The Use of Predictive Analytics in the Canadian Life Insurance Industry

May 2019

The Use of Predictive Analytics in the Canadian Life Insurance Industry

AUTHORS Jean-Yves Rioux, FCIA, FSA, CERA, Deloitte
Arthur Da Silva, ACIA, FSA, Deloitte
Harrison Jones, ASA, Deloitte
Hadi Saleh, Deloitte

SPONSORS Society of Actuaries
Canadian Institute of Actuaries

CONTENTS

The Use of Predictive Analytics in the Canadian Life Insurance Industry

Introduction

The increasing availability of big data and the use of predictive analytics are changing how insurers have traditionally operated. Both the Canadian Institute of Actuaries (CIA) and the Society of Actuaries (SOA) have identified predictive analytics as a strategic priority.

While property and casualty (P&C) insurers have used analytics for many years and have led the way, other insurers are beginning to introduce predictive analytics into their operations and risk management practices. Although predictive modelling techniques have been used for a few years already, their applications are becoming more widespread and more innovative. In recent years, related knowledge has become an official requirement for actuarial professionals obtaining their Associate designation.

This study is another example of the investment by the CIA and SOA into the field of predictive modelling. The goals are to investigate how the Canadian life insurance industry is utilizing predictive modelling and examine potential areas for enhancement. As a result, this study will focus on providing insights into applications used in the Canadian life insurance industry compared to those of other industries and will briefly touch on trends that are likely to impact how companies do analytics.

In this study, *Canadian life insurance industry* means the operations of insurance companies as they relate to all products sold by Canadian life insurers in Canada, including life, accident and sickness, disability, and annuities. P&C products are excluded from the study.

In this study the terms “predictive modelling” and “predictive analytics” have been used interchangeably to refer to the practice of using statistics to predict outcomes.

Use of this Report

The purpose of this study is to survey and benchmark the use of predictive modelling in the Canadian life insurance industry. Deloitte was engaged by the SOA/CIA to perform the survey and prepare the report. Deloitte is not responsible for the consequences of using this report for any other purposes.

No part of this report may be used or presented without reference to it.

Of note about this report:

- Fifteen entities were selected to participate in the survey. This sample includes a mix of direct writers, reinsurers, and bank-owned insurance subsidiaries which represented a large proportion of the life and health insurers and reinsurers.
- The preparer has assumed that the responses provided during the survey were accurate. The results presented in this study are based on what participants willingly shared, and in some instances on what the participants estimated when they did not have an exact answer to a question.
- For the few questions where some participants did not answer, the statistics were adjusted so that the percentages are calculated based on the smaller number of survey participants.
- The summary statistics such as averages are based on a count of participants as opposed to being scaled using a liability or premium metric. Both percentages and assessment rating were averaged in such a manner.

Executive Summary

This research was prepared by surveying the current practices of the Canadian life insurance industry (completed via interviews of 15 direct writers, reinsurers, and bank-owned insurance subsidiaries) and supplementing these findings by researching predictive modelling applications used outside of the Canadian life insurance industry (the Canadian P&C insurance, non-Canadian insurance, and non-insurance industries).

We developed a framework to categorize and group applications of a similar nature that spanned different operational aspects or functional areas of insurers. This allowed for clearer discussions and reporting on applications.

Our survey asked participants to assess the value of their analytics initiatives, as well as the effort to implement them. Respondents reported that the highest-value applications related to simplified underwriting (UW), fraud detection, targeted marketing, and inforce retention management. The second layer of applications seen as moderately valuable included new customer segmentation, application triage/accelerated/automated UW, cross-selling/up-selling, pricing, inforce segmentation, lapse experience studies, operational efficiencies, and claims management optimization.

The research identified applications used outside of the Canadian life insurance industry that could have relevance for the industry, and those included: optimizing marketing, product design, cognitive UW, automated processing of data, improved pricing granularity, operational processing, distribution strategies (including recruiting), robo-advisory sales, improved customer experience, and workforce analytics.

In addition, we also identified broad trends impacting analytics applications and the related processes, and identified the following items that insurers should keep an eye on:

- Increased privacy regulations/transparency requirements/ethics dilemma
- The blurring line between automation and modelling
- Move to more use of cognitive aspects (e.g., image interpretation, voice sentiment detection)
- Virtual assistants
- Focus on customer experience

Based on the survey responses, improvements can be made to data used for predictive analytics with regard to centralization, quality of data, and data access procedures. Additionally, new technologies (for use of augmenting data) have not yet been significantly leveraged by the survey participants, and this is an area that insurers are starting to tap into.

We see from the responses that there is progress to be made with regard to governance and policies around standardization. There appears to be a gap between medium/small survey participants and large participants with respect to governance around data (specifically data updates, data accuracy, and standards). It is likely that the same gap applies to these size groups for rest of the industry. Similarly, there is a gap with respect to governance around modelling (specifically model validation, code review, and version control). Only five out of 15 survey respondents indicated that they have software upgrade policies, which is significantly lower than perhaps warranted.

Although 78% of final decisions surrounding predictive analytics are made by C-suite or higher, only one-fifth of boards, CEOs and EVPs are involved in decisions surrounding predictive analytics (this appears low to us and we would expect more involvement in predictive analytics that is used in decision making).

When analytics is not given highest priority, the key reasons given by the participants include data quality, challenges in aggregating data, and other competing business priorities.

Not surprisingly, the large survey participants are, on average, performing more projects and have more projects planned than medium or small participants.

As expected, large respondents place a much higher emphasis on full-time equivalent (FTE) assignment and tend to have performed analytics for a much longer period than the respondents in other size groups.

The composition of the analytics teams is such that overall close to 50% of the predictive analytics resources are “business experts” versus data architects/statisticians. Of these business experts, 75% are actuaries, so there is a significant involvement of actuaries in analytics.

In terms of capabilities, large and medium respondents assessed their own capabilities for business knowledge as lower than technological capabilities and/or statistics/analytics knowledge. This may indicate a need to invest in training technicians about the business, or hiring such people with analytics capabilities. Small respondents assessed their technological capabilities as lower than their business knowledge, which indicates the need to invest in technology and related training.

Strengths and weaknesses identified

Most survey participants do not have a centralized data repository (either fully or partly centralized), and many participants, particularly medium and small ones, do not rate their data as particularly complete or accurate. Having a comprehensive data repository with clean, complete, and accurate data is important in unlocking the power of analytics, particularly across functional domains (e.g., marketing, risk, valuation).

Most respondents have explored or plan on exploring the use of predictive analytics in experience studies. This process is data rich and the data systems are already in place, so the process is a natural candidate for analytics work.

Most respondents, particularly outside of the larger organizations, have many gaps in the types and breadth of applications, particularly for marketing, retention management, distribution–client matching, and even in accelerated/automated UW.

Most survey participants have not thought about use of predictive analytics to improve internal operations (e.g., workforce analytics or use of Natural Language Processing/Natural Language Generation (NLP/NLG) to speed up processes and reduce error rates).

Aside from large respondents and a couple of medium respondents, analytics has not been placed at the highest priority. Many survey participants do not have strong support from leadership sponsoring analytics initiatives in their companies.

Many respondents indicated that they have not thought about a standardized development environment with automated testing procedures. Only half the survey participants have explored machine learning or deep learning techniques.

Survey participants have been struggling to hire and retain analytics experts that fit business needs. Indeed, most analytics experts (e.g., statisticians and data scientists) have strong technical ability but either do not have the appropriate business knowledge or the ability to communicate ideas to non-technical audiences across the business. Actuaries traditionally have strong technical abilities and understanding of the business, and are able to communicate effectively across the organization. They should therefore be well positioned to expand their technical knowledge to cover predictive analytics and their applications.

Methodology

The underlying work for this engagement followed two main streams:

- A survey of the current practices of the Canadian life insurance industry
- Research on predictive modelling applications outside of the Canadian life insurance industry

The methodology followed in each of those two streams is described hereunder.

1. Survey

1.1 Survey population

Fifteen entities were selected to participate in the survey. This sample includes a mix of direct writers, reinsurers, and bank-owned insurance subsidiaries which represented a large proportion of the life and health insurers and reinsurers. Notable absences from this survey include small insurers and Insurtech start-ups.

Survey participants were categorized in three size groups based on gross written premium information. This resulted in categorizing the participants as three large, seven medium, and five small companies. Where appropriate, we comment on any difference in response for the different groups.

1.2 Sampling process

Survey participants were selected based on overall size of life and health insurers and reinsurers based on total actuarial liabilities for Canadian non-P&C business as reported to OSFI and AMF. This was influenced and approved by the members of the oversight group for this project.

1.3 Survey design

The survey questions were structured around the “DELTA” framework popularized by Harvard professor Thomas Davenport, a framework which measures an organization’s culture, readiness, and maturity as they relate to analytics. DELTA is an acronym whose letters cover the Data, Enterprise, Leadership, Targets, and Analysts aspects. Using this framework ensured the survey covered all relevant aspects.

The survey was administered through a series of interviews with participants. This approach was chosen because not everyone describes analytics using the same language or terminology, and to allow the opportunity to ask for clarifications.

The survey was performed mostly throughout May 2018. The questionnaire is included in Appendix A of this report.

2. Research on predictive modelling applications outside of the Canadian life insurance industry

The second major component of this study was to provide examples of predictive modelling applications that were used in other industries, including the P&C insurance industry, and in life insurance companies outside of Canada. This was accomplished by reviewing existing literature and interviewing subject-matter experts.

We will also discuss identified emerging trends.

3. Structure of the report

The aims of this report are to investigate how the Canadian life insurance industry is utilizing predictive modelling in Canada and examine potential areas for enhancement. The report is therefore structured into two broad sections:

- Findings relating to predictive modelling applications (from both the survey and the research)
- Survey findings

Findings Relating to Predictive Analytics Applications

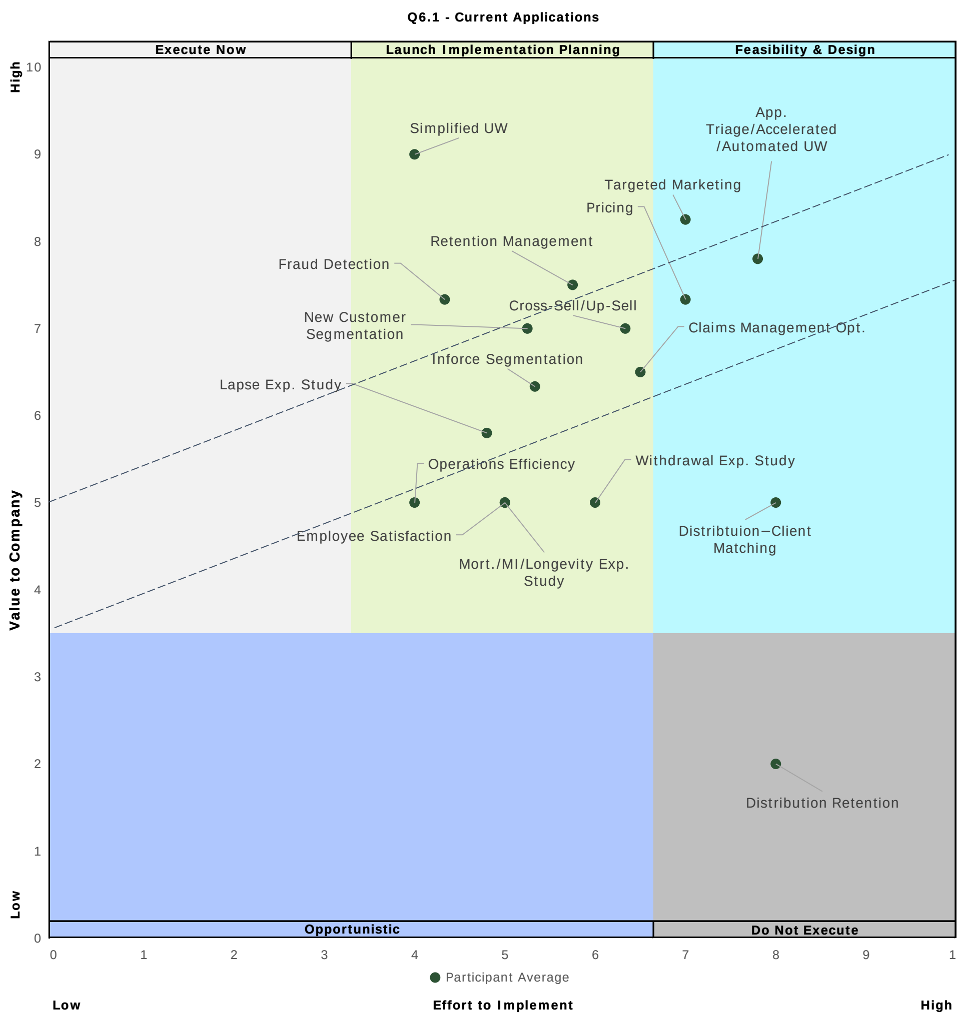
Participants assessed the value of their analytics initiatives as well as the effort to implement them, and concluded that many of those yielded great value in relation to the effort required to implement.

In this section we investigate how the survey participants are utilizing/will utilize predictive modelling, now and in the future, using the findings from the survey. We also examine potential areas for enhancement from some additional research into emerging applications from P&C insurers and other industries.

4. Findings from the survey

4.1 Current applications

The following chart illustrates the current applications as identified by the participants on an effort/value grid. The effort score is illustrated on the x-axis, and therefore applications showing in the left portion of the grid are easier to implement. The value score is illustrated on the y-axis so that applications listed on the top portion of the grid are more valuable to execute. The effort and value scores were based on the judgement of the respondents.



In order to rank and prioritize applications, we assigned a combined value score "V" and a combined effort score "E" based on the average of the respondents' assessments. On the premise that value was more important than effort, and that high-value applications would be pursued despite the effort required to implement, we gave twice the weight to the value score than to the effort score. This allowed us to develop a prioritization framework based on the relative score "2V-E". This is illustrated in the chart above as two lines to separate the applications and assign them to higher (above the top line), medium (between the lines), or lower (beneath the lower line) priority groups. They are listed hereunder.

Higher-priority applications:

Score	Application	Description
2V-E=14 V=9, E=4	Simplified UW	Use of predictive models to simplify the UW application process (minimal UW is performed)
2V-E=10.3 V=7.3, E=4.3	Fraud detection	Develop predictive models to identify potential sources of fraud in activities in certain distribution partners, product portfolios, or specific transactions or policies
2V-E=9.5 V=8.3, E=7	Targeted marketing	Supplement existing internal efforts to identify ideal prospects for new insurance sales
2V-E=9.3 V=7.5, E=5.8	Retention management	Differentiate surrender propensity for individual contracts and identify next best offer

Medium-priority applications:

Score	Application	Description
2V-E=8.8 V=7, E=5.3	New customer segmentation	Profile potential customer base and determine channel preference, ability to cross-sell, product affinity, etc., often by using third-party data
2V-E=7.8 V=7.8, E=7.8	Application triage/accelerated/ automated UW	Solution designed to perform all or some of the screening functions traditionally completed by underwriters, and thus seeks to reduce the manpower, time, and/or data necessary to underwrite an insurance application
2V-E=7.7 V=7, E=6.3	Cross-sell/up-sell	Identify optimal inforce customers for targeted offers for additional life insurance coverage or additional products
2V-E=7.7 V=7.3, E=7	Pricing	Use predictive models to discover and utilize new pricing variables
2V-E=7.3 V=6.3, E=5.3)	Inforce segmentation	Profile inforce customer base and determine channel preference, ability to cross-sell, product affinity, etc., often by using third-party data
2V-E=6.8 V=5.8, E=4.8	Lapse experience study	Use predictive models to understand and model lapse behaviors
2V-E=6.5 V=6.5, E=6.5	Claims management optimization	Case management optimization of resources and intervention efforts, typically done for disability claims

Lower-priority applications:

Score	Application	Description
2V-E=6 V=5, E=4	Operations efficiency	Streamlining of insurance processes (new business, UW, actuarial) reducing or eliminating inefficient procedures
2V-E=5 V=5, E=5	Employee satisfaction	Using predictive models to understand employee satisfaction
2V-E=5 V=5, E=5	Mortality/mortality improvement/longevity experience study	Use predictive models to understand and model mortality and mortality improvement experience
2V-E=4 V=5, E=6	Withdrawal experience study	Use predictive models to understand and model withdrawal behaviors
2V-E=2 V=5, E=8	Distribution–client matching	Use predictive models to develop system-based rules to match compatible distribution partners with new or orphaned clients or with products to sell
2V-E=-4 V=2, E=8	Distribution retention	Identify opportunities to enhance retention of agents

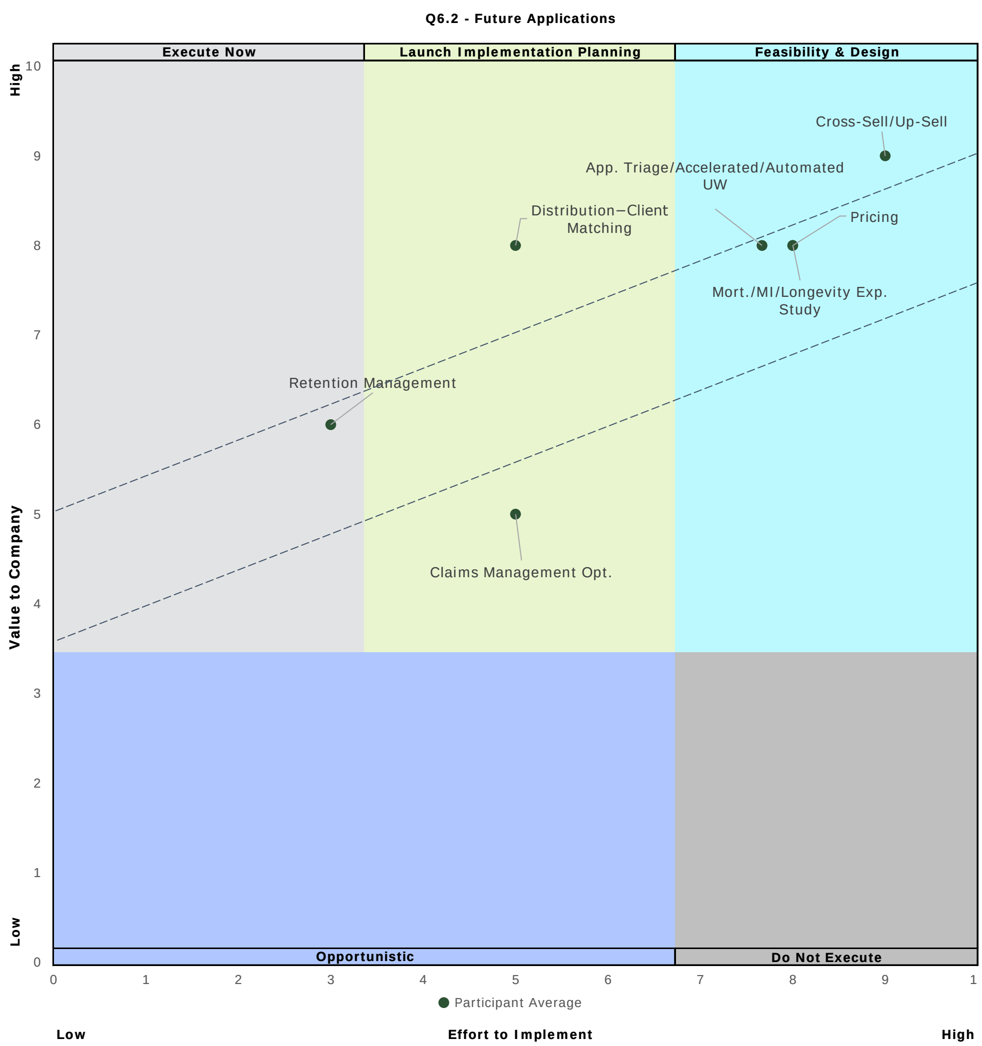
The number of times an application was identified as a current application by survey participants is as follows:

Priority	Application	Count
Higher	Simplified UW	1
	Fraud detection	6
	Targeted marketing	4
	Retention management	4
Medium	New customer segmentation	4
	Application triage/ accelerated/automated UW	5
	Cross-sell/up-sell	6
	Pricing	3
	Inforce segmentation	3
	Lapse experience study	5
	Claims management optimization	2
Lower	Operations efficiency	3
	Employee satisfaction	1
	Mortality/mortality improvement/ longevity experience study	7
	Withdrawal experience study	1
	Distribution–client matching	1
	Distribution retention	1

Perhaps it comes as a surprise that claims management was given a medium priority and is reported as being done by only two respondents, as in our experience this is a high-candidate application for analytics. We suspect that more is currently being executed in this area and that the personnel we reached out to were not aware of these initiatives, a view which is consistent with the actual experience and industry knowledge of the members of the project oversight group.

