

Quarantine Bad Responses

Before making decisions based on your survey data it is a good practice to clean your data. This is particularly true if you used panel respondents or provided an incentive to increase your response rate. Using our Data Cleaning Tool you can identify and remove responses with poor data to ensure you are making decisions based on the best data possible.

Before we get started discussing the tool itself, we should start by discussing the data cleaning. Cleaning survey data is a subjective process. This means it almost always requires a human to make final judgments on the quality of each response. We've provided a tool that helps to automate the decisions you make, but we can't do it without you. Every poor data flag will not apply to every survey. If you choose to check off all the poor data flags you will, most likely, not be satisfied with the results.

Instead, for the best results, you should consider your survey and your respondents. Think about what data and respondent behaviors will skew your results and use the tool to look for those poor data flags.

The Data Cleaning tool is not available for surveys created before the release of this feature on October 24, 2014. If you wish to still use this tool for surveys created prior to the release of this tool simply copy the survey and [import the data](#). With the exception of the speeding tool, all other tools will be available to use after importing the data.

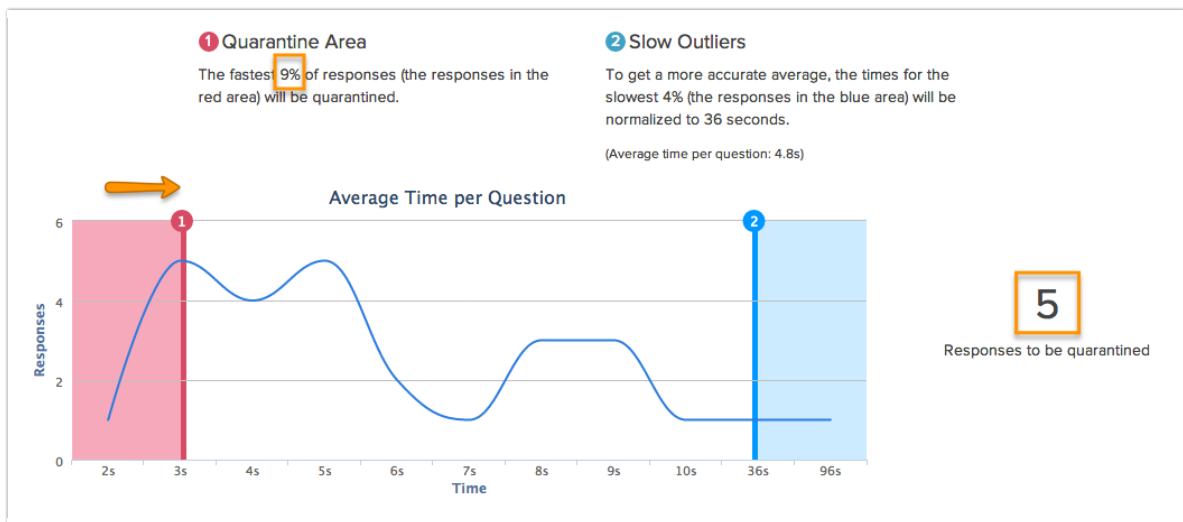
Surveys created after this date will show the option to clean data once you have collected 25 live (non-test) responses. To get started, go to **Results > Quarantine Bad Responses**.

