

The Power of Direct:
Relevance. Responsibility. Results.



10 Important Tips for Successful Model Implementation

Stephen H. Yu



Vice President, Analytics & Insights

April 7, 2011



What we will cover

- Analytical Environment
- Data Ingredients for Modeling
- Data Summarization
- Types of Variables
- Variable Design
- Scoring & QC
- Back-end Analysis
- 10 Important Tips for Successful Model Implementation



Any pain implementing models?

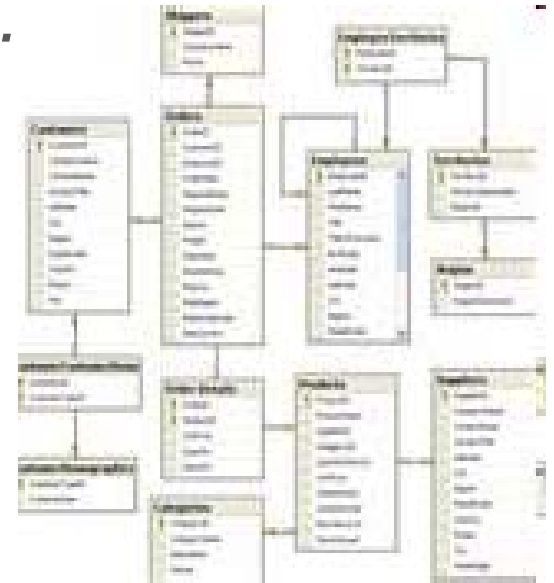
- Not easy to find “Best” customers
- Modelers are fixing data all the time
- Rely on a few popular variables
- Always need more variables
- Takes too long to build models and score
- Inconsistencies shown when scored
- Disappointing results



What does your database support?

If you have a marketing database...

- Order Fulfillment
- Contact Management
- Standard Reports
- Ad hoc Reports and Queries
- Name Selections
- Response Analysis
- **But does it support modeling and scoring?**



Why Model?

- Increase Targeting Accuracy
- Save Cost by Contacting Less / Smart
- Consistent Results
- Reveal hidden patterns in data
- Repeatable – Key for automation
- Expandable
- “Supposedly” save time and effort



Why NOT Model?

- Universe is too small
- Predictable data not available
- 1-to-1 marketing channels not in plan
- Tight budget
- Lack of resources



