An Introduction to Advanced Analytics and Data Mining

Dr Barry Leventhal

Henry Stewart Briefing on Marketing Analytics
19th November 2010

Barry
Analytics
Transforming Data

Agenda

- What are Advanced Analytics and Data Mining?

- The toolkit of data mining techniques

- Some issues to keep in mind
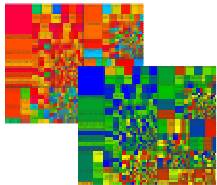
- Which technique should you use?

What is Data Mining?

A process of discovering and interpreting patterns in (often large) data sets in order to solve business problems

Data   Pattern   Information   Action
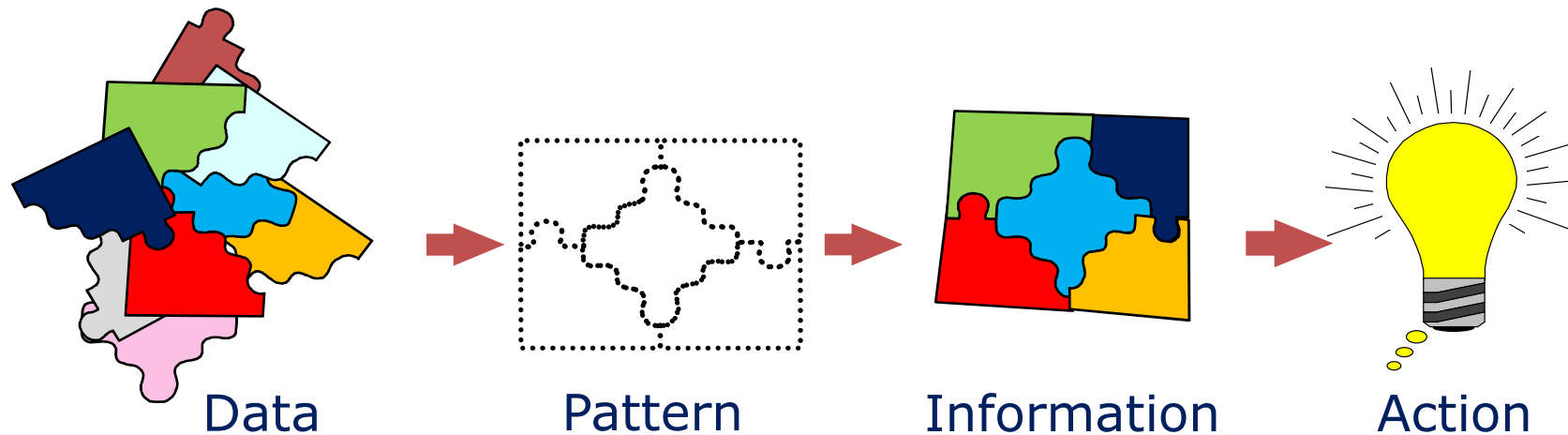
Converts Data into Information

Barry
Analytics
Transforming Data

What is Advanced Analytics?

