

# How To Stop Good Customers Leaving (using SAS)



PROPHET ANALYTICS  
*-know your customers*

Gordon Wallace  
Prophet Analytics

3<sup>rd</sup> April 2008



# The Bank's Credit Card Dilemma

Low Intro Rate  
Rate Starts  
Rewards  
Increased Fees

Attrition results in  
\$20M loss per year

Save millions by retaining high  
value, 'at risk' customers



# The Requirement

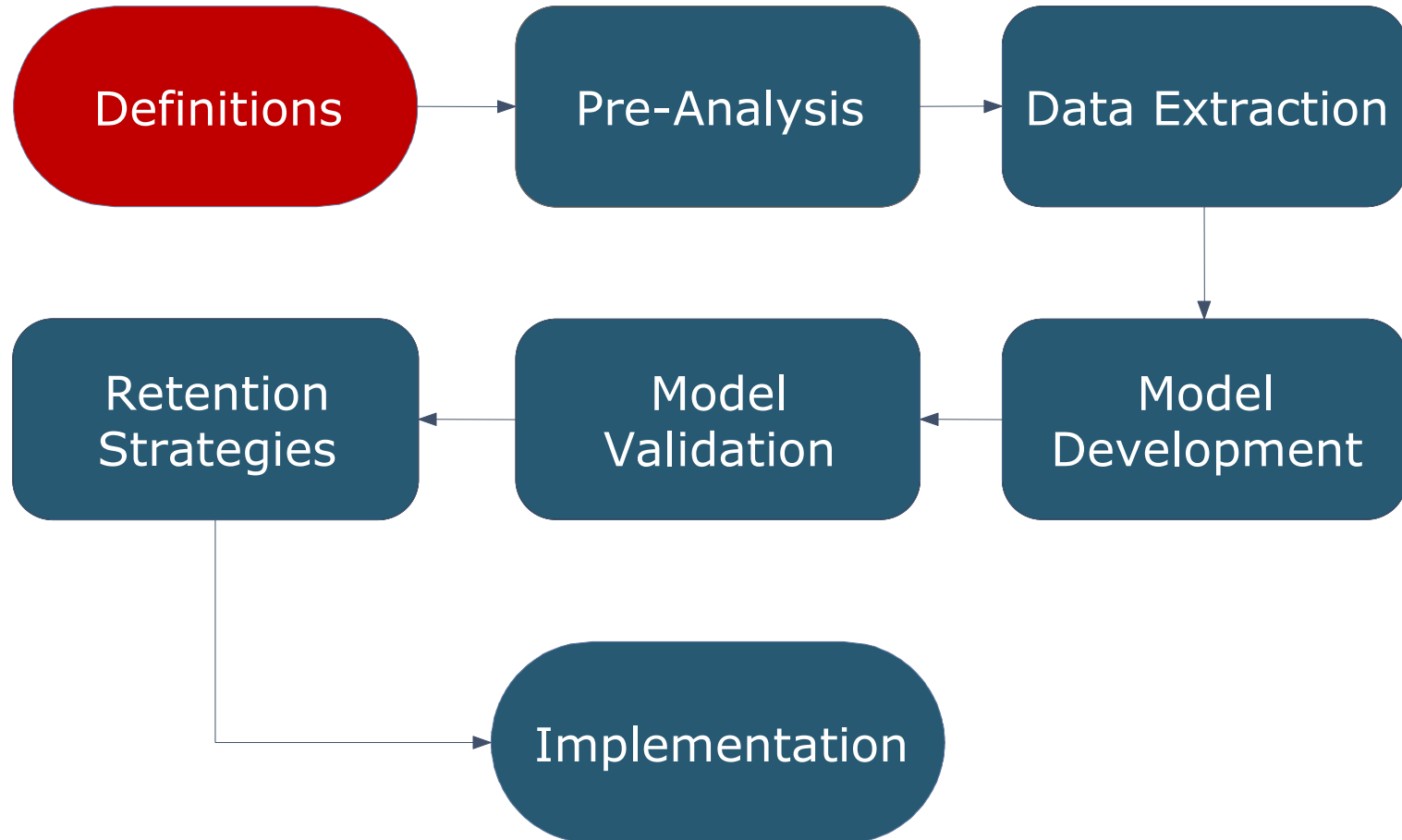
To proactively retain good customers and therefore reduce loss of revenue and marketshare.

To achieve this we need:

- A Retention Model measuring the risk of attrition
- A Segmentation identifying ‘good’ customers
- Sufficient warning to be able to take action
- An ongoing Retention process using latest data



