# Citations Guide for Sellers & Marketers

By DJ Capobianco • djcap.net

## Why This Matters

Any time a statistic appears, it is your responsibility to note where it was found. As you remember from university, this is called a citation. the company has a legal responsibility to do this accurately—and you're responsible for the citations in any work you produce, present or distribute!

Per our friends at the Federal Trade Commission, It's also our job to have confidence that the stats we publish are not outliers—we need to show typical results. If you have questions about the company's confidence that a given statistic is typical, please contact a member of the Market Insights & Analytics group.

## Basic Citations

Citations need to appear as close to each statistic as possible—your goal is to place the citation so it follows the stat wherever it goes.

- Every slide with a statistic should have a citation, so that if a slide is copied into another presentation, the citation goes with it
- Every stat in Blog or social media post should have a footnote in the sentence or paragraph (as appropriate), and each post should have the footnote references at the bottom
- Stats shared visually should have the citation within the graphic, so that people who download/screenshot the image still know how the finding was generated

### The core requirement: statistics need to be findable and replicable.

**A Minimum Viable Citation Has...**

1. The name of the company that provided the statistic
2. Whether or not the company paid for the statistic to be produced
3. The name of the report or project from which the statistic was generated
4. The Month and Year the statistic was found
5. The country in which the study was fielded

Use commas to separate these items, unless a comma appears within these items.

Example with commas:
*Source: Kantar Brand Lift Survey, commissioned by the company, Target Holiday Campaign, January 2024, USA*

Example with semicolons:
*Source: Talkwalker; mentions of "Target", "Walmart", and "Nordstrom"; November 2022, Global*

---

# Item 1: The name of the company that provided the statistic

This is usually self explanatory, but there's a few places where it gets tricky.

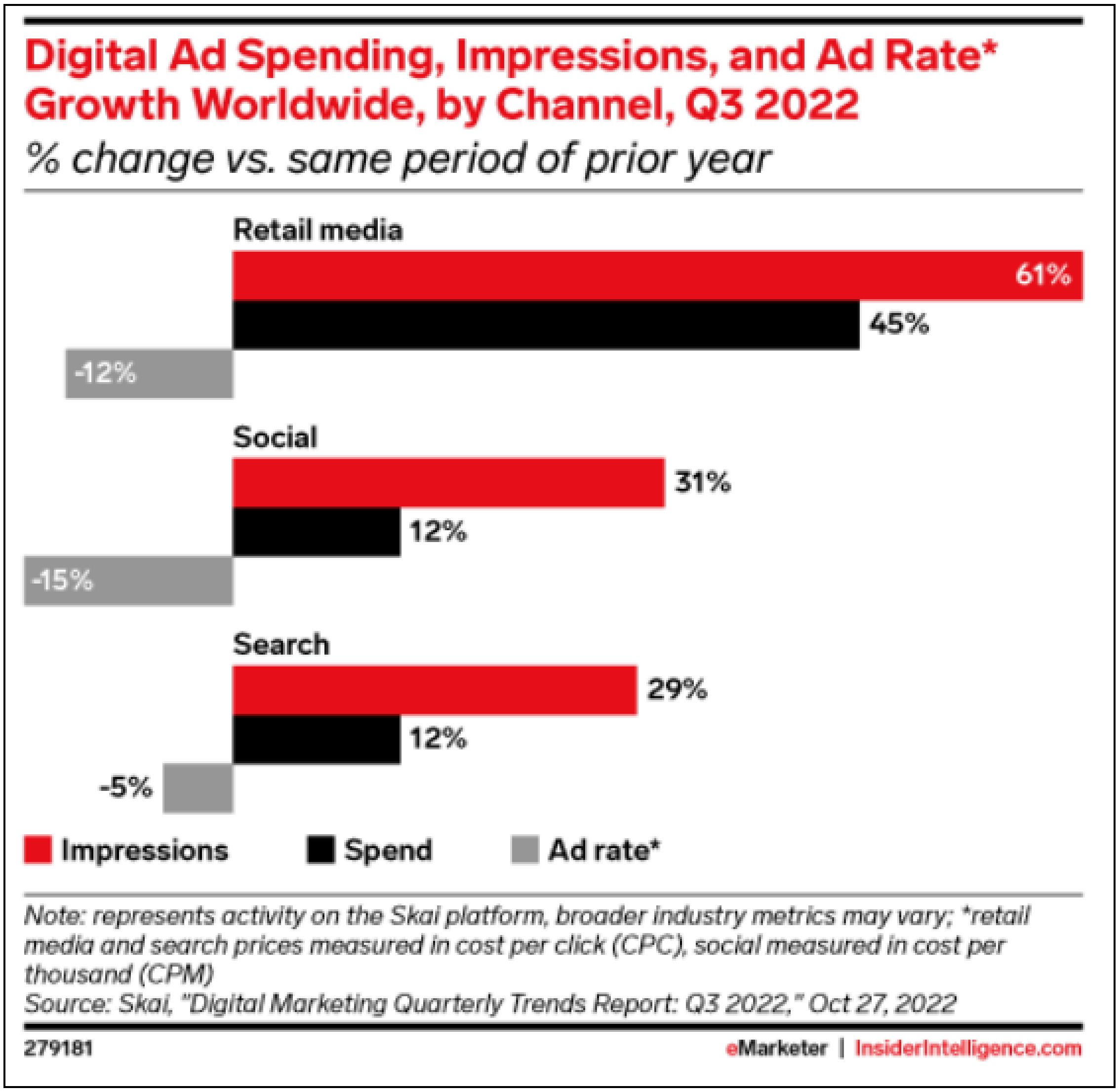
## Stats from Articles and Research Aggregators

