

**THE
POWER
TO KNOW[®]**

Data Mining

Robert Smith, Pre-Sales Consultant

Ad in London Newspapers, 1900

Men wanted for hazardous journey. Small wages, bitter cold, long months of complete darkness, constant danger, safe return doubtful. Honour and recognition in case of success.

—Ernest Shackleton, *Antarctic Explorer*

Did it work?

Too many definitions of data mining

- ..."the nontrivial extraction of implicit, previously unknown, and potentially useful information from data"... http://en.wikipedia.org/wiki/Data_mining
- is the process of analyzing data from different perspectives and summarizing it into useful information.....<http://www.anderson.ucla.edu/faculty/jason.frand/teacher/technologies/palace/datamining.htm>
- ...A class of [database applications](#) that look for hidden patterns in a group of [data](#) that can be used to predict future behavior...
http://www.webopedia.com/TERM/D/data_mining.html
- The process of analyzing data to identify patterns or relationships.
https://iomega-eu-en.custhelp.com/cgi-bin/iomega_eu_en.cfg/php/enduser/std_adp.php
- The ability to query very large databases in order to satisfy a hypothesis ("top down" data mining); or to interrogate a database in order to generate new hypotheses based on rigorous statistical correlations ("bottom-up" data mining).
www.inproteomics.com/nwglosde.html
- Data mining is derogatory. It means sorting through a huge volume of data, extracting decision rules that seem to favor one team over another, but without regard to whether or not there is any cause-and-effect relationship. Data mining is the sports-betting equivalent of sitting a huge number of monkeys down at keyboards, and then reporting on the monkeys who happened to type actual words.
www.basketballfreesportspicks.com/glossary.shtml
- Searching large volumes of data looking for patterns that accurately predict behavior in customers and prospects.
www.adobe.com/products/vdp/glossary.html
- As the term suggests, data mining is the analysis of data to establish relationships and identify patterns.
practice.findlaw.com/glossary.html
- The analysis of data for relationships that have not previously been discovered.
www.creotec.com/index.php
- Nontrivial extraction of implicit, previously unknown and potentially useful information from data, or the search for relationships and global patterns that exist in databases. Bob Klevecz "The Whole EST Catalog" Scientist 12 (2): 22 Jan 18 1999 more... Algorithms & data analysis glossary
www.genomicglossaries.com/content/chemoinformatics_gloss.asp
- Data mining (or data discovery) is the process of autonomously extracting useful information or knowledge ("actionable assets") from large data stores or sets. Data mining can be performed on a variety of data stores, including the World Wide Web, relational databases, transactional databases, internal legacy systems, pdf documents, and data warehouses. Most of XSB, Inc's technologies involve data mining or data discovery.
www.xsb.com/glossary.html
- The process of using statistical techniques to discover subtle relationships between data items, and the construction of predictive models based on them. The process is not the same as just using an OLAP tool to find exceptional items. Generally, data mining is a very different and more specialist application than OLAP, and uses different tools from different vendors. Normally the users are different, too. OLAP vendors have had little success with their data mining efforts.
www.olapreport.com/glossary.htm
- Extraction of useful information from data sets. Data mining serves to find information that is hidden within the available data.
www.pcai.com/web/glossary/pcai_d_f_glossary.html
- The comparison and study of large databases in order to discover new data relationships. Mining a clinical database may produce new insights on outcomes, alternate treatments or effects of treatment on different races and genders.
www.thebusinesssite.com/IT%20Terms/Health%20Terms.htm
- The ability for users of a system to interrogate a database ad hoc.
www.oshpd.cahwnet.gov/MIRCa/glossary/glossaryad.htm
- The unguided (or minimally guided) application of a collection of mathematical procedures to a company's data warehouse in an effort to find "nuggets" in the form of statistical relationships
www.loyaltybuilders.com/resources/glossary.htm