Step 1: Speeding

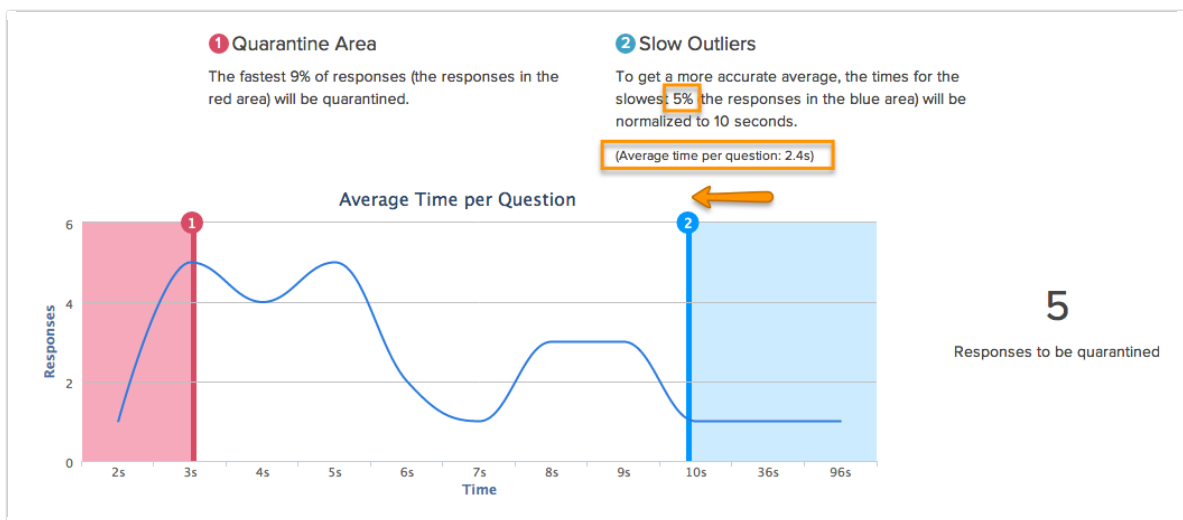
The data cleaning process starts with identifying respondents who sped through your survey. We start with this step because speeding is one of the most serious indicators of poor quality data. Respondents who proceed through your survey faster than others are likely not carefully reading your questions. This generally means that their answers are less accurate.

In the speeding step, the distribution of average response time per question will be charted for you. This is computed for each response by taking the total response time divided by the number of questions that that respondent was shown. The chart will start with a red line, used to eliminate speeders, at the fastest 1% of responses and a blue line, used to normalize slow outliers, at the slowest 10% of responses.

1. To get started, click and drag the red line to quarantine the fastest responses. You will most likely make the decision of the placement of the red line based on the curve of the data, but remember to keep an eye on the percentage and number of responses you are quarantining.



- Next, click and drag the blue line to the left if you wish to normalize the average per question response time. All responses in the blue area will be given the max average per question response time indicated by the blue line. Keep an eye on the percentage of responses you are affecting and the change in the average time per question as you make these adjustments.



- Once you are satisfied with your adjustments to average time per question and the responses that are quarantined as a result, click **Next** to move on to evaluating answer quality.

Step 2: Answer Quality

In the answer quality step of the data cleaning tool, the remaining responses that were not quarantined for speeding can be further evaluated for data quality. There are 8 possible indicators of poor data.

Each poor data quality flag has a default flag weight. If you wish to change the importance of any these poor data flags you can increase the importance by increasing the flag weight or decrease it by assigning a lower flag weight.

Below we discuss each of your poor data quality indicators.

Choose Which Flags to Use

Response quality is evaluated according to the following flags. If a certain flag is more important to your particular survey setup, increase the weighting. [Learn more about how Dirty Data is calculated.](#)

- Straightlining/Patterned Responses (4 instances)** - Responses with all the same answers (straightlining) or answers in a visual pattern (Christmas Tree) in option grids will be flagged.
Flag Weighting:
- Gibberish (23 instances)** - Responses with non-word answers (e.g. "a;oihrg") will be flagged.
Flag Weighting:
- Fake Answers (15 instances)** - Responses that use fake answers (e.g. email@example.com) will be flagged.
Flag Weighting:
- Trap/Red Herring Questions** - Responses that match the following rules will be flagged ([Learn More about setting up Trap Questions](#)).
- Consistency Check** - Responses that match the following rules will be flagged ([Learn More about setting up Consistency Checks](#)).
- All Checkboxes (20 instances)** - Responses that check all checkboxes in a list will be flagged.
Flag Weighting:
- Single Checkbox (37 instances)** - Responses that only check a single checkbox in a checkbox list will be flagged.
Flag Weighting:
- One Word Answer (7 instances)** - Responses that have single word answers for required essay questions will be flagged.

