Using Consumer Data in Research 1. What is Consumer Data?

Welcome, everyone. I'm Dr. Nick Berman. And I'm going to be taking you through these next three videos on using consumer data in research. So we're going to cover three different things over the three separate videos. So firstly, we're going to think a bit about what is consumer data? So what is this data? And what can we actually do with it, what's the sort of thing we might use it for. And then in the second video, we'll go on to a little bit of some examples of what we can do with it. And the third video, we'll talk about about few methods. These are all short videos. So 10-15 minutes long, so hopefully quite easy to dip in and out off.

So well, what is consumer data? So it's a relatively new term, but it's kind of helpful to think about, consumer data arises out of everyday transactions for goods and services carried out between individuals and organisations. So loyalty card data is a great example of this. So you do your shopping, you pay for your shopping, you might have a loyalty card to get points or whatnot. But that data about what you bought, can give us some really, really interesting insights. You but it is wider than loyalty card data. So it's any kind of transactional based data. So we could be looking at how sales, we could be looking at acount transactions, all sorts of different types of transactions. This data can be personal and can be confidential. So we need to look after it and use it in acceptable ways use it in a way that's consistent with good research practice. And it's also an increasing proportion of all data that's collected about systems today. So you know, the we have census data, which is very comprehensive. But this consumer data is much more frequent, and can be much more useful than census data. It's also worth saying, geospatial, most of this data is often geospatial, as well. And I'm a geographer by training. So we'll come back to geospatial in due course, but the geospatial element of consumer data is often the key bit and a really useful proportion of it. Equally, some people are very keen on data is and some people use data are I tend to use them interchangeably, I'll probably go for data is, that's the one I generally use, but feel free to hear whichever you prefer.

So that's a little bit on what consumer data is, but how does it relate to this term, big data that you may well have come across as well? And big data is often defined as data that you cant open with Microsoft Excel, and we usually interpret that to be it's more than a million rows of data, something along those lines. And there are all sorts of big data out there, you know, you get sensor data, Internet of things, some transaction data does count as big data. But a lot of consumer data isn't big data, you know, think about the consumer data that the consumer data research centre hold, you know, about two thirds of that is not big data, you know, we might call it small data, if you like, or you, you can open it in Excel and do some work with it in Excel. The really big thing to think about about consumer data is that it's always secondary data. So the data is not connected, not collected in mind for research it's collected for some other reason, you know, for the company around the loyalty card scheme, or the company processing the transactions or whatever it might be. So we need to bear that in mind when we do some analysis of it. Also, some of these consumer datasets are what we might call analysis ready. So they're

all nice and clean and tidy, you're very clear with what getting, load it up and you can do your analysis. Some of them are very raw data. So they're, you know, they might be very big files, they might be very unstructured data, and they need quite a lot of processing before you can actually do something, some useful analysis with that. So that we've got these kind of two types of consumer data.

So just just thinking about kind of what what data is available. You know, I mentioned a couple of examples already. But loyalty card data, transaction data, house sale data, probably the big one. There, you can get data from Airbnb, and you could theoretically get data from TripAdvisor and that sort of thing. All sorts of stuff to do with travel, social media data, I would argue comes under consumer data, you know, smartphone logs, Smart Energy meter readings, all sorts and that there's lots and lots out there. But one really useful thing that we can do with them. It complements more traditional sources that consumers so we have different surveys. We've got the census, we've got the ONS Longitudinal Survey. And this comparison can be really, really useful because we know that the provenance, we know the representativeness of these big surveys, and we can compare those with the consumer data to say, Okay, do we have? Is our consumer data representative? Is it a good sample, you know, who's included in the data? And more importantly, who's not included in the statute as well.

So where do we actually get it from? So you know, there's, there's a few specific sources. So the Consumer Data Research Centre is one thing, they host a number of data sets, they're funded by the ESRC, to supply these datasets for research, you can get some of it from commercial companies. And, you know, some companies are really, really open with the data they collect. So Zoopla are a great example, they have an API, you can use it to get some of their data about the house sales and rental listings, and do all sorts of interesting stuff with it. You know, on the flip side, you know, some companies who collect consumer data are very, very pleased with it, you know, Google and Facebook could be the prime examples here. And this is because their business model is centred on making use of that consumer data, because it has a monetary value for them. So they can use it to target advertising primarily. So, you know, there's lots of consumer data, and not all of all of it is accessible for us. And there's also, you know, other sources as well. So there's the UK Data Service, Office of National Statistics, all will hold a whole range of datasets and some of those are consumer data. So there's, there's lots of different places you can get this data from, and lots of potential for combining these consumer data and traditional datasets.

