



The new types of data for market research

Carlos Ochoa | Ferran Savin

What's inside...

1

Introduction: Data. The fuel of insight generation | [page 3](#)

2

The evolution of market research data | [page 4](#)

2.1 Traditional data types [page 5](#)

2.2 New data, new classifications [page 6](#)

2.3 Our approach: A new criteria [page 7](#)

3

The new panorama | [page 8](#)

3.1 New data types map [page 9](#)

3.2 When should we ask, when should we observe [page 10](#)

3.3 Data types and insights [page 11](#)

4

A real-life case of data combination: Pernod Ricard | [page 20](#)

5

About the authors | [page 24](#)

Introduction

Data. The fuel of insight generation

Data is fueling market research; good insight generation is impossible without relevant and appropriate data to support it.

Researchers today are lucky: we are seeing tremendous growth in the amount of information available for research, at a low cost that used to be unimaginable. But this eBook is not about quantity, it's about diversity: the Internet and the proliferation of mobile devices have resulted in fascinating new types of data that can be used, individually or in tandem, to solve research problems that were out of researchers' scope just a few years ago.

In this eBook, we'll shed some light on these new data types and the research problems they can solve.



The evolution of market research

Traditional data types

Classifying data for market research didn't used to be a big deal. Methods and data were placed in one of two broad categories: quantitative and qualitative.

For years, survey data occupied a prominent position among quantitative research methods, mainly due to its versatility and ease of comprehension, and because there were plenty of tools we could use to process it.

Traditionally, qualitative research is regarded as more diverse and harder to analyze than quantitative research. Video recordings and transcriptions of conversations are examples of the raw information produced by qualitative research methods, such as focus groups and in-depth interviews.

The arrival of the Internet added a new dimension to this simple approach to classification. The industry started talking about online and offline data. The

Internet certainly had an unprecedented effect on the market research industry, but it did not change the fundamentals of data, just of data collection: an online survey produces basically the same output as its offline counterpart, and the same goes for online focus groups, online personal interviews, and so on.

But the online transformation of traditional research methods was just the tip of the iceberg. New technologies have the potential to generate completely new types of data that have no exact offline equivalent. And this is where today's researchers see their biggest opportunity.



New data, new classifications

The explosion of new types of data that can be used for market research requires a completely new perspective on how we classify data.

The quantitative-qualitative dichotomy is no longer valuable. The Internet has blurred the line between these data types: can we still talk about qualitative research if we implement a sentiment analysis on 1,000 Facebook users? Can data from hundreds of online mystery shoppers be considered qualitative?

If we stick to formal definitions, the distinction between qualitative and quantitative is based on how we analyze the data. Quantitative research relies on statistical analysis, which aims to produce generalized conclusions from a sample of a population.

This practice is very dependent on the size of the sample, as there is no way to make such a generalization

from only ten in-depth interview participants. For that reason, some traditional methods have been classified as qualitative more out of necessity than decision.

But the Internet has drastically reduced costs, making it possible to increase sample sizes for research methods that have traditionally been used with a qualitative approach.

Given these circumstances, researchers must develop new classification criteria that focus on different aspects of data that go beyond its supposedly quantitative-qualitative nature, or even its online-offline origin. To do this, we could think about two new dimensions.

Our approach: a new criteria

Dimension 1

What is the nature of the data?

Two different types of data can be distinguished depending on their nature.

Subjective data

This label encompasses every kind of data that relates to emotions, sentiments, feelings, intentions, preferences, etc. That is, everything related to aspects processed by the human brain.

Subjective data is...

- Unstable, as people may change their minds.
- Difficult to compare between individuals.

Objective data

Data related to unquestionable facts, such as how much a consumer paid for his/her mobile phone, which political party he/she voted for in the last election, or what brand of car he/she drives.

Objective data is...

- Stable: facts are facts; they cannot change depending on when data is collected.
- Easily comparable among individuals.

Dimension 2

How is data collected?

There are two ways to collect data.

Declarative data

When the people being researched are responsible for actively providing data.

Declarative data is...

- Subject to cognitive and social desirability bias.
- Subject to memory failures.
- Flexible: applicable to almost any research problem.
- Cheap.

