Investigating Applications of Data Science in UK and non-UK Actuarial Teams

A benchmarking of actuarial departments' practices
& proposed recommendations based on industry research

EAA e-Conference on
Data Science & Data Ethics

29 June 2021

EAA e-Conference on
Data Science & Data Ethics

29 June 2021

Your speakers for today

XAVIER MARECHAL

CEO, Reacfin
xavier.marechal@reacfin.com
www.reacfin.com

Reacfin

VALERIE DU PREEZ

FIA, Managing Director, Dupro Ltd
valeriedupreez@dupro.co.uk
www.dupro.co.uk |
www.actuartech.com

Dupro

ANJA FRIEDRICH

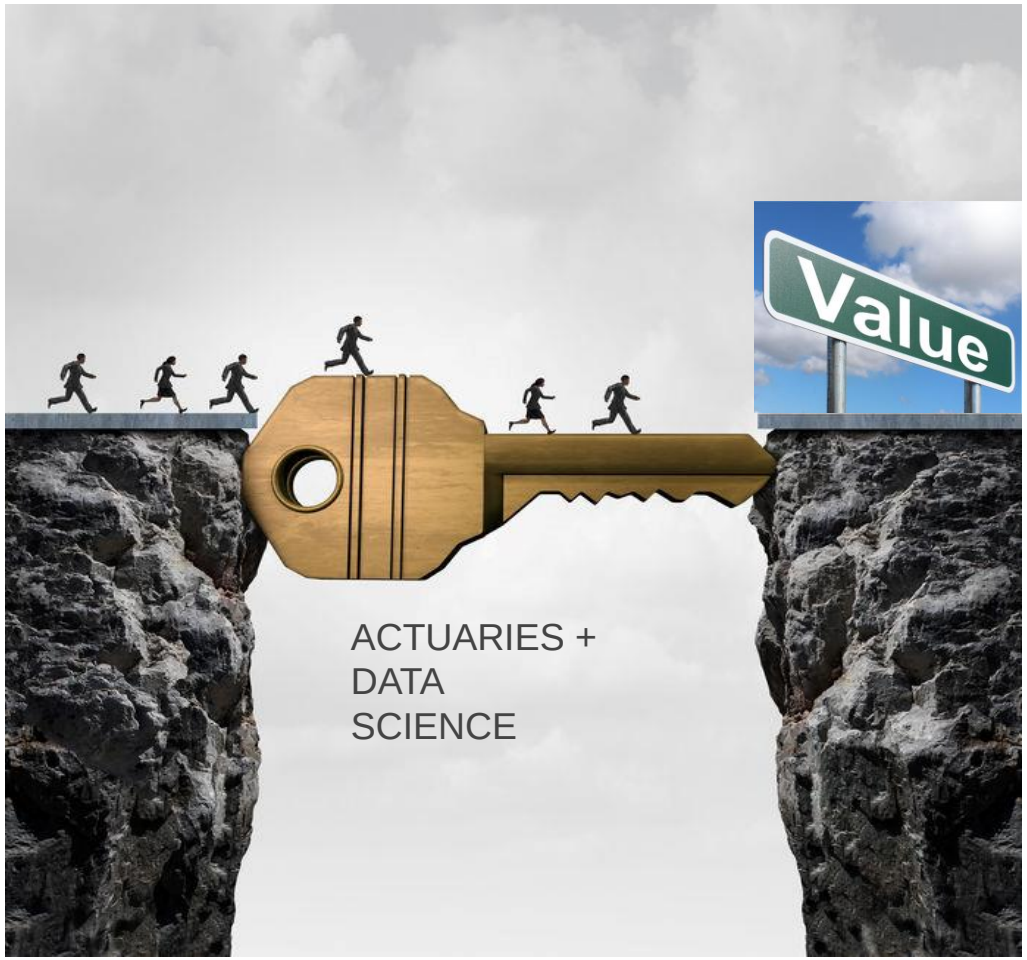
Manager, Synpulse
anja.friedrich@synpulse.com
www.synpulse.com

synpulse

european actuarial academy

AGENDA

- **Part I:** Introduction to benchmarking exercise performed
- **Part II:** How do insurers create value using data science?
  What are the most deployed data science use cases in insurance?
- **Part III:** What is the level of maturity with regards to data?
- **Part IV:** Which tools & techniques are deployed that enable the application of data science?
- **Part V:** How to ensure optimal team performance when it comes to applying data science?
  Where are the upskilling opportunities?
- **Part VI:** What are the main challenges & opportunities in adopting data science?


Value
ACTUARIES +
DATA
SCIENCE

PART I
INTRODUCTION

european actuarial academy

KEY FOCUS AREAS OF THE BENCHMARKING EXERCISE

Actuarial Department Specific

Operating Model

- People including skill set, training
- Processes & Controls
- Systems
- Technical Nature & Maturity of Tools & Techniques
- Policies

Data

- Understanding data science pipeline
- Monetising Data
- Accountability for Data
- Data Science Pipeline Impact
- Marginal Value

Use Cases

- Application within actuarial context
- Effort/Value Analysis
- Impact of COVID-19
- Use of Machine Learning & Visualisation

Strategy

- Role of the Department
- Overcoming barriers to adopting
- Ethics
- Data Science Risks & Risk Management
- Upskilling

Organisation Specific

Benchmarking Diagram 1

BENCHMARKING – SCOPE AND PARTICIPANTS

- Data Science is defined as "The combination of business domain expertise, computer science, and knowledge of mathematics and statistics to extract meaningful insights from data."*
- We contacted around 100 insurance companies of which 42 actuarial departments from health, life and/or non-life organisations based in UK, South Africa, Switzerland, Belgium and Luxembourg participated in our Benchmarking.
- There were varied messages from those whom we contacted who were not included in the exercise including that data science is not a priority for them; that they felt their company or teams were too small; that they felt they don't have enough data; have other more important priorities or in some cases they simply did not answer.

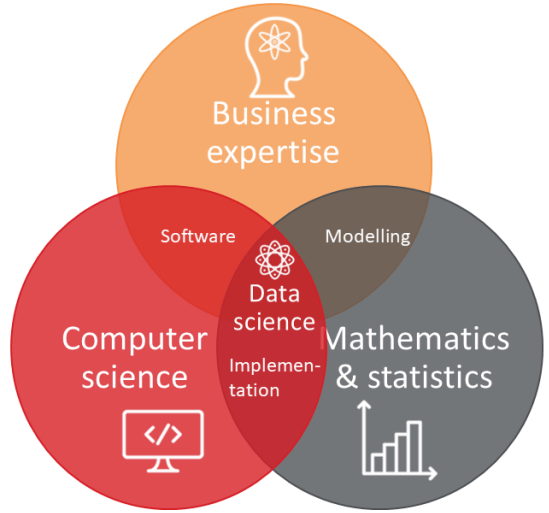
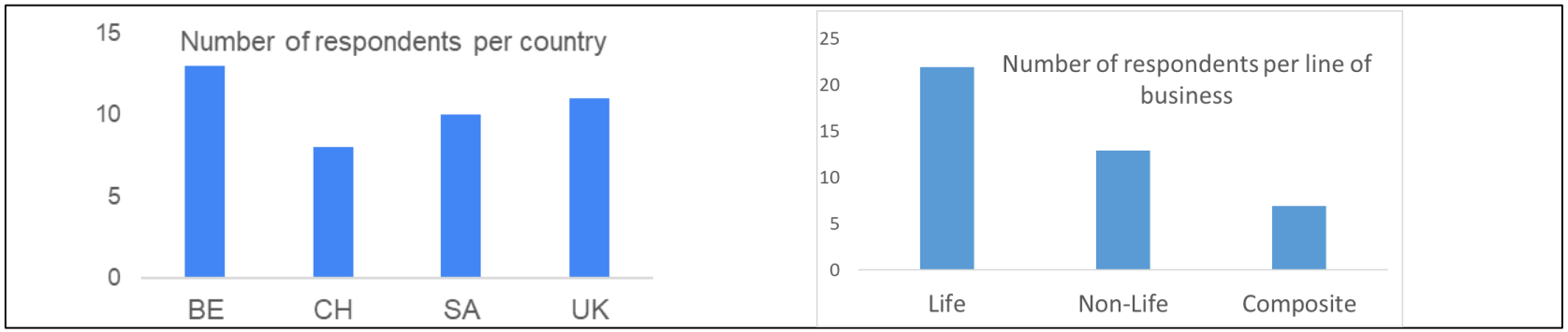