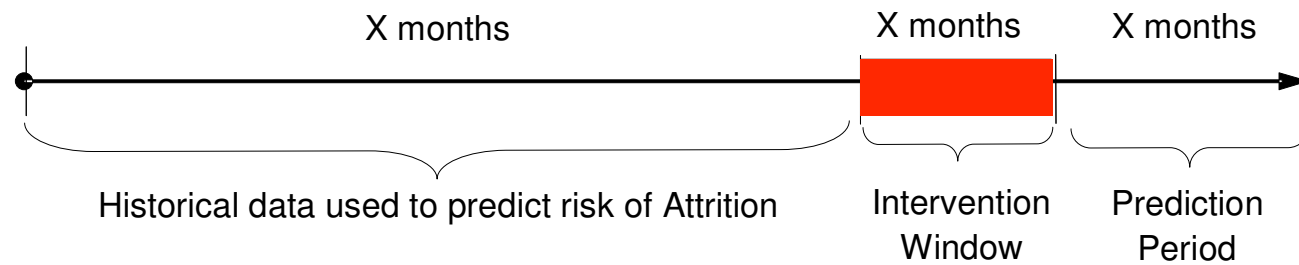
# Retention Modelling Process



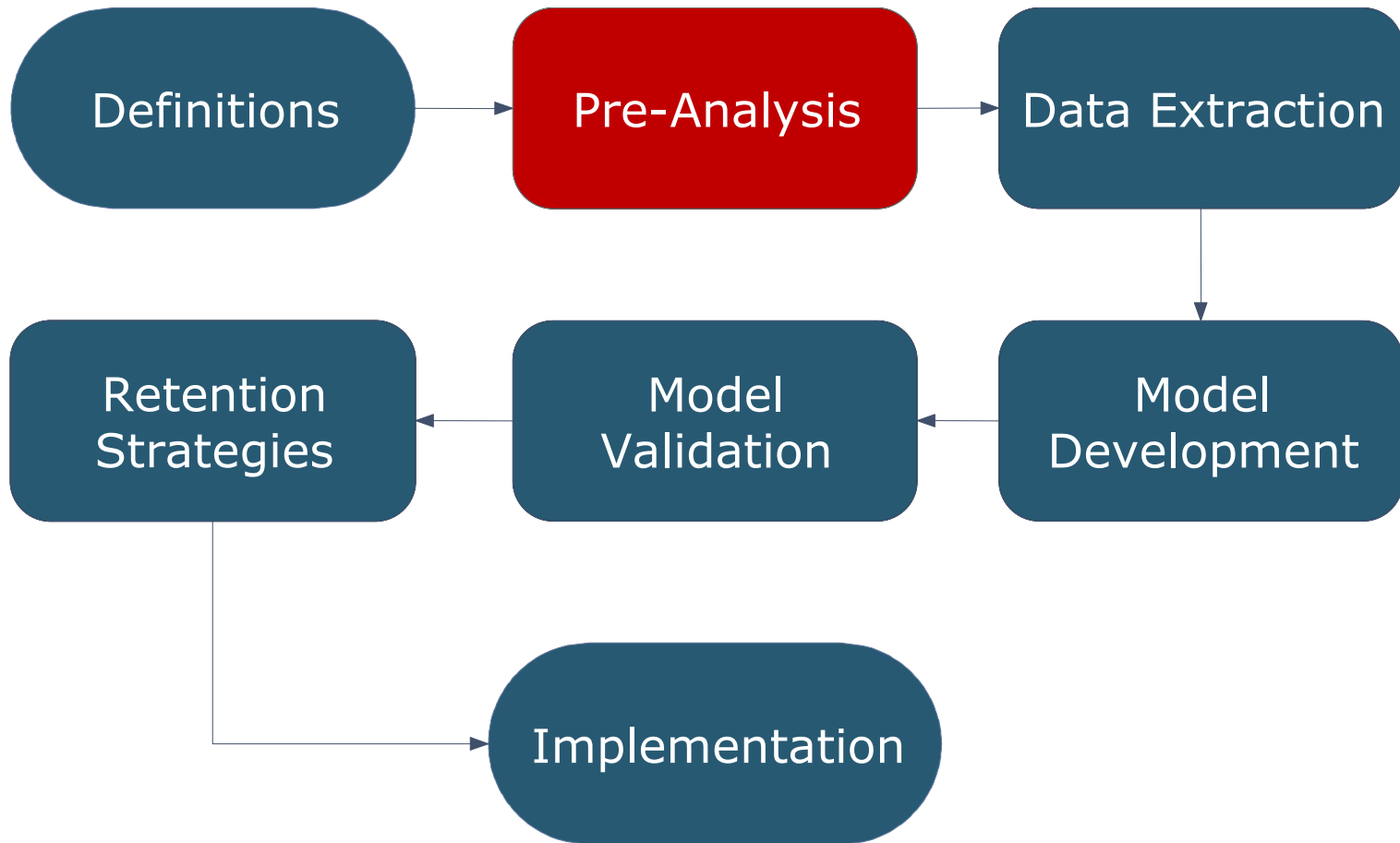
# Definitions

What question is being asked? It is vital to agree on:

- The Attrition (Target) Event
  - Account vs Customer level attrition
  - Account Closure: Forced or Voluntary, Silent Attrition, Balance Transfers
- The Population of Interest
  - Customer Type (e.g. Business, Retail)
  - Card Type & Status
- Timeframes
  - Prediction Period: predicting attrition in the next X months
  - Whether to incorporate an Intervention Window
  - Frequency of attrition scoring in production (e.g. weekly, monthly)



