



Sampling Overview

Conducting research on the social web involves huge numbers of data points, and therefore measuring a smaller sample of the conversation online, rather than measuring 100% of the conversation, is often necessary. The following FAQs outline the most common questions around general sampling in our platform.

Does Crimson Hexagon sample?

Yes, we do. Each metric found in your monitor (except for volume) has the potential to show either the total conversation or a random sample of the total conversation, depending on the number of data points available. The one exception to this rule is volume at monitor level, which is always based on total numbers. However, this will not apply to volume for Saved Filters.

When is something sampled in *ForSight*?

If the conversation you are monitoring has over 10,000 posts in a single day, *ForSight* will take a random sample from each data source (in proportion to the prevalence of that data source in the total conversation) equaling 10,000 total posts. For example, if the conversation you are monitoring finds 4,000 posts on January 1 and 16,000 posts on January 2, all 4,000 posts will be analyzed for day 1 and a random sample of 10,000 posts will be analyzed for day 2, totalling 14,000 posts for the entire date range.

Does this mean that posts outside of the 10,000 per day sample are not analyzed?

The individual posts outside of the sample collected each day are not directly analyzed (however they will be included in the volume metric). However, we will extrapolate the data from the sample to represent results from the overall sample critical for highly accurate results.

How is the sample determined when there are over 10,000 posts per day?

The sample is randomly selected from each data source in proportion to the prevalence of that data source in your total volume.

Is there any way to export a larger amount of posts than what is used in the sample?

No, currently this is not possible through either *ForSight* (UI) or our API.



Does the API use the same sample as *ForSight* (UI)?

Yes, the API will pull from *ForSight* (UI) results and have the same sample.

How accurate is sampling compared to measuring the total data set?

Random sampling is a highly accurate method for measuring large universes of data. Sampling is widely used among political pollsters and market research agencies alike. Moreover, these groups tend to use significantly smaller sample frames (between N=150 and N=1,000) for their research and typically have a 95% confidence interval.

Why do we sample this way?

The main reason for this sampling method has to do with the overall functionality of *ForSight*. If we provided *all* content for the visualizations, loading times would be through the roof. We like giving our customers accurate visualizations in seconds (not minutes or hours). Also, monitor run-times would be affected, because our database would have to collect a ton of more data (most of which our users will not see or care about). It is all about efficiency and accuracy, and we are confident with how we are getting both!