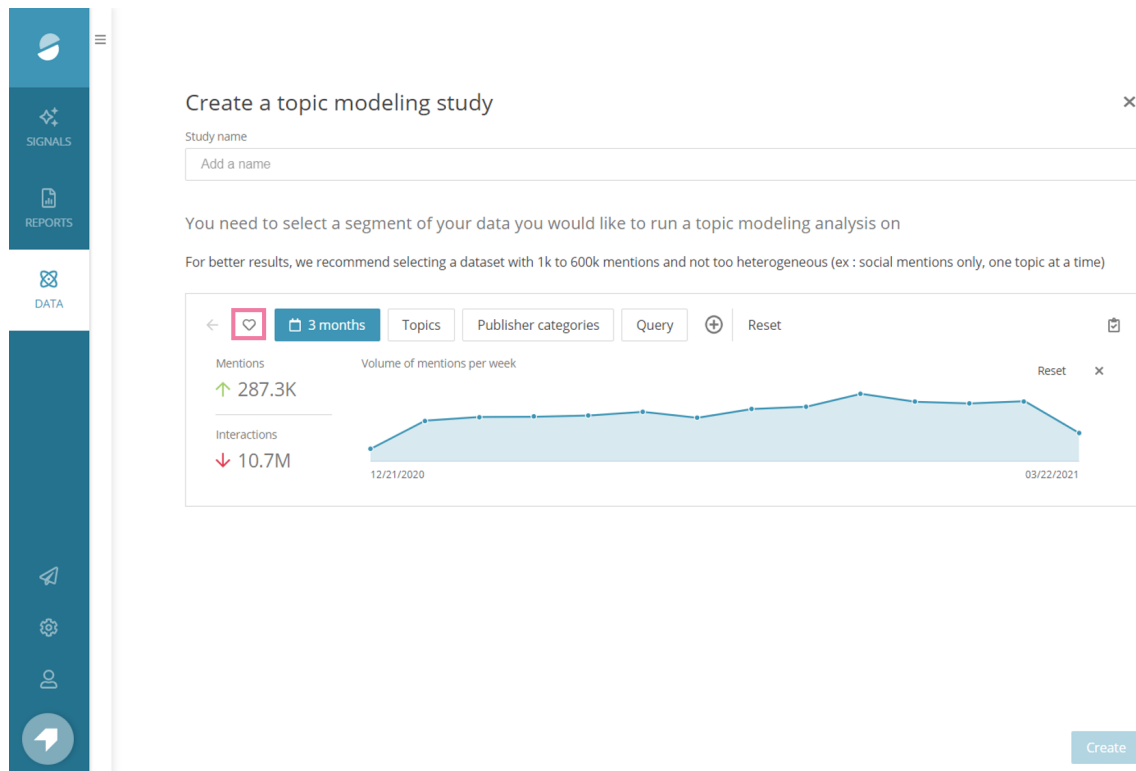
In the data section, click on the Topic Modeling tab.



You land on the Topic Modeling home screen. Click 'create a new study'.



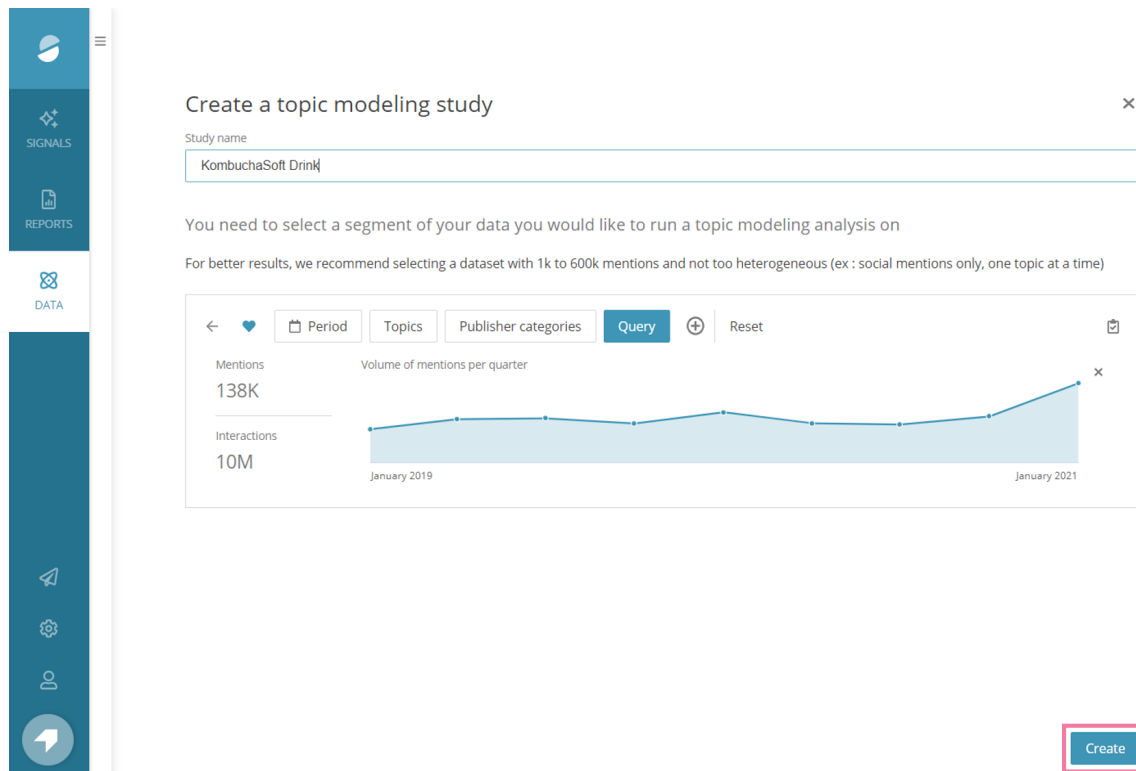
After clicking on Create a new study, you need to define the scope of your study.