4.2 Future applications

The following chart illustrates the future applications as identified by the participants on an effort/value grid.



Using the formerly described methodology we identified the following applications by priority.

Higher-priority applications:

Score	Application	Description
2V-E=11 V=8, E=5	Distribution–client matching	Use predictive models to develop system-based rules to match compatible distribution partners with new or orphaned clients or with products to sell

2V-E=9 V=9, E=9	Cross-sell/up-sell	Identify optimal inforce customers for targeted offers for additional life insurance coverage or additional products
2V-E=9 V=6, E=3	Retention management	Differentiate surrender propensity for individual contracts and identify next best offer

Medium-priority applications:

Score	Application	Description
2V-E=8.3 V=8, E=7.7	Application triage/accelerated/ automated UW	Solution designed to perform all or some of the screening functions traditionally completed by underwriters, and thus seeks to reduce the manpower, time, and/or data necessary to underwrite an insurance application
2V-E=8 V=8, E=8	Mortality/mortality improvement/ longevity experience study	Use predictive models to understand and model mortality and mortality improvement experience
2V-E=8 V=8, E=8	Pricing	Use predictive models to discover and utilize new pricing variables

Lower-priority applications:

Score	Application	Description
2V-E=5 V=5, E=5	Claims management optimization	Case management optimization of resources and intervention efforts, typically done for disability claims

The number of times an application was identified as a future application by survey participants is as follows:

Priority	Application	Count
Higher	Distribution–client matching	1
	Cross-sell/up-sell	1
	Retention management	1
Medium	Application triage/accelerated/ automated UW	3
	Mortality/mortality improvement/ longevity experience study	1
	Pricing	1
Lower	Claims management optimization	1

5. Research

As the identification of applications using analytics was a key focus of this engagement we supplemented the summary of the survey results with research into what applications are used outside Canada and outside of the life insurance industry, and with interviews with analytics experts. To help explain the various applications found from our literature review and interviews, it was useful to develop a broad category framework so that each application can be associated with a broader category. In practice, these applications may cross multiple categories. Applications that are not covered by Canadian life insurance, either currently or in the near future, are expanded further where needed. These findings are detailed hereunder.

5.1 Sales and marketing

Using predictive models for sales and marketing is not unique to the insurance industry. It is common practice to use modelling to understand and predict customer behavior (using clustering techniques), qualify and prioritize sales leads, offer the tailored products/service recommendations, target the right customers, and overall inform marketing strategies.

The range of applications identified in the research for this category includes:

- Customer segmentation/profiling – identification of individual characteristics correlated with purchase decision (such as product affinity, propensity to buy, propensity to renew, health profiles, and lifestyle scores)
- Cross-selling additional products or up-selling additional coverage at the time of sale to a new customer
- Target marketing/lead generation and prioritization
- Customer Lifetime Value (CLV)
- Marketing campaign effectiveness, marketing spend optimization, marketing strategy development
- Product development – development of new products that take advantage of vast amounts of data (such as usage-based insurance, and telematics), predicting customer product demand based on internal and external data
- Tailored product recommendations

Of the above, the following applications were not currently used in a predictive nature, nor planned to be used in the near future, by survey participants:

Customer Lifetime Value

Only one of the survey participants indicated using CLV applications. CLV is a metric that calculates the value that a customer brings to a business, either through an increase in revenue or a decrease in operational costs. One key distinction for the CLV, versus other metrics, is that it accounts for the potential value that a customer brings to the business over their entire lifetime. By using CLV, an insurer can structure its business decisions to target highly profitable customers across numerous lines of business.

Marketing effectiveness and optimization

Numerous measures can be undertaken to increase the effectiveness of marketing campaigns, messaging, and spending. Customer feedback based on support calls and social media posts can be used to develop new marketing campaigns using text analytics and NLP. This data can be used to investigate product appeal, understand customer satisfaction, and tweak the campaigns used to promote products.

Product development

Predictive analytics can be used to analyze internal data (such as customer digital interactions, voice calls, emails, and existing sales data) and external data (such as social media and demographic data) to predict what products and features the market will value and desire in the future. A related example includes Netflix developing success criteria for its originally produced content based on usage data, metadata, and ecosystem data, such as what is trending in the news.

Tailored product recommendations

Customized product recommendations involve targeting the right customers at the right time with the right content. This process leverages both internal data (such as purchase data, customer interaction data, and digital interaction data) and external data (such as lifestyle data, purchasing behaviours, and social media) to understand purchase behaviours in order to match product and service recommendations to a customer’s predicted needs.

5.2 Underwriting

The repetitive and rules-based framework of most UW policies is an option for predictive modelling. The business can create an algorithm which provides a prediction to be used to eliminate unnecessary testing based on customer-provided answers, speed up the issuance process, and raise issue limits without a testing requirement. Analytics can also be used to enhance the information with third-party data such as the Medical Information Bureau database for smoker indications. If the predictive model is built well, the enhanced speed and accuracy is seen as having a high return on investment. As Canadian discrimination law is in effect potential issues exist in regulatory restrictions (around the data that is used for pricing/UW purposes) and reputational risk (when there is public push-back on the UW practices).

The range of applications identified in the research for this category includes:

- Improved risk segmentation – stretching criteria for selecting an UW class (i.e., table shave), discovering new predictors in health using new and alternative data sources (e.g., 23andMe genetic testing)
- Simplified UW/application process streamlining
- Application triage and UW resource prioritization
- Accelerated UW – reducing intrusive UW requirements, automated decisions on UW requirements for each applicant, smoker propensity models (predicting an applicant’s smoking status without fluid testing), straight-through processing
- Optimize UW – raise limits for issuance without tests
- UW resource prioritization
- Cognitive automation to replicate UW decisions
- Automated processing of UW data (e.g., using computer vision/NLP)

Of the above, the following applications were not used currently, nor planned to be used in the near future, by survey participants:

Cognitive automation to replicate UW decisions

AI can be trained to detect patterns in human decision making such that it is able to suggest or make decisions that humans would normally make in simpler scenarios. Such technology could be used to relieve human decision making in these scenarios to allow the people to focus on more complex decision making requiring more complicated human judgement. It could also be used to supplement such a process.

Automated processing of UW data

Text analysis, NLP, and image analysis can ease the burden of extracting information from unstructured UW data (e.g., paper applications, attending physician statements) and export it to a structured format to aid in speeding up the UW process.

5.3 Pricing, experience studies, and reserving

Pricing – Future revenues of an insurance company are directly impacted by product pricing and business volumes, so getting the product price right is of utmost importance. Offering a more granular and specific price can improve the financial stability of the business while maintaining competitive rates with the industry. One major barrier preventing the widespread use of these predictive models is regulatory restrictions. Understandably, regulators are not comfortable with applying an unexplainable model to pricing consumers.

Potential applications in this category for pricing include:

- Improved pricing accuracy and/or granularity through additional pricing variable identification and tier scoring
- Decrease turnaround time in pricing process using text mining/NLP and NLG

Of the above, the following applications were not currently used, nor planned to be used in the near future, by survey participants:

Improved pricing accuracy and granularity

Integrating additional variables, including a range of non-traditional variables, can help with increased pricing granularity, with the goal of more accurate pricing. Non-traditional variables can include lifestyle information (e.g., recreational hobbies, online shopping behaviour) from third-party vendors, credit score (based on payment pattern information, account history, bankruptcies/liens, collections), or proxies to credit score (e.g., telecom payment history). Furthermore, pricing accuracy can be improved by introducing variable interactions to account for correlations between variables.

Decrease turnaround time in pricing process using text mining/NLP and NLG

Turnaround time for pricing can be decreased using modern technologies such as text mining and NLP to retrieve key information from unstructured text documents (such as product information and facultative reinsurance documentation). NLG can be used to intelligently automate the generation of pricing documentation such as memos and approval emails.

Experience studies and reserving – The experience study process is data rich and has been executed for many years, so the expertise is in place. It is therefore understandable that there has been a recent surge in using predictive models in this area to determine whether they can provide more accurate predictions than traditional methods. In cases where the predictive model is optimal, new assumptions can be applied to reserves based on a predictive model.

Potential applications in this category for experience study and reserving include:

- Identifying additional experience drivers (e.g., additional UW data, external party data) or variables correlating with existing drivers of claim events (e.g., risk scores linked to prescription drug usage, lifestyle scores)
- Improve estimates for existing experience drivers (e.g., mortality, policyholder behaviour)
- Granular reserve estimates (e.g., at member level) and economic reserves and capital estimation

Of the above, the following applications were not currently used, nor planned to be used in the near future, by survey participants:

Granular reserve and capital estimation

Reserves that have been traditionally been estimated with low levels of granularity.

5.4 Distribution management

Predictive analytics can be used to optimize the sales force by better understanding behaviours and success factors.

Potential applications in this category include:

- Advisor performance monitoring, management, and retention
- New agent hiring and selection
- Distribution–client matching/optimization
- Distribution strategy
- Robo-advisory

Some reference was made to distribution analytics that appeared generic, so it is possible that some of the more specific applications identified here have been implemented by survey participants, but because this is not widely used they are described here. Of the above, the following applications were not currently used, nor planned to be used in the near future, by survey participants:

New agent hiring/selection

Predictive analytics can be used to identify traits of highly productive advisors and then be used in the recruitment of new advisors.

Distribution strategy

Insurers can leverage additional data, both internal and external, to deepen customer knowledge and open new distribution channels (e.g., WeChat in China).

Robo-advisory

An alternative distribution method to human advisors is known as “robo-advisors”. These advisors can interact with customers in natural language and can make recommendations based on customer needs in simpler circumstances.

5.5 Inforce management

Predictive analytics can be used to improve customer experience, identify the most valuable customers, and develop initiatives around and optimize the sales targeted to those.

Potential applications in this category include:

- Inforce client segmentation – differentiating inforce based on health/risk scores in place of UW, lifestyle, and product affinity
- Retention management – design retention strategies (e.g., targeted conversion, targeted loyalty programs, next best offer) to target high-value customers with a propensity to surrender
- Post-lapse inforce management
- Changes to non-guaranteed (adjustable) policy features
- Cross-sell/targeted product offers, up-sell of additional coverage (this is discussed in the Sales and Marketing section)
- Improve customer experience

Of the above, the following applications were not currently used, nor planned to be used in the near future, by survey participants:

Post-lapse inforce management

For customers that have already lapsed, data on the reason for lapse can be used to predict/offer other products or coverages that the lost customer may be interested in.

Changes to non-guaranteed (adjustable) policy features

Predictive modelling techniques could be used for adjustable policy features (such as adjustable premiums, and crediting rates) using additional data sources.

Improve customer experience

Provide tailored services and relevant information to customers informed by predictive models.

5.6 Claims management and fraud detection

Claims management – The claims management process can be quite involved, and part of this process can be repetitive. This means that there is room for automation in the form of a predictive model. The goal of such models is often to process and manage incoming claims to the business. Although the human element cannot completely be eliminated, models can be used to support the process and enhance the information available. Modern machine learning applications can improve content recognition and prioritize more intelligently, and even increase customer satisfaction by significantly reducing response time.

Potential applications in this category for claims management include:

- Claims management triage/preliminary processing triggering different levels of claims management, claim resource prioritization, and what to do/what to stop doing with the use of claims scores (e.g. disability claims scoring)
- Claims segmentation – determining type of claim and action required
- Operational efficiencies in claims processing

Of the above, the following applications were not currently used, nor planned to be used in the near future, by survey participants:

Operational efficiencies in claims processing

NLP can be used to extract key attributes (e.g., nature of injury, occupation) from unstructured claims notes that can be analyzed to improve data for claims processing and used for predictive modelling to proactively identify claimants that are at risk of prolonged claims.

Also, automated claims processing can be done via virtual assistants (e.g., Trov, Aetna's Ann) that interact in natural language. For example, a recommended claim template can be provided for claim submission based on a customer inquiry.

Fraud detection – Insurance fraud has become a popular predictive modelling topic and application. Although it is difficult to know the magnitude of the impact of fraud for each insurer, there is no doubt that reducing fraud will have a significant and positive impact to both the business and consumers. The precise implementation of a fraud detection model can be quite complex and is reliant on a strong underlying set of data.

Potential applications in this category for fraud include:

- Application fraud – detection using smoker propensity models to reduce false declarations such as that of non-smoker to obtain a discount
- Proactive responses to fraudulent applications
- Claims fraud – fabricated death claims, exaggerated or false disability claims, and other frauds which can be investigated using external data sources (e.g., social media, weather)
- Operational efficiencies in fraud detection

Of the above, the following applications were not currently used, nor planned to be used in the near future, by survey participants:

Proactive responses to fraudulent applications

Application fraud (e.g., fake applications for agent rewards) can be proactively investigated prior to claim using pattern and anomaly detection (e.g., unusual volume of applications, matches to known fraud lists).

Operational efficiencies in fraud detection

Efficiencies can be gained in the investigation of structured data. For example, Lemonade uses almost 20 machine learning algorithms to assist with fraud detection for incoming claims. The algorithms are based on image and video data and a response is provided within minutes.

5.7 Other

Predictive modelling can be used to support the operations more generally. This is equally true for insurance and non-insurance companies. Most of these other applications have not been considered by survey participants. Applications span workforce analytics, risk management, investment/trading, and process improvements.

Potential applications in this category include:

Workforce analytics

Predictive analytics can be used to manage a company's own employees, with methods such as:

- Retaining employees with the right retention offers
- Predicting drivers of employee engagement and impact of human resources policies on performance
- Determining where to invest in real estate (office locations), by predicting where future talent is likely to reside and growth potential of such talent in various locations
- Improve recruiting process by integrating social media information to identify better fit and determining drivers of high employee performance

Risk management

Predictive analytics can be used in risk management for companies under various contexts:

- Safety – determining drivers of workers' safety and health issues and designing policies to manage those risks
- Brand risk – identify and monitor risk to brand by extracting systematic trends from high amounts of noise/chatter in unstructured data sets like social media
- Credit risk modelling – identifying drivers of credit in credit risk in investment securities, using a variety of alternative data (instead of credit scores and internal payment data), such as including telco and transactional information, social media activity, and non-credit bill payment patterns to predict their clients' repayment abilities
- Proactive data quality assessment – proactively identify anomalies and escalate those to modellers and/or data providers; identifying data gaps and data inconsistencies

Investments/trading

Predictive analytics can be used in investments and trading for companies under various contexts:

- Find signals for higher (and uncorrelated) returns for portfolio construction
- Algorithmic trading for purposes of asset-liability management or hedging to reduce head count as seen in portfolio management companies (e.g., BlackRock)

Process improvements

Predictive analytics and AI have been used to modify processes to improve customer experience and operations efficiency, which can be applied to insurers in the following ways:

- Use voice-to-insight algorithms to more quickly diagnose customer needs
- Mining feedback messages from customers using NLP to better predict which sellers will create poor experiences (e.g., eBay/Airbnb)
- Improved website search and information presentation – use AI and machine learning combined with NLP to predict a customer's intent or segment customers into groups that can be better served with tailored information. Virtual shopping assistants can be used to recommend products (e.g., North Face's usage of IBM Watson)
- NLP and text analytics to improve customer service over email or 24-hour chatbots/virtual assistants that can interact using natural language. For example, service departments can use automatic matching of answers to customers' inquiries based on past customers' questions or historical service data
- Speech analytics on customer calls to analyze sentiment and key concerns and recommend next best actions
- Engaging customers' understanding of own risk better (e.g., MyFitnessPal), and proactive advice on healthy behaviours (e.g., Under Armour's UA Record)
- Intelligent automation of repetitive processes such as HR and IT with a goal to reduce the need for human involvement and reduce error rates
- Automation of the analysis of unstructured data (e.g., image, video, scanned documents)
- Contract management and negotiation – data extraction and collection to identify specific clauses and conditions using textual analytics technology (e.g., JP Morgan using its COIN software for commercial loan agreements)
- Forecasting – new business plans, predicting demand for internal resources (people, process, technology)

We also researched the broad trends impacting analytics applications and the related processes, and identified the following as items that insurers should keep an eye on:

- Increased privacy regulations/transparency requirements/ethics dilemma
- The blurring line between automation and modelling where models update themselves automatically
- The move to more use of cognitive aspects (e.g., image interpretation, voice sentiment detection)
- Virtual assistants
- Focus on customer experience

Summary of Findings Relating to Predictive Analytics Applications

Survey participants have assessed the value of their analytics initiatives as well as the effort to implement them and found that many of those applications yielded great value in relation to the effort required to implement.