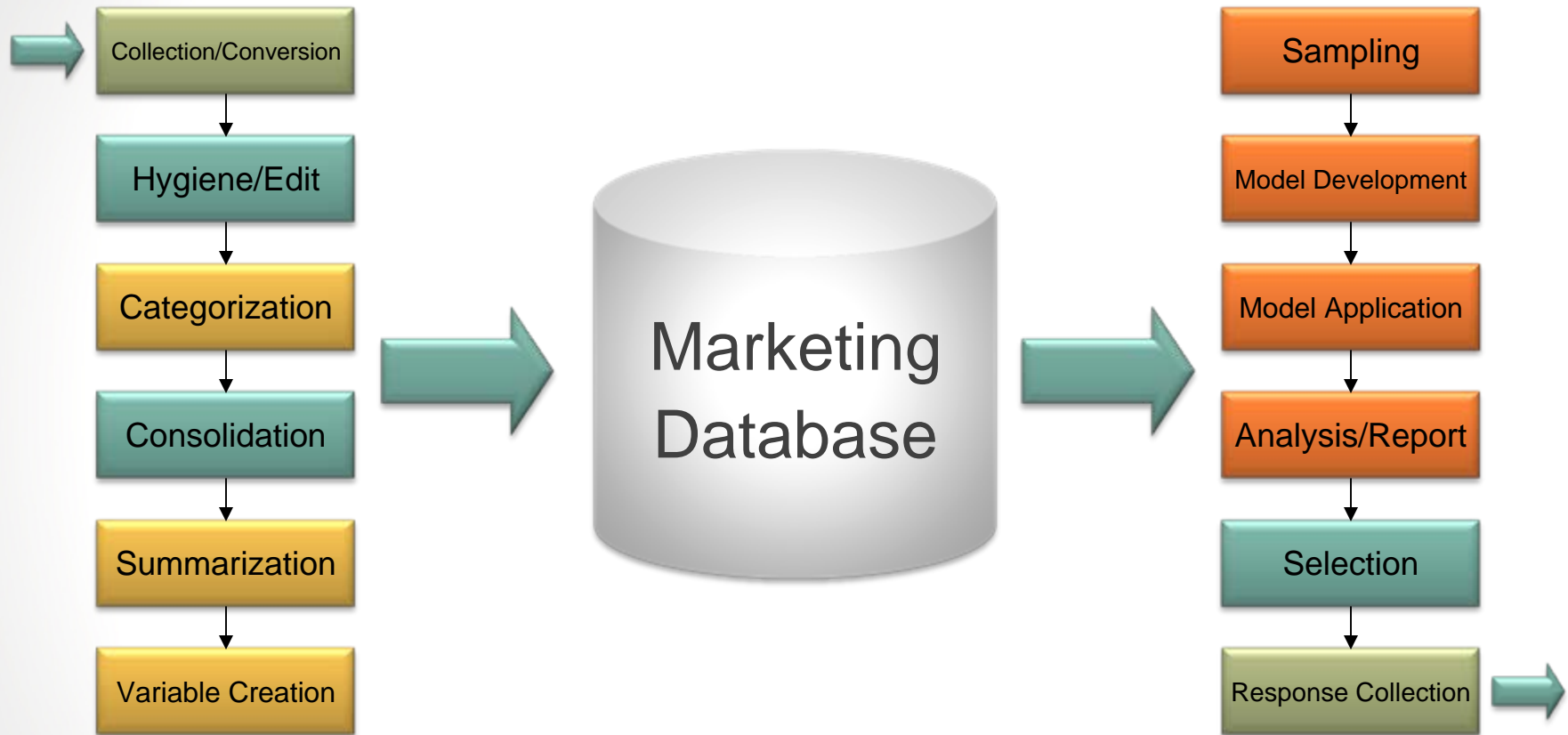
For modeling, clean the data first

“Garbage-in, garbage-out”

- Most data sets are messy & unstructured
- Over 70-80% of model development time goes to data prep work
- Most databases are NOT model-ready
- Modeling & Scoring
 - Extension of database work
 - Consistency is “the” key



“Model-Ready” Environment



“Consistency is the Key”

1. Set Clear Goals

- Short and long-term needs and goals
 - Feasible & attainable
 - Success criteria
 - Know your limitations
 - Technical know-how
 - Resource & budget limitation
 - Divide and conquer
- Plan for the entire marketing cycle



When you set analytical goals

- Consider the end-users first
- Be realistic
 - Budget
 - Resource
 - Data availability
- Put them in phases
 - Most bang for the buck
 - Projects w/ shortest timeline



A Few Analytical Goals

- Rank & select prospect names
- Cross-sell/Up-sell
- Segment the universe for messaging strategy
- Pinpoint attrition point
- Assign lifetime values
- Optimize media/channel spending
- Create product packages
- Project customer value
- Detect fraud



2. Set Roles & Responsibilities

1. Pre-model Data Preparation



DP, Outsource,
Analysts

2. Model Development
& Testing



Statistical Analysts,
Outsource

3. Post-model
implementation



DP, Outsource,
Analysts

Division of Labor

- Clear line of communication – Project Manager
- Set requirements and standards up front
- “ENFORCE” the rules – Internal & External
- Maintain Consistency



Why Front-end DP Important?