# Retention Modelling Process



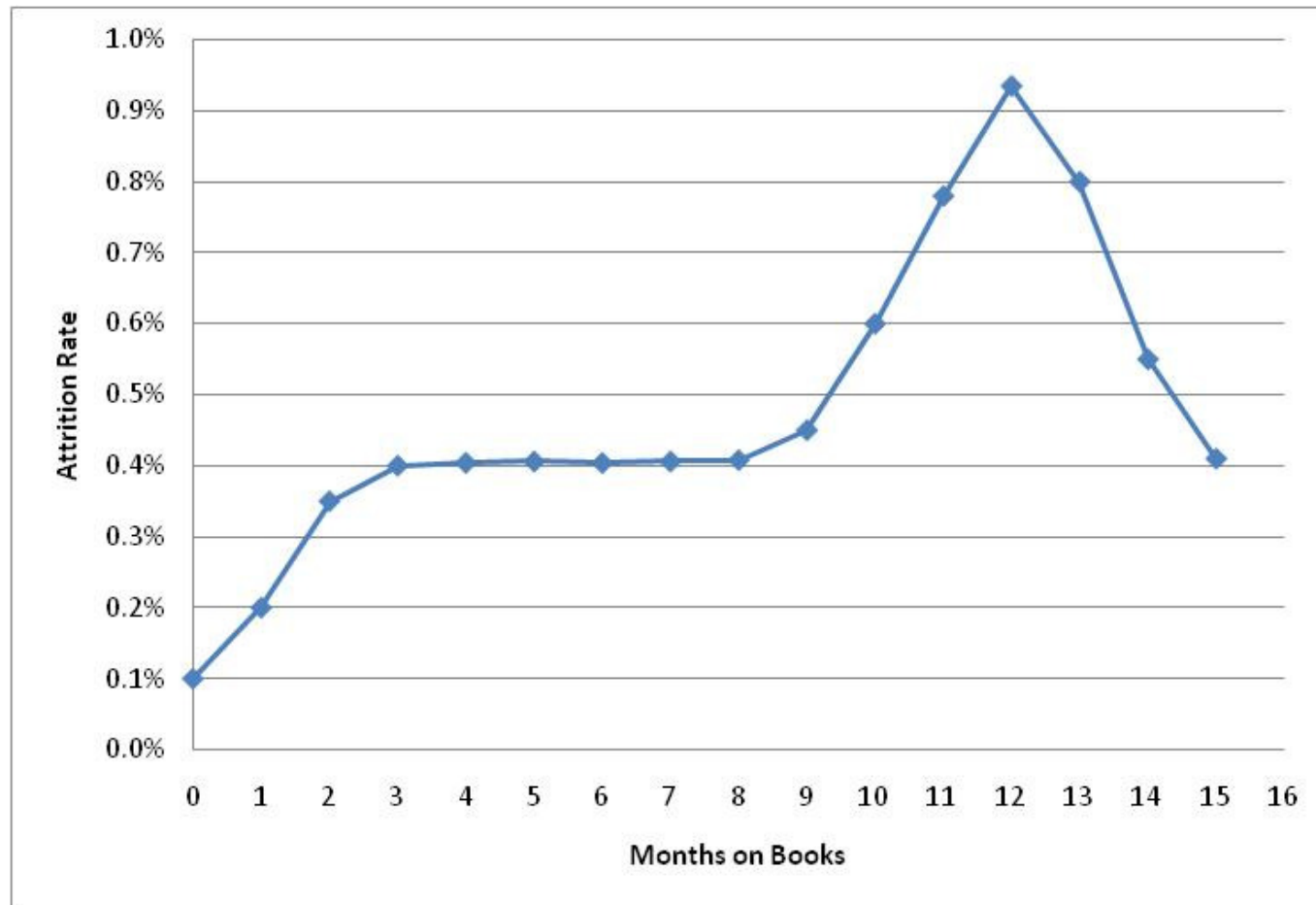
# Pre-Analysis

- Brainstorm causes of attrition with business experts
- Look at past attrition in portfolio:
  - Historical events like introduction of new products
  - Seasonal events (e.g. tax time, year end etc)
  - Life stage (end of introductory rate, annual fees etc)
- Segment by:
  - Card Type (MasterCard, Visa, American Express)
  - Card Status (e.g. Standard, Gold)
  - Cohort (month account opened)

