Monetizing Data Assets

Through Data Refinement and Analytics

Stephen H. Yu

President & Chief Consultant

Willow Data Strategy, LLC

What we will cover

DATA LANDSCAPE

INSIGHTS VIA ANALYTICS

ART OF TARGETING

"ANALYTICS-READY" ENVIRONMENT

PROPER PERSONALIZATION USING PERSONAS

POST CAMPAIGN ANALYTICS & ATTRIIBUTION

Data Landscape

Solution-Oriented Data and Analytics Strategy

The Age of Ubiquitous Data & AI

In the world where every channel is turning into an "addressable" medium…

- Data is everywhere, like in the Matrix
- Every breath you take, every move you make…

→ Are you harnessing the power of data?

Or are you just creating mounds of data?

Advanced analytics is within our reach,
as machines are getting continuously smarter

Database Marketing Landscape

- No guessing game – You MUST know your target
- Vast amount of data collected in the name of Big Data → But, are they being used properly?
- Analytics play a huge role in prospecting & CRM
- Short paced marketing cycle getting shorter
- Huge difference between advanced marketers and those who are falling behind

→ It is about the commitment, not technology

Winners are the ones who know how to wield the power of all available data faster.

Where the Data Movement is Going

- Every channel is, or will be interactive
- Every interaction will be data-based, and in real-time
- Marketers must stay relevant to cut through the noise
- Machine Learning & AI will lead to automation on all fronts
- Huge difference between advanced users and those who are falling behind

It's all about staying relevant through "Personalization", through every channel, with every customer, at all times.

It is NOT about Channels or Technologies

"The Future of Online is Offline" *– Stephen H. Yu, 2002*

- There is no such thing as an "online person"
  - It is almost offensive to the consumer
  - Channel-centric view confuses buyers
  - New channels and technologies in the future

- This data and analytics business should be about "People"
  - No one is one-dimensional – Adopt "Buyer-centric" point of view
    - Never be channel-, product-, division- or brand-centric
      – But most marketers are
    - Buyer-centric data structure leading to proper "Personalization"
  - Never about the technology, but about the people who are looking at screens (or even thin air)
    - And they are in control, not the marketers!

What 1:1 Marketing is about

- **Marketers must know:**
  - Whom to contact, and
  - What to offer, if they decided to contact someone
    - » And through what channel and when
- **Analytics help marketers with:**
  - Targeting
  - Personalization

Everyone is being bombarded with marketing messages through multiple channels everyday

WILLOW DATA STRATEGY
making sense of data

It is about the Data Users, too

For Decision Makers

- Take the data seriously, not just your gut feelings
- Define the goals first, then control the flow of data and technologies
- Don’t blindly trust machine-based solutions
- Be logical, as there is no toolset that can read minds (yet)
- Set specific goals for small successes

For Data Scientists

- Don’t be a “Data Plumber”, but a businessman
- Don’t be technology oriented, but solution oriented
- Don’t do things just because you can
- Don’t be part of arrays of machines, only to be replaced by them

→ Be the one that wields machines with clear objectives

What is a Data Scientist?

But,
who would have
all these skills
at a master level
in one body?

WILLOW DATA STRATEGY
making sense of data

Being an Outstanding Data Scientist

A Data Scientist is a combination of:

- Business Analyst – w/ deep understanding in business challenges and industry landscape
- Master Data Manipulator – w/ functional coding skills
- Statistical Analyst – w/ knowledge in statistical modeling and Machine Learning

Elements of a Good Data Scientist:

1. Technical Skills
2. Positive Attitude
3. Communication Skills
4. Business Acumen

What a Business Analyst does:

- Identify business challenges and issues
- Provide solutions to meet business goals
- Develop analytics projects to produce tangible deliverables
- Identify and examine data sources and procure necessary data
- Consolidate data from disparate sources
- Create technical specs for data refinement and analytical steps
- Define targets using logical expressions
- Act as a project manager to push sub-tasks forward
- Develop reports and data visualization
- Create "stories" based on learnings and share them with business community
- Present "executive summary" and next steps to decision makers

Insights via Analytics

Solution-Oriented Data and Analytics Strategy

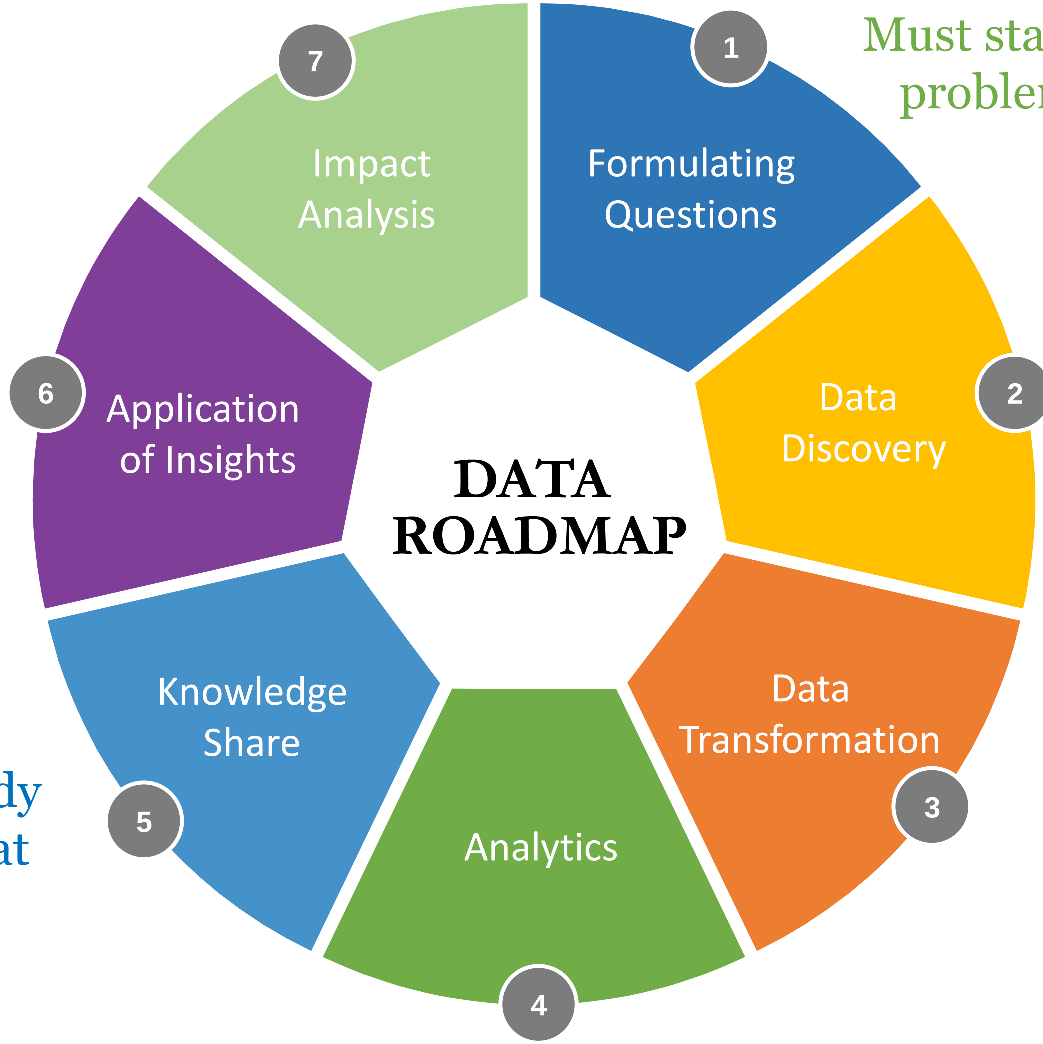
Build the Data Roadmap

A database is not just a sum of all data sources, and Analytics is not just an array of statistical techniques

1. Business Goals
2. Answers via Analytics/Modeling
3. Databases Optimized for Analytics

→ Solution design based on business goals, not around capabilities

Goal-Oriented Step-by-Step Approach



Must start with a clear problem statement

But, get ready to jump in at any stage

WILLOW DATA STRATEGY
making sense of data

Big Data Must Get Smaller

Usable information is:

- Easily understandable without deep knowledge in analytical techniques or mathematics
- Digestible small bits of information, not mounds of data
- Consistent, accurate, effective, and easy-to-use tool
- Useful in most cases, not just limited instances
- Readily accessible through users' preferred devises and channels

Users need "Answers to Questions". So,

1. Cut down the noise
2. Provide insights, not data

through Analytics

WILLOW DATA STRATEGY
making sense of data

Pillars of Data and Analytics

DATA PLAYERS MUST EXCEL IN:

Collection

- Data Collection by Channel
- Rapid Data Retrieval
- Basic Dashboard

Refinement

- Data Hygiene and Standardization
- Consolidation and Summarization
- Advanced Analytics including Statistical Modeling

Delivery

- Comprehensive Dashboard and BI Reporting
- Ad hoc Reports
- Campaign Targeting and Management
- Personalization

Insights are derived from data through the refinement process

Different Types of Analytics

"ANALYTICS" MEANS DIFFERENT THINGS…

Business Intelligence Reporting:

KPI Reports/Dashboards,
Digital Analytics,
Product/Channel Reports,
Campaign Performance Analysis

Predictive Modeling:

Propensity Models, Personas,
Response Models,
Customer Value Projection,
Churn Prediction,
Demand Forecasting

Descriptive Analytics:

Customer Profiling,
Segmentation & Clustering,
Customer Journey Analysis

Optimization:

Channel Optimization,
Marketing Spending Analysis,
Marketing Mix Modeling,
Attribution, Pricing Models

- Plus, Prescriptive Analytics can be applied to any stage

17

Matching Analytics to Marketing Cycle

Acquisition	Relationship Development	Retention	Winback
Acquisition Model	**Cross-Sell Up-Sell Model**	**Retention / Attrition Model**	**Reactivation / Win-back Model**
Who is most likely to respond?	Maximize product/ channel potential?	Who is vulnerable?	Who is likely to come back?
Approval / Screening Model	**Customer Value Model**		
Who is likely to be qualified?	What is the future value of each customer?		

