Institute and Faculty of Actuaries

Enhancing traditional reserving using data science techniques

Tom Durkin, LCP, tom.durkin@lcp.uk.com
Charlie Stone, LCP, charlie.stone@lcp.uk.com

26 April 2019

Agenda

- What are your objectives?
- Defining your analytics strategy
- Case study 1: Multi-factor reserving
- Case study 2: Diagnostics
- Building your business case

What are your objectives?

Earlier identification of trends	More robust results	More time for value added analysis
Quicker results	Avoiding reserving surprises	Saving time and money
Reducing mundane work	Improve management information	Increased confidence in results

What are your objectives?

Earlier identification of trends	More robust results	More time for value added analysis
Quicker results	Avoiding reserving surprises	Saving time and money
Reducing mundane work	Improve management information	Increased confidence in results

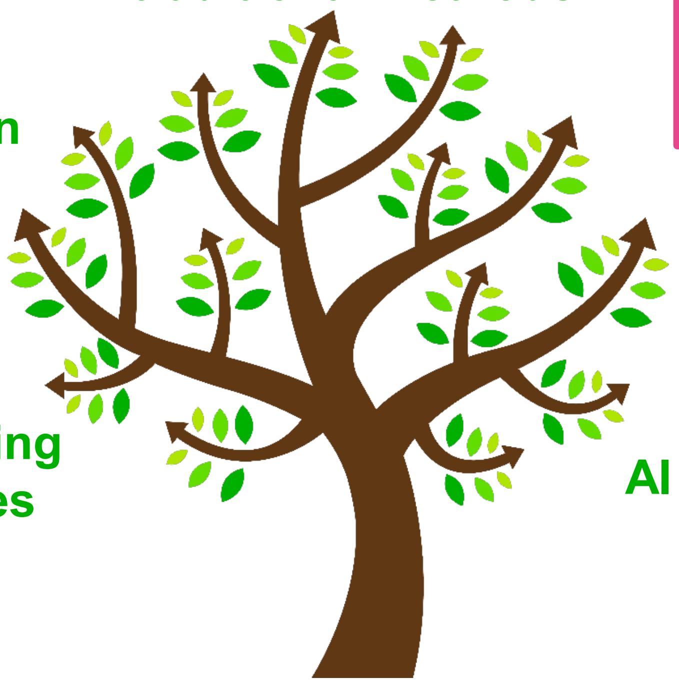
What are your objectives?

Earlier identification of trends	More robust results	More time for value added analysis
Quicker results	Avoiding reserving surprises	Saving time and money
Reducing mundane work	Improve management information	Increased confidence in results

Defining your analytics strategy

Focus for today, using techniques including:

- Clustering
- Hold-out sampling
- Feature engineering



Case study 1: Multi-factor reserving

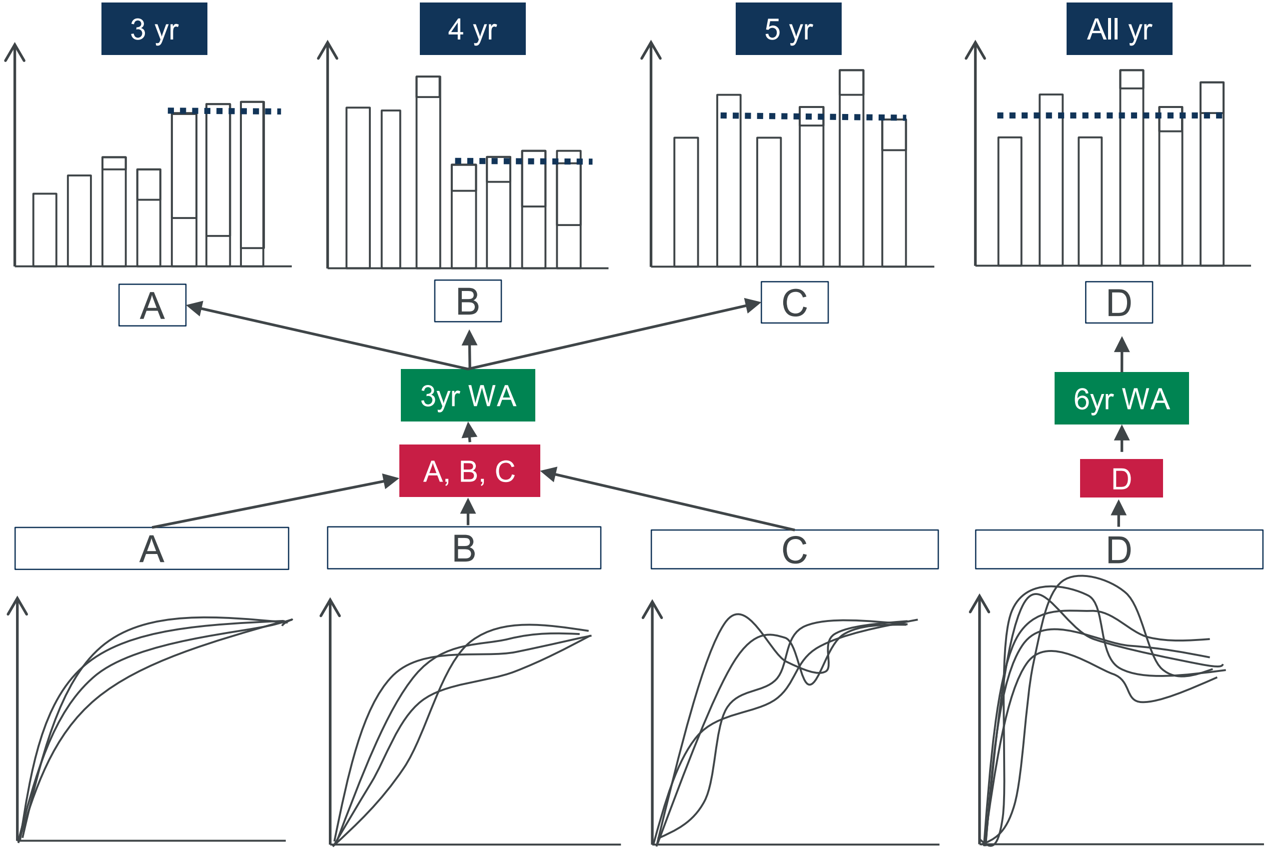
Traditional

- Reserving typically focused on one segmentation of the data
- Time consuming to consider alternative segmentations
- Challenging to give robust rationale for homogeneous risk groups

Enhanced

- A new automated reserving engine, prepares reserves in less than 5 minutes
- Multiple reserve estimates using alternative key risk factors / data segmentations
- Engine scores the quality of each reserve estimate

