



# Data Mining Can Be Fun...

Michael Esposito



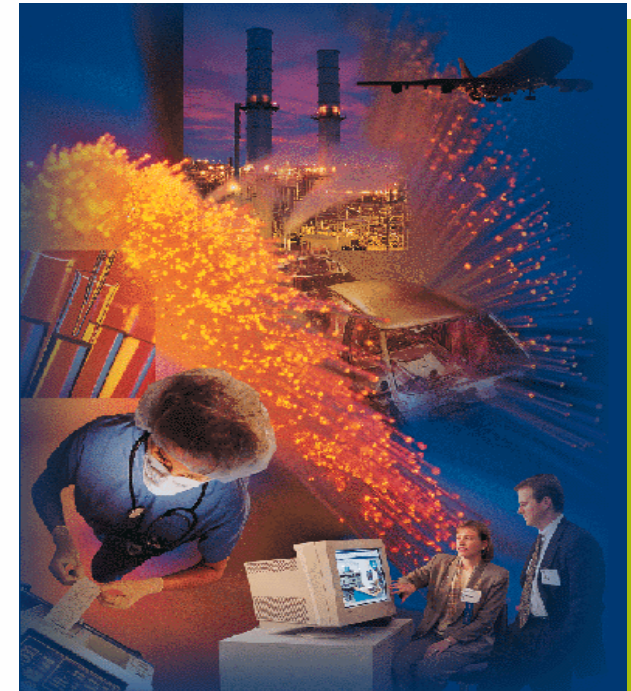
YEAH, RIGHT!!

Michael Esposito



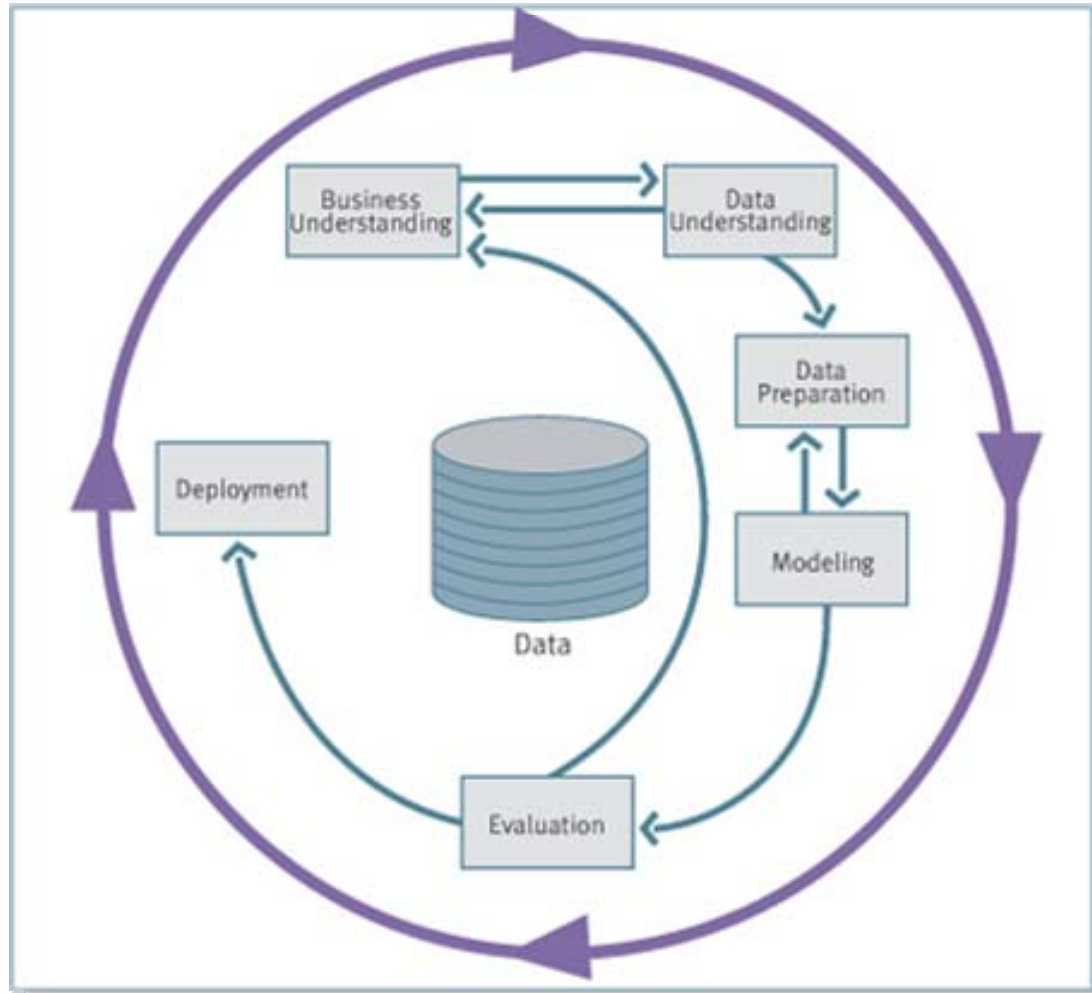
## What IS data mining??

