A GUIDE TO

CUSTOMER DATA PRECISION AND MATCH RATES

A Guide to Customer Data Precision and Match Rates

When you query an **Application Programming Interface (API)** to enhance your customer data, how can you be sure the data you receive is relevant, accurate, and worth paying for? This guide explains what you need to know before you start paying a fee for public data that supplements your existing customer data.

UNDERSTANDING API QUERIES AND MATCH RATES

When you send a query through an API to search for public contact information, it's similar to entering a search phrase in a search engine. In the case of a contact data API, your search query is based on an email address, company domain, phone number, or social handle. When you initiate the search, you're basically asking the search engine—the API—to look through all the available information and match as much relevant information as possible to the information you entered in your query.

The percentage of your queries that return one or more related data points is called a *match rate*. For example, if you query an API with 100 email addresses, and the search algorithm returns seemingly relevant bits of information for 33 of the email addresses in your query, your match rate is 33%.

Sometimes companies charge customers for queries, and sometimes companies charge customers for matches, depending on the nature of the search and the type of data being searched for. Here's how it works:

- When you pay for queries, you may or may not get any data back. You're simply paying for the time it takes the server to conduct the search for you. Regardless of whether your search returns matched data, you still pay for the server time.
- When your pricing plan is based on paying for matches, you don't pay for the queries; you pay for the data resulting from the search.
- Some plans are hybrids based partly on matches and partly on queries. Some plans may charge an additional fee for increasing the rate at which your queries run and matched data is returned, which is referred to as a *rate limit*.

TUNING MATCH RATES WITH PRECISION

Your match rate is dependent on two factors known as recall and precision. Here's the difference:

- *Recall* refers to how much data is found to be at least somewhat relevant to your search query. Recalled data could be *inferred*, meaning that the search algorithm made educated guesses to find relevant and up-to-date data. Inferred recall isn't as accurate as other methods of recall, but it's better than random guessing.
- *Precision* refers to the percentage of confidence in the relevance of recalled data. Precision increases when recalled data is actually observed and *cross-validated*, or checked and re-checked against other precise data points to ensure a high level of confidence in the match.

When paying for data based on a match rate, the best case scenario is to have a high match rate (recall) and high confidence (precision) as shown in Figure 1. That way, you're paying for the most accurate data and you're receiving as much data as possible.
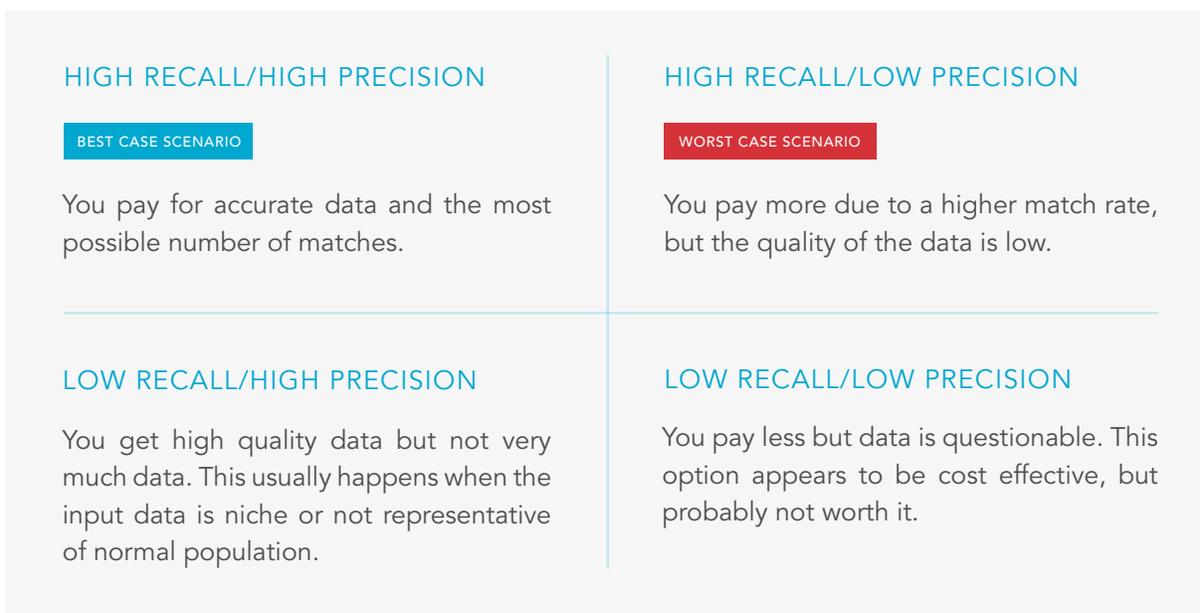
HIGH RECALL/HIGH PRECISION

