Presciient

Getting started with Predictive Analytics in the Public Sector

2012-06-13

Stephen Samild

Director, Presciient

As advertised

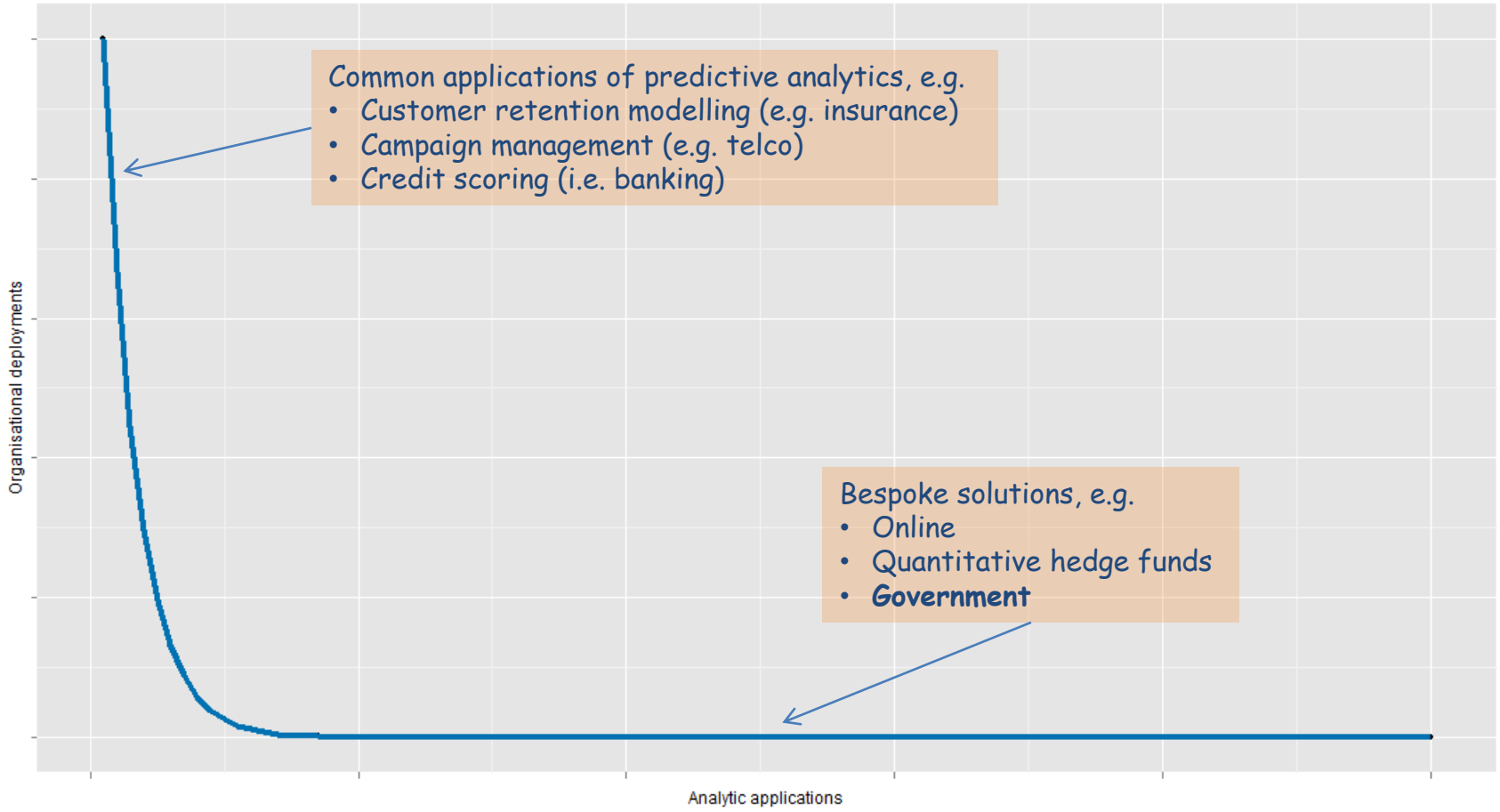
Many agencies are seeking to harness the potential of Predictive Analytics. This session will provide a practical guidance for getting started with Predictive Analytics initiatives, drawing on lessons learned at the Department of Immigration and Citizenship.