Definitions

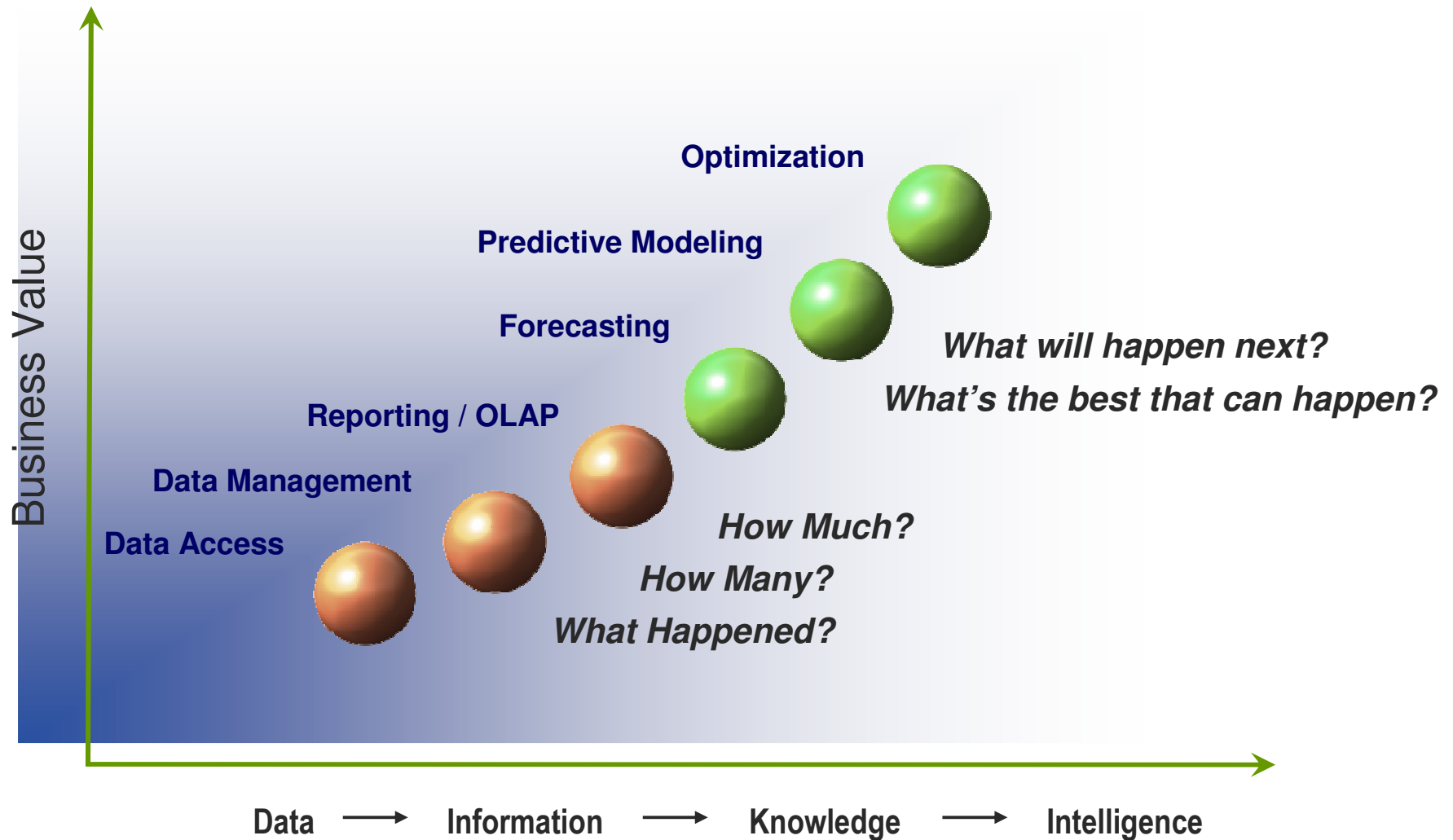
- Data mining is the process of autonomously extracting useful information or knowledge from large data stores or sets.

