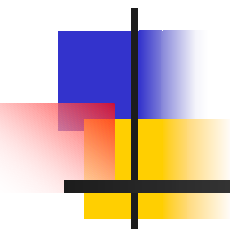


Knowledge Discovery and Data Mining Applied to Engineering Applications



Shirley Williams
The University of Reading

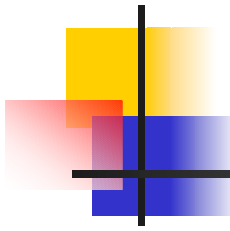


Overview

- The process
 - Understanding the problem
 - Collecting and preparing the data
 - Exploring the data -> Results
 - Modelling -> Knowledge

Iterative

Example using telecomm data



Knowledge Discovery with Engineering Data

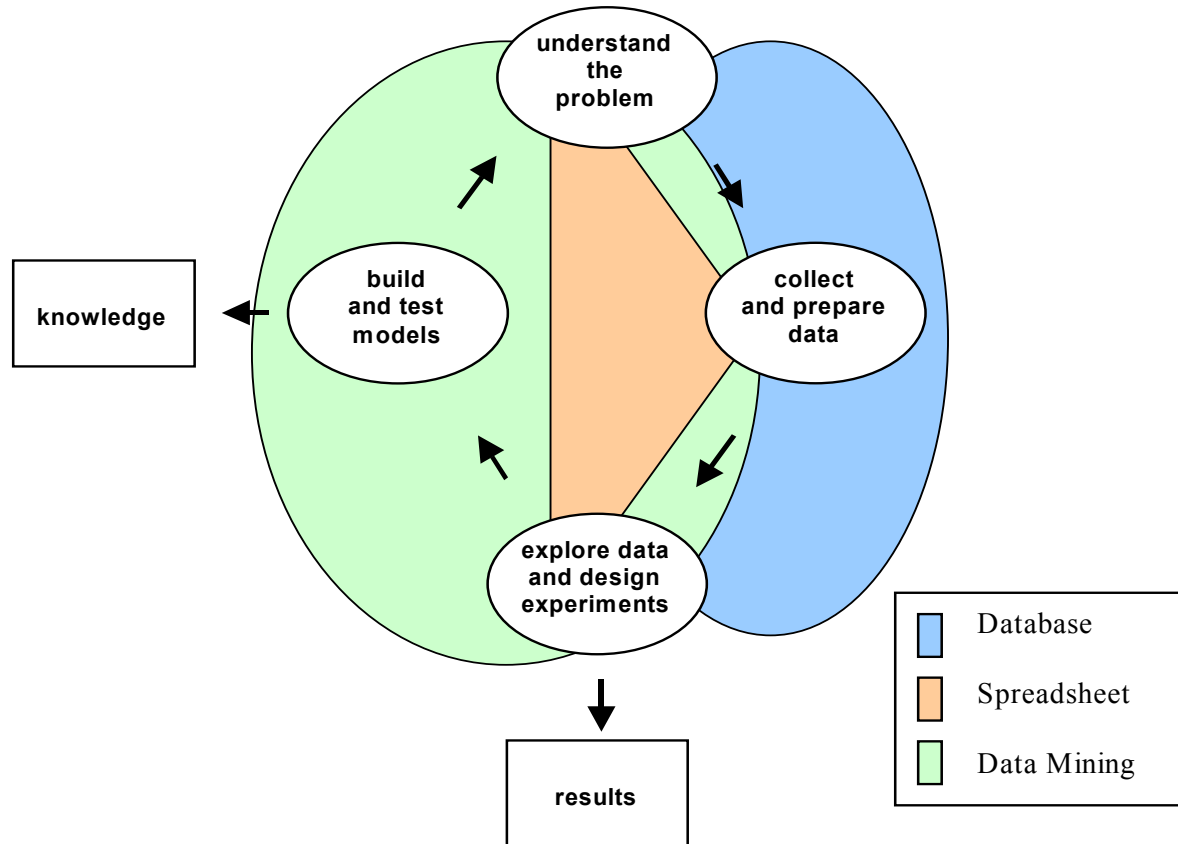
What we want:

- Find anomalous trends
- To get answers to questions related to engineering issues

As opposed to...