Stats found on AdAge, Harvard Business Review, etc are generally straightforward:
*Source: Harvard Business Review, "What PwC Learned from Its Policy of Flexible Work for Everyone," January 2019*

But stats in articles and publications are often originally from another place. When citing statistics that are reported second-hand by a publication or aggregator, use the word "via" to indicate that you found the stat from a secondary source.

*Source: [original research firm] via [where you found the stat], "[name of article in which you found the stat]", [date when the article in which you found the stat was published]*

For example, if you pulled the below statistic from Insider Intelligence/eMarketer from the article "Retail Media Proves its Value Amid Ad Spending Slowdown"




You would cite it by writing:
*Source: Skai via eMarketer/Insider Intelligence, "Retail Media Proves its Value Amid Ad Spending Slowdown," Jan 2023*

## Internal Company Statistics

"Company Internal" should only be used if there's truly no third party that can provide the number. Remember, the market doesn't like when we grade our own homework.

If you have to use an internal number, make sure to provide enough information in your citation for another employee to replicate it.

If this internal stat is going into a presentation, we highly encourage you to put more details on the source in the speaker notes. Helpful details include:

- What search terms you used
- The database you queried
- The date range you queried
- The department that provided the analysis
- The name of the internal document that first surfaced the information
- A URL or intranet link to the original source

## Item 2: Whether or not the company paid for the statistic to be produced

In public settings, we must indicate whether or not a given project was paid for by the company (aka “commissioned”). This is to clarify that a finding was not found via a neutral party, like a university or academic research journal.

Another way to think about this: “would the reader of that stat look at it differently if they knew the company paid for the study?”

Even if we didn’t pay, if we happen to have a material connection with the producer of the study—e.g.,they have invested in the company, the company has invested in them, or we have another relationship that could be considered a source of bias—that should be disclosed within the contents of the materials.

**These types of projects are always commissioned by the company:**

- Campaign Measurement, like Kantar Brand Lift or Clean Room studies
- Major thought leadership projects

**These types of stats are not commissioned by the company**

- Stats found in articles by trade publications, like AdAge or Variety
- Stats from research aggregators like eMarketer, Statista or Packaged Facts
- Company Internal data

**Data from SaaS research platforms may or may not be commissioned by the company**

- It’s not considered commissioned research if you or another employee pull directly from a SaaS platform like Talkwalker, Global Web Index (GWI), YouGov BrandIndex, Comscore MediaMetrix
    - All of these statistics would exist even if we didn’t buy access to those platforms, so data from these platforms are not “commissioned”
- But we need to disclose when we’re not pulling the stats directly
    - Scenario A: The maker of the SaaS tool pulled the data from its own software
        - If the SaaS company pulled the data from its own software on our behalf, cite it as:
            - Source: [Company Name] on behalf of the company, [Name of Project], [Dates Pulled], [Countries from which the data was pulled]

        - In practice, this will look like:
            - Source: Talkwalker on behalf of the company, Big Trend Report, December 2021, France
    - Scenario B: We hired a consultant or agency to use a tool on our behalf
        - If the company that pulls the data from a SaaS tool is different from the company that makes the SaaS tool—like an consultant or agency—cite the stat using “via:”
            - Source: [SaaS tool] via [Agency], [Project Name], [Dates Pulled], [Countries from which the data was pulled]
        - In practice, this will look like:
            - Source: Talkwalker via CapGemini, Makeup Audience Report, July 2024, USA

## Item 3: Name of the report or project from which the statistic was generated

Your objective here is to use a consistent name such that another employee can find the source you’re referencing. Please don’t invent a report name!

### If the stat is from a publication or data aggregator like eMarketer:

…use the name of the article in which you found the stat. A reasonable person should be able to search for the stat and find it in five minutes.

