SMARTVISION Europe

Premier Business Partner IBM

Authorized Software Value Plus
Business Analytics

Business Analytics Award
Winner 2012

# Operational Analytics for Predictive & Proactive Maintenance

# Agenda

- Introduction to operational & predictive analytics
- Worked examples of operational analytics
  - Practical examples
- Break
- Demonstration of capabilities
  - Model development & text mining
- Best practices & maximising success
  - Analytical methodology, resources & deployment
- Summary & conclusion
- Lunch

SMARTVISION Europe

# Predictive Analytics for Smarter Business

Premier Business Partner IBM

- Premium, accredited partner to IBM specialising in the SPSS Advanced Analytics suite.
- Team each has 15 to 20 years of experience working in the predictive analytic space - specifically as senior members of the heritage SPSS team

# What do we mean by ‘Predictive Analytics’?

**Predictive analytics** encompasses a variety of techniques from statistics and data mining that analyze current and historical data to make predictions about future events

Analysis of structured and unstructured information with mining, predictive modelling, and 'what-if?' scenario analysis.

# What is operational analytics for preventative maintenance?

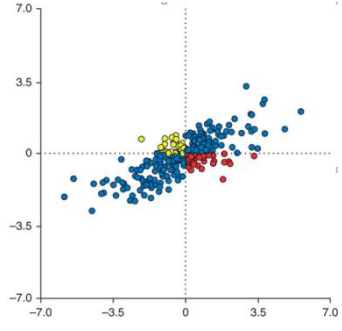
Understanding the patterns in operational data to determine the areas of greatest risk and directing resources *before risk becomes reality.*

SMARTVISION Europe

A SELECT INTERNATIONAL COMPANY

# What do we mean by ‘predictive analytics’?

- It’s different from business intelligence or MI reporting
- Actually, it’s not *always* about prediction
- However, predictive analytics *does* creates important new data
- These data take the form of estimates, probabilities, forecasts, recommendations, propensity scores, classifications or likelihood values
- Which in turn can be incorporated into key operational and/or insight systems



# Predictive operational analytics: industry sectors

SMARTVISION Europe

A SELECT INTERNATIONAL COMPANY

# Predictive operational analytics: common applications

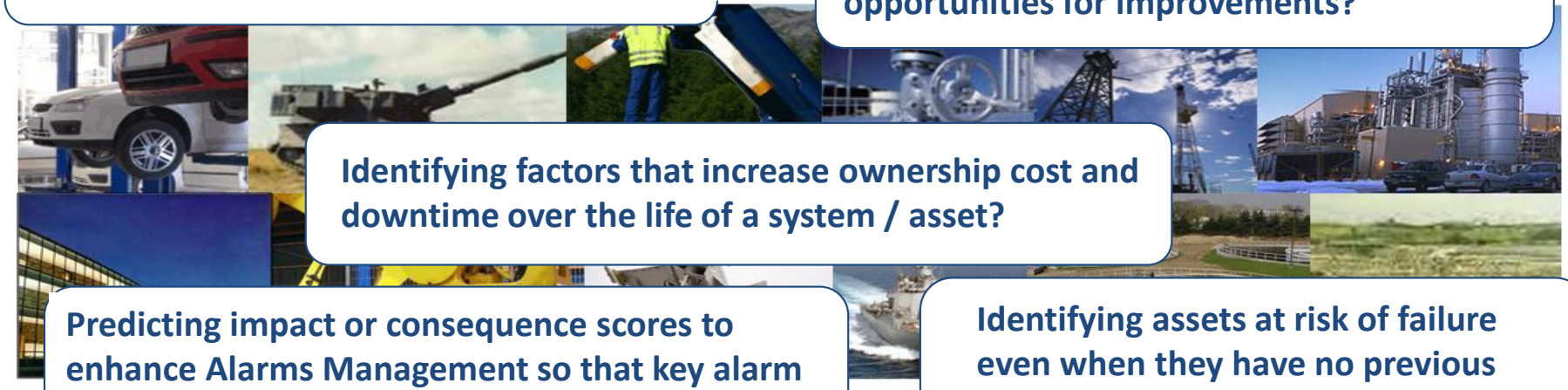
SMARTVISION Europe

A SELECT INTERNATIONAL COMPANY

# How can *operational analytics* help?

Unearthing characteristics that lead to an increased frequency of failures?

Finding patterns in maintenance operations that could point to opportunities for improvements?



Identifying factors that increase ownership cost and downtime over the life of a system / asset?

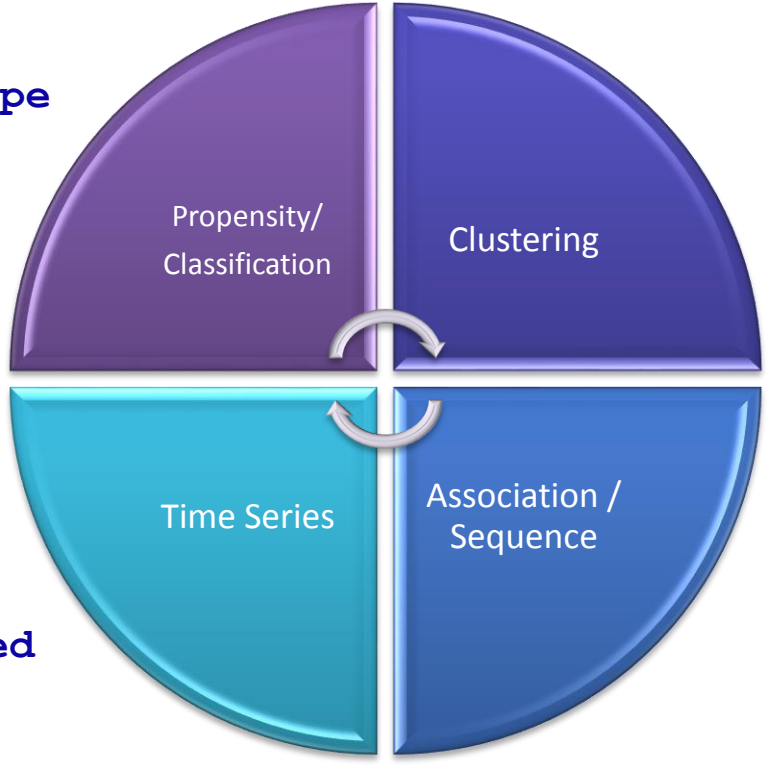
Predicting impact or consequence scores to enhance Alarms Management so that key alarm events are prioritized

Identifying assets at risk of failure even when they have no previous failure history

Mining *free text* from thousands of logs that describe the maintenance performed on systems to accurately categorize maintenance reports and identify areas of risk

# Types of predictive modelling…

**Predict a particular type of outcome**




**Identify groups within a population displaying homogeneity (based on a wide array of data)**

**Forecast a future value over a defined time period**

**Identify repeatable patterns of behaviour or sequence…**

# Types of predictive modelling…

- Classification / propensity
  - How likely is this asset (vehicle / pump / property / meter) to fail / report and issue?
- Clustering
  - How can I divide plant / asset portfolio into meaningful and discernible groups as a framework for proactive maintenance / inspection regimes?
- Association & sequence
  - What is the sequence & cadence of recorded events that can be identified as the antecedents of an asset failure in a specific location?
- Time series
  - What  is production line downtime going to be next month / quarter / year?

# Other SPSS predictive maintenance & quality customers

Scottish Water
Always serving Scotland

BMW

South Staffs Water

ERICSSON

LIFESCAN
a Johnson & Johnson company

YorkshireWater

L&Q
creating places where people want to live

Sikorsky
A United Technologies Company

SMARTVISION
Europe

A SELECT INTERNATIONAL COMPANY

# Operational analytics: worked examples

**Jarlath Quinn**

# Effective operational analytics applications…

Environmental

- Weather Conditions
- Ambient Temperature

Interaction

- Maintenance History
- Notes from inspection
- Customer Feedback

Assets

- Machine
- Material
- Age

Behavioural

- Telemetry
- Alarms
- Events (Failure ,Faults)

## Utilise historical data from multiple sources…

# Effective operational analytics applications…

Environmental

- Weather Conditions
- Ambient Temperature

Interaction

- Maintenance History
- Notes from inspection
- Customer Feedback

Assets

- Machine
- Material
- Age

Behavioural

- Telemetry
- Alarms
- Events (Failure ,Faults)

…to build accurate, testable predictive models…

**19% Likelihood new filter required**

*22% chance of Failure*

0.43 probability of repeat error

Estimated Temperature = 26.2

# Effective operational analytics applications…

Environmental

- Weather Conditions
- Ambient Temperature

Interaction

- Maintenance History
- Notes from inspection
- Customer Feedback

Assets

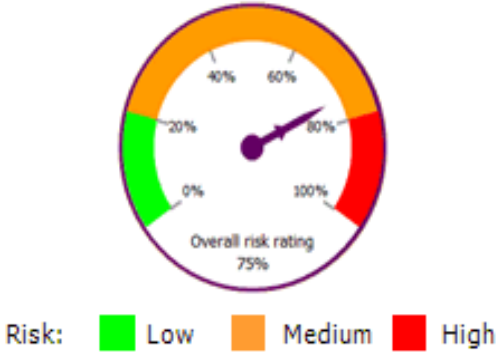
- Machine
- Material
- Age

Behavioural

- Telemetry
- Alarms
- Events (Failure ,Faults)