Yeah, and this combinations can be quite quite useful to derive more more insights. So a nice example of that is looking at the housing market, you know, and the census will give us information on migration every 10 years. And only down to the local authority level, if we look at some of the housing data, we can find migration information annually, we can probably do annual estimates, and we can get much more specific than just to a local authority. So it can be a really, really useful tool. However, it is not a silver bullet, you know, this is consumer data, sometimes the analysis can be very tricky, depending on what the data is, and what sort of work we're doing with it. You know, for example, some providers have done some work already on this raw data to produce some what they call analysis, ready data products. So that can be really, really helpful. And equally, once we do some analysis, we will often find patterns in data. But I would say the reasons why these patterns are here are the interesting bit. In some cases, the patterns might be very obvious, you know, we're looking at transport data, we have a morning rush hour peak and evening rush hour peak, you know, and that makes sense. But sometimes

these reasons may not be very, very obvious. So we need to bear that in mind and think about, okay, why might these relationships exist?

And so how do we actually, you know, work with some of these, you know, we've talked about it being secondary data already. And some of them are quite raw, quite big files, unstructured, difficult to work with data linkage is often key to getting a lot of this information out. So for example, you know, we've got some data from zoopla, on house sales, and we could look at that. And it tells us, you know, a lot about house sales and prices and frequency and so on. But if we combine that with rental data, as well, we can start to tell us a bit about okay, where do we get sales? Where do we get rentals. And if we kind of bring that into some energy performance certificate data and electoral roll data, we can start to tie all that together and look at the housing market as a whole and migration and say, Okay, we've got some rentals, we've got sales, and we know kind of roughly where people are going. So you know, are there certain locations where people tend to go from rented property to purchasing a house, and now the house is getting bigger. So they're going from a one bed flat to a two bed house? Or they downsizing? Yeah, they're going from a four bedroom house to a three bedroom house. Are they're going to different areas as well? So we can get a lot of extra data from that.

And so, you know, the results can give us a lot of information, but not as, not all of it is necessarily useful. And so how do we actually go about interpreting this? So, domain experts are key, you know, as is collaboration, so you know, if it very much depends on what area you're working in. But if you're looking at, you know, rental data, it's really useful to have someone who knows about housing, knows about housing tenure involved in your project. If you're dealing with fuel poverty, you know, for the looking at fuel data, does that include information on prepayment metres, you know, that's an important thing to think about, because they're very different subset of customer. And if they are included or not included, that's gonna have quite a big impact on your data and on your results. And there's a couple of different ways you can look at this, you could you could, you could already be a domain expert in a certain area, and then learn some of the consumer data data analysis techniques. Or you could do it your way around. So you could be an expert in the consumer data analysis techniques and learn the domain as well. So you know, which is most useful really depends on your background and what you want to do. But it's important to have those two aspects in your your research team somewhere. One other key kind of aspect, we're thinking about interesting, this is representativeness of data. So how representative is the consumer data that we're using. And this is where the domain knowledge is key. Because if if you're very stuck into the data, you can almost not see the wood for the trees as it were. So you need to think about, okay, who's in this data, but also who's not in this data?

And so what, what sort of methods can we actually do with these, what do we need? So we talked a little bit about kind of domain knowledge. But we have some specifical technical skills as well. So there's some data science skills that might be needed, particularly for working with some of the rawer forms of data. And GIS or geographical information science is also key element as well, because a lot of the data is spatial in its native nature, and it has a location. So it could be a postcode and address coordinates or local authority. But the geographical element is a key way of bringing multiple sets of data together.

The other aspect is we need to think a bit a little bit about data governance and GDPR. And this is because the data we work with, potentially could be individual's information, individual level records. And so we need to handle this in a in a sensitive and responsible manner. Yeah, and the COVID tracing app was a great example of this, you know, so when when COVID first came up, came upon us back in early 2020. And there was a lot of discussion about tracing apps, and how do you trace who might have potentially infected who, and there was these two big approach, centralised or decentralised, and centralised was pushed by a lot of governments. But that recorded who was in contact with her in a central database. And you know, there's a lot of very sensitive information there in terms of where people have been, and is it was open for abuse, potentially, depending who had access to. But then a decentralised approach was developed where there's no central database, but you still managed to do the same contact tracing, but in a different way. And it's a much more secure way of dealing with because there's no central database, so the data can't breach. So interesting things to think about. And a key element of this is training. And there the training is really important, and many other kinds of tools and techniques that you need to use. And it's available as well. So whether it's data science, and it's GIS, whether it's data governance, GDPR, or any of the other aspects that we've touched a bit about, there's lots of training courses out there. So NCRM do some, universities do some, and there's commercial options as well. So have a look on the website and see what's out there.

So we spent a few minutes looking at what is consumer data. So it's all about data from everyday transactions, the goods and services. There's a lot we can do with it, it can be really, really useful. But it's not a silver bullet tool research questions. Couple of key things, it's secondary data. So it's not collected solely for research. It's collected for some other purpose and we happen to be using for research. You will need some background knowledge in your domain area, whatever you're looking at, and some specific technical skills as well. So it's data science, GIS data governance to bring all those together.

So for the next video, we're going to look a little bit of what sort of things we can do with consumer data. So we've got some great examples from different datasets of work people have done so we'll go through those and then the final video, we'll look a little bit about what skills do we need to work with consumer data Thank you very much and see you in the next video.