

Research for Marketing Decisions

Vaibhav Chawla

Department of Management Studies

Indian Institute of Technology Madras

Week - 03

Lecture - 13

Exploratory Research Techniques: Secondary Data and Thematic Analysis

Okay, let's get started. Good morning everybody. So what we are going to do today is before we discuss what we are going to do today, what we discussed in the last class. So let me take you through all the sessions. The first class we discussed about the market research process.

The six steps of the market or marketing research process whatever you call it. The second class we discussed in detail about the problem definition and developing an approach to the problem. In the uh third session we continued with the step number two and we had certain examples uh. In session number four we discussed about the research design and types of research design including what are the different types of research design. Exploratory and conclusive. Exploratory and conclusive and conclusive there is descriptive and causal. So, in the last session we had a little bit of idea about the research design what it is and what are its various types. In today's session we are going to discuss very very important which which would be very

you know easy to do as well as very very important to do some sort of analysis exploratory analysis we will see today please pay attention because again whenever you are going to solve a business issue or a problem these techniques would come very very handy, it would save you so whatever we are going to discuss today it it would help you analyze the issue so that you get to the problem second thing it would give you lot of ideas if you are into new product development or if you want to know what customers think. feel intend about your product. So exploratory research design what it is ? It is the design that we use for the initial problem definition step we use there exploratory research design because we are still exploring the situation we are exploring the situation we are exploring the situation in the problem definition step we want to go from

symptom to probable cause that's what that's where we use the problem definition that's what where we use the exploratory research design, exploratory as a word means that we do not know what what what are the variables that we need to deal with what is the probable cause what are its constituents we don't have information about all that we don't know what is going on. So we use exploratory research design various methods of exploratory research design problem audit, interview with experts, qualitative research, secondary qualitative research, secondary data all this can help you explore the issue and get to the problem. Now straight away we will look at the what are the exploratory research methods or techniques secondary data and qualitative research these are the broadly two set of techniques that we use in exploratory research. what is secondary data?

What is secondary data? Anybody who knows the answer? That we get from internet or any other secondary sources but not directly from the people that we are intended to ask. Now let's say when you are conducting a research You would have some set of customers.

They are part of Indian population. Let's say census has collected data about them. The data is about them, coming from them. Will it be a primary or secondary? Primary data is the data that you collect first hand for the problem at hand, which did not exist before you define the problem, before you state the problem.

Secondary data is the data which existed before you define the problem. So, census data whether you do market research or not government will do census after each 10 years of gap. So, there the that data will be there likewise many market research agencies like Nielsen, Kantar, GFK these all would collect some data. Let's say Nielsen does it about the FMCG brand consumption the panel data every month they collect it across many cities so whether you conduct the market research or not that data you can use it you can buy that data that is you can buy the data from companies like Nielsen relevant to your product category so the difference between primary and secondary data Clear?

Secondary data, the data that existed before that somebody else collected but that you could use for your problem at hand. You could use for your issue at hand to get to the probable cause. Sir, aren't these two steps already involved in the marketing business probability finding? Sorry? In the task involved, you have already mentioned these two points.

Yeah, yeah, yeah, right. That is what I am talking about. We have, we now know what are the methods, but now we will clarify them more and see how to conduct those. So, knowing something is, knowing that we use secondary data analysis, but how you do secondary data analysis. First of all, knowing there was no clarity about what is primary and secondary data.

So, I had to clarify that. So, that means primary data for one problem can be a secondary data for another problem afterwards. Yeah. So, you collected data for your problem at hand and this data let us say you put in your company repository. Tomorrow there is another issue that can use the data from your research as a secondary data.

Now, secondary data which means will it be quantitative in nature or qualitative ? Qualitative in the forms of in the form of let's say paragraphs, words, sentences, videos both very good yes it can be both. Now qualitative research is the another technique secondary data is first qualitative research is another technique it is a very very strong sort of technique which is used to get more more and more insights from your customers survey is more about validation you try to validate one correlates with another or one is a you know if you try to sometime look at causality also in the correlation supported by some logic and theory and so on. Qualitative research is more for gathering insights. Qualitative research is very rich which means the kind of insights the kind of data that you get in qualitative you would never get in quantitative but there is a catch in qualitative we always do studies with a very small sample, such a small sample that cannot be a representative of the population because let's say if if you are into a moisture you are into a face cream let's say for females adult females so their population would be in crores now when you want to let's say learn about this particular you know face cream buying and behavior

when you are doing qualitative research you will do it with the selected in qualitative research you will do it with the very small sample. That sample let us say if that sample at max let us say if it even if it is 100, 100 cannot be a representative of a population of crores. And in 100 also there would be some differences let us say age, economic status so there further differences are there and it will be very difficult for you to generalize the insights to the population some what this 50 women are speaking about using face cream you cannot generalize you cannot say you cannot say all it would apply to crore of population of the customers because that sample is very small.