An application categorization framework was developed to group applications of similar nature and covered the following categories:

- *Sales and marketing*
- *UW*
- *Pricing, experience studies, and reserving*
- *Distribution management*
- *Inforce management*
- *Claims management and fraud detection*
- *Other*

The Canadian life insurance companies (the survey participants) currently value highly the following current applications:

- *[UW] Simplified UW*
- *[Claims management and fraud detection] Fraud detection*
- *[Sales and marketing] Targeted marketing*
- *[Inforce management] Retention management*

In addition they currently value moderately the following current applications:

- *[UW] Accelerated/automated UW*
- *[Sales and marketing] Cross-sell/up-sell*
- *[Pricing, experience studies and reserving] Pricing*
- *[Inforce management] Inforce segmentation*
- *[Sales and marketing] New customer segmentation*
- *[Pricing, experience studies, and reserving] Lapse experience study*

The main applications that may be worth investigating executed by other companies than Canadian life insurers include:

- *[Sales and marketing] CLV (although mentioned by one respondent)*
- *[Sales and marketing] Marketing effectiveness and optimization*
- *[Sales and marketing] Product development*
- *[Sales and marketing] Tailored product recommendations*
- *[UW] Cognitive automation to replicate UW decisions*
- *[UW] Automated processing of UW data*
- *[Pricing, experience study, and reserving] Improved pricing accuracy and granularity*
- *[Pricing, experience study, and reserving] Dynamic price optimization*
- *[Pricing, experience study, and reserving] Decrease turnaround time in process using text mining/NLP and NLG*

- *[Pricing, experience study, and reserving] Granular reserve and capital estimation*
- *[Distribution management] New agent hiring/selection*
- *[Distribution management] Distribution strategy*
- *[Distribution management] Robo-advisory*
- *[Inforce management] Post-lapse inforce management*
- *[Inforce management] Changes to non-guaranteed (adjustable) policy features*
- *[Inforce management] Improve customer experience*
- *[Claims management and fraud detection] Claims segmentation*
- *[Claims management and fraud detection] Operational efficiencies in claims processing*
- *[Claims management and fraud detection] Proactive responses to fraudulent applications*
- *[Claims management and fraud detection] Operational efficiencies in fraud detection*
- *[Other] Workforce analytics (e.g., employee retention, employee engagement, hiring using third-party data)*
- *[Other] Risk management (e.g., drivers of workers' safety, brand risk monitoring, proactive data assessment)*
- *[Other] Investment/trading (e.g., identifying drivers of performance, algorithmic trading)*
- *[Other] Process improvements (e.g., voice processing, NLP processing)*

We also researched the broad trends impacting analytics applications and the related processes and identified the following as being items that insurers should keep an eye on:

- *Increased privacy regulations/transparency requirements/ethics dilemma*
- *Blurring line between automation and modelling*
- *Move to more use of cognitive aspects (e.g., image interpretation, voice sentiment detection)*
- *Virtual assistants*
- *Focus on customer experience*

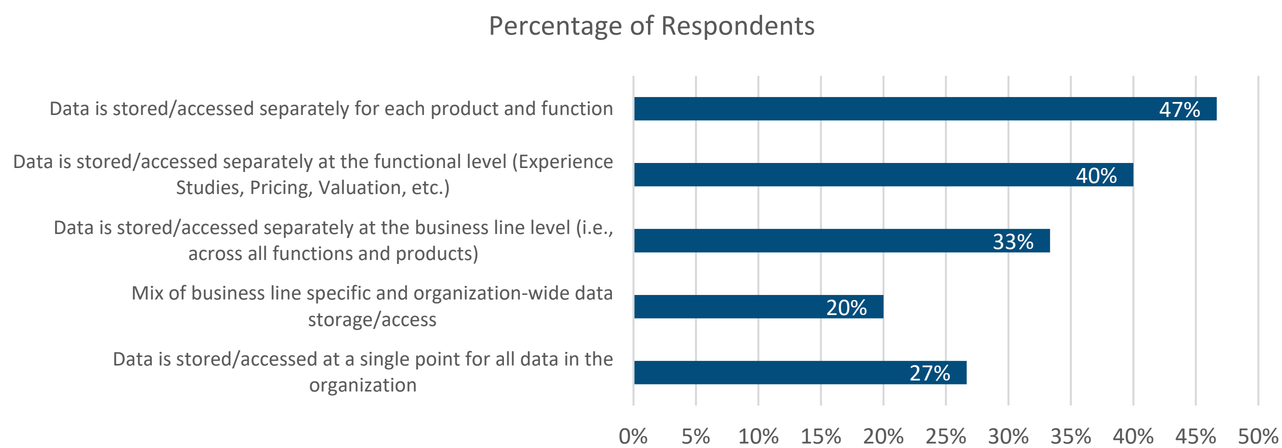
Survey Findings

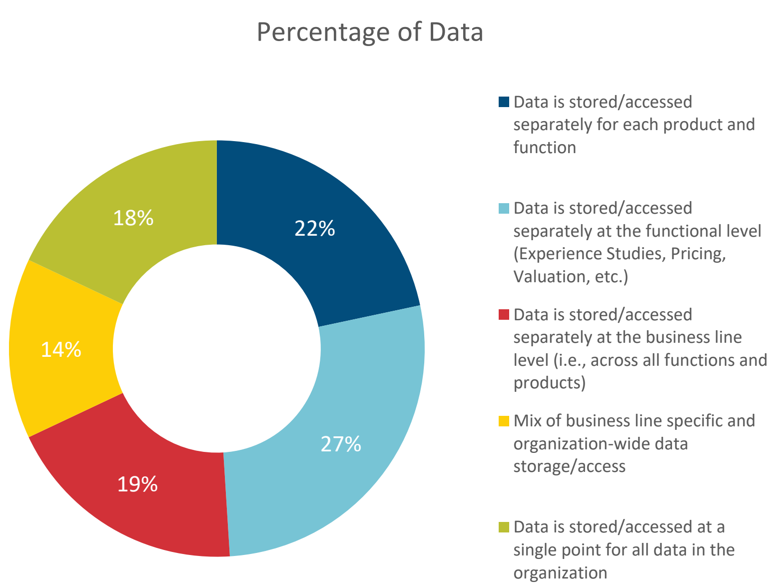
In the following section we outline the questions and responses to the survey questionnaire, grouped into the main themes.

6.1 Data

Data is the starting point for all analytical work at an organization. Is the data clean, easy to access, and effective?

Question 3.1 - What best describes the centralization of your data (as self-assessed by respondents)?

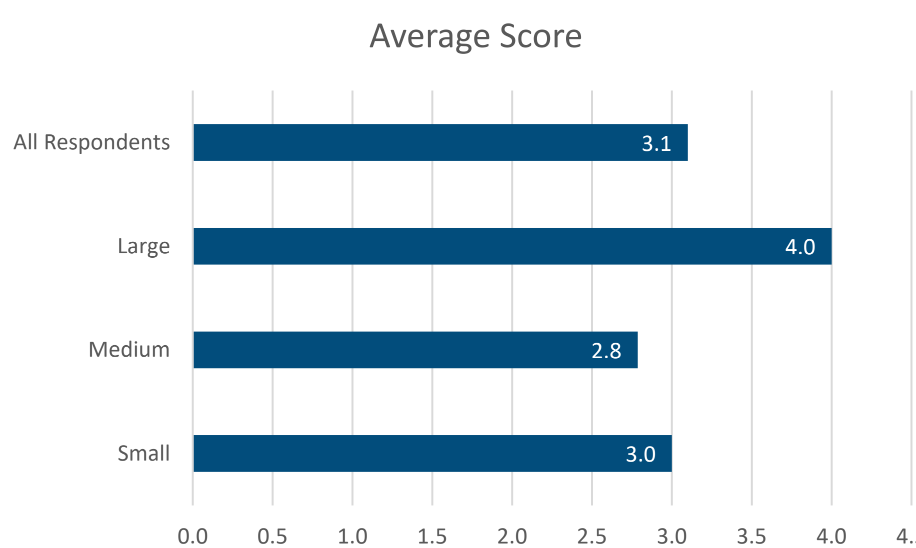




- Survey participants appear to be split on how data is organized across the business.
- Almost half (47%) of the respondents have segmented data at the product and function level.

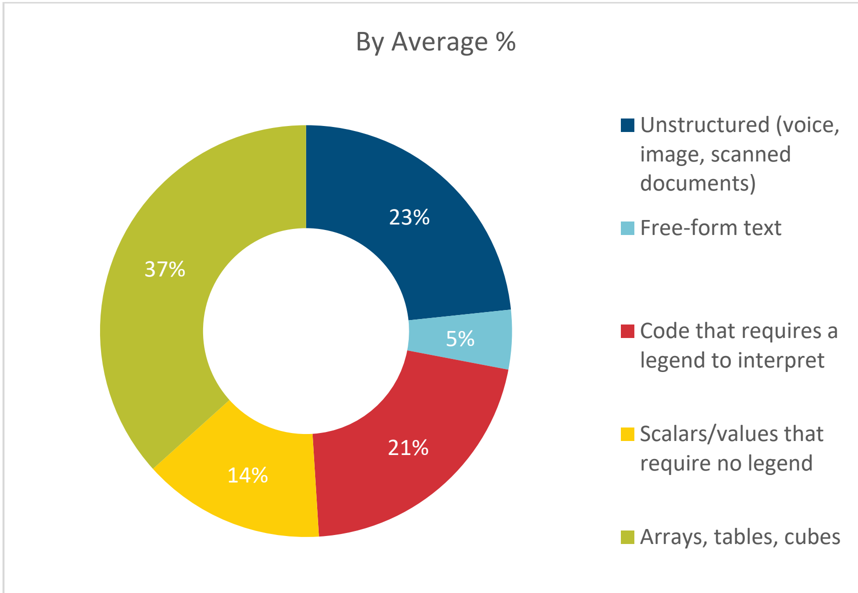
- Approximately one-third (27%) of all respondents indicated that they have some centralized data at organization level.
- Overall, 18% of all data owned by survey respondents is centralized. In our experience data centralization is likely to be more difficult to achieve for a large company with many legacy systems than for a smaller company.
- Sixty-eight percent of data is decentralized at the Business Unit (BU) or lower levels. We also observe that reinsurers have less centralized data than direct insurers.
- From our experience we found that the best practice is to reduce the number of access points and increase standardization by centralizing the data. This makes things easier for the users of data to extract the information they need, which could span multiple lines of business or functions. However, survey respondents also said that through various technology initiatives, acquisitions, and priorities, it can be difficult to harmonize all data. As the business grows, so does the complexity of the data structure and storage.

Question 3.2 - On a scale from 1 to 5, how would end users rate the completeness and accuracy of your data? [1 = lowest / 5 = highest]

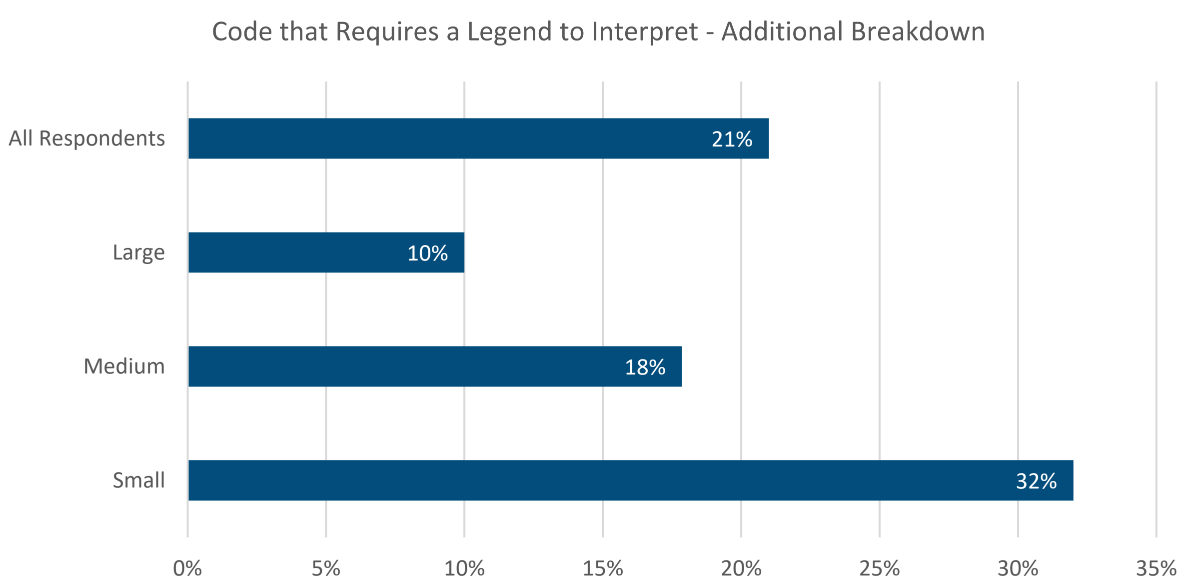


- Survey participants have rated the completeness and accuracy of their data as 3.1 out of 5. While this is not a bad score, it also shows that there is significant room for improvement. No respondent answered that they view their data as perfect (5 out of 5), which signals that every respondent believes that their data completeness and accuracy can be improved.
- There is a minimal difference between the average score indicated by medium survey participants versus small participants. Large participants assessed themselves with a much higher average score (4) than the remainder of the respondents (2.8 and 3.0 for medium and small respectively).
- There is only a minor difference between the rating for direct writers (3.14 out of 5) and that for reinsurers (3.00 out of 5).

Question 3.3 - Provide an approximate percentage of your data that is in each of the following categories.



- Regardless of the size grouping, there is a fair amount of unstructured data with 23% of the data being of that type.
- Only 5% of the data was identified as free-form text.
- As a result, less than one-third (28%) of respondents manage data in either unstructured or free-form text format.

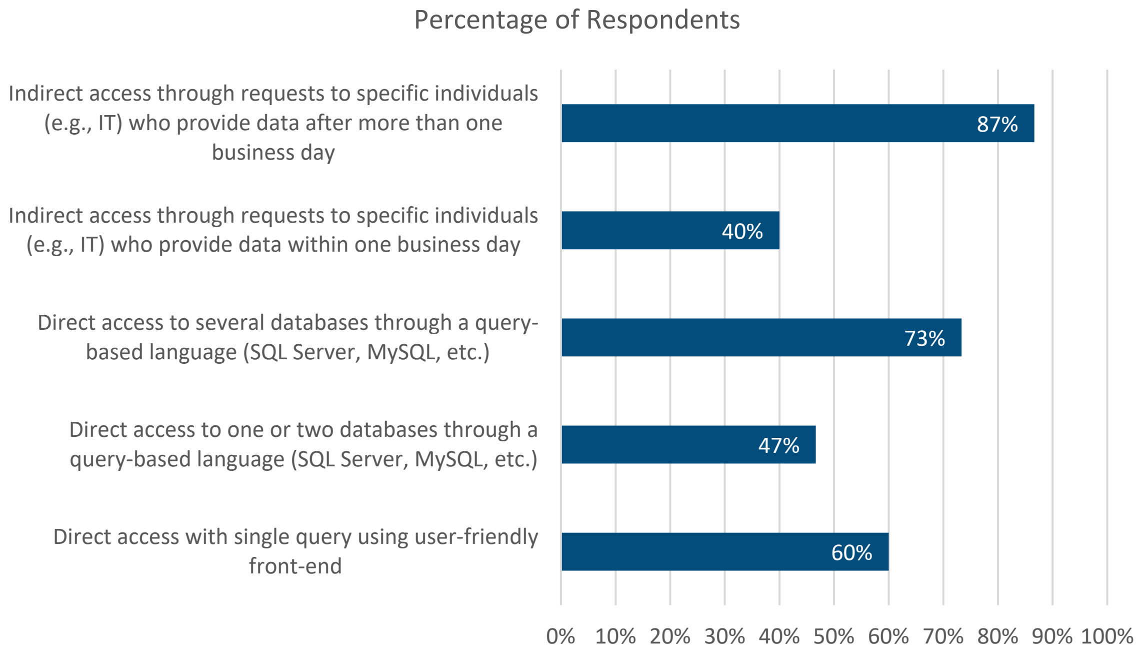


- Approximately one-fifth (21%) of data is stored in the form of 'Code that requires a legend to interpret'. There is a marked increase in percentage of that type of data as the group sizing decreases among respondents.

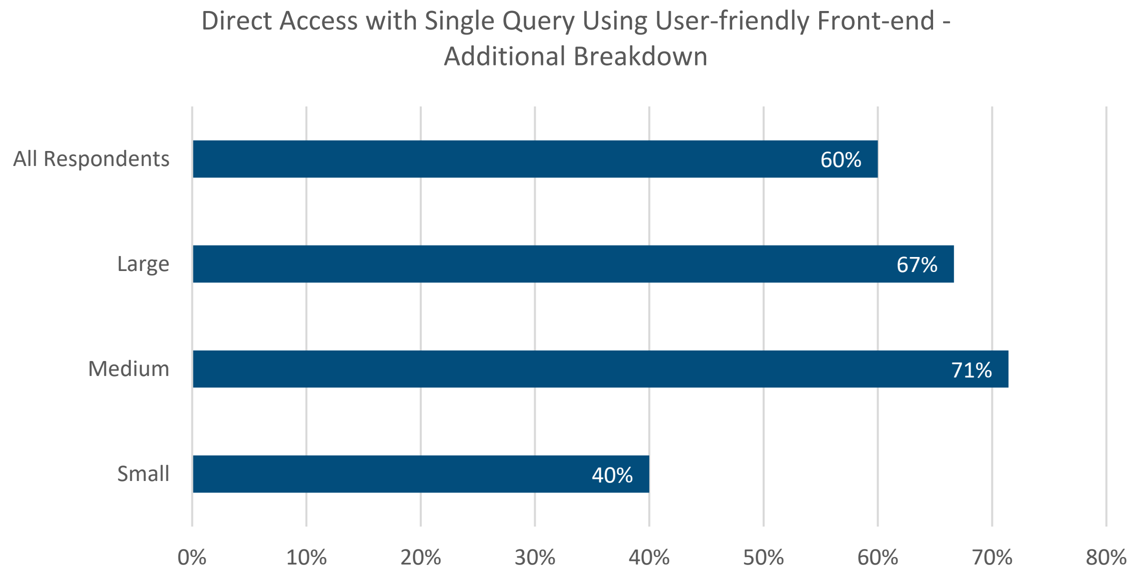
Each distinct color corresponds to one of the respondents in the survey.

- Survey participants store a wide range of data types, and the proportions vary greatly between respondents.
- In our experience the wide range of data types is at the basis of the challenges companies are facing by having to manage multiple data sources that have different underlying structures (or lack thereof).
- We can see from the above chart that for some participants there is a high percentage of the data that is unstructured. Historically, this would represent data such as claims forms and miscellaneous registration documentation (such as a picture of a driver's license). This data can be difficult to process and might not be leveraged in predictive modelling. However, the question remains whether or not this data leads to a wealth of information that could potentially improve predictive power.

Question 3.4 - How is data accessed across the organization? (Select all that apply)

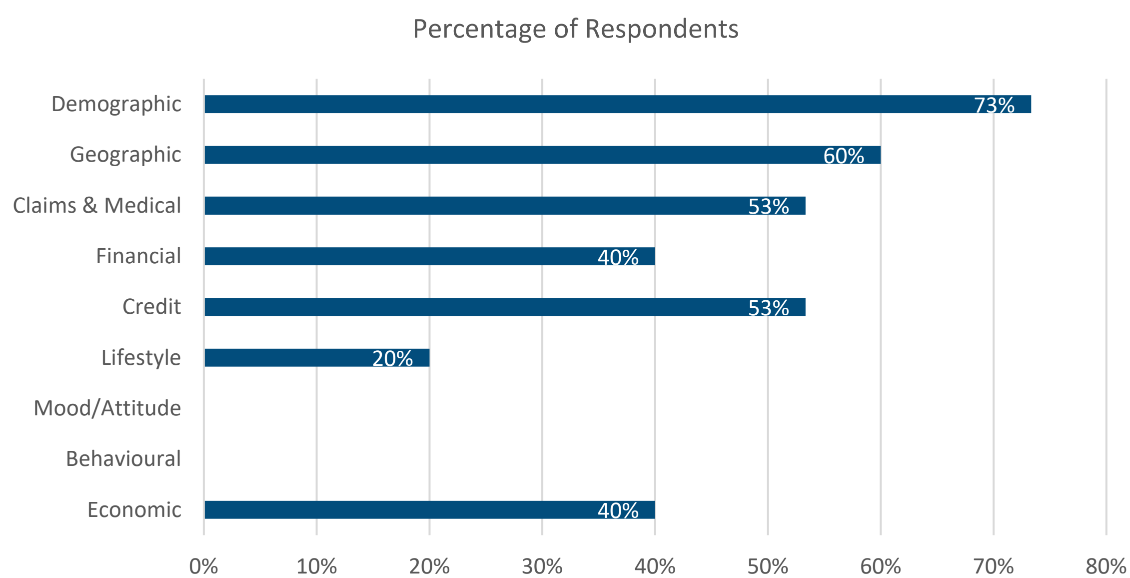


- The majority of the survey participants still request at least some data through specific individuals who are not members of their analytics group (such as a separate IT team).
- Of these respondents that are requesting through another source, most (87%) have a wait time that is longer than one business day (compared to 40% who wait one day or less). All large and medium respondents have such a delayed process, whereas a fraction of the small respondents do not.
- Indirect delayed access could present operational issues, especially in situations where the goal is to accelerate/automate extraction and reporting.
- In our experience the best practice for a business will depend on the users who require access to data. For those businesses that have experts in managing and extracting from databases, direct access through a query-based language is often the most efficient process. If the business does not have these experts, then a front-end access point would be ideal.
- Also of note is that maintaining multiple databases might be a necessity, but can also create complications when trying to join multiple sources.



