

Analytics in Revenue

Daniel Sinnott
Chief Analytics Officer, Revenue

What I will cover this morning

Organisational set-up

- Governance & infrastructure
- Capability development

Analytical approaches

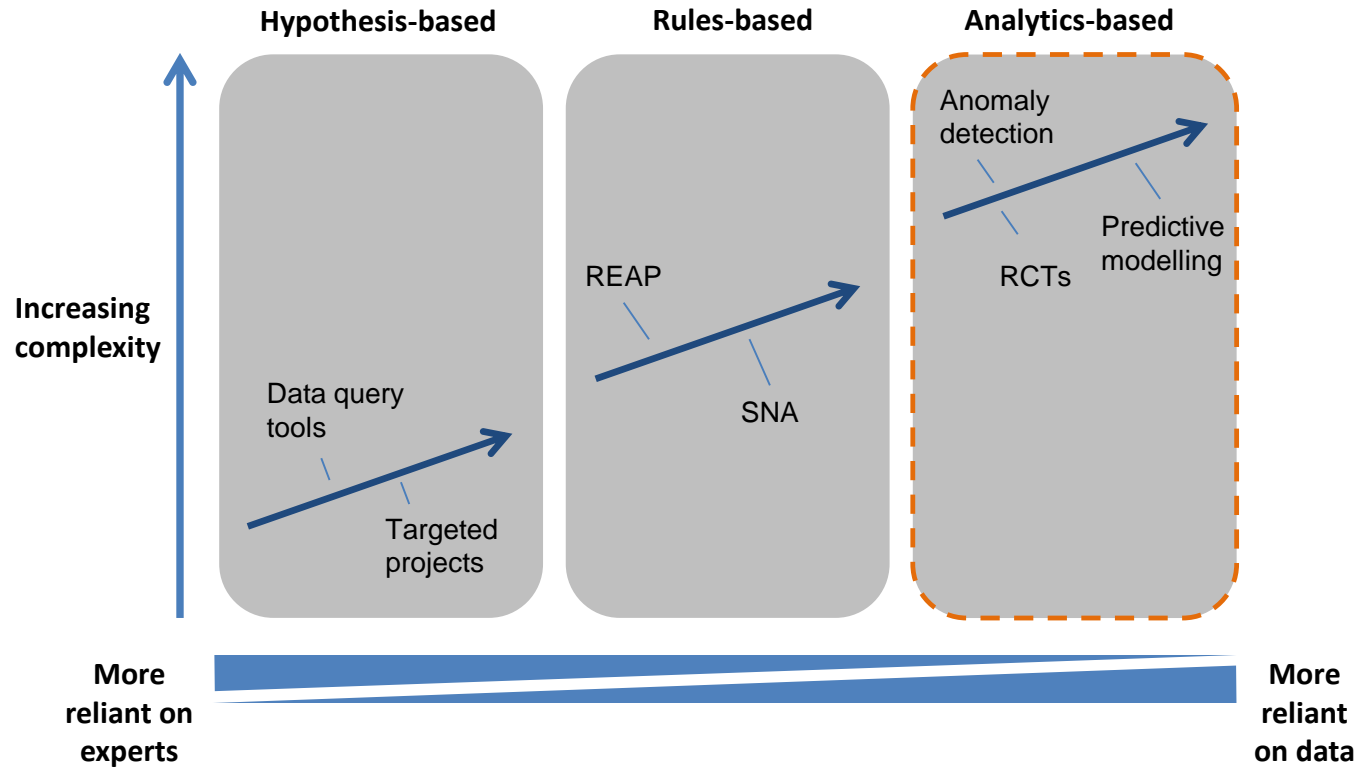
- Data sources
- Analytical methods

Challenges & opportunities

- Data quality & representativeness
- Natural taxation

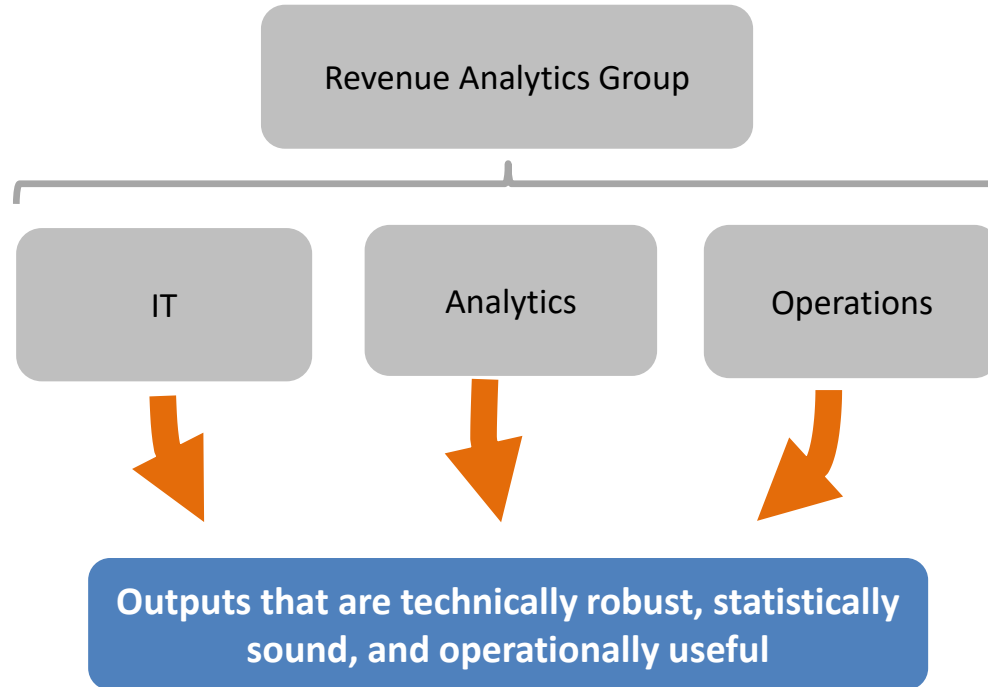
**Theme: Just because it's quantitative,
doesn't mean it's informative!**

Types of operational data use in Revenue



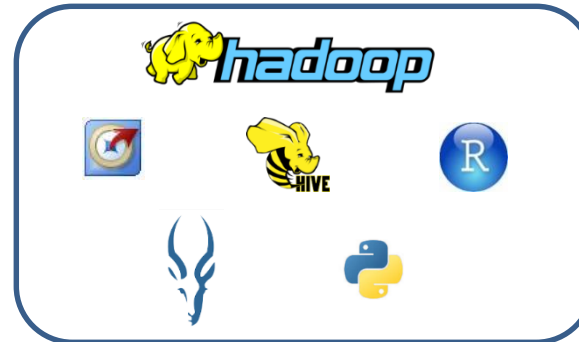
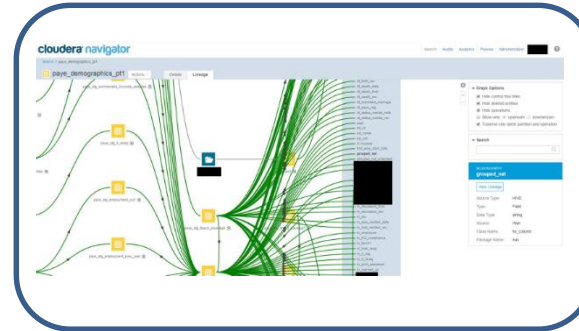
Organisational set-up

Strong governance ensures IT, operations, and analytics work together effectively



Data processes & warehouse designed specifically for analytics

- **Metadata key to realising value from diverse data-holdings**
 - Full tracking of data lineage in place
 - Populate metadata as tables are created
 - Working with dev teams to ensure metadata is created at source where possible
- **Software platform meets specific needs of analytics function**
 - Performance & reliability
 - Handles unstructured and semi-structured data
 - Access to a wide range of tools for data exploration and modelling
 - Strong data governance



Developing capabilities in-house



Identify and Recruit:

- Seek out suitable talent in-house
- Look for enthusiasm, and a background in natural or social sciences

Develop and Retain:

- Focus on developing programming skills
- Blend online and classroom training
- Provide diverse opportunities