Marketing analysis and fraud detection
the traditional uses in telecommunications

An Investigation Oriented KD Method



Understanding the Problem

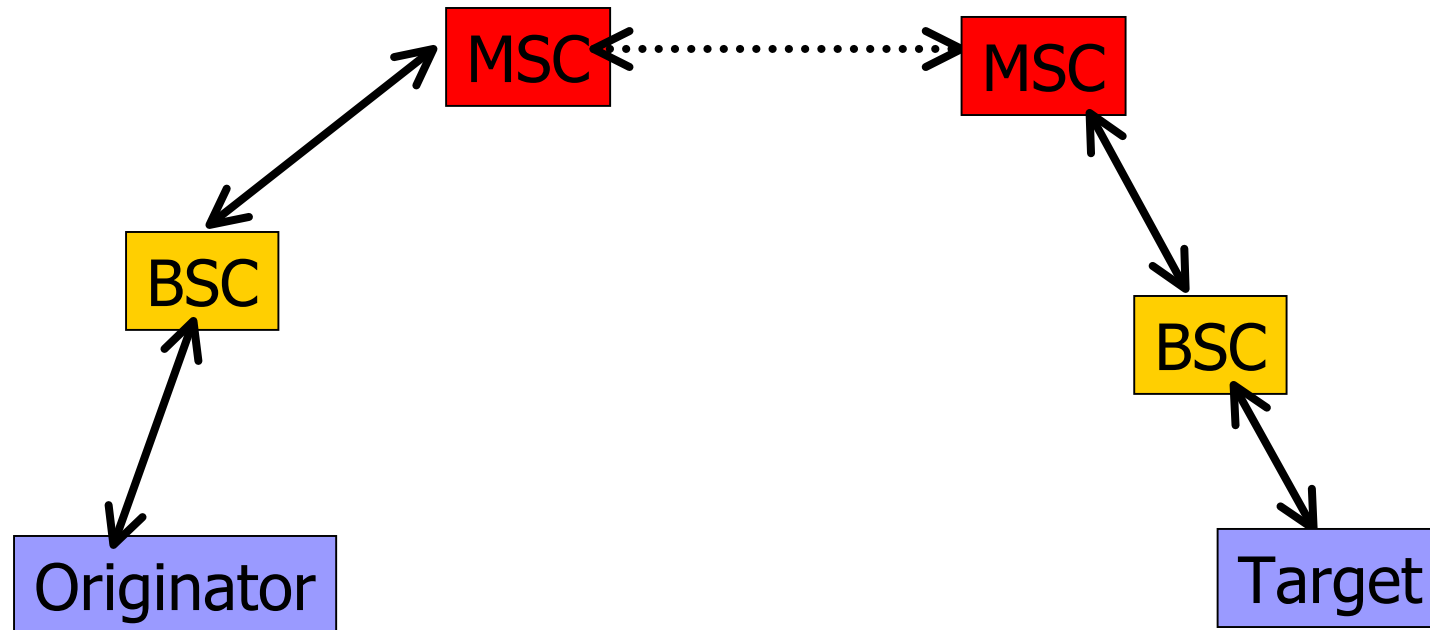
EOS (End of Selection)

- EOS codes are raised to indicate a call related event
- Providing immediate information:
 - Call progress
 - Why a call has failed
 - Dropped, Set up problems or Other

"Access is easy but analysis techniques are complex with many EOS to focus on and many mobile switches to analyse."

Understanding the Problem

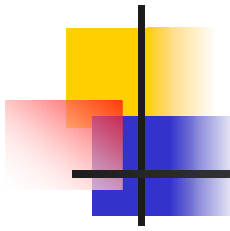
EOS – Mobile to Mobile



Understanding the Problem

Potential Gains

- Nearly 2 million EOS events associated with problems at call set up
 - Several may be associated with a single unsuccessful call
 - Quarter million unhappy customers?
 - Helping their calls succeed may stop them churning



Collecting and Preparing the Data

First Steps

- Accessing the data
- Accessing the right databases
- Understanding the fields
- Getting the engineers to classify the EOS codes
- Creating summative data