Behavioral data

When data is observed from individuals, without the individuals playing an active role in the data delivery process.

Behavioral data is...

- Not susceptible to self-reporting bias.
- Not susceptible against memory failures.
- Rigid: several behavioral data collection solutions may be needed for different research problems.
- Costly.

The new panorama

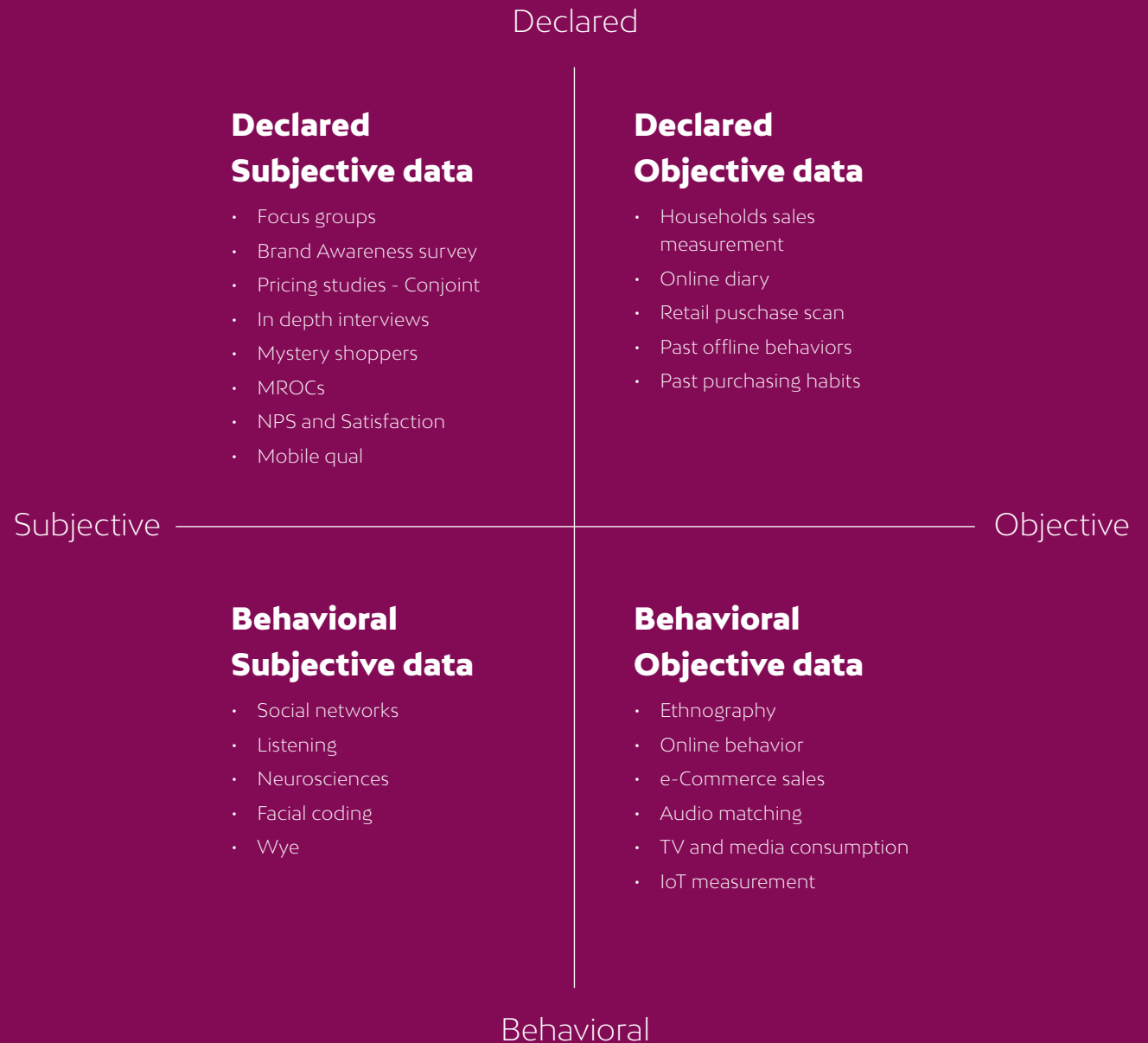
The new data type map

By combining the above dimensions, we can classify almost any new data type we can think of on the following map.