www.xsb.com/glossary.html

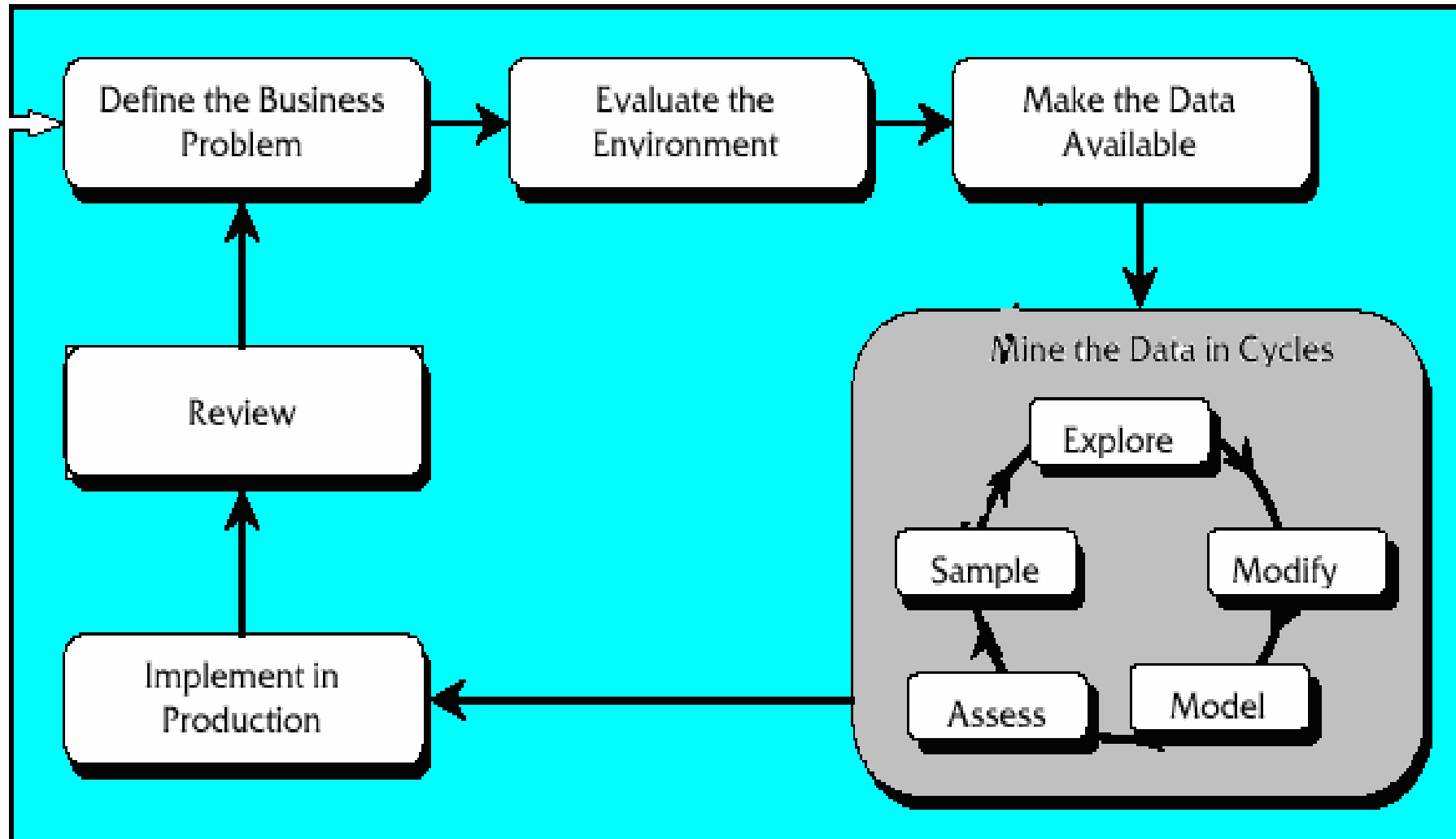
1980's brought us

- **The fall of the Berlin wall**
- **The IBM PC**
- **Lotus 1-2-3**
- **C++**
- **Immon**
- **Codd**
- **The ability to analyse volumes of data**