Conventional Business Intelligence provides a largely historic view of enterprise data, but its volume and complexity often obscure additional insights. Predictive Analytic methods can uncover highly specific predictive patterns, trends and relationships. These data-driven insights can be utilised by public sector agencies to create risk and behavioural profiles, detect and counteract fraud and non-compliance, and enrich strategic and operational intelligence products.

The session will cover:

- How Predictive Analytics makes use of data
- What comprises a Predictive Analytics capability, and how to develop one
- How to develop a business case for Predictive Analytics
- What you can do to get started with Predictive Analytics, right now, with the resources and data you already have

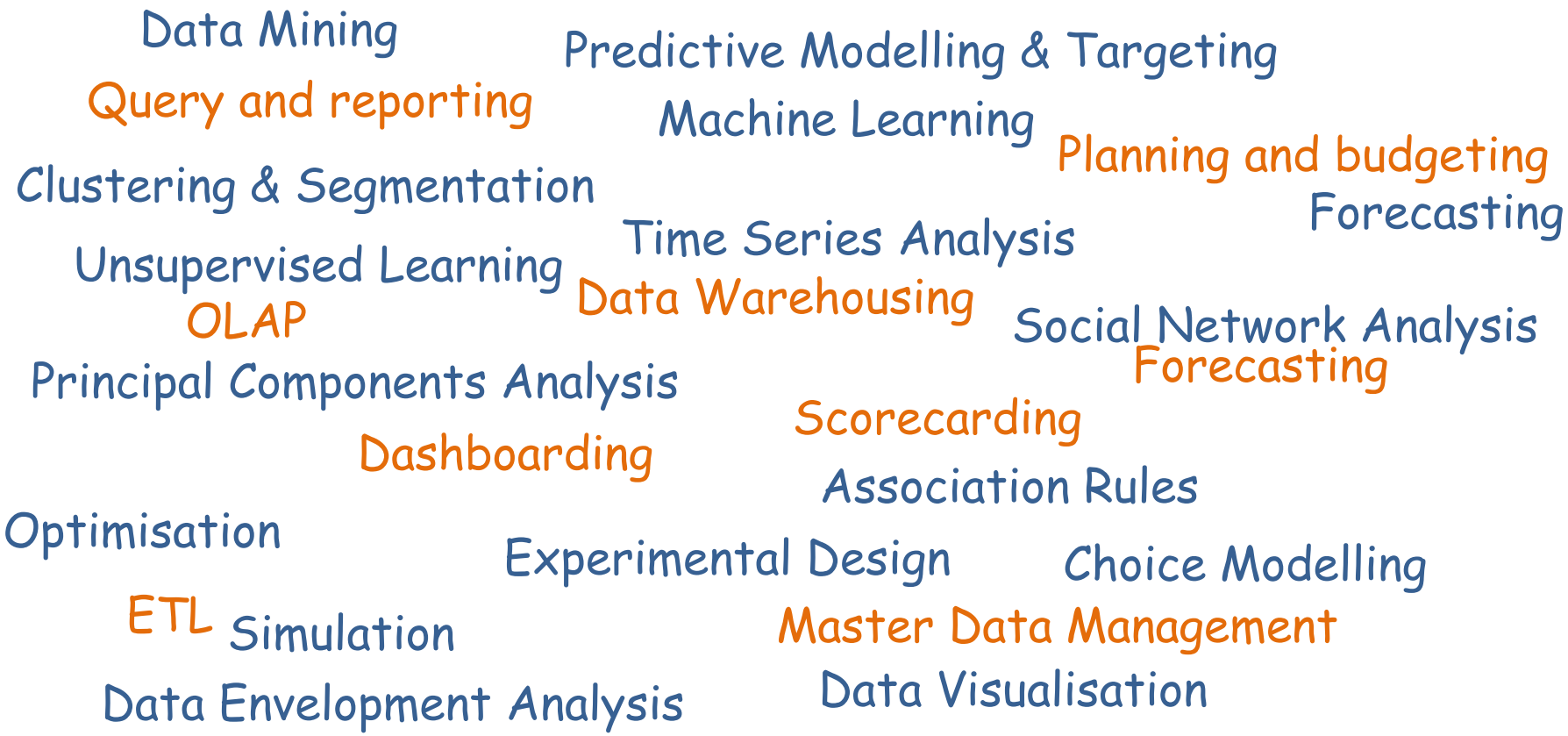
Presciient services the 'long tail' of analytics