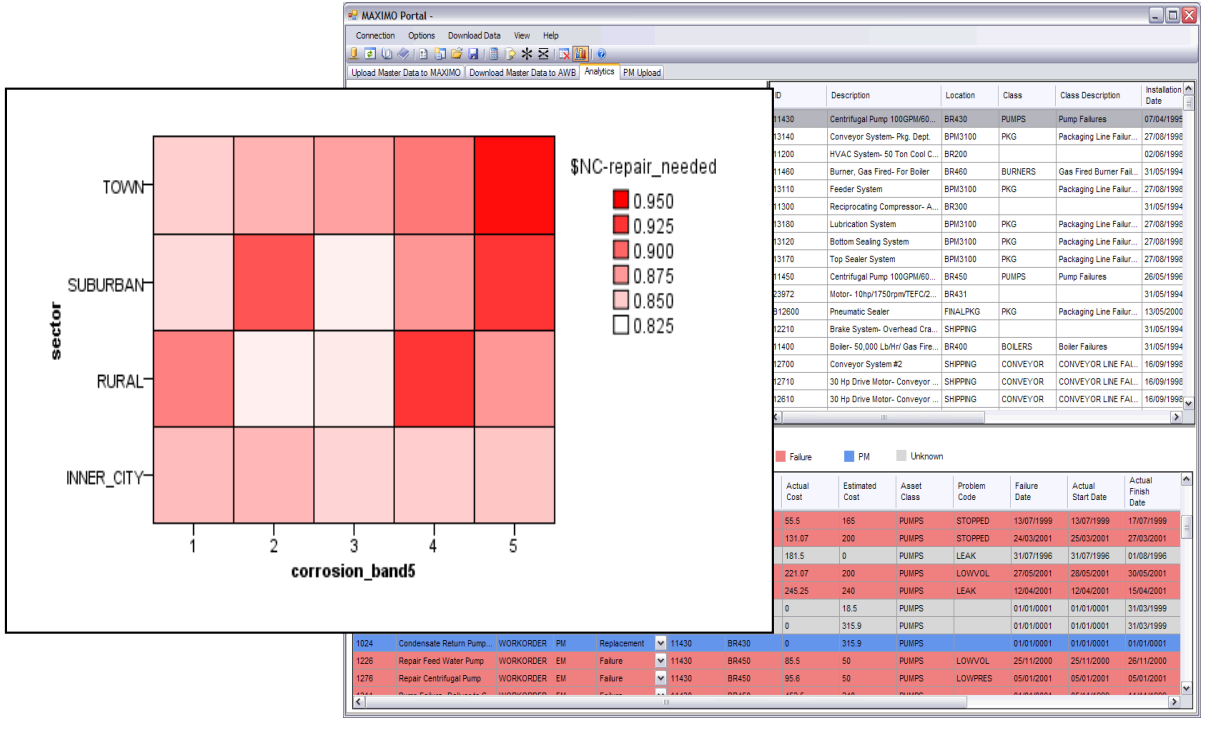
…to generate predictions and risk scores ….

Model



# Effective operational analytics applications…

…that can be deployed into operational systems and other insight/reporting platforms




SMARTVISION Europe

A SELECT INTERNATIONAL COMPANY

# Effective operational analytics applications…

## …to make smarter decisions

| | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| 1 | Turbine ID | Previous Risk Score | Current Risk Score | Previous Risk Score_pct_rank | Current Risk Score_pct_rank | Risk_rank_change | Current Predicted Asset Failure |
| 231 | 230 | 54 | 54 | 27 | 31 | -4 | T |
| 232 | 231 | 15 | 9 | 67 | 75 | -8 | F |
| 233 | 232 | 24 | 20 | 34 | 44 | -10 | F |
| 234 | 233 | 15 | 53 | 67 | 32 | 35 | T |
| 235 | 234 | 15 | 20 | 67 | 44 | 23 | F |
| 236 | 235 | 15 | 9 | 67 | 75 | -8 | F |
| 237 | 236 | 15 | 2 | 67 | 98 | -31 | F |
| 238 | 237 | 15 | 9 | 67 | 75 | -8 | F |
| 239 | 238 | 62 | 62 | 23 | 24 | -1 | T |
| 240 | 239 | 62 | 62 | 23 | 24 | -1 | T |
| 241 | 240 | 17 | 17 | 38 | 51 | -13 | F |
| 242 | 241 | 15 | 20 | 67 | 44 | 23 | F |
| 243 | 242 | 62 | 62 | 23 | 24 | -1 | T |
| 244 | 243 | 15 | 21 | 67 | 39 | 28 | F |
| 245 | 244 | 15 | 15 | 67 | 55 | 12 | F |
| 246 | 245 | 83 | 72 | 13 | 18 | -5 | T |
| 247 | 246 | 15 | 8 | 67 | 90 | -23 | F |
| 248 | 247 | 14 | 14 | 96 | 57 | 39 | F |
| 249 | 248 | 90 | 90 | 6 | 7 | -1 | T |
| 250 | 249 | 90 | 90 | 6 | 7 | -1 | T |
| 251 | 250 | 90 | 90 | 6 | 7 | -1 | T |
| 252 | 251 | 90 | 90 | 6 | 7 | -1 | T |

Sheet1 Sheet2

Select destination and press ENTER or choose Paste

100%

# Consolidate the data that seems most relevant to the application

Asset Register

Meteorological/Location Data

Maintenance History

Load/Monitoring Data

| Field | Sample Graph | Measurement | Min | Max | Mean | Std. Dev | Skewness | Unique | Valid |
|---|---|---|---|---|---|---|---|---|---|
| Month recorded | | Nominal | -- | -- | -- | -- | -- | 12 | 3660 |
| Section | | Continuous | 2.000 | 4866.000 | 2433.925 | 1420.193 | -0.002 | -- | 3660 |
| Zone | | Nominal | -- | -- | -- | -- | -- | 4 | 3660 |
| Previous Over current Trips | | Continuous | 0.000 | 13.000 | 0.725 | 1.484 | 3.131 | -- | 3660 |
| Lubrication Type | | Continuous | 0.000 | 21.000 | 0.238 | 1.120 | 8.734 | -- | 3660 |
| New Seal | | Continuous | 0.000 | 15.000 | 0.841 | 1.358 | 2.658 | -- | 3660 |
| Planned shutdowns | | Continuous | 0.000 | 30.000 | 2.069 | 2.910 | 2.476 | -- | 3660 |
| Average RPM | | Continuous | 0.000 | 20195.000 | 3815.406 | 2397.339 | 0.815 | -- | 3660 |
| Asset Age Score | | Continuous | 0.000 | 1458.000 | 170.622 | 224.106 | 2.138 | -- | 3660 |
| Average Vibration Score | | Continuous | 0.000 | 1613.000 | 359.111 | 293.049 | 1.019 | -- | 3660 |
| Torque_rating | | Continuous | 0.000 | 73.000 | 3.451 | 5.020 | 3.046 | -- | 3660 |
| Bearing Weight Score | | Continuous | 0.000 | 653.000 | 53.145 | 81.452 | 3.097 | -- | 3660 |
| Insulation Rating | | Continuous | 0.000 | 327.000 | 65.013 | 46.979 | 0.909 | -- | 3660 |
| Fatigue Score | | Continuous | 0.000 | 35.114 | 4.797 | 7.747 | 1.657 | -- | 3660 |
| Asset Failure | | Flag | -- | -- | -- | -- | -- | 2 | 3660 |
| Temperature_Class_One_Week_Before | | Nominal | -- | -- | -- | -- | -- | 4 | 3660 |
| Average Pressure Score | | Continuous | 5.615 | 42.692 | 17.248 | 6.957 | 0.686 | -- | 3660 |

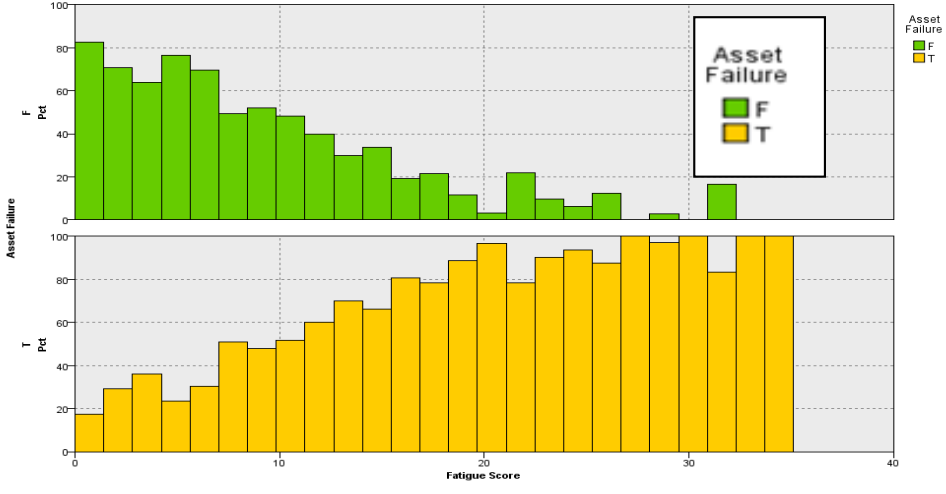
[1] Indicates a multimode result [2] Indicates a sampled result

