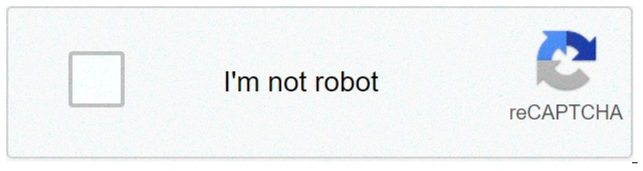

Continue

Best way to analyze qualitative data

Qualitative data analysis involves the identification, examination, and interpretation of patterns and themes in data and determines how these patterns and themes help answer the research questions at hand. Download a print version (PDF, 80 kB) of the guide to qualitative data or view the online version below. Qualitative analysis focuses on cases (rather than variables). A case could be an individual person, a whole class of students or an instance of a particular phenomenon (e.g. bullying). Comparative analysis between and across cases can be a powerful way of identifying patterns in the data. What patterns/common themes emerge around specific items in the data? How do these patterns (or lack thereof) help to shed light on the broader study question(s)? Are there any deviations from these patterns? If, yes, what factors could explain these atypical responses? What interesting stories emerge from the data? How can these stories help to shed light on the broader study question? Do any of the patterns/emergent themes suggest that additional data needs to be collected? Do the patterns that emerge support the findings of other corresponding qualitative and quantitative analyses that have been conducted? Documentation of the data and the process of data collection Organisation/categorisation of the data into concepts/themes Connection of the data to show how one concept/theme may influence another Testing theories, by evaluating alternative explanations and searching for negative cases. The final report of any qualitative analysis may include a number of formats such as text, maps, charts, images and/or sound. More information on analysing qualitative data, see this overview of qualitative data. Authored by Alyona Medelyan, PhD – Natural Language Processing & Machine Learning How many hours have you spent sitting in front of Excel spreadsheets trying to find new insights from customer feedback? You know that asking open-ended survey questions gives you more actionable insights than asking your customers for just a numerical Net Promoter Score (NPS). But when you ask open-ended, free-text questions, you end up with hundreds (or even thousands) of free-text responses. How can you turn all of that text into quantifiable, applicable information about your customers' needs and expectations? By coding qualitative data. Keep reading to learn: What coding qualitative data means (and why it's important)Different methods of coding qualitative dataHow to manually code qualitative data to find significant themes in your data What is coding in qualitative research? Coding is the process of labeling and organizing your qualitative data to identify different themes and the relationships between them. When coding customer feedback, you assign labels to words or phrases that represent important (and recurring) themes in each response. These labels can be words, phrases, or numbers; we recommend using words or short phrases, since they're easier to remember, skim, and organize. Coding qualitative research to find common themes and concepts is part of thematic analysis, which is part of qualitative data analysis. Thematic analysis extracts themes from text by analyzing the word and sentence structure. What is qualitative data analysis? Qualitative data analysis is the process of examining and interpreting qualitative data to understand what it represents. Qualitative data is defined as any non-numerical and unstructured data; when looking at customer feedback, qualitative data usually refers to any verbatim or text-based feedback such as reviews, open-ended responses in surveys, complaints, chat messages, customer interviews, case notes or social media posts For example, NPS metric can be strictly quantitative, but when you ask customers why they gave you a rating a score, you will need qualitative data analysis methods in place to understand the comments that customers leave alongside numerical responses. Types of qualitative data analysis Content analysis: This is the most common example of qualitative data analysis. It refers to the categorization, tagging and thematic analysis of qualitative data. This can include combining the results of the analysis with behavioural data for deeper insights.Narrative analysis: Some qualitative data, such as interviews or field notes may contain a story. For example, the process of choosing a product, using it, evaluating its quality and decision to buy or not buy this product next time. Narrative analysis helps understand the underlying events and their effect on the overall outcome. Discourse analysis: This refers to analysis of what people say in social and cultural context. It's particularly useful when your focus is on building or strengthening a brand.Framework analysis: When performing qualitative data analysis, it is useful to have a framework. A code frame (a hierarchical set of themes used in coding qualitative data) is an example of such framework. Grounded theory: This method of analysis starts by formulating a theory around a single data case. Therefore the theory is "grounded' in actual data. Then additional cases can be examined to see if they are relevant and can add to the original theory. Qualitative data analysis software Advances in natural language processing & machine learning have made it possible to automate the analysis of qualitative data, in particular content and framework analysis While manual human analysis is still popular due to its perceived high accuracy, automating the analysis is quickly becoming the preferred choice. Unlike manual analysis, which is prone to bias and doesn't scale to the amount of qualitative data that is generated today, automating analysis is not only more consistent and therefore can be more accurate, but can also save a ton of time, and therefore money. The most commonly used software for automated qualitative data analysis is