Data-driven solutions in the public sector are typically bespoke

- Agencies are legislated monopolies
- How they do business differs substantively from equivalent agencies interstate and overseas, e.g:
  - ATO assessing taxation returns
  - Centrelink assessing welfare claims
  - Customs assessing cargo and passenger movements
  - DIAC assessing visa applications
  - Medicare assessing healthcare refunds

Predictive analytics can be meaningfully distinguished from business intelligence



More often than not, “Predictive Analytics” means “predictive modelling” or “statistical forecasting”

- Inferring the unseen from the seen using a model
  - Where the model’s definition has come from within the data, not from assumptions we’ve made outside the data

The challenges that apply to business intelligence are all shared by analytics (in fact they become more pronounced)

"Between 70% to 80% of corporate business intelligence projects fail, according to research by analyst firm Gartner."

- 10 January 2011, Computer Weekly,
  'Poor communication to blame for business intelligence failure, says Gartner'

"Tried and true best practices for enterprise software development and support just don't work for business intelligence (BI)."

- 22 July 2011, Forrester,
  'AGILE BUSINESS INTELLIGENCE SOLUTION CENTERS ARE MORE THAN JUST COMPETENCY CENTERS'

"One in six big IT projects go over-budget by an average of 200%, according to new research."

- 26 August 2011, BBC News TECHNOLOGY,
  ''Black Swans' busting IT budgets'

Presciient

Decision making in the public sector is driven by power and authority, not data

Authority

- Bureaucratic mechanisms:
  - Policies, rules
  - Standard operating procedures
  - Formal, rational, explicit
- *Following the rules*

Power

- Political power:
  - Influence, relationships
  - Strategic behaviour & coalition building
  - Informal, disorderly, tacit
- *Bargaining*

Power in Organizations, 1981
Jeffrey Pfeffer, Thomas D. Dee II Professor of Organizational Behavior at the Graduate School of Business, Stanford University,

Presciient

Initatives in the public sector require upfront permission, and sometimes upfront investment

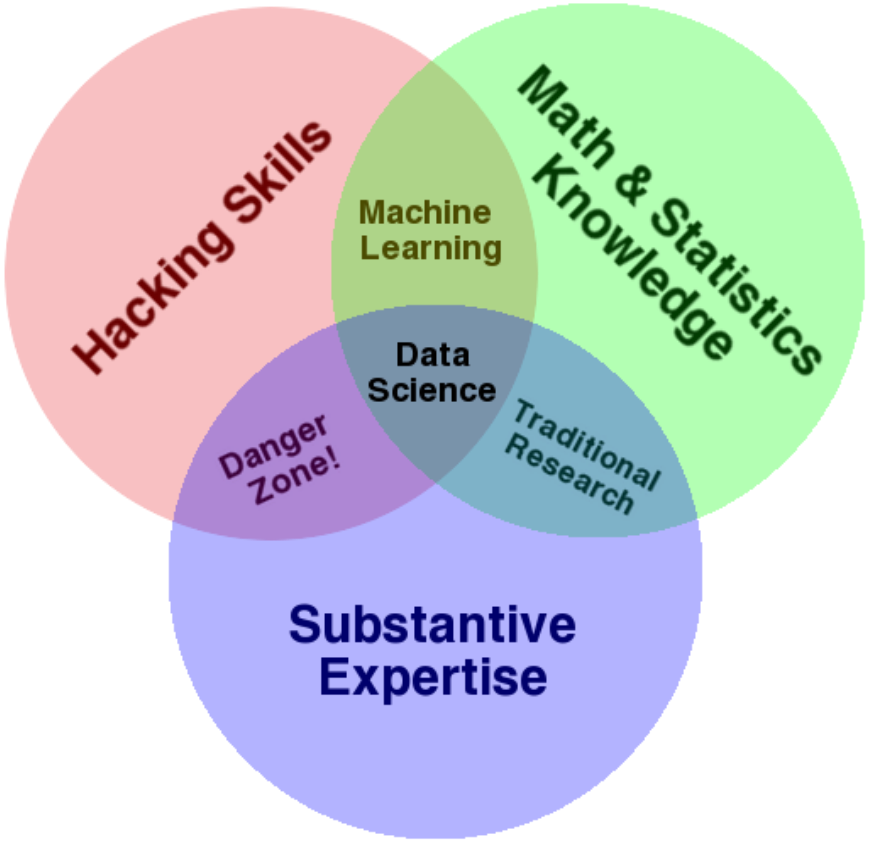
- Analytics is software enabled, data reliant
- Most analytic applications require cross-functional cooperation, some cross-agency
- **One way or the other, you need a business case**