So, after gathering insights from qualitative then you conduct a large sample study survey which where you use a representative samples in several thousands let us say 0.1 percent of the population or 1 percent of the population. And then you try to just say that from the survey that this applies to my population. From qualitative you get tentative things which may or may not apply to population. That is why you conduct quantitative for validating and for confirming the probable cause and then solving it. So, let us move further about the secondary data.

So, sometimes you would require to do this also for collecting secondary data. So there are certain criteria that one has to use to evaluate whether some secondary data is useful or not. So the criteria so you cannot use every bit of secondary data that is not that is on internet because you don't know the credibility of the source you don't know what is the data collection methodology whether it was scientifically rigorous or whether somebody has stated their opinions but has put in the form of a research study which means if you are relying on some secondary data for your business issue at hand you have to first of all check its specifications which means whether in that secondary data it is written how the data was collected whether the data collection process is mentioned how rigorous was it, how representative was it for its population so all those things how the data was collected was it bias-free in the way it was collected?

So all those, whether the methodology is mentioned and whether the methodology looks fair and unbiased. That is the specification. First of all, for secondary data, you have to run all these checks before you use it for your business issue. The second one is error. You have to look at whether the data is accurate

That how you will get again by looking at the methodology. Sometimes when you look at that secondary data you might get some very very different figures that then what you expected. So, which might hint that there might be an error. So, you can validate it from other source as well. Currency means how current is the data.

Suppose you are conducting market research today for a business issue and you are looking at 10 years back census data for secondary data as a secondary data source. Now currency means how current is the secondary data if you are doing something today market research and you are looking at data which is old 10 years or even 2 years back COVID was there now COVID is not there. The data that you would look at 2 years back and now would be would might or might not apply. So how current is the data for your use. Next one is objective for what per objective the data was collected.

the secondary data was collected. So let's say when the secondary data was collected, let's say it was retail, aggregate retail data about the movement of different brands. The objective was to see the movement of brands in organized retail stores. And your purpose is to look at the consumer behavior with respect to your particular brand. or why people buy this particular brand now from that aggregate data you would not be able to get that data will not be helpful because your question is why people buy your brand but the data that you are getting is an aggregate data that won't be helpful so the objectives are different. Next was in the nature and content of the data let's say you want to study the individual

consumption behavior but the data that is secondary data that is available is about household consumption so the nature and the content of the data is not useful for your study at hand. The last one is dependability how dependable is the data which means look at the source whether the source is credible, whether the data is dependable you can rely on that data. So, whenever you are looking at some secondary data just look at where is the source ? Is that a credible company or agency or research firm that is giving you that particular data. That would be one of the ideas regarding the dependability of secondary data. Then if the clear method is given that is another way of seeing how dependable is the secondary data. Some of the e-commerce websites which are not monitored properly by the e-commerce firm, you would see reviews written by their own employees.

So, kind of reviews written by every e-commerce firm, e-commerce firms website, you would for definitely some of the websites, you do not know whether they are credible because sometimes the seller's own people could also write the review if the correct checks and balances are not put up by the e-commerce firm. So, these are some of the criteria to evaluate the quality of secondary data. So, what are some good solutions to evaluate this data? Depends upon what is the kind of research you are doing.

There are lot of government and non-government sources from where you can get the data. You have we have the in library we have both you know academic and corporate sources to get the data. So, we subscribe to CMI database likewise we we we subscribe to Bloomberg I think one terminal they are going to set up. So, likewise there are many financial and non-financial corporate sources of data of or databases that are available through our library and academic resources also.

One just has to know what is the right kind of resource for their secondary data resource for their problem at hand. The broad market research data you could get from the reports.

Market based data, size and the trends data that you can get from secondary data resources. So I will not go into what is the internal secondary data, external secondary data and this and that. We will directly now look at one particular method of doing secondary data analysis.

