Insurance Claims Analytics Solution Brief

At the heart of the insurance business is the ability to accurately price policies according to risk. Failing to do so poses serious consequences. If a product is underpriced, insurers are more liable and must recoup losses; if a product is overpriced, the company risks losing market share. It's a delicate balance, and the actuaries responsible must develop complex models leveraging huge volumes of historical claims data, third-party data, and estimation data to get it right.

In today's volatile market, the importance of accurate policies has only increased. Interest rates have been steadily declining for the past eight years, which threatens income from rate-sensitive products and investments and places, and necessitates reliable underwriting profits. To compensate, many insurance companies have sought out new types of data to bolster their claims analytics and offer more personalized policies. Yet these data sources, from geospatial to social media to device data, are often unstructured, leading to challenges in efficiency and accuracy when preparing that data for analytics.

Preparing Data for Claims Analytics

The most jarring hurdle that actuaries face when working with new data sources is that spreadsheet tools like Excel aren't equipped to deal with their complexity or size. Excel is susceptible to stalling under large datasets, and its manual transformations can allow faulty data to go undetected. The more technical data preparation tools, such as SAS or R, are great at building models, but not so much at visually manipulating and preparing data of any shape and scale.

For insurance companies, relying on these tools has begun impacting performance. Actuaries aren't able to move swiftly, nor with confidence, when shoehorning new and larger volumes of data into Excel tools. And for most insurance companies, scaling SAS or R developers is out of the question. In both cases, highly-skilled people are spending the majority of their time preparing data with tools that inhibit collaboration, even though policies often necessitate input from multiple parties.

> "Each key event in the customer cycle, we have to capture that information and make it available immediately. And we can now analyse everything from flood patterns to driving patterns, with sensors on aircraft and cars – big data is about using as much of that information as quickly as possible to make the right decisions about risk."
>
> -Simon Gratton, Chief Data Officer, Information Age

Data Wrangling: The Trifacta Approach

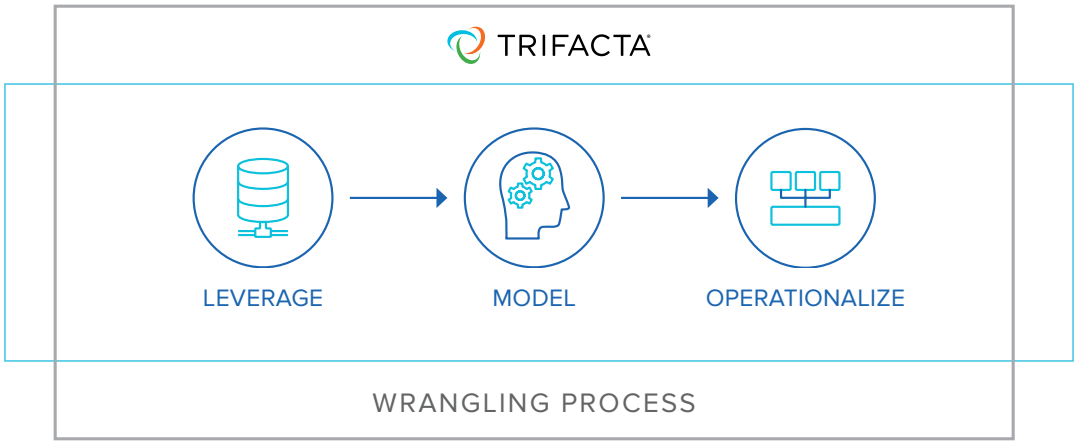
Trifacta's unique approach to data preparation applies the latest techniques in machine learning, data visualization and human-computer interaction, allowing users greater productivity and more autonomy over their data. What was once relegated to those with programming skills or put at risk in Excel with its known limitations around accuracy and scale is now automated and accessible to actuaries and other business-oriented professionals. Trifacta is the self-service solution for modern insurance companies, providing the following key benefits.

Wrangling Data of All Shapes and Sizes

In the past, it was a safe bet to assume that insurance companies dealt with data in structured formats and of reasonable volume, either stored within the enterprise or supplied by a few specialized vendors. However, the last few years have seen an explosion of data in both complexity and volume, providing insurance companies with ample opportunity for new insights, but increased data wrangling challenges.

Trifacta's intuitive interface automatically infers any form of data (structured, semi-structured, unstructured, hierarchical, and more) to offer a visual representation to the raw data so that business professionals can slice and dice the data as needed, allowing for more accurate downstream analytics. For users that typically would have relied on an army of data engineers in order to translate complex data, this is a huge gain.

With regard to data volume, Trifacta **leverages** advanced sampling techniques in order to process any volume of information required to **model**, train and validate results with confidence. With the ability to leverage diverse data sets that have the maximum amount of attributes and years of history, actuaries can now segment down to any level in record time.



Data Quality First

The predictive models that determine pricing or risk accuracy are only as good as the data that feeds them. Analysts devote a huge percentage of their time remediating data inaccuracies, but even still, bad data can often slip through the model undetected.

To offer actuaries a high level of trust in their data, Trifacta automates data quality checks along the steps of the process to expose data anomalies for users to remediate. For example, Trifacta automatically identifies that an attribute is a birthdate and determines whether the date is well-formatted or has missing dates. On top of that, Trifacta examines the information in context, not only verifying that an age should be an integer, in this case, but also points out that an age of 250 years needs attention. Each action taken to visually format, structure, clean and validate the data at scale offers immediate feedback to guide the user with certainty toward an accurate data outcome.

"It's impossible to overstress this: 80% of the work in any data project is in cleaning the data."

-DJ Patil, Former U.S. Chief Data Scientist under Obama's Administration

Flexibility, Collaboration & Faster Delivery

An important exercise in modeling is evaluating various options by enriching the data with numerous attributes to find correlation and causation. With a self-service data preparation technology like Trifacta, this process is accelerated, allowing analysts to rapidly explore several simulations until they discover the most robust outcome. And from there, analysts are just few clicks away to automating and operationalizing this process to integrate it into a repeatable data pipeline.

Data wrangling is rarely a singular responsibility—most organizations will recruit multiple collaborators to work on specific areas of the data based upon their own specialization. With Trifacta, users can work together on entire workflows, recipes and datasets for enhanced collaboration and partnership across the organization. And because Trifacta leverages the power of the most advanced large-scale data processing frameworks, users can collaborate at scale, testing various hypotheses, building models, and applying changes on huge amounts of data in record times.

"Trifacta has allowed us to gain a better understanding of our data in less time and much earlier on in the process. We are extremely excited about its potential, and plan for company-wide adoption of Trifacta as one of our core technologies enabling self-service."

-Mark Bendall, Lead Program Architect

Governance and Auditability

With increased regulations in the insurance industry, such as Solvency II and IFRS 17, companies must be able to demonstrate how a model or report has been calculated at any moment. However, most analysts are buried under Excel spreadsheets and ill-equipped to answer time-sensitive data audit questions.

With Trifacta's interactive data lineage visualizations, analysts can trace data across its lifecycle—from data sourcing, through blending, aggregation, and reporting—to understand downstream dependencies, as well as filter on specific types of dependencies. Not only does this allow insurance companies to address regulatory requirements, but this visibility is also useful for maintaining a model and sharing with others for reuse and best practices.

Conclusion

In today's evolving market, leading insurance companies strive for efficiency and accuracy in claims analytics. With Trifacta, actuaries can build better targeted and more accurately priced products by leveraging an intuitive data preparation tool that drives deeper insights. Trifacta has helped companies reduce the process of preparing data by up to 90%, which allows actuaries to spend less time manually preparing data and more time delivering analysis and insights.