FOUR PRACTICAL PRINCIPLES

1. **Start in the lab** (not in production)

- Analytics is exploratory, eclectic, specialised*
- Trial and error
- Finding comes before building
- Deployment takes many forms**
- Insights must lead***
- Analytics is continuous****
- Questions will change

1. Start in the lab > Analytics is exploratory, eclectic, specialised

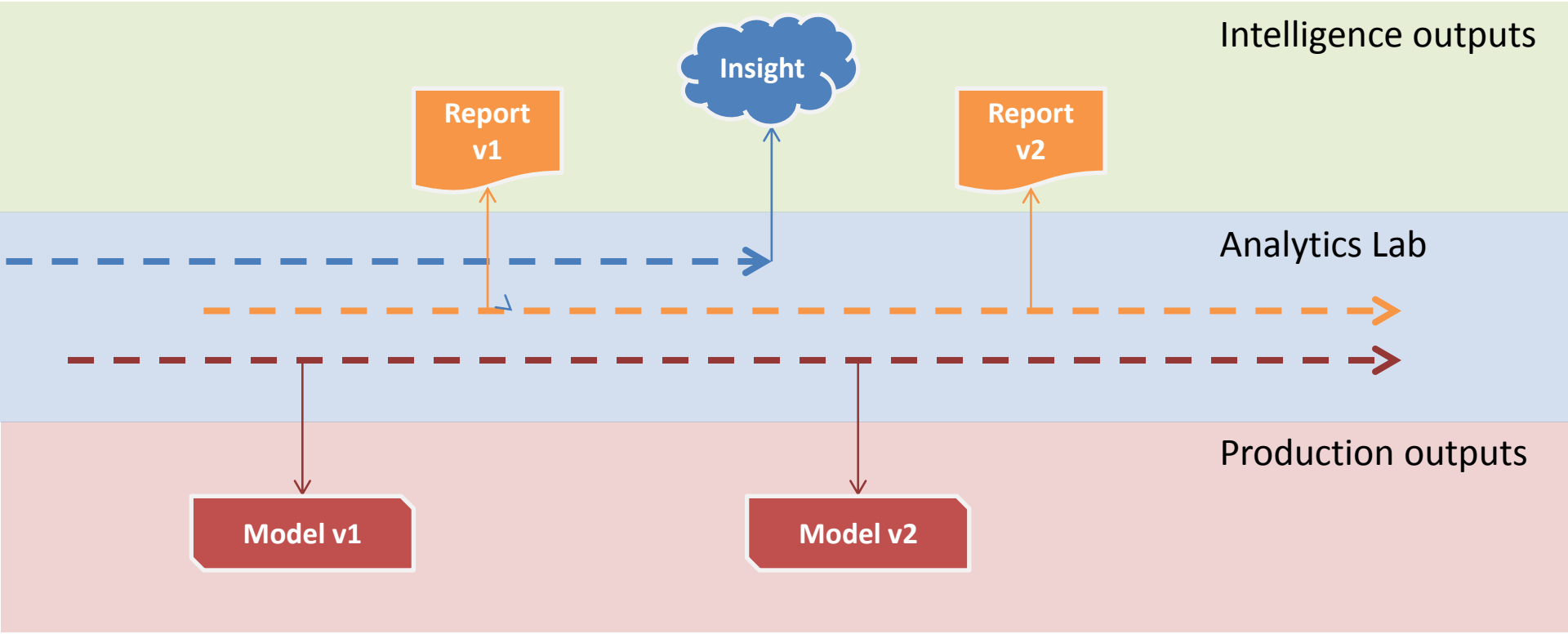


'The Data Science Venn Diagram'
Drew Conway, 30 September 2010, '*Zero Intelligence Agents*'