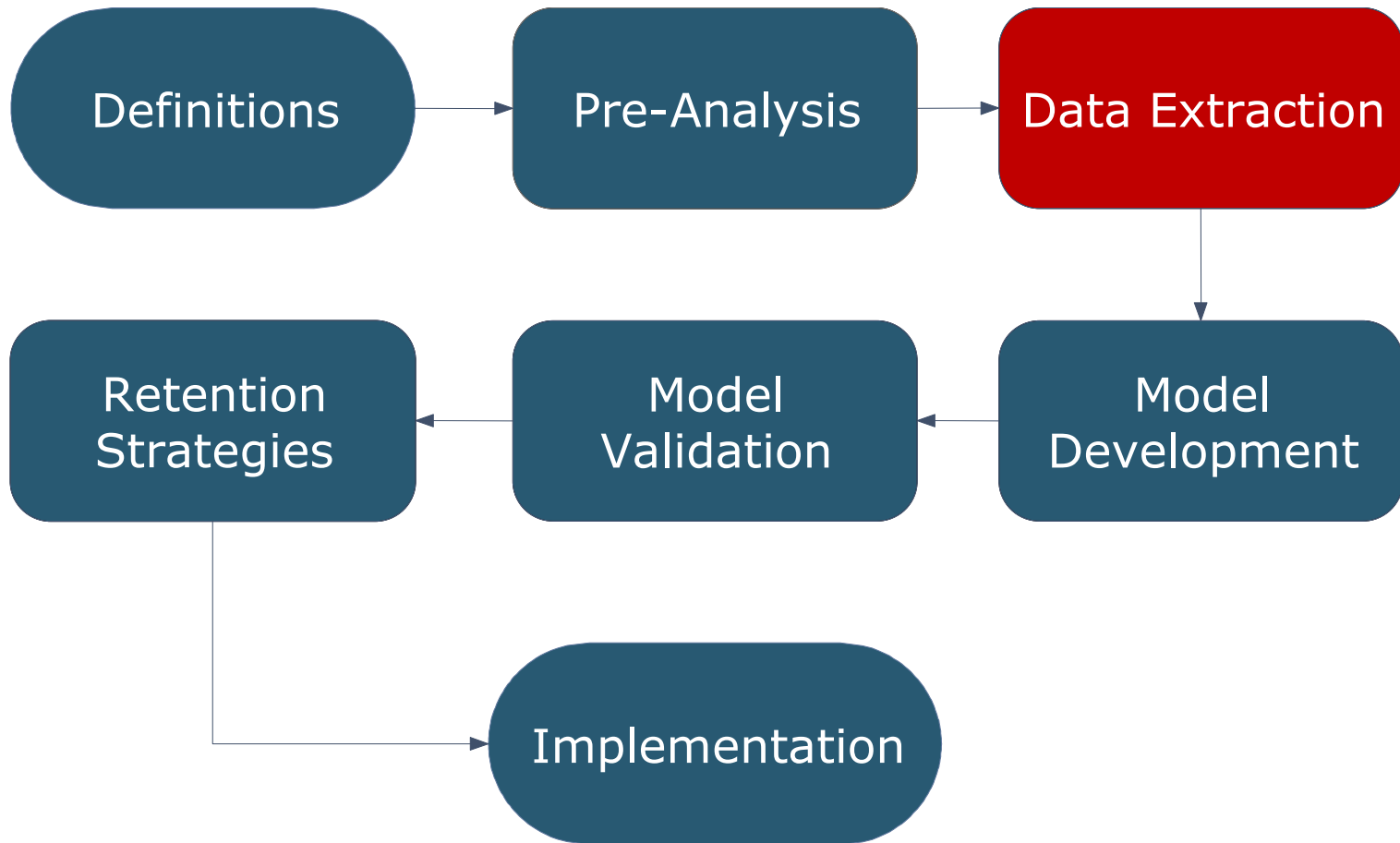


# Pre-Analysis

Attrition peaks around the first annual fee



# Retention Modelling Process



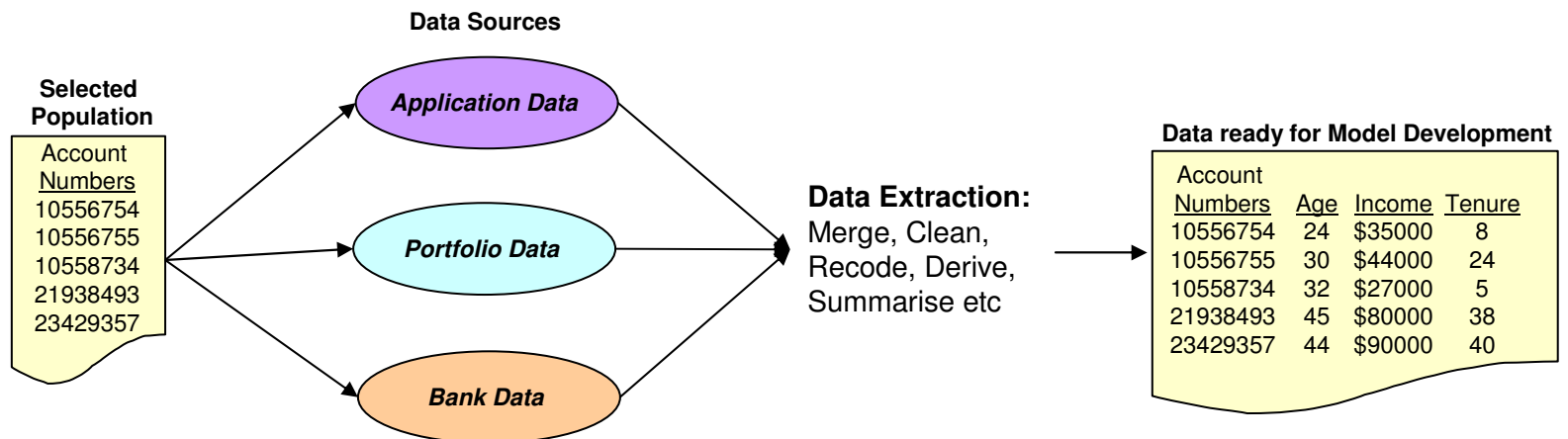
# Data Extraction

- 80% of time is spent extracting and manipulating data
- Decide historical data window to use (e.g. last 12 months)
- System Specific Considerations
  - Identify all applicable data sources
  - Ensure data used is available in production environment
  - Merge disparate data feeds (e.g. match different customer ids)
- Sampling - taking a random sample is a common process used to overcome computational limitations.
- Oversampling – as attrition is a rare event, it may be more appropriate to select all customers who attrited and an equal number of randomly selected to non-attritors.