SMARTVISION Europe

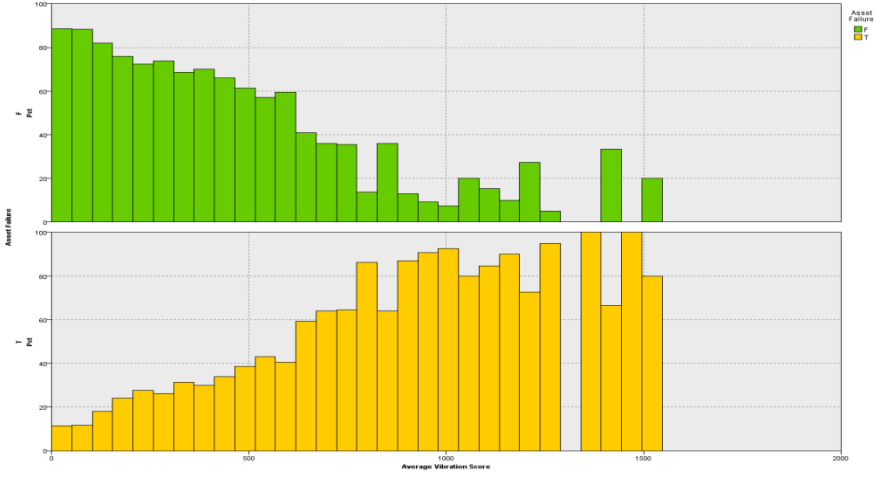
A SELECT INTERNATIONAL COMPANY

# Visualise the data and identify potential predictive indicators

- **Corrosion/fatigue score**
- Higher the degree of corrosion
- Higher the risk of asset failure




- **Average gas pressure score**
- Lower the sustained pressure score
- Higher the risk of failure/discharge




SMARTVISION Europe

A SELECT INTERNATIONAL COMPANY

# Don't ignore unstructured data

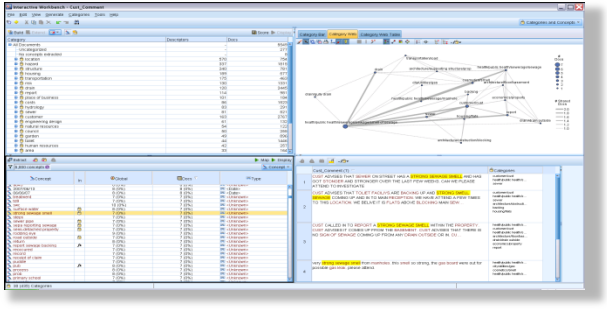
**Text mining produces *structured data from unstructured:***

***example from Water Industry***

- "Tried to clear but they reckon its on the main sewer line - causing backup inside toilet - neighbour across the back has been having similar problems and we found a blockage on the main - can we check?"

Text mining gives

- Main sewer
- Backup
- Blockage



- "Possible discharge of cooking fat from lateral into main sewer as there is a block outside the takeaway."