Researchers have little choice when it comes to deciding which type of data to collect (dimension 1), as it depends on the research problem they face. But they do have some leeway when it comes to how the data is collected. For instance, TV audiences (objective data) can be measured in two different ways:

1. Using the behavioral method, by installing an electronic device (known as meter) in a sample of individuals' households and tracking the channels being watched.
2. Using the declarative method, by asking for the same information through a regular survey of the same individuals.

When should we use declarative data and when should we use behavioral data? Or, in other words: when should we ask, and when should we observe?



When should we ask, when should we observe?

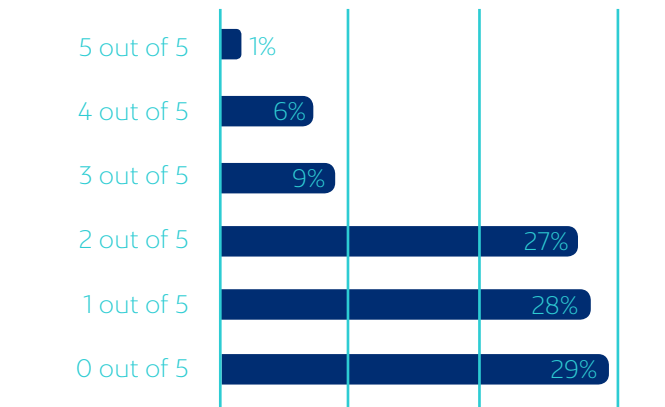
The answer to this question is straightforward: observe whenever possible, ask in all other cases.

Some methodological research has shown that behavioral data has a clear advantage over declarative data. People have serious difficulty accurately reporting past activities, and especially online activities. What's more, people are often tempted to report data that matches an image they have of themselves that differs from the reality.

However, things are not that easy. Observation is not always possible; it is more like a rare opportunity. In most cases, behavioral data collection solutions must be built to serve specific research purposes.

Traditionally, these solutions are expensive, and operationally and technologically complex. Nevertheless, recent circumstances have favored collecting such data. Internet, and especially the mobile Internet, is behind these opportunities.

Respondents recall their last 5 visited websites:



Does memory match reality?

As proved in the paper "**When should we ask, when should we measure**" (M. Revilla, C.Ochoa, R. Voorend & G. Loewe), survey questions that involve recalling past activities are quite difficult to answer because of the limitations of human memory. On the chart above, the graph shows research results: Most panelist (29,0%) did not declare any of the 5 websites observed.

Data types & insights

As we saw before, by combining the two data dimensions, we get four categories of data:

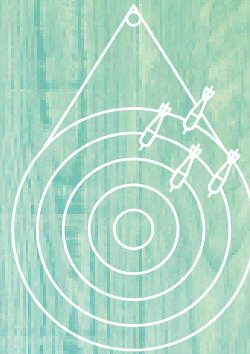
Declared subjective data	Declared Objective data
Behavioral subjective data	Behavioral Objective data

In the following sections, we will delve deeper into these classifications to define what kind of insights we can extract from each data category.

Depending on the nature of the study, it will be obvious what type of data must be used. Whenever several data types are available for the same purpose, it's important to keep data quality in mind when choosing the best option. Data quality is defined as the product of reliability and validity.



“Data quality is defined as the **product of reliability and validity**”



Reliable
Not valid



Not reliable
Valid



Not reliable
Not valid



Reliable
Valid

Declared Subjective Data

Data types that fall under this category are related to emotions, sentiments, feelings, intentions, preferences and opinions collected through active participation of the people under research.

Declared subjective data has been widely used for many years to make crucial decisions on new product development, pricing, and brand positioning, among other things.

To what extent is such data reliable? Can we base crucial business decisions on people's self-reported preferences? Some methodological research raises doubts about the quality of this type of data. However, as we will see later, there are not many alternatives.

When do you need declared subjective data?