Straightlining/Patterned Responses

Responses in option grids with all the same answers (straightlining) or answers in a pattern may indicate that the data may not represent the respondent's actual opinion. Both straightlining and patterned response behavior are methods used by respondents to complete surveys as quickly as possible. This behavior is usually motivated by incentives.

Responses with instances of this behavior will be flagged by default. The number of instances detected across all responses will be available to the right of the flag label. This flag will even work for grid questions with rows that are randomized as the flag is set as part of the response itself.

If you wish to turn this off, you can do so by unchecking the checkbox.

Questions evaluated: [Radio Button](#) and [Checkbox](#) grid questions with more than 4 rows.

How it works: A response receives a *straightline* flag for each grid question where more than half of the rows answered the same. A response receives a patterned response flag for each grid question where every answer is within one step from the previous row's answer.

Straightlining/Patterned Responses (4 instances) flagged.
Flag Weighting:

Gibberish

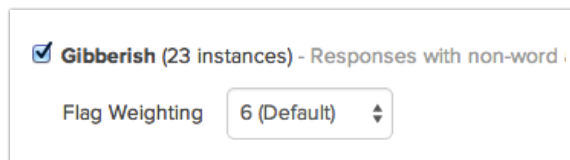
Nonsense answers to open-ended questions may indicate that the respondent was keyboard mashing and was not engaged.

Responses with instances of this behavior will be flagged by default. The number of instances detected across all responses will be available to the right of the flag label.

If you wish to turn this off, you can do so by unchecking the checkbox.

Questions Evaluated: All open-text fields.

How it works: Open-text responses with less than 7 words are evaluated and matched against an English language dictionary. Responses receive a gibberish flag for each open-text field answer with less than 7 words that contains non-words, for example, "lkaghsd." This flag will not work as well with non-English language responses, especially non-roman character sets. No need to worry though about false positives, responses with non-roman character sets will be ignored.



A screenshot of a configuration box for the 'Gibberish' flag. It features a checked checkbox followed by the text 'Gibberish (23 instances) - Responses with non-word'. Below this, there is a 'Flag Weighting' label and a dropdown menu showing '6 (Default)' with up and down arrows.

Fake Answers

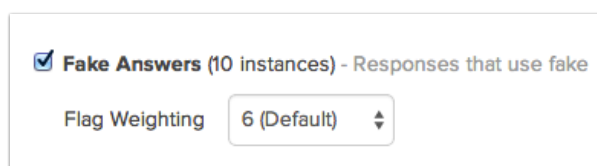
Fake answers (e.g. email@example.com, lorem ipsum, test, etc.) to open-ended questions may indicate respondent did not provide truthful answers elsewhere in the survey.

Responses with instances of this behavior will be flagged by default. The number of instances detected across all responses will be available to the right of the flag label.

If you wish to turn this off, you can do so by unchecking the checkbox.

Questions Evaluated: All open-text fields.

How it works: Open-text responses with less than 7 words are evaluated and matched against typical dummy answers, such as, lorem ipsum, test, etc. Responses receive a fake answer flag for each open-text field answer with less than 7 words that contains dummy words.



A screenshot of a configuration box for the 'Fake Answers' flag. It features a checked checkbox followed by the text 'Fake Answers (10 instances) - Responses that use fake'. Below this, there is a 'Flag Weighting' label and a dropdown menu showing '6 (Default)' with up and down arrows.

Trap/Red Herring Questions

Trap or red herring questions can be added to surveys and used to filter out responses from disengaged respondents.

To use this data cleaning flag, check the checkbox. Next, use the logic builder to set up the rules for responses you wish to flag.

Say, for example, we added a question to a survey that asks respondents to agree or disagree with the statement "The sun rotates around the earth."

3. The sun rotates around the Earth:

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Then we can set up a logic rule to flag any responses to this red herring question that are anything other than Strongly Disagree.