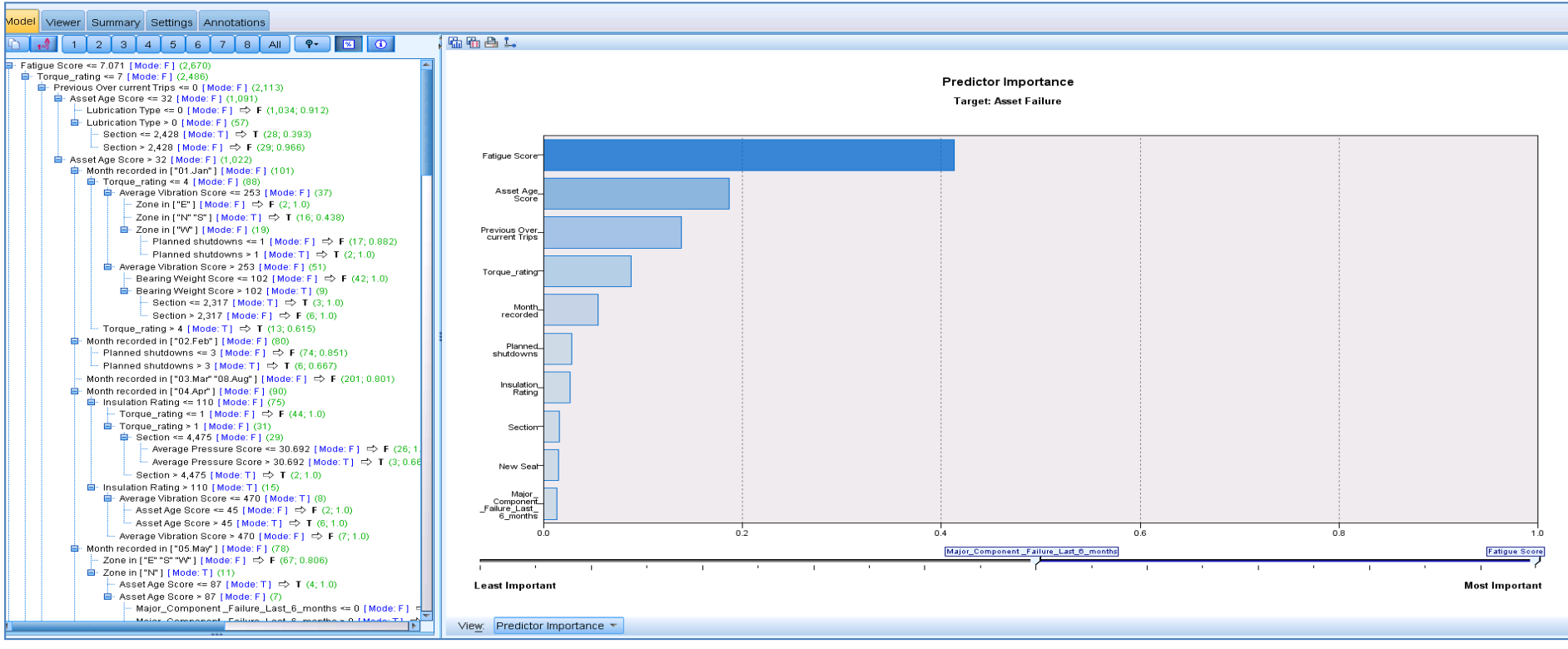
Text mining gives

- Fat problem
- Lateral sewer
- Property type

SMARTVISION Europe

A SELECT INTERNATIONAL COMPANY

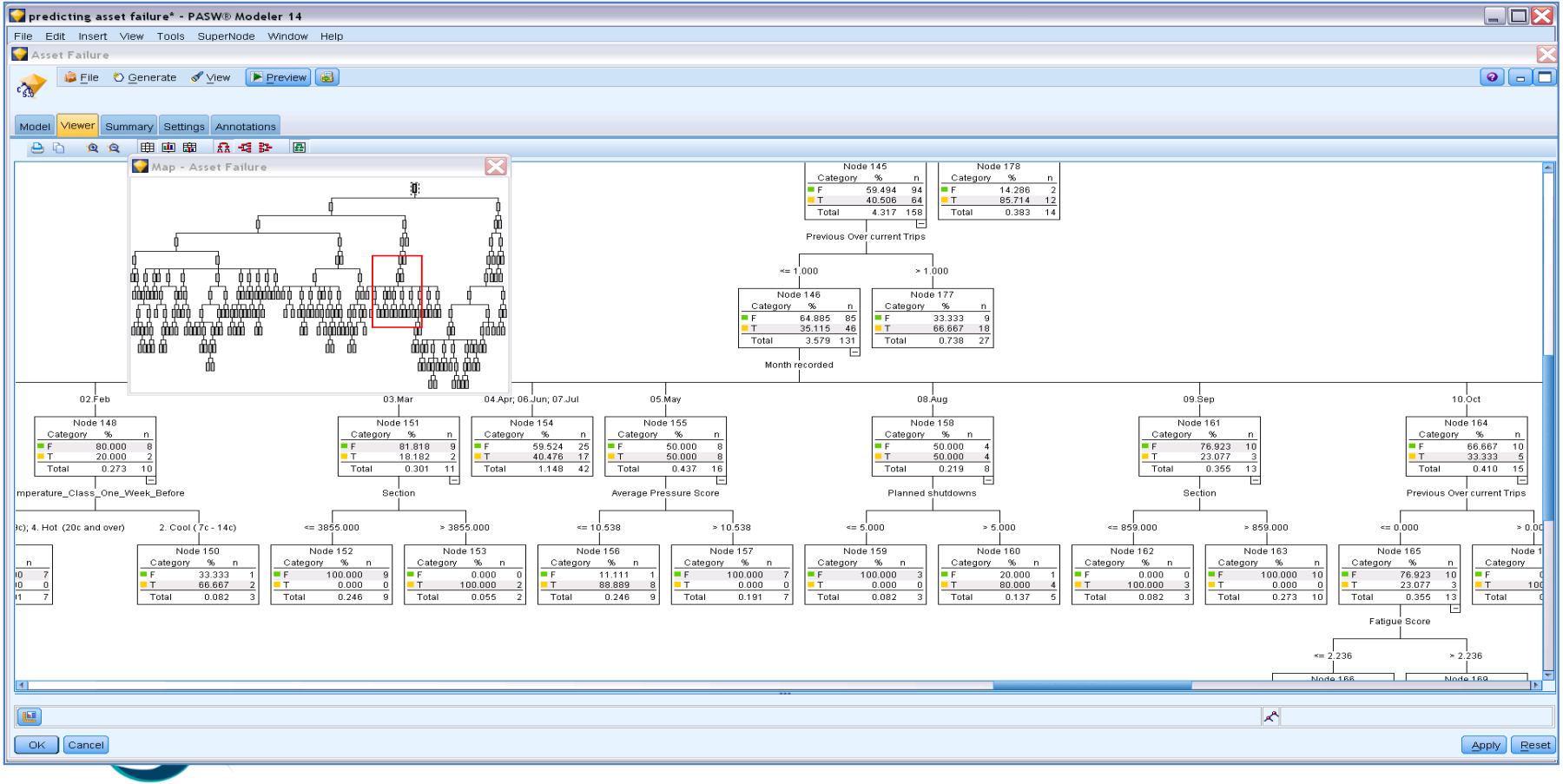
# Make sure the model makes sense




SMARTVISION Europe

A SELECT INTERNATIONAL COMPANY

# Example of an actual reusable predictive model



# Model evaluation: what does 'success' look like?

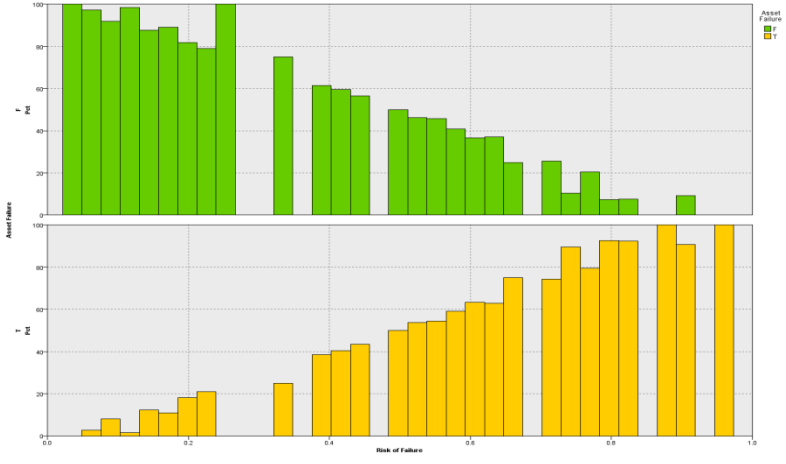
Results for output field Asset Failure

Comparing $C-Asset Failure with Asset Failure

| | | |
|---|---|---|
| **Correct** | 3,079 | 84.13% |
| **Wrong** | 581 | 15.87% |
| **Total** | 3,660 | |

Coincidence Matrix for $C-Asset Failure (rows show actuals)

| | F | T |
|---|---|---|
| **F** | 2,067 | 363 |
| **T** | 218 | 1,012 |




## Model classification

- 84% accuracy in predicting asset failure
- Chart shows strong correlation between estimated risk of failure and actual failures

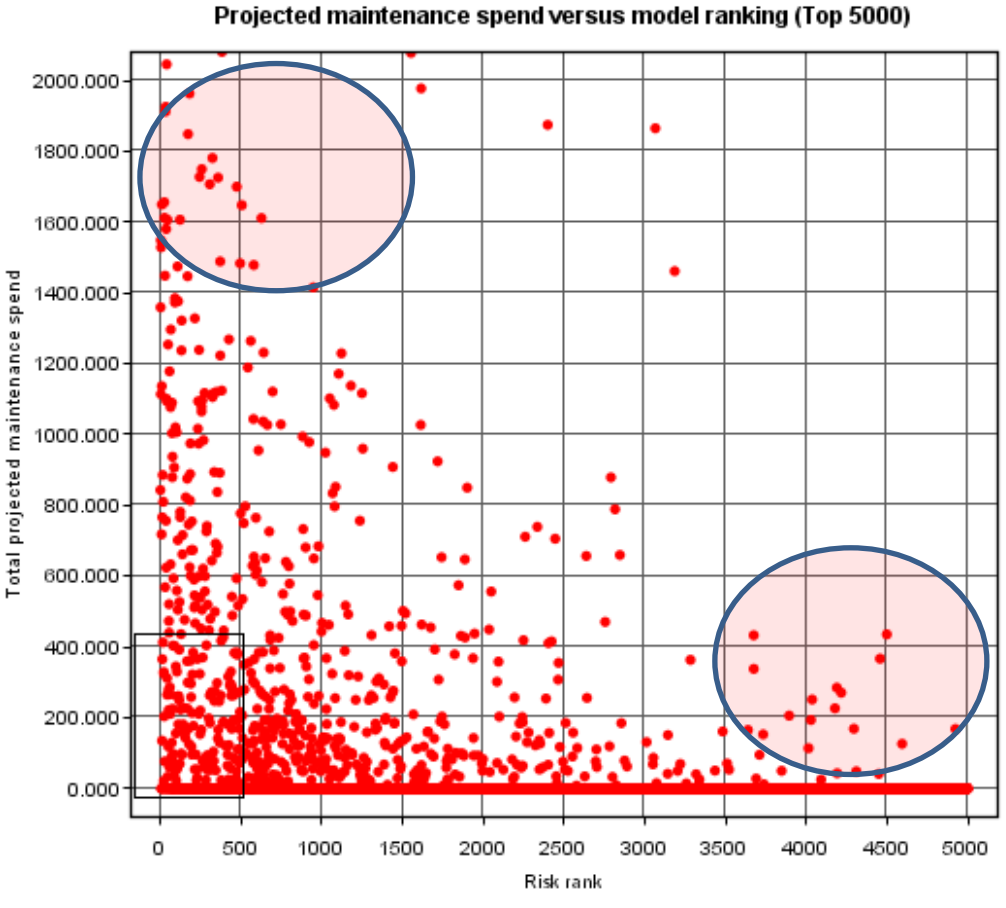
# What does ‘deployment’ look like?

- Assets in red have a high risk profile but no previous issues

| Ranking | Asset information | | Prioritisation of siltation | | | History of incidents | | | | Recency | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ranking | Asset id | Type of asset | Likelihood of siltation | Consequence | Total Risk - Siltation | Pollution Count | Internal Count | External Count | Back Up Count | Last reported siltation issue | Estimated siltation frequency (months) | Number of historic reported issues |
| 1 | 43723952 | Main | 0.64 | 107.1 | **68.8** | 0 | 21 | 2 | 0 | 17/12/2007 | 4 | 8 |
| 2 | 101405798 | Main | 0.99 | 56.3 | **55.7** | 0 | 11 | 0 | 0 | 13/12/2006 | 8 | 2 |
| 3 | 78583469 | Main | 0.97 | 55.1 | **53.5** | 0 | 11 | 0 | 0 | 06/02/2007 | 8 | 3 |
| 4 | 43804534 | Main | 0.80 | 55.1 | **44.0** | 0 | 11 | 0 | 0 | 22/04/2006 | 8 | 3 |
| 5 | APKUHQ8H5CL44HU083_lat | Public lateral | 0.65 | 66.0 | **43.1** | 2 | 11 | 1 | 0 | 22/01/2005 | | 1 |
| 6 | 43724270 | Main | 0.74 | 58.1 | **42.9** | 0 | 11 | 3 | 0 | | | 0 |
| 7 | AP203W8M5Y153GCGNM_lat | Public lateral | 0.78 | 55.0 | **42.7** | 7 | 9 | 1 | 0 | 12/09/2007 | 32 | 2 |
| 8 | APPMB88Q5MP44WU02H_lat | Public lateral | 0.45 | 94.0 | **42.5** | 0 | 18 | 0 | 0 | 09/08/2007 | 1 | 7 |
| 9 | APE4AA8W5Y442TW075_lat | Public lateral | 0.77 | 54.0 | **41.7** | 0 | 10 | 0 | 0 | 20/08/2006 | 12 | 3 |
| 10 | APF4E98W5R044KTGLR_lat | Public lateral | 0.74 | 55.0 | **40.9** | 1 | 10 | 1 | 0 | | | 0 |
| 11 | 43724570 | Main | 0.96 | 41.1 | **39.6** | 0 | 8 | 1 | 0 | | | 0 |
| 12 | APM1HT8H53E43GU0Q0_lat | Public lateral | 0.79 | 50.0 | **39.3** | 0 | 9 | 1 | 0 | | | 0 |
| 13 | AP58F18C5B955LE0WX_lat | Public lateral | 0.41 | 94.0 | **38.4** | 0 | 18 | 0 | 0 | 23/02/2007 | 4 | 6 |
| 14 | APXR988M5Y8420T0GM_lat | Public lateral | 0.77 | 49.0 | **37.8** | 0 | 9 | 0 | 0 | 02/11/2006 | 10 | 3 |
| 15 | APJVEC8R5L04YCUGYH_drain | Private drain | 0.80 | 46.0 | **37.0** | 0 | 6 | 0 | 6 | | | 0 |
| 16 | 43723196 | Main | 0.71 | 50.1 | **35.4** | 0 | 10 | 0 | 0 | | | 0 |
| 17 | APNV9L8M5TL427T0YM_lat | Public lateral | 0.39 | 90.0 | **35.3** | 0 | 17 | 1 | 0 | 08/05/2007 | 1 | 7 |
| 18 | AP7VG38R8R14A6W0NG_lat | Public lateral | 0.79 | 44.0 | **34.6** | 0 | 8 | 0 | 0 | | | 0 |
| 19 | 44051115 | Main | 0.76 | 45.1 | **34.4** | 0 | 9 | 0 | 0 | 05/11/2007 | 10 | 4 |
| 20 | APKQEV8R56K437TGQH_lat | Public lateral | 0.79 | 42.9 | **33.7** | 1 | 7 | 0 | 0 | 06/08/2007 | 11 | 4 |

# Model scores open new doors of insight

- Risk becomes a new *dynamic* metric
- Risk can be viewed in terms of –
  - projected spend
  - asset value
  - failure consequence
  - maintenance cost



