LexisNexis® RISK SOLUTIONS

Insurance

Optimal data quality means better data-driven decisions

Insights from LexisNexis® Risk Solutions about data quality benefits and best practices.

FEBRUARY 2021

Table of Contents

Executive Summary

A wealth of data is available to insurance carriers today, from credit data and police reports, to roof condition information and more. As a direct consequence, carriers are investing millions of dollars into sophisticated analytics, cloud computing, artificial intelligence (AI) and machine learning (ML) to turn that data into insights that will inform decisions at every level of the business. However, despite data's critical role in enabling AI and ML, the quality of the data itself is often an afterthought.

Data quality can vary significantly. Data might contain inaccuracies or unidentified biases, or even be incorrect in nature. For carriers, identifying the best sources of data, merging data sets, integrating the data into workflows and analyzing the data to develop actionable insights is no easy task. And the more data available, the more important it is that carriers invest in ensuring that the data they use is clean, accurate and properly linked.

This report describes the key data quality factors carriers need to consider when they assess existing and new sources of data. It also looks at the quality of data generally used across the industry today, how poor-quality data has an outsized effect on insurance customers and how optimal data quality can help carriers across the entire policy lifecycle. Furthermore, the report suggests several actions carriers can take to avoid bias within their data sets and to address common data quality pitfalls.

Data of optimal quality that has been cleansed, normalized and standardized, among other steps, is a critical basis for the data-driven decisions that carriers make every day—decisions that will have a lasting impact. Yes, it's an investment. But it's one that will affect their profitability and competitiveness in the long term.

Understanding the main factors that impact data quality

Savvy insurance carriers know that data is the key to making good decisions. From accurately assessing risk to pricing policies for profitability, carriers need access to data to achieve positive results across the entire policy lifecycle. They also need quality data to understand historical trends and to make accurate forecasts for the future.

However, decisions based on these analyses are only as good as the underlying data. With new sources of data coming online every day, it can be difficult to know which sources to trust and which offer the most value. That's why we recommend considering the following important factors.

Data quality factor

Questions to ask

Data quality factor	Questions to ask
Source	• Is the data factual or modeled? • Is it interpreted? • Is it manufactured or organic? • Is it reliable? • What was the original need for the data?
Accuracy	• Is the data correct? • Can it be validated or matched against another data set?
Completeness	• How robust is the data? • Missing values?
Coverage	• Are there gaps in the data, such as segments, time or geography? • Is there a way to validate gaps? • Are industry benchmarks available?
Contestable	• Can the data be corrected if necessary?
Legal/compliance	• Who holds the rights to or has ownership of the data? • Does the data meet requirements such as those of the Fair Credit Reporting Act (FCRA) and Drivers Privacy Protection Act (DPPA)?
Historical	• Has the data been consistently sourced over time? • How often is the data updated? • Does the data travel in time? • Can we access the same data available today in archives? • Do archive dates line up with our decision points?
Filters	• Has the data been screened or filtered?
Bias	• Is the data biased in some way? • How were outcomes generated? • Who was responsible for generating outcomes? • Do objectives align with the expected outcomes?
Enrichment	• Is it possible to add appropriate data to provide deeper insights for decision making? • Are you maximizing the ability to link to other data sources?
Accessible	• Does the data provide batch or transactional consistency?
Benchmarks	• Are benchmarks available for drawing comparisons?

An additional point for carriers to consider is whether they're taking advantage of all the data the business has purchased. Sometimes carriers acquire data sets, but don't follow through by incorporating them into their workflow or analysis. Investing in the right data—of optimal quality—can make all the difference for carriers seeking to differentiate themselves from their competitors.

Data quality across the insurance industry today

Carriers have an ever-increasing hunger for data, a growing capacity to process larger files using AI and ML and an evolving sophistication in how they incorporate new types of data across the business. However, in general, as they add more data from different sources, they're seeing their data quality decline rather than improve.

While no data sources are perfect, data from scraped web sites and open-source resources can be of inconsistent quality. Further, disparate sources of data often have no underlying linkages and variations make it difficult to merge or match entities—people, vehicles, businesses and locations—producing duplication and redundancies. At the same time, vendors are offering new sources of data, such as satellite or fixed wing aircraft imagery. While this data may hold appeal to carriers, these data sets aren't all of equal value.