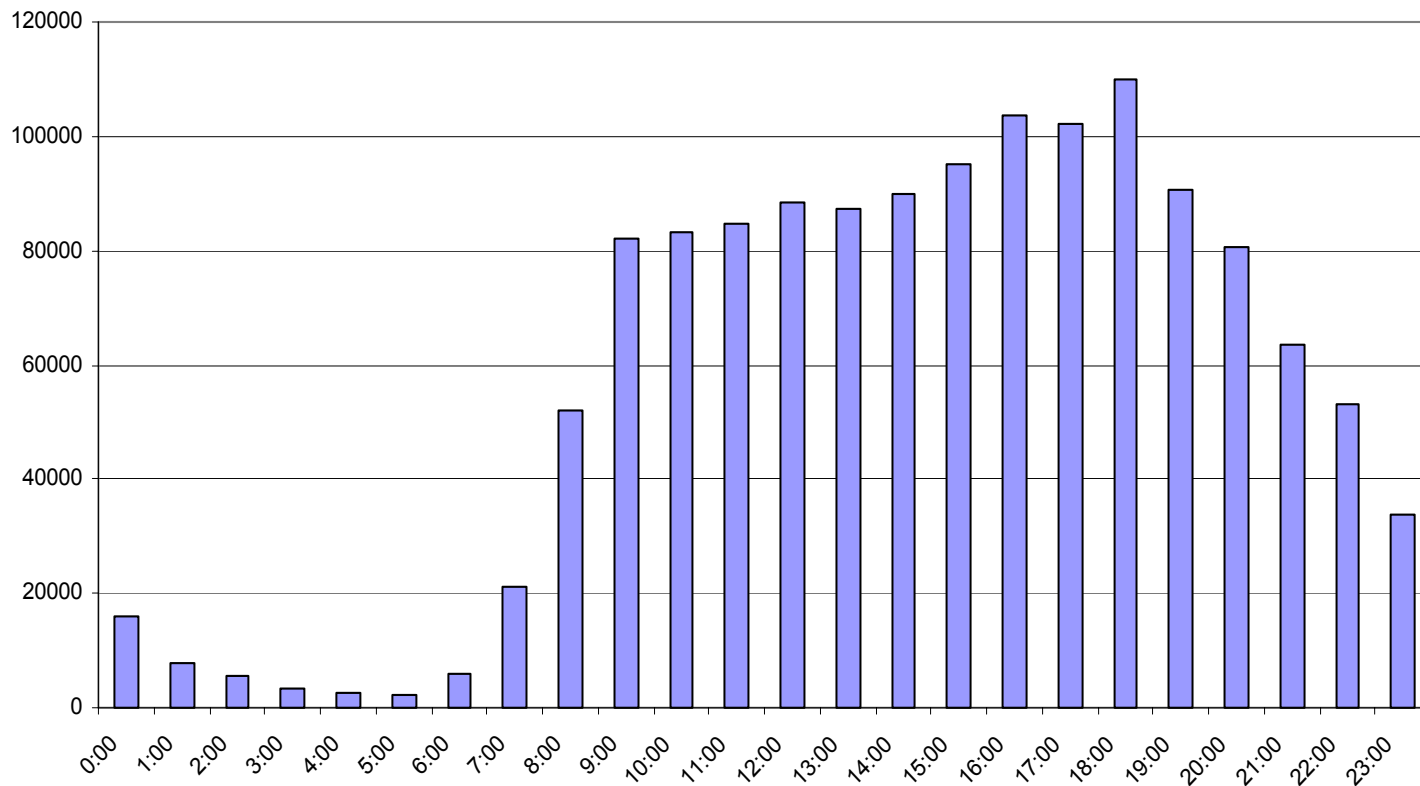
Collecting and Preparing the Data

Magnitude of the Problem

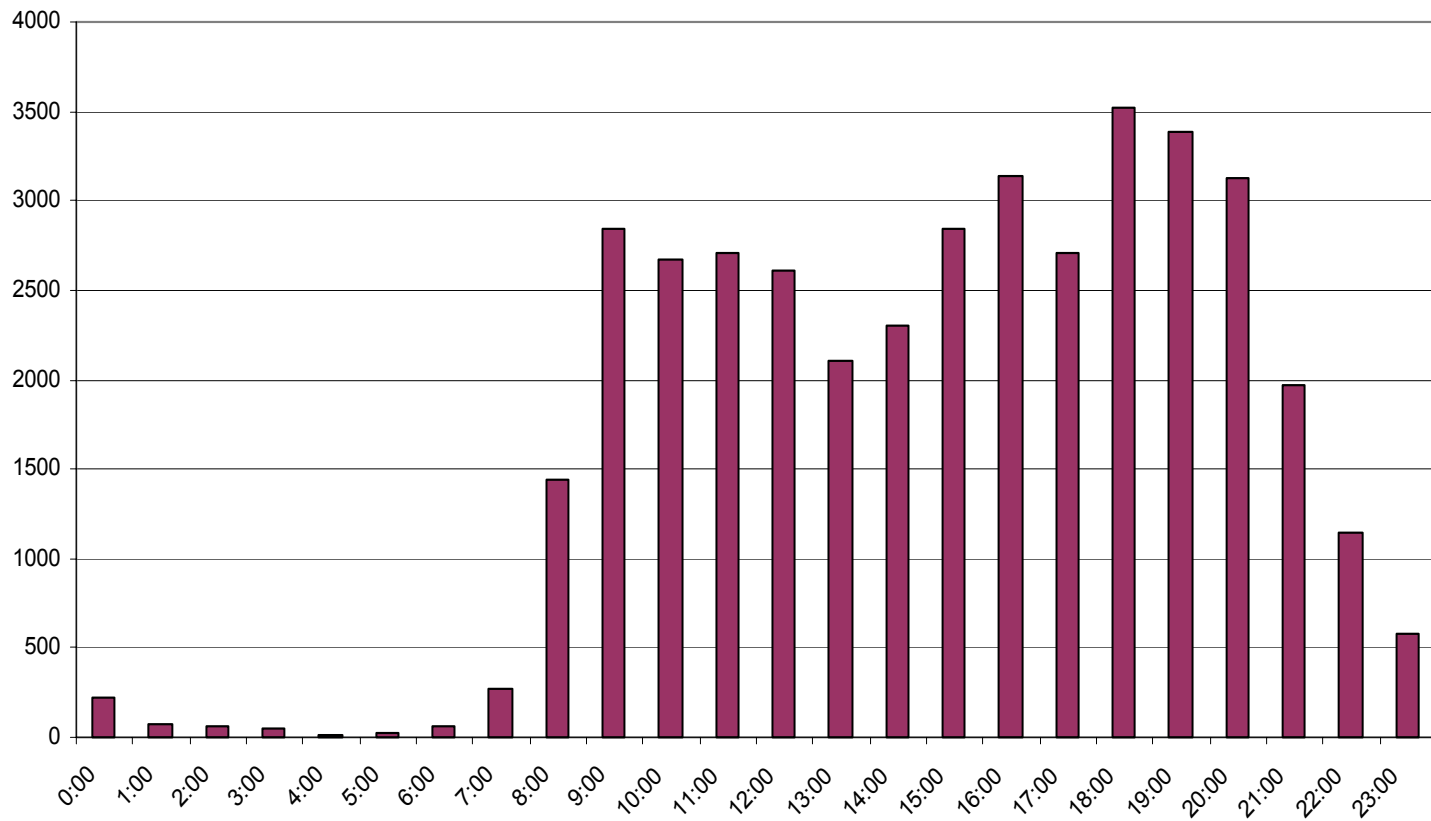
- 100 million plus calls per day
- Over 10 million EOS events
 - Many of these are *subscriber busy*
 - We can't fix that
 - But marketing may be able to sell them *call waiting*