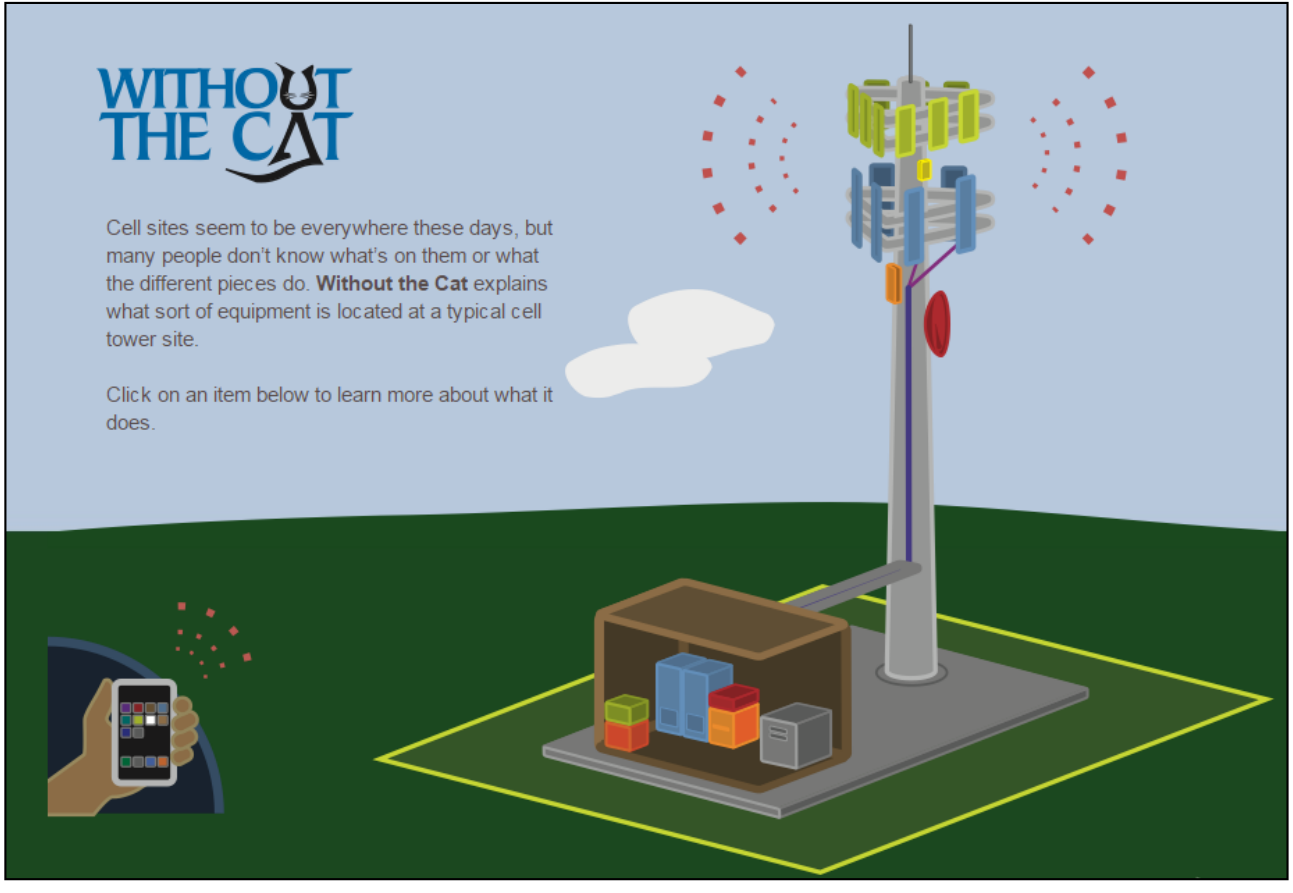

SMARTVISION Europe

Y

SMARTVISION
Europe

# Let’s See A Demonstration

# Cell site maintenance



SMARTVISION
Europe

# Best Practice

## What are the common ingredients of successful applications?

### Using multiple data sources

- Fixed attributes
  - Asset data
    - Model type/class
    - Specification
      - Weight
      - Size
      - Range
- Dynamic attributes
  - Maintenance history
  - Usage history
  - Part replacements
  - Maintenance reports (free text )
  - Operating environment

Asset data

Usage

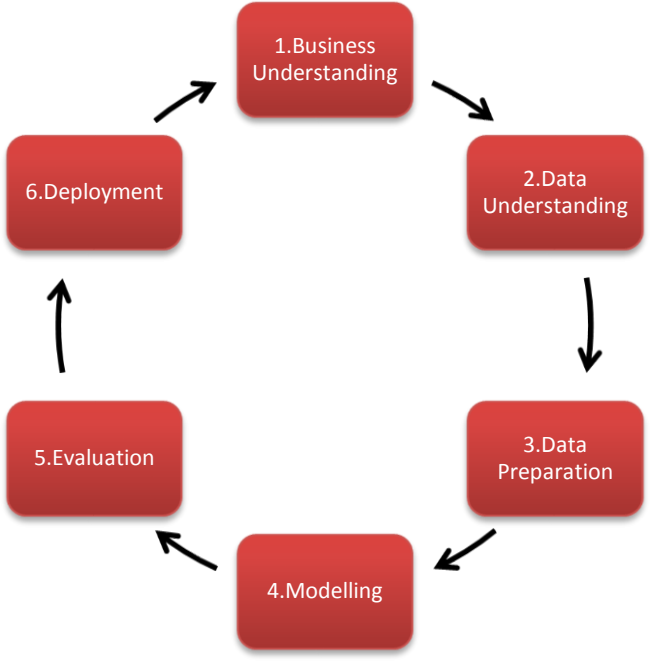
Maintenance history

Environmental/ telematics

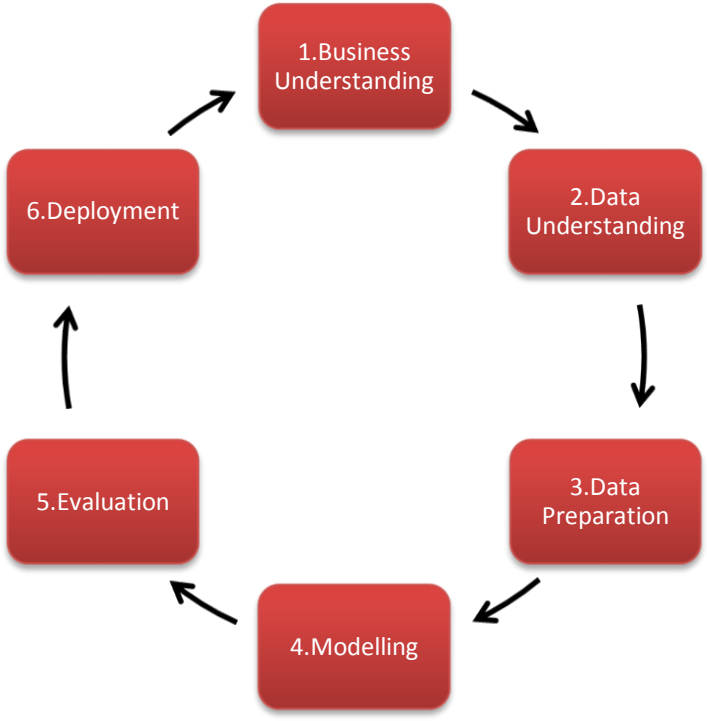
# What are the common ingredients of successful applications?

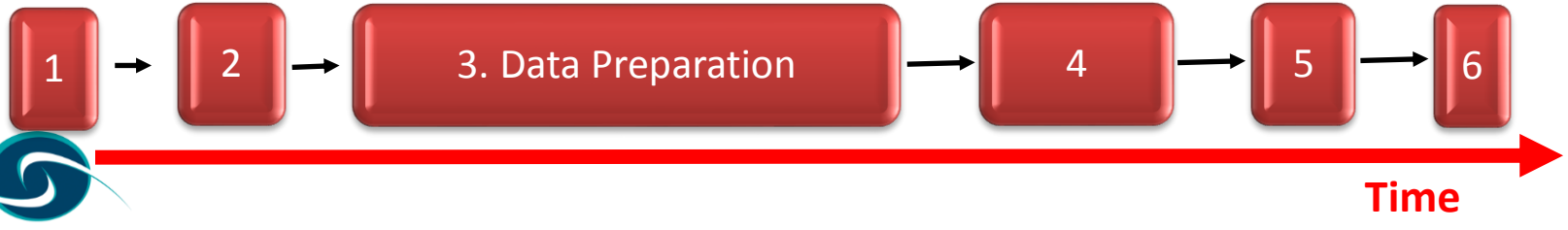
## Utilising a powerful, proven methodology