There's an additional challenge for carriers that choose to purchase data attributes rather than raw data. Unless they understand the logic or filters applied to the data to define the attributes, they might not be aware of underlying biases, such as those that relate to geography, ethnicity or socio-economic status. If they use raw data instead, they need to have the skills to filter, clean and cap the data. They also need the ability to manage these changes going forward so that any decisions based on the data can

be reviewed and potentially revised. Modeled data can also present issues if it uses an unknown or black box approach; targets, bias, predictors or proxy variables may exist but be impossible to identify. The more data harvested, the greater the chance that findings could be skewed by poor quality data.

Under pressure to gain a competitive advantage, some data scientists working for carriers might choose to trust data rather than spending the significant time and effort needed to verify its origins and veracity. They might not have the expertise or luxury of time to invest in optimizing third-party data to ensure accuracy and consistency, or to integrate it with their existing data sources. Or they might follow inconsistent data quality practices. In fact, many companies think that hiring talented data scientists by themselves solves all of their analytic challenges. However, it's the combination of real world experience and domain knowledge along with the technical skills that will determine the future leaders.

Without a thoughtful strategy and robust quality assurance processes in place, carriers could be undermining the data-driven decisions they use for underwriting and developing their brand and customer relationships. Just like with human resources and technology capabilities, carriers need to make a serious investment in data or find vendors they can trust to manage it for them.

Poor data quality affects customers directly

It should come as no surprise to carriers that the quality of data they use impacts their customers. Data affects every interaction, from how long customers spend on a call or meeting in person to their general perception of the carrier's brand. Do they have to answer dozens of questions about the characteristics of their property? How difficult is it to update their coverage? How long does it take to handle their claims?

Customers purchasing new coverage expect a fast and seamless experience. At the time of quote, if carriers don't have access to the most current data such as vehicle identification numbers (VINs), prior coverage details and property characteristics—and they have to rely on their customers to supply this information—they risk frustrating customers and losing sales.

Asking customers to provide information impacts the speed and quality of underwriting too. It increases the potential to place individuals in an incorrect rating tier as well as the need to change quoted prices and services when relevant details are uncovered later in the process. This can cause needless friction early in the carrier/ customer relationship when trust is paramount.

Less than optimal data can also give rise to customers being confused about their coverage, dissatisfied with claims handling and frustrated by how long it takes to receive their first payment. All these things can lead to a loss of customer trust and damage to their brands, with customers questioning every decision and disputing claim amounts.

With better data, you can cultivate more positive experiences, build stronger customer relationships and enhance the brand.

Optimal data quality can help carriers at every stage of the policy lifecycle

Data quality can have a huge effect on other areas of the insurance business too. For instance, for marketing and direct mail campaigns, carriers need accurate contact information for customers. If their records are not current, carriers risk wasting valuable marketing spend. Furthermore, inaccurate marketing attributes and models can lead to ill-timed and suboptimal offers and pricing that's not competitive or is unprofitable.

At point of quote, incorrect data entry, such as inaccurate or non-current personally identifiable information (PII) or address information, can hinder a carrier's ability to validate underwriting downstream, which exposes the company to unforeseen risk. Likewise, fraudulent information and inaccurate insurance scores can affect premium calculations adversely.

Quality data from trusted sources can minimize a carrier's exposure to gaps in assessments of insured assets. Home insurance carriers, for example, with accurate Coverage A values and knowledge of roof condition, property characteristics and use of the home for rentals or home-based businesses can make better decisions than those with inaccurate data.

Automobile insurance carriers need correct vehicle mileage and information about advanced driver assistance systems. They need to know if there are any young people in the household who hold a new license and are not on the policy, and if any drivers on the policy have had driving violations. For vehicle fleets, to assess risk accurately, carriers need to know if drivers have driving under the influence (DUI) charges or personal auto claims.

Credit data often has better PII and is more up to date than typical marketing data sources.

9

When providing insurance coverage for businesses, whether commercial property insurance or insurance for business assets, carriers need to know if there has been a history of claims or unsafe conditions, what security measures are in place and the location of the business.