Data Visualisation

Web Analytics

Text Analysis

Social Network Analysis

"*Any solution that supports the identification of meaningful patterns and correlations among variables in complex, structured and unstructured, historical, and potential future data sets for the purposes of predicting future events and assessing the attractiveness of various courses of action. Advanced Analytics typically incorporate such functionality as data mining, descriptive modelling, econometrics, forecasting, operations research optimisation, predictive modelling, simulations, statistics and text analytics.*" (Source: Forrester Research)
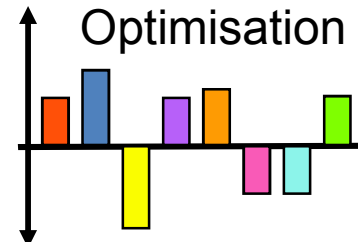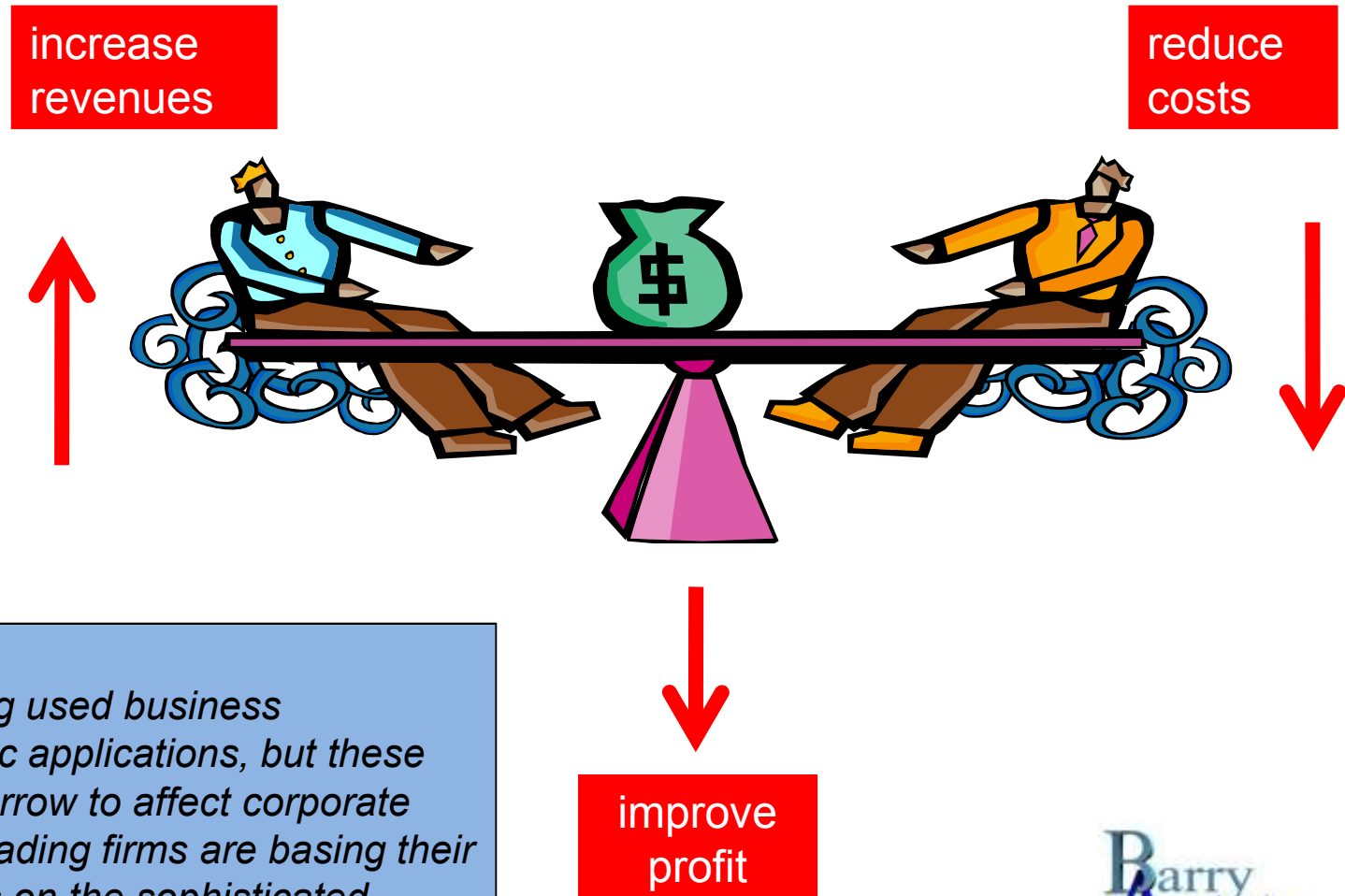
MINING

Data Mining

Simulation

Contact Optimisation

Barry Analytics

Transforming Data

How can Advanced Analytics help?

- By helping companies to increase revenues or reduce costs

increase revenues

reduce costs

improve profit

Tom Davenport:
"Companies have long used business intelligence for specific applications, but these initiatives were too narrow to affect corporate performance. Now, leading firms are basing their competitive strategies on the sophisticated analysis of business data."

Barry **A**nalytics
Transforming Data

Where can Advanced Analytics add Value?

The Toolkit of Data Mining Techniques
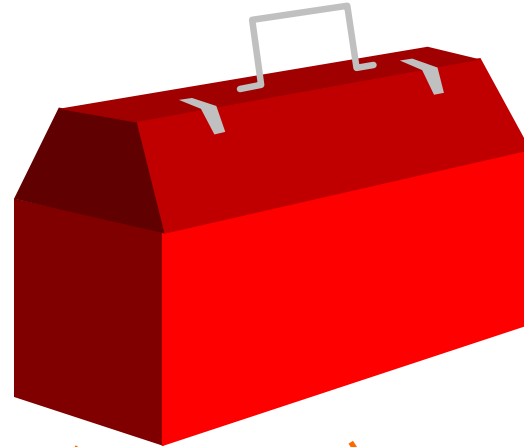
Traditional Statistics

- Regression Models
- Survival Analysis
- Factor Analysis
- Cluster Analysis
- CHAID

Machine Learning

- Rule Induction
- Neural Networks
- Genetic Algorithms

Barry Analytics

Transforming Data

Two main types of Analytical Model

Type 1: Models driven by a Target Variable

e.g. Which customers to cross sell?