Identify "Valuable-Vulnerable" segment with CLV and Churn Prediction models

WILLOW DATA STRATEGY
making sense of data

Analytics Solutions by Marketing Cycle

Types of Analytics	Acquisition	Relationship Development	Retention	Winback
Predictive Analytics	Affinity / Clone / Prospecting Model	Cross Sell / Up-sell Model	Retention / Renewal Model	Reactivation / Winback Model
	Lead Scoring	Spending Estimation	Attrition / Churn Model	
	Approval / Screening Model	Lifetime Value Model		

Predictive Analytics

- "Personas" – Personal Attributes
- Forecasting / Optimization / Attribution / Media Mix Modeling

Descriptive Analytics

- Customer Profiling
- Customer Segmentation / Clustering
- Market Basket Analysis

Business Intelligence

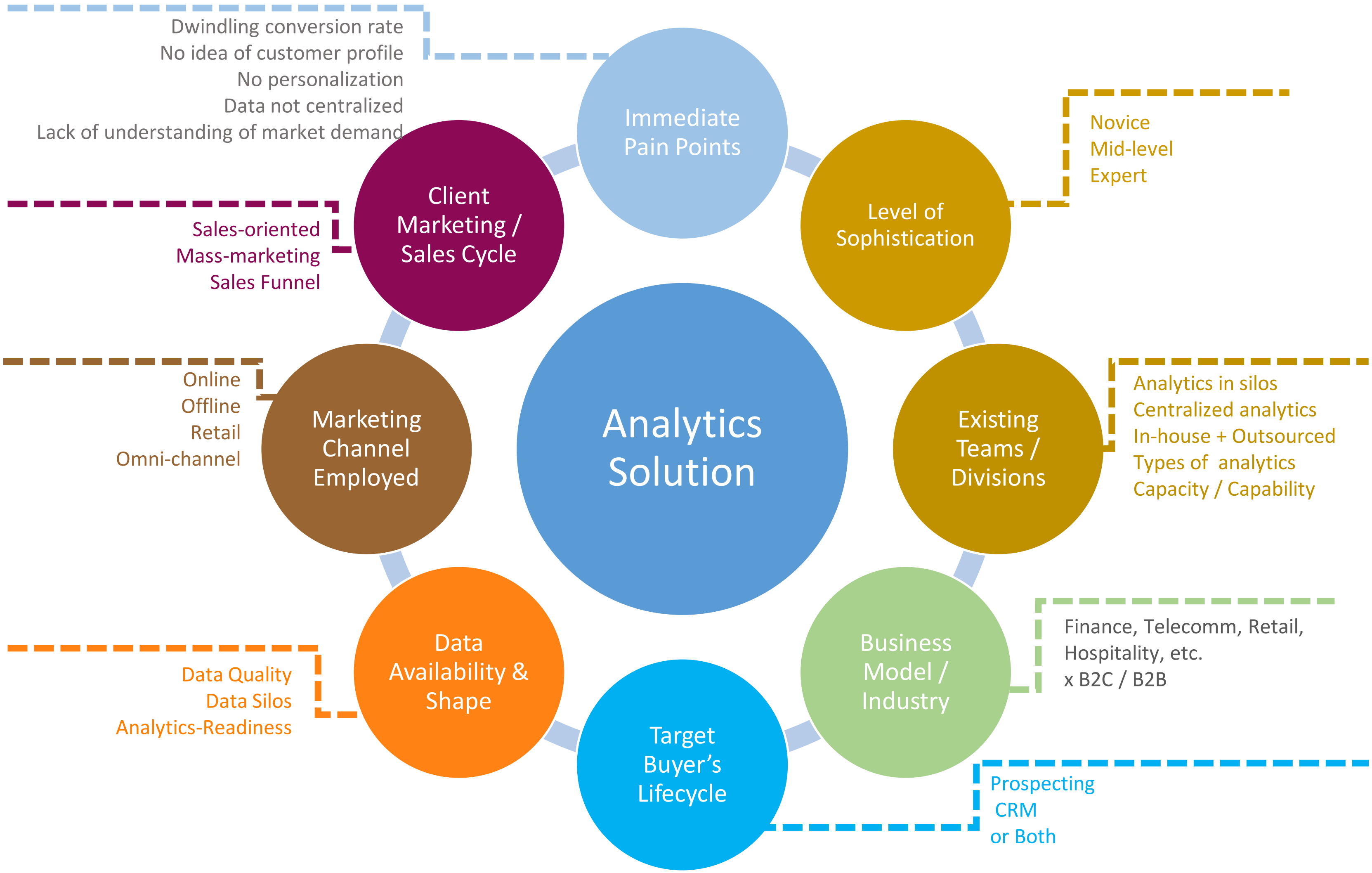
- Digital Analytics / BI Reporting
- A/B Testing
- Competitive Intelligence

Data Strategy / Management

- "Analytics Sandbox" – Analytics-Ready Data Environment
- 3rd Party Data Evaluation

WILLOW DATA STRATEGY
making sense of data

8 Dimensions of Solutioning



WILLOW DATA STRATEGY
making sense of data

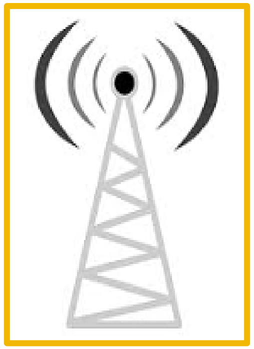
Different Goals & Data for Different Industries

Banking, Finance & Credit Card

Travel, Hospitality & Entertainment

Retail – Online & Offline

Publications



Telecommunications & Utilities

Non-Profit

Catalog

B2B Paths to Conversion

End-to-End B2B Sales Journey – from Prospecting to Nurturing

Key Questions

What are the key marketing levers that attract more prospects?

Of the respondents, who are the most likely to convert?

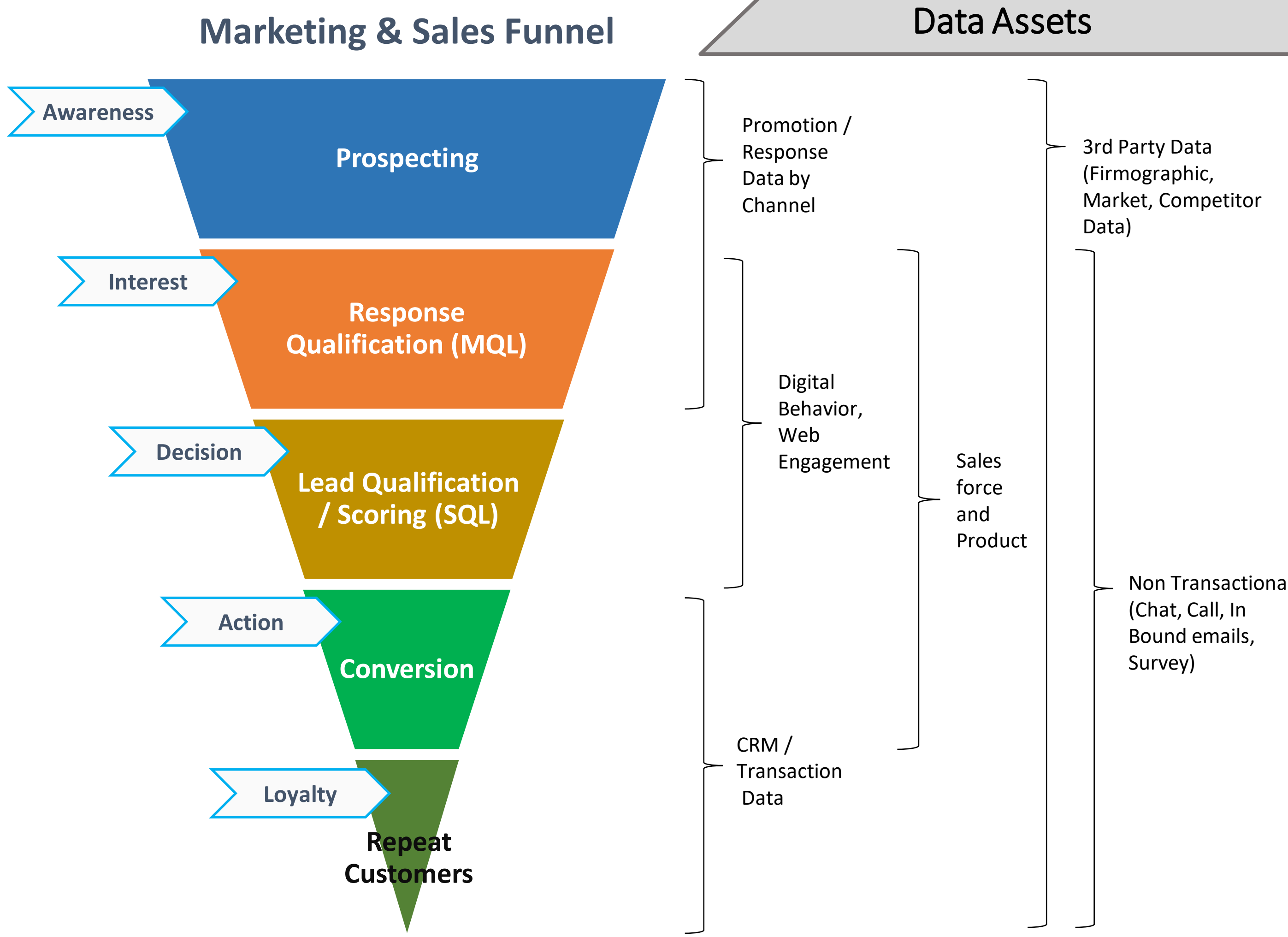
What % of leads are active, stagnant, or lost in the funnel? What are their profiles, and which factors move them forward?