BEST CASE SCENARIO

You pay for accurate data and the most possible number of matches.

HIGH RECALL/LOW PRECISION

WORST CASE SCENARIO

You pay more due to a higher match rate, but the quality of the data is low.

LOW RECALL/HIGH PRECISION

You get high quality data but not very much data. This usually happens when the input data is niche or not representative of normal population.

LOW RECALL/LOW PRECISION

You pay less but data is questionable. This option appears to be cost effective, but probably not worth it.

FIGURE 1 — RECALL VERSUS PRECISION

You might be happy with a match rate of 33% or more, but did you know that match rates can be adjusted up or down according to how much confidence there is in the relevance of the matched data? Whether your match rate is high or low, you need to be sure that the data you're paying for is relevant, accurate, and up to date before you agree to pay based on a match rate.

HOW TO BALANCE PRECISION AND MATCH RATES FOR YOUR BUSINESS USE CASE

It may seem obvious that a high match rate with high precision is a good deal. However, based on your use case, you may be willing to sacrifice some precision in order to receive more data. Companies like FullContact that are able to tune its level of precision using cross-validation can result in data confidence of up to 99%. But most businesses prefer to pay for data at 95% precision by allowing a certain level of inference in the matched data. That way, match rates can be increased to a tolerable level without sacrificing too much data quality.

FullContact matches your data at or above 95% precision, because that's what our customers typically ask for. When you use FullContact's APIs, you can be sure that your match rate is not artificially high and that the data you're paying for is cross-validated and not wasted on too many inferences or unvalidated matches.

CALCULATING THE REAL COST OF LOW PRECISION MATCH RATES

If you're paying for a high match rate and questionable precision, you or your customers are assuming the cost of data quality in terms of your own time and effort. For example,

- If you're building low precision contact information into a product feature or an app, your customers will need to spend time and energy correcting and updating any inaccurate data they receive through your product.
- If you're using low precision contact data in your own systems and processes, your sales, marketing, IT, or customer support teams will need to spend time and energy correcting and updating any inaccurate data you introduce into your systems and processes.

To make matters worse, there's a snowball effect when you pay for low precision data over time. Many bad data inferences made repeatedly could result in a database that's infected with invalid matches, which could in turn result in more invalid matches or inferred matches with each new query.

Make sure you discuss match rates and precision with your contact data provider so you can maximize the return on your data enrichment investments and keep your data and operational efficiencies sound over the long term.

BE EXTRA DILIGENT WHEN DATA PROVIDERS ALLOW YOU TO KEEP ALL MATCHED DATA FOREVER

Contact information is always changing and goes out of date frequently. If your database contains rapidly changing data and data matched with lower precision results, your overall results can deteriorate rapidly over time.

It can sound like a good deal when a data provider allows you to keep the data they sell. But the hidden costs of storing rapidly changing data are high. Instead of paying based on how much data you can keep, make sure you're paying based on how much data your provider can keep up to date for you. It's far better to pay a licensing fee to access data that's frequently refreshed than to pay once to acquire and keep data that deteriorates in quality.

MORE ABOUT FULLCONTACT PRECISION AND MATCH RATES

To learn more about how FullContact's precision data can make your match rates worth paying for, contact a FullContact representative today and request a custom report to find out how your actual data matches up.

Get yours today at:
www.fullcontact.com/enterprise