You have two main options:

1. Either you already created a favorite filter with the scope for your study. In this case, you just need to apply it here by clicking on the "Heart" icon
2. Or you can play with the filters here and narrow down your scope to maximum 600K mentions

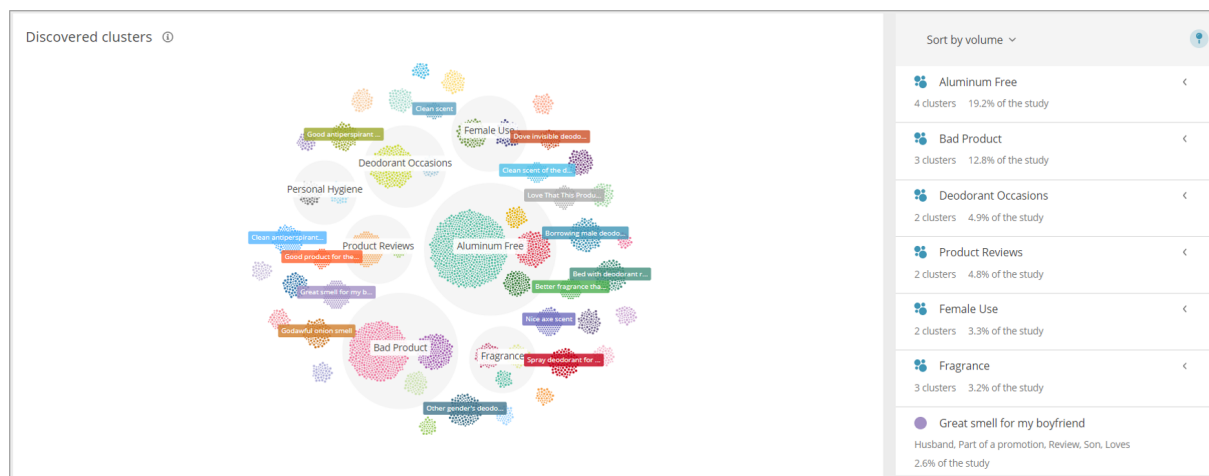
Don't forget to name your study before clicking "Create".



Once launched, your study will take 7 to 30 minutes to run. You can still use other components of the dashboard experience while your study runs.

## Results of the study

When you access the results of your study, you can see the different clusters that have been detected in your set of mentions.



On the left, you see a two-level visualization showing you how the mentions are grouped into clusters and groups of clusters. Each cluster and group of clusters are automatically named by the algorithm. You can rename the clusters and groups of clusters on the fly if the names are not relevant to you.

On the right, you have the list of groups of clusters with:

- The name
- The number of clusters in the group
- The weight of the group in the study (percentage of mentions compared to the whole study)

For clusters that do not belong to any group, you have the list of clusters with:

- The name
- The number of mentions
- The weight of the cluster in the study (percentage of mentions compared to the whole study)
- Main keywords detected in the cluster

### **Warning:**

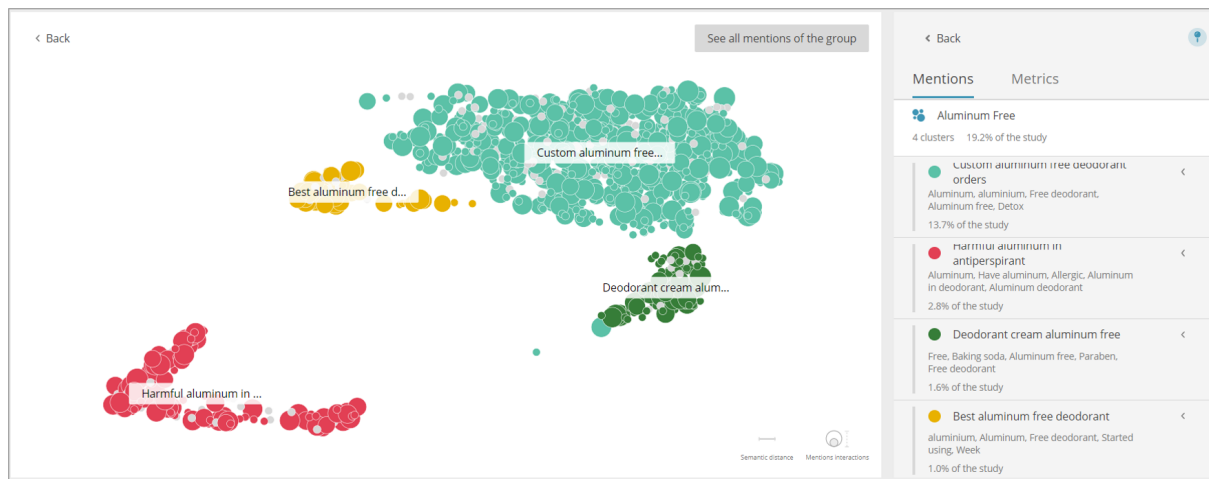
If you sum the mentions in each cluster, it won't match the total number of mentions you chose as an input of your study.

This is due to multiple factors:

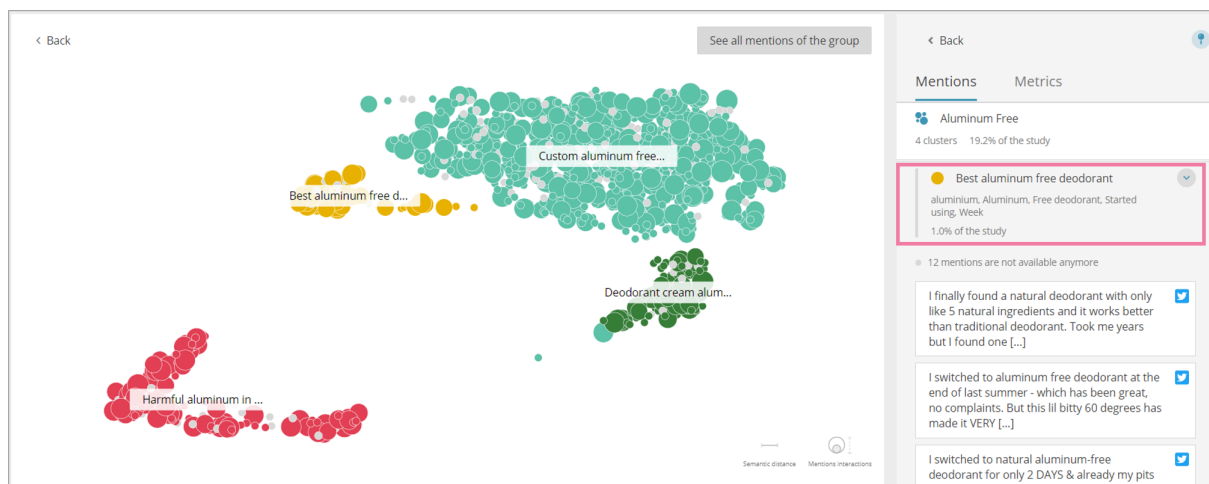
- As said above, a Topic Modeling study is run on a sample of the total set of mentions chosen by the user.
- Pre-processing and post-processing phases remove mentions that are not relevant.
- Mentions that are relevant but that do not match any of the detected clusters will also be left aside.