Automated reserving engine

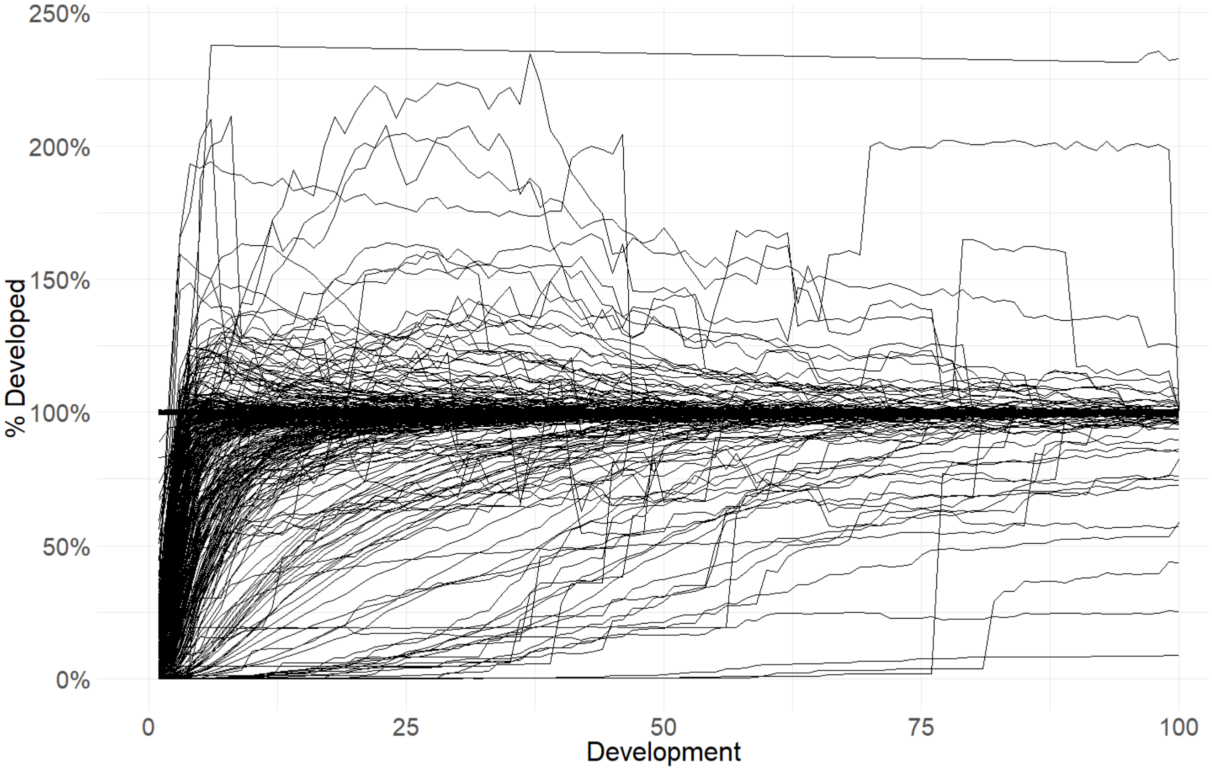


C: What averaging period should be used for initial expected loss ratio?

B: What averaging period should be used for each development pattern?

A: Which segments should be grouped when calculating development patterns?

A: Which segments should be grouped when calculating development patterns?



A: Which segments should be grouped when calculating development patterns?

Which linkage function?

Which clustering method?

What is the optimal number of clusters?

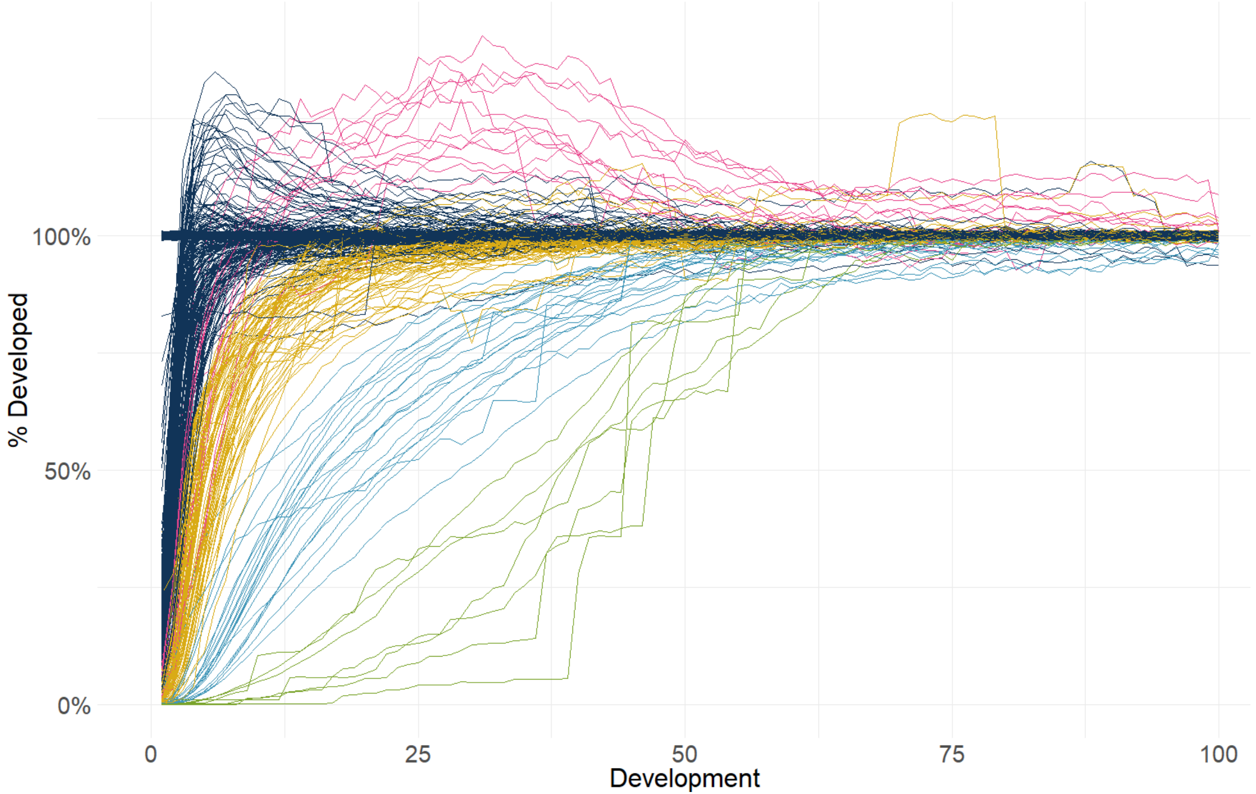
Should the variables be standardised?

How stable are the clusters?