Trap/Red Herring Questions - Responses that match the following rules will be flagged ([Learn More about setting up Trap Questions](#)).

Flag responses matching these rules (toggle):

[Remove All Logic](#)

Q 44. The sun rotates around the earth: Strongly Disagree

Disagree

Neutral

Agree

Strongly Agree

[+ Add Condition](#)

Flag Weighting

Consistency Check

Similar to trap/red herring questions, consistency checks can be added to surveys and used to filter out responses from disengaged respondents. While red herring questions can be cognitively taxing and might foster distrust, consistency checks are less conspicuous to survey respondents.

To use this data cleaning flag check the checkbox. Next, use the logic builder to set up the rules for responses you wish to flag.

Say, for example, we asked an open-text age question at the beginning of the survey.

1. Please enter your age:

And, later in the survey, we asked an age question with age ranges as answer options.

14. What is your age?

We can set up logic rules to flag responses in which the answers to these two questions were inconsistent.

Consistency Check - Responses that match the following rules will be flagged ([Learn More about setting up Consistency Checks](#)).

Flag responses matching these rules:

Group 1

Q 1. Please enter your age:	< (less than)	18
and		
Q 47. What is your age?	is one of the following	<input type="checkbox"/> under 18 <input checked="" type="checkbox"/> 18-24 <input checked="" type="checkbox"/> 25-34 <input checked="" type="checkbox"/> 35-54 <input checked="" type="checkbox"/> 55+

[+ Add Condition](#)

or

All Checkboxes

Another possible indicator of poor data quality is checkbox questions with all answers selected. For example, if your survey is about the NFL you might include a screener question to ensure that your respondents are NFL fans. Respondents, particularly panel respondents and respondents who are motivated by an incentive, will often select all options in a screener question so that they qualify to continue the survey.

5. Which professional sports leagues do you regularly follow?

- NFL
- NBA
- MLB
- NHL
- MLS
- WNBA

Responses with all checkboxes checked are may also indicate that web-browser form-fill tools have been used. This is especially true when the answer list includes *Not applicable*, *Other*, or *None of the above*.

To use this data cleaning flag check the checkbox. The number of instances detected across all responses will be available to the right of the flag label.

Questions Evaluated: [Checkbox questions](#)

All Checkboxes (20 instances) - Responses that check all checkboxes |
Flag Weighting

Single Checkbox

Checkbox questions with a single option selected in specific scenarios may indicate poor data quality. Professional survey panelists often understand that some surveys will disqualify them if they don't check anything on a screening question. To circumvent this, they will just mark one answer to ensure they can proceed.

If you decide to use this flag it is best to be used in conjunction with other flags as a single checkbox response alone is not usually enough to quarantine a response.

To use this data cleaning flag check the checkbox. The number of instances detected across all responses will be available to the right of the flag label.

Questions Evaluated: [Checkbox questions](#)

Single Checkbox (36 instances) - Responses that only check

Flag Weighting

One Word Answer

Single-word answers to open-ended questions may indicate either that a web-based form-filler has been used to record a response or that your respondent was not engaged.

To use this data cleaning flag, check the checkbox. The number of instances detected across all responses will be available to the right of the flag label.