What are the optimal number of touchpoints required for conversion? What are the key channels and offers for these touchpoints? Is there a specific path that results in higher conversion?

What are the key drivers of sales? (e.g., promotion, price, offer, sales engagement, channels, creative, deal size, etc.)

What specific cross-sell / up-sell opportunities should we pursue on a customer level?

Who are the customers more likely to churn? Where are 'valuable-vulnerable' segments?



WILLOW DATA STRATEGY
making sense of data

Why Model?

Converting Complex Data into Answers to Questions

Why Model?

- Increase targeting accuracy
- Reduce costs by contacting less/smart
- Stay relevant with target customers
- Achieve consistent results
- Reveal hidden patterns in data
- Reach marketing automation faster
- Expand the target universe
- “Supposedly” save time and effort

Models summarize complex data into simple-to-use “scores",
and fill in the gaps by converting “unknowns” to “potentials”

Why NOT Model?

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

Really? Remember

1:1 Marketing is about:

- Knowing whom to engage
- Knowing what to offer if you decided to engage someone

Models provide answers for both

Data to Answers via Modeling

Raw Data	Actionable Answers
» Demographic / Firmographic	» Likely to buy a luxury car
» Transaction Data / RFM Data	» Likely to be an early-adopter
» Products & Services Used	» Likely to be a wine enthusiast
» Promotion / Response History	» Likely to have a home office
» Channel Usage Data	» Likely to be a risk-averse investor
» Lifestyle / Survey Responses	» Likely to respond to free shipping offers
» Delinquency history	» Likely to be a high value customer
» Call / Communication Log	» Likely to be qualified for credit
» Movement Data	» Likely to upgrade/leave/come back
» Survey / Social Media / Sentiments	

Formulate the answers through modeling

Problems with Rule-based Segmentation

Variable Selection

- What variables to use? Same old RFM?
- Data overload – Too many variables to choose from
- Resorting back to a few popular variables like Income & Age

Weight Factor

- Not all variables are equally important
- Variables interact – Dynamics hard to detect by intuition
- Some are even negatively correlated

Banding

- How high is high enough?
- Popular banding (e.g., income over $100k) may not work
- Grouping categorical variables is difficult

Do not just rely on your gut feelings – Employ advanced analytics

What is a Model?

"Model is a mathematical expression of <u>differences</u> between groups"

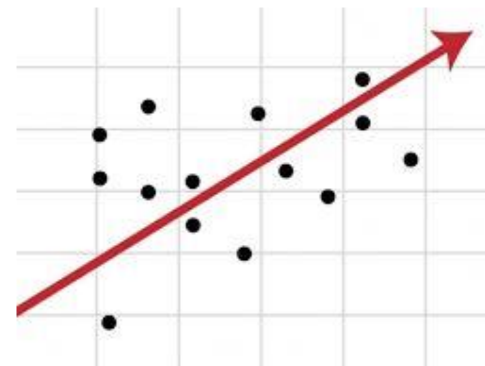
Target vs. Non-Target, such as

- Buyer vs. Non-Buyer
- Responder vs. Non-Responder
- Loyal vs. Attrition
- High Value vs. Low Value

→ Defining target and non-target are equally critical

Regression - Overview

- Powerful Statistical Technique
    - Less Predictive Error
    - Greater Discrimination
- Identifies Combinations of Characteristics that best predicts a specific behavior
- Result - Regression Equation
    - ***Tool to Score & Rank Customers or Prospects***
- Types of Regression Techniques
    - *Logistic Regression* - for Dichotomous Dependent Variable
        - Yes/No Type of Behavior (*e.g.* Response/Non-Response, Buyer/Non-Buyer, etc.)
    - *Multiple Regression* - Dependent Variable is Not Dichotomous
        - Continuous or Ranges (*e.g.* Revenue $ Amount)



Modeling Strategies for Marketing

- Acquire New Customers
  - "Clone" Current/Best Customers
  - Maximize Response/Net Acquisition
- Predict Approval/Payment/Activation
- Increase Revenue/Donation Amount
- Increase Profitability – per Contact or Customer
- CRM Applications
  - Increase Retention
  - Reduce Attrition
  - Predict Switching
- Minimize Risk
  - Bad Debt Prediction
  - Fraud Detection

WILLOW DATA STRATEGY
making sense of data

Sample Regression Equation

Effect	***Weight (Coefficient)***		***Model Variables***	***Variable Description & Recode***	***Data Source***
=	- 5.4821		Constant		
-	0.6630	X	Length of Residence (LOR)	Actual value if LOR ≤ 5 years, if LOR > 5 years then it is set = 5.	*3rd Party* Consumer Database
+	0.2235	X	% HH Income $30,000+	The percentage of HH's in neighborhood earning incomes ≥ $30,000 is recoded by taking the square root of the value.	Census Data
+	0.000865	X	PPI – Purchasing Power Indicator	A statistical estimate of the probable spending power of each HH based on regional cost of living variables. Actual values are used in the equation.	*3rd Party* Consumer Database
+	0.0259	X	% Some College	The % of persons aged 25+ that have completed 1-3 years of college is recoded by setting all values ≥ 50% = 50. If % < 50, actual value used.	Census Data
+	0.6448	X	Presence of Mobile Age Person	If one or more family members are aged 18 - 29 years, then value set = 1, otherwise the value set = 0.	*3rd Party* Consumer Database

WILLOW DATA STRATEGY
making sense of data

Regression – Sample Gains Chart

ACQUISITION MODEL

Development Sample

Decile Group	Count	# New Customers	Acquisition Rate	Acquisition Index	Cumulative Count	Cumulative # New Customers	Cumulative Acquisition Rate	Cumulative Acquisition Index
1	10,049	479	4.77	448	10,049	479	4.77	448
2	10,048	212	2.11	199	20,097	691	3.44	323
3	10,049	127	1.26	119	30,146	818	2.71	255
4	10,047	59	0.59	55	40,193	877	2.18	205
5	10,049	49	0.49	46	50,242	926	1.84	173
6	10,049	37	0.37	35	60,291	963	1.60	150
7	10,048	32	0.32	30	70,339	995	1.41	133
8	10,047	33	0.33	31	80,386	1,028	1.28	120
9	10,048	26	0.26	24	90,434	1,054	1.17	110
10	10,048	14	0.14	13	100,482	1,068	1.06	100

Regression – Sample Efficiency Curves

WILLOW DATA STRATEGY
making sense of data

When Outsourcing Analytics, Consider

1. **Consulting Capability:** Translate marketing goals into mathematics
2. **Data Manipulation:** Standardization, hygiene, categorization, consolidation
3. **Track Record in the Industry**: Not in rocket science, but in marketing
4. **Types of Models Supported**: Watch out for one-trick ponies
5. **Speed of Execution:** Turnaround time measured in days, not weeks
6. **Pricing Structure**: Model development is only one part
7. **Documentation**: Full disclosure of algorithms, gains charts and reports
8. **Scoring Validation**: Job not done until fully scored and validated
9. **Back-end Analysis:** For true "Closed-loop" marketing
10. **Ongoing Support**: Periodic review and update

The Art of Targeting

Targeting is a Business Decision

Art of Targeting

Remember T, C, M

1. Target

2. Comparison Universe

3. Methodology

- **Defining the proper target is most critical, even more than the methodology**
  - Yes, even for Machine Learning
- **Marketers must get involved in Target Definition**
  – State the goals and usages clearly
  – Don't be a bad patient demanding specific prescriptions

"Some targets are not what they seem…"

Start by hanging the target in the right place

WILLOW DATA STRATEGY
making sense of data

Defining the Target (or Targets) (1/4)

- **Continuous Target**
  - How frequent is frequent enough?
  - How much is high enough value?
  - How big is the size of the ideal target?