Diagram 3

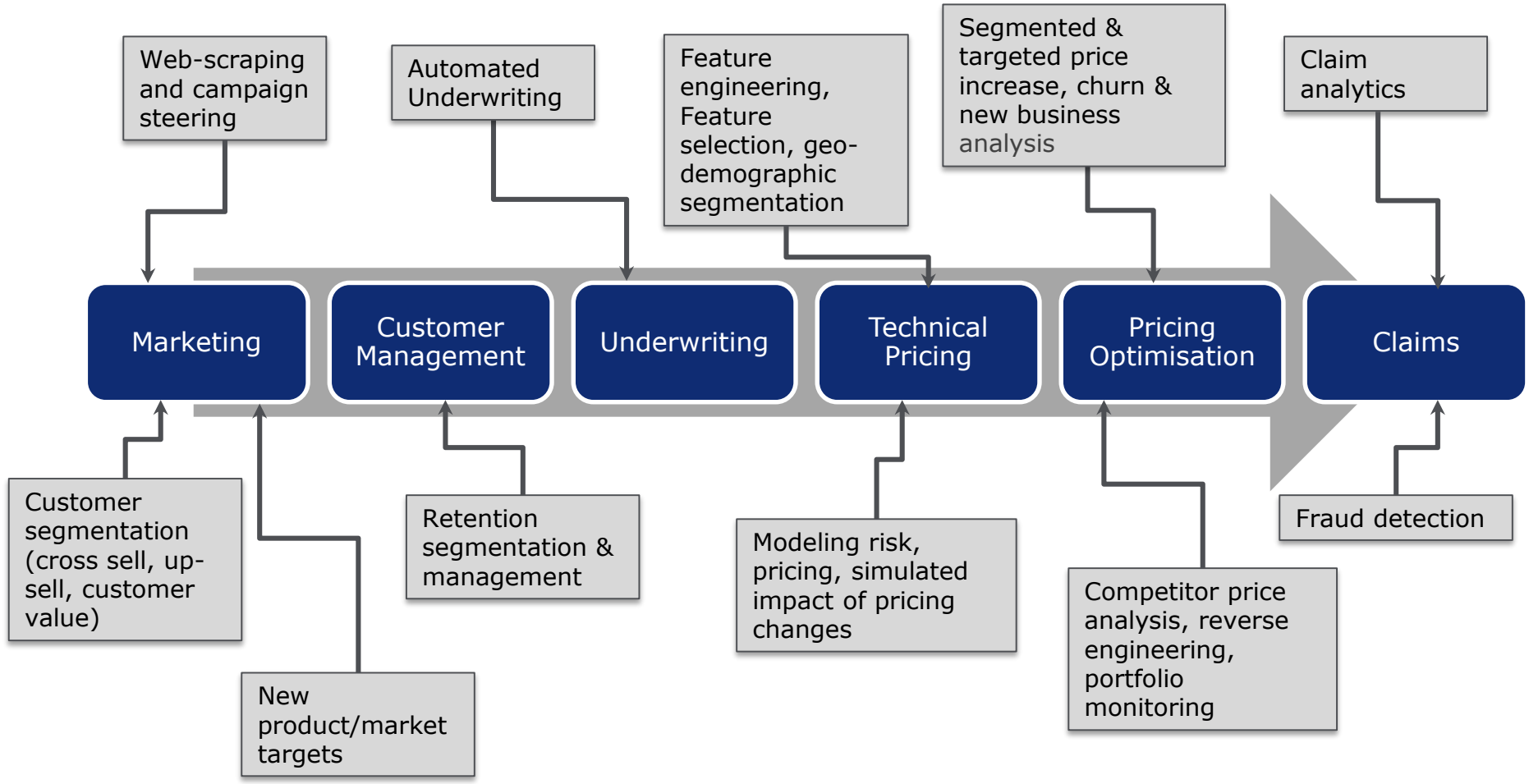


Benchmarking Diagram 2

**Institute and Faculty of Actuaries : What is Data Science?*

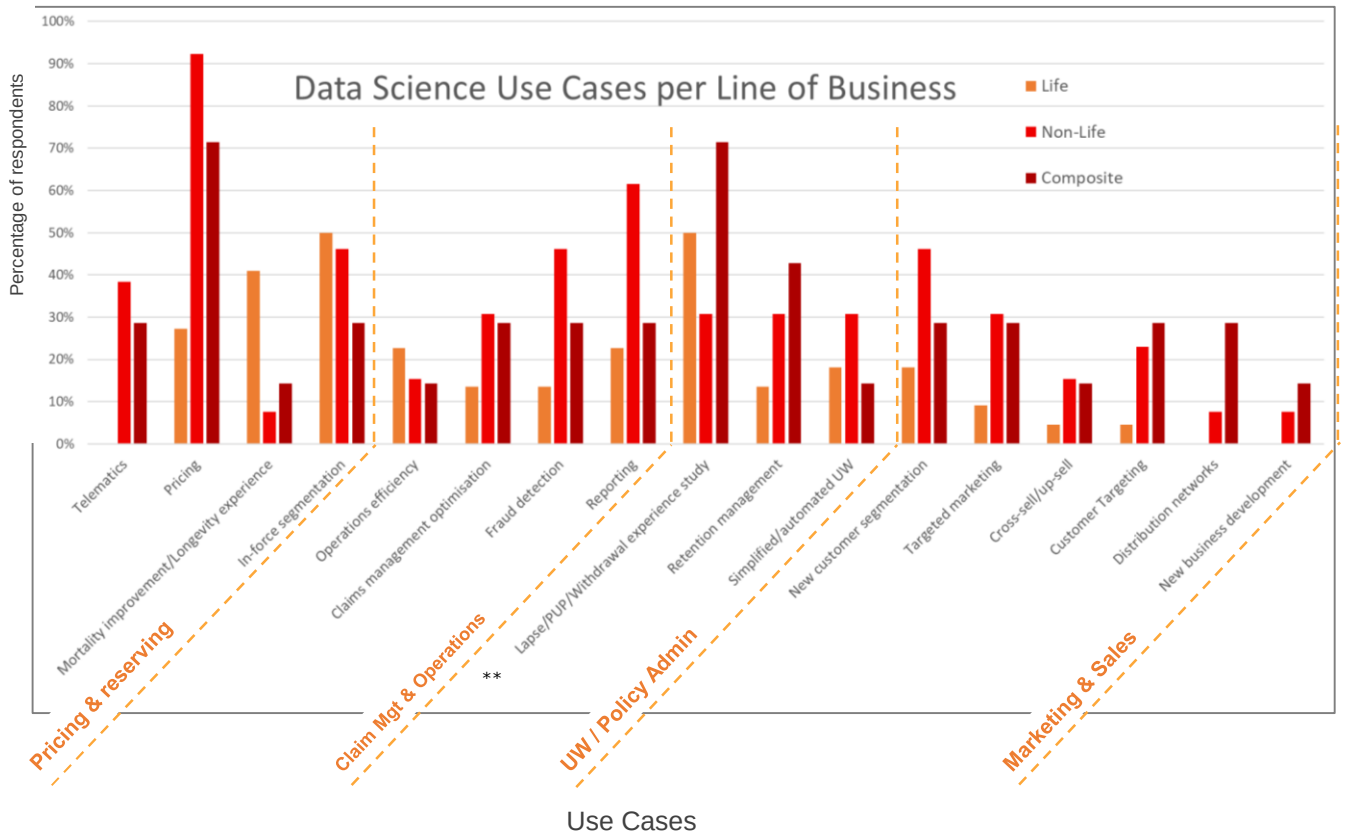
PART II
CREATING VALUE THROUGH DATA

DATA IS EVERYWHERE IN THE INSURANCE VALUE CHAIN… AND SO TOO THE OPPORTUNITIES FOR USING DATA SCIENCE



Benchmarking Diagram 3

HOW DO INSURERS CREATE VALUE USING DATA SCIENCE?