1. Start in the lab > Deployment takes many forms

Analytics consumer	Item of deployment	Delivery method
• Senior management	• One-off insight • Nuanced answer(s) to a complex question • New piece of context	• Verbal • Post-it note • Powerpoint presentation
• Middle management • Knowledge workers	• Enriched business intelligence, e.g. with segments, automated forecasts, confidence intervals • Targeted analyses	• Model outputs integrated into existing BI platform • BI prototypes • Written reports & studies
• Frontline staff	• Operational decision support or decision automation, in real time, e.g. with scores	• Models embedded into core systems

1. Start in the lab > Analytics is continuous



1. Start in the lab > Insights must lead

"A very early piece of work that we did on some travel data identified that at a particular point in time in 2006, that anybody with a Belgian passport trying to get on a Cathay Pacific flight in Hong Kong and wanting to fly to Brisbane or Perth was 80% likely to be an imposter. This had never been highlighted in that way before, and we demonstrated it just through a simple decision tree, but it was enough to highlight to people just what the power of this particular approach was. It was a data-derived pattern… It was a really important part of the process because that decision tree went up all the way to the senior executive as an example of what could be done. **This then gave us the permission to go and explore further."**

-**Klaus Felsche**, Director of Intent Management & Analytics, Australian Department of Immigration and Citizenship (DIAC),