Exploring

Calls at Bath on a Thursday

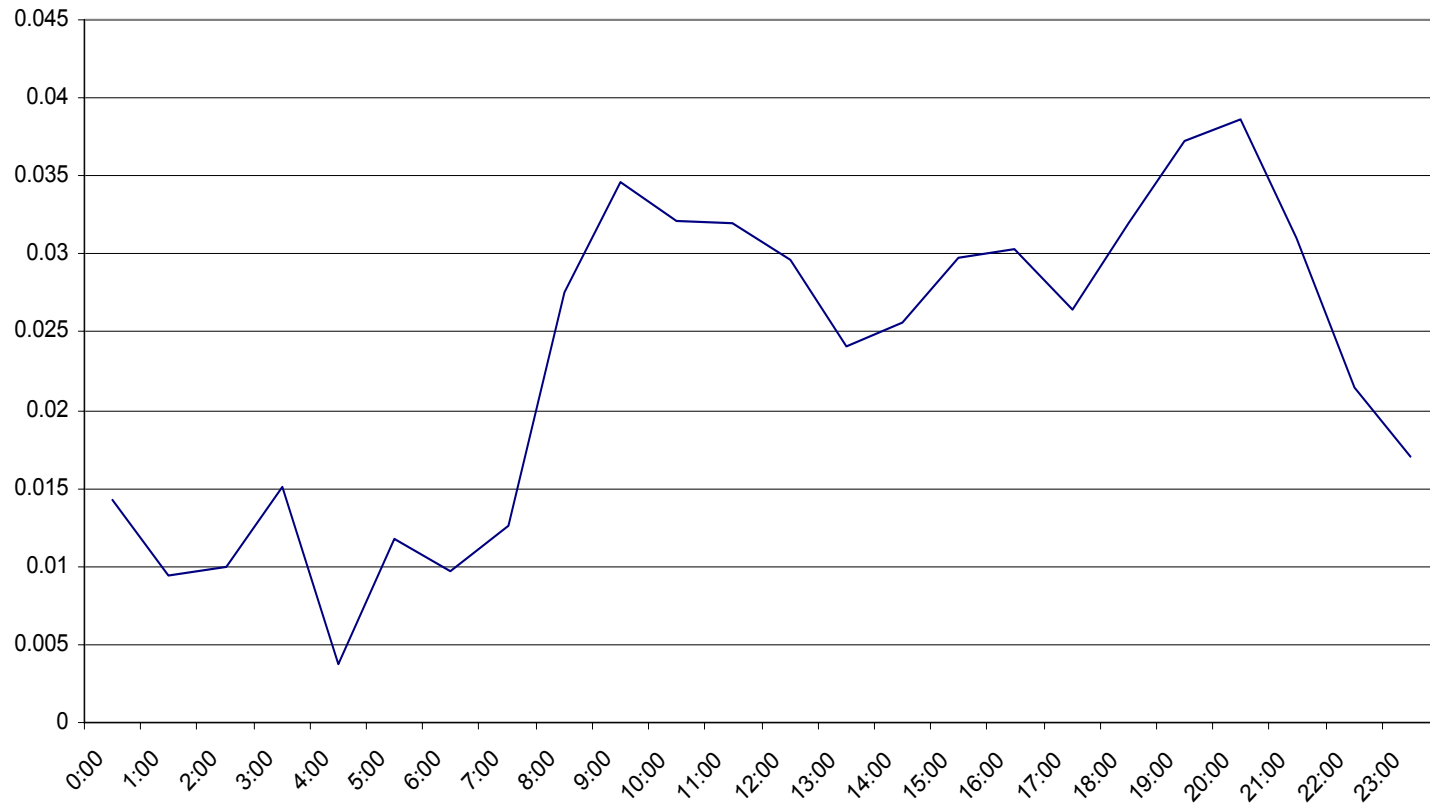


Exploring EOS 33

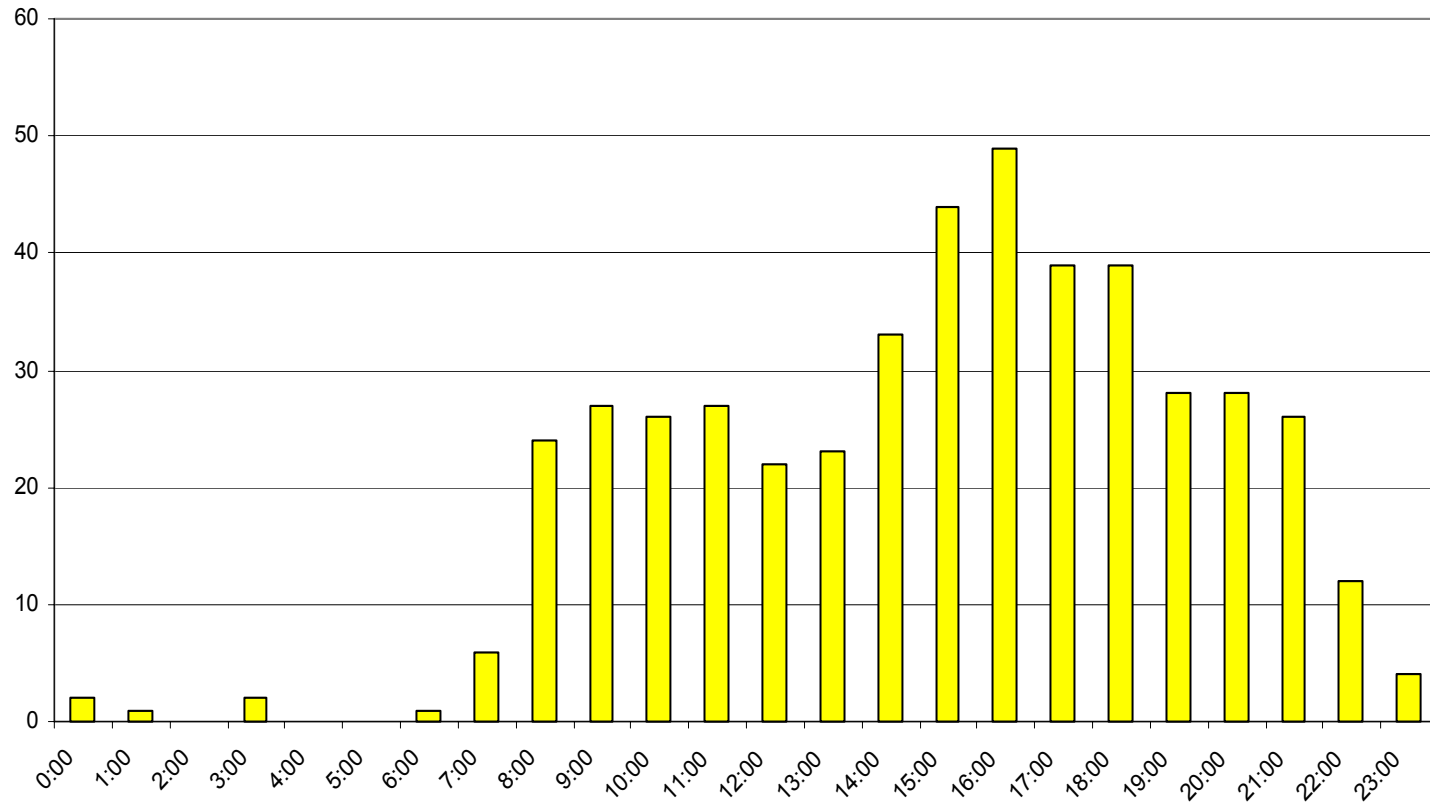


Exploring

Ratio EOS 33 to Calls

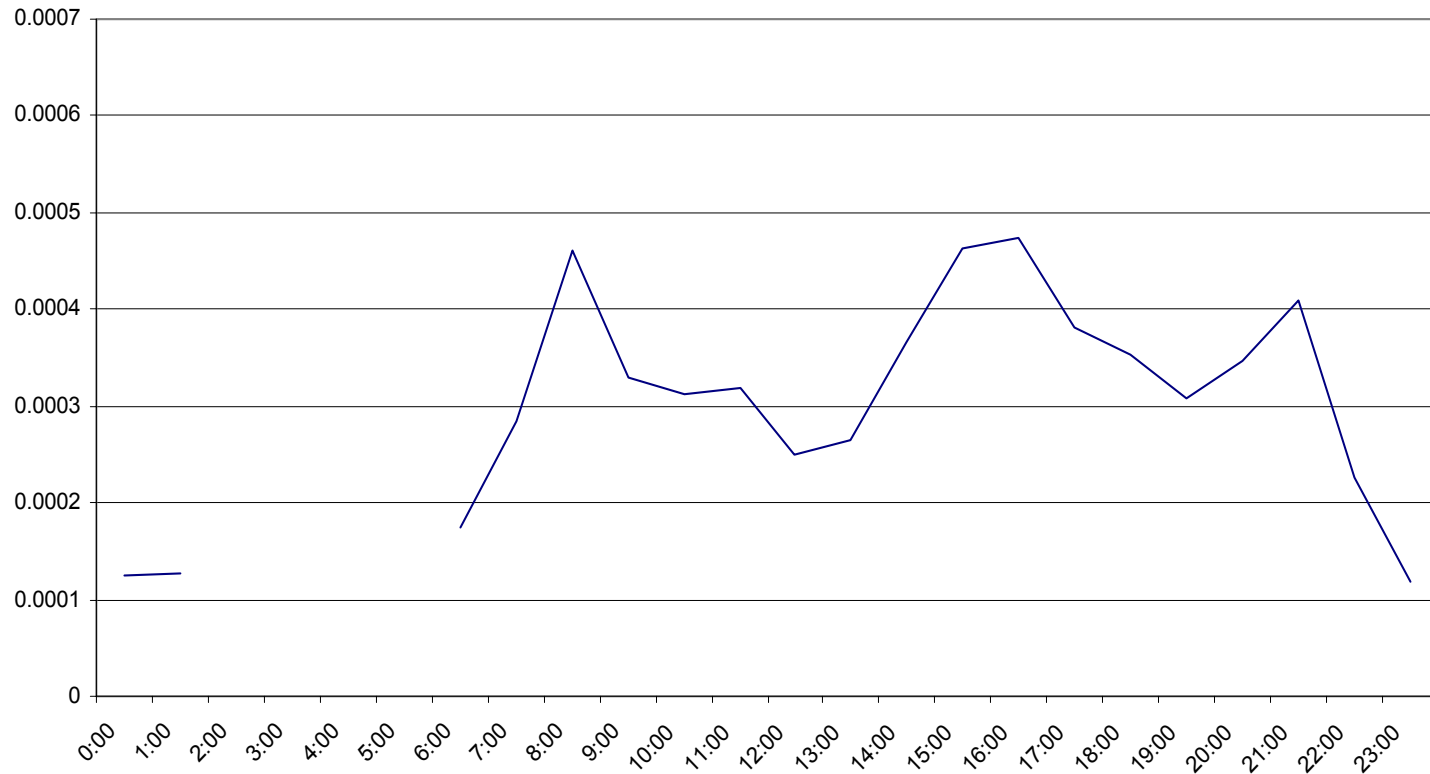


Exploring EOS 160



Exploring

Ratio EOS 160 to Calls

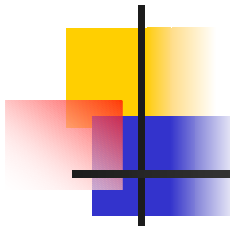




Exploring

Results

- There are different peaks for calls and some EOS codes
- Ratios of EOS to Calls vary from place to place and code to code

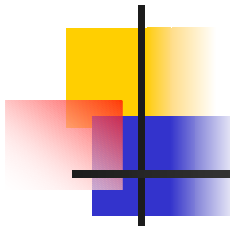


Exploring

Experiment Design

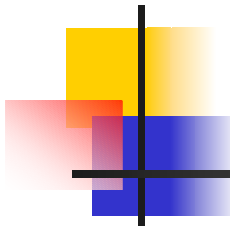
Investigate what leads to set up problems

- Do high values of EOS codes match high values of Calls?