- Implies building a Predictive Model

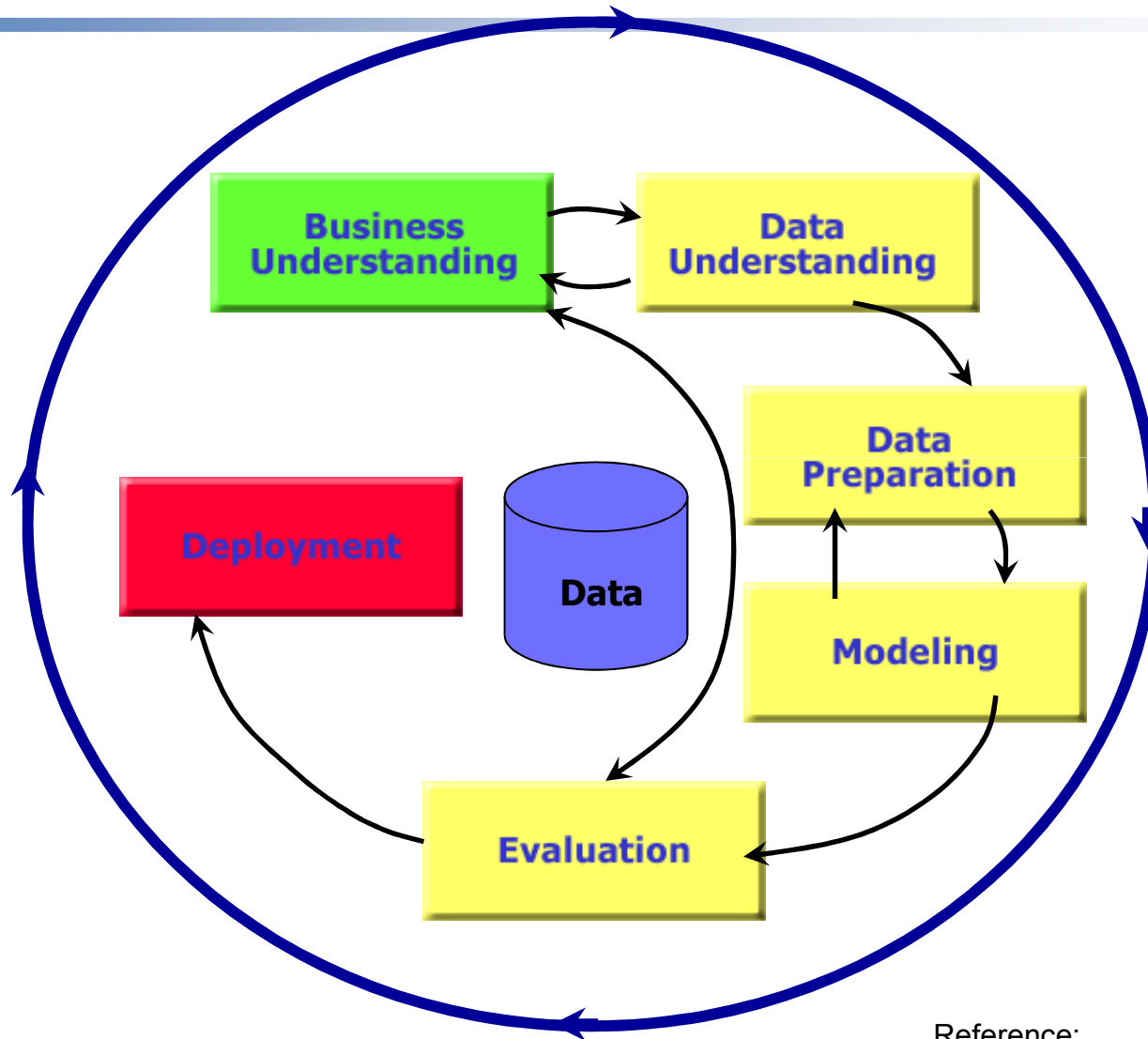**- 'Directed' Data Mining Techniques**

Type 2: Models with no Target Variable

e.g. What are our most important customer segments?