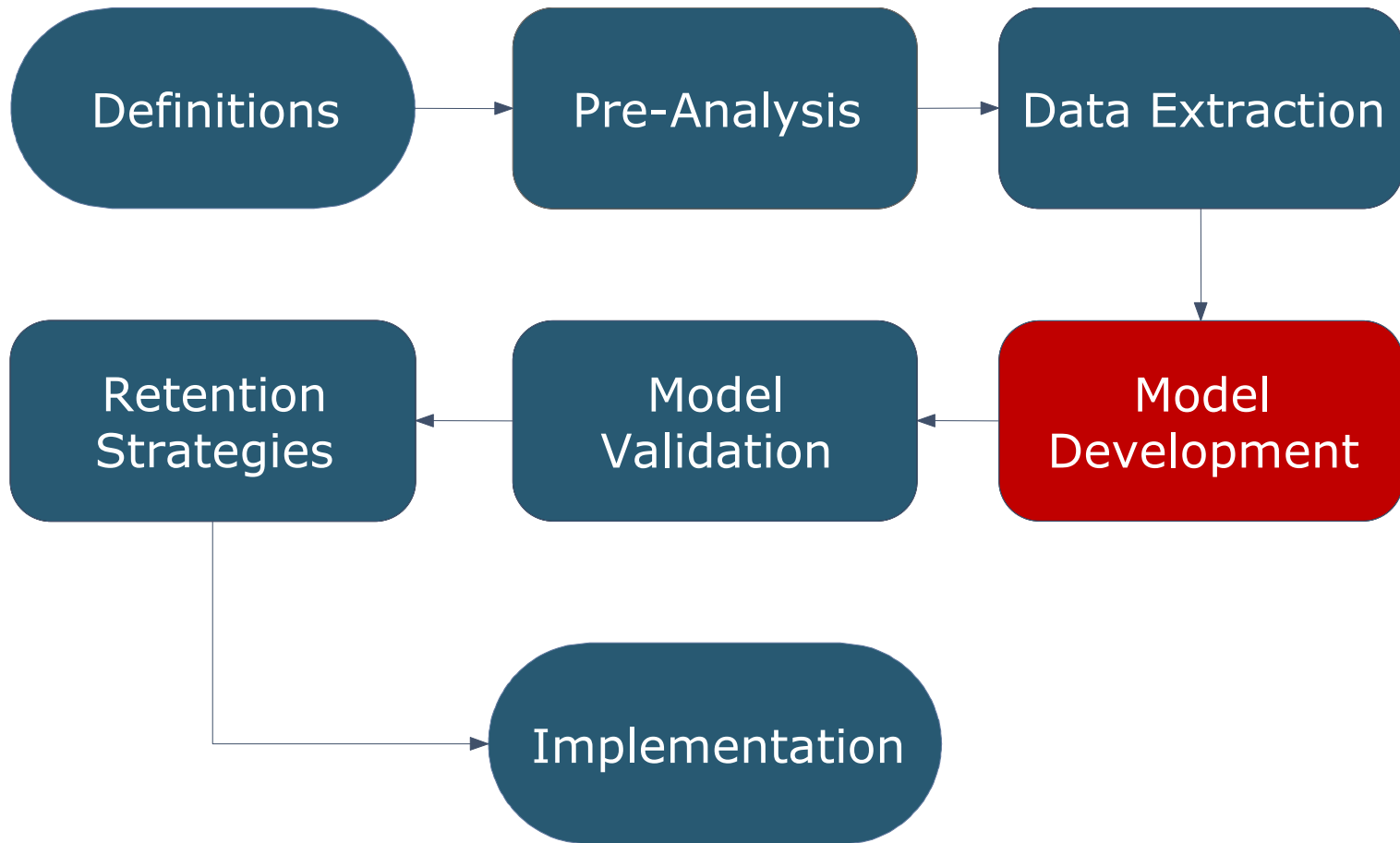
# Data Extraction

- Extract potentially predictive variables
  - Check data quality (e.g. how well is customer age field populated)
  - Recode (e.g. simplify credit card categories)
  - Transform (e.g. take log of dollar value fields)
  - Derive new variables (e.g. time to annual fee due)
  - Roll-up variables (e.g. number of transactions in last month)
  - Summarise variables (e.g. max balance in last 3 months)
  - Include variables known to be predictive from previous models



- Partition into Training (e.g. 60%) and Validation (e.g. 40%) data.

# Retention Modelling Process



# Model Development – Variable Evaluation

Using Training Data, examine the predictive value of variables:

- Correlations
- Other metrics (Max Lift, Information Value)
- Stepwise Discriminant Analysis (using PROC STEPDISC)
- Interactions: Identify using Decision Trees in Eminer

Interaction example:

- Those with Annual fee due are more likely to attrite
- Older customers are less likely to attrite.
- However, having an annual fee due has a greater negative effect on older customers than younger ones in terms of attrition risk.