→ Do not make arbitrary cut-off lines, and follow the real value

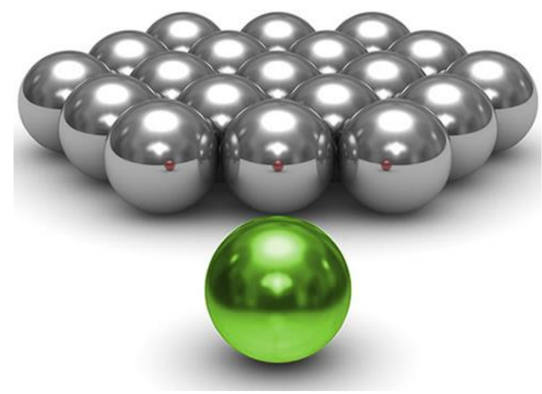
→ Employ Multiple Regression

Defining the Target (or Targets) (2/4)

➢ Multiple Targets

Multiple distinctive segments such as:

- Infrequent Big Spenders
- Frequent Small Spenders
- Introductory Product Buyers
- Loyal Customers
- Recent Buyers
- Dormant Customers
- Geographic Targets (e.g., regions, cities, etc.)
- Demographic Segments (e.g., Millennials, Young & Upcoming, Soccer Moms, Middle-age Blues, Empty-nesters, Golden Years, etc.)



Defining the Target (or Targets) (3/4)

- **Target within a Target**
  - Multi-step approach for multi-step sales/marketing
    - B-to-B Sales pipeline (various stages of lead qualification, sales, and conversion)
    - Open, Click, Browse, Convert, Repeat Purchase cycle
  - Very narrow target in a big universe
  - Sub-targets within major segments

Defining the Target (or Targets) (4/4)

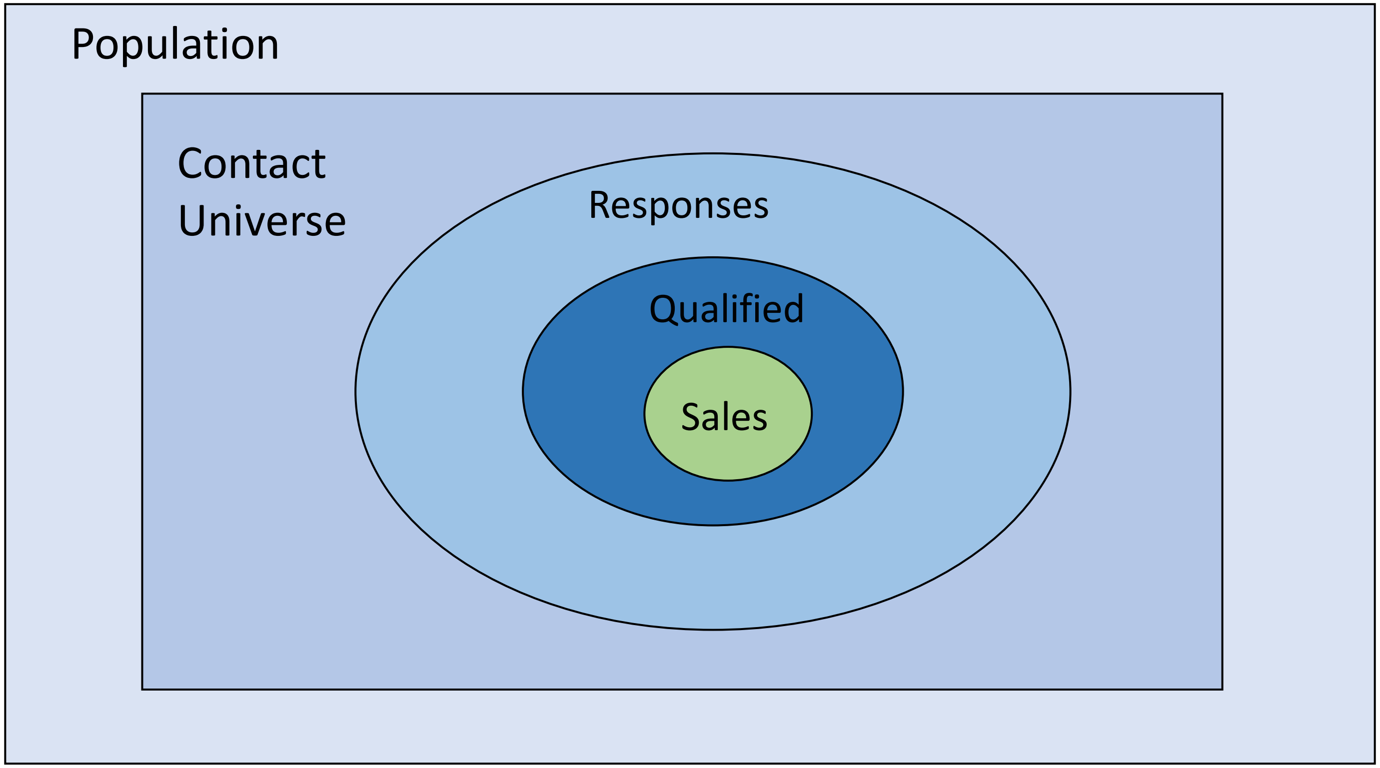
➢ Inversely Related Targets

For example,

- Frequent shoppers with low average spending
- Responsive prospects with bad credit

→ Build multiple models and find cross-sections

Multiple Model Approach Illustrated



More than 1 model may be necessary, if there is any inverse relationship among subsets

For example:

- Response Model: Target=Responders Comparison=Non-Responders
- Qualification Model: Target=Qualified Comparison=Non-Qualified among Responders

Sample Double Model Matrix

Response x Revenue										
Response Decile	**Revenue Model Decile**									
	1	2	3	4	5	6	7	8	9	10
1										
2	High Response, High Revenue							High Response, Low Revenue		
3										
4										
5			Average Response & Revenue							
6										
7										
8	Low Response, High Revenue							Low Response, Low Revenue		
9										
10										

- High-High is an obvious target, and Low-Low may be ignored. The rest depends on the targeting and messaging strategies.

WILLOW DATA STRATEGY
making sense of data

Analytics-Ready Environment

Analytics Sandbox for Speed and Accuracy

Any Pain Implementing Models?

- Modelers are fixing data all the time
- Repeatedly rely on a few popular variables
- Always need more variables
- Takes too long to build models and deploy them
- Inconsistencies shown when scored
- Disappointing results, even with machine learning

Most troubles happen <u>before</u> or <u>after</u> the modeling process.

What does your database support?

You must have databases for…

- Order Fulfillment
- Contact Management
- Standard Reports / Dashboards
- Ad hoc Reports and Queries
- Campaign Execution
- Response Analysis
- Trend Analysis

But does any of them support predictive modeling and scoring?

Predictive Modeling is all about "Ranking"

Ultimately, models must properly "Rank"

- Households
- Individuals
- Companies
- Email Addresses
- Products

Define the level of data accordingly

- Relational or unstructured databases won't cut it
- Must create "Descriptors" that fit the level that needs to be ranked