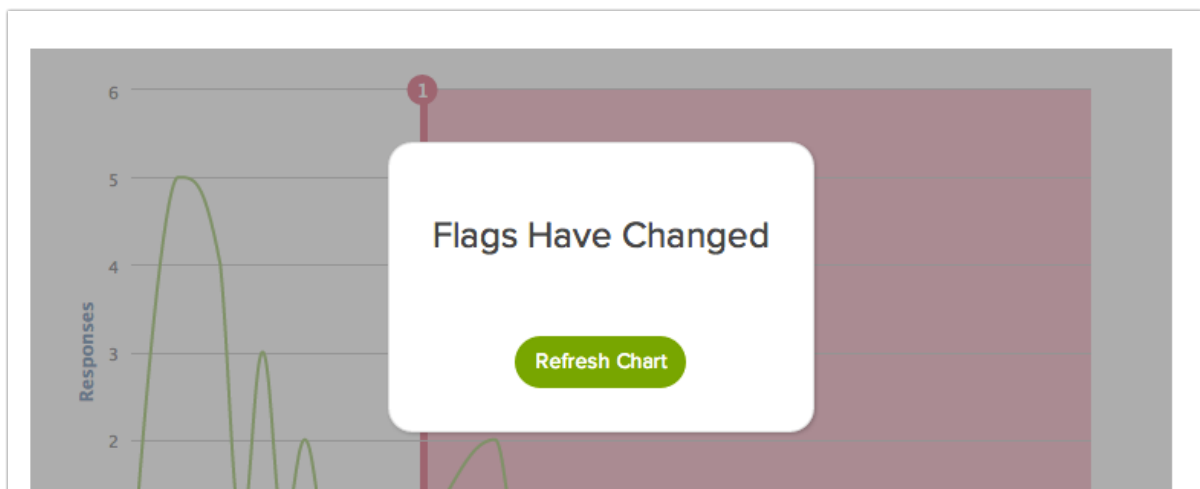
Questions Evaluated: Required [Essay/Long Answer questions](#)

One Word Answer (6 instances) - Responses that have single word

Flag Weighting

Choosing Quarantine Threshold

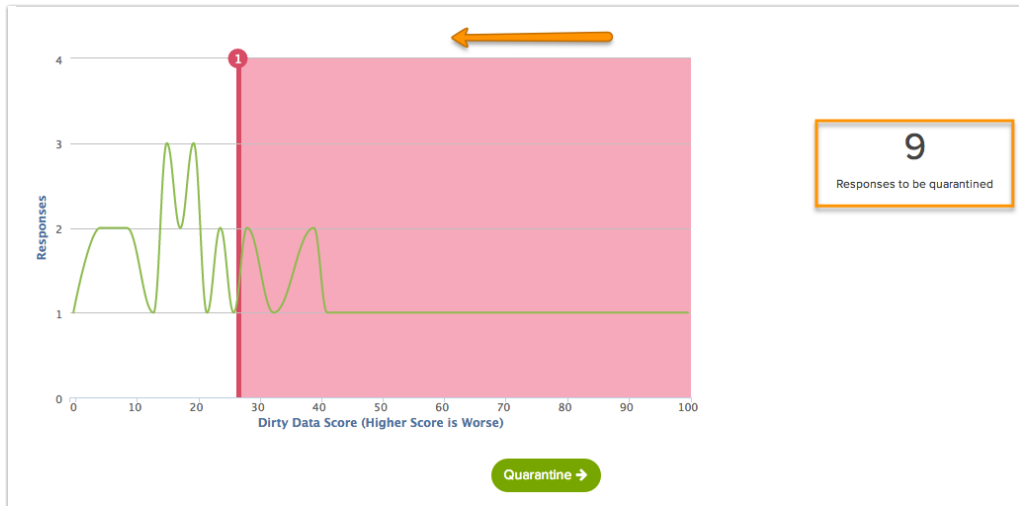
Once you are finished choosing the flags you wish to use and customizing their weights, scroll down to choose your quarantine threshold. In most cases, you will have made changes to the data flags that require the chart to be refreshed.



The Dirty Data Scores for your data set will be charted on the x-axis with the number of responses

charted on the y-axis.

1. Click and drag the red line to quarantine responses with high Dirty Data Scores. Responses in the red area will be quarantined and removed from your results.
2. When you are satisfied with the responses to be quarantined, click the **Quarantine** button.



How are the dirty data scores calculated?

A Dirty Data Score is calculated for each response using the flags that you enabled. Each flag is multiplied times the weights that you selected for each flag type. All Dirty Data Scores are normalized on a curve so that the response with the highest score receives a 100 and the response with the lowest score receives a 0.