** This could be as a result of the nature of our respondents' profiles (our respondents were mainly from actuarial departments and hence may not be fully representative of the true situation at each individual company).*
*** Lapse experience study are also applicable to pricing & reserving projects*

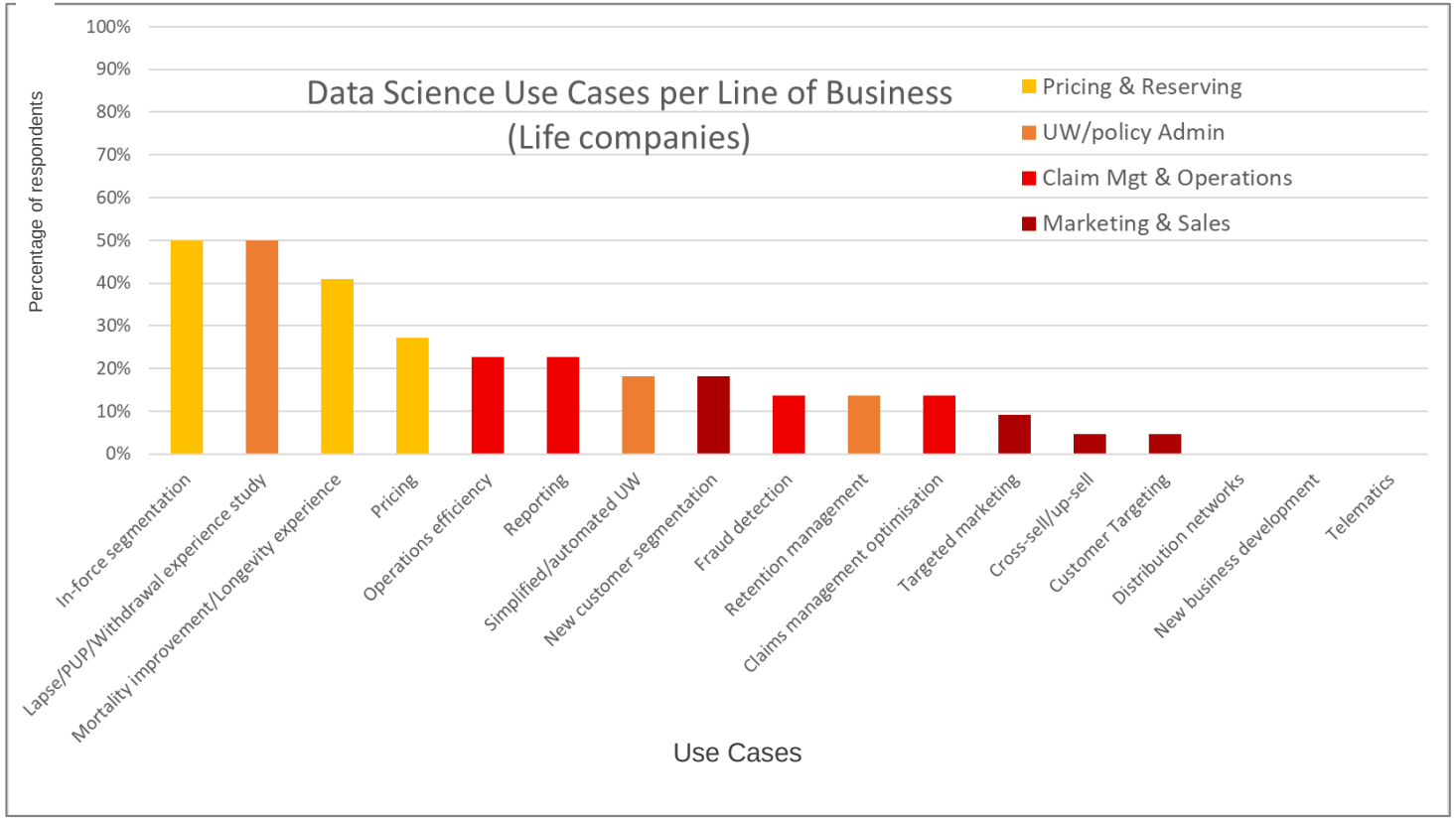
Benchmarking Diagram 4

Key take-aways

- What are the most deployed data science use cases in insurance?
  - Data science use cases include those related to insurance core functions (experience analysis, pricing, underwriting, reserving)
  - Data science use cases are not yet so widespread in upstream (marketing, sales) and downstream (claim management) activities*
- The extent to which Data Science is applied in these use cases depends on the nature of the department and specific challenges faced; as well as the specific techniques and skills applied
- Interactions with the wider business is essential to the success of use cases. Actuaries and data scientists collaboration is essential in optimal application

HOW DO INSURERS CREATE VALUE USING DATA SCIENCE?

LIFE INSURERS



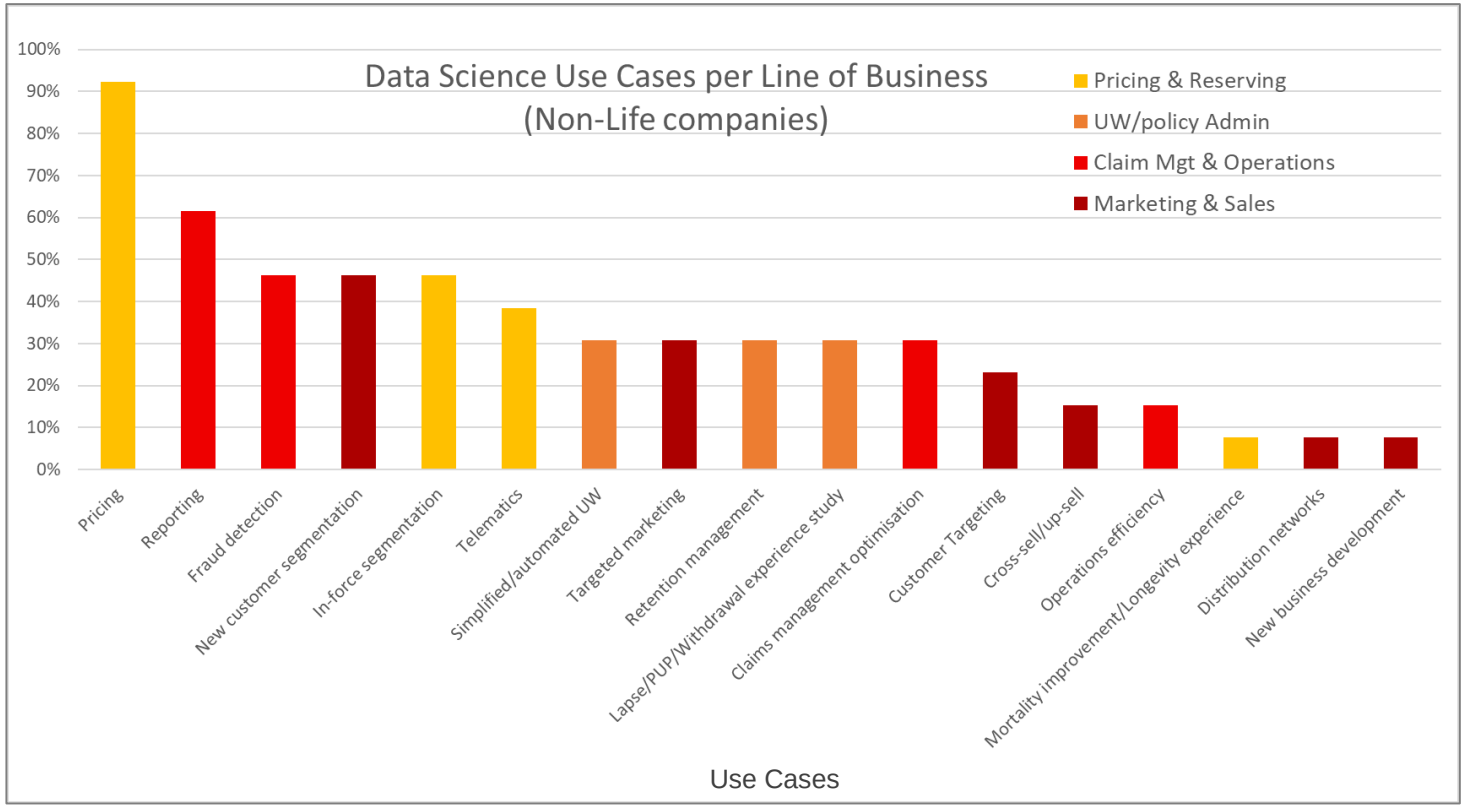
Key take-aways

- Life insurance respondents reported that pricing & reserving-related applications are popular applications of Data Science in their company.
- Lapses and mortality modelling (for pricing and reserving but also risk management) in particular; are common use cases for life insurance companies.
- Marketing-related topics are currently less developed (or the respondents are less aware of these use cases in their company).

Benchmarking Diagram 5

HOW DO INSURERS CREATE VALUE USING DATA SCIENCE?