“The nontrivial extraction of implicit, previously unknown, and potentially useful information from data. It involves statistical and visualisation techniques to discover and present knowledge in a form that may be easily comprehended.”





# Data Mining Methodology (CRISP – DM)

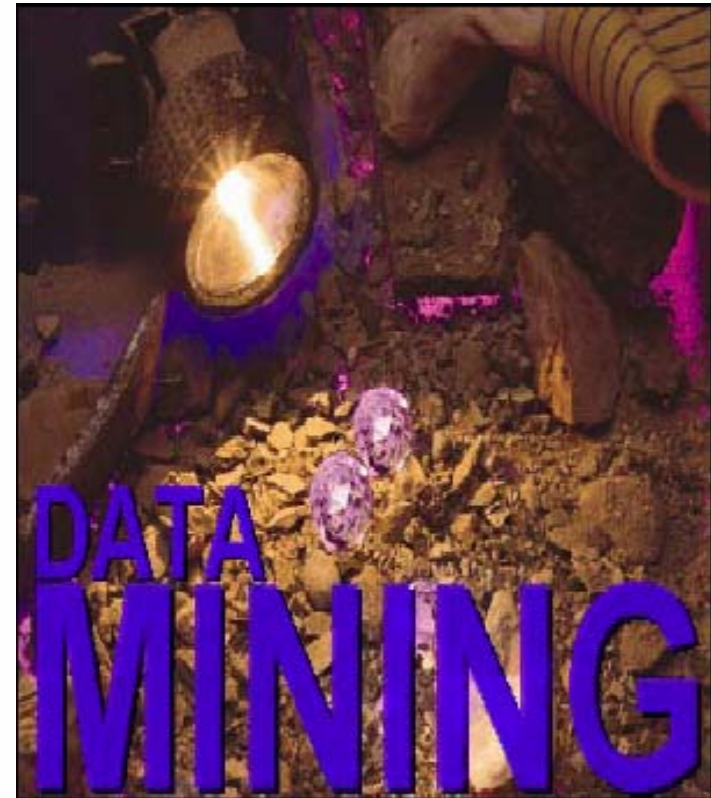


<http://www.crisp-dm.org/>



## What IS data mining?? (SAS version)

***"The process of selecting, exploring and modeling large amounts of data to uncover previously unknown actionable information for a business benefit."***