- Implies a Descriptive Model

- 'Undirected' Data Mining Techniques

Barry Analytics

Transforming Data

The Data Mining Process

Cross Industry Standard Process for
Data Mining

Reference:
Step-by-step data mining guide
CRISP-DM 1.0

Barry
Analytics
Transforming Data

# Some issues to keep in mind:
Issue 1: Use an appropriate technique

- Some years ago, the DMA Targeting & Statistics Group held a seminar to explain and compare four analytical techniques:
  - Cluster Analysis
  - Decision Tree
  - Neural Network (supervised)
  - Regression Model

- The four techniques were applied to a sample of lifestyle data in order to predict private healthcare cover

Some issues to keep in mind:
Issue 1: Use an appropriate technique

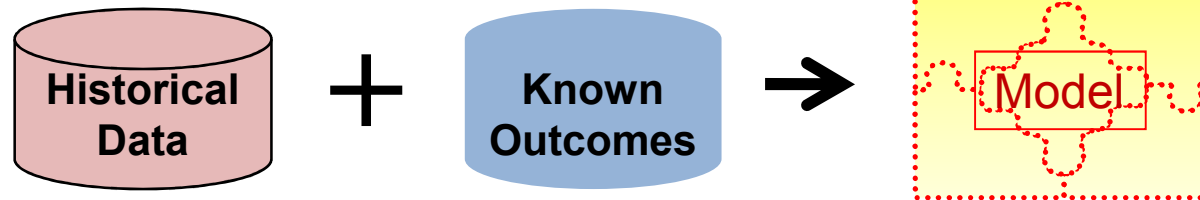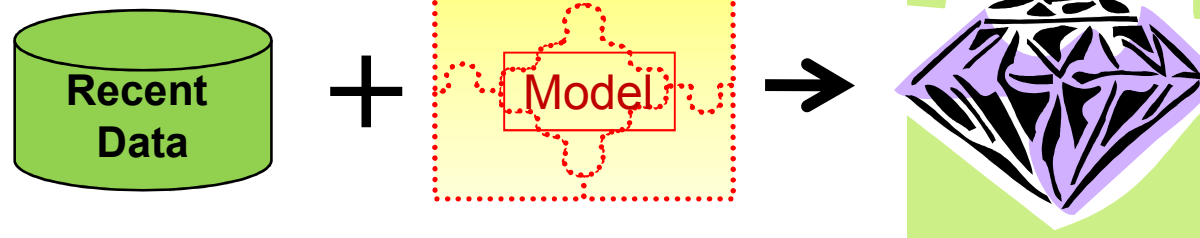Comparison of private healthcare targeting via four analytical techniques
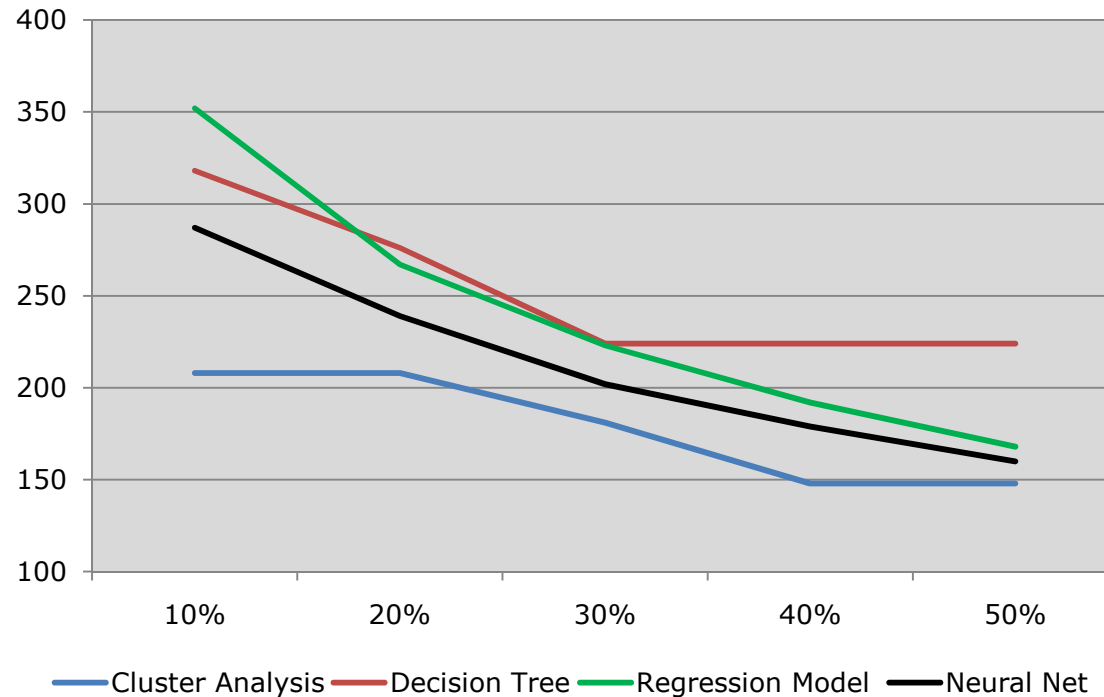
Source:

CMT/ DMA Targeting & Statistics Interest Group

Issue 2: Modelling and Deploying are separate stages in the data mining process

Modelling

Historical Data + Known Outcomes → Model

Deploying

Recent Data + Model → Predictions

- "Modelling" is 'one-off' – until model requires rebuild
- "Deploying" takes place repeatedly
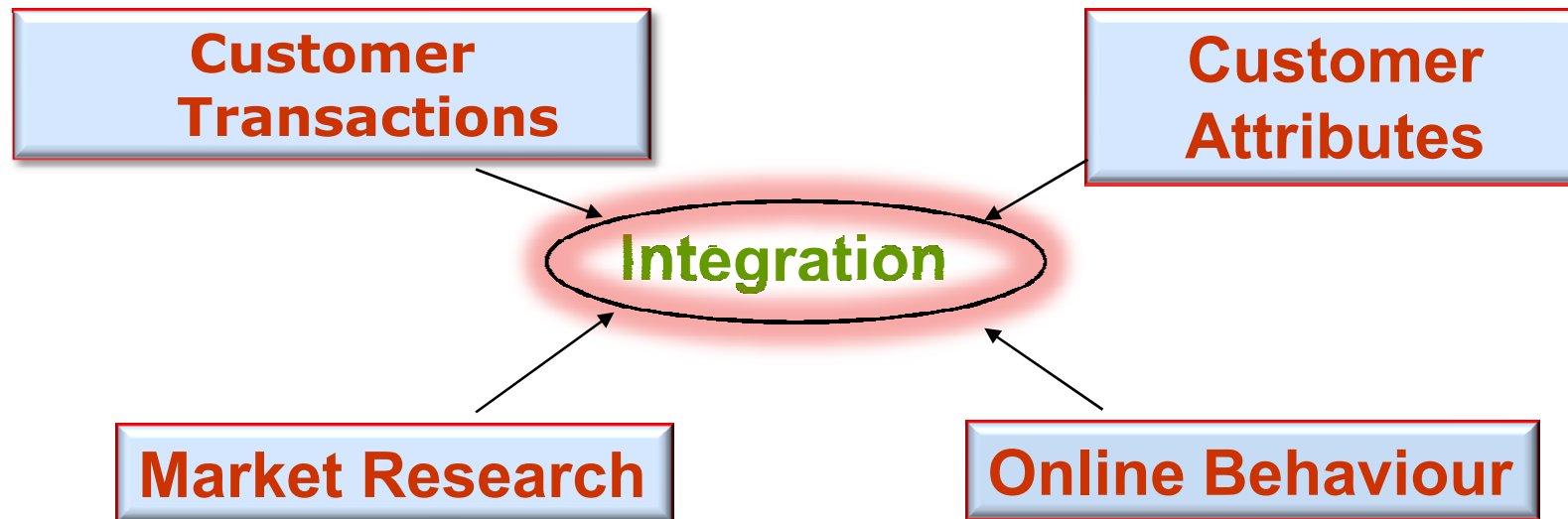
Issue 3: Do <u>not</u> forget your data!

Your data is the key to gaining value from analytics and modelling – essentials to consider:

- Data quality

- Data predictivity

- Data integration

- Data governance

The Importance of Data Integration

- Business value increased by integrating complementary datasets
- New insights may be created by data integration, e.g.

Many applications of data integration...

- Predictive models to target behaviours identified by research

- Integration of web, email and traditional offline channels

- Tracking across channels, e.g. Attribution of media effects

Which analytical technique should you use ?

The choice generally depends on…

- business problem
- whether problem is predictive or descriptive
- underlying data environment
- variables to be predicted or described
- ability to implement solution
- whether key statistical assumptions hold

- Obtain help from a Statistician or Data Mining Consultant
- All about the problem, not the technique
- Combination of approaches works best

Thank you!

Barry Leventhal

+44 (0)7803 231870
Barry@barryanalytics.com