- Direct access with single query using a friendly front-end exists more for large (67%) and medium (71%) respondents than for small respondents (40%).

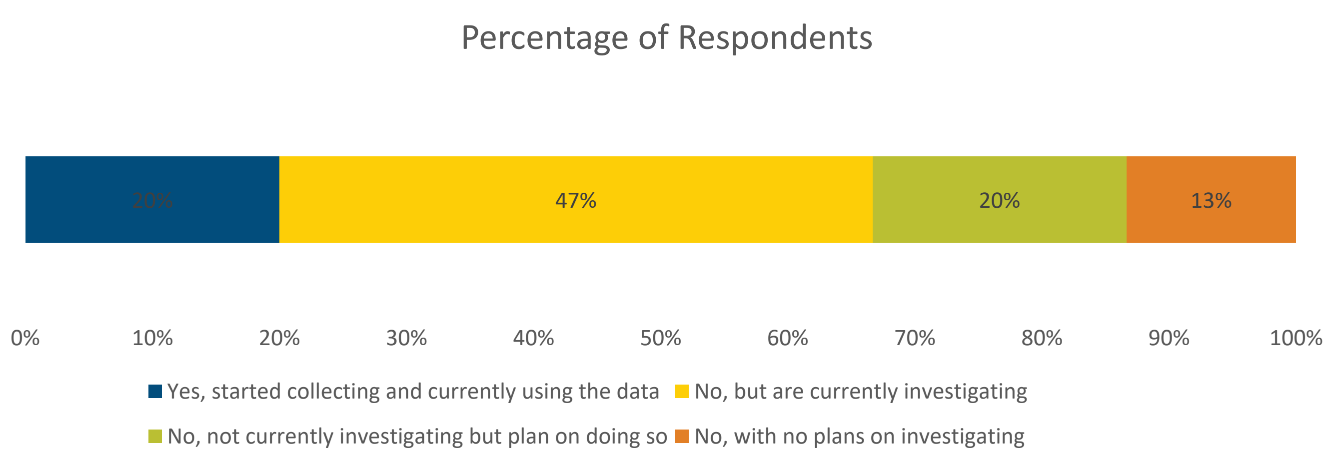
<u>Question 3.5 - What types of third-party data does your organization currently use for analytics? (Select all that apply)</u>



- Of those performing predictive analytics, all respondents indicated that they are currently using third-party data sources to supplement their own data. This is a strong signal that survey participants are willing to invest in their own analytics capabilities, and that they trust these third-party sources.

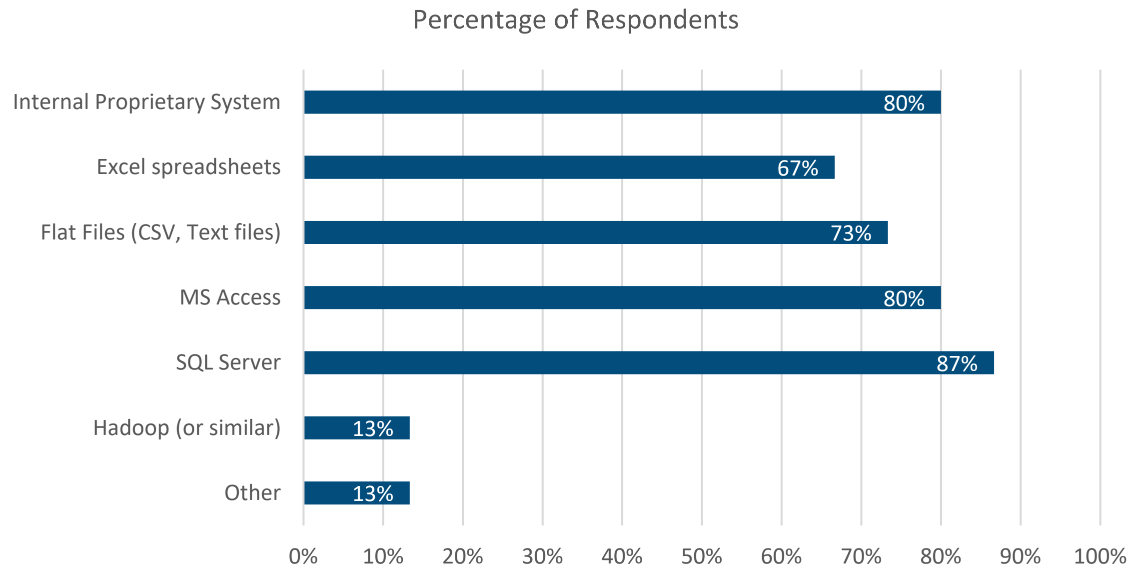
- Demographic data ranks as the most popular type of third-party data used among respondents, at 73% by company count, followed by Geographic data at 60%, Claims & Medical (53%), Credit (53%), Financial (40%), and Economic (40%).
- No respondent indicated that they used any third-party data for Mood/Attitude or Behavioural.
- For Claims & Medical the fact that not all large respondents indicated using such a type of data could be interpreted as them possibly presuming their own data is credible enough for current applications.
- Two-thirds of the large respondents are using third-party Lifestyle data, as opposed to only one small respondent.
- None of the small group respondents are using third-party Financial data. Only one large respondent is using third-party credit data, whereas higher usage exists for the medium and small respondents.
- Generally, the larger the respondent, the more it tends to use third-party Demographic, Geographic, Financial, and Economic data.
- Additionally, Credit and Economic data are used more by direct writers than by reinsurers.
- Due to the analysis in which these data sources are typically used, we can infer that third-party sources are mainly used for experience analysis (such as mortality rates, lapse rates, and conversion rates).

Question 3.6 - Have you started collecting data through new technologies (i.e., Fitbit) to augment the use of traditional sources of data?



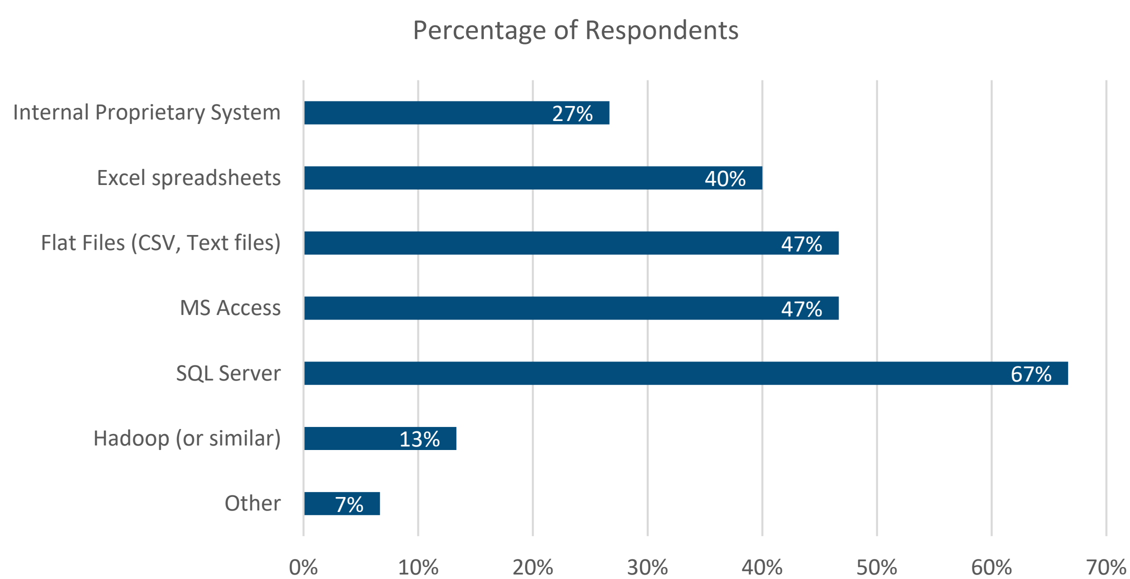
- One respondent from each size group (large, medium, small) has indicated that they have collected this type of data and are currently using it.
- No survey participant has indicated that they collected data and are not used it, nor have any investigated and decided not to collect/use this type of data.
- All large respondents that have not collected such data yet are investigating the value of augmenting data through new technologies. The same is true of all but two medium respondents.
- Generally, the majority of the survey participants are either using, or looking to use, data from new technologies. This confirms a growing sentiment that alternative data sources could be valuable to insurers and reinsurers. However, it is unclear at this time whether or not the use of this data is providing a competitive advantage for those companies.

Question 3.7 - How do you store your internal data?



- The most popular platforms to store internal data included: SQL Server (87%), Internal Proprietary System (80%), MS Access (80%), Flat Files (73%), and Excel (67%). Hadoop was also mentioned.
- Other software mentioned included DB2 and FileNet.

Question 3.8 - How do you store your external data?



- As expected, the storage platform used for internal data and external data is quite different.
- The most popular platforms to store external data included: SQL Server (67%), MS Access (47%), Flat Files (47%), and Excel (40%). Internal systems and Hadoop were also used.
- Other software mentioned included PDF and Word.
- Examples of other software are DB2, Oracle, and FileNet.

- There are advantages and disadvantages to using any of the software listed in this question. The best option for a company will, again, depend on the end users who are accessing the data. That being said, there are data management principles which are applied more easily in certain software over others. For example:
    - Microsoft SQL Server allows you to create primary and foreign keys which relate to the tables in your database. Microsoft Excel does not inherently have this feature.
    - Other considerations such as the size of your data, security protocols, user friendliness, and processing speed will depend heavily on the software you choose to store your data.
- There is still a large portion of the survey participants that are using “less than ideal” software to manage their data across the business. This would include any data that is maintained specifically within Excel or another flat file extension (such as csv and txt files). The high percentages in other “ideal” software indicate that most businesses are using it as well, just not for all their data sources.
- Third-party data is stored more regularly within a relational database. Typically, third-party data already comes in a structured format, where the relation between tables is spelled out, so it is easier to store in a relational database. However, this could also be an indication of entities purchasing third-party data being more technically advanced with data management and thus preferring to store it within a relational database.
- Hadoop (or similar software) is still only used by a smaller portion of the survey participants. We may see this percentage grow in coming years as the survey participants seem to handle a high level of unstructured data.

Summary – Data

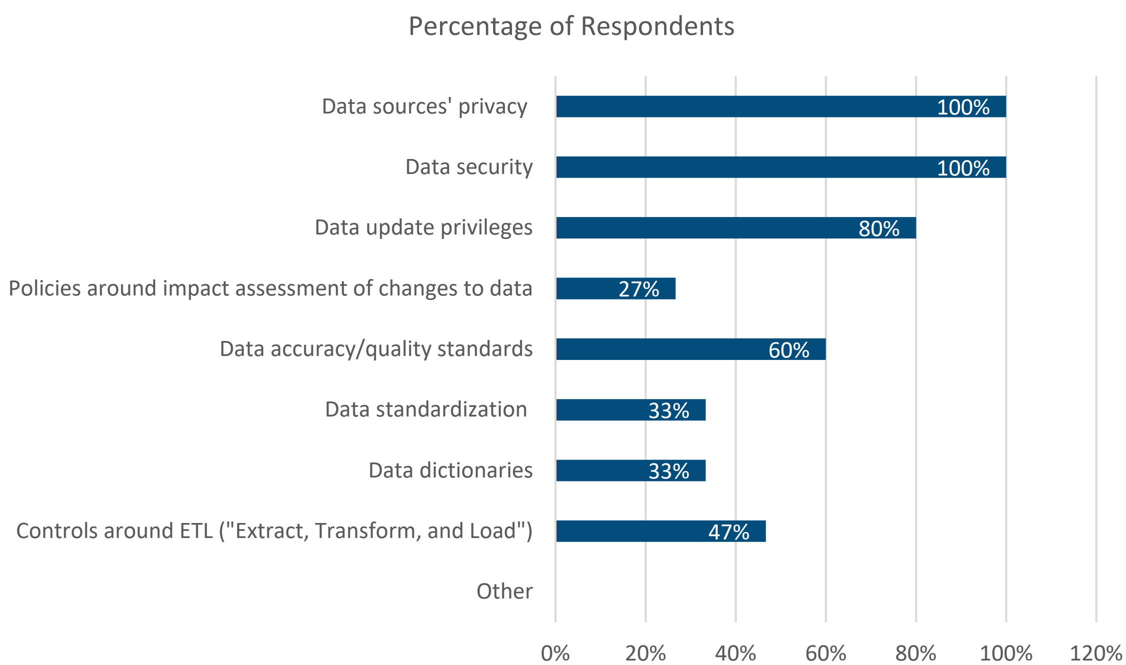
Based on the survey responses there are improvements to data for predictive analytics to be made with regard to centralization, quality of data, and access procedures to relevant data.

Additionally, new technologies (for use of augmenting data) have not yet been significantly leveraged by the survey participants and this is an area that insurers are starting to tap into.

6.2 Enterprise

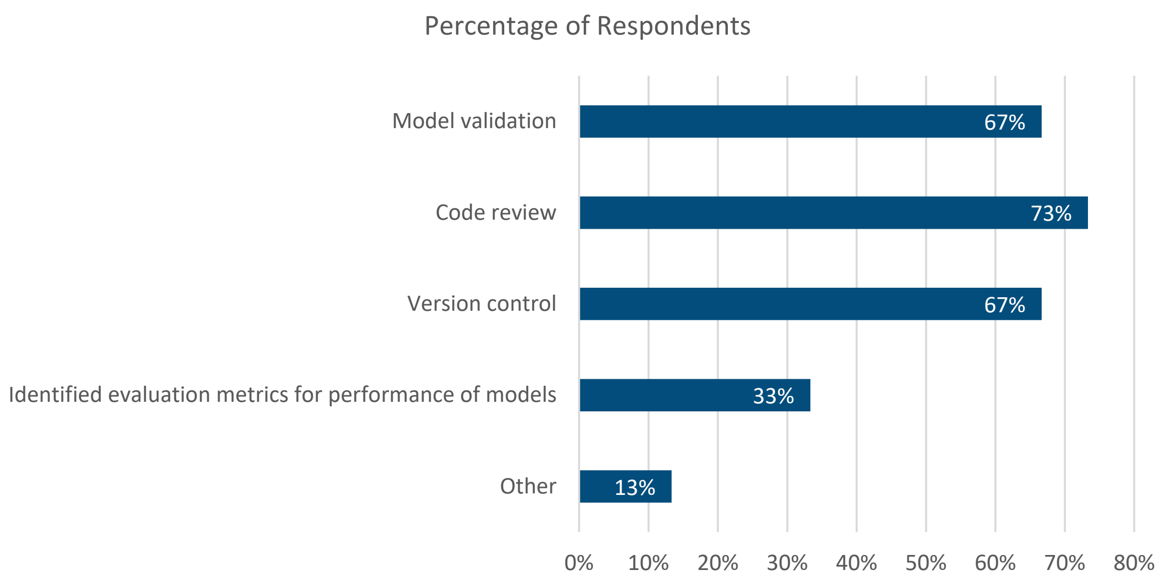
It is important to have proper hardware, software, and structure to manage the data. How is the data stored, managed, and accessed by analysts at the organization?

Question 4.1 - Select the governance aspects and policies surrounding data used in analytics for your organization. (Select all that apply)



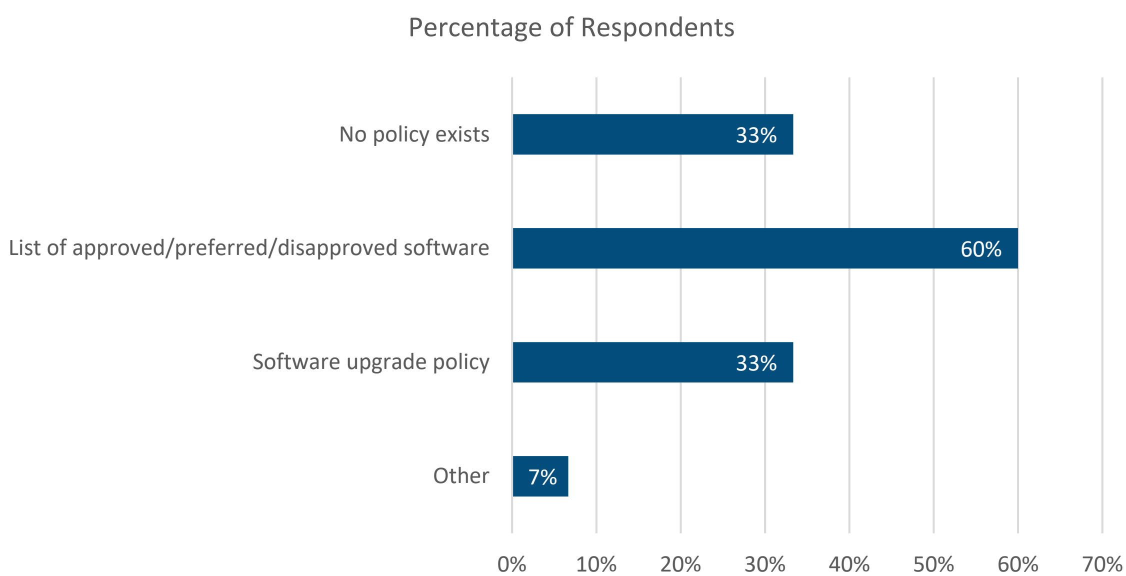
- All respondents indicated having governance/policies around data privacy and data security. Other popular governance/policies in place included: data update privileges (80%), data accuracy/quality standards (60%), and controls around ETL (47%).
- Few respondents (approximately one-third) have governance/policies around standardization, producing data dictionaries, and data change management.
- There is a higher focus on governance/policies for data update privileges, data accuracy, and quality standards for larger respondents.
- The highest emphasis is placed on data privacy and data security. This is logical given the data points that are often handled by direct writers and reinsurers. There is a high risk associated with exposing sensitive information, and the survey participants are actively taking steps towards mitigating this.
- There is less emphasis on governance policies that can assist with predictive modelling. It appears that since there is a lower risk associated with these policies, some respondents are not enacting them. That being said, there is still risk associated with lack of knowledge regarding one's own data. It is regarded as best practice to at least acknowledge procedures for each of the items listed above.

