

## **Data Collection & Scraping**

### **Frequency & Speed**

The first time you add a new product, it will typically take between five minutes to an hour to fully collect all of the reviews and load them into the dashboard (there are some exceptions to this for a small number of sites we cover with a Socialgist). After the reviews have been collected for the first time, the dashboards and reviews load instantly in future.

In our standard license, we crawl the data once a day (this is configurable at a project level for clients that have brought a mixed cadence model). Specifically, we will collect the data every 24 hours since the collection for that product is finished. This means that if you created different products at different times during the week, the data for those products will refresh at different times.

### **Refreshing Data**

Every 24 hours, we will collect all of the reviews, remove duplicates, and add the new full reviews. This way we can ensure 100% data coverage of publicly available data.

### **Historical Data**

We bring back all the publicly available reviews for the given product URL. You can use the date filter at the top of the dashboard to look at any period of time historically.

### **Data Completeness**

The analysis includes star ratings that also contain reviews. In other words, if a star rating does not have a text review, we do not include it in our analysis. Each product is capped at 20,000 text reviews. Note that it is extremely rare for a product to exceed this limit, and the limit is subject to change if needed.

### **Crawlers**

We use a combination of in-house crawlers and supplemental crawlers from our trusted data partners.

### **Adding sites to be crawled**

In the URL pricing model, all sites we currently cover are included. However, they will need to be activated on your account. Adding additional sites is an option in the product pricing.

Please note that not all sites are crawlable. For example, if the data is behind a login screen, private, or the terms and conditions of the site prohibit it, we will not be able to obtain the reviews data. If you'd like coverage of a site we don't currently cover, please raise this request with your CSM, and we can run a complementary feasibility check to ensure it's possible to compliantly collect data from the site in question. If you'd like them to prioritize development of a crawler for this site, there is a small one-time fee.

### **Amazon API**

We do not use the Amazon API as only very basic product details are available through it. Instead we collect data directly from the web page to get the maximum amount of data and detail.