## **Deep-dive in clusters and groups of clusters**

You can select a group of clusters to see it in more detail. On the left, you see how the different clusters are positioned, and on the left you have the details of each cluster.

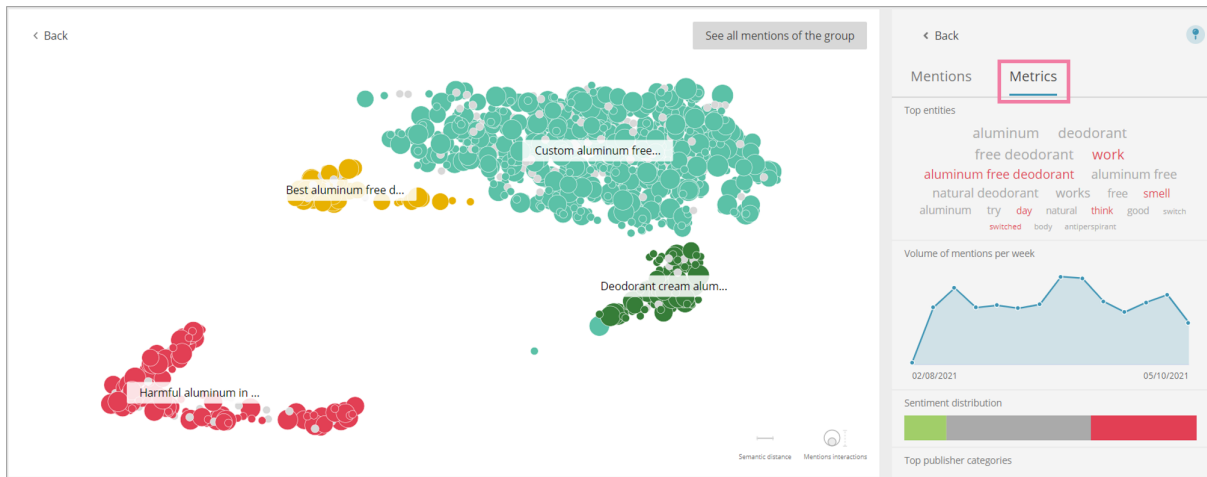


You can click on each cluster of the group to see the mentions associated with it.

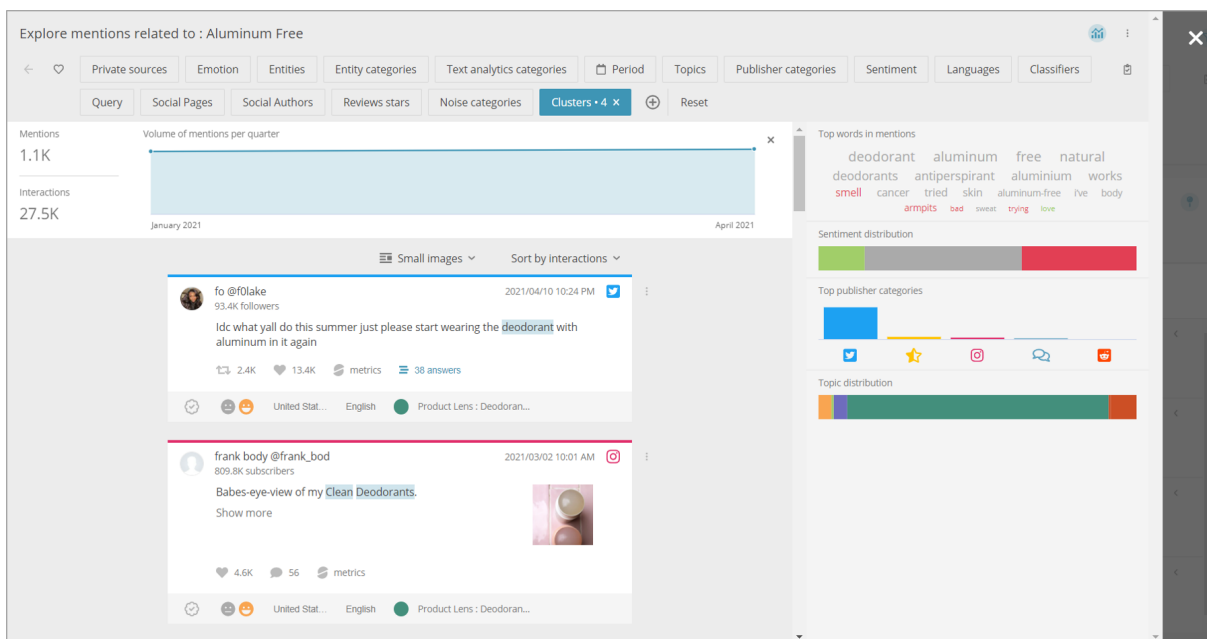


If you click on Metrics, you access the main metrics of the group at a glance:

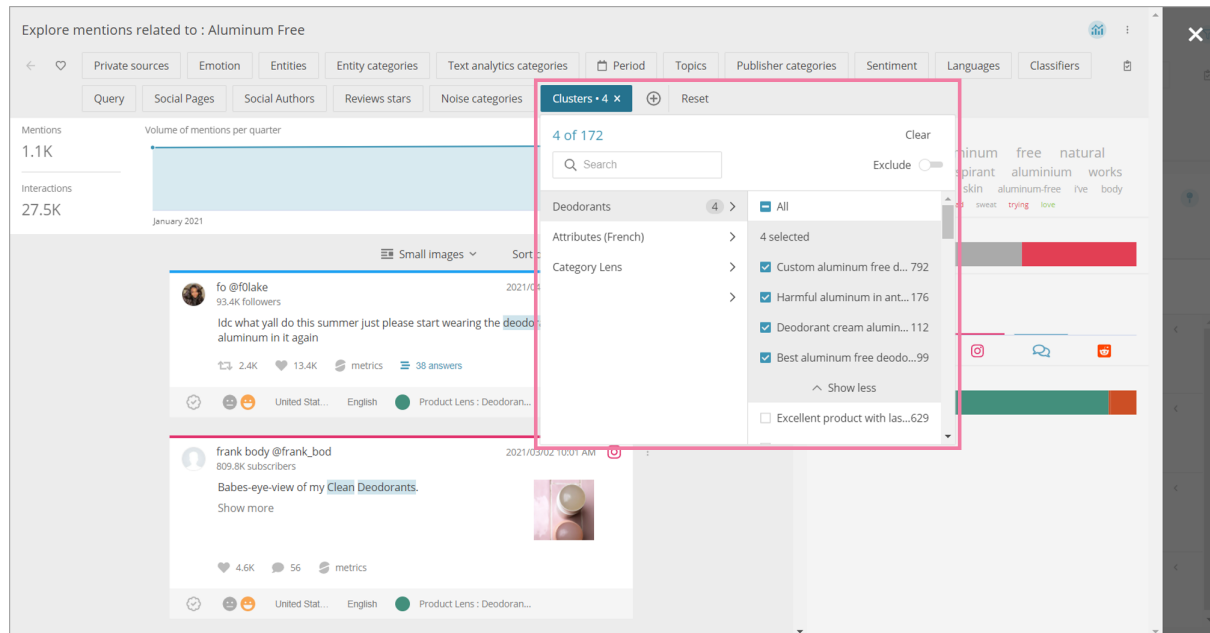
- Main expressions and keywords associated
- Evolution of volume of mentions over time
- Sentiment distribution
- Top publishers
- Topic distribution



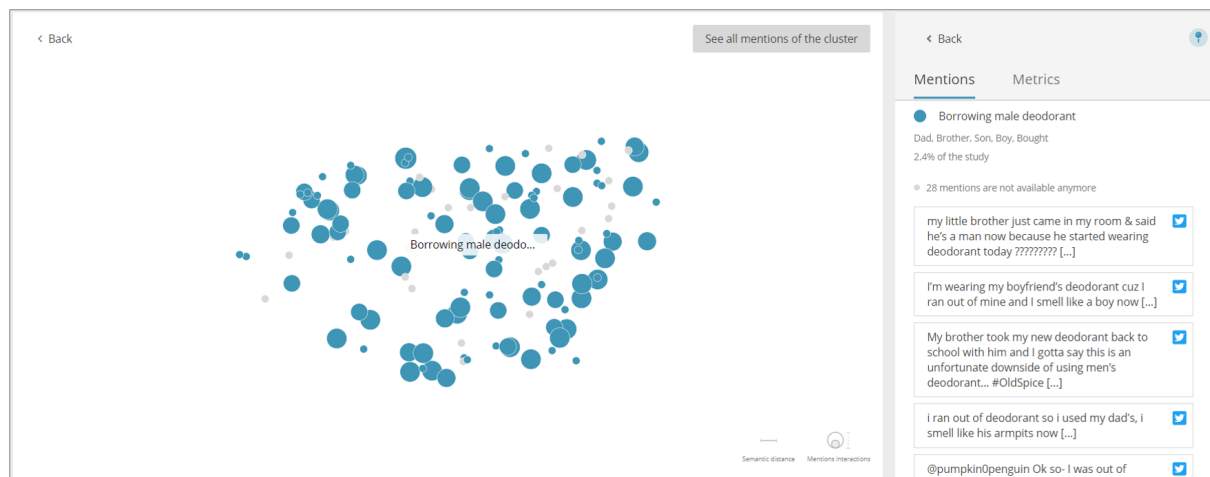
Finally, you can access all the mentions associated to the group of clusters.



When clicking on "See all mentions of the group", you will open the Listen Tab, automatically filtered to display only the mentions associated with the clusters of the group. However, you can use the "Clusters" filter to display the mentions associated with one cluster only, or to hide some clusters.



You can of course also select a cluster that does not belong to any group. When clicking on it, you access the exact same information as in a group of clusters.



## Settings

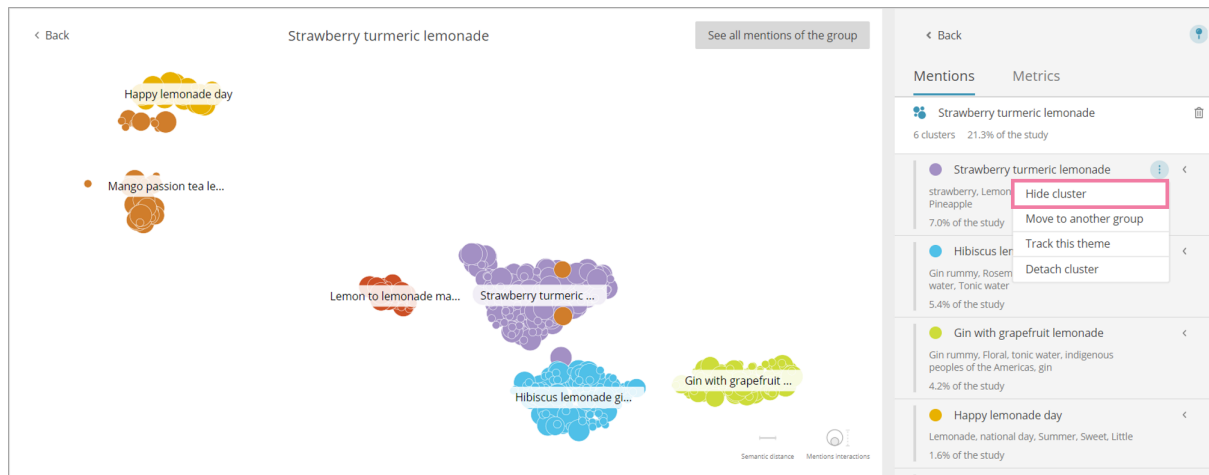
At any time you can decide to change the parameters and name of your study by clicking on Settings.

After choosing your new parameters, click on apply on your study will be launched again and will need the same time (7 to 30 minutes) to run.