# Data Mining Methodology (SAS version)

## Sample



Input Data Source



Sampling



Data Partition

## Explore



Distribution Explorer



Multiplot



Insight



Association



Variable Selection



Link Analysis

## Modify



Data Set Attributes



Transform Variables



Filter Outliers



Replacement



Clustering



SOM/Kohonen

## Model



Regression



Tree



Neural Network



Princomp/  
Dmneural



User Defined Model



Ensemble



Memory-Based Reasoning



Two Stage Model

## Assess



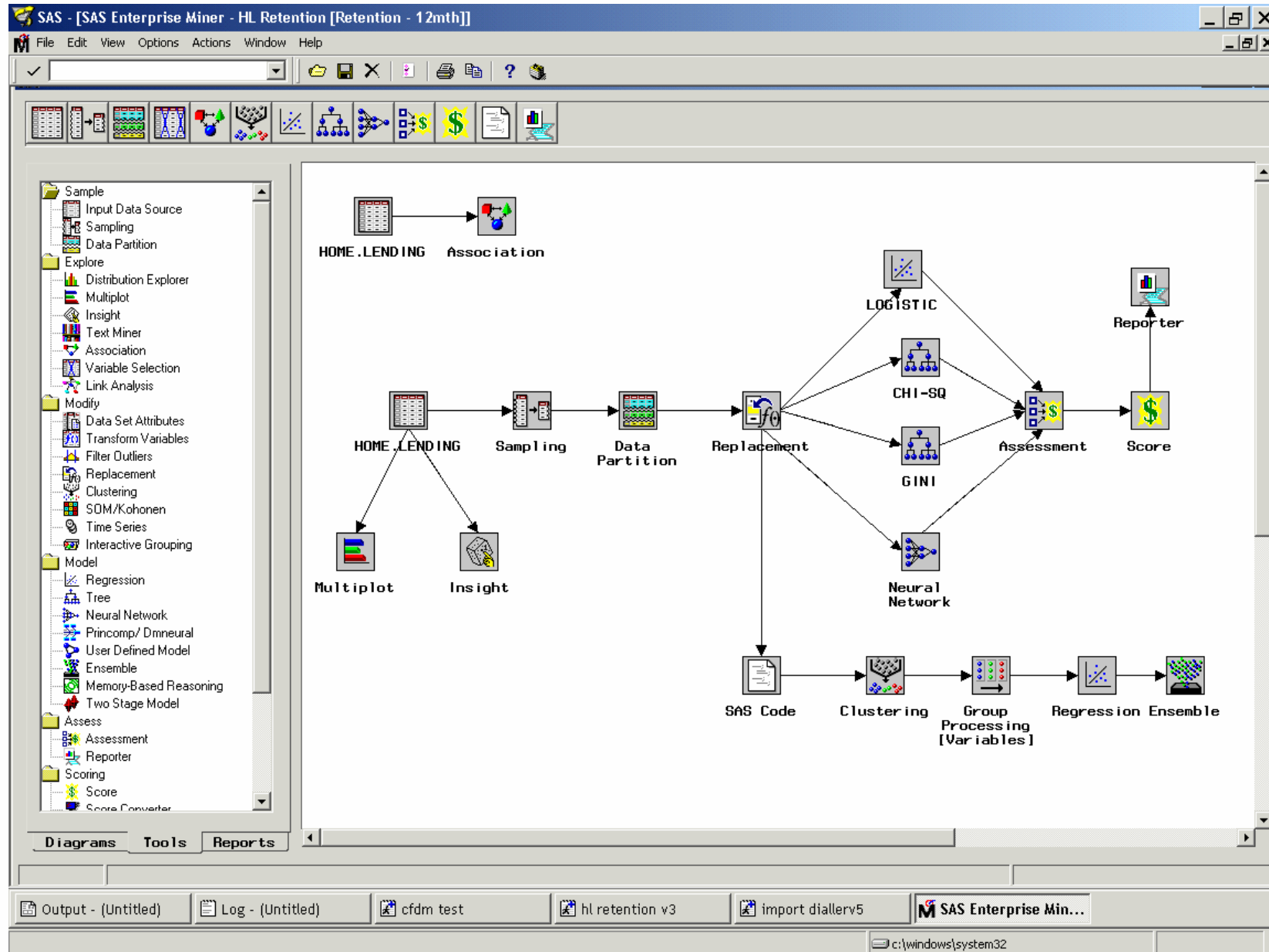
Assessment



Reporter



# Typical Data Mining Project





# Interesting Applications of Data Mining

**Netflix Prize**

Home Rules Leaderboard Register Update Submit Download

NETFLIX

Browse Recommendations Friends Queue Buy DVDs

Home Genres New Releases Previews Netflix Top 100 Crit

### Welcome!

The Netflix Prize seeks to substantially improve the accuracy of predictions about how much someone is going to love a movie based on their movie preferences. Improve it enough and you win one (or more) Prizes. Winning the Netflix Prize improves our ability to connect people to the movies they love.

Read the [Rules](#) to see what is required to win the Prizes. If you are interested in joining the quest, you should [register a team](#).

You should also read the [frequently-asked questions](#) about the Prize. And check out how various teams are doing on the [Leaderboard](#).

Good luck and thanks for helping!

**Movies For You**

Randy, the following movies were chosen based on your interest in:  
[Bowling for Columbine](#)  
[Carnivale: Season 1](#)  
[Fahrenheit 9/11](#)

**You really liked it...**

Now owned for just \$5.99

**The Big One**

★★★★☆

Original art

Lewis Black: Re and Score

★★★★☆

Not Interested

Not Interested

Member Favorites  
Easter Eggs  
By Decade  
By Studio  
Movies You've Seen

Give a friend

<http://www.netflixprize.com/>



# Interesting Applications of Data Mining

## Challenge:

New Zealand Cricket identified the need to help team coaches and the captain determine their match play strategy in real time; to make decisions – at any point during the innings – most likely to produce a win.



## Solution:

New Zealand Cricket now has a single data mart with models that are easy for coaches and others to use for real-time match data analysis and decision making.

## Benefits:

While the SAS solution provides New Zealand Cricket with a competitive edge in matches, it is also helping to develop players' skills.



<http://www.sas.com/offices/asiapacific/sp/successes/NZCricket.html>



# Demonstration Of Enterprise Miner



# Want to learn more?



## Customer Value Management Analyst

With over 17 000 talented employees across Australia and New Zealand, Suncorp prides itself on helping people build and protect their dreams.

A unique new role for a CVM Analyst is now available within the Customer Analytics and Database Marketing department. Working in a team of three this role supports the development of customer value management programs based on analysis of customer behavioural change and defined business requirements.

Key responsibilities of this role include:

- identifying actionable business approaches for marketing initiatives by analysing and interpreting customer/product and transactional data
- ongoing measurement of CVM/Direct Marketing programs success and progress
- analysing and reporting on Customer behavioural changes affecting the Group's customer base.

To be successful in this role you will have a strong statistical background with appropriate tertiary qualifications in a quantitative discipline and ideally Marketing. Working closely with business units you will establish key relationships and ideally have experience with SAS Enterprise Miner.



# Questions ???