# Model Development – Final model

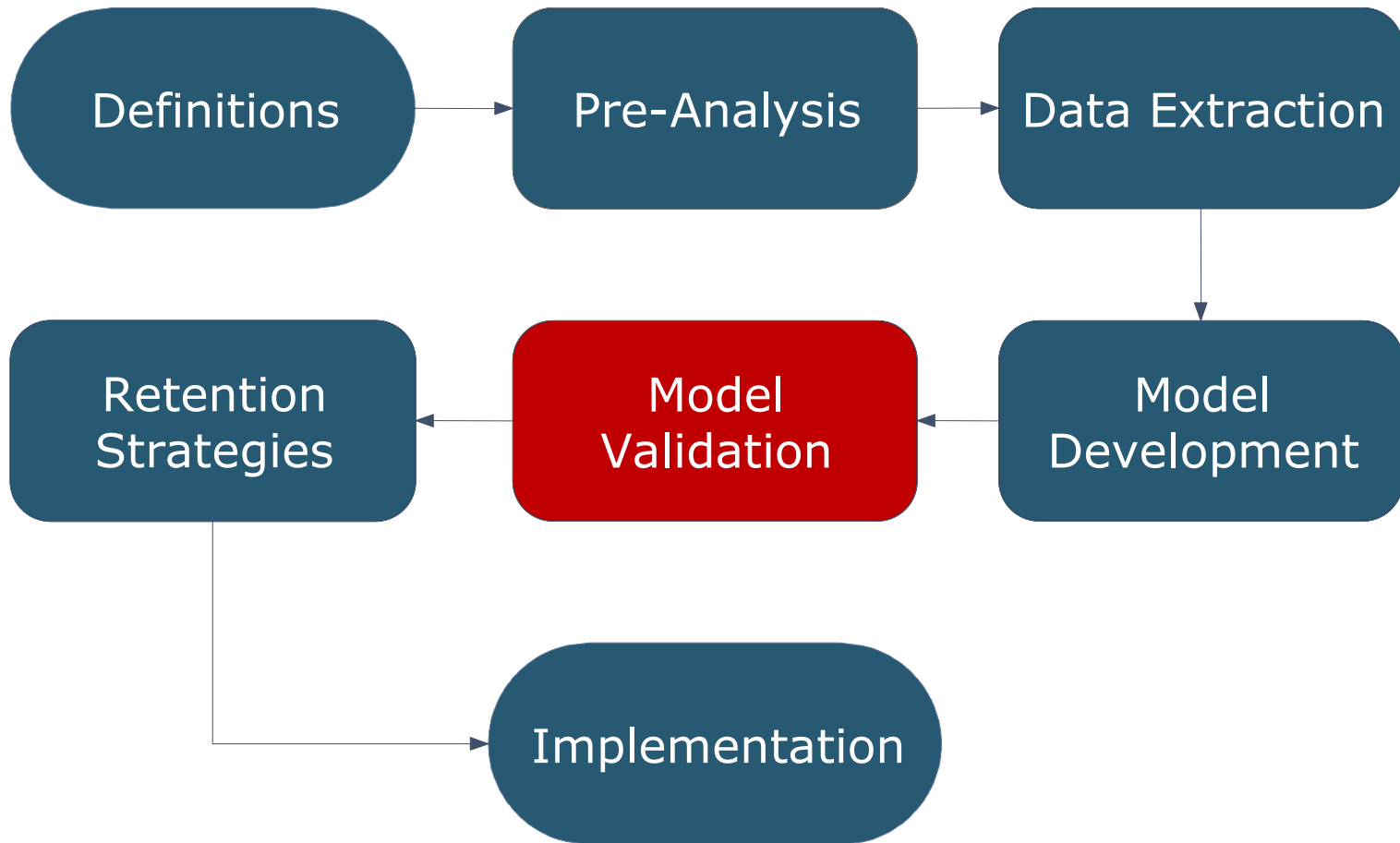
The final model equation is the result of an iterative process including:

- Creating a variable shortlist using evaluation metrics.
- Identifying and addressing the issue of Collinearity using PROC VARCLUS
- Refining the model equation using various PROC LOGISTIC methods and diagnostics
- Selecting the Final Model based on model assessment statistics

Note: With too much obsession it is possible to ‘overfit’, resulting in a model that may not be robust over time.



# Retention Modelling Process



# Model Validation

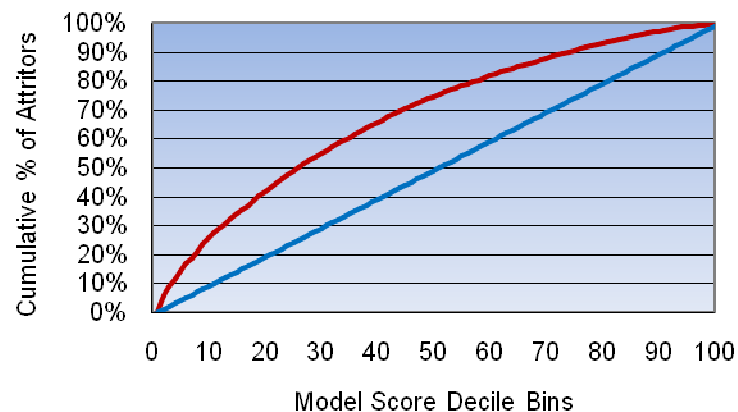
Model Validation is used to check that the final model is robust, ie that it is appropriate:

- For independent data (Validation data)
- Over time ('Out of Time' data)

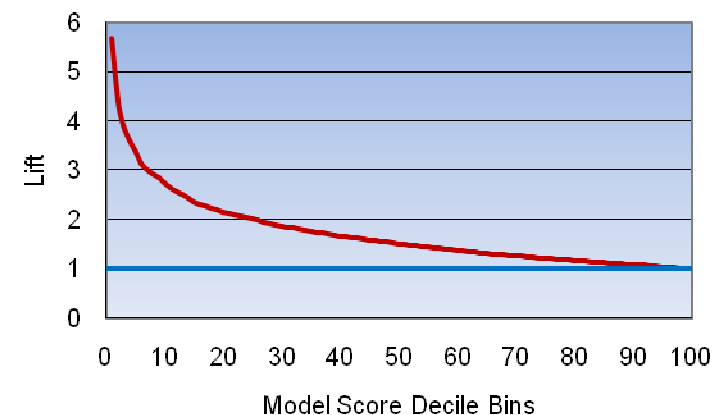
Model validation measures include:

- Gains and Lift charts
- Gini coefficient & K-S statistic

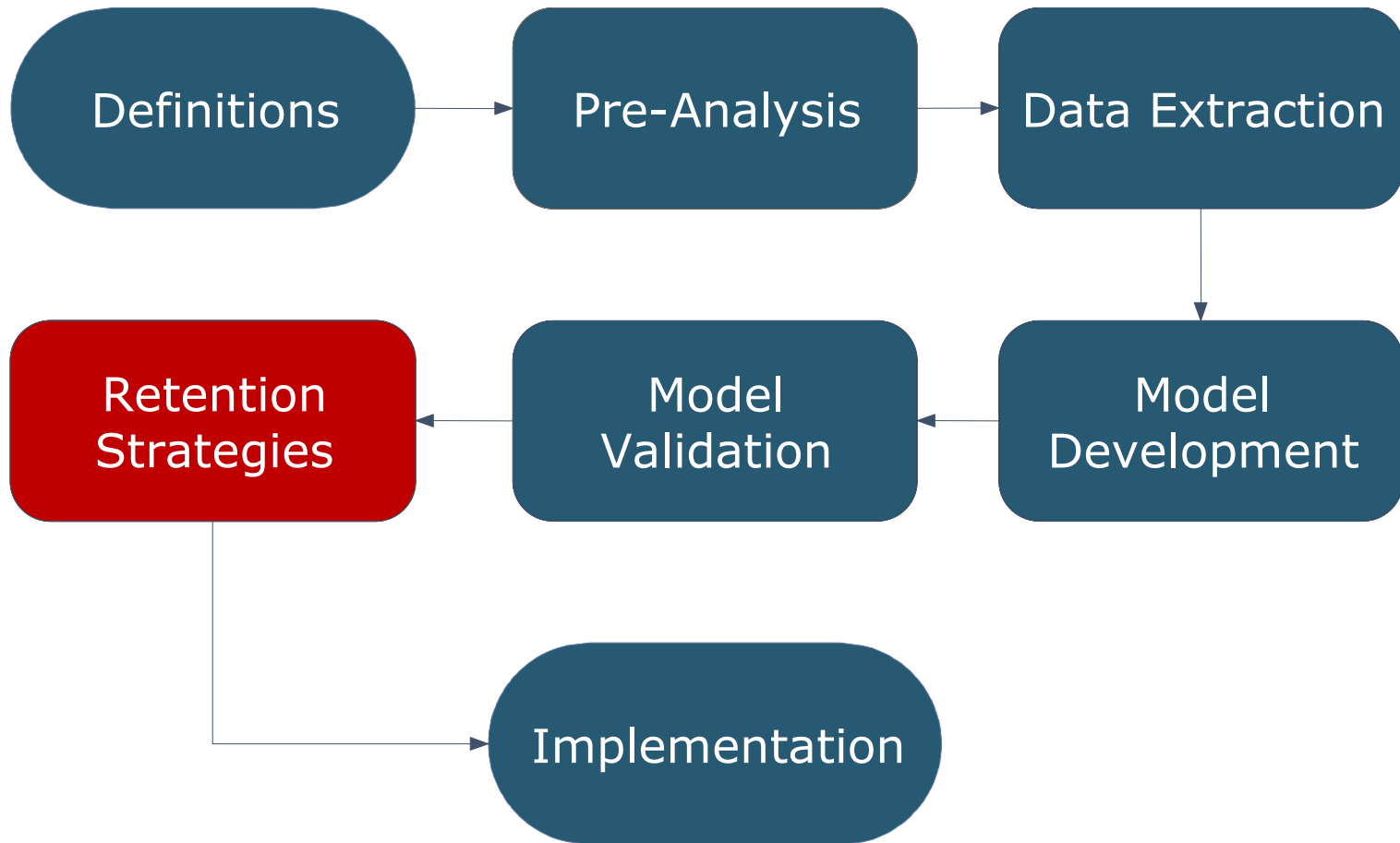
Validation Sample Gains Chart



Validation Sample Lift Chart

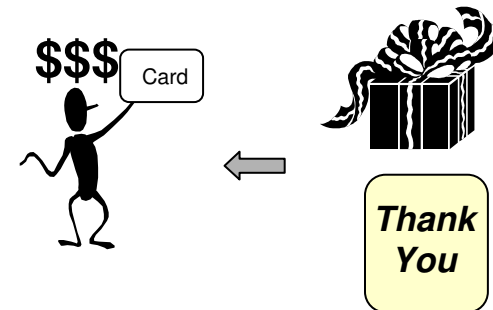


# Retention Modelling Process

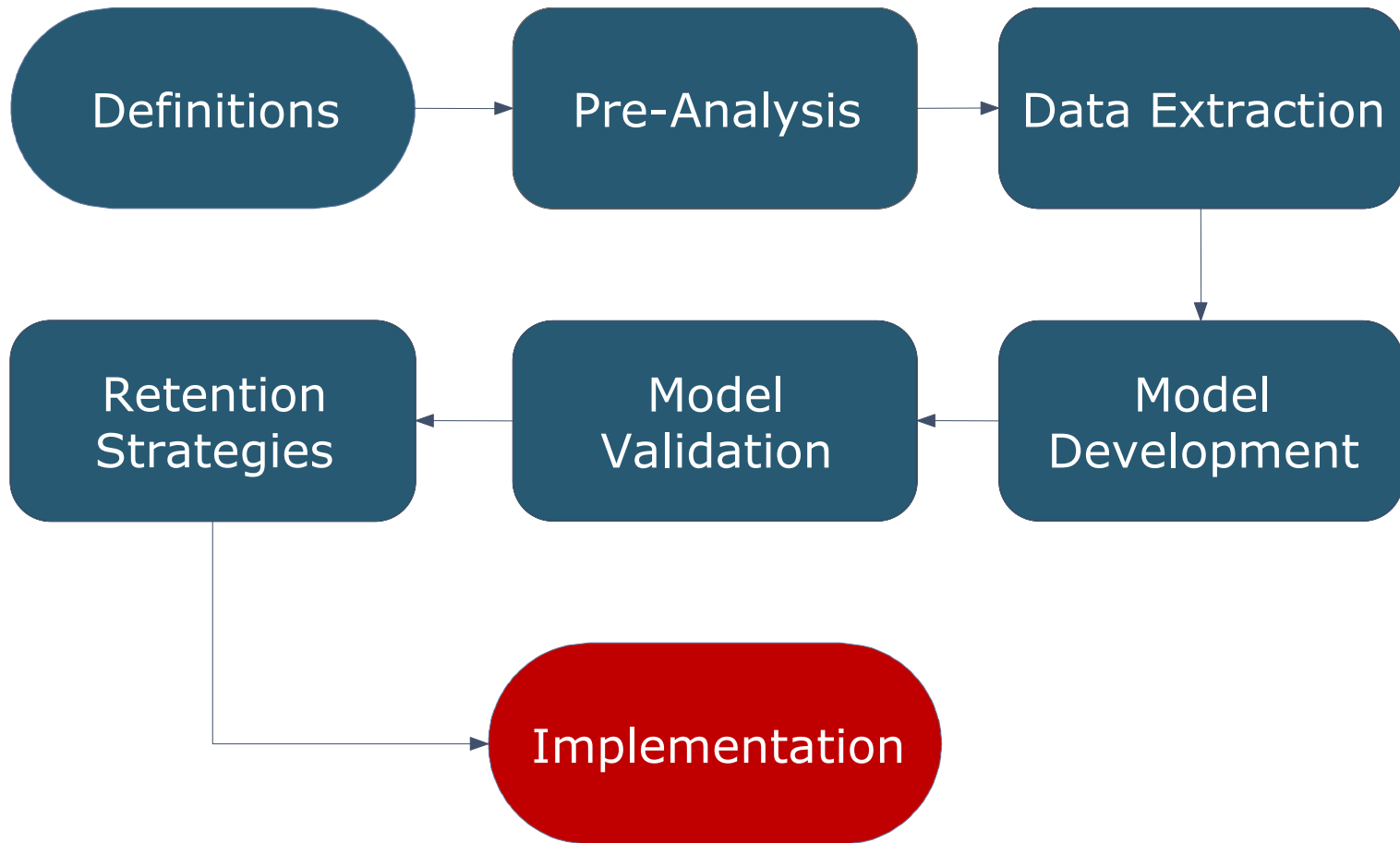


# Retention Strategies

- Create Deciles and Percentiles for risk of Attrition and examine by Segments
  - Profitability: unprofitable, low, high
  - Behaviour Groups
- Target High Profit and/or favoured Behaviour Groups
  - Waive Annual Fee
  - Reward them (e.g. magazine subscription)
- Unprofitable/Undesirable left alone (i.e. no action)
- Moderate deals for others
  - Increase credit limit
  - Small reward (e.g. movie ticket)



# Retention Modelling Process



# Implementation

- To be run every month in Teradata DWH
  - Cut down SQL extraction code to variables in final model
  - Convert SAS data manipulation code to SQL
  - Code Logistic model equation in SQL
- Embed in trigger environment for automatic lead generation (with offer appropriate to profitability and behaviour segment)

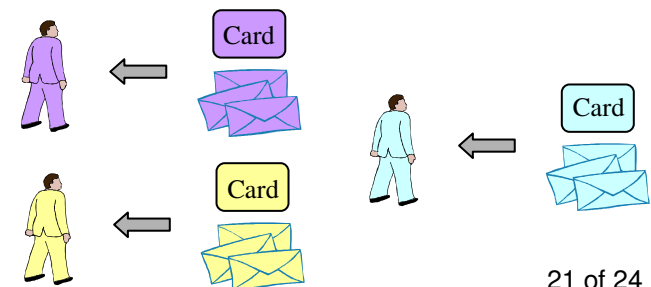


# Results/Outcome

- An understanding of the factors effecting Attrition
- Automated system for predicting Attrition risk with early enough warning to intervene
- Customer segmentation according to profitability and behaviour group
- Strategy on how to deal with 'at risk' customers from different segments

These all resulted in:

- Increased retention of 'good' customers
- Reduced revenue loss



# Tips & Tricks for modelling

- Being able to 'sell' the idea of a model to non-technical stakeholders is just as important as modelling skills
- Involve and pick the brains of subject matter experts