– CRISP-DM: Cross Industry Standard Process for Data Mining

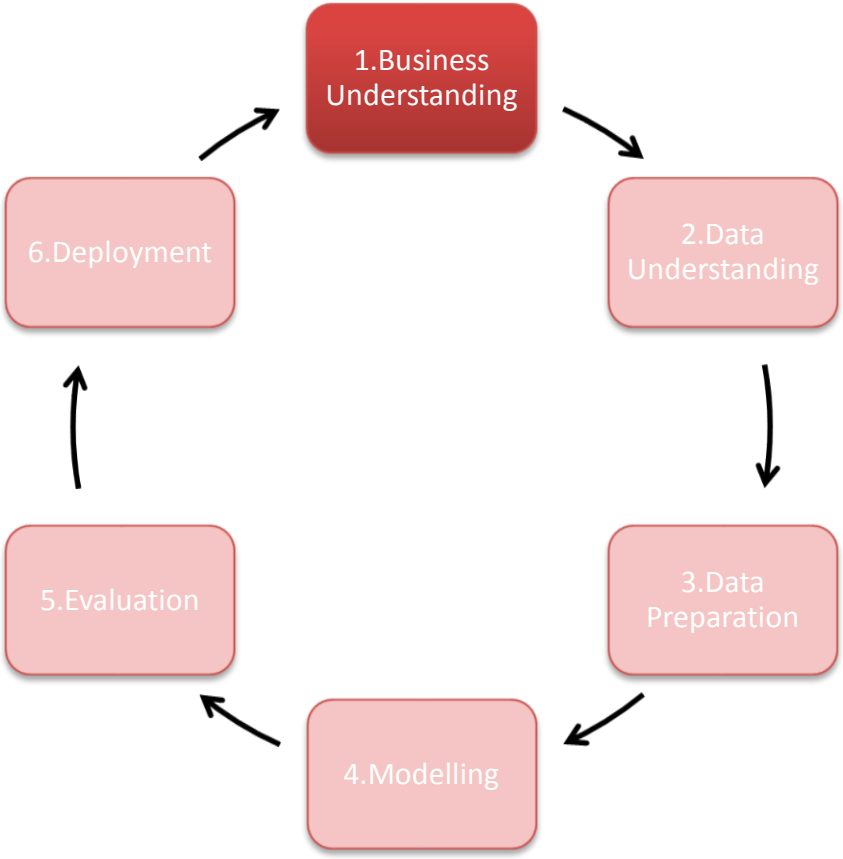



http://crisp-dm.eu/


1.Business Understanding
2.Data Understanding
3.Data Preparation
4.Modelling
5.Evaluation
6.Deployment


1
2
3. Data Preparation
4
5
6
Time

# The CRISP-DM process



# Example business objectives

- A water company wants to reduce pollution
- A Telco company wants to improve coverage
- An on-line gaming company want to identify fraudulent bets
- A charity wants to increase donor lifetime value
- A multi-channel subscription-based magazine want to improve renewal rate
- Local government planners want to know how likely a ward is to sustain next year
- A shipping company wants to identify containers that are likely to contain smuggled items
- A coffee retailer wants to understand what effect price changes will have on demand
- A hospital wants to know how many A&E staff to deploy on each shift
- An on-line retailer wants to increase their repurchase rates

# Example business objectives – more specifically

1. How do I reduce downtime?
2. How do I improve my SLA performance?
3. How do I reduce time to repair?
4. What is the effect of preventative maintenance?
5. What is the correlation between the fault diagnostic and its closure/outcome
6. What drives delays in fixing?
7. Which sites require the highest maintenance and why?
8. What equipment requires the highest maintenance (repeat corrective tickets) and why?
9. What causes additional/multiple work orders/repair tasks within a ticket?
10. Which replacement parts do I need to have in stock? And where?

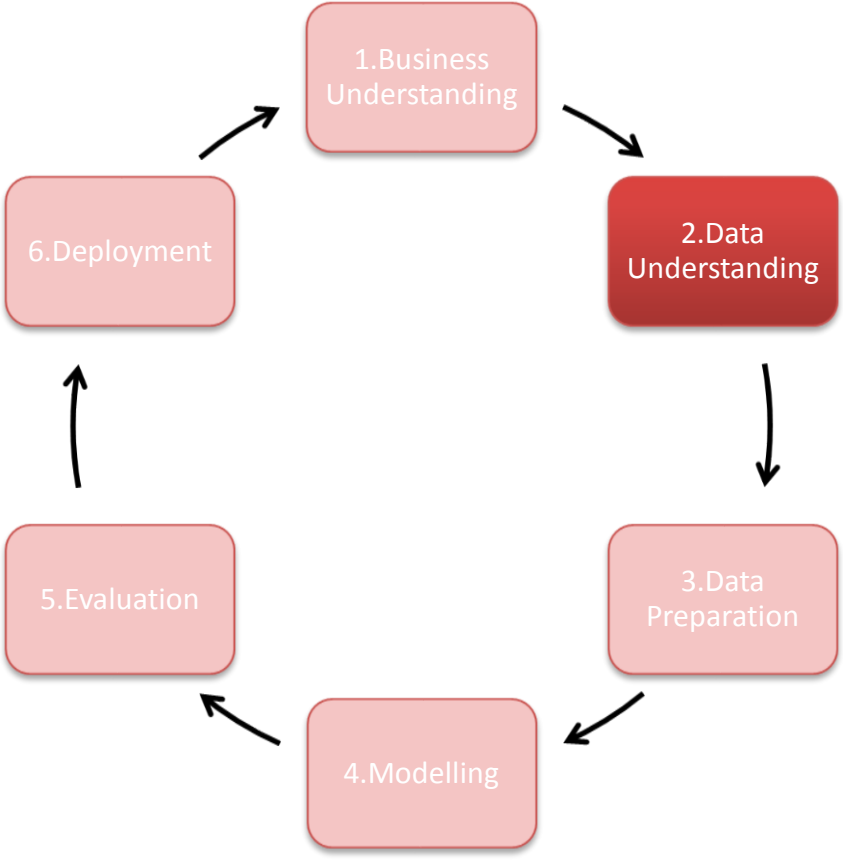
# Potential success criteria

- Reduce downtime by 10%
- Improve SLA performance to 99% for severity 1s
- Reduce average repair time by 1 hour
- Reduce overall repair costs by 20%

OR

- Develop a model that can accurately predict 4 time out of 5 when an asset will fail in the next 3 months

# The CRISP-DM process



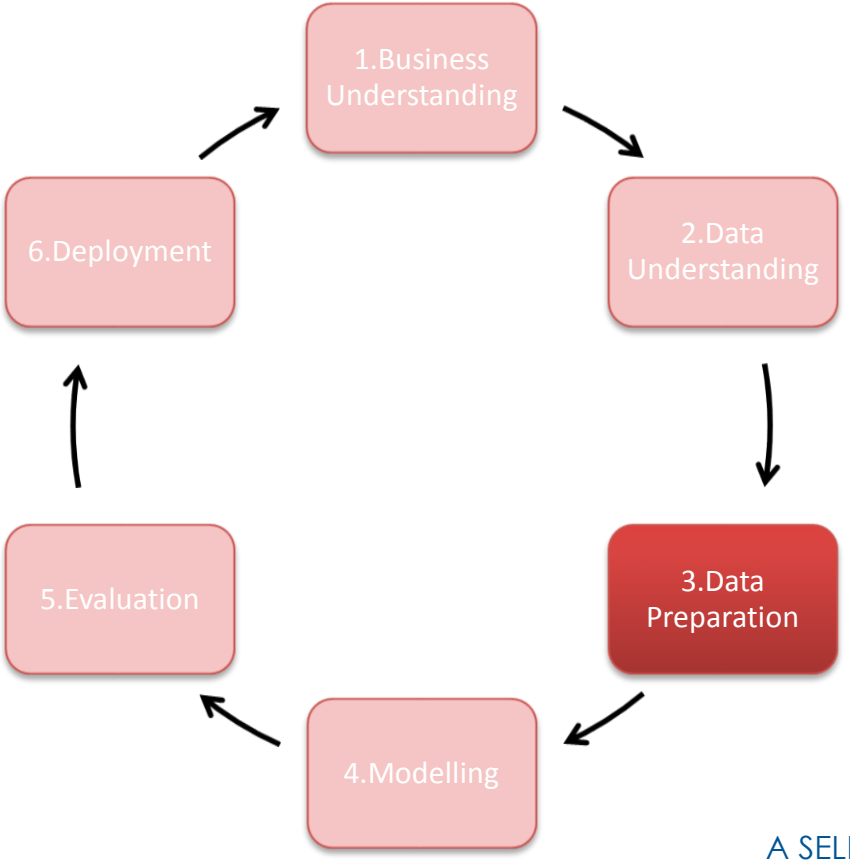
# 2. Data understanding – high level

- Identify the **data sources** and **fields** which may have a bearing on the business/analytical objectives
- Review data schemas and any other data documentation
- What looks relevant?
- What are the formats?
  - Databases, text files, excel, etc.
- What are the fieldnames?
  - Metadata
- Crucially … what is the likely **target** field that maps to the business objective e.g.
  - Repair time
  - Machinery failing
  - Assigning the right engineer
  - Identifying the right fix
  - Identifying the right parts

# 2. Data understanding – low level

- **Explore** the data
- Typically looking for **patterns** between fields
- Using uni- and bi-variate analyses
  - Examine fields one-by-one or in pairs
  - Often using visualisation tools
- Test hypotheses
  - E.g. High Pressure is the root cause of failure
  - Travel time is the most significant delay in the repair cycle
- Validate data
  - Identifies any issues involving anomalies
- Develops understanding and informs modelling