'Klaus Felsche on Analytics at DIAC', 4 April 2012, Analyst First

Don’t take this literally. It’s an anonymised example and it’s not from DIAC

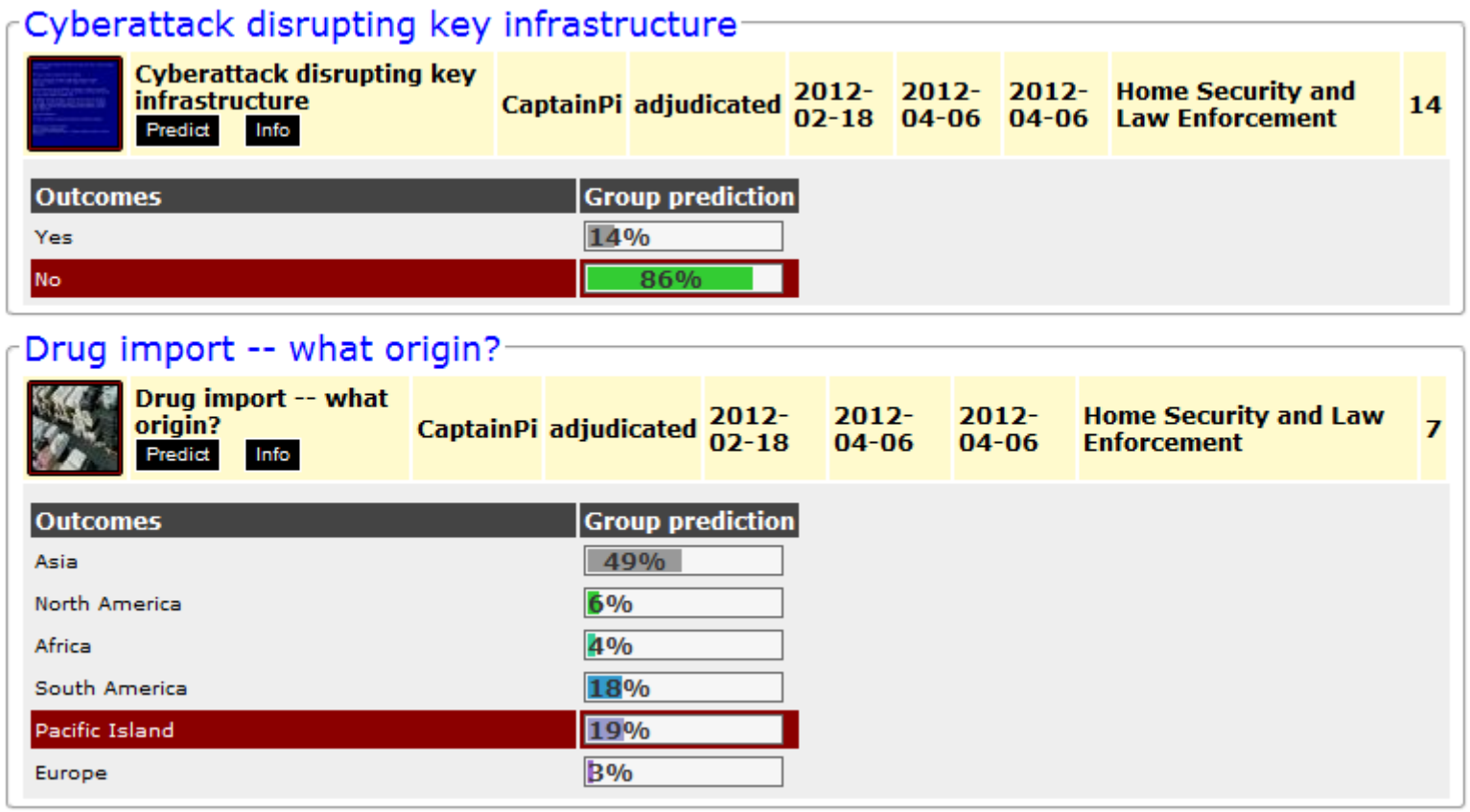
```
node), split, n, yval, (yprob)                    * denotes terminal node

1) root 1032 1 (0.50000000 0.50000000)
     2) days_since_last_transaction <  68  444  0 (0.74258023 0.25741977) *
     3) days_since_last_transaction >= 68  588 1 (0.37429006 0.62570994)
          6) state_of_residence=NSW, VIC, QLD, WA, SA 279 0 0.53998107
             0.46001893)
               12) state_of_residence=NSW, VIC, WA 52 0 (0.93891542
                   0.06108458) *
               13) state_of_residence=QLD, SA 227 1 (0.47924997 0.52075003)
                    26) household_type=Homeowner, Self-funded_Retiree 39
                        0 (0.84792406 0.15207594) *
                    27) household_type=Renter, LAFH, Student 188
                        1 (0.42788050 0.57211950) *
          7) state_of_residence=ACT, NT 309 1 (0.26676397 0.73323603)
               14) employer=Government, Self_Employed 0 (0.63265171
                   0.36734829) *
               15) employer=Private, Sole_Trader, Other 262 1 (0.22619956
                   0.77380044) *
```

