

THE IMPACT OF AUTOMATED EXPLORATORY DATA ANALYSIS IN LITIGATION

Speeding up the exploratory data analysis process using flexible automation and consistent reporting allows analysts to deliver analyses quickly while ensuring precise, accurate results.

BY MIKE HOROHO AND ANTHONY KELLY, FTI CONSULTING

In most litigation matters requiring complex data analytics, our consultants receive vast quantities of data with little to no information regarding its contents. The volume of data alone makes a comprehensive manual review nearly impossible, especially under the tight time constraints imposed by many litigation scenarios. While data is often critical in mounting a successful legal strategy, large quantities of data without context creates an impenetrable black box, obscuring any data calculation processes.

Asking the Right Questions

Effective analytics requires a thorough understanding of the data being analyzed and can be achieved through exploratory data analysis (EDA). EDA starts by asking some basic questions of data sets to make an initial assessment of the data's structure and quality to facilitate a plan of attack. These questions include:



- Is there an existing unique record id?
 - Are fields populated with numbers, dates, or alphanumeric characters?
 - What is the distribution of numeric values?
 - How many values exist for categorical fields?
 - How many fields have been left blank?
 - How are multiple data sets related?
- Once this initial assessment has been completed, EDA continues

by asking questions that examine what illogical conditions might exist that provide an indication of the overall quality and reliability of the data. Some hypothetical examples of such questions are:

- Why are there accounting entries with date stamps that precede the existence of the company being analyzed?
- Why are 95 percent of the records blank for a field?
- Why is one accounting entry 1000x the size of the next largest?

- Why are a smaller number of values negative instead of positive for a specific field?

Answering these questions is a simple endeavor when dealing with a small number of fields and records; however, the task becomes much more complex and time-consuming when dealing with millions of records and hundreds of fields. When data analytics in a litigation setting requires precision under accelerated timelines, EDA becomes critical to delivering timely analysis.

Applying the Right Technology

Data analytics experts use computer code that can be deployed on any data set, speeding up the EDA process and enhancing efficiency without sacrificing precision. Custom coding systems can be connected to data sets, automating a large portion of the EDA process. Since data can be stored in many ways, these custom coding systems are flexible enough to access information across a variety of formats including:

- Numerous relational database management systems (Microsoft SQL Server, Oracle Database, MySQL, etc.)
- Text files
- Microsoft Excel workbooks

Once the program is connected to a dataset, the information can be fed into the system in a standard format. Standardization allows these coding systems to run the same set of

analyses regardless of how the data is natively stored.

One best practice is to create a reporting document that can be distributed to both technical and non-technical personnel, thus increasing the number of people who have “eyes on the data.” Review of the standard reporting document reveals important features of a dataset and helps raise critical questions about illogical data conditions. Further, it can quickly reveal columns with less importance, allowing analysts to narrow their focus on the most impactful fields which creates downstream efficiencies during production of the analysis at hand.

Analyzing a Company’s Data

Folding EDA into a broader engagement strategy is critical to the success of projects that require analytics on complex data sets. Without the capabilities of automated EDA, the task of deriving insights across a high number of variables of unknown type and meaning would be challenging and time-consuming.

Deploying automated EDA tools upon acquisition of a company’s data set and review of standard reporting packages should be a required step. Upon review of these reports, it may be discovered that, while the dataset appeared robust at the outset, many fields were not viable for analysis. Through review

of date fields, reports may also reveal large, critical portions of a data set are missing. These issues, while quickly and painlessly resolved with EDA, could invalidate a large amount of work if left undiscovered until later. By reviewing reporting packages with the client, the EDA process can identify and fill gaps in the data, focus the analysis and provide valuable, timely insights.

Speeding up the EDA process using flexible automation and consistent reporting allows analysts to deliver analyses quickly while ensuring precise, accurate results.

Mike Horoho is a Senior Director and Anthony Kelly is a Senior Consultant, both with the Forensic & Litigation Consulting segment of FTI Consulting, Inc. The views expressed herein are those of the author(s) and not necessarily the views of FTI Consulting, Inc., its management, its subsidiaries, its affiliates, or its other professionals. FTI Consulting, Inc., including its subsidiaries and affiliates, is a consulting firm and is not a certified public accounting firm or a law firm.