Which variables to use in clustering?

How to cluster segments with similar trending?

A: Which segments should be grouped when calculating development patterns?

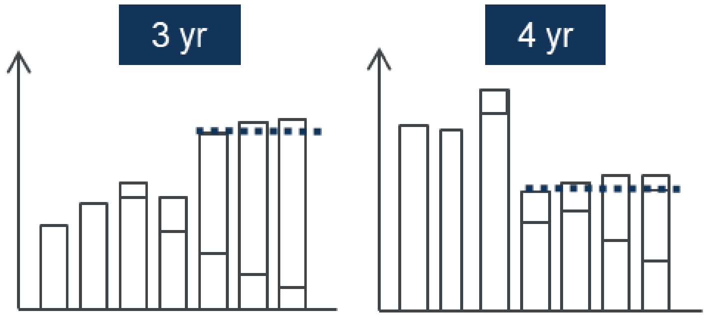


B: What averaging period should be used for each development pattern?

- Averaging period selected that best predicts emerging experience across 10 **hold-out samples**
- "Best" is defined as the smallest average difference between projected vs actual position

C: What averaging period should be used for initial expected loss ratio?

- Which cut-off point splits the ultimates into two groups that are most similar?
- Use **measure of entropy**, KL divergence, to compare different cut-off points. Choose cut-off point with smallest entropy.

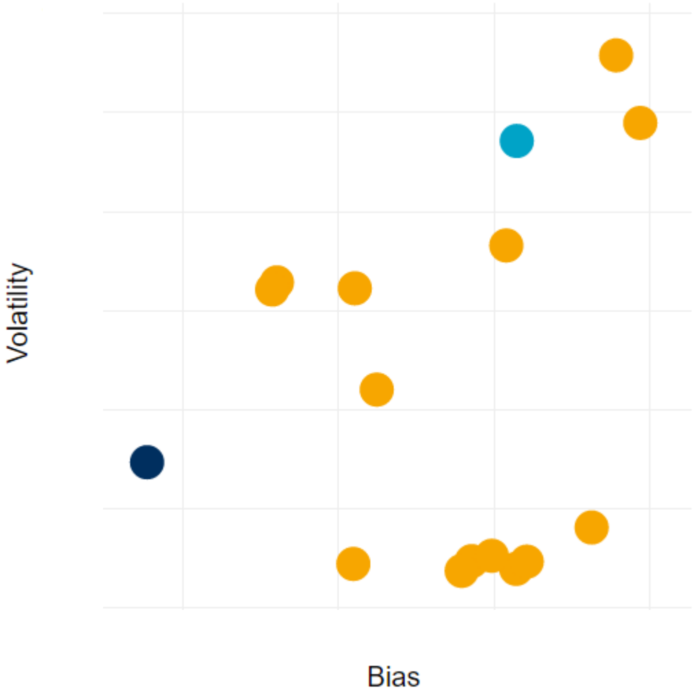


Assessing the quality of each reserve estimate

Volatility

Options include:

- Stochastic Mack
- ODP bootstrap
- Rollforward volatility
  - Movement in reserve estimates between analyses
- Sub sample volatility
  - Repeat the reserving process on samples from the full set of individual claims data



Bias

Options include:

- 'Recent experience will continue'
  - Emerging experience error
- 'Nothing has changed'
  - Back-testing error
- 'Wisdom of crowds'
  - Difference from other segmentations

The automated engine assesses a reserve estimate as "high quality" where it has low volatility and low bias – with options for how each is measured

Case study 1: Key outcomes

- Insights from investigating new key risk factors
- Easy to communicate final estimates, as based on well understood techniques
- Data driven rationale for selected segmentation