Declared subjective data is especially appropriate when we need to gather ideas for new product development, product testing, new concept testing, brand equity evaluation, understanding client satisfaction, ad testing, ad effectiveness measurement, and, in general, when we want to learn the **why** behind a certain behavior.

Which data collection methodologies produce this type of data?

- **Traditional opinion surveys**
- **In-depth interviews**
- **Focus groups**
- **Online communities**
- **Mobile qualitative & quantitative research**
- **Micro-surveys**
- **Virtual Environments / Virtual Reality**
- **Offline / Online mystery shoppers**
- **Conjoint and Maxdiff**
- **Client satisfaction measurement (i.e. NPS)**

Featured innovative methodologies

Online communities.

A market research online community, also known as MROC, is a data collection technique that allows researchers to uncover qualitative insights in real time.

In MROCs, participants are invited to take part in a series of conversations and exercises around a given topic, interacting with each other and with the researcher.

Conjoint online (choice data).

Conjoint analysis (CA) is a statistical technique developed in the 1960s that determines how people value the different attributes of a given product. CA is used to design products and services, measure price sensitivity, etc. The Internet has breathed new life into this technique: conjoint questionnaires (a set of repetitive multiple choice questions specifically designed to address a research problem) can be easily programmed online, and they are more flexible and versatile than their offline counterparts. They enable researchers to see results almost immediately. Thanks to the Internet, CA has become a real option for almost any market researcher.

Online Mystery shoppers.

Mystery shopping is a well-known method of measuring quality of service, as well as compliance with regulations. It is also used to validate product distribution at the point of sale (PoS) and collect information from competitors in their distribution channel. Data collection for mystery shopping studies has traditionally been conducted by professional researchers pretending to be regular costumers at the point of sale.

But with the advent of the Internet and the ubiquity of mobile device use, and in particular with the emergence of online panels, mystery shopping can be done by regular consumers following specific instructions from the researcher, collecting data on the service of interest and, of course, being properly paid for doing so. This new approach allows the industry to collect data in real time, on a broad scale, and from the point of view of real consumers.

Uses & applications

U&A case.

Declared subjective data is still valuable. There are several kinds of Usage and Attitude studies (U&A), but the general aim of the studies is to understand the marketplace and the opportunities for a product or service within a targeted consumer group. Usage and attitude studies try to define the frequency of product use, frequency of product purchase, attitudes towards merchandise in terms of strengths and weaknesses and features a product lacks that would be desirable. These studies also touch brand image and loyalty matters.

U&A are complex studies where it is beneficial to combine different types of data.

Nevertheless, in many parts of these research projects we need declared subjective data. For instance, when we need to understand why consumers use and buy certain products, we need to use declarative subjective data.



Declared objective data

Objective data can be collected in the same ways as subjective data: self-reporting from a sample of individuals.

In fact, questionnaires are still widely used to collect self-reported objective data, and there is nothing wrong with this: a questionnaire is an extremely versatile tool; it is capable of collecting virtually any sort of data. In the same questionnaire, an individual may be asked about her/his current car brand (objective) as well as which car brand he/she is planning to buy (subjective).

However, problems arise when collecting two specific types of objective data:

1. **Hard-to-remember data:** do you remember how much you paid last time you went to a restaurant? How many hours you spent watching TV last week? Exactly which products you bought when shopping last week?
2. **Sensitive information:** it is well known that people tend to misreport objective data related to some specific topics, such as politics, sexuality, and personal hygiene.

Why would researchers prefer to use self-reported objective data?

Very often there are no alternatives, or the alternatives are expensive. However, things are starting to change, thanks to the emergence of new technologies: millions of consumers are now connected via the Internet, and this opens up a world of opportunities. In fact, behavioral data collection is gradually replacing declared objective data in many research problems.

When should you use declared objective data?

Ideally, we would always like to collect objective data through observation (behavioral data). So the answer to this question is straightforward: use declared objective data if there is not a feasible way to use behavioral data.

Which data collection methodologies produce this type of data?

- **Online diaries**
- **Mobile diaries**
- **Households consumption measurement**
- **Past purchase habits**
- **Past offline behaviors**

Featured innovative methodologies

Mobile diaries.