2. Start with questions (not with data)

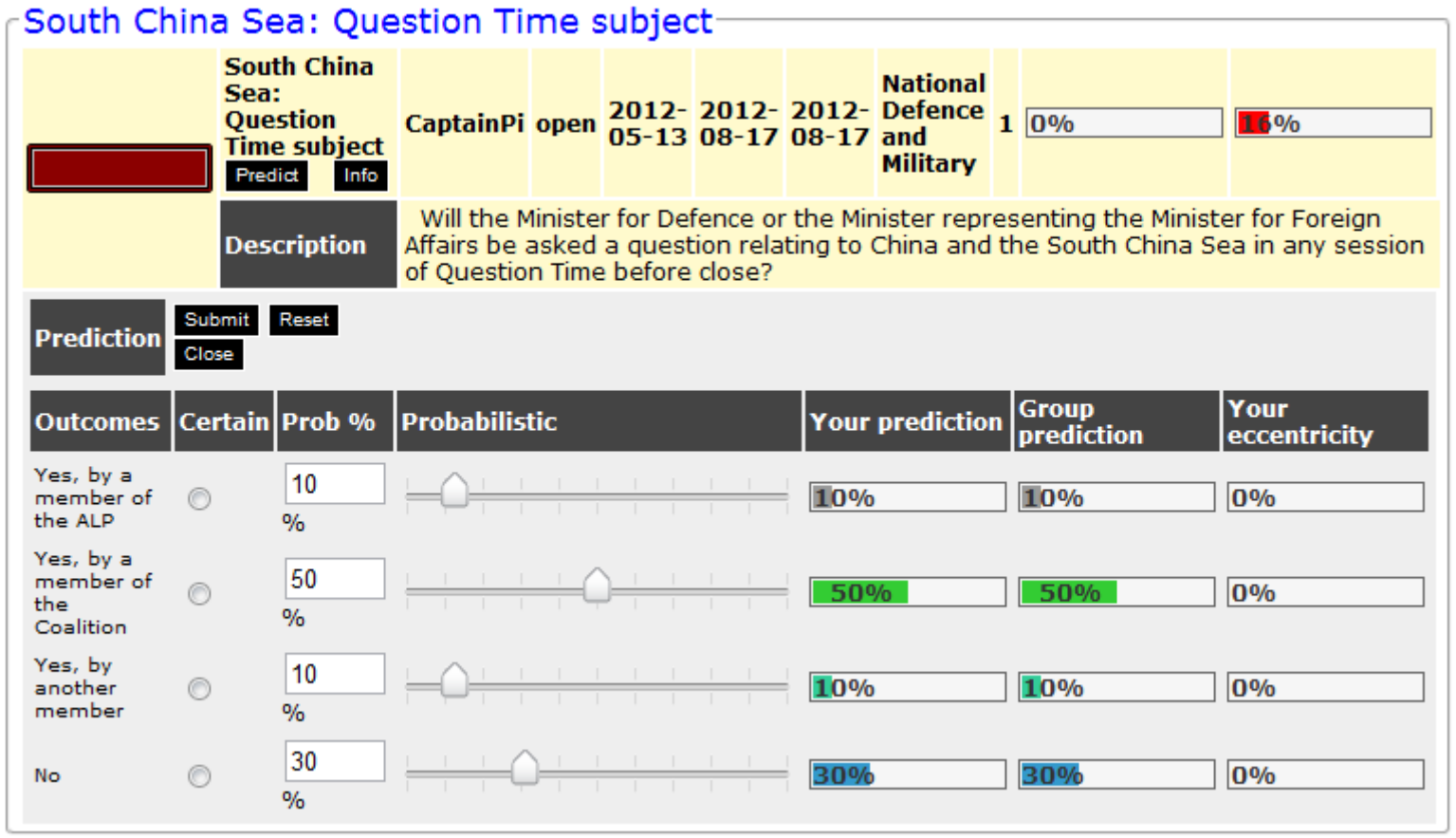
- Questions frame analysis, drive decisions
- Knowing what you don't yet know, or can't know, is a form of knowing
- Data always lags*
- Analysis is read-write**

2. Start with questions > Data always lags



AIPIO Forecasting Competition

2. Start with questions > Data always lags



AIPIO Forecasting Competition

2. Start with questions > Analysis is read-write

Routine analytical tasks I do:

- Entering hitherto tacit data
- Codifying business knowledge
- Finding and synthesising data from outside sources
- Creating dummy and randomised data
- Capturing novel assumptions
- Imposing new categories on existing categorical data
- Enriching existing data by deriving or devising on-the-fly metadata
- Building scenarios and constructing counterfactuals
- Drafting and adding commentary, interpretation, and notes
- Formalising and detailing new questions and follow-on analyses

3. **Start with value** (not with ‘best practices’)

- Solutions are bespoke
- What matters for you will be highly specific
- What worked for others may not work for you, or matter for you
- Insights can be trivial
- Sponsors need to be kept happy
- Payoff comes from core business*

3. Start with value > Payoff comes from core business

"The CIO was very clear that we had to be focused on risk identification in the visa caseload…

"[T]he main business driver was efficiency: identifying risks and then aligning differentiated treatments appropriate to treat them…