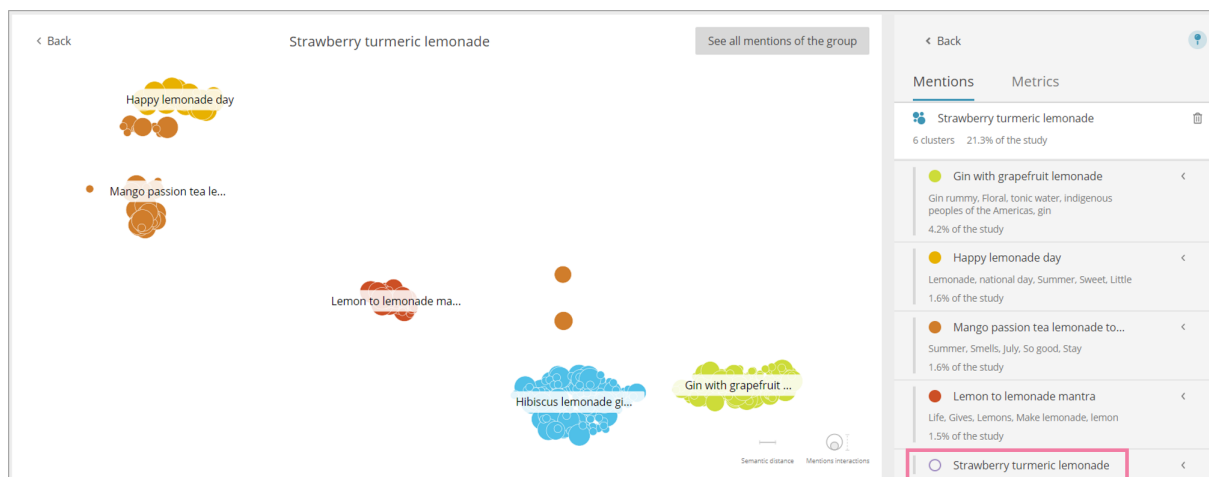
## Manage the clusters

### Hide a cluster

If a cluster does not seem relevant, you can hide it.



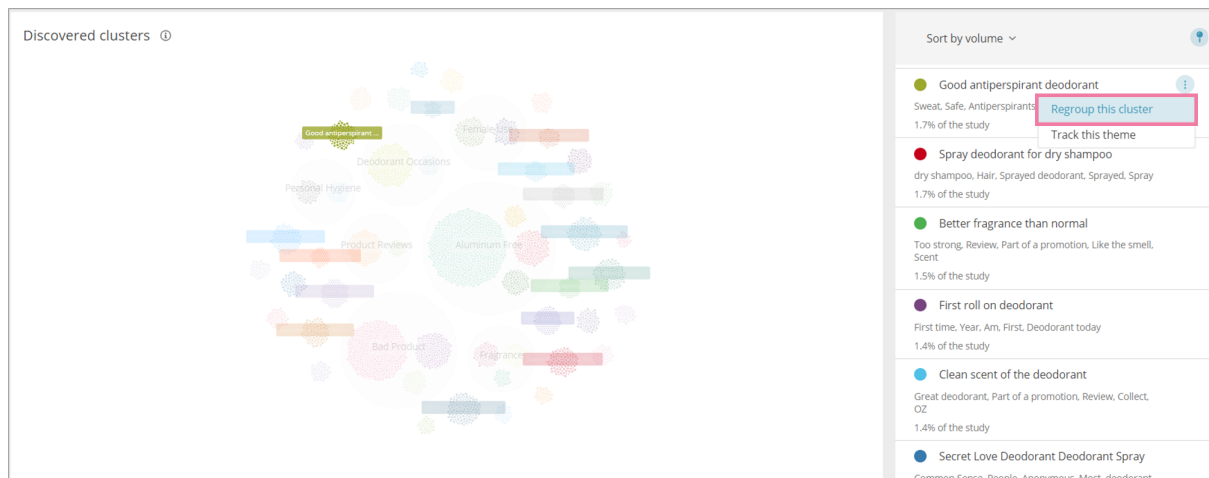
It will no longer appear in the visualization and will be hidden in the stream of clusters.



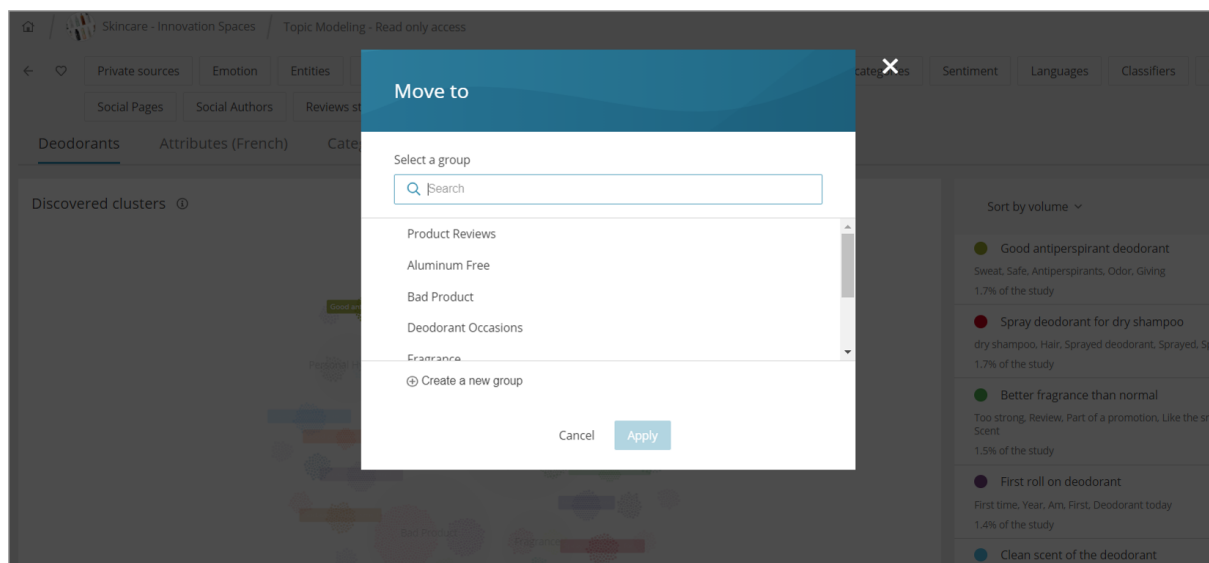
### Group and detach clusters

It can happen that you find in your Topic Modeling study two or more clusters that are similar and that have not been automatically grouped by the algorithm (we have to admit that the machine will never be able to deliver perfect results).