That will be very very useful. And rest of the theory about secondary data you can actually read from your book where it is very well given about what are the different sources of secondary data, what are the different websites, directories where you can go and fetch the secondary data and so on which is very very useful. You know self explanatory you can just read and find out the sources. What we are going to do now is just to look at how we can do secondary data analysis. This is one of the secondary data analysis exercise that I did for just for the demonstration purpose for one of the restaurants in Chennai the market the business issue was the sales were going down and the management of that particular restaurant did not know why the sales are going down. So one of the ways is secondary data analysis, right?

Now where is the secondary data? You go to the Zomato and reviews are there. right ? Reviews are there so that is the kind of secondary data now once you get to that secondary data what what to do with that data so with the help of only five reviews I am going to analyze that secondary data to get to the probable cause so these five first five reviews are given so the ratings of that particular restaurant were below three out of five. So you want to see what is the probable cause. One way is you look at, you filter the reviews and look at first 50 negative reviews or ratings below 3 and below.

The other way is you just take the first, the latest 50 reviews and try to analyze them to see What could be the probable cause? Now in this case I have done with only 5. So these are 5 are the reviews given for that particular restaurant. You can quickly read 1 or 2.

The first one is rating is 3. The food is decent but the quality looks on the downside compared to what they charge. Ambience was ok. Not on the hygienic side. In the brackets putting spoon directly on the mop table.

The second one is rating 2. What is written is we picked this place for lunch during office hours as most of us were vegetarians and it had decent rating near our office so you can read what is given we picked the buffet overall over a la carte as it was supposed to be their speciality the soups provided were not exactly flavorful but felt good for the weather the kathi roll blah blah blah the others were just mediocre and the main course was again mediocre so till the main course so there is a story written about their experience.

So once you have all these five all the reviews let's say you analyze 50 reviews and their ratings are given, next process is Next process is you need to define your unit of analysis. In this case, unit of analysis which means what will you consider as one unit which you will analyze.

Now, in this case, for example, if you look at the reviews, you would see paragraphs are written. Generally, each sentence or a couple of sentences sentences that are written together are considered as a unit of analysis. So what you will do or let me reframe so there are these different terminologies are there so one would say that once you have these paragraphs the data about the reviews these reviews are there the coding unit you will look at the unit from which you will extract the meetings you will look at the unit from which you will extract the meanings and the minimum unit the generally accepted unit is you look at the sentences and take out the meaning right in one word or two words what is the code code means in one word or two words what is the meaning of this entire sentence in one or two words that is what is called first level coding, open coding. So don't get you know

carried away by the terms the simple idea is once you have the reviews each review entire review will be a unit of analysis which means for entire review you will analyze now in each review each sentence you will derive the meaning by writing you will derive for each sentence you will write a meaning in very brief that will be the open code for example For first user, what is written is the food is decent, but the quality looks on the downside compared to what they charge. So I have written low in value for money. So for these two sentences that were coupled together, that were more or like same sentence, you have to write their brief meaning in the form of open code. Open code means it is giving the same meaning in one or two words about what is written in the sentence.

Next one ambience was okay so ambience okay next code is written then not on the hygienic side putting spoon directly on the hygiene poor so these three codes are coming from the first review so you can do do this exercise on word document by review and edit. So, review show comments by putting comments or track change, but by best is to do by showing comments. So, you know it is like manual process. No no see there are softwares available to help you do this job, but one should also know what goes into those softwares right ? So, this is the basic process this is called open code coding where you pick up that your

unit of this the unit the coding unit is a sentence or couple of sentences that are together, now in this case for first review we put the we put three codes that can summarize the idea of the review, you can say these are the unit of analysis the sentence or a couple of sentences. Now we will complicate it little later after we complete this exercise. So wherein let us say for example let me discuss that itself here. Let us say if you have conducted an interview where there are six questions.

So, we look at how to code them. We will come to it later. Let us not complicate it here. So, now we will go to second one. Second review, if you look at, we picked this place for lunch during office hours, lunch meal.

As most of us were vegetarians, for vegetarians, and it had decent rating, ratings fairly okay. We picked the buffet over a la carte as it was supposed to be their speciality. The soups provided were not exactly flavorful but felt good for the weather. Soup quality not good. The kathi roll in the starter was the only one we enjoyed the most.