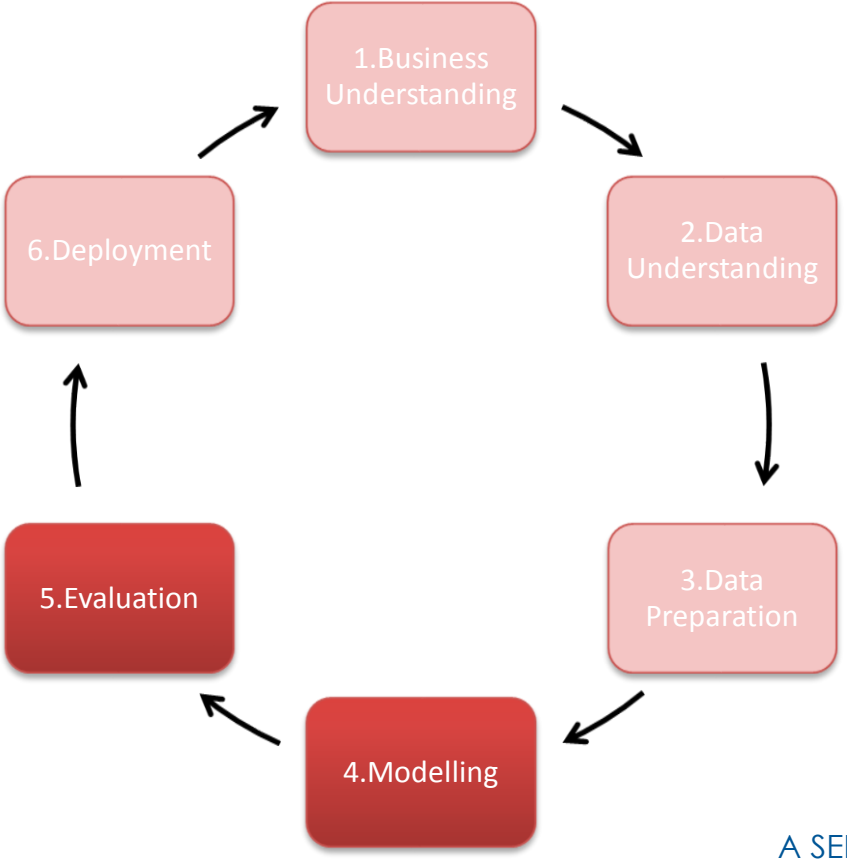
# The CRISP-DM process



# 3. Data preparation

- Data understanding effectively designs this step
- Together with data understanding this can be more time consuming than expected
  - Sometimes 80% of a project
  - Especially for newer projects
- Typically integrates data from different sources
  - **Often operational sources that haven't been analysed in this way before**
- Aggregates data
- Create composite measures
  - E.G. Band variables
  - Apply formulae e.G. Compute annualised figures and other ratios
- Comparable to ETL (extract transform load)

# The CRISP-DM process



# 4.Modelling

- Apply a variety of modelling techniques
- Candidate list identified during understanding phase
  - Driven by data types (see later)
  - Constrained by available tools
- 2 broad styles:

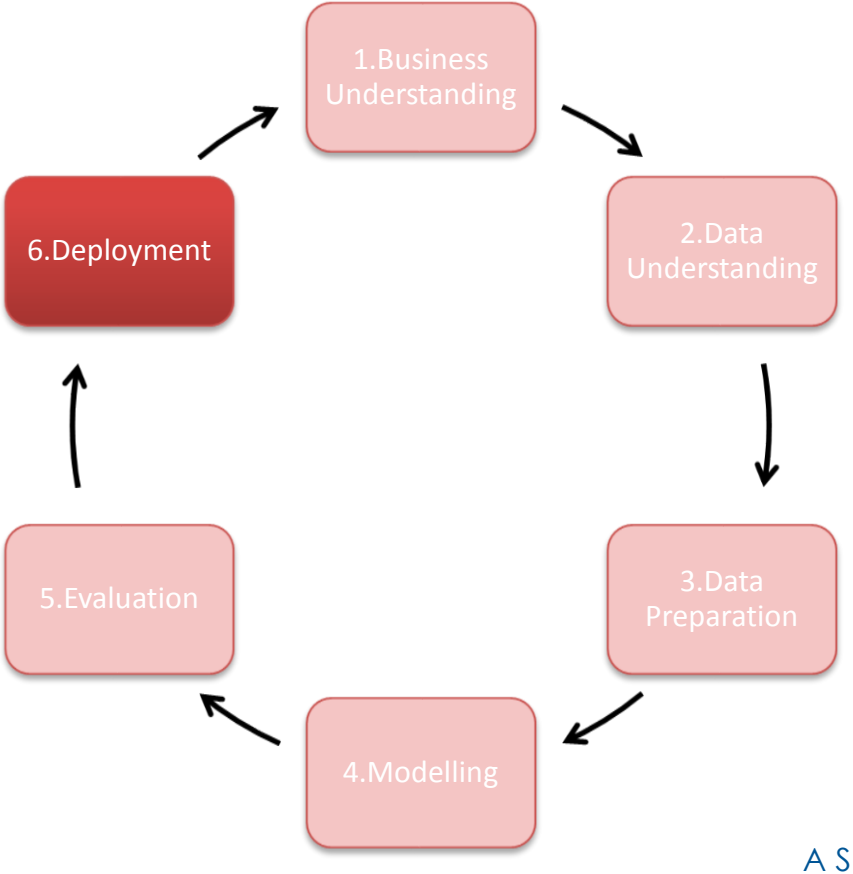
  a) **Hypothesis led**. Add the fields/predictors that we believe are driving the outcome

  b) **Data led**. Add more fields at the beginning and incrementally reduce (and/or let the algorithms do that)
- The best performing modelling algorithm is a function of the specific data/problem

# 5.Evaluation

- Essential that the models are tested against unseen data
- Typically the data is partitioned into 2 (or 3) sets at random  e.g. 70%:30%
  1. Training (modelling) set
  2. Test (holdout) set
  3. Evaluation set
- Evaluate against the success criteria agreed in the understanding phase
- Often it is about how well the model performs against a given value criteria e.g. revenue
  - Defined in d**ata understanding** phase

# The CRISP-DM process



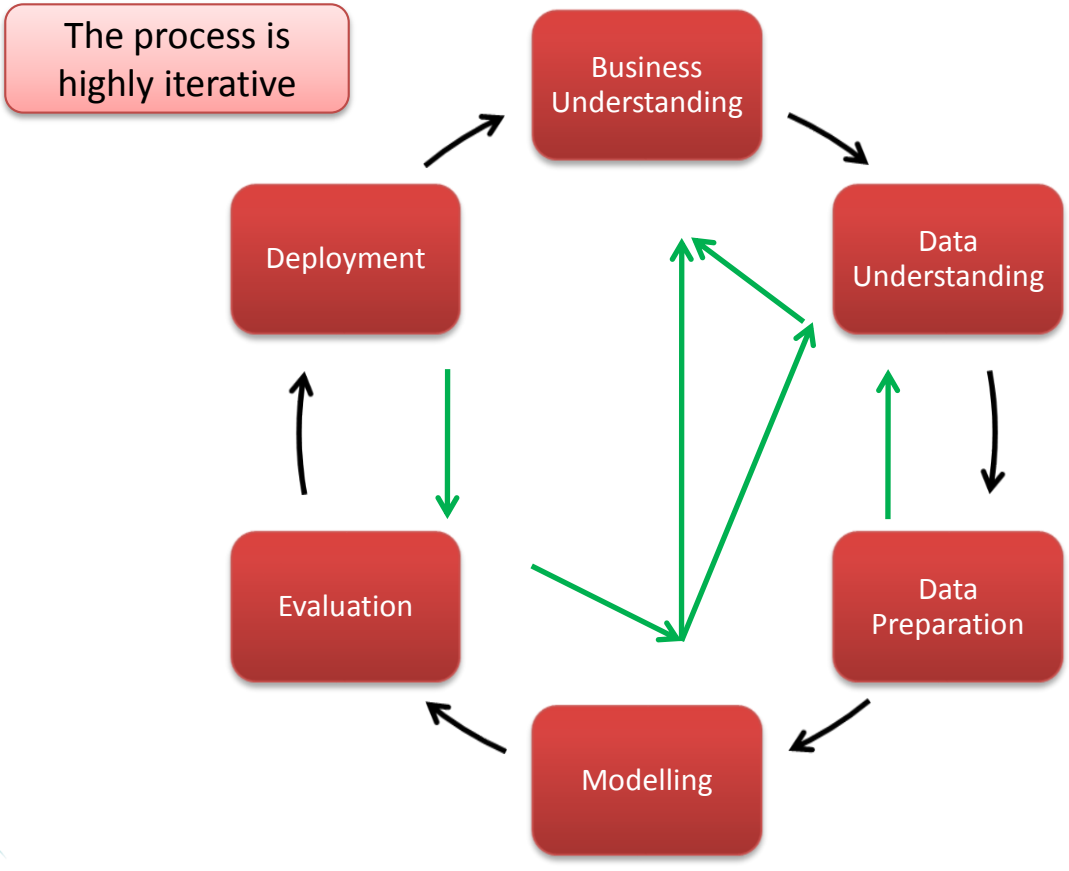
# 6.Deployment (1)

- Could be as simple as a list of names and predictions/scores
  - E.g. a prioritised fix list
- Could be as complex as a model encapsulated as a computer program and embedded in an operational system to predict in real time and automate decisions
  - E.g. a model embedded in a system which sends alerts and triggers
- Could be embedded in a **What-if?** simulator
- Important to distinguish between a model in the modelling and deployment phases
- Typically…
  - In the modelling phase many different models and modelling options are built and evaluated
  - In the deployment phase the winning model(s) are fixed
    - E.g. we deploy a decision tree with a fixed shape