- Do some places have more set up problems than others?
- Are some times better than others?
- Are some days better than others?



Discovery

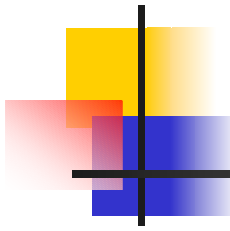
- The Health of a switch can be represented numerically as:
 - The ratio of occurrence of certain EOS codes to the number of calls
 - The ratio of combinations of EOS codes to the number of calls
 - Other quantities



Modelling

Running Experiments

- Data Mining then used with targets of good health to identify particularly poor and good
 - Places
 - Dates
 - Times



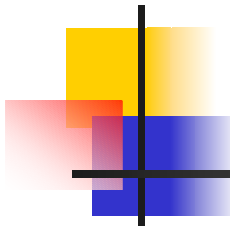
Modelling

Approaches

Many approaches were tried using SAS Enterprise Miner TM, using different parameters, including:

- Regression
- Neural Networks
- Decision Trees

It was a great help having a statistician on the team



Modelling

Results

The Engineers preferred the results they could easily understand

- Decision Trees

But similar indications where found using other techniques



Where Next?

- The Engineers can then carry out a detailed analysis of why certain, dates, times and places are *unhealthy*
- Further iterations of the process need to be undertaken to answer new questions



Conclusions

- Knowledge Discovery is applicable to engineering data
- Lots of results can be found using simple tools and techniques
- Deeper knowledge can be found with advanced techniques
- The knowledge needs to be acted upon