- The most 'statistically correct' model is not necessarily the best - the final model may be selected based on business reasoning in preference to statistical accuracy
- Project Management
  - Do not underestimate the time it takes to extract/manipulate data
  - Allow time for delays outside of your control (e.g. DWH delays, subject experts leaving company, server issues, source system problems)
  - Spend the time to define all parts of modelling project so it is clear before any work takes place
- Modelling is an iterative process of discovery, you will not get the final model the first time round



# Tools of the Trade

- Teradata Datawarehouse
  - Data extraction, final automated process & strategies in production system
- Base SAS
  - Data Extraction/Manipulation, Data Exploration and Model Development/Validation
- SAS Enterprise Miner
  - Data Exploration, Variable Selection
- SAS Enterprise Guide
  - Data Exploration/Visualisation



# References

## **SAS Course: Predictive Modelling with Logistic Regression**

<http://www.sas.com/offices/asiapacific/sp/training/courses/pmlr.html>

## **Communicating the Results of Predictive Models to Non-Technical Audiences**

<http://www2.sas.com/proceedings/sugi27/p124-27.pdf>

## **SAS 9.1 Online Documentation**

<http://support.sas.com/onlinedoc/913/docMainpage.jsp>

## **Statsoft Electronic Statistics Textbook**

<http://www.statsoft.com/textbook/stathome.html>

## **Application of Proc Discrim & Proc Logistic in Credit Risk Modeling**

<http://www2.sas.com/proceedings/sugi31/081-31.pdf>



## **Prophet Analytics – Know Your Customers**

[www.prophetanalytics.com.au](http://www.prophetanalytics.com.au)