<u>Question 4.2 - Select the governance aspects and policies that impact modelling efforts for your organization. (Select all that apply)</u>



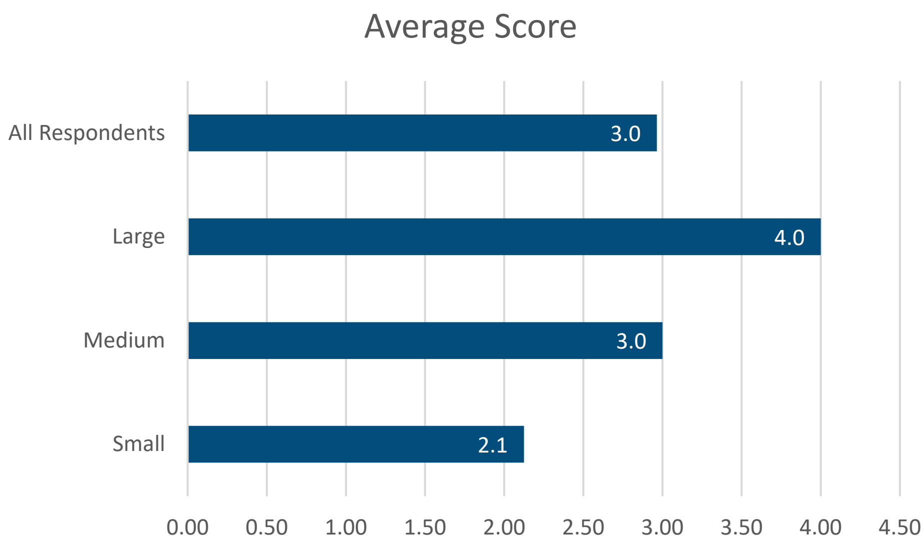
- About two-thirds of the survey participants have governance policies for their predictive models. The remaining one-third of the survey participants perform very minimal predictive modelling work to begin with and so governance policies in this field are a lower priority.
- The most popular governance/policies impacting the modelling included: code review (73%), model validation (67%), and version control (67%), which are all important to the modelling process.
- Other responses mentioned included documentation standards, issue tracking, and code documentation.
- One interesting result is that only about one-third of the survey participants have a governance policy regarding the evaluation metrics for performance of models.
- All of the large respondents have model validation, code review, and version control but only two-thirds of them have identified metrics. It is our opinion that the identification of appropriate evaluation metrics and the definition of the criteria is an important consideration for selecting the optimal model, and companies would benefit from adopting such policies.

<u>Question 4.3 - Select the governance aspects and policies relating to software and technology used in analytics applications for your organization. (Select all that apply)</u>



- Popular governance/policies around software and technology included: list of approved software (60%) and software upgrade policy (33%).
- As many as one-third of respondents had no governance/policies in place regarding software and technology. We would have expected very few not to have such policies in place.
- Other governance/policies indicated by respondents included: deployment guidelines and pre-production/production environment policy.
- Two-thirds of large respondents have a list of approved/preferred/disapproved software. Only one large respondent, two medium respondents, and two small respondents have software upgrade policies. Thus, only five out of 15 survey respondents indicated they have software upgrade policies.
- Overall there is less governance in place for software, when compared to both data and models. One hypothesis for why this is the case is based on the phenomena of open-source software. As later questions in this survey have demonstrated, open-source software (R, for example) has become the most commonly used by survey participants. Governance policies would need to be reviewed on a frequent basis if they are to stay up to date with all the developments for open-source software and their packages.
- Special consideration will need to be given in the future with regard to user-written packages in open-source software. While the base code may be tested on a regular basis, user-written code might not be tested appropriately. It may be worth identifying if there will be a potential shift where businesses start to write more governance policies on user-written packages.

<u>Question 4.4 - If the analytics functions are not fully centralized, please assess the strength of the skills and resources coordination across the organization. [1 = extremely weak / 5 = extremely strong]</u>



- Overall, respondents self-assessed the strength of the skills and the resource coordination with a score of 3.0 out of 5.
- Large respondents have self-assessed much higher skills and resource coordination with a score of 4 out of 5, followed by medium respondents with 3 out of 5, and small respondents scoring 2.1 out of 5.
- The overall score suggests that there is a certain level of acceptable coordination occurring across the survey participants, yet there is still room for improvement.

Summary – Enterprise

We see from the responses that there is progress to be made with regard to governance and policies around standardization.

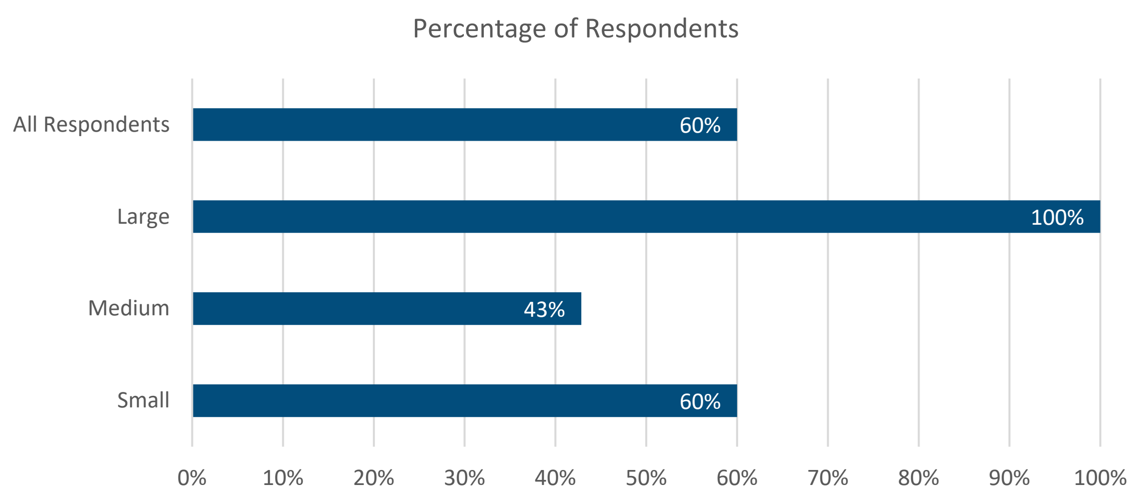
Medium/small survey participants need to bridge the current gap with large participants with respect to governance around data (specifically data updates, data accuracy, and standards). It is likely that the same gap applies to these size groups for rest of the industry. Similarly, the gap should be addressed with respect to governance around modelling (specifically model validation, code review, and version control).

Only five out of 15 survey respondents indicated they have software upgrade policies, which is significantly lower than is perhaps warranted.

6.3 Leadership

Senior leadership want to unlock the power of analytics within their organisations. Are they willing to hire the right people, spend time building the correct systems and processes?

Question 5.1 - Does your organization have an executive responsible for data and what is his/her title?

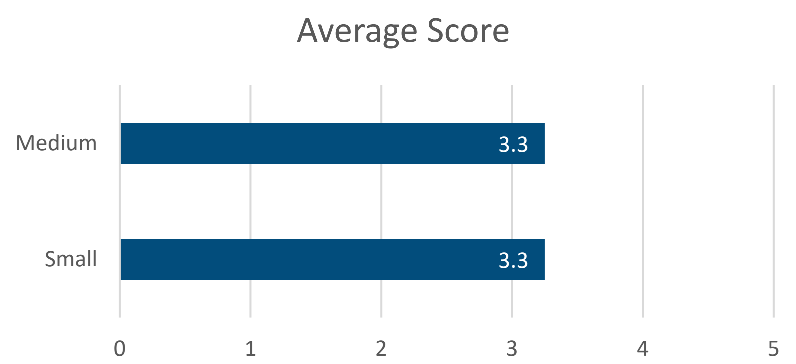


- Overall 60% of the survey participants have an executive responsible for data, with all large respondents indicating the existence of such a role. Surprisingly, more of the small respondents than medium respondents had this role, although it is unclear if this executive had other responsibilities (e.g., CIO).

The following summarizes the titles of the executives responsible for data:

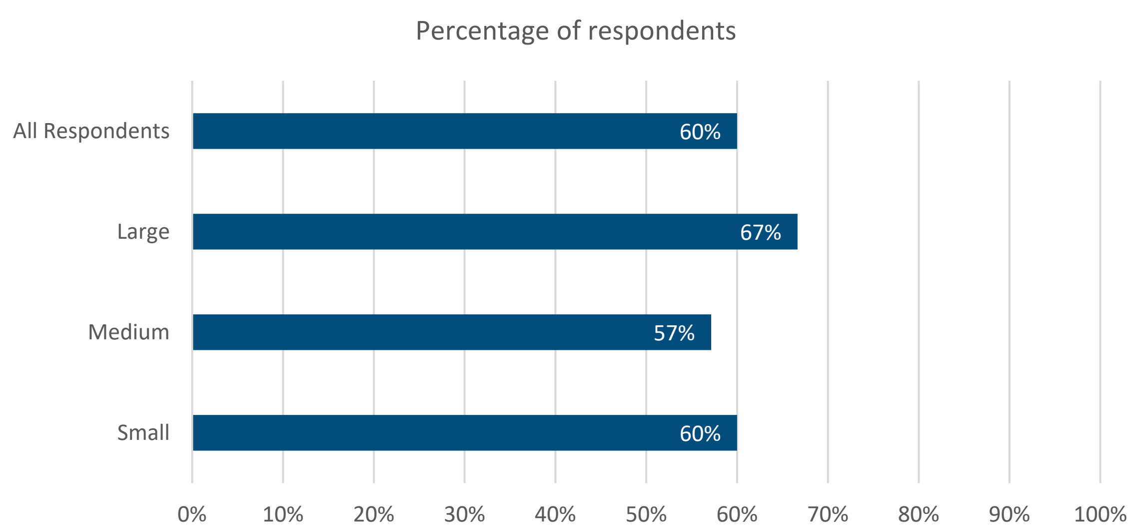
Title	Count
Chief Data Officer	3
Chief Vice President/Executive Vice President	1
Vice President	3
[Exists but title unknown]	1
None	7

Question 5.1a - If there is no executive responsible for data, how much does leadership understand data (quality, structure, safeguards, etc.)? [1 = No understanding / 5 = Full understanding]



- In cases where there is no executive responsible for data, the self-assessed leadership understanding of the data was 3.3 out of 5.
- In our experience the use of high-level dashboards showing the state of the data is effective in increasing leadership's understanding of the data.

Question 5.2 - Does your organization have an executive responsible for predictive analytics and what is his/her title?

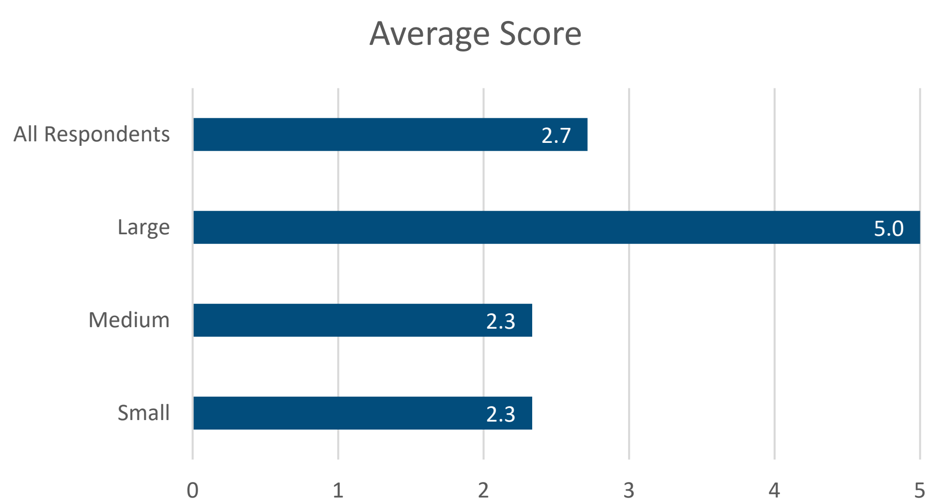


- Overall 60% of the survey participants have an executive responsible for predictive analytics, with more (67%) of the large respondents indicating the existence of such a role.

The following summarizes the titles of the executives responsible for analytics:

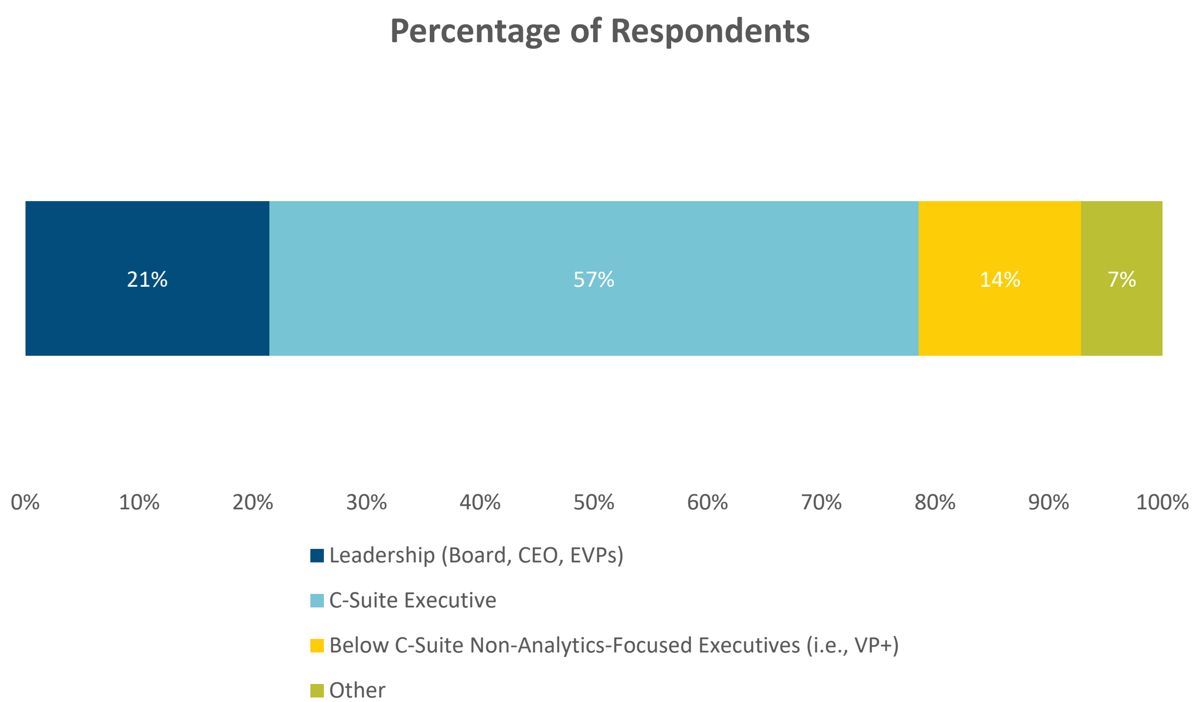
Title	Count
Chief Analytics Officer	1
Chief Vice President/Executive Vice President	1
Senior Vice President	2
Vice President	3
Director	1
[Exists but title unknown]	1
None	6

Question 5.2a - If there is no executive responsible for predictive analytics, how much does leadership understand its value-added opportunities and competitor's initiatives? [1 = No understanding / 5 = Full understanding]



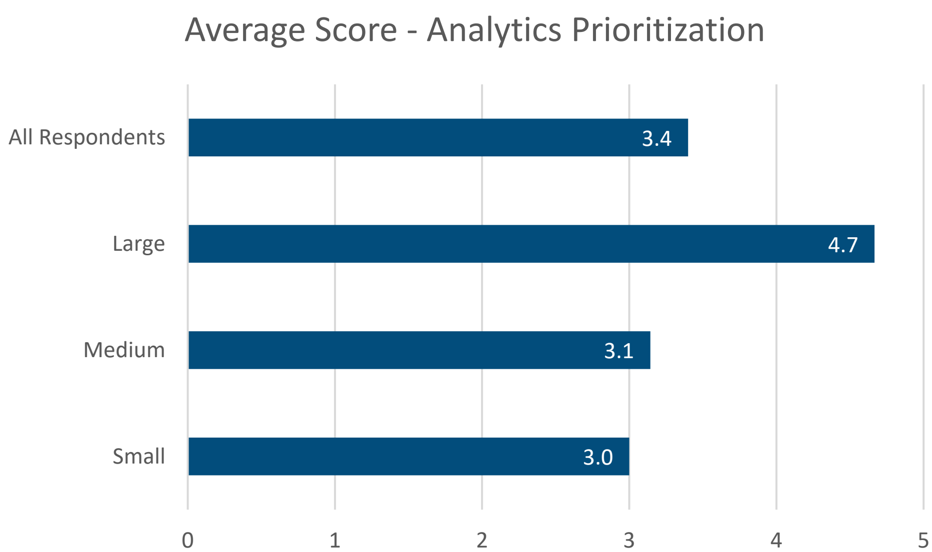
- In cases where there is no executive responsible for analytics, the self-assessed leadership understanding of the value-added opportunities and competitor's initiatives was 2.7 out of 5.
- Large respondents indicated a perfect understanding.
- This implies that, for other than large companies, it can sometimes be a struggle to get buy-in for the need for/results of predictive models from senior executives, especially at the early stages of a company's predictive analytics journey. Similarly, this could be moved forward through easy-to-use dashboards or, in the case of predictive modelling, converting to a quantifiable result that is easily understood by senior executives.