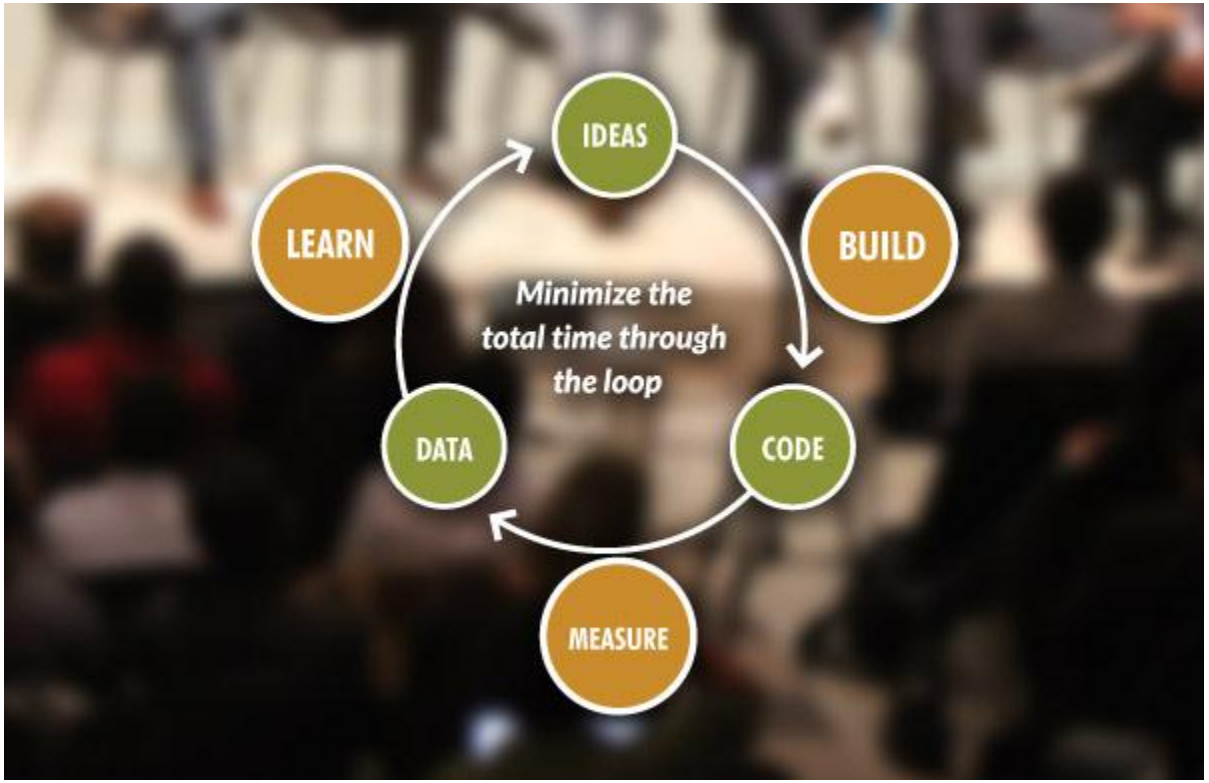
"We have to get on the front foot, pointing out that in four years' time there will be a million more visas to process given current growth rates. How are we going to do that with current methods and our current processes? Analytics offers a solution that can help you manage the growing caseload without a loss of integrity."

-**Klaus Felsche**, Director of Intent Management & Analytics, Australian Department of Immigration and Citizenship (DIAC),
'Klaus Felsche on Analytics at DIAC', 4 April 2012, Analyst First

4. **Start the Lean way** (not with a project plan)

- Don’t let the perfect (e.g. accuracy) be the enemy of the good (e.g. clarity, speed)
- Analytics is intrapreneurial*
- Analytical literacy must be developed iteratively**
- Analytics is a capability***

4. Start the Lean way > Analytics is intrapreneurial



'The Lean Startup Methodology'
Eric Ries

4. Start the Lean way > Analytical literacy must be developed iteratively

1. Lab
2. Test
3. Prototype…

"Once we've got all of that boxed up we've effectively done the design and development stages of any project and we're ready to put a really strong business case to the organisation which says: this thing works; it has the following limitations; the following people like it; the following people hate it; and it's going to cost this much money to turn into an enterprise solution."

4. …Production

-**Klaus Felsche**, Director of Intent Management & Analytics, Australian Department of Immigration and Citizenship (DIAC),

'Klaus Felsche on Analytics at DIAC', 4 April 2012, Analyst First

4. Start the Lean way > Analytics is a capability

- An analytics capability is both an Intelligence and a Production function:
  - Refines analysis objectives
  - Develops analysis methods
  - Produces and deploys analysis products
  - Monitors and manages their use

Additional reading

- The Lean Startup / Eric Ries
- What is data science?, O'Reilly Radar
- The Analytics Lab, Analyst First
- Against best practices in Business Analytics, Analyst First