For example:

- + 1 straightlining instance x flag weight of 8
- + 3 gibberish instances x flag weight of 6
- + 2 fake answers x flag weight of 6
- + 5 all checkboxes x flag weight of 4
- + 6 one word answers x flag weight of 4

= Dirty Data Score 82 (before normalization)

Step 3: Results

Your data cleaning results will show how many responses have been quarantined.

1 Speeding
2 Answer Quality
3 Results

Your responses have been cleaned! 3 responses have been quarantined.

- All responses in the limits you have set have been tagged as "Quarantined"
- By default, all quarantined responses will be filtered out of reports.
- You can individually add and remove responses to the quarantine list through the Data Quality tab on each response.

[View Quarantined List](#)
[Export Quarantined Responses](#)
[Restore quarantined responses](#)

Start Over

From here you can choose to:

View Quarantined List

This link will take you to the **Individual Responses** page where the quarantined responses will be flagged as **Quarantined**.

Responses: 25

Partials - 0 / Disqualified - 0 / Quarantined - 2 27 New!

<input type="checkbox"/>	ID	Status	Date Submitted	Name
<input type="checkbox"/>	27	New Quarantined	May 18, 2015 9:37 AM	fdasklj
<input type="checkbox"/>	26	New Quarantined	May 18, 2015 9:37 AM	Jane Doe
<input type="checkbox"/>	25	New	May 18, 2015 9:36 AM	Jon Doe

Export Quarantined Responses

This link will download your quarantined responses to an Excel spreadsheet.

R	S	T	U	V	W	X
Fake Answers Instances	Gibberish Instances	All Checkboxes Instances	One Checkbo	Name	Email Address	Gender
1				Jane Doe	test@test.co	male
	3			fdasklj	fgdsafsjd;@s	female

Each Data Cleaning category is displayed as a column in the export with a count of instances for each response. For example, in the screenshot above the response in row one has one instance of fake answers whereas the response in row two has three instances of gibberish answers.

Restore Quarantined Responses

This link will take the quarantine flag off of all responses that were quarantined during the data cleaning process. You can go through the quarantine process once more and adjust your flags if needed by clicking on the **Start Over** button on the data cleaning results page.

If you wish to add and remove individual responses from quarantine you can do so on the **Data**

Quality tab of each response under **Results > Individual Responses**.

Response #29 Customize

DATA DETAILS ACTION LOG **DATA QUALITY** NEED HELP?

Select an Action ▼

Possible Straightliner/Patterned Responses
The answer pattern and the response speed suggest that the respondent may have answered (Christmas Tree) through the questions without reading.

46. Indicate your level of satisfaction with the ice cream brands that you like.

	Very Dissatisfied	Dissatisfied	Neutral	Satisfied	Very Satisfied
Dairy Cream			X		
Tasty Treat			X		
Wholesome			X		

If you would like to adjust your data cleaning results, including how many responses are quarantined and the flags that determine dirty data, you can click the option to **Start Over**.

1 Speeding 2 Answer Quality 3 Results

Your responses have been cleaned! 32 responses have been quarantined.

- All responses in the limits you have set have been tagged as "Quarantined"
- By default, all quarantined responses will be filtered out of reports.
- You can individually add and remove responses to the quarantine list through the Data Quality tab on each response.

[View Quarantined List](#) [Export Quarantined Responses](#) [Restore quarantined responses](#)

Start Over

Reports and Exports of Cleaned Data

Once you have cleaned your data, responses flagged as quarantined will be automatically filtered out of your reports and exports. If you wish to include* these, go to the Filter tab of your report or export and uncheck the option to **Remove Responses Flagged as Quarantined**.

*The option to include quarantined responses is available in Legacy Summary Reports only.

FAQ

The option for Data Cleaning Tool is not available on my survey. What gives?

The data cleaning tool is available only for surveys created after the release of this feature on October 24, 2014. For surveys created after this date the option to clean data will only be available once you have collected 25 responses.

The Data Cleaning Tool is not compatible with the [data encryption feature](#) on the Survey Settings

page.

What about randomization? Can the Pattern Flag still detect patterns in grid questions where rows are randomized?