- Good citation: *Skai via eMarketer/Insider Intelligence, “Retail Media Proves its Value Amid Ad Spending Slowdown,” Jan 2023*
- Bad citation: *eMarketer/Insider Intelligence, Jan 2023*

### If the stat is from a one-time study commissioned by the company:

…use the name that appears in the original report that came from the vendor. Don’t just cite the most recent presentation/blog post from which you found the statistic.

### If there’s no report name because it’s from an internal source or SaaS tool:

…describe what was pulled. If possible, insert all search terms or parameters that were used within the citation.

- Examples:
    - *Source: Talkwalker; mentions of “Target”, “Nordstrom”, and “Walmart”/”Wal Mart”; Sep. 30, 2015 to Nov. 21, 2015; USA*
    - *Source: ComScore MediaMetrix; visitors to TikTok and other social networks, ages 25-34; June 2023; USA*

    - *Source: Company Internal, promoted impressions from @Target and @Walmart, Jan 2024 - Feb 2024, Global*

- If the data pull is complex, put details in an easily accessible appendix or speaker notes:
    - *Source: Talkwalker; mentions of Target and competitors (search terms in Appendix); Sep. 30, 2015 to Nov. 21, 2015; USA*
    - *Source: Talkwalker; mentions of Target and competitors (search terms in Appendix); Sep. 30, 2015 to Nov. 21, 2015; USA; Global; posts from among 4000 users who self identify as mothers in their social media bio (search terms in Appendix)*
    - *Source: Company Internal promoted impressions from US Retail profiles (list of brands included in the Appendix), July 2022 - September 2022, USA*

## Item 4: The month and year the statistic was found

**Avoid writing dates with numbers alone!** While Americans write dates as Month/Day/Year, every other country on the planet writes dates as Day/Month/Year. To avoid confusion as your materials travel to employees around the world, use “Jan 10 2018” or “10 Jan 2018” instead of “1/10/18.”

Also, be sure your stat is reasonably recent. Our platform changes materially every year, so use good judgment as to whether the stat you’re citing is recent enough to still apply.

### If the stat is from a publication or data aggregator like eMarketer...

Use the date the article was published, not the date you found the stat. If you can’t find a month and year on the article, it’s probably an outdated stat and you should remove it from your materials.

### If the stat is from a one-time study commissioned by the company...

Do: Use the date the research was “in field” -- e.g., when a survey was being completed by respondents, or when participants were physically in a lab Don’t: use the date when we finalized our research deliverable.

### If the stat is from a SaaS tool...

Use the dates that were part of the query, not the date you retrieved the data.
Example: *Source: Talkwalker, mentions of “Target”, Dec. 4, 2020 to Jan. 20, 2021, Canada*

## Item 5: The country in which the study was fielded

Behaviors and attitudes found in the US should not be assumed to be global, and while you’re welcome to share statistics across countries, we should be clear about the nationality of the folks included in the sample.

# Advanced Citations

The problem with minimum citations as described above: they don't truly allow for another employee to find a specific stat. There's no proof within the citation that it is indeed the source; even if it was, it may be hard to track down a specific finding.

A better citation will have all of the information discussed above, as well as more details for where within a given report a stat was found.

Our clients are smart—they're going to want to know more about our work. Make it easy to build trust with them by including all the relevant information right on the slide!

**There are five elements of bias behind each statistic, and they greatly influence its outcome**

A. How the data was collected -- i.e., via a survey, in a lab, or with tracking software
B. How we calculated the result
C. Who participated in the study, and why we chose them
D. How many people were part of the study

**Here's an ideal citation for survey-based data:**

- Statistic: Compared to the general population, TikTok users are 7% more likely to consider buying a VR headset
  - *Source: Kantar Millward Brown, commissioned by the company, Tech Survey 2023, April 2023, USA, Q45. "How likely are you to consider purchasing the following?", Net "Somewhat Likely" and "Very Likely", Base = TikTok Users who are planning on buying consumer tech in the next two months, n=700*

**Here's an ideal citation for behavioral data:**

- Statistic: 40% of Viewers on Netflix saw the entire ad
  - *Source: NeuroInsight, commissioned by the company, Unconscious Associations, May 2023, USA, views calculated with eye tracking software, base = 18-65 y/o streaming users who use Netflix, Disney+, and Max at least 3x weekly, n=1235*

Although these "201" citations are longer, they should remain on the slide or blog post for a viewer to see. If you truly run out of room, you can put them in the speaker notes or appendix. If you move a citation to the speaker notes, indicate that on the slide.

