Our data analytics team is frequently tasked to use qualitative and quantitative techniques on data to identify common behavior, reveal key facts, and understand patterns; all leveraged to draw vital conclusions related to a matter. Not a small task without a ‘user manual’ to inform you what the data means and how it fits together. Add multiple data sources, like relational databases and NoSQL databases, JSON and CSV files, XMLs, spreadsheets, etc., to the equation and the challenge grows even larger.

So what do you do when your client gives you these huge data sets and asks you to help make their case? Go into puzzle-solving mode.

In a recent wage & labor class action suit, focused on allegations of inconsistent calculations of pay resulting in underpayments and inaccurate wage statements, the dataset for analysis covered millions of records, for more than one-hundred thousand employees, spanning nearly a decade. We were tasked with identifying potential class members, calculating how many people were affected, determining how long the problem existed and the cause, as well as how to fix the issues. With no prior insight into the data, that was a tall order. And a settlement of perhaps hundreds of millions of dollars was at stake.

Determining the common thread between all of them to perform a comprehensive analysis is like solving a puzzle. Once you begin to put the pieces in place, details begin to emerge and the picture starts to reveal itself.

We began by gaining an understanding of what each of the clients earn codes represented. Earn codes are used to compute the individual line items on a paystub. This gave us the ability to validate and re-calculate line items for overtime, double overtime, holiday pay, incentive pay, etc., which is where part of problem existed. With payment calculations in hand, we identified patterns in paystub generation to determine the time period in which inconsistencies occurred. This allowed us to narrow the scope of analysis and ultimately led to the identification of class members. From the class members, we were able to zero-in on geographical locations, employee divisions, employment types, and time of the year (could it be a seasonal problem?) which aided in determining the root cause of the issue.

Other challenges we solved:

- Design a repeatable workflow, despite ever-changing criteria, to identify potential class members
- Develop an algorithm to de-dupe newer data against itself, and prior datasets, before running our analysis
- Create custom reporting, to visualize data patterns and calculations, for production
- Customize calculations if certain criteria are met, or eliminate if criteria were not

In the end, iDS’ team of experts were able to put all the pieces together, thanks to careful step-by-step analysis and a streamlined process. We were able to provide consistent results each time the criteria changed or additional data was collected, determined the cause of the miscalculations, and our more granular paystub calculations showed substantially fewer regular rate of pay miscalculations.
As for the client? iDS’ ability to quickly update and adjust the analyses allowed counsel to understand the potential exposure for their client under various conditions and tailor their arguments for best and worst case scenarios. Our innovative solution to the data issues allowed our client to reduce potential damages by more than 25%.