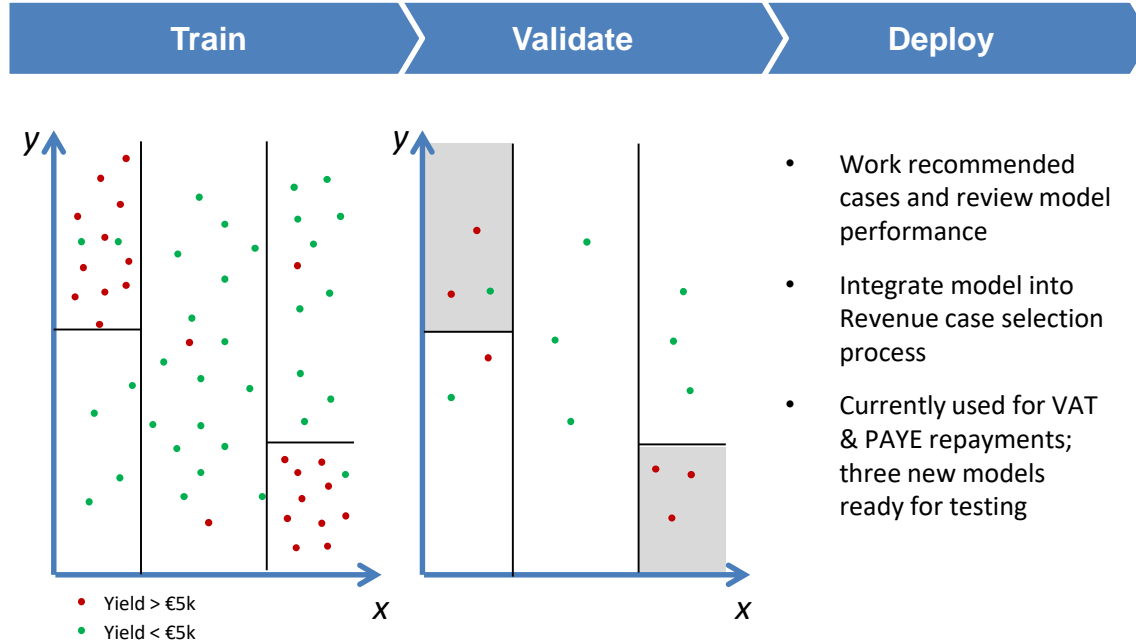
Analytical approaches

Overview of Revenue Data Sources

	Internal	External - Domestic	External - Foreign
Structured	<ul style="list-style-type: none">• Tax returns• Intervention outcomes• Filing behaviour• Registrations• Payments	<ul style="list-style-type: none">• Government bodies• Banks• Merchant acquirers• Letting agents• General requirements – e.g., Form 46G	<ul style="list-style-type: none">• Automatic exchange of information:<ul style="list-style-type: none">• Income & assets• Breakdown of corporate activities
Unstructured	<ul style="list-style-type: none">• Phone calls• Emails• Letters• Case notes	<ul style="list-style-type: none">• Suspicious Transaction Reports• Good Citizen Reports	<ul style="list-style-type: none">• Tax rulings• Spontaneous exchanges• Sundry other (eg. Panama Papers)

Revenue draws in millions of records annually – only selected sources shown here

Our ideal project: Models supervised by past intervention outcomes

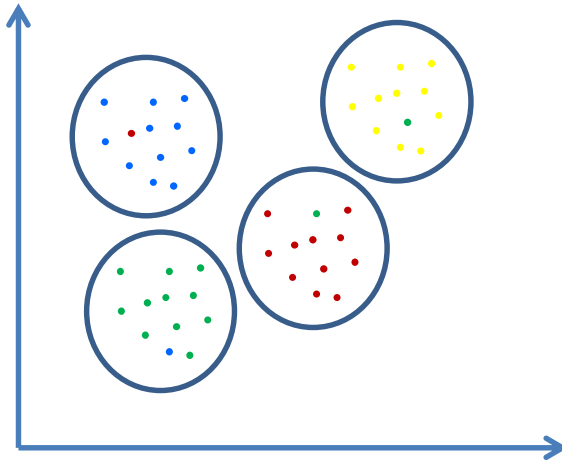


- Work recommended cases and review model performance
- Integrate model into Revenue case selection process
- Currently used for VAT & PAYE repayments; three new models ready for testing

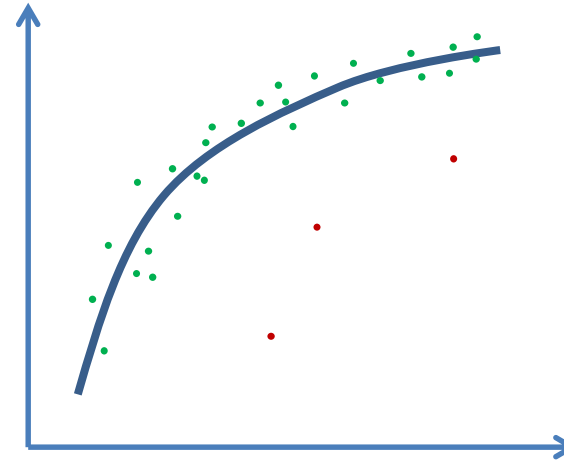
But case selection process may introduce substantial bias...