WILLOW DATA STRATEGY
making sense of data

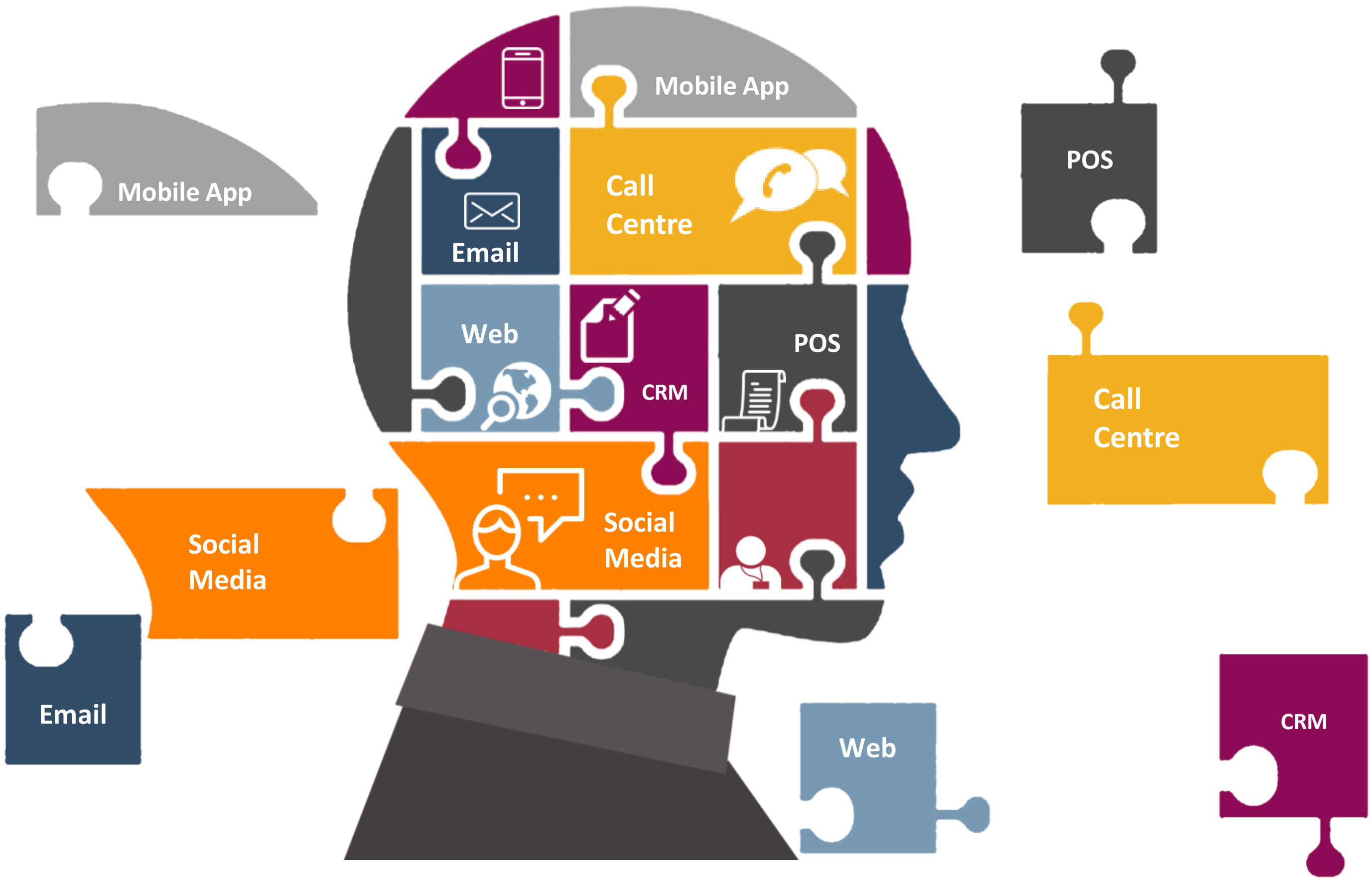
For Analytics, Clean the Data First

"Garbage-in, garbage-out"

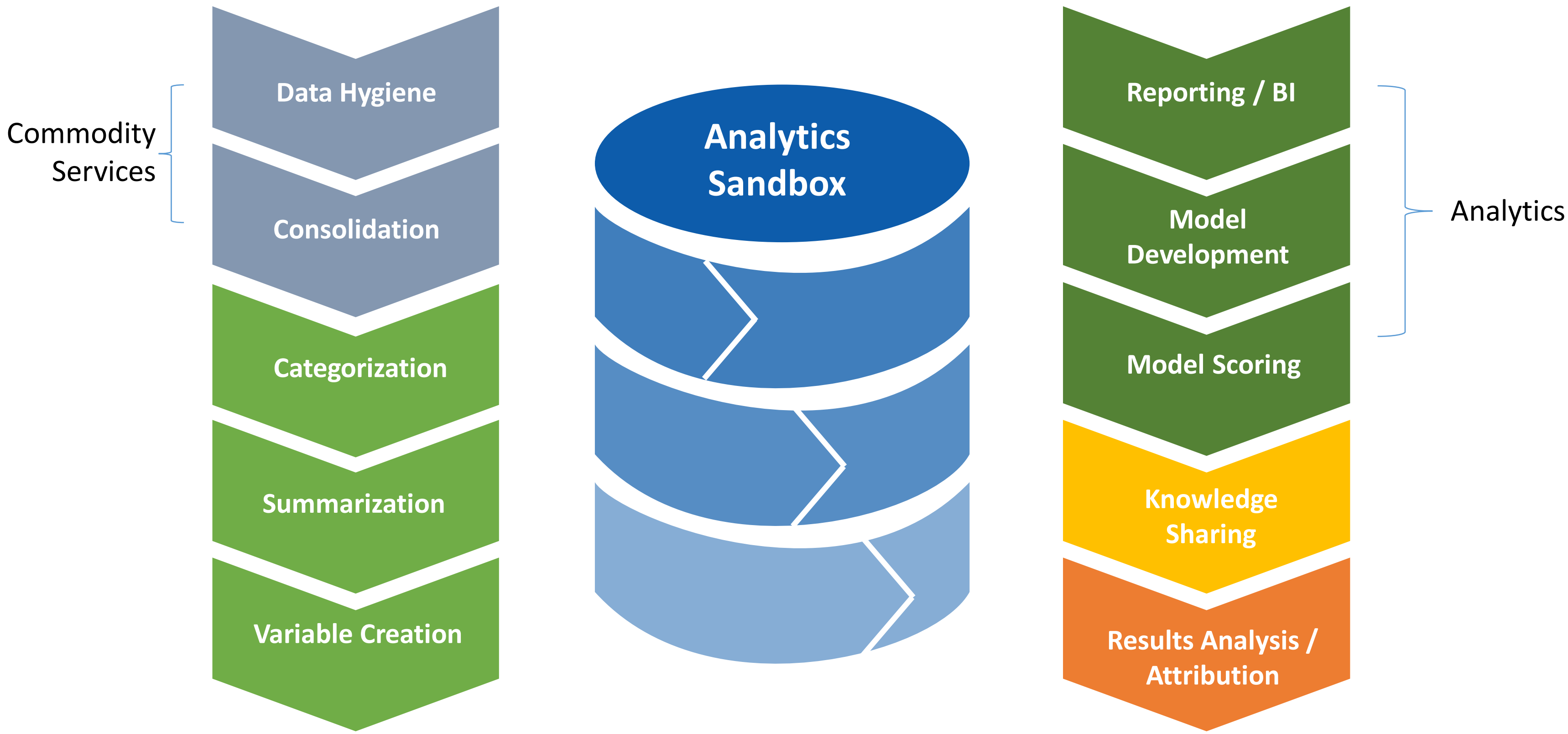
- Most data sets are "unrefined" and "unstructured"
- Over 80% of model development time goes to data prep work
    - Most databases are NOT analytics-ready
- Modeling & Scoring
    - Extension of database work
    - Consistency is "the" key

WILLOW DATA STRATEGY
making sense of data

Single View of Customer for CRM



Holistic View of Analytics-Readiness



More than 80% of analytics work goes into data refinement

Why Front-end Data Work Important?

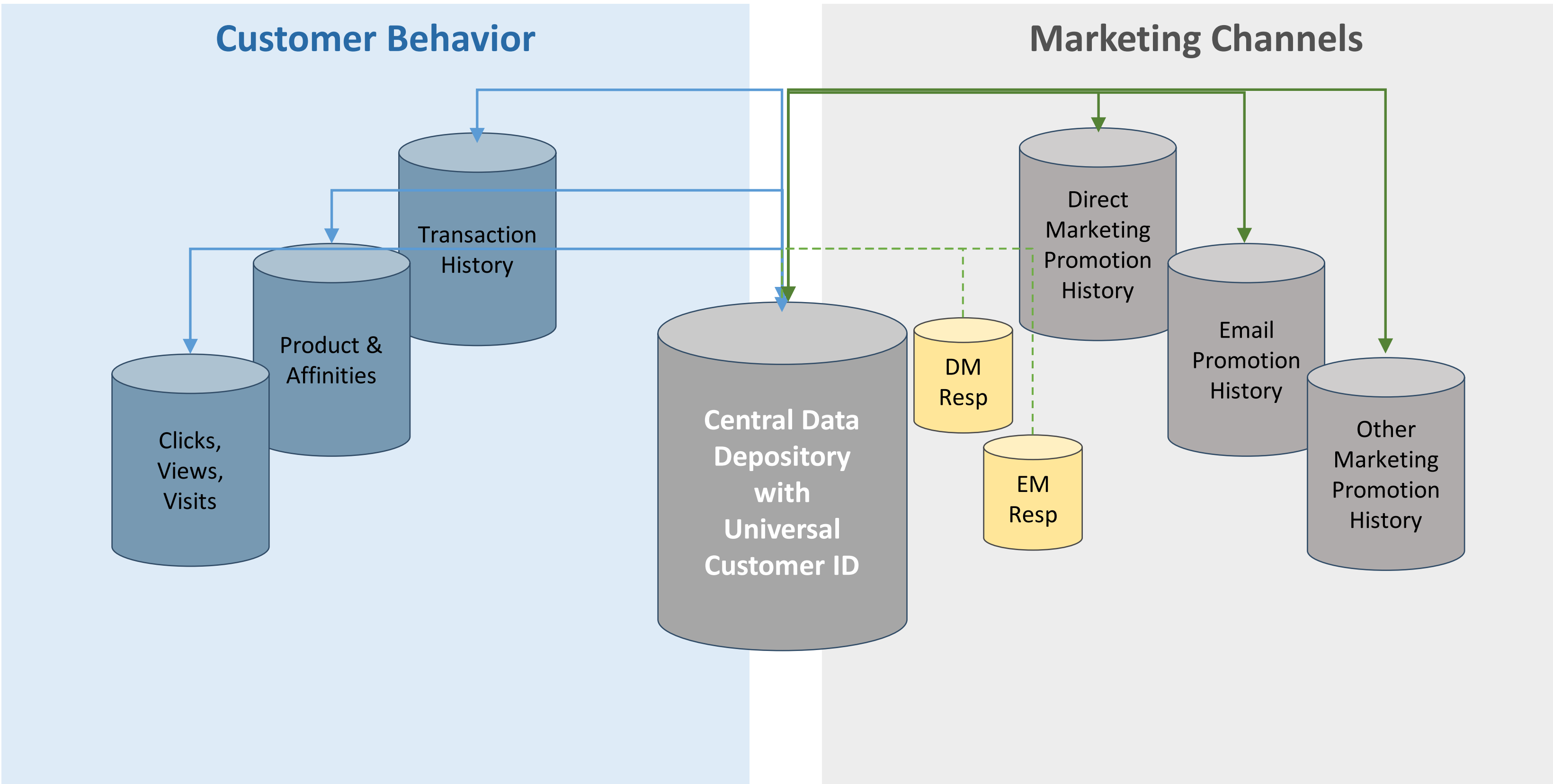
Many analysts spend majority of time doing the data work
→ Modeling work at the last minute!