text analytics software such as Thematic. Qualitative data analysis example: Thematic categorizes qualitative data into themes Why is it important to code qualitative data? Coding qualitative data makes it easier to interpret customer feedback. Assigning codes to words and phrases in each response helps capture what the response is about which, in turn, helps you better analyze and summarize the results of the entire survey. Researchers use coding and other qualitative data analysis processes to help them make data-driven decisions based on customer feedback. When you use coding to analyze your customer feedback, you can quantify the common themes in customer language. This makes it easier to accurately interpret and analyze customer satisfaction. Automated vs. Manual coding of qualitative data Methods of coding qualitative data fall into two categories: automated coding and manual coding. You can automate the coding of your qualitative data with thematic analysis software. Thematic analysis and qualitative data analysis software use machine learning, artificial intelligence (AI), and natural language processing (NLP) to code your qualitative data and break text up into themes. Thematic analysis software is autonomous, which means… You don't need to set up themes or categories in advance.You don't need to train the algorithm — it learns on its own.You can easily capture the "unknown unknowns" to identify themes you may not have spotted on your own. …all of which will save you time (and lots of unnecessary headaches) when analyzing your customer feedback. Recently, thematic analysis software has been categorised as Unified Data Analytics. What is thematic coding? Thematic coding, also called thematic analysis, is a type of qualitative data analysis that finds themes in text by analyzing the meaning of words and sentence structure. When you use thematic coding to analyze customer feedback for example, you can learn which themes are most frequent in feedback. This helps you understand what drives customer satisfaction in an accurate, actionable way. To learn more about how thematic analysis software helps you automate the data coding process, check out this article. How to manually code qualitative data For the rest of this post, we'll focus on manual coding. Different researchers have different processes, but manual coding usually looks something like this: Choose whether you'll use deductive or inductive coding.Read through your data to get a sense of what it looks like. Assign your first set of codes.Go through your data line-by-line to code as much as possible. Your codes should become more detailed at this step. Categorize your codes and figure out how they fit into your coding frame.Identify which themes come up the most — and act on them. Let's break it down a little further… Deductive coding vs. inductive coding Before you start qualitative data coding, you need to decide which codes you'll use. What is Deductive Coding? Deductive coding means you start with a predefined set of codes, then assign those codes to the new qualitative data. These codes might come from previous research, or you might already know what themes you're interested in analyzing. Deductive coding is also called concept-driven coding. For example, let's say you're conducting a survey on customer experience. You want to understand the problems that arise from long call wait times, so you choose to make "wait time" one of your codes before you start looking at the data. The deductive approach can save time and help guarantee that your areas of interest are coded. But you also need to be careful of bias; when you start with predefined codes, you have a bias as to what the answers will be. Make sure you don't miss other important themes by focusing too hard on proving your own hypothesis. What is Inductive Coding? Inductive coding, also called open coding, starts from scratch and creates codes based on the qualitative data itself. You don't have a set codebook; all codes arise directly from the survey responses. Here's how inductive coding works: Break your qualitative dataset into smaller samples.Read a sample of the data.Create codes that will cover the sample.Reread the sample and apply the codes.Read a new sample of data, applying the codes you created for the first sample.Note where codes don't match or where you need additional codes.Create new codes based on the second sample.Go back and recode all responses again.Repeat from step 5 until you've coded all of your data. If you add a new code, split an existing code into two, or change the description of a code, make sure to review how this change will affect the coding of all responses. Otherwise, the same responses at different points in the survey could end up with different codes. Sounds like a lot of work, right? Inductive coding is an iterative process, which means it takes longer and is more thorough than deductive coding. But it also gives you a more complete, unbiased look at the themes throughout your data. Categorize your codes with coding frames Once you create your codes, you need to put them into a coding frame. A coding frame represents the organizational structure of the themes in your research. There are two types of coding frames: flat and hierarchical. Flat Coding Frame A flat coding frame assigns the same level of specificity and importance to each code. While this might feel like an easier and faster method for manual coding, it can be difficult to organize and navigate the themes and concepts as you create more and more codes. It also makes it hard to figure out which themes are most important, which can slow down decision making. Hierarchical Coding Frame Hierarchical frames help you organize codes based on how they relate to one another. For example, you can organize the codes based on your customers' feelings on a certain topic: Hierarchical Coding Frame example In this example: The top-level code describes the topic (customer service)The