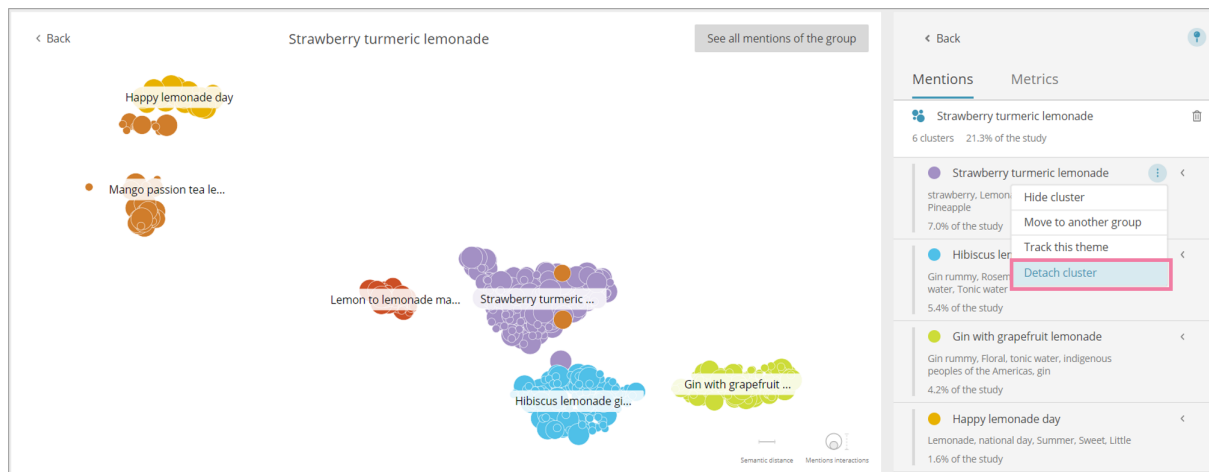
In this case, you can manually group clusters by clicking on "Regroup this cluster".



You can then choose to group this cluster with an already existing group or to create a new one.



On the opposite, if some clusters have been gathered in a group but are not similar enough to you, you can choose to detach them, and let them as stand-alone clusters or to create new groups.