- Inexperienced analysts spend majority of time doing DP work
 - Modeling work at the last minute!
- Creative variables enhance models
- Inconsistent data creates a chain reaction to melt-downs
- Data append/match becomes ineffective



Internal Challenges

- Staff – Locate, Hire, and Retain
 - Doers vs. Vendor Managers
- Justify the cost
 - Salary, Software, Hardware, Support staff
- Keep analysts challenged
 - # Projects
 - Diversity of projects
 - Mentoring & training
 - Turn mathematicians into business people



Outsourcing Challenges

- Where to go?
 - Individual Consultants
 - Standalone Analytical Service Providers
 - Database Service Providers
 - Direct Marketing Agencies



Individual Consultants

Pros

- Like employees
- Dedicated
- Full access
- Adapt to company culture and systems

Cons

- Temporary
- One foot out the door all the time
- Walks away with knowledge
- Can be expensive



Analytical Service Providers

Pros

- Focused – this is what they do
- Experts in the fields
- Broad industry knowledge
- Entrepreneurs

Cons

- Often small – stretched staff
- Standalone pricing – not many freebies
- May lack secure data sources
- May not have resources to handle large datasets



Database Service Providers

Pros

- One-stop shop – Support all functions
- Data knowledge
- Deep resource pool
- Friendly package pricing
- Broad industry knowledge

Cons

- Large organizations may move slowly
- Analytics may take back seat
- Emphasis on their own data



Direct Marketing Agencies

Pros

- Strategic
- Controls creative process via segmentation

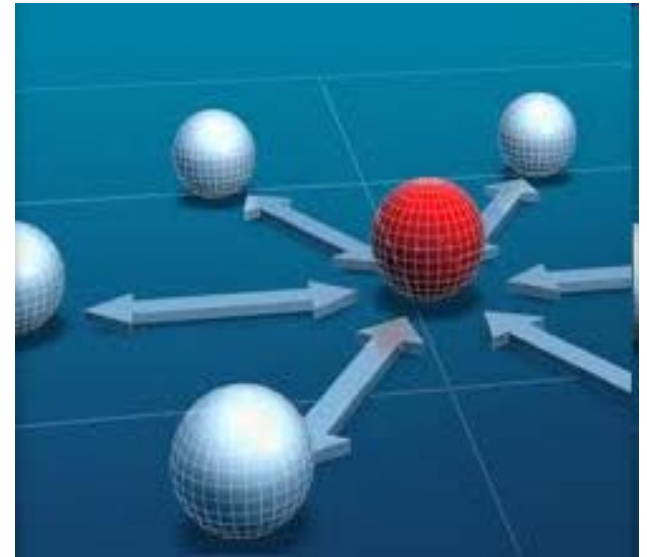
Cons

- May not be core competency
- They outsource too



When outsourcing, consider:

1. Consulting Capability
2. Data Processing Capabilities
3. Track Record in the Industry
4. Types of Models Supported
5. Documentation
6. Scoring Validation
7. Back-end Analysis
8. Speed of Execution
9. Pricing Structure
10. Ongoing Support



Consistency is the key

- Constant communication among all players
 - Common goals
 - Shared documents
- Not just words; data values must be consistent
 - Development Sample = Database Environment
- Enforce rules
 - Make no exceptions



3. Data Inventory

- “Modeling is making the best of what we know”
- Beyond obvious RFM data
- Get Deeper
 - Product/Service Level Data
 - Historical Data
 - Channel Data – Inbound & Outbound
- External Data



Create Data Menu

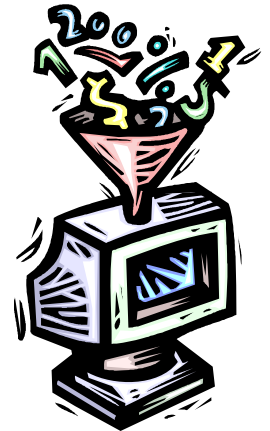
- Base it on Companywide Need-Analysis
- Ask the Analysts first
- What type of models are in the plan?
 - Affinity/Look-alike Models
 - Promotion/Response Models
 - Time-series Models
 - Attrition Models
- Consider non-analytical departments
- Maintain the ones that fit the objective