<u>Question 5.3 - Who makes the final decision on the prioritization of predictive analytics initiatives?</u>



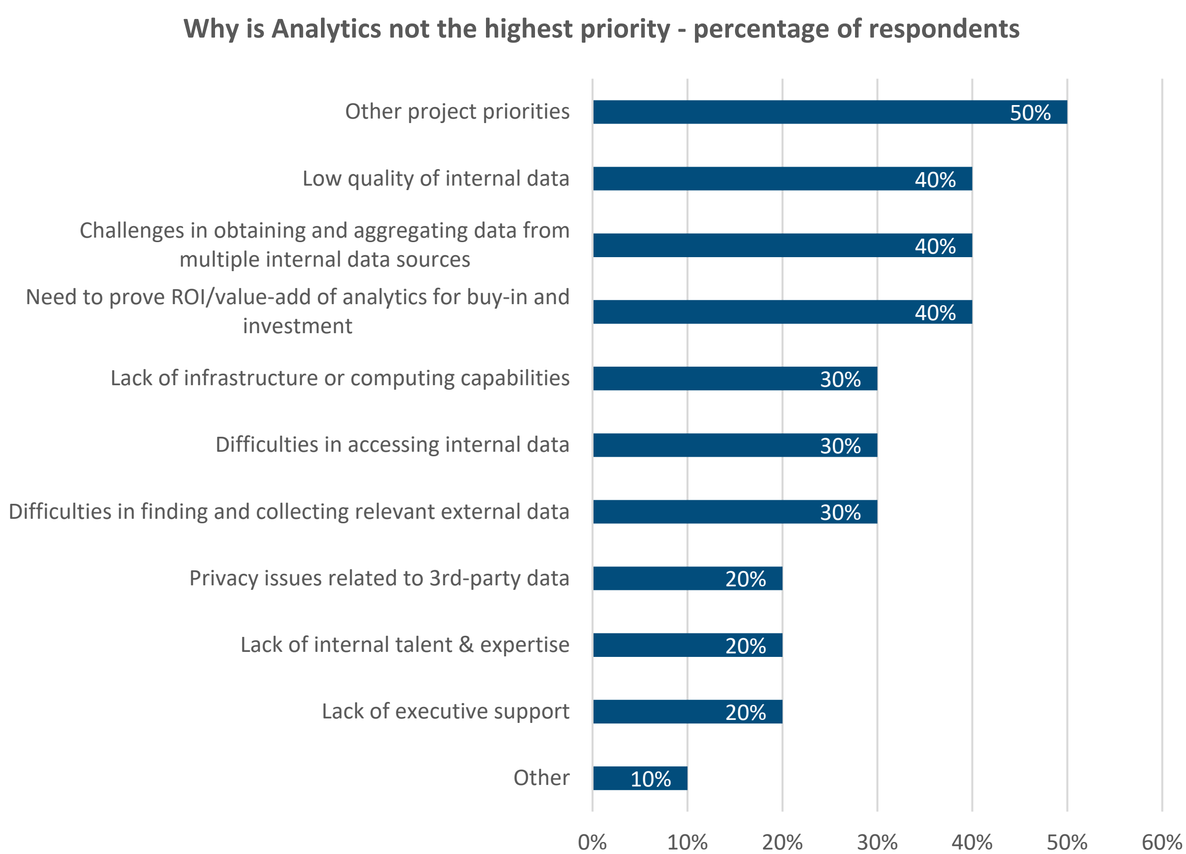
- The majority of the survey participants (78%) have indicated that the senior executive leaders (i.e., C-suite or higher) make the financial decisions when it comes to predictive modelling initiatives. Close to one-fifth (21%) of boards, CEOs, and EVPs are involved in such decisions.
- Two-thirds of decisions are made at the C-suite level for large and medium respondents, while all decisions are taken by that level for small respondents.
- While this may be expected, it also shows a disconnect between the financial decisions and the general understanding of data and modelling practices. What this likely means is that trust is being placed on the middle management that they know what they are doing for each predictive modelling initiative.

<u>Question 5.4 - How does predictive analytics rank within your organization with respect to project priorities? [1 = Low priority / 5 = most important investment for the future]</u>



- On average, large respondents have indicated a high prioritization of analytics initiatives. The prioritization given by medium and small respondents was lower and not materially different for those size groups. It is worth noting that, on average, reinsurers have placed a noticeably higher priority (3.8) on analytics initiatives than direct writers (3.3).

Question 5.4a - If predictive analytics is not the highest priority (i.e., scored as 5 in prior question), what are the key reasons (check all that apply)?



- As many as 10 respondents indicated analytics as not being at highest priority.
- The most common reasons indicated by them included existence of other project priorities (50%), the low quality of internal data (40%), the challenges in obtaining data from multiple internal data sources (40%), and the need to prove ROI (40%).
- Lack of infrastructure or computing capabilities, difficulty in accessing internal data, and difficulties in collecting external data were also mentioned.
- There does not seem to be a generalized lack of executive support.

Summary – Leadership

Sixty percent of all respondents have a dedicated executive for both data and predictive analytics; all large respondents have one for data and two-thirds have one for predictive analytics.

Seventy-eight percent of final decisions surrounding predictive analytics are made by C-suite or higher.

Only one-fifth of boards, CEOs and EVPs are involved in decisions surrounding predictive analytics (this appears low to us and we would have expected more involvement for something as important as predictive analytics).

Key reasons for lower prioritization include other competing business priorities, data quality, challenges in aggregating data, and the need to prove ROI of analytics efforts.

6.4 Targets

Goals and targets should be identified for the analytics work. Have the strategic decisions and users been identified?

Question 6.1 - For what current and past applications are you using analytics? “Applications” are defined as cases where a predictive (or ML, AI, etc.) model has been calibrated for prediction purposes, to identify drivers/explanatory variables, etc.

How do you rate each of these applications in terms of perceived effort to implement and perceived value to the company?

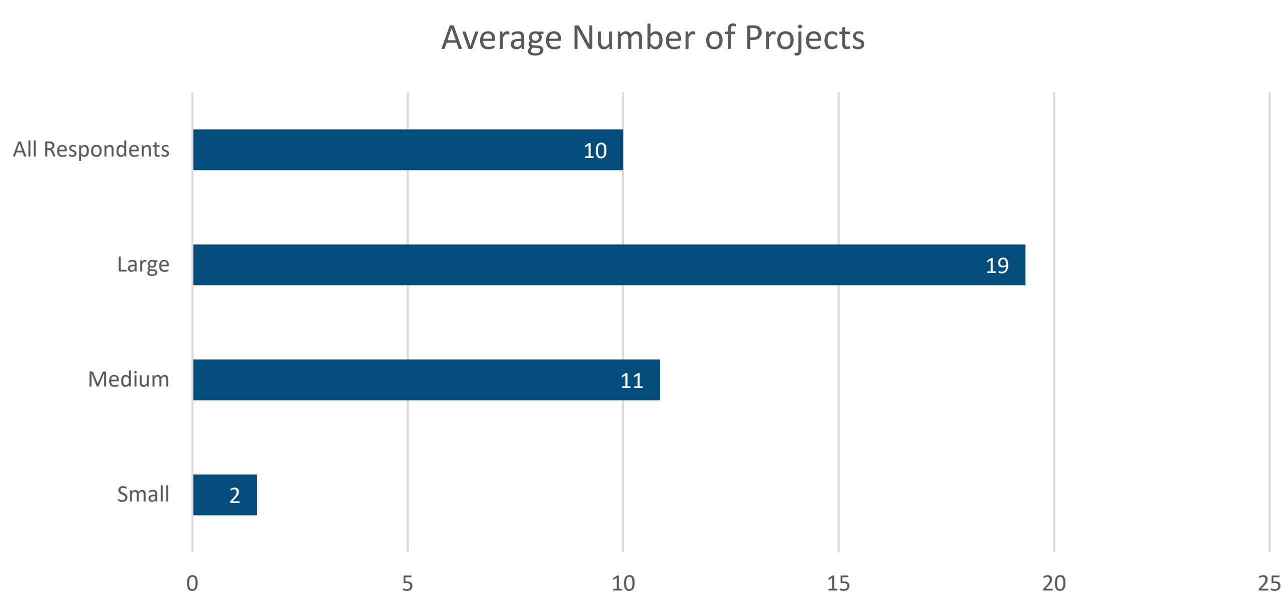
This is summarized on pages 11 to 15.

Question 6.2 - For what applications are you planning on using analytics in the next year?

How do you rate each of these applications in terms of perceived effort to implement and perceived value to the company?

This is summarized on pages 16 to 17.

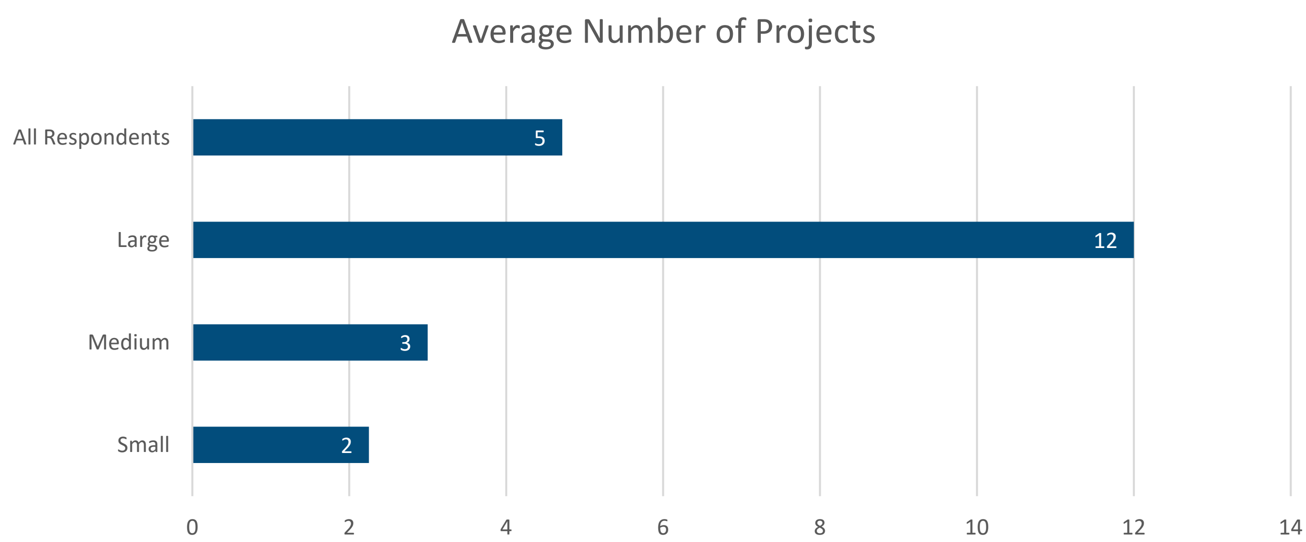
Question 6.3 - Approximately how many analytics projects have you attempted in the last year?
[Participants were asked to provide a count at the application/product level, such as Term 10 Lapse, as an example.]



- The survey participants averaged 10 projects in the last year, with 19 for large participants, 11 for medium participants and two for small participants. One of the large respondents indicated a large number of projects, which increased the average for that category significantly.
- One-fifth of respondents have not performed any projects in the last year.
- As many as 42% had three or fewer projects, including those that had none.
- None of the small respondents performed more than three projects.

- It appears that survey participants are either deeply involved in predictive analytics assignment, or still relatively new to it, with a small number of respondents attempting a moderate amount of work in the field.

Question 6.4 - Approximately how many new analytics applications that were not previously executed, are anticipated to be tackled in the next year?



For this question we asked the respondents to include new types of analytics and exclude the simple application of a prior analytics initiative to another business group.

- All but one of the respondents have projects planned for the future.
- An average of five projects are planned across the survey participants, with 12 projects for large participants, three for medium participants and two for small participants.
- Most respondents will have more than one new project.

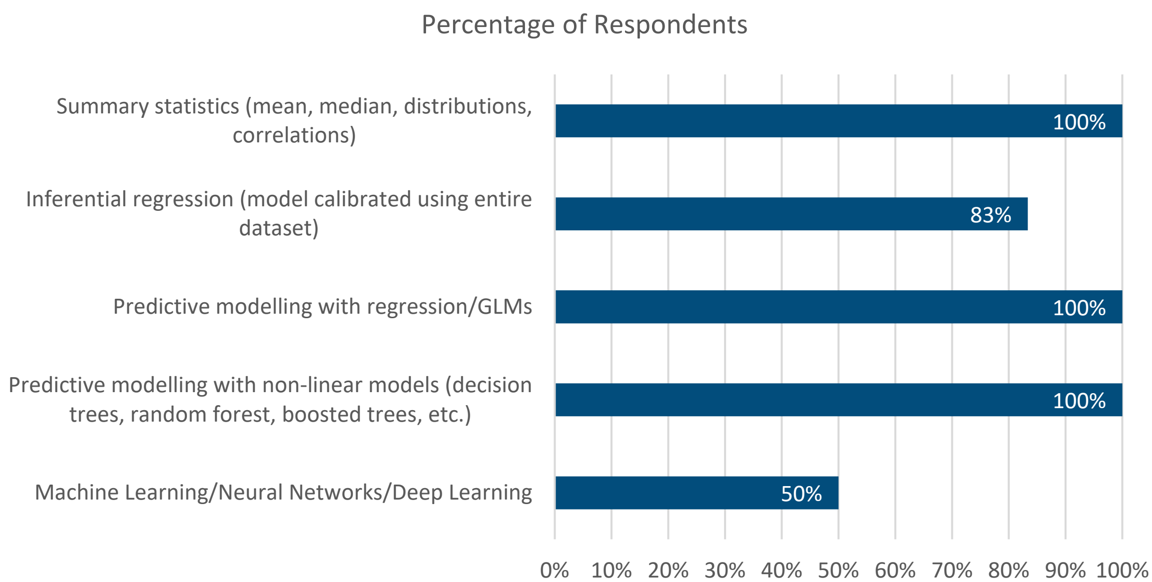
Summary – Targets

The large survey participants are, on average, performing more projects and have more projects planned for the future than medium or small participants.

6.5 Analysts

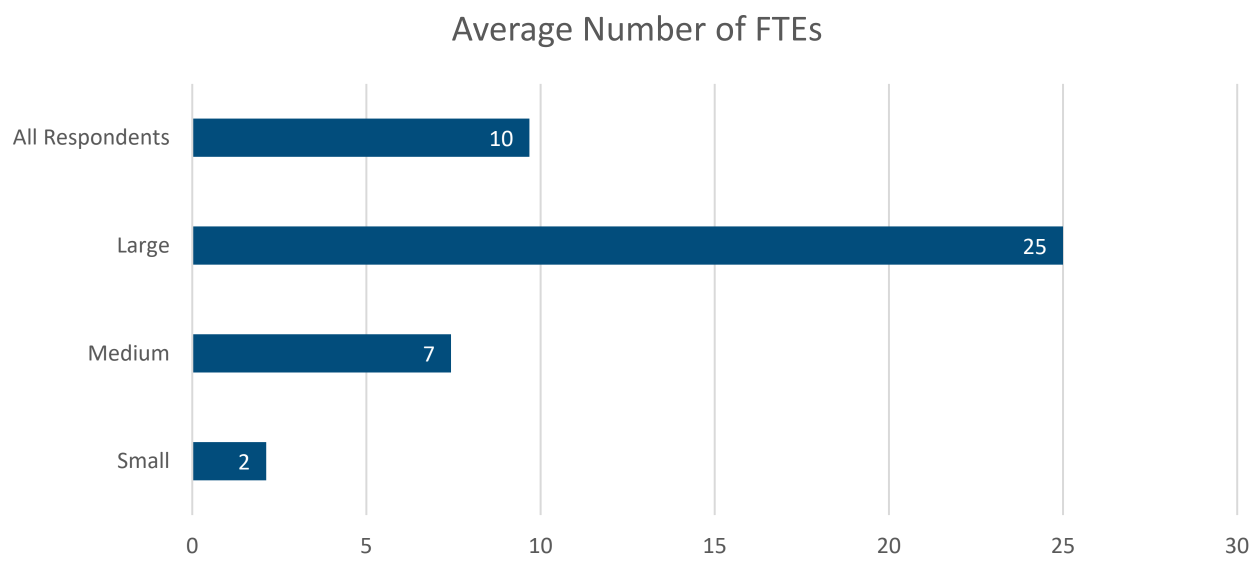
No two analysts are the same and it is important to identify the correct fit for each organization. Have analysts been hired that fit the business requirements to succeed in doing analytics work?

Question 7.1 - What techniques are currently being used?



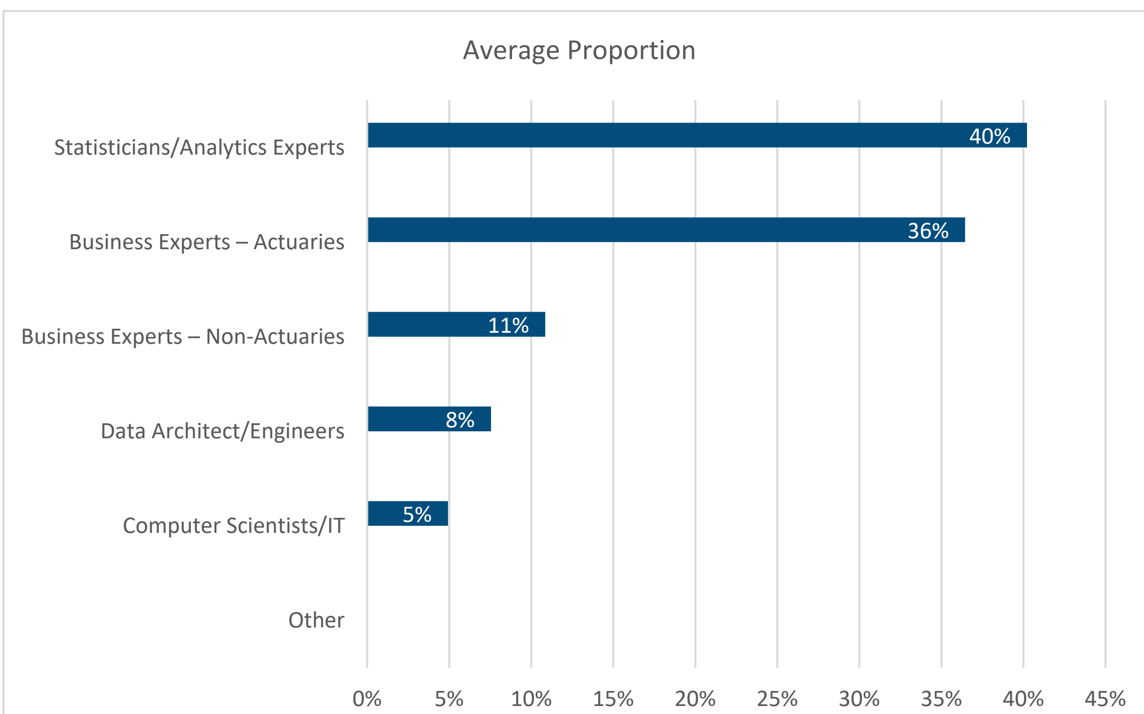
- Of those who are performing predictive analytics (this question excludes three respondents who had yet to perform analytics modelling at the time of this interview), all respondents indicated they are implementing summary statistics, GLM, and decision tree techniques.
- Most survey participants are using multiple approaches, with five respondents indicating all techniques.
- The least-used techniques included Machine Learning/Neural Networks/Deep Learning, with only 50% of respondents using such techniques.
- The traditional approach to understanding the data is the production of some type of summary statistics, so it is not surprising that companies widely included this as part of their analytics process, which also includes understanding of the data.

Question 7.2 - How many full-time equivalents (FTE) are working on analytics?



- The average FTE dedicated to working on analytics is 10, broken down into: 25 for large respondents, seven for medium respondents and two for small respondents.
- It is clear from the above that large respondents have dedicated more resources to analytics, whereas some of the medium/small respondents have not yet dedicated any.
- Respondents that are direct writers have dedicated more resources than reinsurers.