Creative “intelligent” variables enhance model performance
→ Even for machines

Inconsistent data create chain reactions to melt-downs
→ Most trouble happens before and after modeling steps

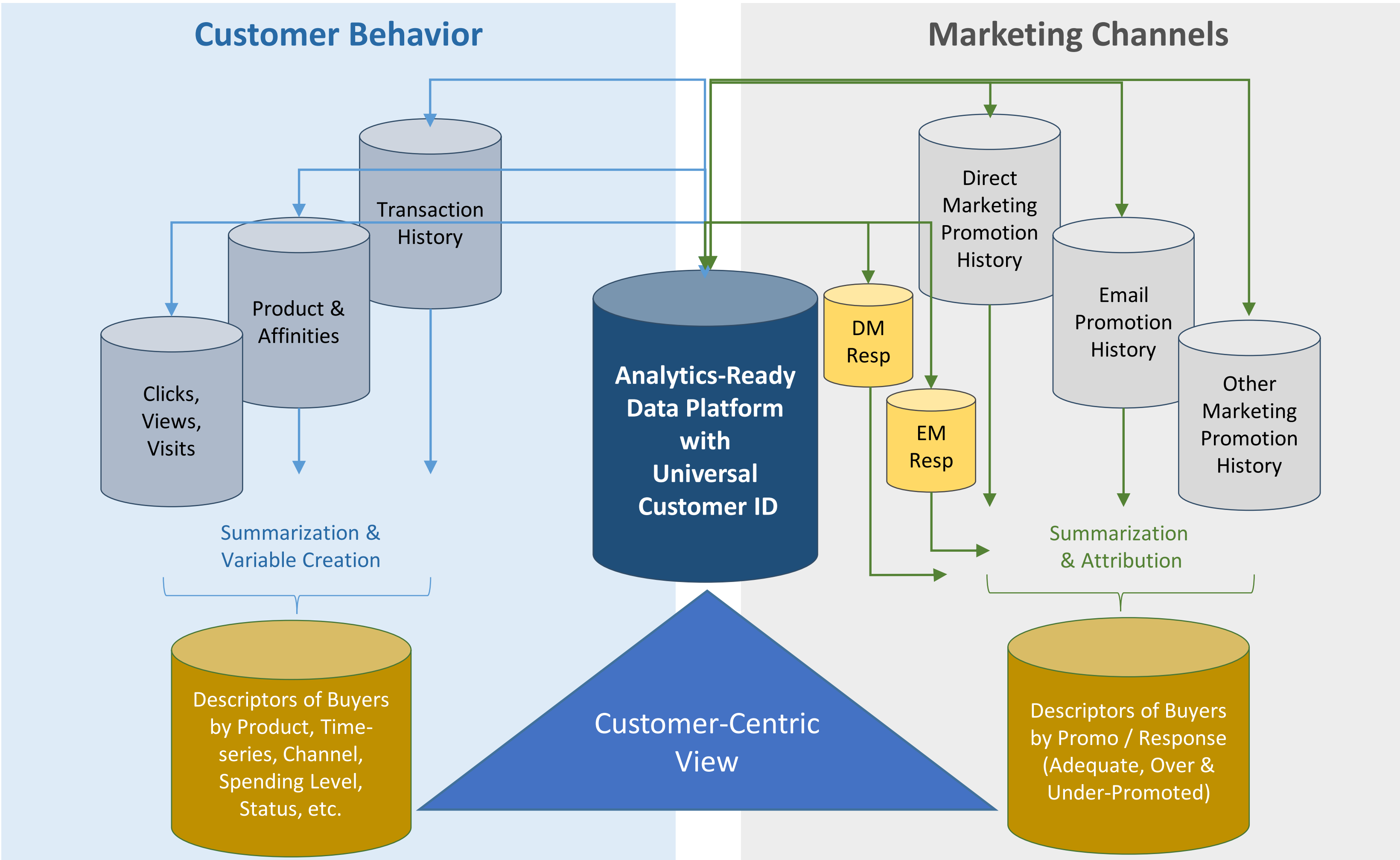
Data match and consolidation become ineffective
→ Consistent target identification and tracking is the key

Basic CDP Data Modeling



! Datasets are in one place, but the work isn't nearly done

Analytics-Ready CDP Data Modeling



WILLOW DATA STRATEGY
making sense of data

What CDPs Must be

- Cover omnichannel sources
- Connected on an individual level
- Clean and reliable
- Up-to-date
- "Analytics-Ready"
- Campaign-ready
- Designed to support omnichannel attribution

Holistic data journey
from Data Hygiene and Consolidation
to Model-Based Targeting and Attribution

Beware of One-Size-Fits-All CDP Solutions

Your business goals and challenges are unique:

- Don't forget the basics
  - Clean PII is a-must, but the definition of "clean" varies
- Match sensitivity must be customized
  - Tight, loose, or medium?
  - Combinations of Name, Address, Email, Phone, and Digital IDs
- Transaction data are unique to your business
  - Who, what, when, how much, through what channel
  - Unique to industries (e.g., subscription model)
- Behavioral data are complex and diverse
  - Online and Offline (i.e., Digital vs. POS)
  - Heavy dependency on employed toolsets

"Exception Management" is the key

Benefits of Analytics Sandbox

Analytics-Ready

Refined datasets with analytics-ready variables

Hundreds of Intelligent variables maximizing predictive power

Zero effort for data hygiene in analytics stages

360-degree View

Summary of all activities around the individual target

Variables as descriptors of customers, not transactions or events

Consistent metrics for casual users and advanced analytics

Speed

Over 10x speed gain in deployment of analytics

Expedited data-based decision making, campaign support and attribution

Ready for customer-centric personalization in real-time

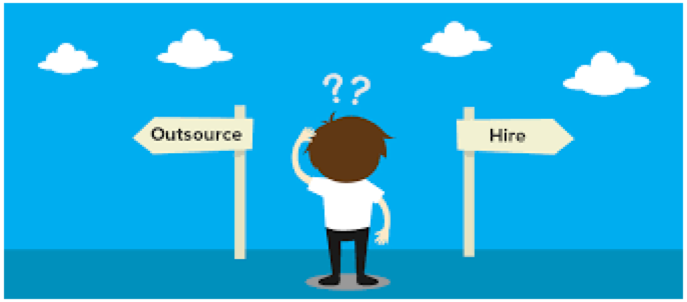
Where to Begin with Analytics Sandbox

Spec it out:

- Project Goals
- Data Source List (as detailed as possible)
- Final Variable List (for the analysts)
- Project Flow:
  - Data Collection
  - Conversion / Standardization
  - Categorization
  - Consolidation / Summarization
  - Variable Creation
  - External Data Append
  - Sampling
  - Modeling
  - Scoring
  - Storage / Sharing
  - Application / Campaign
  - Backend Analysis / Attribution

Who Will Build the Sandbox?

- In-house vs. Outsourcing – Consider:
  - Platform
  - Software
  - Programming
  - Staffing – Never the analysts!
- Cost it out
  - Don't forget the Update & Maintenance Cost
- Involve Analysts for Variable List Review
- **Don't be shy and ask for help from specialists or consultants**



Scope It Out

- Don't try to boil the ocean right away
  → "Analytics is making the most of available data"

- Take a phased approach
  - Watch your budget, and start with low hanging fruits
    (Hint: Transaction data are the most powerful predictors)
  - Phased approach: **Proof of Concept** to full commitment in steps
  - **Think Long-term** – Avoid redesigns and redevelopments
  - Maintain **Consistency**
  - Keep the Historical Data

Data Refinement & Management

Going beyond Simple Event & Transaction Data

3 Major Types of Data for Marketing

Descriptive Data

- Demographic Data
- Firmographic Data
- Geo-demographic Data

(All from 3rd-Party Sources)

Behavioral Data

- Transaction Data
- Online Behavior Data
- Movement Data
- Lifestyle Data

Attitudinal Data

- Surveys
- Primary Research
- Sentiments (via Social Media, etc.)

3-Dimensions in Predictive Analytics

Data Source Evaluation Criteria

1. **Depth**: Content & uniqueness of data
2. **Width**: Data coverage
3. **Accuracy**: Free of errors or false positives
4. **Recency**: Fresh data / minimal data atrophy
5. **Consistency**: Matters more than sheer accuracy
6. **Connectivity**: To other data sources/systems
7. **Delivery Mechanism**: Query, drilldown, visualization
8. **User-friendliness**: Intuitive & meaningful data values
9. **Cost**: Development & maintenance costs

Not all data sources are created equal

Create Data Menu

- Base it on Companywide Need-Analysis
- Ask the Users & Analysts first:

  What types of predictions are required?

  - Affinity/Look-alike Models
  - Promotion/Response Models
  - Time-series Models
  - Attrition Models
  - Etc., etc.,…
- Consider non-analytics departments and users
- Maintain only the ones that fit the objective

→ Don't be afraid to throw out "noises"

Data Menu

Check the ingredients

- What do you have today?
- What can be bought?
- What can be created?