At renewal, carriers need to know if any factors have changed that would affect underwriting risk or policy pricing. They need data that answers questions such as: Have any claims been filed against the policy? Have there been any policy violations? Has the insured asset, such as the home or vehicle, been damaged or in an accident since the original policy was authorized? Has the household composition changed? They need to train employees to gather these changes in household information without unsettling customers because knowing as quickly as possible what has changed or what active insights can be delivered about a customer is key.

In each of these instances, data is available to carriers to help them make decisions. But they must ensure that the data they're using is valid, reliable and accurate.

Carriers that use optimal quality data can:

Boost marketing effectiveness and improve conversion rates.

Protect their brand by making accurate, fair and final decisions.

Keep customer experiences positive and improve retention.

Expedite internal and external processes to increase operating efficiency.

Comply with regulatory requirements with confidence.

Improve risk assessment.

Reduce the number and cost of legal actions.

Best practices for avoiding bias and other data quality pitfalls

Carriers don't have to settle for making long-term decisions based on poor quality data. In fact, if they're willing to make an investment in their data, there are several actions they can take to improve the quality of the data they use starting today.

We recommend carriers:

- **Take a team approach**. Leaders from across all areas of the business should collaborate to ensure that data quality meets the needs of legal, compliance, business analytics, technology, customer service, distribution partners and others.

- **Make sure someone is assigned ownership of the data**. This person or group should be an expert at understanding data, analyzing data and be able to determine the best uses of data.

- **Identify data experts**. Develop relationships with industry data experts to consistently get trusted data.

- **Set up an internal review process before developing data models**. Bias can occur at every stage, from data gathering to processing, so it's important to consider this upfront and seek to mitigate it.

- **Develop a common language and definitions around data**. Make a plan to track data quality. There's potential for data quality to drift or decrease over time.

- **Define consistent rules for new data**. Consider how to bring new data into the business, and how it will be validated by the legal, compliance, analytical and technical teams.

- **Improve data literacy**. This will help raise data quality across the organization and reduce bias. Data quality should be the responsibility of every employee, not just a single team. Often, it's primarily left in the hands of the quality control team— that's a good starting point, but it's better if everyone in the business does their part.

- **Implement data monitoring**. Maintaining quality data is an ongoing process. Define metrics and triage processes ahead of time and set up regular check points and consistent monitoring to ensure quality is preserved.

- **Use data from trusted sources**. Data analytics providers like LexisNexis® Risk Solutions can offer a full range of outstanding data products and powerful data analytics capabilities to help insurance carriers increase efficiency across all lines of business.

Conclusion

Data underpins all decisions and processes across the insurance enterprise. Fortunately, carriers have an increasing amount of data from various sources at their fingertips to inform their business strategies, risk assessments and pricing. However, the data that carriers are currently using is not uniformly of optimal quality and, in some cases, it suffers from some serious shortcomings.

Because poor data quality can have a direct impact on the customer experience and lasting consequences for underwriting and pricing, it's critically important that carriers:

- Understand the impact data quality has on the company's brand.

- Make the use of optimal quality data a priority.

- Develop benchmarking capabilities.

- Compare their organization's data use with that of their competitors.

- Consider working with a partner with data analytics expertise to optimize data quality and consistency, and to leverage quality data sources within the underwriting workflow.

Leading insurance carriers use optimal data quality—to differentiate themselves from competitors, develop more positive interactions with customers, accelerate decision making, build more accurate pricing and achieve positive results across the entire policy lifecycle. At LexisNexis® Risk Solutions, we're here to help.

How LexisNexis Risk Solutions can help carriers improve data quality

Want to optimize the data you use to improve the customer experience and make better decisions? At LexisNexis® Risk Solutions, we have expertise in the areas of data acquisition, data engineering, linking, regulatory review, analytics product management, analytical attribute development, data science and auditing. Leveraging our capabilities and tools, you can focus on your core business of insuring against risk.

LexisNexis can provide contributory and actionable data and insights, including:

- Property characteristics with data cleansing.

- Roof condition data that uses machine learning and artificial intelligence to combine multiple sources, claims clusters for auto and home, weather data and imagery.