NON-LIFE INSURERS

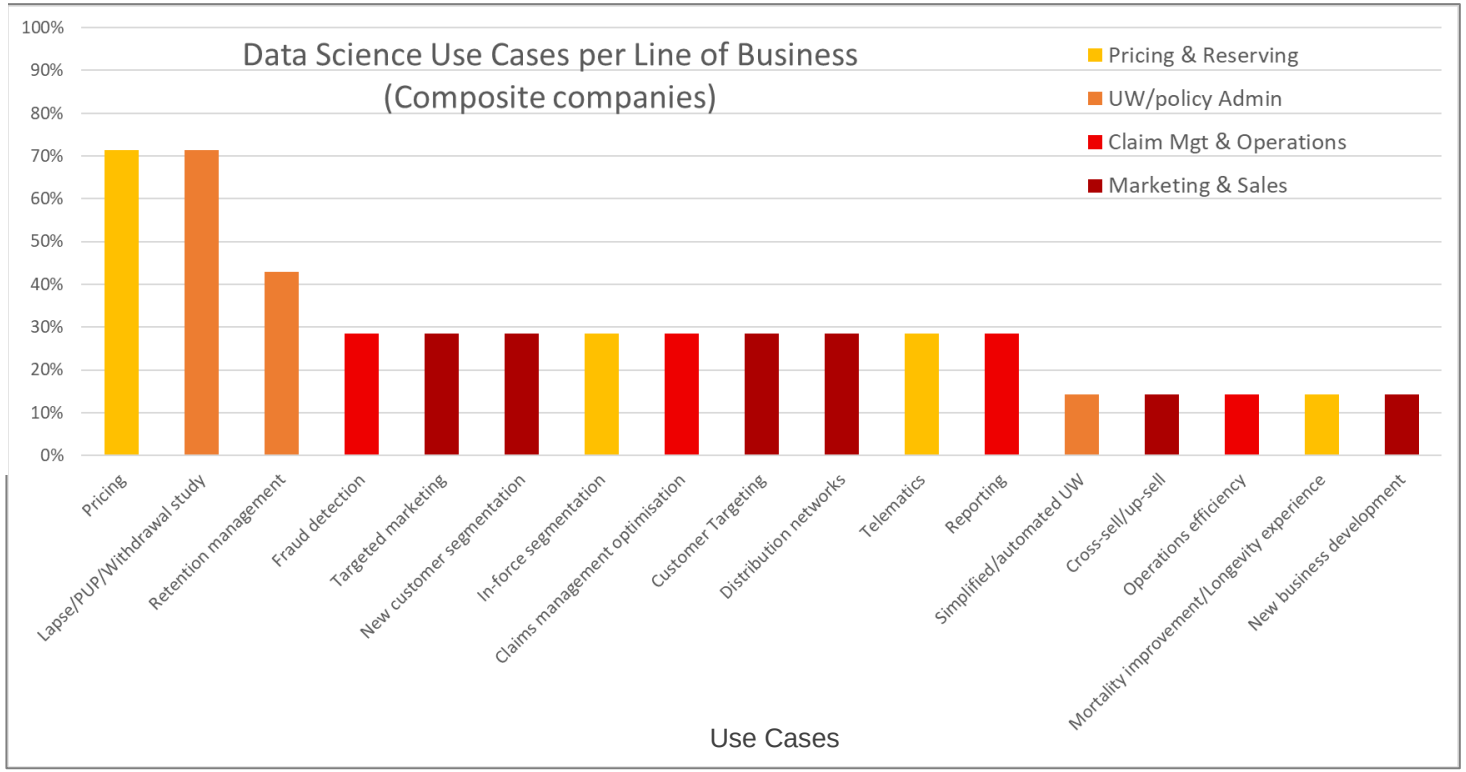


Key take-aways

- In non-life insurance, pricing is the most applied use case. The large use of Data Science in pricing departments reflects a trend observed for many years (increased segmentation, use of new data sources ,etc.)
- Claims management & claims related use cases (e.g. fraud detection) are more developed in non-life insurance given the number of claims

Benchmarking Diagram 6

COMPOSITE INSURERS



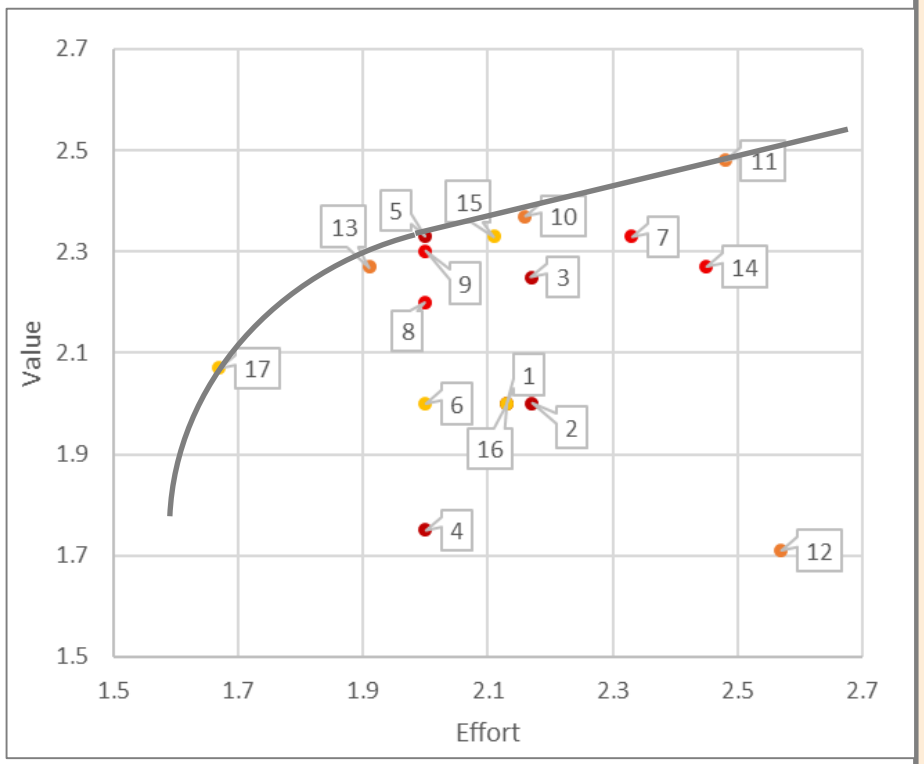
Key take-aways

- For composite (life & non-life) companies, the types of deployed uses cases are similar to those observed on the previous slides but the relative popularity are different.

Benchmarking Diagram 7

PERCEIVED VALUE GAINED FROM DATA SCIENCE RELATIVE TO EFFORT

Data Science Application of Value relative to Effort



Key take-aways

- We aimed to identify what the perceived value created by using data science relative to the effort required to reach this value, were.
- Value and effort are assessed by participants on a qualitative scale (Low = 1, Medium = 2 and High = 3). Averages of qualitative scores are then computed.

Benchmarking Diagram 8

PART III
DATA MATURITY

WHAT'S THE LEVEL OF MATURITY WITH REGARDS TO DATA?

Key take-aways

1. Use of Internal and External Data

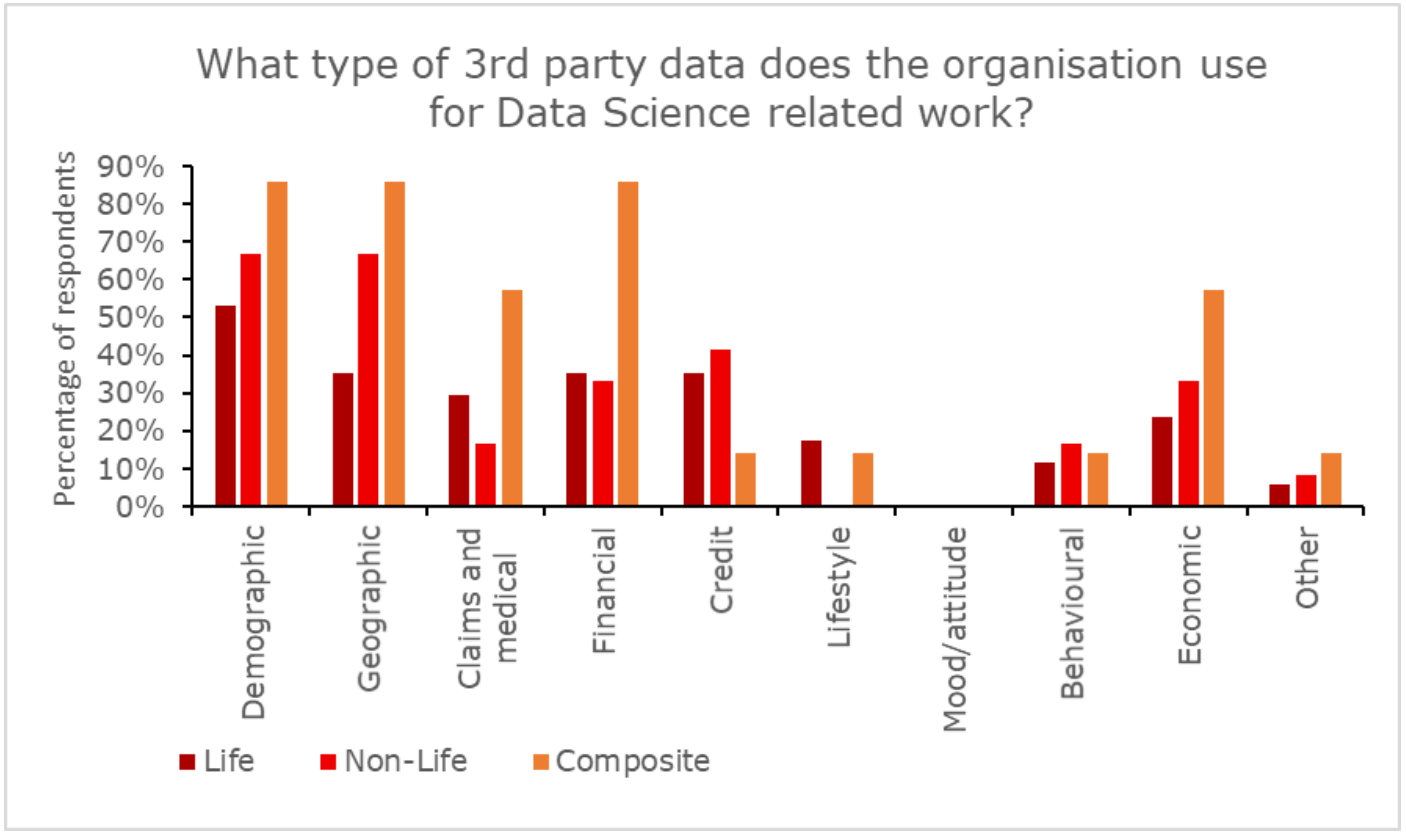
Internal
Insurers still struggle to make full use of internal data due to:

- Low quality of internal data
- Difficulties in accessing internal data
- Obtaining and aggregating data from multiple internal data sources

External
Overall goal to enrich internal data with external data

2. Most common types of External Data

- Demographic
- Geographic
- Financial
- Claims and medical
- Economic



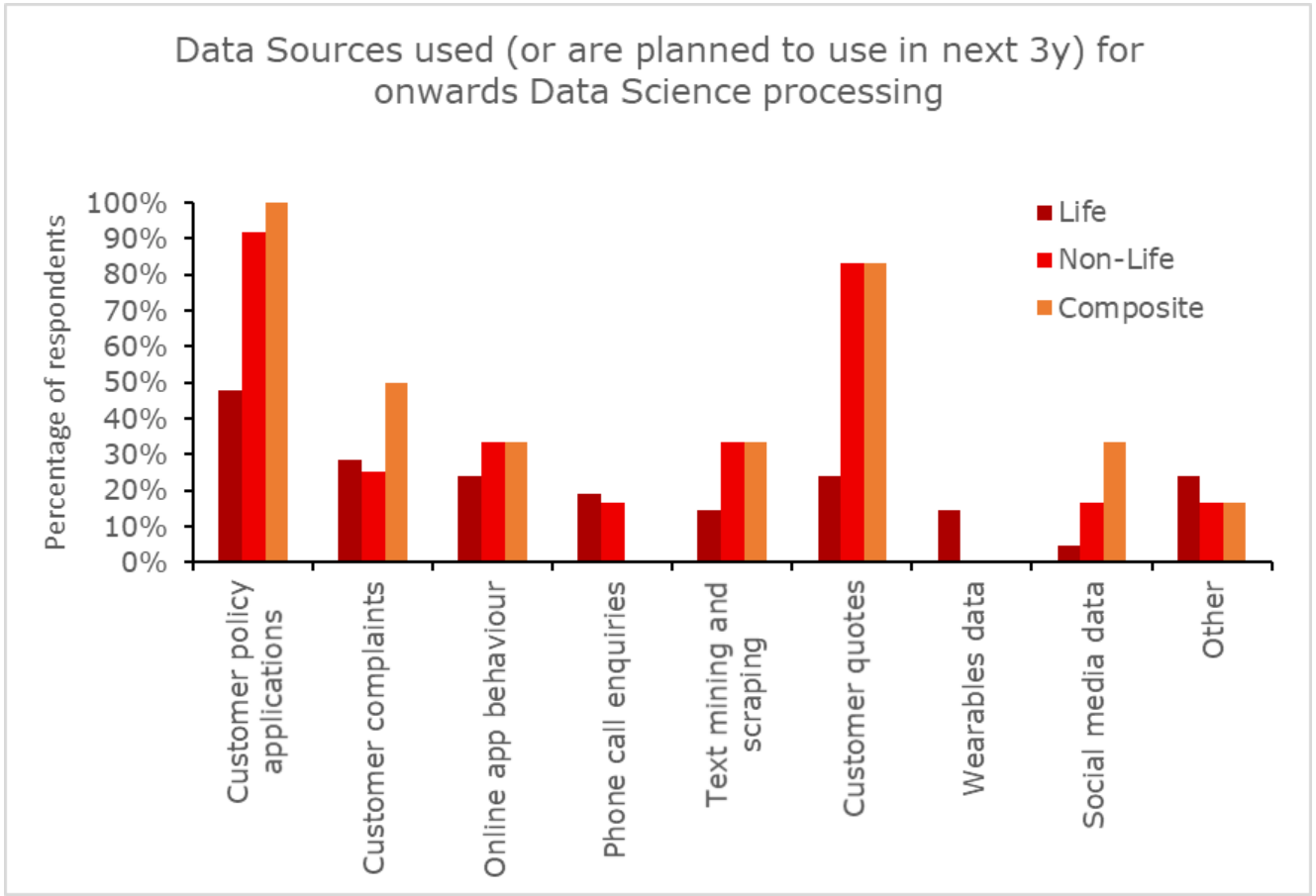
Benchmarking Diagram 9

WHAT'S THE LEVEL OF MATURITY WITH REGARDS TO DATA?

Key take-aways

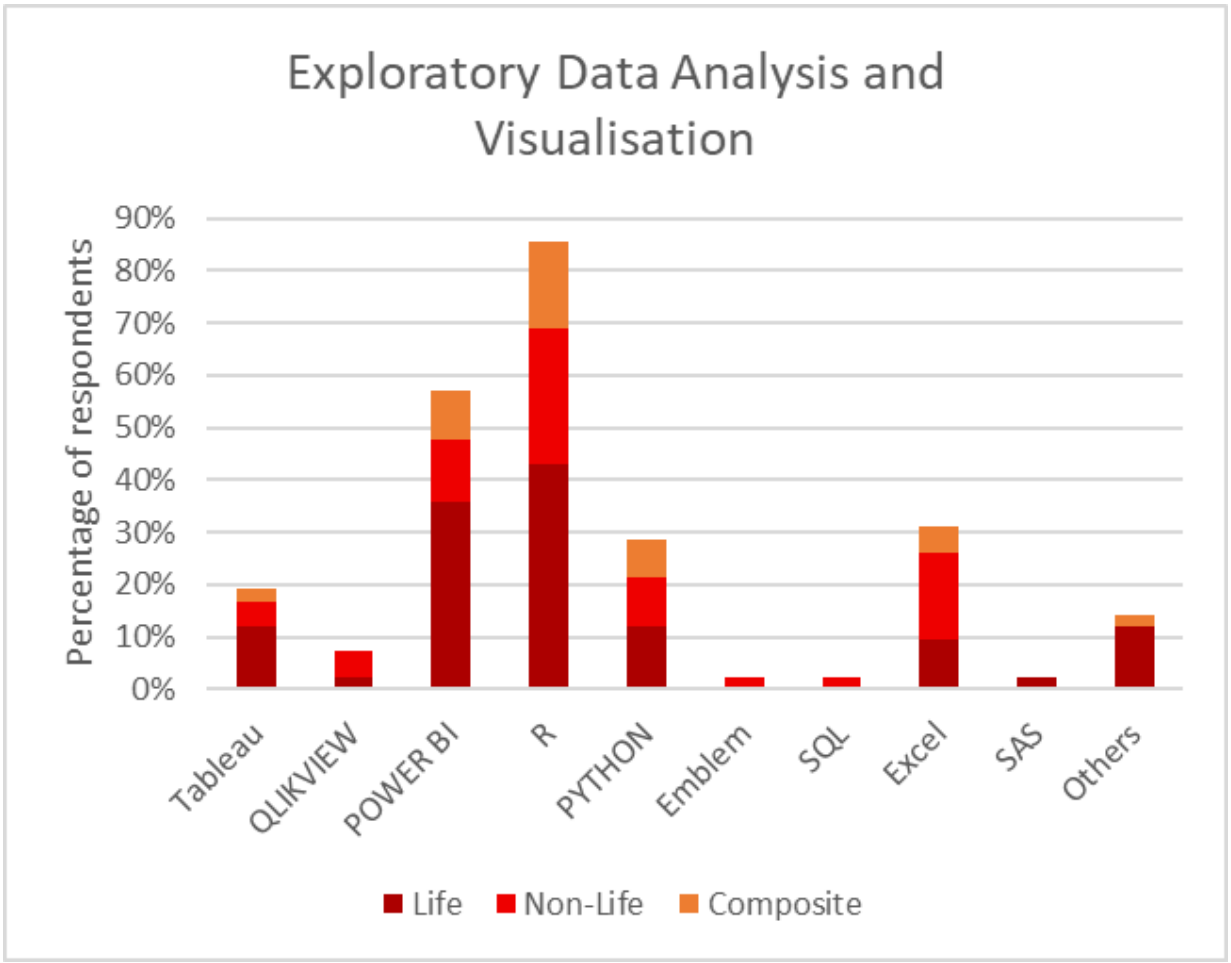
3. Data Sources

- Not surprisingly; the most used data sources are *customer policy application* and *customer quotes* to ensure an understanding of customer needs and risks.
- Quite trending are data sources to understand *customer behaviour* like online app behaviour or complaints' data; for example.
- In many cases the insurer also relies on the reinsurer to providing additional insight where data is limited.
- Less popular sources were *social media* data as well as *wearables data*