Research diaries are studies that aim to measure consumers' habits on a specific category of product. For instance, a group of individuals can be asked to report their beverage consumption for a week, providing detailed information of each single intake: day of the week, time, location, type of beverage, amount, etc. Data for research diaries used to be collected on paper notebooks.

With the advent of the Internet, notebooks gave way to online questionnaires, but in both cases data was liable to the same problem: data collection occurred at the end of the day, so people had to recall activities that had happened hours earlier (i.e. there was a risk of misreporting). Now, mobile diaries allow people to deliver the data from their mobile devices, almost in real time, which is more convenient for them and produces more reliable data.

Uses & applications

Mobile diary as a mystery shopping project.

Mystery shopping studies and mobile diaries can be used together, thanks to high Internet penetration.

To do so, members of an online panel are asked to visit a point of sale and take pictures of products, gather price information, promotional materials, etc.

This objective data is collected through participant's smartphones and, of course, can be complemented with declared subjective data.



Behavioral subjective data

But is there any way to collect subjective data by observation?

Can we infer emotions, opinions and preferences from people's behavior?

The answer is yes and no.

Recent scientific and technological developments in the field of neuroscience have allowed us to connect physical reactions, such as changes in heart rate, respiration, and blood flow to emotions. This connection can be used to collect subjective data from individuals without asking them and, therefore, eliminating the risk of self-reporting bias.

In a broad sense, we could also consider comments that people willingly post on social media to be behavioral data: this data is observed without asking people to actively deliver it for market research purposes. Opinions published on social media can have a major impact on brands, so researchers would do well to be mindful of recent development in these different regions of the web.

However, both sources of behavioral subjective data have some limitations, which makes declared subjective data even more valuable and even harder to replace.

Neuroscience studies use sophisticated technical equipment, and this usually requires participants to go to a research lab, with costs impossible to afford. To overcome this limitation, many companies have developed software-hardware solutions for performing such studies at home. Nevertheless, remote neuroscience studies are simplified versions of the original techniques.

Data drawn from social media is a useful source for insights, but it can only be used for certain topics:

people are willing to share their opinions on social media on subjects that they are interested in, which are not necessarily subjects that the researcher is interested in.

When should you use behavioral subjective data?

Use behavioral subjective data to make business decisions based on customer's emotions when you don't trust the declarative data alternative. For instance, **ad testing (i.e. evaluating the future performance of a commercial)** is more effective when conducted using neuroscience techniques than traditional declarative solutions (surveys).

Which data collection methodologies produce this type of data?

- **Social media listening**
- **Offline/online facial coding**
- **Offline/Online eye tracking**
- **Other neuroscience techniques**

Featured innovative methodologies

Online facial coding / Eye tracking.

Eye tracking is a research technique that monitors people's eyes movement to understand where they are looking (e.g. a TV commercial); facial coding measures emotions by analyzing people's facial expressions. Thanks to recent technological developments, it is now possible to conduct this sort of research through regular web cams and user's browsing devices. This technology, in tandem with online, makes behavioral subjective data a plausible solution to many research problems at an affordable cost.

Social media listening.

Social media listening is in the same category as buzz monitoring. Data is collected from a combination of sources such as online forums, blogs, and all major social networks. One key benefit of this method is that data can be collected in real time, so marketers can react instantly to people's opinions, complaints and demands. As a drawback, social media data is sometimes complex to analyze and affected by several external factors that must be take into consideration.

Uses & applications

Video testing for inbound marketing.

Inbound marketing has become a key marketing strategy that is focused on attracting new customers through the creation of relevant content for specific buyer personas.

Inbound marketing campaigns can generate a large number of leads and social media interactions.

Posting short videos with key content is one of the most successful ways to engage with potential leads. Technology has made eye tracking and facial coding research affordable, so videos can be effectively tested with a sample of individuals to select those with higher potential interest and to improve video sequences that do not snag the user's interest.



Behavioral objective data

Self-reported information can be inaccurate and biased, depending on the specific type of data we are interested in.