Examples:

- *Source: Kantar Brand Lift Survey, commissioned by the company, Target Holiday Campaign, January 2024, USA,* question language in speaker notes
- Source: Talkwalker, mentions of "Samsung", July 3, 2023 to Nov. 4, 2023, Brazil, full boolean query in appendix

# Item A: How the data was collected

## For surveys: How the question was asked

*Source: Ipsos, commissioned by the company, Autos p2p 2024, November 2023, USA, Q65. “How likely are you to consider purchasing or leasing the following vehicle types?”, Net “Somewhat Likely” and “Very Likely”, Base = platform users who are planning on buying a car in the next six months, n=700*

It’s really easy, and unfortunately common, for marketers to write leading and unfair questions so they get the best outcome. That means that a smart audience is going to want to see the specific language we used in our surveys. That means that your citation should include the specific language in the question. And to make the stat easy to double check or reproduce, it should also include the question number.

## For behavioral data: how the behavior was tracked

*Source: NeuroInsight, commissioned by the company, Unconscious Associations, May 2023, USA, views calculated with eye tracking software, base = 18-65 y/o streaming users who use Netflix, Disney+, and Max at least 3x weekly, n=1235*

The guidelines for citing behavioral studies are, in spirit, the same as those for citing survey data. The specifics are a little different: here, we want to concisely note the methodology or equipment used.

# Item B: How we calculated the result

*Source: Ipsos, commissioned by the company, Autos p2p 2024, November 2023, USA, Q65. “How likely are you to consider purchasing or leasing the following vehicle types?”, Net “Somewhat Likely” and “Very Likely”, Base = platform users who are planning on buying a car in the next six months, n=700*

As you know, statistics are the result of how people react to a question—either by answering a question in a survey, or by behaving in a certain way after a stimulus in a behavioral survey.

For clarity, simplicity, and storytelling, we often will report a statistic that adds together the results of multiple reactions—this is called a “net”. In those cases, we should indicate what steps we took to calculate that number.

In the above example, we’re reporting the number of people who chose either “somewhat likely” or “very likely” in the survey.

## Item C: Who participated in the study, and why we chose them

*Source: Ipsos, commissioned by the company, Autos p2p 2024, November 2023, USA, Q65. "How likely are you to consider purchasing or leasing the following vehicle types?", Net "Somewhat Likely" and "Very Likely", Base = platform users who are planning on buying a car in the next six months, n=700*

Another way that dishonest or careless marketers goose their research is by only including people who may react well. For example, if someone asked "How favorable is your view of the company" to a group of people who use our product every day, we can safely assume their response will be different than from the average person who doesn't use the product or uses it less frequently.

(Similarly, people who use our product frequently—and are therefore familiar with it—would likely behave differently in a laboratory study.)

That means that your citation should include a description of the "base:" what attributes make up the group of people from which the stat is calculated. Use your best copywriting skills to describe the group(s) concisely, but make sure we're not obfuscating the spirit of the base.

### Do I need to describe my base when I pull data from a SaaS too?

If you're pulling from among a non-obvious population, yes you do. For example, if you're pulling behavior from a certain type of user, or looking at website visits from a certain group (like Gen Z), be sure to describe the details of the base definition.

- Statistic: Pins about Kraft from Moms on Pinterest are 20% more commented on than overall Pins about Kraft
    - Source: Talkwalker; mentions of Kraft and competitors (search terms in Appendix); Aug 15, 2022 - Aug 15, 2023; Global; Pins from among 7000 users who self identify as mothers in their Pinterest bio (search terms in Appendix)
- Statistic: TikTok reaches 90% of Gen Z
    - Source: ComScore MediaMetrix; visitors to TikTok and other social networks, ages 13-24; March 2024; USA

## Item D: How many people were part of the study

*Source: Ipsos, commissioned by the company, Autos p2p 2024, November 2023, USA, Q65. "How likely are you to consider purchasing or leasing the following vehicle types?", Net "Somewhat Likely" and "Very Likely", Base = platform users who are planning on buying a car in the next six months, n=700*

As you know, the number of people from which a statistic is calculated is called the sample size, sometimes referred to as the “N” (short for number). When citing, include either “sample size = __” or “n=__” in your citation to prove that the stat is from a reasonable size sample.

## What’s a reasonable sample?

Our friends at the FTC like to remind us that we’re legally obligated to only report what could be reasonably considered a “typical result.” Lower sample sizes are statistically less likely to produce typical results.

**For surveys:** a sample to be projectable to a large population, like an entire country or our user base, we should strive for a sample of at least 500 people. For brand studies and subcuts of demographic research, samples should be at least in the hundreds. We try our very best to avoid reporting from samples fewer than 150; if we do, it should be visually indicated in both the citation as well as in the design of the slide or post so it’s obvious to our audience.

**For behavioral or laboratory studies**: Studies that aim to be projectable to a very large population—for example, an offline sales study—should follow the same guidelines as above. Laboratory studies will by nature be smaller, which makes them less projectable to larger populations.