Case study 2: Diagnostics

Traditional

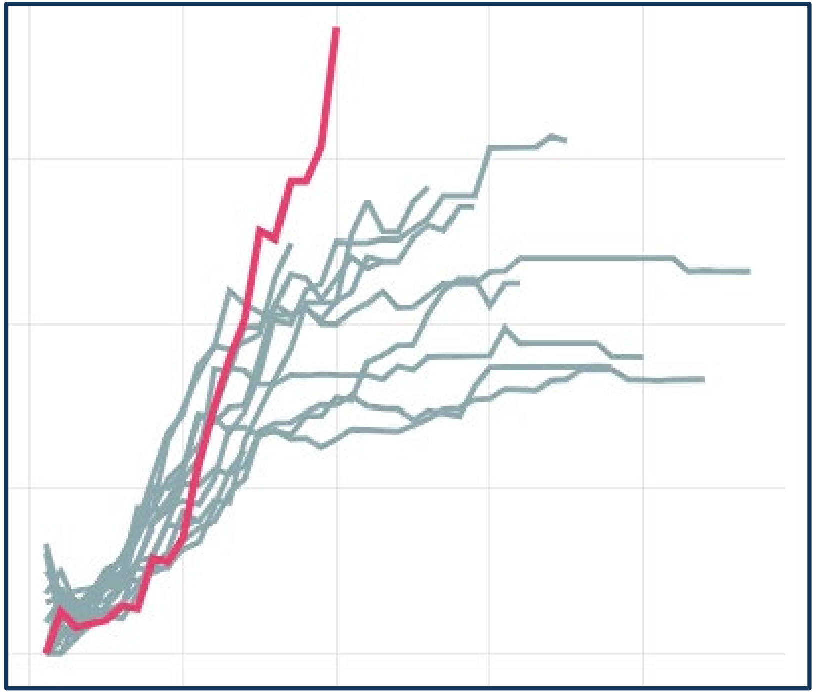
- Manual review of triangles
- Time consuming (ie, hours) to review all triangles, so typically consider a selected sample
- Potential to miss features

Enhanced

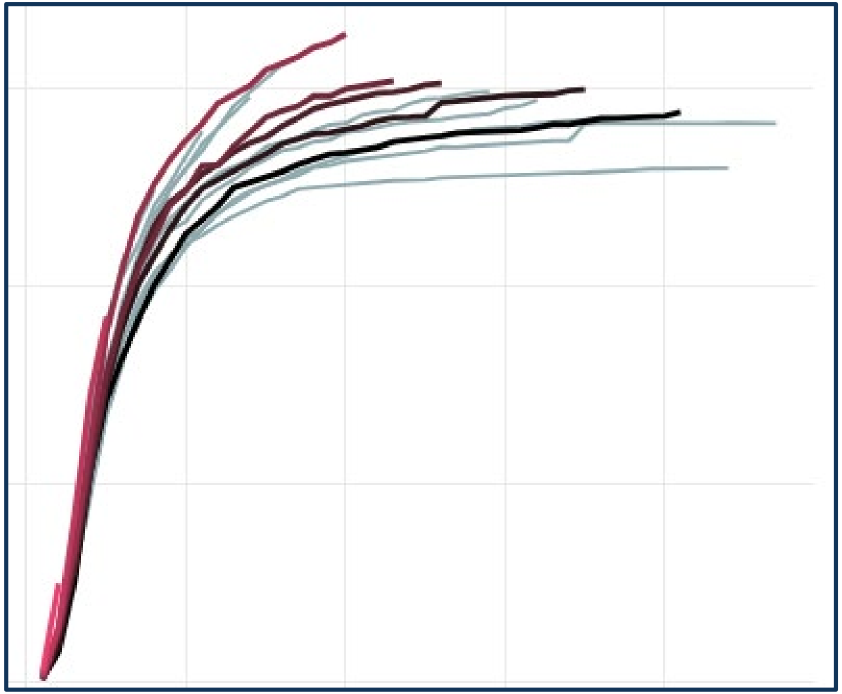
- Automated approach, which priorities top triangles to review
- Quick (ie, minutes) and scalable
- More time to understand the “why?”

What are we looking for?

“Sticking out”



“Fanning out”



Feature engineering

Feature engineering:

Using domain <u>knowledge of the data to create features</u> that make machine learning algorithms work

"Coming up with features is difficult, time-consuming, requires expert knowledge. Applied machine learning is basically feature engineering."

Andrew Ng, Co-founder of Google Brain, Stanford Professor

Feature engineering

Raw data

- **Transactions**: Payment and reserve transactions by claim
- **Claim details:** (incident date, reporting date, head of damage etc)
- 10^6 rows, 10^2 columns

Triangles

Triangle: total values by cohort and development (10 rows, 10^2 columns)

Value: Paid, Incurred, claim count, average claim size (10)

Segmentation: Line of business, head of damage, claim size (10^2)

Features and metrics

Features
(eg, 5 to 10)