Fortunately, new ways of collecting behavioral objective data have been developed in recent years. The Internet has played a key role in this development. It is this data category that offers the most benefits, thanks to new technologies and methodologies.

Generally speaking, **behavioral data has higher quality than self-reported data**. This is particularly true when we talk about online behaviors. For instance, if we want to evaluate which websites are most visited by a target population (for instance, to assess where an ad campaign should be placed), we could use a survey asking the participants about the websites they have recently visited.

People visit a lot of websites every day, and they jump from one to the other in a matter of seconds; it is unlikely that we would be able to collect reliable data on this topic. What's more, some people will likely under-report some of the websites they have visited (e.g. pornographic websites).

Consider the behavioral alternatives to collecting the same data. We could ask a sample from an online panel to install an application, known as a “meter,” on their browsing devices; this enables researchers to collect the URLs of all the websites visited, as well as other relevant information (search terms, app usage, etc.). This automated data collection overcomes problems related to human memory limitations and under-reporting of some information. Behavioral objective data is currently in the middle of a technological transformation that may reap benefits for your research goals.

When should you use behavioral objective data?

Use behavioral objective data whenever possible, and ask a sample of the target population in all other cases.

Which data collection methodologies produce this type of data?

- **Geolocation research**
- **Online behavioral research**
- **E-commerce sales measurement**
- **Internet of Things (IoT) data**
- **Audio matching**
- **STV & radio audiences**

Featured innovative methodologies

Geolocation research.

Modern smartphones can record geolocation data (e.g. to help the user reach a location using Google Maps). This data can be used for market research purposes: we can understand how people travel, estimate consumption habits in offline stores, identify the most suitable places to open a store or to place an ad, and so on.

Online behavioral research.

Thanks to the development of passive metering technology, data related to online behaviors can be collected. Current metering technology collects data such as URLs visited, search terms and app usage from several devices owned by the same individual (cross-device).

E-commerce sales measurement.

The same online behavioral data, in combination with advanced data processing techniques, can produce estimates related to e-commerce activity: share of traffic per player, sources of traffic, conversion rates, number of orders, prices, etc.

Internet of Things (IoT) data.

Internet connectivity spreads to almost every electronic device we use. We are not only talking about personal computers and smartphones, but also washing machines, refrigerators, cars and so on. This is known as the Internet of Things (IoT). IoT has the potential to produce massive amounts of behavioral objective data to be analyzed for market research purposes.

Audio matching.

Mobile devices also have the capability to record audio from their environment. This data can be analyzed with advanced machine learning algorithms to identify which TV/radio channels are being listened to by the user. The installation of this technology on an online panel allows measurement of TV/radio audiences in a reliable and cost-effective way.

Uses & applications

Online path to purchase

As commerce is increasingly migrating to the Internet, more and more research is being conducted on how people buy online.

Competition online is merciless; hundreds and thousands of e-commerce websites are selling similar or even identical products at similar prices. For owners of online businesses, understanding customers' decision-making process is crucial.

Behavioral cross-device data is helping marketers understand the key factors in the purchase process in order to optimize ad campaigns, pricing, online checkout processes, etc.



A real-life case of data combination: Pernod Ricard

A Digital Path to Purchase in the Liquor and Spirits Industry.

Paper presented by Pernod Ricard in Esomar Latam 2017 with data from Netquest's Behavioral panel.