mid-level code specifies whether the sentiment is positive or negative The third level details the attribute or specific theme associated with the topic Hierarchical framing supports a larger code frame and lets you organize codes based on organizational structure. It also allows for different levels of granularity in your coding. Whether your code frames are hierarchical or flat, your code frames should be flexible. Manually analyzing survey data takes a lot of time and effort; make sure you can use your results in different contexts. For example, if your survey asks customers about customer service, you might only use codes that capture answers about customer service. Then you realize that the same survey responses have a lot of comments about your company's products. To learn more about what people say about your products, you may have to code all of the responses from scratch! A flexible coding frame covers different topics and insights, which lets you reuse the results later on. Tips for coding qualitative data Now that you know the basics of coding your qualitative data, here are some tips on making the most of your qualitative research. Use a codebook to keep track of your codes As you code more and more data, it can be hard to remember all of your codes off the top of your head. Tracking your codes in a codebook helps keep you organized throughout the data analysis process. Your codebook can be as simple as an Excel spreadsheet or word processor document. As you code new data, add new codes to your codebook and reorganize categories and themes as needed. Make sure to track: The label used for each codeA description of the concept or theme the code refers toWho originally coded itThe date that it was originally coded or updatedAny notes on how the code relates to other codes in your analysis Create high-quality codes Your codes should do these 4 things: Cover as many survey responses as possible. The code should be generic enough to apply to multiple comments, but specific enough to be useful in your analysis. For example, "Product" is a broad code that will cover a variety of responses — but it's also pretty vague. What about the product? On the other hand, "Product stops working after using it for 3 hours" is very specific and probably won't apply to many responses. "Poor product quality" or "short product lifespan" might be a happy medium.Avoid commonalities. Having similar codes is okay as long as they serve different purposes. "Customer service" and "Product" are different enough from one another, while "Customer service" and "Customer support" may have subtle differences but should likely be combined into one code.Capture the positive and the negative. Try to create codes that contrast with each other to track both the positive and negative elements of a topic separately. For example, "Useful product features" and "Unnecessary product features" would be two different codes to capture two different themes.Reduce data — to a point. Let's look at the two extremes: There are as many codes as there are responses, or each code applies to every single response. In both cases, the coding exercise is pointless; you don't learn anything new about your data or your customers. To make your analysis as useful as possible, try to find a balance between having too many and too few codes. Group responses based on themes, not wording Make sure to group responses with the same themes under the same code, even if they don't use the same exact wording. For example, a code such as "cleanliness" could cover responses including words and phrases like: CleanTidyDirtyDustyLooked like a dumpCould eat off the floor Having only a few codes and hierarchical framing makes it easier to group different words and phrases under one code. If you have too many codes, especially in a flat frame, your results can become ambiguous and themes can overlap. Manual coding also requires the coder to remember or be able to find all of the relevant codes; the more codes you have, the harder it is to find the ones you need, no matter how organized your codebook is. Make accuracy a priority Manually coding qualitative data means that the coder's cognitive biases can influence the coding process. For each study, make sure you have coding guidelines and training in place to keep coding reliable, consistent, and accurate. One thing to watch out for is definitional drift, which occurs when the data at the beginning of the data set is coded differently than the material coded later. Check for definitional drift across the entire dataset and keep notes with descriptions of how the codes vary across the results. If you have multiple coders working on one team, have them check one another's coding to help eliminate cognitive biases. Conclusion: 6 main takeaways for coding qualitative data Here are 6 final takeaways for manually coding your qualitative data: Coding is the process of labeling and organizing your qualitative data to identify themes. After you code your qualitative data, you can analyze it just like numerical data. Inductive coding (without a predefined code frame) is more difficult, but less prone to bias, than deductive coding.Code frames can be flat (easier and faster to use) or hierarchical (more powerful and organized).Your code frames need to be flexible enough that you can make the most of your results and use them in different contexts.When creating codes, make sure they cover several responses, contrast one another, and strike a balance between too much and too little information.Consistent coding = accuracy. Establish coding procedures and guidelines and keep an eye out for definitional drift in your qualitative data analysis. Best practises for analyzing open-ended questions If you've made it this far, you'll likely be interested in this free guide: Best practises for analyzing open-ended questions. If you have questions, you can type to our chatbot and we'll do our best to get back to you. Happy coding!