Check Your Data Inventory

Let's start with what you have

- Name & Address: Key to Geo/Demographic Data
- Order Transaction Data: "RFM", Payment Methods
- Item/SKU Level Data: Products, Price, Units
- Mailing/Response History: Source, Channel, Offer
- Life-to-Date/Past "X" Month Summary Data
- Customer Level Status Flags
- Surveys/Product Registration Forms
- Customer Communication History Data



Need Conversion, Categorization, & Summarization

4. *Data Summarization*

- Make the Best of RFM Data
- Be Creative with Product & Channel Data
- Be Careful with Time-sensitive Data
- Consider the Alternative
 - Easier to summarize scores later?



Modeling is all about “Ranking”

- Ultimately, Models must properly “Rank”
 - Households
 - Individuals
 - Products
- Determine the level of data accordingly
 - Relational databases won’t cut it
 - Must create “Descriptors” that fit the level that needs to be ranked



Variables as Descriptors

If you are ranking individuals, describe the individuals

- Behavioral:

 - ✓ Transaction Data – RFM, Product Purchase

 - ✓ Response Data – Offer, Source, Channel

- Geo-Demographic: Household/Geo

- Attitudinal: Surveys, Inquiries

➤ Match the Level of Data via “Summarization”



Data Summarization – Matching the Level of Data

Order Table

Cust ID	Order #	Order Date	\$ Amount
000123	100011	2007-05-06	\$199.99
000123	100128	2008-08-30	\$50.49
000123	103082	2009-12-21	\$128.60
003859	100036	2008-06-06	\$43.99
003859	101658	2009-01-20	\$43.99
003859	102189	2009-04-15	\$119.45
003859	106458	2010-02-18	\$43.99
004593	104535	2010-07-30	\$354.72
016899	107296	2009-07-14	\$199.99
019872	102982	2008-09-07	\$128.60
019872	103826	2009-04-30	\$499.99
019872	109056	2010-03-12	\$59.99

Order Summary Table

Cust ID	# Orders	\$ Total	First Order Date	Last Order Date
000123	3	\$379.08	2007-05-06	2009-12-21
003859	4	\$251.42	2008-06-06	2010-02-18
004593	1	\$354.72	2010-07-30	2010-07-30
016899	1	\$199.99	2009-07-14	2009-07-14
019872	3	\$688.58	2008-09-07	2010-03-12



Maximize Power of Transaction Data

- RFM Data must be Summarized
(or De-normalized)
- Turn RFM data into Individual / Household Level
“Descriptors”
- Combine with essential categorical variables
(e.g., Product, Offer, Channel, etc.)



Sample Variables after Summarization

<u>Before</u>	<u>After Summarization</u>
Recency	<ul style="list-style-type: none">• Weeks since last online purchase• Years since member sign up• Days since last delinquent date• Months since last response date
Frequency	<ul style="list-style-type: none">• Orders by offer type• Orders by product/service type• Payments by pay method• Average days between transactions
Monetary	<ul style="list-style-type: none">• Total \$ past 24 months• Life-to-date spending• Average dollars by channel• Average dollars by product type



RFM Data Summary – Timeline

- Life-to-date Summary provides the historical view



May create bias towards tenured customers

- Put time limit on variables (e.g. 12-month, 24-month, etc.)



May require higher number of variables and complicate the process

- For Lifetime Value & Time Series Models



Must create historical arrays (daily, weekly, monthly counts of events)

Who does the summary work?

Answer: Not the statistician!