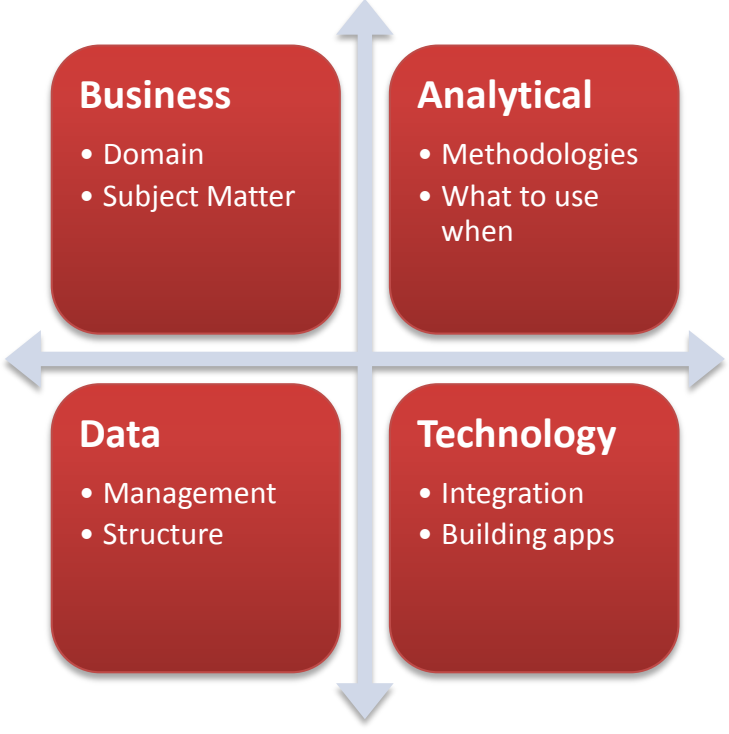
# 6.Deployment (2) (monitoring)

- If we did our job properly then the deployed model should correspond to what we saw in evaluation
  - Other factors may intervene
- Ongoing evaluation ("monitoring") still needs to happen if models are to be used over time
  - Some models have a longer shelf life than others
- More recently there has been some development of models which adapt/correct themselves to changing circumstances
  - Some level of re-modelling to improve accuracy
    - "Self adapting"
  - More commonly this is achieved through the concept of **champion/challenger** modelling or **model refresh** approaches

# The CRISP-DM process



# People and roles

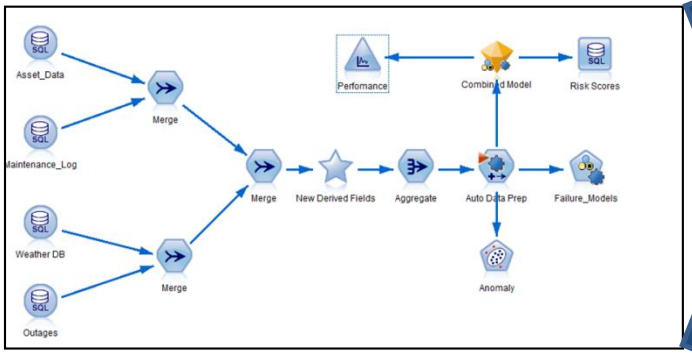


# Executing a predictive project - summary

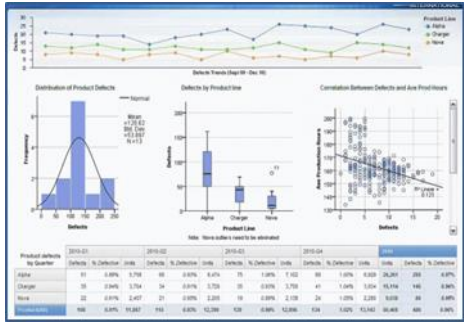
- A predictive analytics project can be more like a **research & development** project
  - Can we build a successful model?
  - Has anyone done this before?
  - What is the **risk** that we cannot achieve the objectives?
- Hence projects can fail
- It isn't just about the analyst
  - Larger projects usually need a larger (multi-disciplined) team
- IBM/SPSS Modeler visually maps functions and data flows to the CRISP process

# What are the common ingredients of successful applications?

## Incorporating results in both operational and reporting platforms




Predictive insight developed through analysis & modelling



Key Performance Predictors as part of enterprise MI – ensure awareness

Scores / predictions / likelihoods out to operational systems, rules & workflow - to change behaviours

SMARTVISION Europe

A SELECT INTERNATIONAL COMPANY

# What are the common ingredients of successful applications?

## Integrating the resultant insight with existing systems

Cognos. software

infosphere™

WebSphere

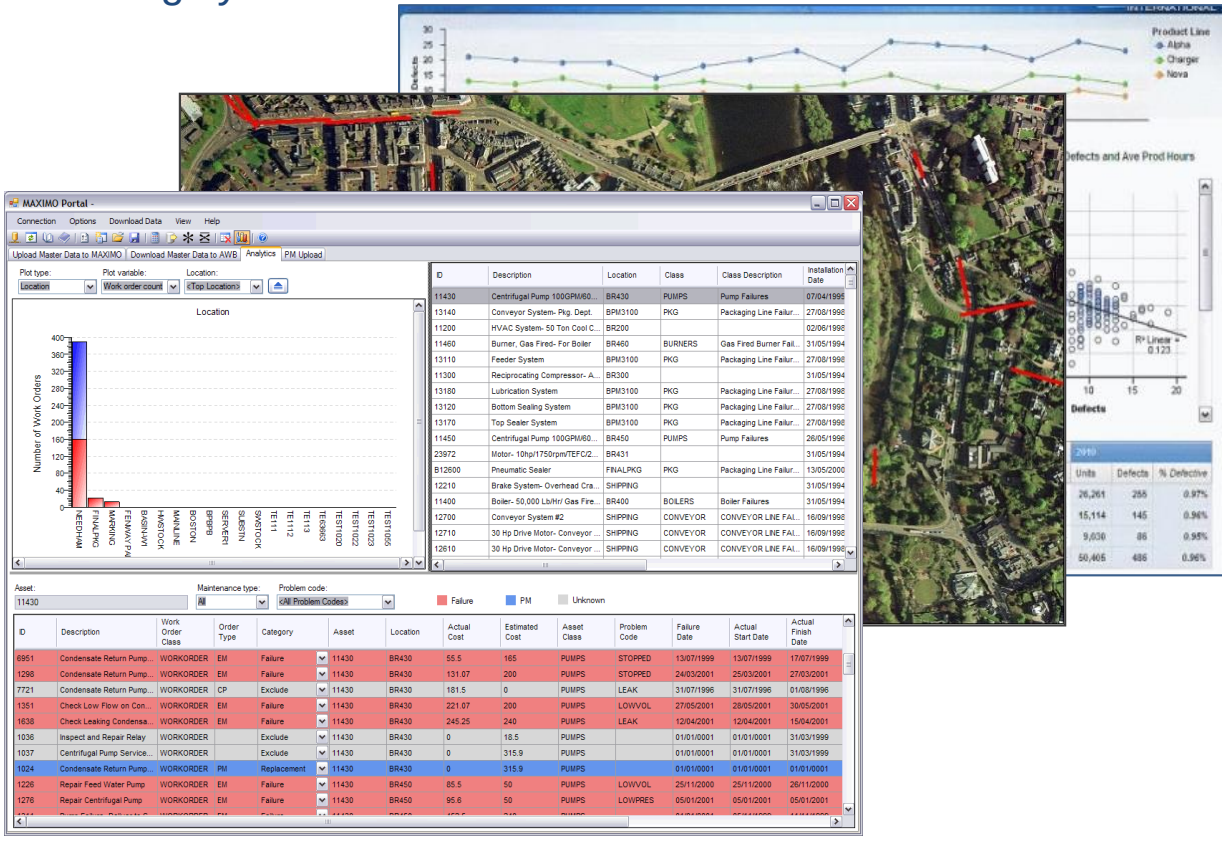
SPSS

NETEZZA

ILOG
Changing the rules of business®

maximo

COREMETRICS




SMARTVISION Europe

A SELECT INTERNATIONAL COMPANY

# Summary, next steps & close

# Common misunderstandings

- Revolutionary results overnight!
- You'll need a Ph.D.
  - In fact , data–literate, business focussed people learn how to do this all the time.
- The more *accurate* the model the better
- You need a clean, single-customer-view warehouse

# Advice to get started

- **Build internal credibility:** *Think* about where you would get *biggest impact* for the *least effort*.
  - How can we “prove it” quickly and efficiently?
- Consider adopting a proven methodology e.g. CRISP-DM
- Don’t get hung up on modelling techniques - focus on *Business Understanding* and *Deployment*
- Consider the full data landscape
- Consider the sorts of roles involved /impacted
- Consider integration with other business insight systems (e.g. MI/BI)
- How will you know its worked? Focus on measuring the benefit – e.g. response rate lift, increased cross-sell, revenue/profit impact

# Options to get started…

| | | |
|---|---|---|
| Operational Analytics (Starter Pack) | Single client software tools, set up, training & first project deployed | £18K |
| Operational Analytics (Enterprise Automation Pack) | 2 x client plus server software tools, automation plus set up, training & first projects | £80K |
| Operational Analytics (Predictive Maintenance & Quality) | Full enterprise deployment including predictive analytics, client server, full automation, database technology etc. | POA |

SMARTVISION Europe

A SELECT INTERNATIONAL COMPANY

# Working with Smart Vision Europe Ltd

- As a premier partner we sell the IBM SPSS suite of software to you directly
  - We're agile, responsive and generally easier to deal with
- As experts in SPSS / analytics / predictive analytics we will
  - Deliver classroom training courses
  - Offer side by side training support
  - Offer "skills transfer" consulting
  - Run booster and refresher sessions to get more from your SPSS licences
  - Give no strings attached advice
- We are a support providing partner so if you already have SPSS you can source your technical support directly from us (identical costs to IBM)
  - We offer telephone support with real people as well as web tickets / email queries
  - We offer "how to" support to help you get moving  on your project quickly

SMARTVISION Europe

Contact us:

+44 (0)207 786 3568
info@sv-europe.com
Twitter: @sveurope
Follow us on LinkedIN
Sign up for our Newsletter

# Thank you