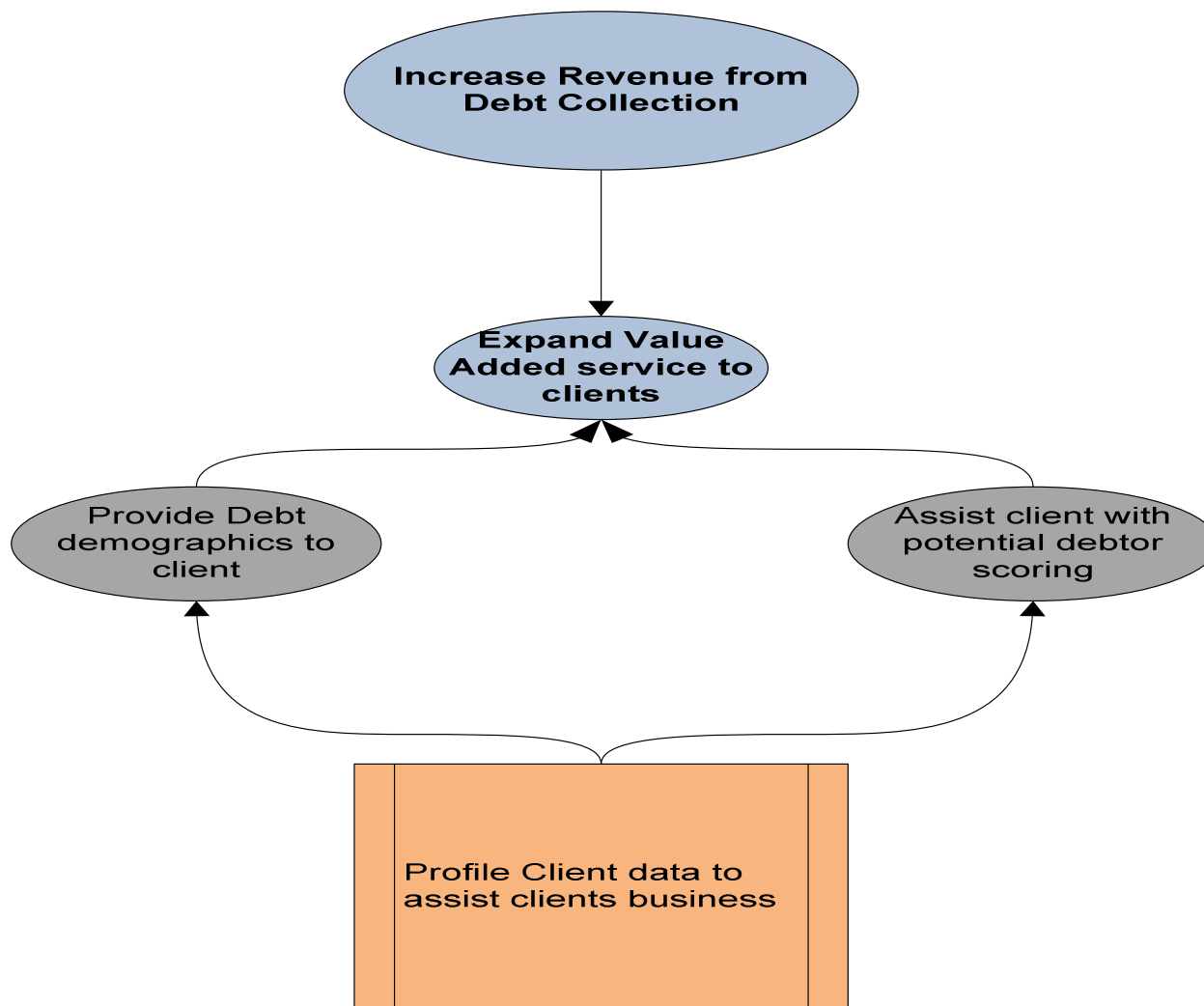
Beyond Traditional BI to Business Analytics