Key Takeaway

The data variables must be consistent everywhere

- Main database
 - Model Development Sample
 - Pool of records to be scored
- Pre-build summary variables in the database



5. Data Categorization

Free-form data come to life through categorization

– *Don't Give Up!*

- Hidden data in:
 - Product, Service, Offer, Channel, Source, Status, Titles, Surveys, etc.
- Have categorization guideline?
- Who will do it?
 - Consider text mining techniques



Collecting Categorical Data

- Any Non-numeric Data
 - Product
 - Service
 - Offer
 - Channel
 - Source
 - Market
 - Region
 - Business Title
 - Member Status
 - Payment Status
 - etc...
- **Categorize as much as possible at the data collection stage**

Example:

Offer Code

- A Flat Dollar Discount
- B % Discount
- C Buy 1, Get 1 Free
- D Free Shipping
- E No Payment Until...
- F Free Gift
- G etc.



Categorization Guidelines

- Be consistent throughout
 - Survey Form
 - Data Entry
 - Inventory Database
 - Data Collection & Compilation
 - Summarization
 - Modeling, and Scoring
- Create “Code” structure
 - NEVER allow free-form answers



Categorization Guidelines (Continued)

- Create Rules and DON'T Deviate from them
- More specific the better
- But, don't allow too many variations (over 20) in one code
- Don't forget the end goal



Training & Automation



Combine them later



Break into multiple codes if necessary



Must be "relevant"

6. *Data Hygiene & Data Append*

- Data conversion
 - Create consistency
 - Standardization
 - Edit
 - Purge
- Cover all bases – PII & RFM Data
- Create rules and be consistent



PII – Gateway to External Data

What is hidden behind simple name & address?

- Standardize Name & Address first
 - Maintain PII (Personally Identifiable Information)
 - Hygiene via periodic NCOA and standardization
- First & Last Name – Ethnic, Gender
- Name, Address, Email – Demographic Data
- Address – Geo-demographic, Census Data
- Zip – County, Market Region, DMA



External Data

- Always consider buying data before collecting and building
 - Compiled Demographic/Firmagraphic Data
 - Behavioral Data
 - Attitudinal/Lifestyle Data
 - Census Data (Census Tract/Block Group or Zip Level)



External Data Check List

- Test multiple data sources
 - Friendly variable definitions for analysts
 - Coverage/Match Rate
 - Price
 - Who will do the match & append?
- Learn about the data sources
 - What's real and what's imputed?
 - Don't stop at Demographic: always add "Behavioral" data



7. Missing Values

Missing values are inevitable...

- For Numeric Data (e.g., \$, Counters, Dates, etc.)
 - Incalculable vs. Data-append Non-matches
 - Missing is missing: DO NOT fill in with 0's
- For Categorical Data (e.g., Codes, Text, etc.)
 - Leave room for "N/A" (e.g., blank, "N/A", "0", ".", etc.)
 - Code "Non-matches" to external files differently

"Missing Data can be meaningful."



More on Missing Data

- Agree on Imputation Rules
 - Do it upfront
 - Must be part of scoring codes
- Educate non-analysts
 - Hard to undo when combined with other values
 - Train data-append vendors
- Always check % missing
 - Development Sample vs. Life Databases



8. Sampling

- Pre-select
 - Be consistent
 - Don't over do it
- Method
 - Nth: Better for production environment
 - Random
 - Set standard programs
- Who does it?
 - Analysts
 - Part of database system?



Database vs. Sample

- Development Sample vs. Live Database
 - Database Structure
 - Variable List/Name
 - Variable Value
 - Imputation Assumption

Lead to disasters if “anything” is different

- Do NOT play with model groups that are set in the development sample



9. Scoring QC & Installment

Most troubles happen after the models are built...