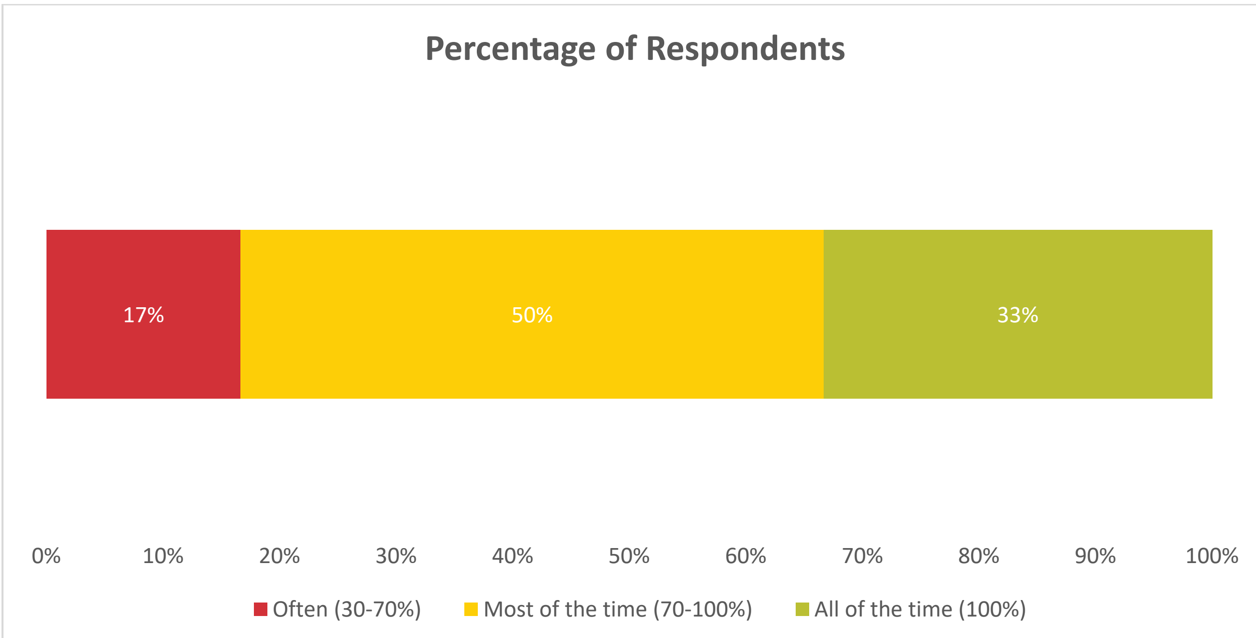
Question 7.3 - Approximately what are the proportions of the FTEs in each of the following categories?



- Overall close to 50% of FTE are business experts as compared to data architects and statisticians.

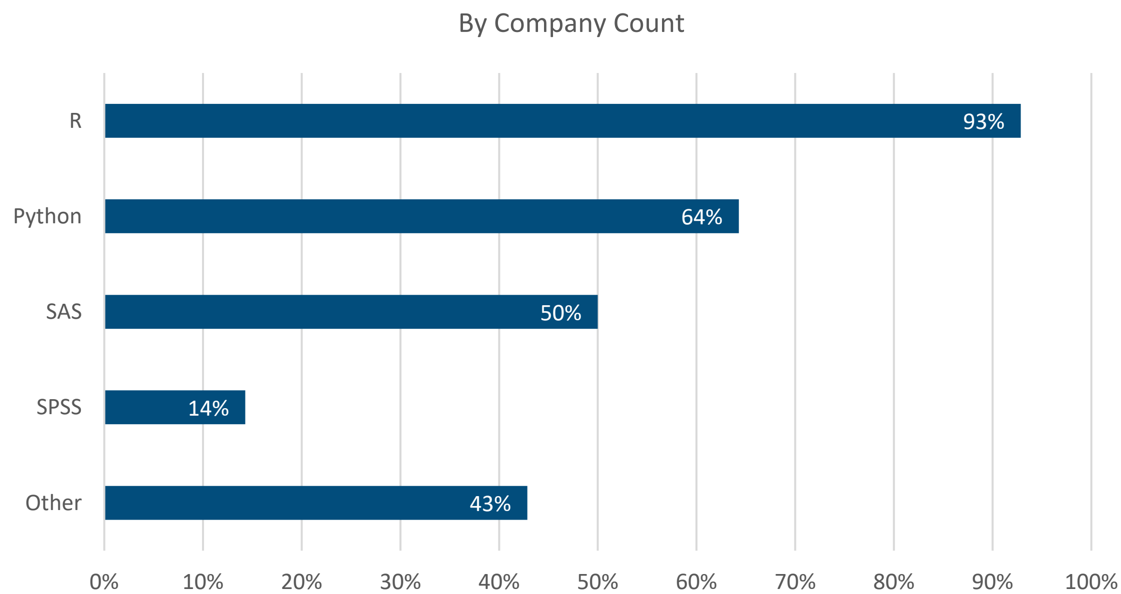
- Where business experts are used, approximately three-quarters of them are actuaries.
- Large survey participants tend to use more data architects/statisticians (at 78%), whereas small participants tend to use more business experts (actuarial/non-actuarial) (at 81%).
- Reinsurers tend to use more business experts than direct writers (83% vs 35%, respectively).
- Finally, limited use is made of computer scientists/IT specialists at the time of this survey (5% of resources only).

Question 7.4 - On each predictive analytics project, to what degree do you integrate the skillsets of Business Experts (both Actuarial & Non-Actuarial), Statisticians/Analytics Experts, and IT personnel?



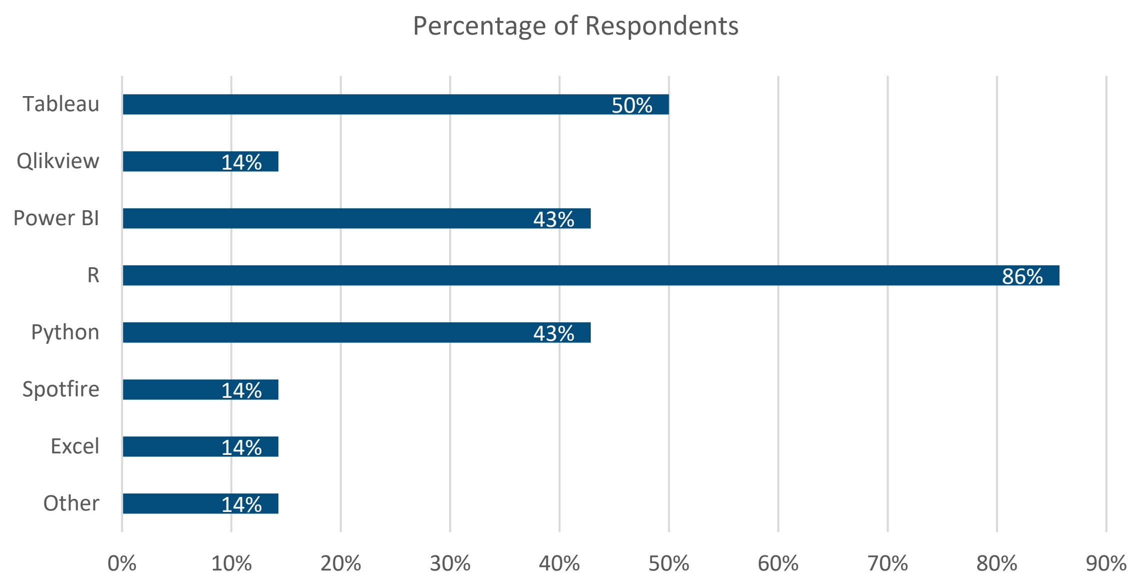
- 33% of respondents indicated using an integrated team all the time and 50% of respondents indicated using them most of the time. This translates into 83% of respondents indicating they integrate skillsets most of the time or more often.
- This leads to better business knowledge integration and is likely to lead to higher predictive power in modelling and higher success rates in solving relevant business issues.

Question 7.5 - What software are you using for predictive modelling? (Select all that apply)



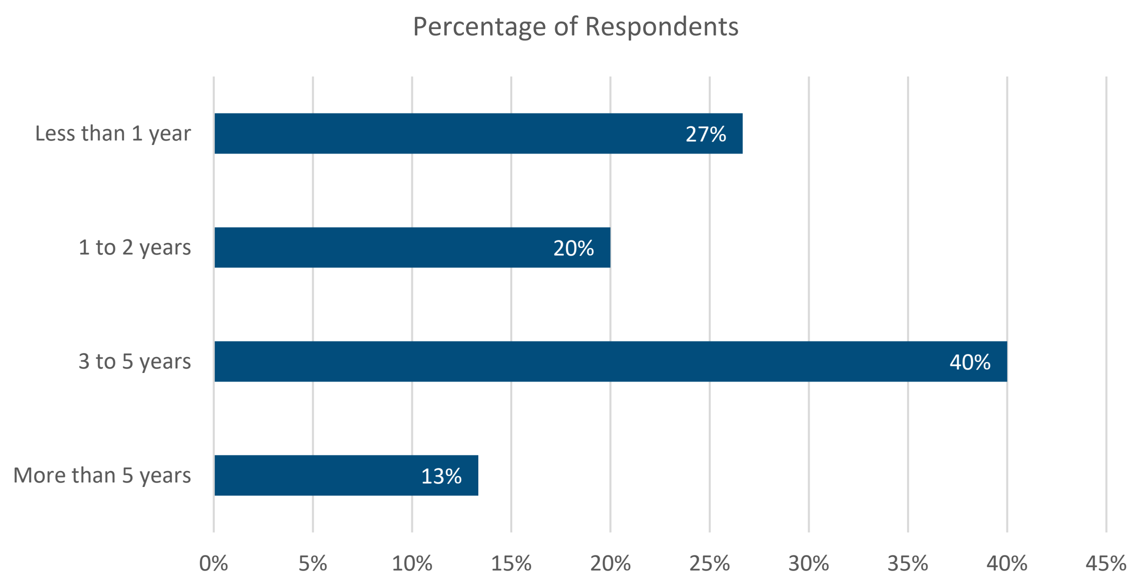
- The open-source software R holds a strong lead in popularity at 93% usage (13 participants) across the survey participants. Other popular software included Python (64%, or nine participants) and SAS (50%, or seven participants).
- Survey participants indicated using other software, such as: Data Meer, MS Power BI, Statistica, SPSS, Matlab, and IBM Watson.
- SAS is used more by medium survey participants than by larger participants.

Question 7.6 - What software are you using for exploratory data analysis and visualization? (Select all that apply)



- R continues to reign as the most popular in terms of usage for analysis and visualization at 86%, or 12 survey participants.
- Other popular software at the time of the survey included Tableau (50%), Power BI (43%), and Python (43%).
- Another piece of software mentioned was SAS.

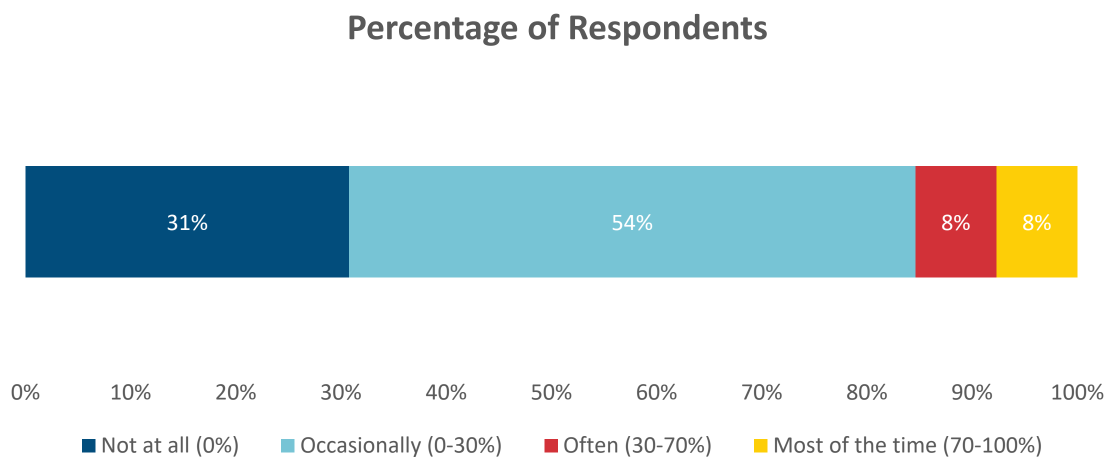
Question 7.7 - How long has your company been performing predictive modelling?



- The average number of years the survey participants have been doing analytics was 3.3 years.

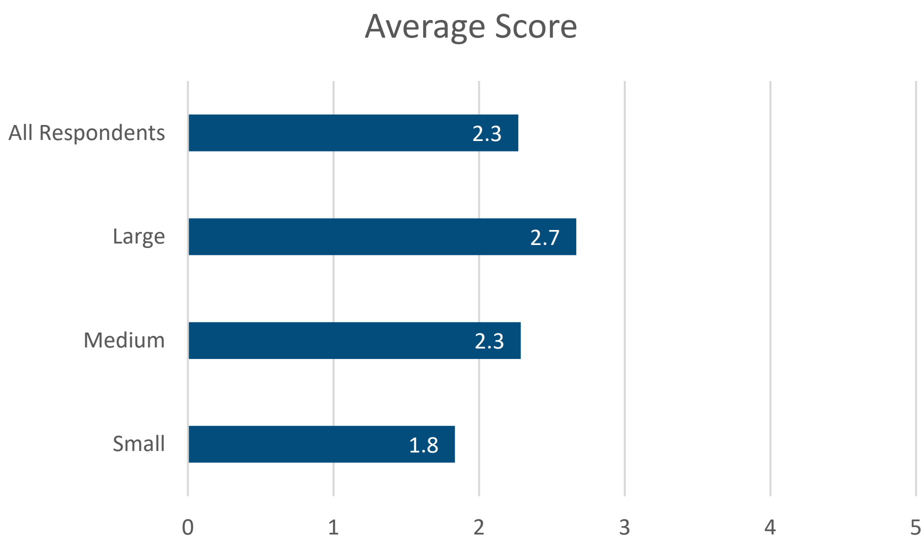
- Larger respondents tend to be performing for longer (average 4 years), and all larger respondents have been performing more than 3 years (with 67% between 3 and 5 years and 33% more than 5 years).
- None of the small respondents have been performing more than 5 years (with 80% less than 2 years).

Question 7.8 - When performing a new analytics application that was not previously executed, to what extent are you using external consultants?



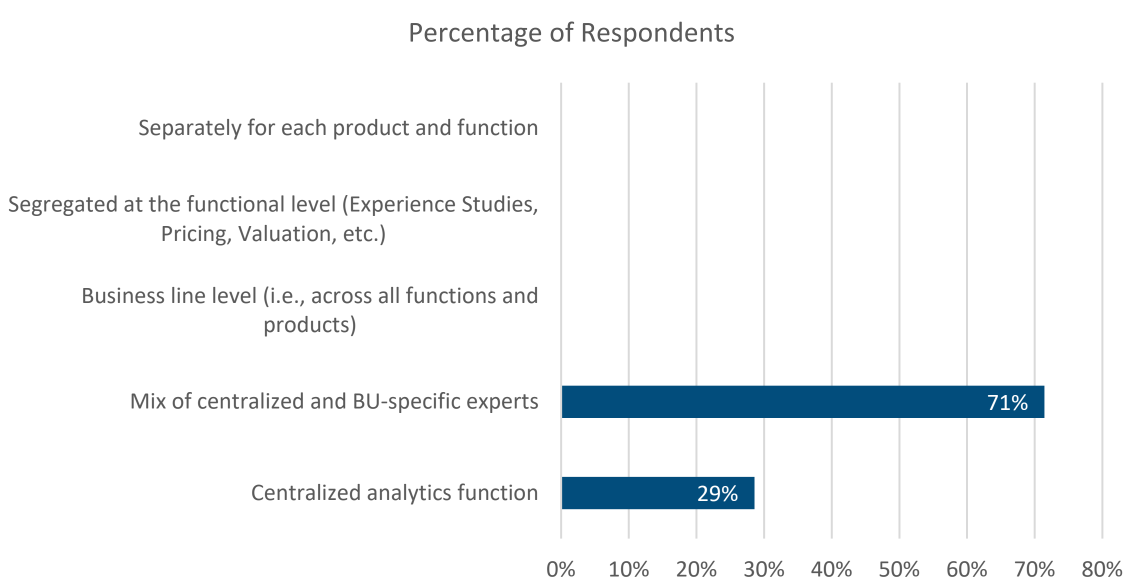
- Large survey participants will only occasionally hire consultants, whereas small participants are more likely to hire them.
- For the majority of the survey participants, the predictive modelling function has stayed internal for most of their modelling efforts, instead of being outsourced to an external consultant/IT vendor. This implies that predictive modelling is becoming a necessity for the actuarial toolkit. Actuaries will be performing this work, and given the general trends of the industry, it appears that it will be a valuable skill to have in the future.
- However, almost 70% of participants have at least occasionally used external consultants when performing a new analytics application.

Question 7.9 - How difficult is it to hire and retain analytics experts for internal positions? [1 = Extremely Difficult / 5 = Easy]



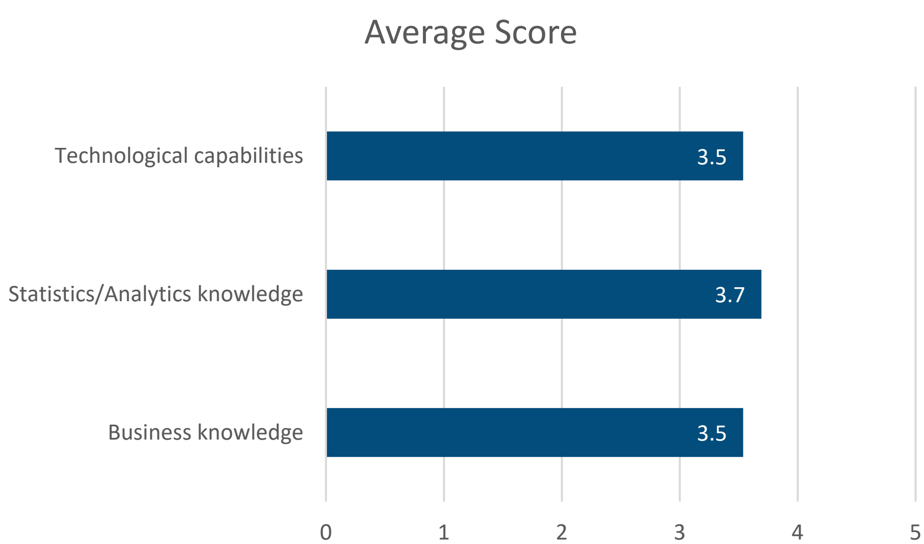
- Overall, the difficulty to hire and retain analytics experts had a score of 2.3. The general consensus is that it is moderate/difficult to find talent for predictive modelling.
- It is observed to be easier for larger respondents than smaller respondents to hire analytics experts.
- Many participants indicated that it is harder to hire the right people (with the correct balance of technical abilities and business knowledge) than it is to retain them.
- Although it was not the case for large survey participants, a few respondents indicated difficulties in hiring experts.

Question 7.10 - How are your analytics capabilities organized throughout your organization?



- Seventy-one percent of respondents indicated using a mix of centralized and BU-specific (specialists in decentralized groups) experts and 29% indicated using a centralized function.
- No significant difference was noted between size groups.
- Reinsurers tend to be more centralized than direct writers.
- It does not appear that there is much segregation in the analytics functions for the survey participants. It is worth noting that even though the data was indicated as being highly segmented, the analytics function is not.

<u>Question 7.11 - How would you rank the following aspects of your analytics talent:</u>



- Overall the analytics talent was assessed with scores of 3.5 for technological capabilities, 3.7 for analytics knowledge, and 3.5 for business knowledge.
- Larger respondents tend to have better scores, with 4.3, 4.7, and 3.3 respectively.
- While large and medium respondents assessed the business knowledge lower than other aspects, the small respondents assessed their technological capabilities lower than their business knowledge.

Summary – Analysts

Large respondents place a much higher emphasis on FTE assignment. Large respondents tend to have performed analytics for much longer than the respondents in other size groups. These facts appear to point to higher priority being given by larger respondents than by the respondents from other size groups.