- Telematics with normalization, filtering and validation (removing corrupted or incomplete data and duplicate events), and data standardized by format and unit.

- Connected car data, including updates from the vehicle itself, advanced driver assistance systems (ADAS) availability and usage, and vehicle damage.

- Motor vehicle records with standardized data, violations and court records.

- Police reports delivered quickly and in a more ingestible format.

- Data verification to alert you to discrepancies in PII such as invalid and deceased social security numbers, unverified dates of birth and invalid addresses.

In addition, using proprietary linking technology in combination with our unique identifier, LexID®, we're able to resolve, match and manage information for more than 283 million U.S. unique consumer identities, 78 billion records and over 10,000 sources of data. We can score models based on data availability—our model segmentation leverages most predictive factors for each segment, with increased scoreable rates and consistent score outcomes. And we can provide dedicated regulatory resources to ensure regulatory questions are answered preproduction, while we monitor regulatory changes for impacts on products.

Our core business is bringing the highest quality, most robust data products and analytics to the insurance industry and helping carriers integrate them into their workflow. We've been servicing the industry for over 30 years, developing processes, trends and benchmarks and tracking them for preproduction, production and postproduction quality products. Our insurance industry focus means we can implement new model versions as necessary when there are changes in data, new data sources and evolving regulatory requirements, or when there are changes in the industry direction based on market conditions and events.

How can we help you meet your data quality objectives?

Contributors

Jon Atteberry
Director, Data Science
LexisNexis® Risk Solutions

Jon Atteberry is a Director of Data Science at LexisNexis® Risk Solutions, leading a team responsible for developing Acquisition and Retention models and attributes for Auto, Homeowners, Life and Commercial lines.

Jon has over 10 years of experience working in data analysis and analytics, working as Director of Web Traffic and Engagement Analytics at TechTarget, Inc. prior to joining LexisNexis. He holds a Master's degree in Predictive Analytics from Northwestern University as well as an MBA from Georgia State University.

Shewjen Davis
Principal Data Scientist
LexisNexis® Risk Solutions

Shewjen Davis is a Principal Data Scientist at LexisNexis® Risk Solutions. Since joining LexisNexis in March 2016, she innovated the new generation of National Credit File and credit attributes for the insurance industry. Shewjen is regarded as a prominent subject matter expert in consumer credit data, lien and judgment public record data and Insurance Payment History data.

Prior to LexisNexis, Shewjen worked for Equifax, where she was the Sr. Director of Enterprise Analytics & Data Sciences. Shewjen has two decades of work experience in the field of tri-bureau attribution and analytics and model development. Shewjen graduated from the University of Georgia with two degrees, a master's in Statistics and a master's in Computer Science.

Harsh Sharma
Data Scientist
LexisNexis® Risk Solutions

Harsh Sharma is a Data Scientist at LexisNexis® Risk Solutions. In his current role, Harsh develops analytic solutions for the Insurance Claims business. Some of his recent work includes producing actionable data views, benchmarking insurance carriers against the total industry and analyzing the impact of COVID-19 on personal auto claims.

Harsh has over eight years of experience in IT consulting, business analysis and building data-driven solutions in the Insurance domain. Prior to joining LexisNexis® Risk Solutions, Harsh worked at Tata Consultancy Services where he helped European customers transition to new insurance business applications by identifying gaps and performing complete business analysis. Harsh holds a Master's degree in IT with a specialization in Data Science and Analytics from the University of North Carolina at Charlotte.

For more information, call 800.458.9197, or
email insurance.sales@lexisnexisrisk.com

[f] [in] [twitter]

LexisNexis®
RISK SOLUTIONS

About LexisNexis Risk Solutions

LexisNexis® Risk Solutions harnesses the power of data and advanced analytics to provide insights that help businesses and governmental entities reduce risk and improve decisions to benefit people around the globe. We provide data and technology solutions for a wide range of industries including insurance, financial services, healthcare and government. Headquartered in metro Atlanta, Georgia, we have offices throughout the world and are part of RELX (LSE: REL/NYSE: RELX), a global provider of information-based analytics and decision tools for professional and business customers. For more information, please visit www.risk.lexisnexis.com and www.relx.com.