Benchmarking Diagram 10

PART IV
TOOLS AND TECHNIQUES



Benchmarking Diagram 11

Key take-aways

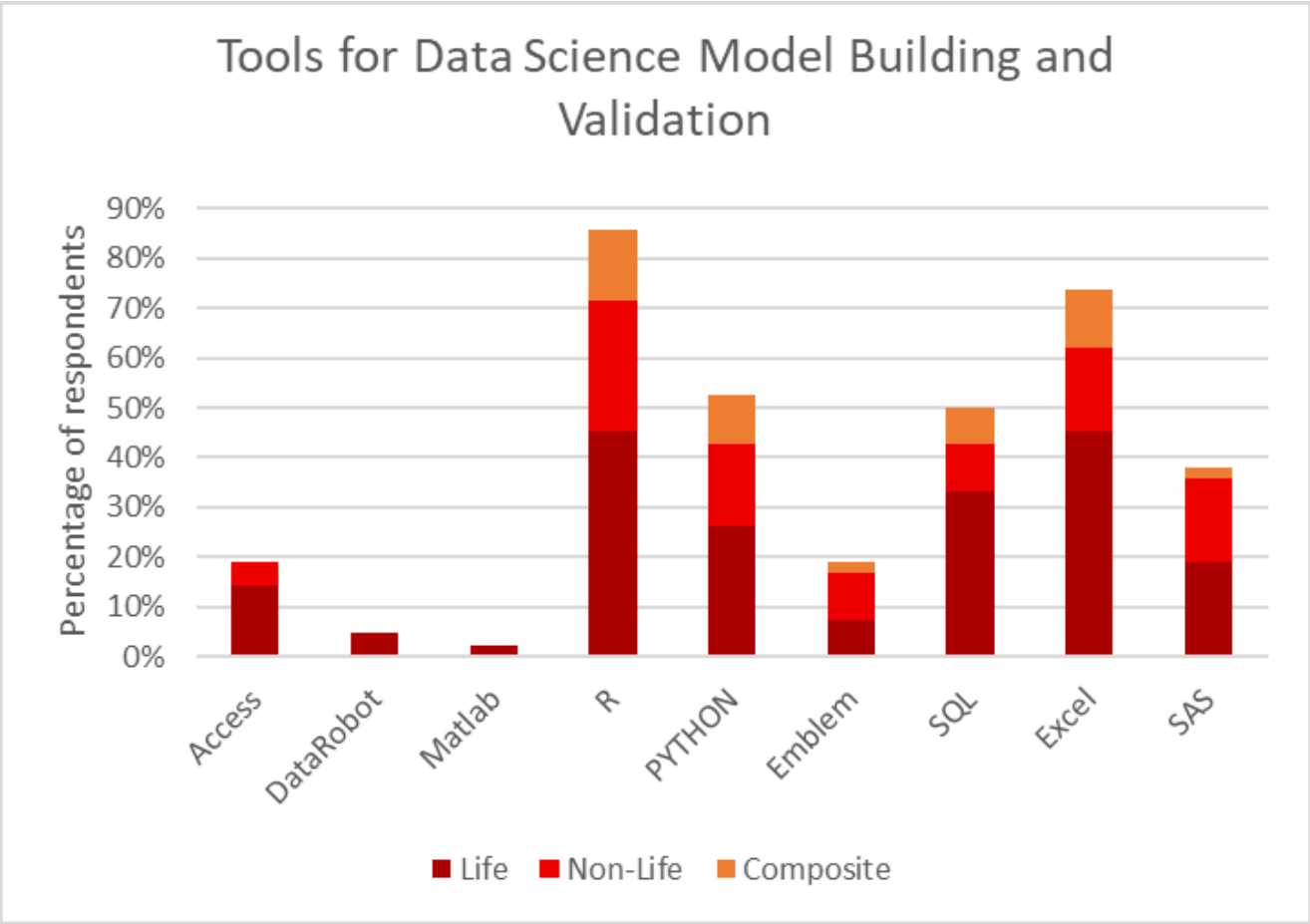
Overall observations

- Most insurers using R and roughly 30% using Python as well; for exploratory data analysis and visualisations.
- Commercial tools: Power BI is leading

Main take-aways:

- Excel is still quite common for data checks and exploration
- R is the preferred tool with most actuarial specific functions and packages

DATA SCIENCE TOOLS: MODEL BUILDING

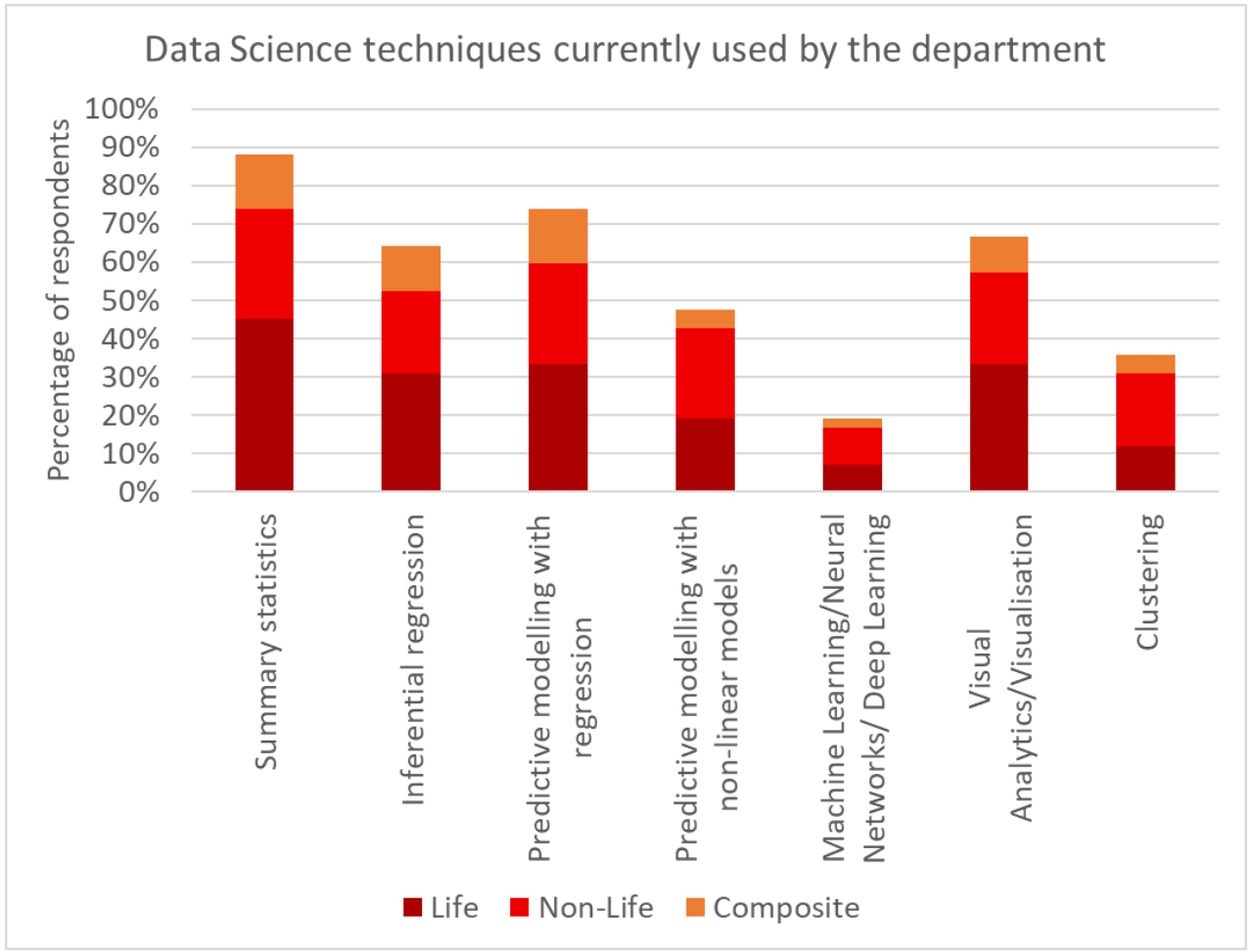


Benchmarking Diagram 12

Key take-aways

- Research is done with Excel or non-commercial tools like R and Python
- SAS: seen as standard software which is common by wider actuarial teams as well
- SQL: is used as an enabler for model building to access data and to manage data
- Implementation of data science models tend to be performed with commercial tools for specific actuarial tasks (due to traceability, compliance)
  e.g. Emblem for pricing
- But: add-on tools to existing software solution are also common.