Overall, close to 50% of the predictive analytics resources are business experts versus data architects/statisticians. Of these business experts, 75% are actuaries.

The most popular software is R and Python for modelling work and R and Tableau for visualization work.

Large and medium respondents assessed capabilities for business knowledge lower than other aspects (technical capabilities and statistics/analytics knowledge), which indicates the need to invest in training people about the business or hiring such people with analytics capabilities.

Small respondents assessed their technological capabilities lower than their business knowledge, which indicates the need to invest in technology and related training.

Strengths and Weaknesses Identified

- Most survey participants do not have a centralized data repository (either fully centralized or partly centralized), and many participants, particularly medium and small ones, do not rate their data as particularly complete or accurate (3 out of 5).
- Most respondents have explored or plan on exploring the use of predictive analytics in experience studies.
- Most respondents, particularly outside of the large ones, have many gaps in the types and breadth of applications, particularly on marketing, retention management, and distribution–client matching, and even in accelerated/automated UW.
- Most survey participants have not thought about the use of predictive analytics to improve internal operations (e.g., workforce analytics, use of NLP/NLG to speed up processes and reduce error rates).
- Many respondents have not thought about a standardized development environment with automated testing procedures.
- Aside from large respondents and a couple of medium respondents, analytics has not been rated as the highest priority (generally ranking 3 of 5).
- Many survey participants do not have strong support from leadership pushing analytics initiatives in their companies.
- Only half of the survey participants have explored machine learning or deep learning techniques.
- Survey participants have been struggling to hire analytics experts that fit business needs.
- Most analytics experts (e.g., statisticians and data scientists) have strong technical ability but do not have either the appropriate business knowledge or the ability to communicate ideas to non-technical audiences across the business.

References

Field	Specialization	Name	Journal/ Periodical/ Conference Title	Author	Date
All	All	How Artificial Intelligence and Machine Learning Can Impact Market Design	National Bureau of Economic Research	Paul R. Milgrom, Steve Tadelis	Jan 2018
All	All	Predicting Emergency Room Frequent Flyers; Producing Actionable Insights from Predictive Models Built Upon Condensed Electronic Medical Records; Risk Segmentation: Application of Predictive Modeling in Life Underwriting	Predictive Analytics 2014 Call For Articles	Joseph Randazzo, J. Patrick Kinney; Sheamus Kee Parkes; Richard Xu, Minyu Cao, Scott Rushing	2014
Insurance – All	All	Advanced Analytics for Insurance	Ernst & Young	Ernst & Young	2013
Insurance – All	All	Anticipating Events – Using Member-level Predictive Models to Calculate IBNR Reserves	*The Actuary Magazine*	Anders Larson, Jack Leemhuis, and Michael Niemerg	July 2018
Insurance – All	All	Data Science Landscape in the Insurance Industry	ETH Zurich	Stefano Perfetti	Dec 2017
Insurance – All	All	Predictive Data Analytics for Claims in Insurance Industry	Infosys	Infosys	June 2018
Insurance – All	All	Predictive Analytics White Paper	The Digital Insurer	Charles Nyce	2007
Insurance – All	All	300 Years of Data Analytics in Life Insurance	Financial Services Forum, Actuaries Institute	Matt Ralph and Avanti Patki	May 2016
Insurance – Life & Health	All	Predictive Modeling Applications for Life and Annuity Pricing and Underwriting	SOA 2013 Life & Annuity Symposium	Jonathan P. Polon, Qichun (Richard) Xu	May 2013
Insurance – Life & Health	All	Predictive Analytics and Accelerated Underwriting Survey Report	SOA	Predictive Analytics and Accelerated Underwriting Subcommittee of the SOA Committee on Life Insurance Mortality and Underwriting Surveys	May 2017

Insurance – Life & Health	All	Report of the SOA Predictive Modeling Survey Subcommittee	SOA	SOA Predictive Modeling Survey Subcommittee	Jan 2012
Insurance – Life & Health	All	Comparing Policyholder Efficiency in Variable Annuity Lapses; Insurance Product Recommendation System; Machine Reserving: Integrating Machine Learning into Your Reserve Estimates; Variable Selection Using Parallel Random Forest for Mortality Prediction in Highly Imbalanced Data	Predictive Analytics 2016 Call For Essays	Jenny Jin, Vincent Embser; Kailan Shang; Dale Cap; Mahmoud Shehadeh, Rebecca Kokes, Guizhou Hu	2016
Insurance – Life & Health	All	Predictive Modeling for Life Insurance – Ways Life Insurers Can Participate in the Business Analytics Revolution	Deloitte Consulting	Mike Batty, Arun Tripathi, Alice Kroll, Cheng-sheng Peter Wu, David Moore, Chris Stehno, Lucas Lau, Jim Guszcza, Mitch Katcher	April 2010
Insurance – Life & Health	All	Application of Predictive Modeling Techniques in Measuring Policyholder Behavior in Variable Annuity Contracts	Insights (Towers Watson)	Guillaume Briere-Giroux, Jean-Felix Huet, Robert Spaul, Andy Staudt, David Weinsier	April 2010
Insurance – Life & Health	All	Predictive Analytics in Life Insurance	Advances in Predictive Analytic Conference	Jean-Yves Rioux, Kevin Pledge, Ian Bancroft, Eugene Wen	Dec 2017
Insurance – Life & Health	All	Predictive Analytics in Life Insurance; Predictive Modeling with Prescription Histories	ACLI Annual Conference	Sam Nandi; Eric Carlson	Oct 2017
Insurance – Life & Health	All	Predictive Analytics Global Survey Results – Still Room to Grow for Life & Health Insurers	Gen Re	Guizhou Hu	Oct 2017
Insurance – Life & Health	All	Transforming the Life Insurance Industry – Lifestyle based Analytics	Actuaries Summit	John King and Kim Cohen	May 2013
Insurance – Life & Health	All	Analytics: A Powerful Tool for the Life Insurance Industry Using Analytics to Acquire and Retain Customers	Capgemini	Capgemini	2011

Insurance – Life & Health	All	From Mystery to Mastery: Unlocking the Business Value of Artificial Intelligence in the Insurance Industry	Deloitte Digital	Deloitte Digital	Nov 2017
Insurance – Life & Health	All	Predictive Analytics in Life Insurance – Predictive Modeling with Prescription Histories	ACLI Annual Conference	Sam Nandi and Eric Carlson	Oct 2017
Insurance – Life & Health	All	Predictive Analytics and Accelerated Underwriting Follow-up Survey Report	SOA	Allen M. Klein, Roland P. Fawthrop, Gordon A. Gibbins, William M. Tilford, David N. Wylde	March 2018
Insurance – Life & Health	All	Modeling of Policyholder Behavior for Life Insurance and Annuity Products – A Survey and Literature Review	SOA, LIMRA	PwC (Jason Campbell, Michael Chan, Kate Li, Louis Lombardi, Lucian Lombardi, Marianne Purushotham)	2014
Insurance – Life & Health	All	Accuracy of Claims-Based Risk Scoring Models	SOA	Geof Hileman, Spenser Steele	Oct 2016
Insurance – Life & Health	All	Insurers Flock to Analytics to Stretch Underwriting Limits	*Insurance Journal*	Alain Theriault	Mar 2017
Insurance – Life & Health	Fraud	Machine Learning Applications in Insurance Underwriting Predicting Applicant's Smoking Propensity	CIA 2017 Annual Meeting	Nitin Nayak, Swiss Re	2017
Insurance – Life & Health	Health Care	10 Promising AI Applications in Health Care	*Harvard Business Review*	Brian Kalis, Matt Collier, Richard Fu	May 2018
Insurance – Life & Health	Health Care	Predictive Modeling with Consumer Data	*The Actuary*	Ksenia Draaghtel	Nov 2011
Insurance – Life & Health	Health Care	Healthcare and Artificial Intelligence: Saving Lives and Costs	Datameer Blog Post	Samantha Leggat	May 2018
Insurance – P&C	All	Predictive Modeling in P&C Insurance	Iowa Actuaries Club	Gilbert Korthals	Dec 2016
Insurance – P&C	All	2016 Property-Casualty Industry Observations and Market Trends in Analytics	BACE Fall Meeting	Ward Group (Aon)	Sept 2016
Insurance – P&C	All	Property and Casualty Insurance Predictive Analytics in SAS	SAS	Mei Najim	2017
Insurance & Banking	All	Artificial Intelligence and Machine Learning in Financial Services	Financial Stability Board	Financial Stability Board	Nov 2017

Other	All	Artificial Intelligence and Digital Banking	Mapa Research	Mapa Research	Nov 2016
Other	All	Predictive Modeling – Turning Big Data Into Big Opportunities	CIA	CIA Predictive Modeling Committee	Jun 2018
Other	All	Predictive Analytics for Marketing – What's Possible and How it Works	Tech Emergence	Daniel Faggella	Jun 2018

Acknowledgements of Survey Participants and Interviewees

We would like to thank the participants of this survey, without whom we would not have been able to obtain an overall view of current and future practices in the Canadian life and health insurance industry and summarize the insights provided in this report. The participants included:

- BMO Insurance
- Desjardins
- Empire Life
- Great-West Life
- Industrial Alliance
- Ivari
- Manulife Financial
- Munich Re
- Optimum Re
- Partner Re
- RBC Insurance
- RGA
- SCOR
- SSQ Financial Group
- Sun Life Financial

We also extend our thank you to those who agreed to be interviewed for the research section:

- Denis Repin
- Shanil Ebrahim
- Peter Wu

Appendix A: Survey Questionnaire

Profiling

#	Question	Response
0	What company do you represent?	Pre-filled by Deloitte
1.1.	What is/are your name(s)?	
1.2.	What is/are your current title(s)?	
1.3.	In which department(s) do you currently work?	

General

#	Question	Response
2.1.	Identify your key challenges in the effective use of predictive analytics and how you have addressed them	
2.2.	What is your company trying to learn through predictive analytics that cannot be obtained by traditional methods?	

Data

#	Question	Response
3.1.	What best describes the centralization of your data? If a mix, what is the rough proportion allocation across each?	a. Data is stored/accessed separately for each product and function b. Data is stored/accessed separately at the functional level (Experience Studies, Pricing, Valuation, etc.) c. Data is stored/accessed separately at the business line level (i.e., across all functions and products) d. Combination of (c) and (e) – mix of business line specific and organization-wide data storage/access e. Data is stored/accessed at a single point for all data in the organization
3.2.	On a scale from 1 to 5, how would end users rate the completeness and accuracy of your data?	1 = lowest 5 = highest
3.3.	Provide an approximate percentage of your data (excluding P&C, if writing both) that is in each of the following categories	a. Unstructured (voice, image, scanned documents) b. Free-form text c. Code that requires a legend to interpret

		d. Scalars/Values that requires no legend e. Arrays, tables, cubes
3.4.	How is data accessed across the organization? (Select all that apply)	a. Indirect access through requests to specific individuals (e.g.,, IT) who provide data after more than one business day b. Indirect access through requests to specific individuals (e.g.,, IT) who provide data within one business day c. Direct access to several databases through a query-based language (e.g.,, SQL Server, MySQL, etc.) d. Direct access to one or two databases through a query-based language (e.g.,, SQL Server, MySQL, etc.) e. Direct access with single query using user-friendly front-end
3.5.	What types of 3rd party data does your organization currently use for analytics? (Select all that apply)	a. Demographic b. Geographic c. Claims & Medical d. Financial e. Credit f. Lifestyle g. Mood/Attitude h. Behavioural i. Economic
3.6.	Have you started collecting data through new technologies (e.g.,, Fitbit) to augment the use of traditional application/underwriting/reinsurance sources of data?	a. Yes, started collecting and currently using the data b. Yes, started collecting but not using the data yet c. No, have already investigated and decided not to collect/use this type of data d. No, but are currently investigating e. No, not currently investigating but plan on doing so f. No, with no plans on investigating

3.7.	How do you store your internal data (check all that apply)?	a. Internal Proprietary System b. Excel spreadsheets c. Flat Files (CSV, Text files) d. MS Access e. SQL Server f. Hadoop (or similar) g. Other (please specify)
3.8.	How do you store your external data (check all that apply)?	a. Internal Proprietary System b. Excel spreadsheets c. Flat Files (CSV, Text files) d. MS Access e. SQL Server f. Hadoop (or similar) g. Other (please specify)

Enterprise

#	Question	Response
4.1.	Select the governance aspects and policies surrounding data used in analytics for your organization (check all that apply):	a. Data sources' privacy b. Data security c. Data update privileges d. Policies around impact assessment of changes to data e. Data accuracy/quality standards f. Data standardization g. Data dictionaries h. Controls around ETL ("Extract, Transform, and Load") i. Other (please specify)
4.2.	Select the governance aspects and policies that impact the analytics modelling efforts for your organization (check all that apply):	a. Model validation b. Code review c. Version control d. Identified evaluation metrics for performance of models e. Other (please specify)
4.3.	Select the governance aspects and policies relating to software and technology used in analytics applications for your organization (check all that apply):	a. No policy exists b. List of approved/preferred/disapproved software c. Software upgrade policy d. Other (please specify)
4.4.	If the analytics function is not fully centralized, please assess the strength of the skills & resources coordination across the organization	1 = extremely weak 5 = extremely strong

Leadership

#	Question	Response
5.1.	Does your organization have an executive responsible for data and what is his/her title?	
5.1a	If there is no executive responsible for data, how much does leadership understand data (quality, structure, safeguards, etc.)?	1 = Leadership has little to no understanding of data 5 = Leadership has full understanding of data
5.2.	Does your organization have an executive responsible for predictive analytics and what is his/her title?	
5.2a	If there is no executive responsible for predictive analytics, how much does leadership understand its value-added opportunities and the related competitors' initiatives?	1 = Leadership has little to no understanding of predictive analytics 5 = Leadership has full understanding of predictive analytics
5.3.	Who makes the final decision on the prioritization of predictive analytics initiatives?	a. Leadership (Board, CEO, EVPs) b. C-Suite Executive c. Below C-Suite Analytics Executives (i.e., VP+) d. Below C-Suite Non-Analytics focused Executives (i.e., VP+) e. Other
5.4.	How does predictive analytics rank within your organization with respect to project priorities	1 = Low priority 5 = Most important investment for the future
5.4a	If predictive analytics is not the highest priority (i.e., scored as 5 in prior question), what are the key reasons (check all that apply)?	a. Lack of infrastructure or computing capabilities b. Low quality of internal data c. Difficulties in accessing internal data d. Challenges in obtaining and aggregating data from multiple internal data sources e. Difficulties in finding and collecting relevant external data f. Privacy issues related to 3rd-party data g. Lack of internal talent & expertise h. Lack of executive support i. Other (please specify)

Targets

#	Question	Response
6.1.	For what current and past applications are you using analytics? "Applications" are defined as cases where a predictive (or ML, AI, etc.) model has been calibrated for prediction purposes, to identify drivers/explanatory variables, etc. How do you rate each of these applications in terms of perceived effort to implement and perceived value to the company?	List of applications Effort and value rated using: L- L L+ M- M M+ H- H H+
6.2.	For what applications are you planning on using analytics in the next year? How do you rate each of these applications in terms of perceived effort to implement and perceived value to the company?	List of applications Effort and value rated using: L- L L+ M- M M+ H- H H+
6.3.	Approximately how many analytics projects have you attempted in the last year?	
6.4.	Approximately how many new analytics applications that were not previously executed (e.g., producing lapse analytics for the first time) are anticipated to be tackled in the next year?	

Analysts

#	Question	Response
7.1.	What techniques are currently being used?	a. Summary statistics (mean, median, distributions, correlations) b. Inferential regression (model calibrated using entire dataset) c. Predictive modelling with regression/GLMs d. Predictive modelling with non-linear models (decision trees, random forest, boosted trees, etc.) e. Machine Learning/Neural Networks/ Deep Learning
7.2.	How many Full Time Equivalents are working on analytics?	
7.3.	Approximately what are the proportions of FTEs in each of the following categories? Include resources borrowed from head office, other BUs, etc.	a. Business experts – Actuaries b. Business experts – non-Actuaries c. Data Architect/Engineers d. Statisticians/Analytics experts e. Computer scientists/IT f. Other (please specify)
7.4.	On each predictive analytics project, to what degree do you integrate the skillsets of Business experts (both Actuarial & Non-Actuarial), Statisticians/Analytics experts, and IT?	a. Not at all (0%) b. Occasionally (0-30%) c. Often (30-70%) d. Most of the time (70-100%) e. All of the time (100%)
7.5.	What software are you using for predictive analytics modelling (select all that apply)?	a. R b. Python c. SAS d. Other (please specify)
7.6.	What software are you using for exploratory data analysis and visualization (select all that apply)?	a. Tableau b. Qlikview c. Power BI d. R e. Python f. Other (please specify)
7.7.	How long has your company been performing predictive analytics?	

7.8.	When performing a new analytics application that was not previously executed (e.g., producing lapse analytics for the first time), to what extent are you using external consultants?	a. Not at all (0%) b. Occasionally (0-30%) c. Often (30-70%) d. Most of the time (70-100%) e. All of the time (100%)
7.9.	How difficult is it to hire & retain analytics experts for internal positions?	1 = Extremely difficult 5 = Easy n/a – no internal positions
7.10.	How are your analytics capabilities organized throughout your organization?	a. Separately for each product and function b. Segregated at the functional level (Experience Studies, Pricing, Valuation, etc.) c. Business line level (i.e., across all functions and products) d. Mix of centralized and BU-specific experts e. Centralized analytics function
7.11.	How would you rank the following aspects of your analytics talent: a. Technological capabilities b. Statistics/Analytics knowledge c. Business knowledge	1 = Low through 5 = High I don't know n/a – no FTEs working on analytics

Appendix B: Definitions

For the purposes of the report, we will use the following definitions:

Deep Learning: in the context of this survey, *deep learning* refers to enhancing an artificial neural network through multiple hidden layers of nodes.

External Data: data that was not originally captured by your own organization and is typically collected or purchased from a third party. This survey uses *external data* and *third-party data* interchangeably.

Free-form text: data that is stored without any restrictions on form. For example, First Name and Last Name are treated as free-form text since there is no predefined set of values.

Front-end: an access point to a database (a database is sometimes referred to as "back-end") which does not require programming knowledge and is typically user-friendly. This allows various personnel within the firm to have access to the same data without requiring the use of the back-end query language.

Generalized Linear Model (GLM): the broad group of linear regression models that allow for dependent variables to have an error distribution that is not normal. This includes ordinary least squares (OLS), logistic regression (logit), probit regression, poisson regression, and many others.

Inferential Statistics: using historical data for general inferences regarding past events. This differs from predictive analytics in that you are not using the results to predict a future event. You are simply using the results of your analysis to understand what happened previously.

Neural Network: in the context of this survey, a neural network is any model which is theoretically inspired by the biological neural network of an animal brain.

Predictive Analytics: applying statistical techniques to calibrate a model using historical data to make predictions about future or otherwise unknown events. This survey uses *predictive analytics*, *predictive modelling*, and *machine learning* interchangeably. In some cases *analytics* was used to refer to the broader practice of modelling historical data, which contains *predictive analytics*.

Unstructured Data: information that does not have a defined model, structure, or relationship to the other data in the database. Unstructured data would not be stored in a relational database where information can be linked through keys. Examples include: voice recordings, images, and scanned documents.