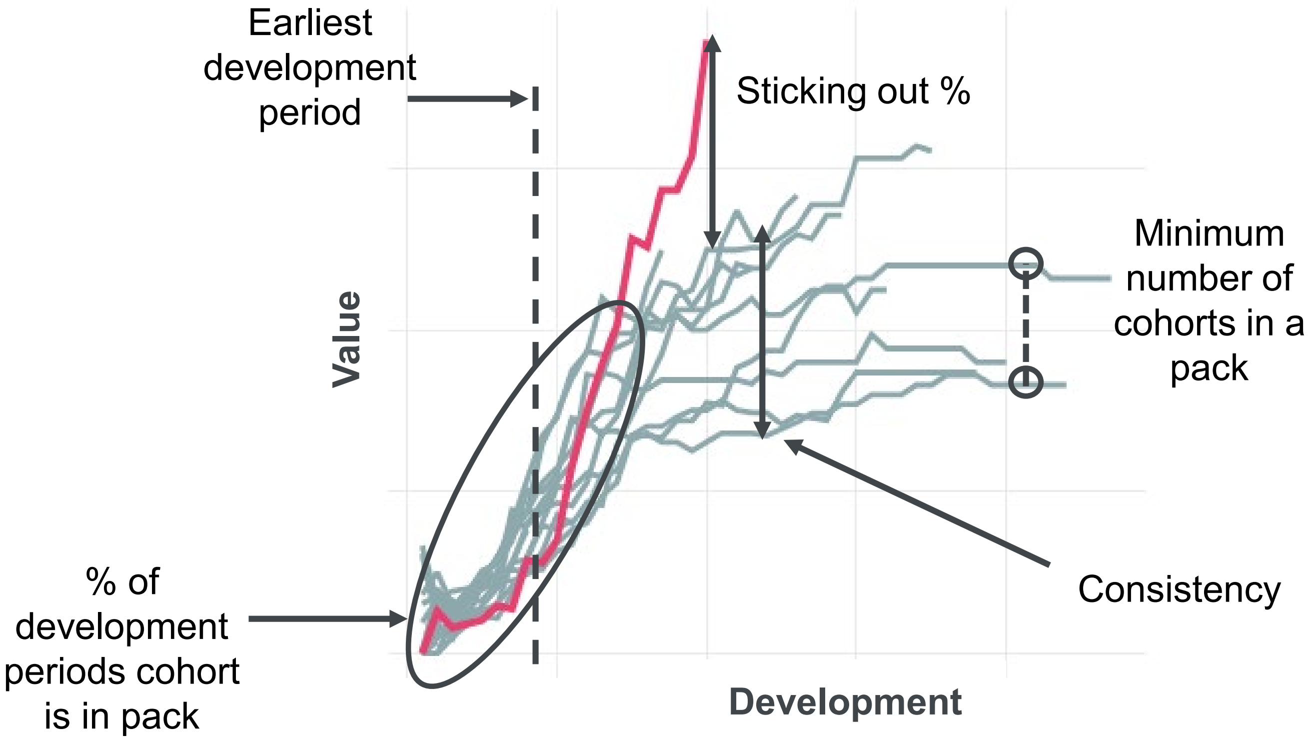
Feature metrics
Anomaly/trend metrics for one anomaly/trend (eg, 5 to 10)

Feature engineering

Fanning out

Feature engineering

Sticking out



Constructing feature metrics

Key steps

1. Brainstorm key metrics
2. Decide which feature metrics to create
3. Create the metrics
4. Check how well the metrics work
5. Improve metrics where needed
6. Go back to brainstorming and creating metrics if necessary

Helpful tips

- Manually classify a large number yourself
- Separate into distinct groups and choose best from each group
- Work on code for one triangle with interim charts and checks. Then ‘map’ across all triangles
- Order by one metric, chart a few examples from top, middle and bottom and see if you agree with ranking
- Pick out examples that don’t work well and chart them. Are there common reasons for why the metric isn’t working?
- Revisit the metrics discarded in step 2.

R notebooks are very useful!

Case study 2: Key outcomes

- Efficient identification of key features across multiple triangles
- Increased confidence that key trends and anomalies have been identified
- Structured approach to building knowledge of a book

Developing your business case

Virtuous engagement cycle	Your objectives	Business objectives
Return on investment	Cost calculations	Intermediate goals

Questions

Comments

The views expressed in this presentation are those of invited contributors and not necessarily those of the IFoA. The IFoA do not endorse any of the views stated, nor any claims or representations made in this presentation and accept no responsibility or liability to any person for loss or damage suffered as a consequence of their placing reliance upon any view, claim or representation made in this presentation.

The information and expressions of opinion contained in this publication are not intended to be a comprehensive study, nor to provide actuarial advice or advice of any nature and should not be treated as a substitute for specific advice concerning individual situations. On no account may any part of this presentation be reproduced without the written permission of the authors.