Cost - *Can you afford to maintain it?*

- Storage/Platform – Consider the scoring part, too
- Programming/Processing Time
- Software
- **Update/Maintenance**
- External Data

Organizational Data Inventory

You may have more than you thought:

- **PII** (Personally Identifiable Data): Name, Address, Email, Phone Number, etc. Key to Geo/Demographic Data
- **Order/Transaction Data**: "RFM", Payment Methods
- **Item/SKU Level Data**: Products, Price, Units
- **Promotion/Response History**: Source, Channel, Offer
- Life-to-Date/Past "x" Months **Summary Data**
- **Customer Status** Flags: Active, Dormant, Delinquent
- **Surveys**/Product Registration: Attitudinal/Lifestyle
- **Customer Communication** History Data: Call-center, Web
- **Online Behavior**: Open, Click-through, Page views, etc.
- **Social Media**: Sentiment/Intentions

Need Standardization, Categorization, & Summarization

Maximize the Power of Transaction Data

Past Behavior is the Best Predictor of the Future Behavior

Most databases describe shopping baskets

→ Start describing your targets

- 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.)

Data Summarization – Matching the Level of Data

Transactions

Cust ID	Trans ID	Order Date	$ Amount
000123	100011	2016-05-06	$199.99
000123	100128	2017-08-30	$50.49
000123	103082	2018-12-21	$128.60
003859	100036	2017-06-06	$43.99
003859	101658	2018-01-20	$43.99
003859	102189	2018-04-15	$119.45
003859	106458	2019-02-18	$43.99
004593	104535	2019-04-30	$354.72
016899	107296	2018-07-14	$199.99
019872	102982	2017-09-07	$128.60
019872	103826	2018-04-30	$499.99
019872	109056	2019-03-12	$59.99

Individual Level Summary

Cust ID	# Orders	$ Total	First Order Date	Last Order Date
000123	3	$379.08	2016-05-06	2018-12-21
003859	4	$251.42	2017-06-06	2019-02-18
004593	1	$354.72	2019-04-30	2014-04-30
016899	1	$199.99	2018-07-14	2018-07-14
019872	3	$688.58	2017-09-07	2019-03-12

WILLOW DATA STRATEGY
making sense of data

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)

Raw Data Coming into Summarization

Merchandizing

- Who: Customer ID / PII
- What: Product SKU / Category
- When: Purchase Date
- How much: Total Paid, Net Price, Discount/Coupon, Tax, Shipping, Return
- Channel/Device: Store, Web, App, etc.
- Payment Method

Promotion Data

- Promotion Channel
- Source of Data / List
- Offer Type
- Creative Details
- Segment/Model (for selection/targeting)

Subscription

- Who: Subscriber ID / PII
- Brand / Title / Property
- Dates: First Subscription, Renewal, Payment, Delinquent, Cancelation, Reactivation, etc.
- Paid Amounts by Pay Period
- Number of Payments / Turns
- Payment Method
- Auto Payment Status
- Subscription Status
- Number of Renewals
- Subscription Terms
- Acquisition Channel / Device
- Acquisition Source

Hotel/Hospitality

- Who: Guest ID / PII
- Brand / Property
- Region
- Booking Site / Source
- Transaction Channel / Device
- Booking Date/Time/Day of Week
- Travel(Arrival) Date/Time
- Travel Duration
- Transaction Amount: Total Paid, Net Price, Discount, Coupon, Fees, Taxes
- Number of Rooms/Parties
- Room Class/Price Band
- Payment Method
- Corporate Discount Code
- Special Requests

Sample Variables after Summarization

Recency

- Weeks since first transaction date
- Weeks since last transaction date
- Years since first sign-up
- Last transaction channel
- Average number of days between transactions
- Transactions done in past 12, 24, 36 months

Frequency

- Number of transactions
- Number of items purchased
- Average transactions by month
- LTD returns and adjustments
- Transactions by payment method
- Transactions by dollar range [$1- $49.99], [$50- $99.99] ..
- # Full-price transactions

Monetary

- Life-to-date $ spending
- Total $ past 12, 24, 48 months
- Average dollars by channel
- Average dollars by payment method
- Highest transaction $ amount
- Lowest transaction $ amount
- Last transaction $ amount

Brand Category

- Transactions made by brand
- Total $ past 48 months by brand
- Average dollars by brand
- Weeks since first & last transaction date by brand

Product Category

- Transactions made by product category
- Total $ past 48 months by category
- Average dollars by product category
- Weeks since first & last transaction date by product category

Data Categorization & Tagging

Freeform data comes to life through categorization

- Hidden data in:
  - Product, Service, Offer, Channel, Source, Status, Titles, Surveys, etc.
- Have categorization guideline?
- Who will do it?
  - **Consider Machine Learning**
- What to throw out?
  - Keep data mainly for analytics and predictive modeling

Categorical Data

Any Non-numeric Data

- Product
- Service
- **Offer**
- Transaction Channel
- Promotion Channel
- Source
- Market
- Region
- Business Title
- Member Status
- Payment Method
- etc…

Offer Code

For example:

- Flat Dollar Discount
- % Discount
- Buy 1, Get 1 Free
- Free Shipping
- No Payment Until…
- Free Gift
- Discount Coupon
- Friends & Family Discount
- etc…

→ Create standard codes

Categorize as much as possible at the <u>data collection stage</u>

Categorization Guidelines

- Define the key categories first
- Categorize during data collection
- Categorize <u>buyers</u>, not products
  - Not the products, but the intentions
  - Products change, but affinities rarely do
- More specific the better
  - Don’t settle with general categories
  - Be as specific as possible (e.g., “Home Theater Speakers”, not just “Home Electronics”)
- <u>Consistency</u> over accuracy
- Automate as much as possible → Via Machine Learning

Categorization Guidelines *(continued)*

Create rules and DON'T deviate from them	➡	Create "Code" Structure Training & Automation
Be consistent throughout	➡	Surveys, Data Entry, Product Taxonomy, Database, Analytics
Don't allow too many variations (over 20) in one category	➡	Break into multiple categorical variables if necessary
Don't forget the end goals and don't over-do it	➡	Must be "relevant"

Data Hygiene & QC

Maintain accuracy and consistency

- Create consistency
  - Consistency via Standardization
  - Proper data conversion & edit
  - Purge bad data mercilessly
- Cover all basis
  - PII (Name, Address, Email, Phone)
  - Transaction data (#, $, Dates)
  - Categorical data (everything else)
- Develop firm rules and enforce them
  - Data players – All of them
  - Everyone who touches data (e.g., retail stores)

Missing Values

"Missing Data can be meaningful"

Showing up as predictors in models

- For Numeric Data (e.g., $, Counters, Dates, etc.)
  - Incalculable vs. Non-matches when joined
  - Missing is missing: <u>DO NOT</u> 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

Caution! Some "unstructured" non-relational databases only store what's available.

More on Missing Data

- Agree on Imputation Rules
  - Across divisions and user groups
  - Do it upfront
  - Must be part of scoring codes
- Educate non-analysts
  - Hard to undo when combined with other values
  - Train data partners and vendors
- Always check % missing
  - Development Samples vs. Main Databases

WILLOW DATA STRATEGY
making sense of data

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

Watch out for Privacy Laws such as GDPR, when exchanging PII

Use Cases for 3rd-Party Demographic Data

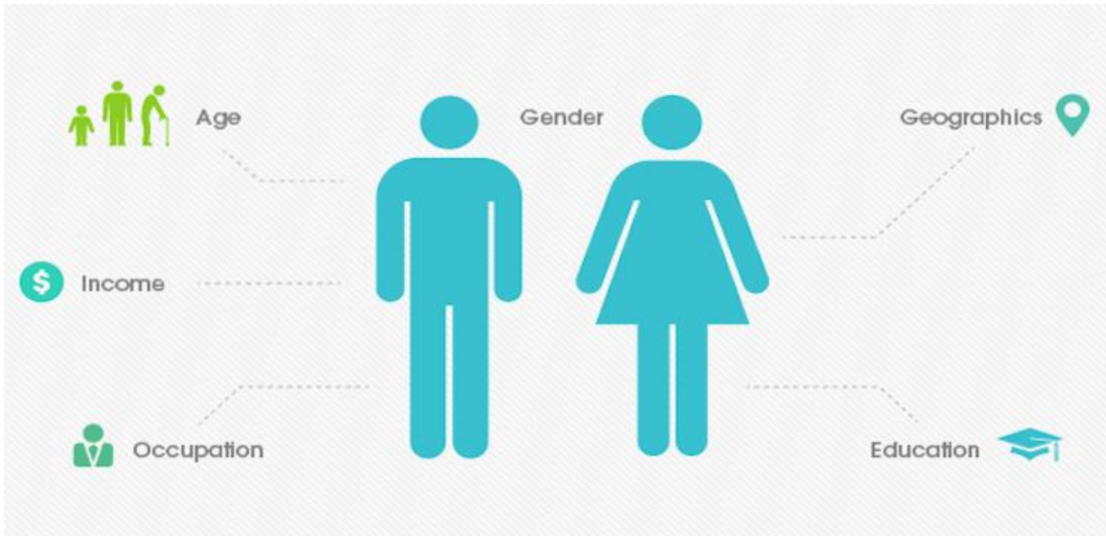
Data Append

- Add extra dimensions to existing customer data for:
  - Customer Profiling – with views against normative data (U.S. sample)
  - Modeling & Segmentation – for cross-sell/upsell, CLV, loyalty, etc.
  - Bring consistency factors to advanced analytics
  - Filtering based on key variables such as age, income, family type, etc.