A compromise: Models for anomaly detection

Peer Groups (Mineral Oils, Construction)



Predicted Values (Income-Consumption)



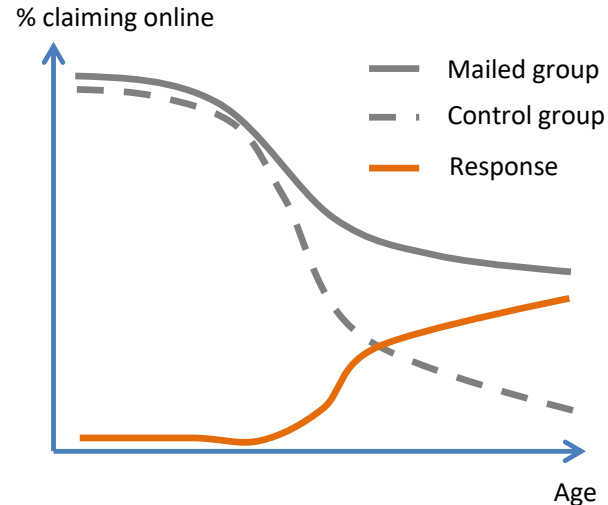
Analytics allows us to make sophisticated comparisons between taxpayers to identify outliers

A sideline: Use analytics to predict response to intervention

Approach taken

- Business objective to target campaigns aimed at persuading taxpayers to claim expenses online
- Initial hypothesis was that younger taxpayers should be targeted
- Controlled experiment run to assess incremental impact; model(s) built to 'predict' experimental results
- Found that older taxpayers responded more strongly

Model output



Predicting *outcomes* is not the same as predicting *response*

Challenges & opportunities

Why we are embracing a 'low-tech' approach

Challenges

- Unrepresentative training sets
- Variation in system usage, etc

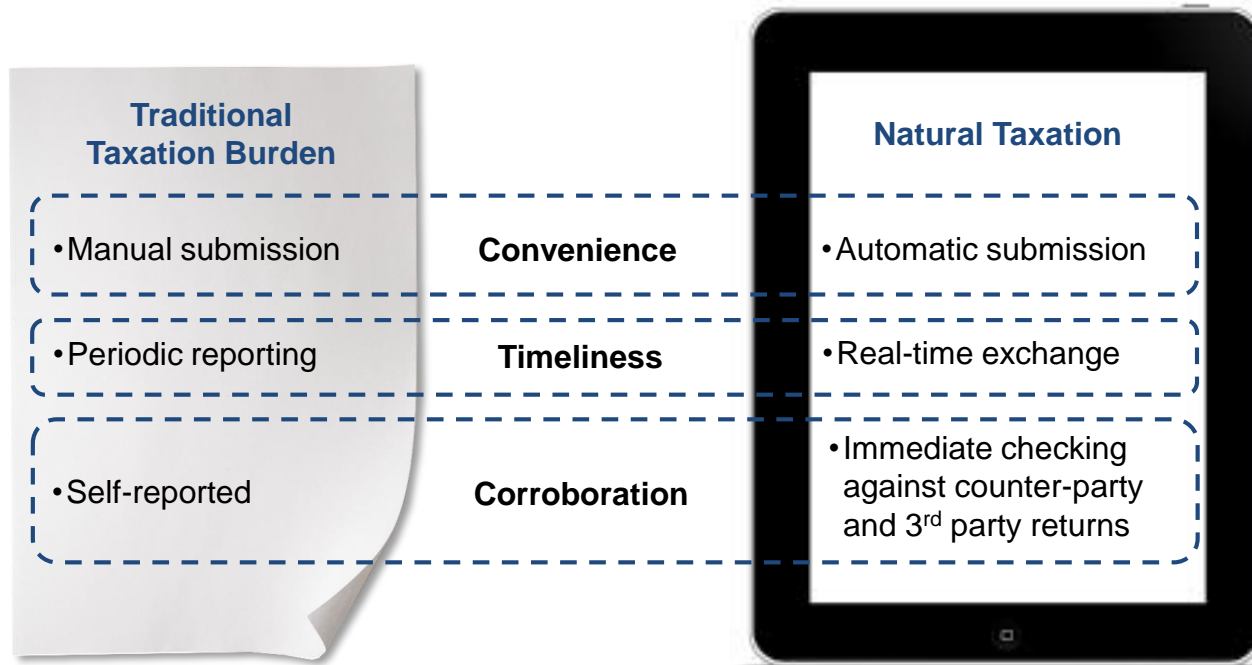
Implications

- Many relationships are just artefacts of data
- Can't just automate search for predictive patterns

(Attempted) Solution

'Low-tech' methods that business experts can review and understand make it much easier to weed out spurious patterns

Data without analytics?



Already under way through eRCT and PAYE real-time;
Opportunity to make compliance the default setting