The others were just mediocre and the main course was again mediocre. Food mediocre quality and because what is written is the kathi roll in the starter was the only one we enjoyed the most. The other was just mediocre, and the main course was again mediocre, food mediocre quality. So likewise again the codes the open codes are written for again the second review, third review, fourth review, fifth review. Once you have these codes written now simultaneously what you have to do is here I have written the actual names of the people you have to hide their names choose the some pseudo names maybe just like this is written A.S. you can write instead of full name, so you have to hide those names. So once you have you give the pseudo names here I have not given for each reviewer read for each review you

did the open coding for you write all the open codes just like for first person three open codes are written for the second, third, fourth, fifth codes are written now after you write these codes you color code the ones that are similar that you think are talking about the same issue same thing for example if you look at the orange one it is staff service poor, no service staff, no staff to greet. So across the users give a color code to the ones that you think are similar are talking about the same thing. So here it is about staff service poor, no service staff, no staff to greet, poor prompt service, service not giving equal preference to all customer, poor ordering service. So you know this is all about poor service of the staff.

Then there are others, this dark green, desert quality poor, lack variety in deserts. This is about the quality of deserts. Likewise for every, for all the codes you have to do this color coding. Give the same color code to the ones that are similar.

Then in the next step, for all the codes that were put in the same color you gather them under the sub themes, in this case orange in the orange these all sub themes for poor staff service is written. Now what is the underlying idea behind all these codes that you have that is the primary theme poor service quality. Likewise next one desert quality poor all these three codes what is the underlying idea this is poor desert quality and variety likewise for all the similar color sub themes you need to write the underlying idea which becomes the primary theme. The next one also. Now you see these themes are written.

And what is written in the brackets is the how many users, how many reviews are talking about this particular theme. Which means these codes come from out of 5 from how many people? These come from 3, this come from 1, this come from 4, 3, 2, 1. So what is the 4 one? Bad experience and recommendation is coming from 4.

So you got the primary themes that are very similar to your probable causes, but in this case you see there could be many probable causes because there are many themes you need to look at the ones with large frequency so out of 5 which has the most frequency when you do it for 50 or 100 so when you will be doing it for 50 or 100 reviews you will get primary themes and you will also get to know which has the largest frequency. So those 1 or 2 would become your probable causes. Now in this case you could also see the relationship between variables if you want to. So you have poor service quality, mediocre buffet food quality, both are 3-3.

And bad experience and recommendation coming from. So which means... Poor service quality and mediocre buffet food quality might be leading to bad experience and recommendation. So in a way you can also tentatively look at the relationships. That is one of the ways of doing secondary data thematic analysis.

This is the most basic form of thematic analysis. This is called inductive coding. From the data you try to look at the you try to see how things happen, you try to understand a part of reality from the data to understanding what is happening. So, this you could use for many for whatever firm there is if there are reviews available

you could use this technique to get to the probable cause. This is how you put secondary data analysis into practice. Now, what if you have a let us say interview that we will

come in the qualitative interviews. This is just how reviews are analyzed using secondary data method. Let us say you have secondary data about

two three different issues from the same document. let's say here we have taken reviews or let me talk about little bit about the interview data analysis as well when you conduct some interview you have multiple questions, let's say each question is catering to a different business issue. Let's say with one of my executive MBA students, we conducted about nine interviews. The questions were related to, one question was how the sales, how the selling is changing or has changed in the COVID and post COVID world ? In the same interview, it was asked what leads to poor salesperson performance in the IT industry. In the same interview, it was asked what leads to achievement and overachievement of quota for salespeople in the IT industry.

So these are three different questions in the same interview. Now each question you will consider as a separate data document. You will not do the analysis together. You will take each question for a particular business issue separately just like you did for reviews and that you will analyze separately because that is talking about different business issue. In this case each review was a data document.

In the other case when we will be learning about the interviews there will be many questions when we ask in an interview. The questions that are talking about the same issue, you can consider them as a part of the same data document. Whereas, the questions that are talking about different business issue, you take it separately. That is how you do the data analysis for interview. Now, interview is primary.

This is thematic analysis that you can use for secondary data, that you can use for primary data. It is just that in this case, the restaurant, the review data was secondary because it was used, it existed before the problem was, the business issue came into picture. So this is one of the method. Any questions about this? Now there are many softwares that would help you do this.

But if you do not want to use any softwares you can just use the word document comments and do this exercise. Now there are softwares that do sentiment analysis also they will look at they are little advanced version of this one where codes they will also look at whether these codes are positive or negative and therefore they will give you sentiments of customers regarding a particular restaurant, let's say. They will go just, you can do sentiment here as well. You can look at the sentiments and give positive, negative and evaluate the overall sentiment whether it is positive or negative about the restaurant.