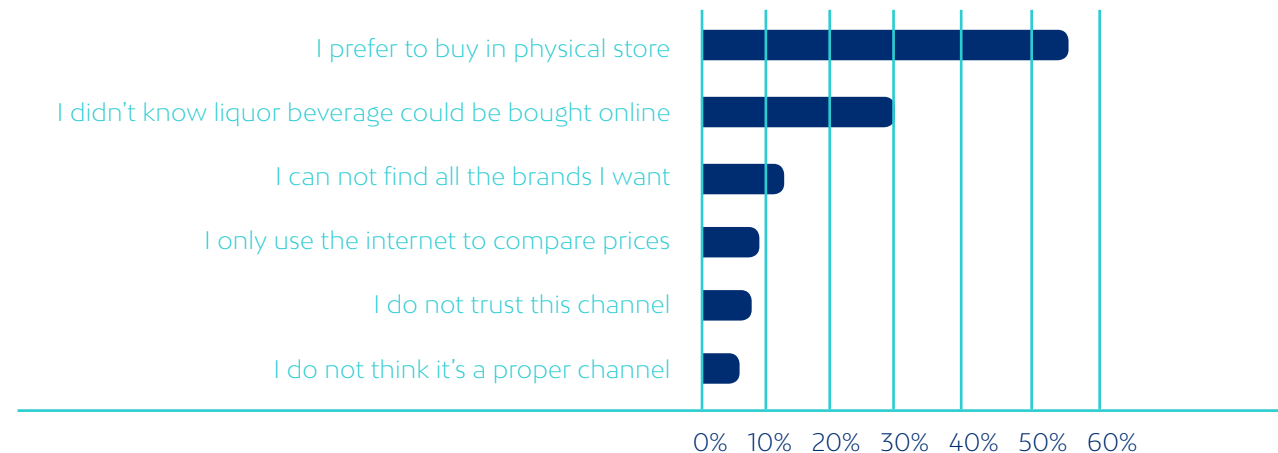
With this paper, María Alejandra and Gabriela Parias from Pernod Ricard offered a real example of the value of combining different types of data.

Research objectives

- Identify the online behavior of the target audience.
- Identify the “buyer’s path to purchase”.
- Develop an action plan to impact the buyers and influence them in the right place.

Design of the investigation

- Methodology: Digital ethnography using passive metering technology.
- Target: Frequent consumers of alcoholic beverages.
- Sample = 661 people from stratum 1-6 Colombia.
- One-month of online navigation data.



Thanks to an analysis of users' digital behavior (behavioral objective data), it was possible to demonstrate, for example, that consumers of alcoholic beverages [1] are not very far along in the adoption of e-commerce as the preferred buying method yet, but they do use the Internet as a source of information.

What is happening tonight?