SAS Data Mining Projects Methodology



Supporting a strategy



Enterprise Miner Framework

Sample



Explore



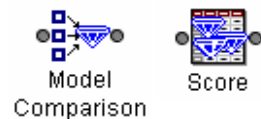
Modify



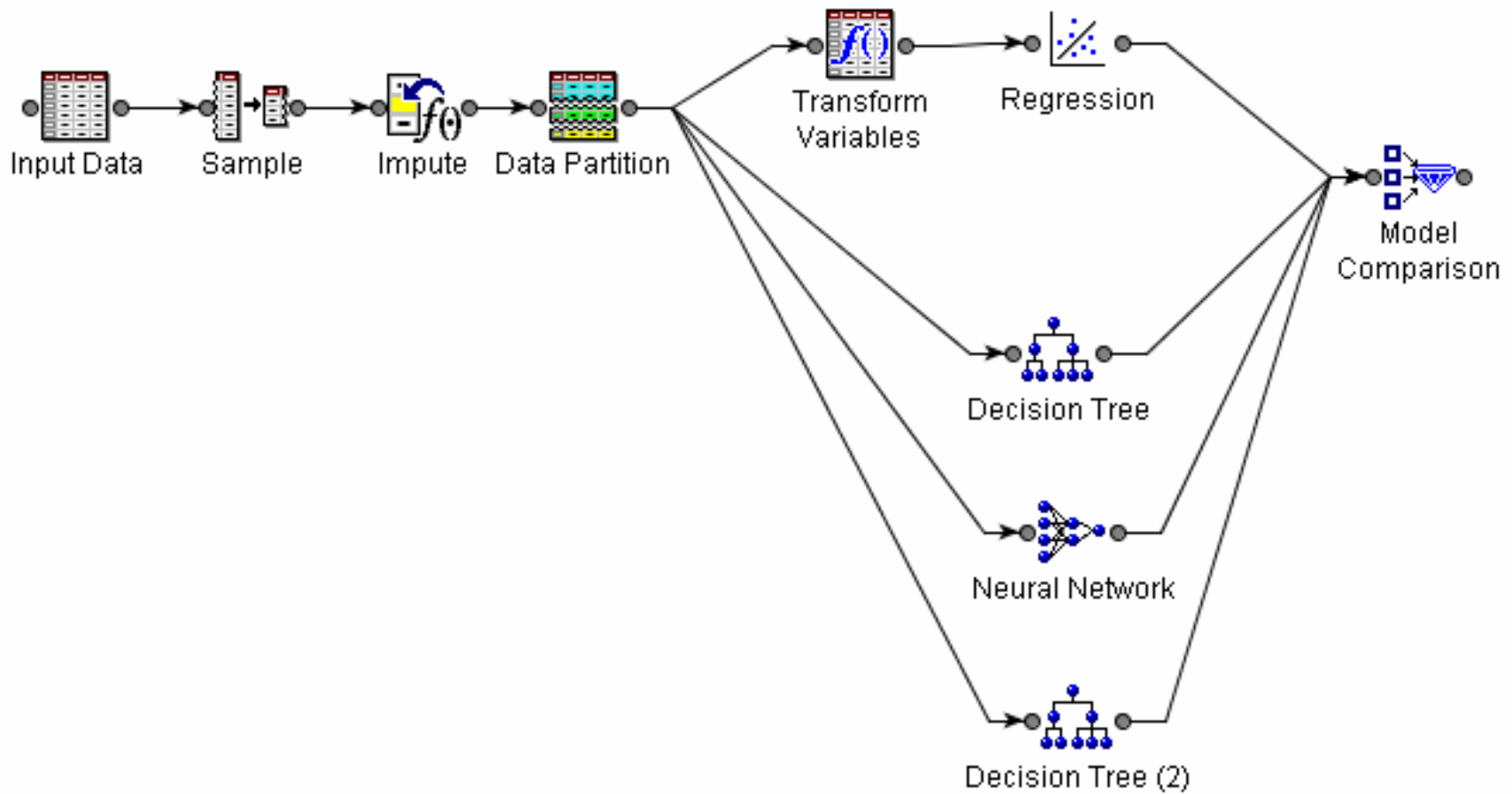
Model



Assess



Create Models



Demo time ...

The ease of use of EM

The use of one of the visualisation charts - ROC

The use of segmentation charts

Two models

Regression techniques – identifying who will purchase goods

Segmentation Model – vehicle engines and attributes



THE
POWER
TO KNOW®