Yes! Because the Data Cleaning Tool sets these flags in realtime as part of the response it can still detect patterns in grid questions with randomized rows. If you check out the Data Quality tab of an individual response you can verify that the response has been flagged for a pattern based on the order the question was displayed in the given response.

Possible Straightliner/Patterned Responses

The answer pattern suggests that the respondent may have straightlined or answered in a visual pattern (Christmas Tree) through the questions without reading.

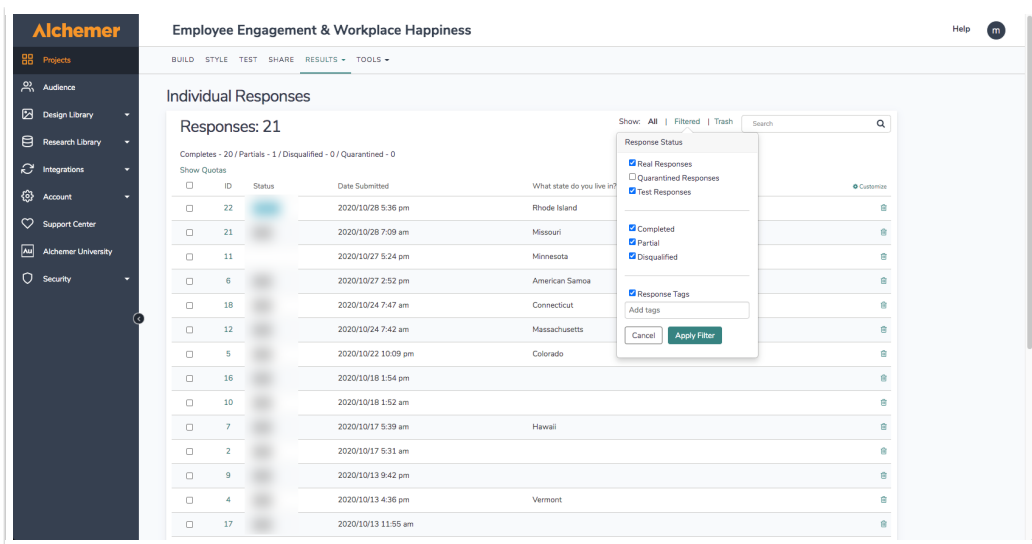
14. Indicate your level of satisfaction with the ice cream brands that you purchase:

	Very Dissatisfied	Dissatisfied	Neutral	Satisfied	Very Satisfied
Dairy Cream					X
Tasty Treat					X
Wholesome					X
Sugar Cow					X
Ice Dream					X
Jill's					X
Mike and Sammy's					X
Cold Cream					X
Homemade					X

Can I remove the Quarantined flag from Individual Responses?

Yes! If you need to remove the *Quarantined* flag from a response, you can do so via **Results > Individual Responses**.

1. First, you will need to make sure that you display the Quarantined Responses. Click on the **Filtered** link next to the search bar at the top of your response list.
2. Next, check the box associated with Quarantined Responses and click the Apply button. Your quarantined responses will now show in your response list.



- Next, access the response in question by clicking on that response in the list. This will open the response details.
- Finally, access the **Data Quality** tab and use the **Select an Action** menu to restore your response.

Other Tools for Ensuring Good Data Quality

- Reduce respondent burden and fatigue by creating branches in lengthier surveys. Learn to set up [Percent Branches](#).
- Use logic to reduce respondent burden and fatigue by only showing relevant questions. Learn to set up [Question Logic](#) or [Page Logic](#).
- Provide survey respondents with the ability to opt out of a question when it does not apply to them to eliminate bias in your data. Learn to set up [Answer Option Special Settings](#).
- On longer surveys, you may want to allow respondents to come back later to finish their response. Learn how to set up [Save and Continue](#).
- Require respondents to select a minimum number of options or set a cap on the maximum selectable options for a question. Learn how to set the [Min and/or Max Selectable Options](#).

Related Articles