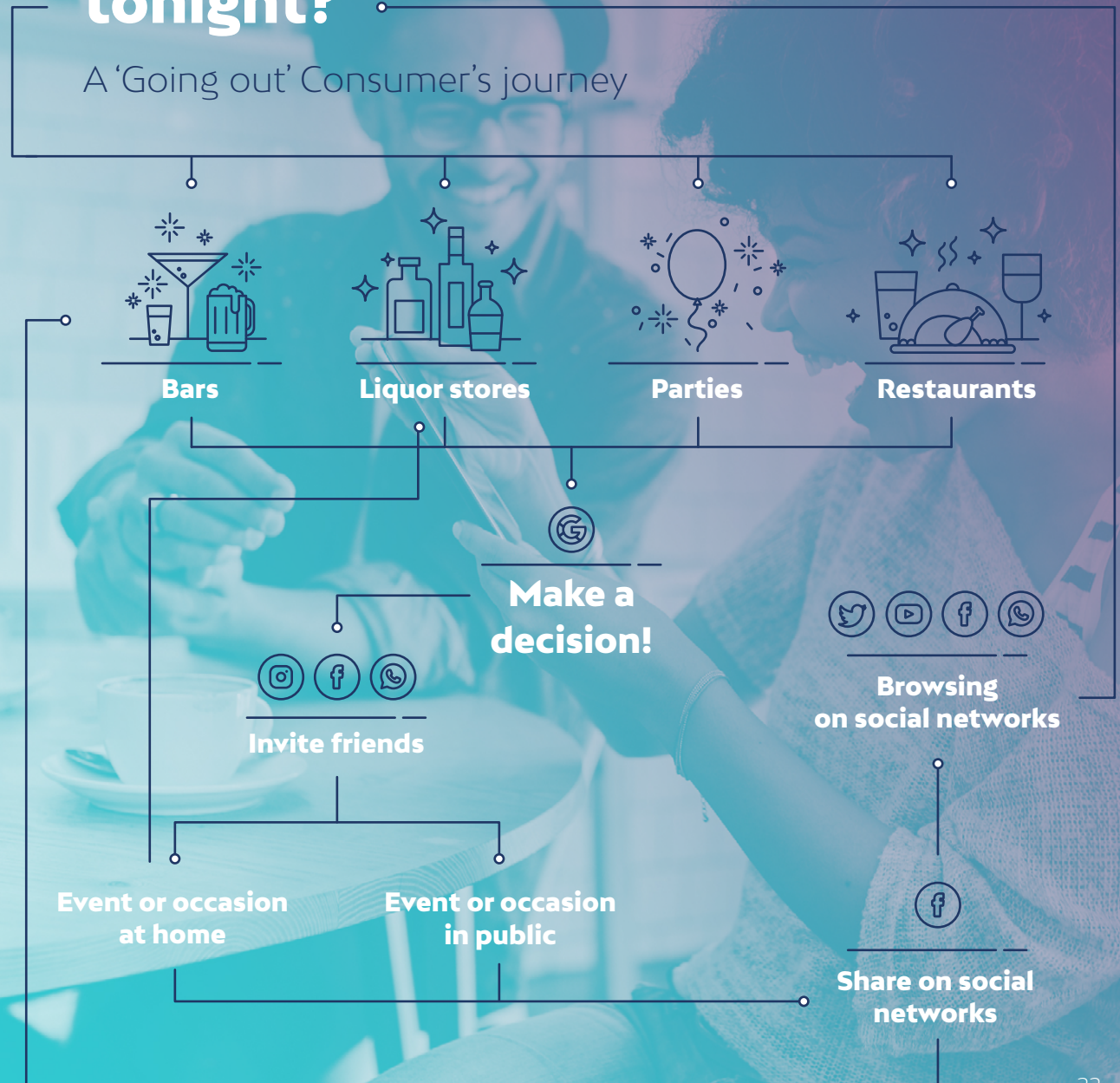
A 'Going out' Consumer's journey

By analyzing over 96,300 keywords, the brand was able to identify the logic behind searches within a certain category of the research, as well as the effect of those searches on online purchasing. This analysis was key in identifying different consumer journeys within this category.

These results revealed that, unlike other categories, such as fashion, beauty, or technology, liquor is not spontaneously searched for online. This understanding help researchers to define 5 diferent liquor journeys:

- What is happening tonight?
- Where can I find alcohol at an affordable price?
- What is the best liquor brand?
- I am looking for a specific brand.
- How to prepare cocktails.

On the right, you can see the Customer's Journey: **What is happening tonight?**



This research, based on the combination of two types of data, allowed the brand to:

- Understand each phase of the alcohol purchase process.
- Identify main consumer journeys.
- Define the main opportunities:
 - › Creating triggers by activating touch points and connecting to occasional consumption.
 - › Educating users about the benefits of the online channel and providing information about the different types and brands of alcohol.
 - › Building trust around the channel.
 - › Integrating online and offline channels.
 - › Generating spaces where recommendations can be searched for and shared, allowing constant feedback.

Results

- By the end of 2016, with only two strategies activated for two journeys, **sales grew 135%** through the e-commerce channels, thanks to efforts to reach consumers at the right time, with the right message, and by offering solutions and value.
- With SEO and SEM actions, **their conversion was increased from 4% to 16%**.
- The brand reached **more than 9 million** people per month, generating a **12% engagement** in their publications.
- By understanding the buyer, brand awareness increased, **purchasing behavior changed**, and their brand grew in the e-commerce channel.

The new types of data
for market research

About the authors



Carlos Ochoa | Chief Client Officer

Carlos is the Chief Client Officer at Netquest, as well as an active collaborator of the joint research program that Netquest holds together with the RECSM (Research and Expertise Centre for Survey Methodology, UPF, Barcelona). He is a regular contributor to several Survey Methodology Publications and Market Research events.



Ferran Savin | Global Product Manager

Ferran is the Global Product Manager at Netquest. Before joining the company, he had been the Product Manager at several international consumer goods firms. He holds university degrees in Business Administration and Market Research Techniques.

**Thank you for
reading.**

**Stay tuned
with us.**



Sample Calculator

Know the number of interviews
you have to conduct

Calculate now





A Symphony of Data

Orchestrating combinations of data to
fine tune your consumer insights.

genuine data