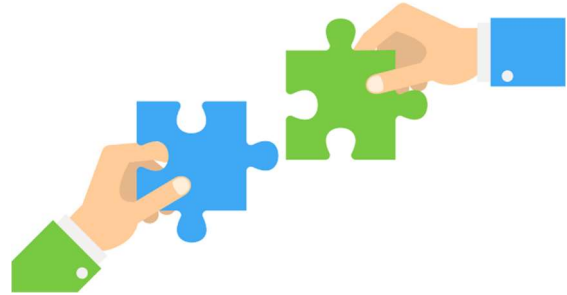


Matching two databases on company names

Keeping your data lake up to date with information on decision makers, potential needs, and financials databases is an incredibly valuable tool to bolster your customer database and deepen your understanding of the market.

The good news is that you've already identified the top data sources for your activity. The bad news? You must aggregate all of these data sets into one to unlock their true potential.

Most of us will admit that aggregating databases can be an overwhelmingly humbling experience - one that highlights errors and approximations in our datasets. The process is time-consuming, difficult to automate, and requires expertise and often perspiration!



That is precisely why we have compiled eight essential tips we have discovered through matching thousands of corporate data files on company names over the past decade.

We hope this informative guide will help you streamline your data aggregation process and encourage you to take advantage of all the useful insights your datasets have to offer.

1. The steps

By combining expert knowledge and algorithmic assistance, matching two databases on company names is more manageable than it might seem. Here are the general steps to follow:

- First, an automated process is used to catch exact matches. Typically, getting 60% of exact matches is a commendable feat.
- But as you dive deeper, you'll soon realize that approximated matching is required to find additional matches. At this point, a blend of manual and automated processes with some domain expertise may lead to 90% to 98% matching accuracy. For most datasets, it is achievable within just a few hours.
- The final 2% of matching requires comprehensive research and will involve digging through records at a very detailed level. It could potentially become a time-consuming process.



Matching two databases based on company names is an iterative and manual process that requires business expertise, but it can be assisted by algorithms.

2. Different company names for the same company

The naming conventions surrounding companies can take on different forms, from

- brand names versus legal structures (think Apple versus Apple Inc.), to
- full names versus abbreviated names (like Amazon Web Services versus AWS).

These names may also include any combination of

- legal abbreviations (like AG, Co., Corp., Inc, LLC, LLP, Ltd or SA),
- geographical markers (like US, U.S., USA, or U.S.A.),
- descriptors of activity (such as REIT, Trust or Holding), and
- even rare characters (like ü, è, ò, ©, or ™).

When it comes to finding the perfect match, even tiny differences in company names can create major noise. Fortunately, most of these differences can be easily eliminated through proper data formatting and leveraging best practices for approximate matching (see below).



Knowing the edition rules of the company names of your two data sets before matching them is an invaluable investment.

3. Company names often change

Business entities often evolve and undergo structural changes, with even iconic companies like Facebook rebranding as Meta Platforms or Google restructuring under Alphabet.

Our records indicate that over 15% of company names change every year. These changes can range from minor tweaks to complete overhauls and cause headaches it comes to matching data sets.



Selecting two high-quality, recently updated data sets can save considerable time and resources.

4. Spotting identifiers

Start by looking at the data and identifying any unique identifiers that could link the two datasets together such as company name, stock ticker, web site or other key characteristics.

Matching multiple fields will improve accuracy and verify consistency.

Our 5 (recommended) unique identifiers:

1. Company name
2. Stock ticker
3. Company web site URL
4. Current CEO's last name
5. Company switchboard phone number

Beyond the unique identifiers, cluster identifiers will help assessing whether two companies with the same name belong to the same industry, country, or segment. Take Alphabet, for example: it's a listed US technology company and the parent company of Google. However, there's also a German automotive dealer company with the same name, completely unrelated to Google's parent.

Our 3 (recommended) cluster identifiers

1. Company country
2. Company industry
3. Company size proxy

Measuring company size proxy can be done with various KPIs ranging from revenue and assets to market value and Fortune 500 ranking. By sorting these KPIs into categories like huge, large, medium, small, and tiny, you can make accurate comparisons.



These recommended identifiers are publicly available information, so they're easy to verify and likely to be present in any databases.

5. Formatting your identifiers

When it comes to record linkage, consistency is key.

Properly formatting your selected identifier will go a long way in ensuring accurate results. Take the time to search for missing or incorrect values, outliers, and other anomalies that could impact matching outcomes. Along the way, pay special attention to spelling company names correctly and using consistent abbreviations.

Once proper formatting is complete, it's time to choose the ideal matching method. When dealing with two databases with different company names, side-by-side record linking is a great starting point to enhance flexibility and precision.

To get an exact match, you can utilize Excel's [VLookup](#) function with exact match on company name, stock ticker, and corporate website URL.



To begin, attempt to achieve exact matches on as many records as possible using your unique identifiers.

6. Approximate matching

Fuzzy matching algorithms can be very helpful when comparing large amounts of data. By adjusting the level of approximation, you can control the amount of variation that is acceptable for each data point. This makes it particularly useful when working with databases that may have errors or typos.

Weighting is a technique that assigns different levels of importance to certain data points based on their uniqueness. This enables the algorithm to focus more on the significant data points, resulting in more precise matches.

On Excel, one simple way to get started is to insert an asterisk before or after a company name to expand the search. For example, "Amazon*" will match with both "Amazon.com" and "Amazon Retail." For advanced Excel users, the [fuzzy match](#) function is highly appreciated.



To maximize record matching, try out different approximate matching techniques.

7. Recording correspondences

When it comes to recording correspondences, ensuring that all results are accurate is non-negotiable after completing the matching process. The ultimate objective is to catch any unexplainable deviations.

And where to begin? Well, your cluster identifiers (country, industry, size) are excellent starting points, but don't stop there. Dig deeper. Look into area codes on company phone numbers, nationality of family names in executive teams, location of the owning company, and social media accounts amongst other data fields. Let your creativity roam and spot suspicious matching.

To ensure maximum benefit from our findings, it is essential to document correspondences on permanent record identifiers for each data set. This documentation allows us to refresh our data set and authenticate matches accurately. Relying solely on correspondences based on company names is not sufficient, as they will inevitably change.



Assessing and documenting correspondence is decisive for the effectiveness of any future data matching.

8. Learning by doing: The art of matching records

Databases are like publications: they keep their readers entertained with carefully curated content and adhere to strict publishing rules. Mastering these rules can help you streamline your workflow.

When it comes to record linkage, you have two options: automated or manual. While automated techniques may seem like a quicker solution, it's important to note that manual methods assisted by algorithms are preferred due to their superior accuracy. Plus, the flexible processing of data makes it easier to customize your approach.

Your customers are always evolving, which means any change is a potential business opportunity. The more files you match, the more skilled you become at detecting valuable patterns and insights.



Hone your record linkage skills to improve your business outcomes and don't miss out on the hidden gems in your data !