Check:

- Model Group Distributions
- Variable distributions (values and indices)
- Missing Values
- Match rate for appended data
- Scoring codes, including score breaks
- Compare to previous runs – Check Deterioration

Set parameters for acceptable differences and Enforce



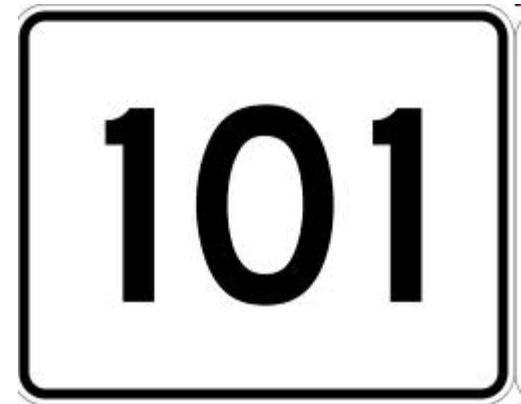
Score Installment in the Database

- Plan ahead to store model scores in the database
- Reserve space for future models
- Store raw scores, not just model groups
- Match the level
 - Household
 - Individual
 - Email
 - Product



10. Back-end Analysis

- **Close the Loop Properly!**
1-to-1 MKT 101: “Learn from past campaigns”
- Must Plan ahead
 - No excuse for not doing it
 - Schedule ahead
 - Budget
- Set Metrics Upfront
 - List Source
 - By Offer, Creative, Season, Product, etc.



Match-Back

- Do not rely on anecdotal results
- Maintain the Master Mail File w/ Source Codes
- Cover all response channels – Off & Online
- Leave it to professionals for “soft” matching
 - Be consistent with match rules
 - Reconcile conflicting data
 - Rules for allocations of mismatches
- Set time window



Match-Back (continued)

For “List Source” Analysis

- Do NOT use Output of Merge/Purge
 - M/P picks one lucky list source from multi’s
 - Random M/P is not fair for smaller lists
 - List prioritization favors top lists
- Use Pre-Merge/Purge file
 - Use “Responder File” as the base
 - Examine “All” lists contributing to responses



Response Reports

- Start with “Canned” Reports from vendors
- Get ready to pay for “Custom” reports
 - Prioritize what you want
- Format and Delivery
- Timing and Interval
- Timeline to be covered (YTD, 12-mo, etc.)
- If you want Cost Analysis, share the details



Key ROI Metrics

Set ROI Metrics, such as:

- Open, Click-through, Conversion Rates
 - “Denominator” in each?
- Revenue
 - Per 1,000 mailed/calls
 - Per Order
 - Per Display, Email, Click-through, Conversion
- By
 - Source, Campaign, Time Period, Model Group, Offer, Creative, Targeting Criteria, Channel (in & outbound), Ad server, Publisher, Key word, Script, Daypart, etc.



Response Data

- Be mindful about “Statistical Validity”
 - Response Rates are NOT baseball scores
 - Size Matters: Segment Size & Size of Differences
 - Set Confidence Level
 - Consult Statisticians
- Get ready for another round
 - Plan for Response Models
 - Maintain Promotion History, not just Responses



Next Steps

- Spec it out
 - Project Goal
 - Data Source List (as detailed as possible)
 - Final Variable List
 - Project Flow:
 - Data Collection
 - Conversion & Categorization
 - Summarization
 - Matching / Data Append
 - Development Sample
 - Scoring
 - Storage
 - Backend Analysis



Plan for Marketing Database

- Know what you need, but don't over do it
 - “Modeling is making the best of what's available”
- Take a phased approach
 - If budget is tight, start with low hanging fruits
 - Proof of concepts without full database commitment in the beginning
 - **Maintain consistency**
 - **Keep Historical Data**



Key Takeaways

- Don't lost sight of long-term goals
- Constant communication among key players
- Consistent data every step of the way
- Match the levels of data (Data Summary)
- Don't over-do it – Phased approach
- Ask for help
- Close the Loop Properly



Q&A

Stephen H. Yu



Vice President, Analytics &
Insights

stephen.yu@infogroup.com

www.infogroup.com





Thank You!

The Power of Direct:
Relevance. Responsibility. Results.

DMA 
Direct Marketing Association