DATA SCIENCE TECHNIQUES APPLIED



Benchmarking Diagram 13

Key take-aways

Overall observations

- Exploration with advanced techniques like machine learning/neural networks/deep learning are still only done by a small number of respondents

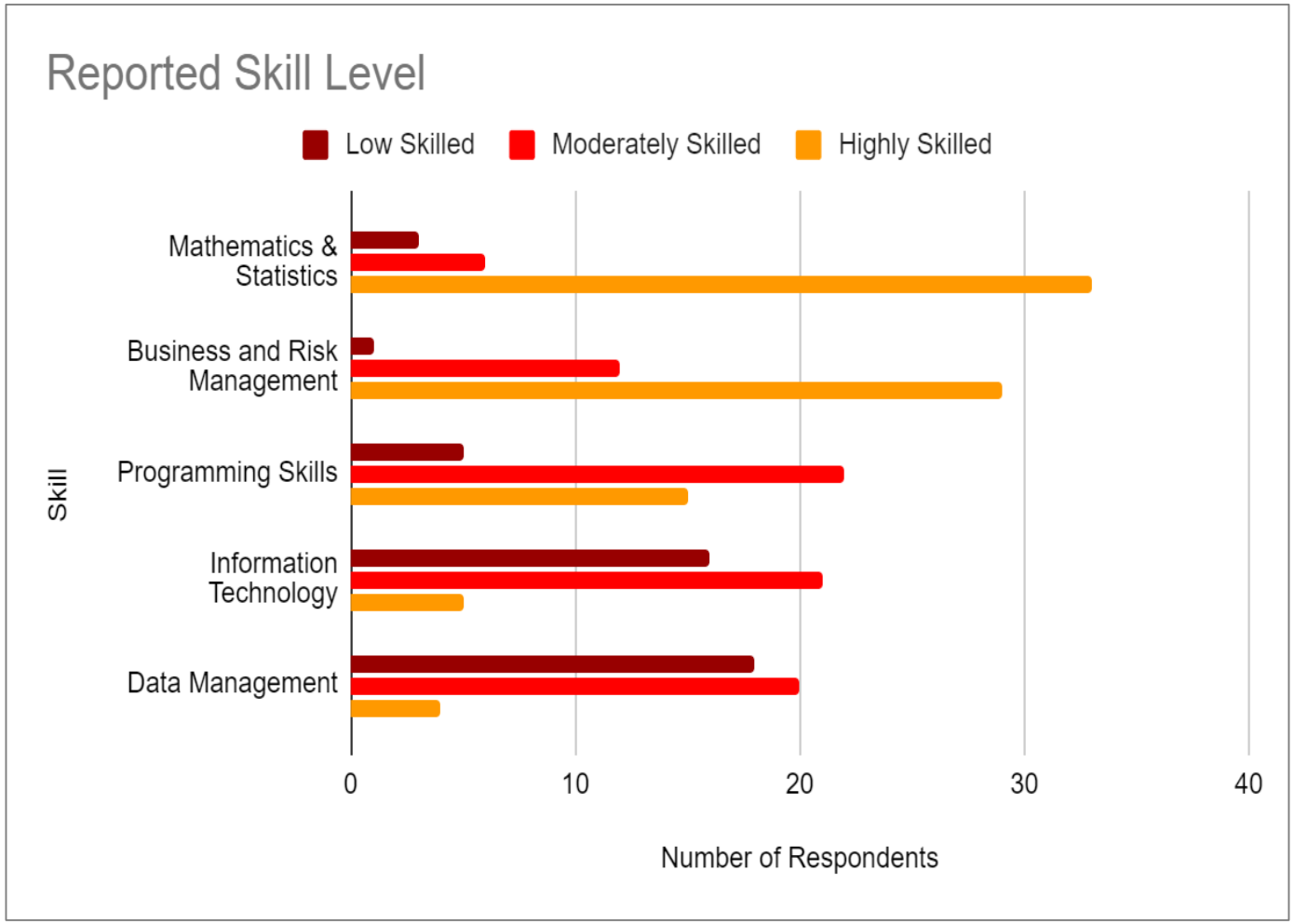
Most common techniques

- Mostly still conventional techniques such as
  - Classical summary statistics
  - Inferential regression
  - Predictive modelling with regression
  - Visualisation techniques

Crucial criteria for applying new techniques

- Enough data and data quality
- Predictability vs interpretability of model
- Costs vs. effort and frequency of use
- Significant improvement expected; compared to models currently used
- The objective of the data science exercise

PART V
UPSKILLING

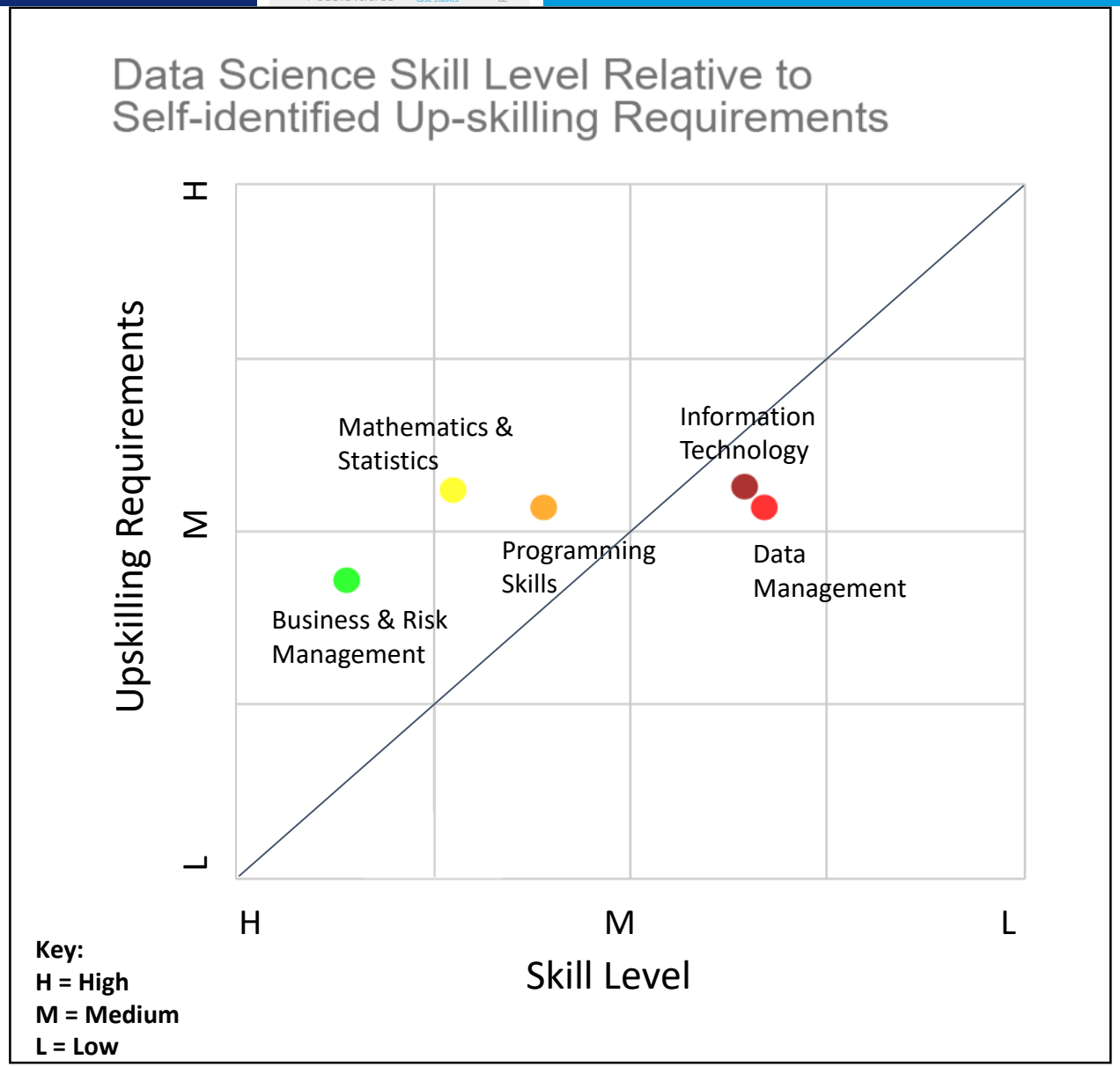


Benchmarking Diagram 14

Key take-aways

In respect of actuarial departments; respondents reported:

- As expected, high skill levels in respect of mathematics & statistics and business & risk management knowledge;
- Relatively lower skill levels in respect of data management and IT;
- Medium level of skills in respect of programming
- The business and risk management skills category included skills related to communication, risk management, validation and reporting



Benchmarking Diagram 15

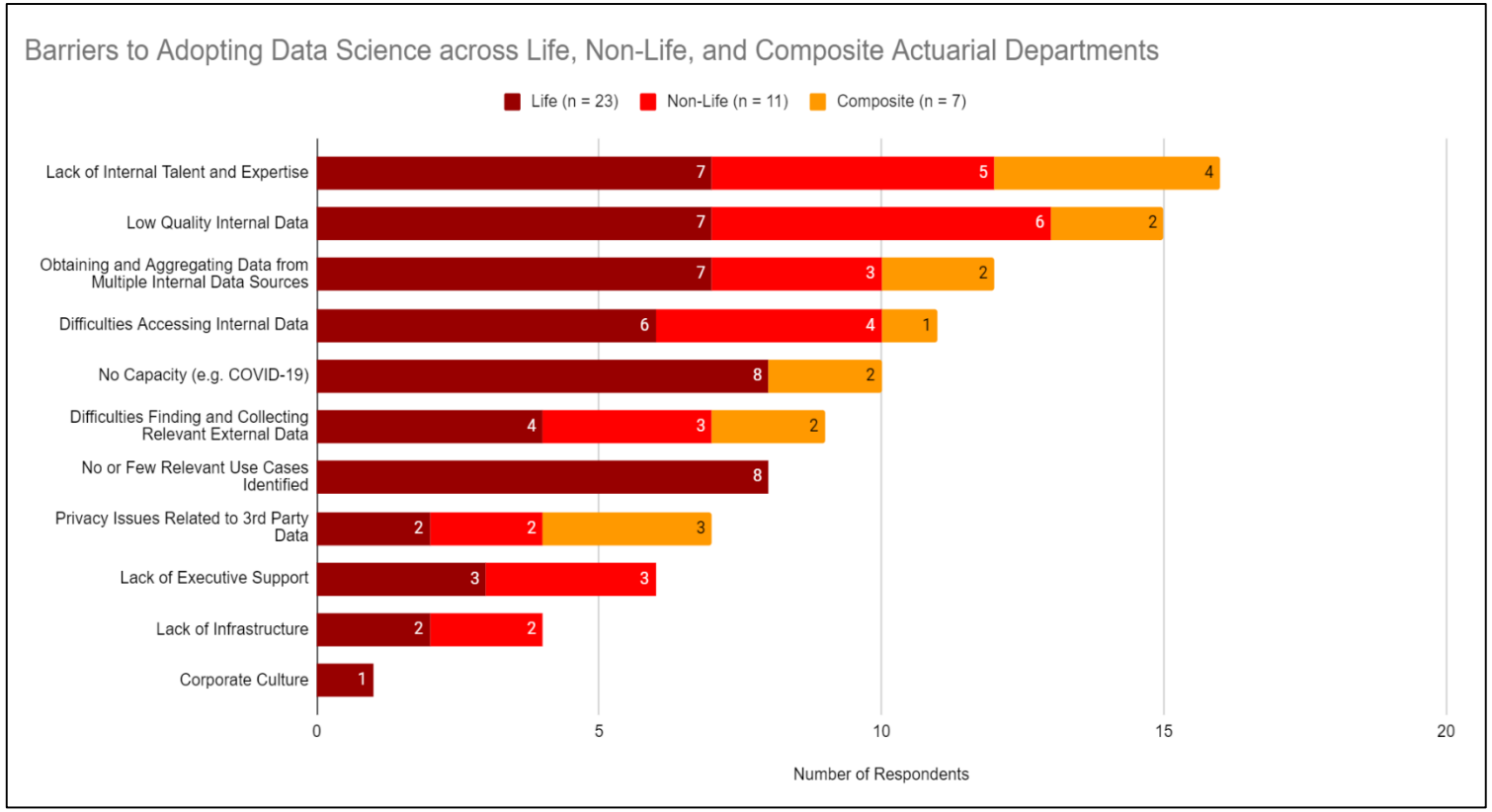
Key take-aways

This graph plots the average data science skill level (from high to low) against the average reported desire to upskilling (from low to high) across five broad skill categories.

- We *expect* the categories to follow along the diagonal (low skill corresponding to high upskilling requirements and *vice-versa)*.
- Categories that fall in the upper-left triangle indicate areas teams are reported to be strong and want to further improve. **Business & Risk Management**, **Mathematical & Statistical** skills are all areas actuaries are traditionally skilled in.
- Areas below the diagonal indicate where teams have reported lower skill levels but have not necessarily indicated a higher upskilling requirement.
- **IT** and **data management** skills are reportedly weaker with a relatively low upskill requirement identified

PART VI
DATA SCIENCE CHALLENGES & OPPORTUNITIES

LIFE VS NON-LIFE VS COMPOSITE COMPANIES



Key take-aways

- **Lack of internal talent** and **low-quality internal data** are the biggest barriers for life, non-life, and composite insurer's actuarial departments
- **For Life** departments the biggest challenge appear to be **the lack of relevant use cases** identified and lack of **capacity** to perform data science related activities.
- **Non-Life** departments appear to have greater difficulties **accessing and using data resourcefully**.
- **Composite** departments report **privacy issues related to 3rd party data** as a barrier.

Benchmarking Diagram 16

Data Science Maturity Level (Benchmarking Actuarial Departments)

Data Science Considerations	Low (I) / Medium (II) / High (III)	Levels Key
Vision & Strategy for Implementation		**I**: No formal strategy reported, or in early stages. **II**: Structured plan in place. Needs further refinement **III**: Specific vision & plan with internal governance for A.I. and Big Data
Extent of Application beyond traditional actuarial workflow		**I**: Limited data science application (data science activities related to data management & reporting processes) **II**: Assumption setting and pricing strategies including Proof of Concept. **III**: Key business decisions, fraud detection, consumer behaviour, value-added initiatives
Non-Traditional Data Sources Used		**I**: Traditional such as policyholder, financial/credit. Limited external data **II**: Mix of internal and external sources. Internal data reflects their experience, risks, and the market they operate in **III**: Also includes text mining, customer behaviour, telematics data
Data and Software Policies in Place		**I**: No formal policies beyond Organisational **II**: Approved software & package. Department-specific data policies around using & access. **III**: Specifying department's IT governance. Focus on validation protocol in addition to data governance - particularly when using advanced models
Data Science Techniques Applied		**I**: Only summary statistics and simple visualisation applied **II**: Fitting GLMs and advanced visualisations (dashboarding) **III**: ML predictive models, automated processes, incorporating AI
Integration with other Domain Experts		**I**: No integration **II**: Integration where appropriate **III**: Specific roles designated (IT; data management; integration)
Infrastructure Control (Data, Hardware, Shared Resources)		**I**: Limited control. Aligned to IT standards **II**: Department team leaders coordinate shared resources **III**: Full control; integrated based on department's needs
Tools Used		I: Mainly proprietary software (Excel, Prophet, Emblem, etc.) **II**: Proprietary legacy systems, some open-source (R/Python) **III**: Mainly open-source. Innovate in-house software packages
Training and Upskilling Strategies		**I**: No formal strategy to improve team's skills. Some individuals may choose to upskill in their own capacity **II**: Some training is offered, however it is either not prioritised or too generic for specific actuarial work **III**: Relevant Actuarial data science upskilling is integrated into the department.

= 5 Respondents

Benchmarking Diagram 17

EAA e-Conference on
Data Science & Data Ethics

29 June 2021

XAVIER MARECHAL

CEO, Reacfin
xavier.marechal@reacfin.com
www.reacfin.com

VALERIE DU PREEZ

FIA, Managing Director, Dupro Ltd
valeriedupreez@dupro.co.uk
www.dupro.co.uk |
www.actuartech.com

ANJA FRIEDRICH

Manager, Synpulse
anja.friedrich@synpulse.com
www.synpulse.com