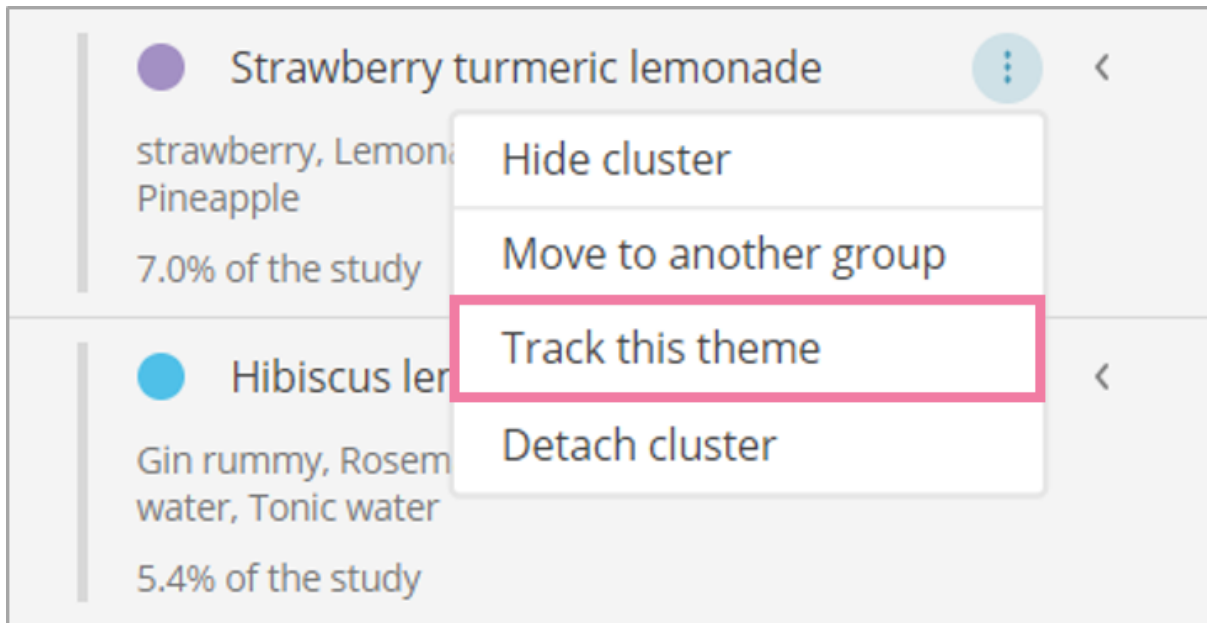


## Convert a cluster theme into a 'Topic' or track the cluster theme as a 'Classifier'

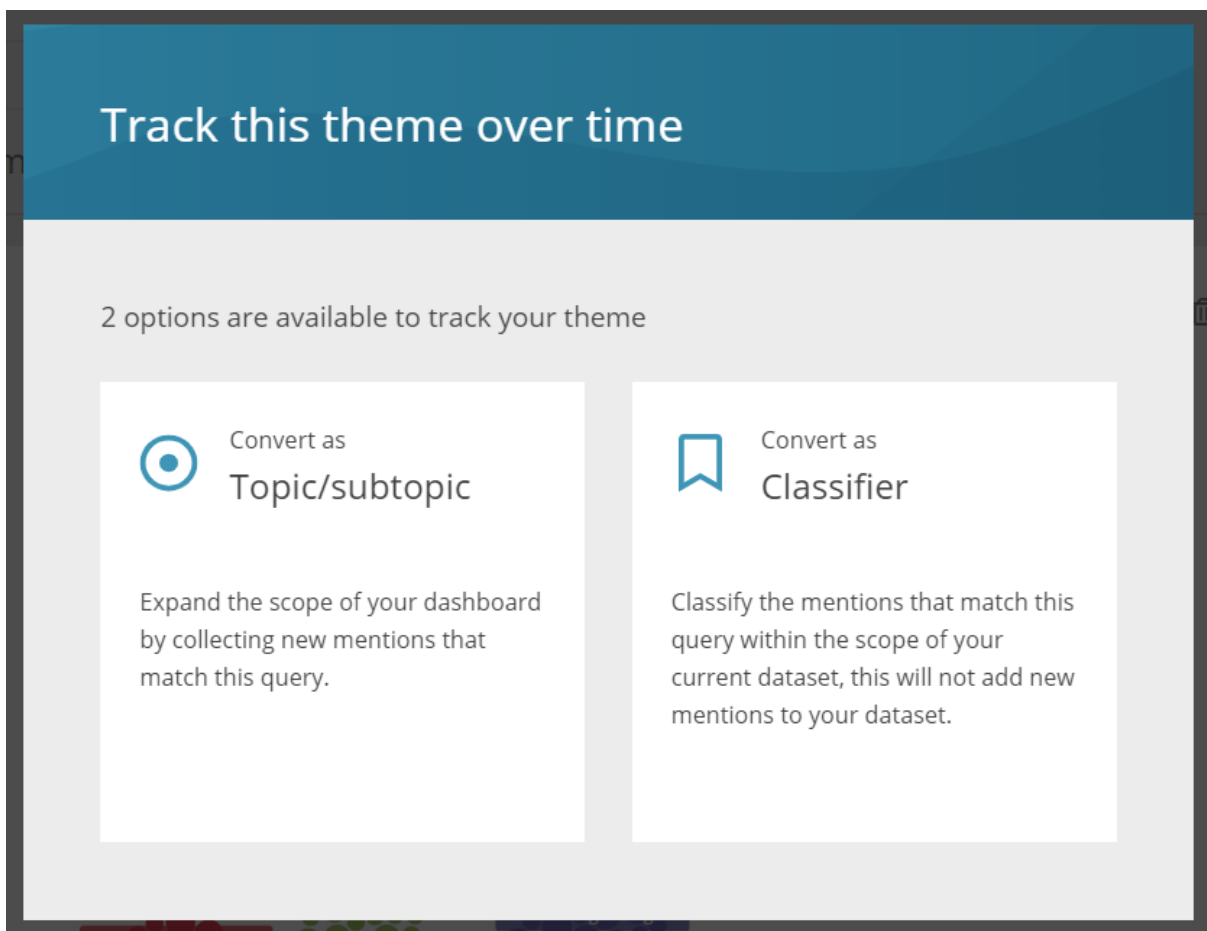
Differently, if you find a cluster valuable, you can either convert it into a 'Topic' or a 'Classifier' for further tracking and incorporation into your reporting. This new feature is powered by AI, as Synthesio will convert the cluster theme into a query-based filter, which can either be used as a Topic or Classifier. We will show you what it looks like below. Let's first cover the steps.

### Step 1 - Click 'Track this theme'

From the right side of the Topic Modeling Study view, click the hamburger icon to reveal four options.



**Step 2 - Either select to convert as a Topic or Classifier**



**What are the main differences between both options?**

**Classifiers:** A classifier is an analytical layer that aims at segmenting and analyzing conversations under the lens of specific business questions (e.g. Brand Intelligence, Product Innovation...). The objective is to provide better, richer, and thinner consumer insights.

In the case of converting a theme into a classifier, it means that the algorithm will auto-generate a query in order to match and map all mentions within the scope of the existing topics:

- You are NOT expanding the scope of the dashboard. Instead, a Classifier buckets the mentions that are currently defined within your existing Topic/subtopic structure.
- It classifies all the mentions matching the query (both existing and new incoming mentions) in the classifier
- Thus, it enables you to analyze your data set under the lens of these classifiers.
- You can assess, track, monitor and report about your classifiers.

**Topics/Subtopics:** The other option is to add an additional topic and/or subtopic to your dashboard.

It means that the auto-generated query will therefore bring new mentions within the dashboard:

- You are expanding the scope of your dashboard. In other words, you will start ingesting new mentions into your dashboard that meet the criteria found within the newly created Topic/subtopic.
- It enables tracking new mentions coming from consumer-generated themes.
- You can then incorporate these new topics into your analysis, or potentially replace existing topics
- One pro-tip we can give you is to re-launch Topic Modeling studies on the newly created topics to create classifiers

- We recommend that you act carefully with the auto-generated queries when creating new topics:
  - They are default designed for classifiers in order to match mentions within the same dashboard scope
  - Indeed, they often use the OR Boolean operator, which makes a lot of sense when you stick to your initial scope
  - However, when you open the door to bringing new conversations into your dashboard, the OR Boolean operator is likely to bring too many new mentions, and especially mentions that are not relevant to your business question.

This is why we recommend that you get in touch with your AM/CSM as modifying the scope of a dashboard can affect your results.

When converted as a Topic/subtopic, you can view the newly created topic within the Setup section of your dashboard. You will need to give it a name. Likewise, when creating as a Classifier, you'll need to give the new classifier a name. In each scenario, you will see the AI-powered query that Synthesio generates. Let's take an example when creating a new classifier.

## Save as a new classifier

Enter a new classifier name

Good taste in cheap scotch

50 char. max

Generated query (preview)

How it works ⓘ

```
malt OR easy OR smooth OR #singlemaltwhisky OR macallan OR full  
OR neat OR #beer
```

