



## Beware of Legal Analytics Impostors

---

*It's worse to have bad analytics than no analytics, because decisions based on inaccurate or incomplete data can be harmful and wrong.*

---

### Don't fall for pretty pictures posing as legal analytics

After Lex Machina coined the term “legal analytics,” the field exploded. Now, there are a lot of tools claiming to provide legal analytics. But analytics is more than just visualizations and pretty graphs. Anybody can slap a pie chart on a dataset. Real analytics are tools for making data-driven decisions. The key is having the right data, and ensuring that it's complete and accurate. That's a lot harder than making a pie chart. Many “analytics” providers don't do the required hard work, and they're hoping you won't notice.

Meaningful analytics have to be based on relevant data, with two requirements: 1) the right kind of data, and 2) complete and accurate data. For example, if you were planning to sell your car and wanted to know a fair price, you wouldn't care about the average resale price of all cars combined. You need the price to be specific to your car, or at least as close as possible. A tool might provide complete and accurate resale prices for all cars, but if it doesn't break down the cars by make and model, it won't help you sell your car. Even worse, a tool may claim to have make and model information, but the pricing data is incomplete or inaccurate. Just like car prices, accurate legal analytics is all about finding and analyzing cases like yours.

Beware of statements like “Nothing is perfect, but these analytics are better than nothing.” That's actually not true. It's worse to have bad analytics rather than no analytics, because bad analytics provide false confidence for bad decisions. Would you use medical diagnosis software that provided an incorrect diagnosis half the time? Of course not. That's not “better than nothing” - that's dangerous.

The following sections review categories of data necessary for meaningful legal analytics for federal district courts, and they compare Lex Machina to a new entrant in the field, Westlaw Edge's Litigation Analytics.

## Case Types: Find the Complete and Accurate Set of Cases in Your Practice Area

---

*To perform meaningful legal analytics you need to be able to create sets of cases that are just like your case.*

---

Focusing on the case types you care about is the first step in generating actionable legal analytics. If you're interested in *product liability* cases, data from *patent* cases isn't relevant and would skew your results.

Lex Machina classifies each case using a combination of automated technologies (keyword searches of pleadings documents, recognition of patent numbers, NOS codes, etc.) and human oversight. In addition, Lex Machina allows a case to have more than one type, which reflects the reality that cases may have multiple types of claims, such as a mixed *patent* and *trademark* case.

By contrast, Westlaw Edge relies exclusively on PACER's Nature of Suit (NOS) codes, which will lead to incomplete and inaccurate results. NOS codes are entered by attorneys when filing a case, and neither PACER nor the courts enforce any accuracy requirements or correct errors. Consequently, NOS codes are frequently incorrect. Moreover, PACER only allows one NOS code per case.

For example, over 25% of Lex Machina's *trademark* cases are found outside of NOS 840, the trademark NOS code. If you only used NOS 840 to focus on *trademark* cases, you'd be missing thousands of cases. As another example, *trade secret* is an important area of law, but there is no NOS code for it. Consequently, if you wanted analytics about *trade secret* cases, you can't find them in Westlaw Edge. You can find them in Lex Machina, because Lex Machina created a case type for *trade secret*.

## Case Tags: Granular Data Refinement for Meaningful Analytics

---

*To create sets of cases like yours, you need practice-specific information available with each case.*

---

After identifying the broad case type that you care about, you typically want to further refine your case list. *Employment* covers a wide range of claims; if you're interested in *Fair Labor Standards Act (FLSA)* cases, you don't want to include *age discrimination* cases that would skew your results.

Lex Machina provides a variety of case tags that are specific to individual practice areas. We create these tags after discussions with experts and thought leaders in each practice area. As with case types, we use a combination of automated systems and keyword searches, combined with human oversight, to ensure the accuracy of the tags.

By contrast, Westlaw Edge offers no further specificity beyond NOS codes. Edge lacks the right kind of data to narrow your case list appropriately.

For example, if you wanted to know the average time to termination in *FLSA employment* cases, you'd get a misleading answer if you could only see time to termination in *employment* cases generally.

## Beyond Counting Cases: Analyze Damages, Remedies, and Findings

---

*If you are relying on tools that don't have the required data to find cases like yours, the resulting analysis will be misleading or wrong.*

---

Cases vary widely in terms of outcomes. Most cases settle, but others go to trial, and can generate billions of dollars in damages awards or rulings with far-reaching consequences. You need to be able to find and then focus on these cases in order to uncover meaningful insights.

Lex Machina has deliberately expanded one practice area at a time, because we put in the work and dedication to understand each practice area and its specific features that are of critical importance to practitioners. For each practice area, we have created a unique set of specific damage types, remedies, and findings. These data are largely annotated by human legal experts, and they allow you to have fine-grained control of which kinds of cases you want to investigate.

By contrast, Westlaw Edge does not provide any of this. Like most other “analytics” tools, their offering is an inch deep and a mile wide. While they can claim broad coverage, they don't have the right kind of data about each case. Broad coverage isn't worth much if the data you need just isn't there.

For example, with a few clicks in Lex Machina, you can see all the *product liability* cases in which damages have been awarded or in which a manufacturer won a Finding of *No Breach of Warranty*. Not only can you identify these cases directly, but you can uncover insights in whatever case set interests you. For example, if you're trying to understand a judge's experience with *product liability* cases, Westlaw Edge can only tell you the number of *product liability* cases before the judge. For those same cases, Lex Machina will show you whether or not the judge has ever awarded damages, what kind of specific *product liability* issues the judge has ruled on, and more.

## Counsel Data: Complete and Accurate Data Is Critical to Understanding an Attorney's Track Record

---

*If your analytics are based on bad counsel data, your conclusions will be inaccurate and potentially dangerous.*

---

One of the most exciting insights provided by legal analytics is understanding the litigation track record of attorneys and law firms. No matter what question you're trying to answer about counsel, the answer requires a solid foundation of reliable and accurate data. PACER provides some counsel information at the top of the docket sheet, but this data is often inaccurate and incomplete. And we're not talking about small mistakes or omissions. PACER's counsel data has systemic problems. Over 45% of the cases Lex Machina gets from PACER have incorrect counsel information, and in some districts, 60% of cases are missing some counsel information.

Lex Machina created the unique *Attorney Data Engine* to address PACER's shortcomings and to provide accurate and complete counsel data. The *Attorney Data Engine* draws on three primary sources of counsel data. First, it saves historical snapshots of the PACER docket to correctly attribute cases to law firms, because PACER only shows an attorney's current law firm. Second, the *Attorney Data Engine* parses the signature blocks of certain filings, because often attorneys filing documents in the case are not listed in PACER. Finally, the *Attorney Data Engine* examines pro hac vice (PHV) docket entries to find additional attorneys.

By contrast, Westlaw Edge simply provides the PACER counsel information from the top of the docket sheet, which is often inaccurate and incomplete.

For example, if you wanted to understand Fish & Richardson's experience with *patent* cases in New Jersey, you'd miss over half of their cases if you used Westlaw Edge. If you're relying on bad counsel information, you might underestimate your opponent's experience or overlook ideal counsel for your own case.