☐ Use same filters as the study

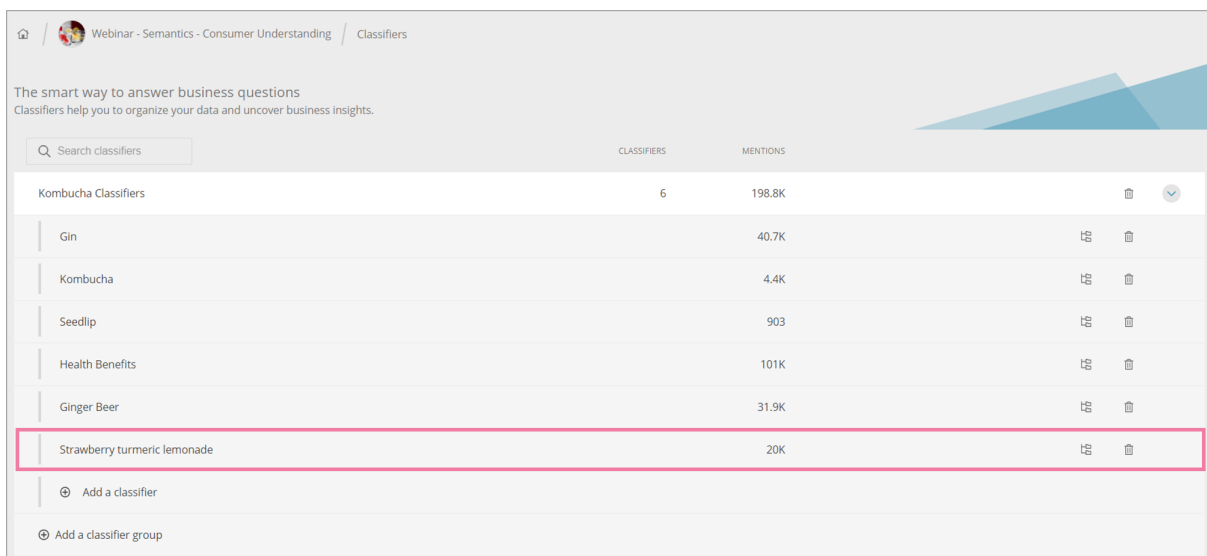
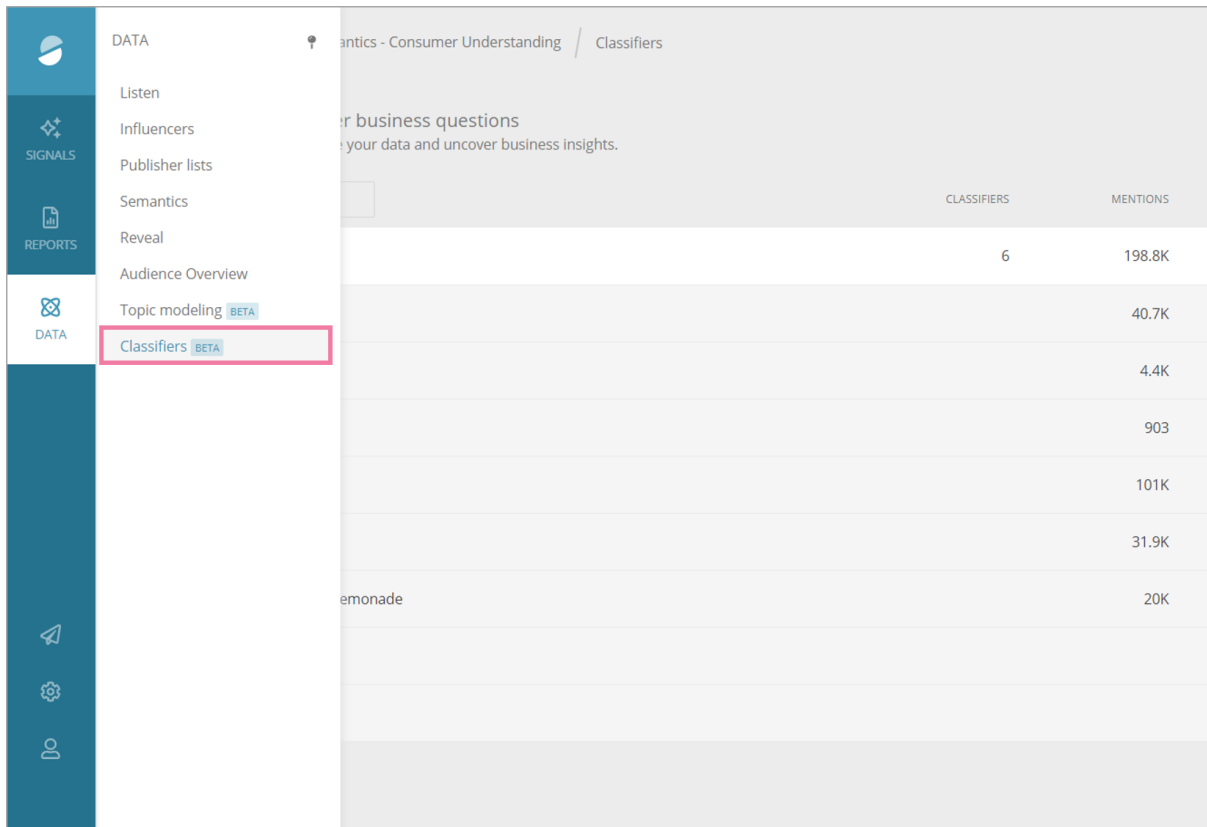
Cancel

Create and go to edit

Create

### Step 3 - Edit and enrich your classifier

When you create a classifier, it is automatically saved in your Classifiers page that you can access in the Data section.



Your classifiers default aggregate the mentions matching the query auto-generated by the algorithm. However, you may want to adapt and/or enrich this query.

On this page, you can:

- Edit the query
- Apply filters to the query

- Choose to automatically update the classifier as new mentions enter your dashboard

Should you want to get all the details about how to use a classifier, [please read this article](#) that explains it in more detail.

#### **Step 4 - Incorporate the newly created Topic or Classifier into your reporting**

So far, all the steps described above are part of the exploration phase, aiming at identifying the most relevant cluster themes that you should track over time. The last step is to transfer your findings to other parts of your dashboard to measure their impact.

Here are the main areas of the product where you should use your newly created topics and classifiers.

#### **Uncover key attributes with the Semantic Explorer:**

When running your classifier or topic through the Semantic Explorer, you can find the relationships between entities (like products) and expressions within the newly created classifier or topic. You can better understand what people are saying about particular themes. In doing so, you find relevant insights related to key product attributes and buying considerations.

#### **Launching Signals on your classifier or topic:**

- Signals is an AI assistant, like a notification center.
- It allows you to aggregate the most significant statistical events from your data set.
- Thanks to launching Signals on your topic or classifier that directly emanates from your Topic Modeling findings, it gives you indications about the shifts and the correlations between words.
- It's important to discover a trend, but it's even more important to understand what makes it statistically meaningful.

#### **Building Reports pages on your topic or classifier:**

- BI-style reporting, this is where you can really measure and report the importance of a trend. It makes the findings accessible across your organization.
- Reports is really the place that enables you to enhance decision-making, this is where you can measure and assess the relevancy of the insights found in Topic Modeling
- For any spike or decrease of mentions, you access Instant Signals that combines statistical and semantic analysis to understand the reasons behind the evolution of a trend.

Should you need more information about Topic Modeling, feel free to reach out to your dedicated Account Manager.