Prospecting

- Create new customer base using all available data sources
  - Define various "Best Customer" types using data assets in CDP
  - 3rd-party vendor builds "look-alike" models using the Best Customers as targets
  - Have the vendor deploy email campaigns and/or produce contact lists for targeted campaigns

External Data



Always consider buying data before collecting and building

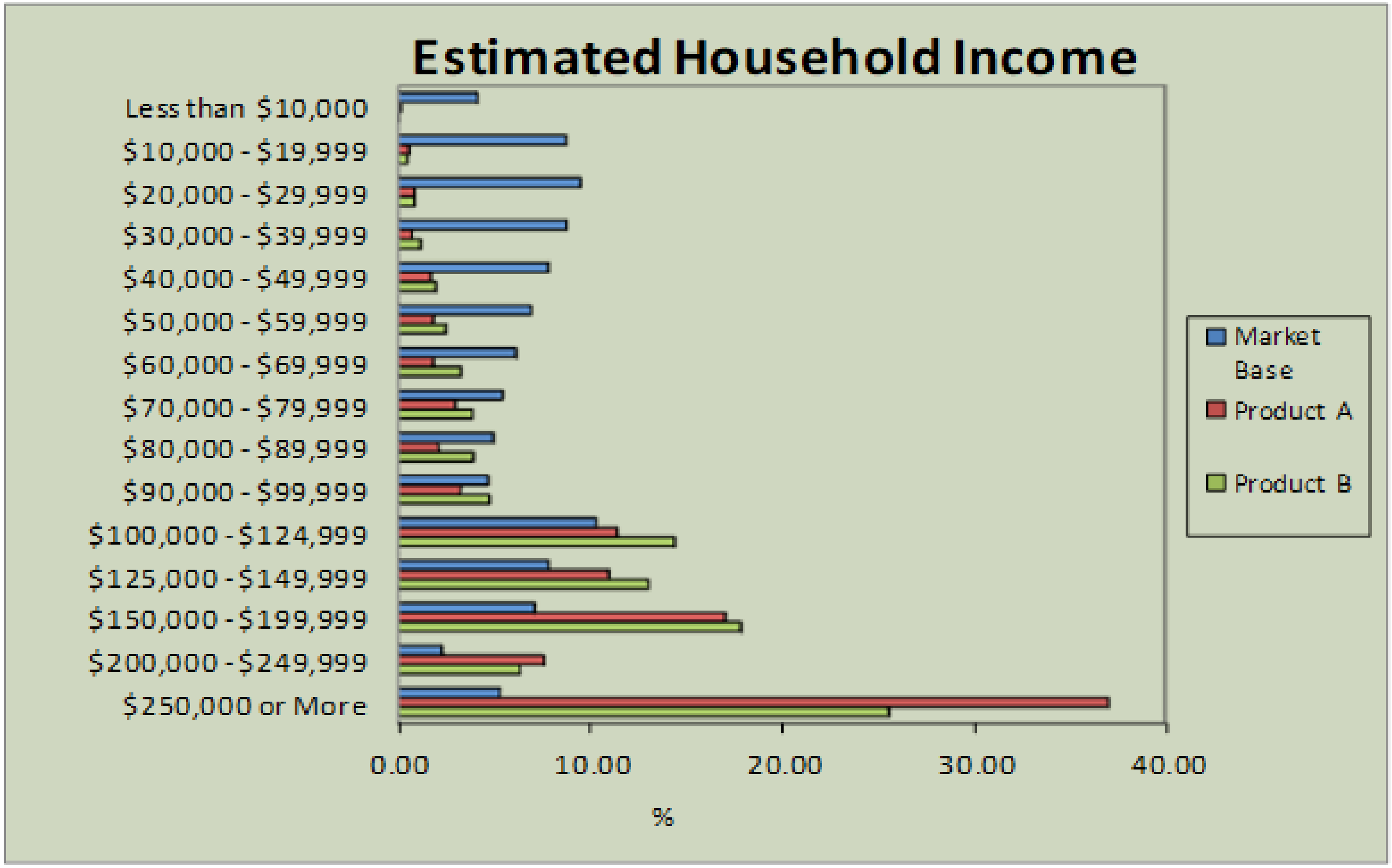
- Compiled Demographic / Firmographic Data
- Behavioral / Transaction / Co-op Data
- Lifestyle / Attitudinal / Survey
- Census / Geo-Demographic Data

Evaluating External Data

- Test multiple data sources
    1. Depth of information / Uniqueness
    2. Coverage / Match Rate
    3. Consistency / Update Cycle
    4. Accuracy
    5. Price
    6. User-friendliness
    7. Delivery Options – Real-time via API

- Learn about the data sources
    - What's real and what's imputed?
    - Don't stop at Demographic: Consider "Behavioral" data

Sample Demographic Profile Report



WILLOW DATA STRATEGY
making sense of data

Sample Multi-Column Profile Report

		Customers				Prospects					
	Market Base	Post 2002		Pre 2002		Lost to Competition		Still Interested		Special Project	
	%	%	Index	%	Index	%	Index	%	Index	%	Index
ESTIMATED HOUSEHOLD INCOME											
Less Than $30,000	10.40	1.06	10	0.83	8	1.42	14	1.69	16	0.79	8
$30,000 - $39,999	10.12	1.07	11	0.82	8	0.91	9	1.61	16	2.38	24
$40,000 - $49,999	8.70	1.53	18	1.36	16	1.04	12	1.61	19	3.17	36
$50,000 - $59,999	8.00	2.02	25	1.78	22	2.85	36	2.55	32	1.59	20
$60,000 - $69,999	7.36	2.34	32	2.28	31	2.72	37	3.23	44	4.76	65
$70,000 - $79,999	6.68	2.98	45	2.67	40	2.72	41	4.50	67	4.76	71
$80,000 - $89,999	6.41	3.57	56	3.40	53	3.17	49	4.33	68	4.76	74
$90,000 - $99,999	5.91	3.93	66	3.85	65	4.08	69	5.27	89	8.73	148
$100,000 - $124,999	11.86	11.44	96	11.18	94	12.89	109	12.91	109	26.98	227
$125,000 - $149,999	8.40	12.25	146	11.81	141	13.47	160	14.87	177	14.29	170
$150,000 - $199,999	7.24	15.89	219	13.46	186	20.34	281	17.76	245	19.84	274
$200,000 - $249,999	4.45	13.73	309	15.10	339	13.08	294	13.17	296	5.56	125
$250,000 or More	4.46	28.21	633	31.49	706	21.31	478	16.48	370	2.38	53
PURCHASING POWER INDICATOR (PPI)											
Less than $20,000	6.44	0.68	11	0.58	9	1.03	16	1.10	17	0.79	12
$20,000 - $29,999	13.26	1.17	9	0.86	6	1.10	8	2.04	15	2.38	18
$30,000 - $39,999	11.16	2.05	18	1.68	15	1.88	17	2.12	19	3.17	28
$40,000 - $49,999	10.11	2.58	26	2.52	25	3.37	33	3.06	30	5.56	55
$50,000 - $59,999	8.95	3.75	42	3.26	36	3.69	41	4.50	50	3.97	44
$60,000 - $69,999	8.44	4.61	55	4.45	53	4.40	52	7.14	85	5.56	66
$70,000 - $79,999	7.37	5.42	74	5.10	69	5.70	77	6.37	86	14.29	194
$80,000 - $89,999	6.42	6.07	95	5.93	92	6.80	106	7.31	114	14.29	223
$90,000 - $99,999	5.45	6.55	120	6.32	116	7.19	132	7.65	140	9.52	175
$100,000 - $124,999	9.56	15.94	167	15.94	167	17.49	183	17.08	179	18.25	191
$125,000 - $149,999	5.69	13.93	245	14.44	254	15.87	279	15.12	266	16.67	293
$150,000 - $199,999	4.81	18.11	377	19.63	408	18.52	385	16.06	334	3.97	83
$200,000 - $249,999	1.33	9.08	683	9.56	719	7.12	535	5.52	415	0.79	59
$250,000 or More	1.00	10.07	1007	9.73	973	5.83	583	4.93	493	0.79	79

WILLOW
DATA STRATEGY
making sense of data

Model Score Quality Control

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
- <u>Compare to previous runs</u> – Check Deterioration

Set parameters for acceptable differences and Enforce them

Scoring – Sample vs. Database

- Development Sample vs. Live Database
  - Database Structure
  - Variable List/Names
  - Variable Values
  - Imputation Assumptions

Lead to disasters if "anything" is different

- **Never** alter model group definitions that are set in the development samples

Share the Model Scores

Model scores are summaries of large and complex data in a compact form

- Sync model scores with other databases and data-marts
- Plan ahead:
  - Reserve spaces
  - Educate the users & *Evangelize*
- Store raw scores, not just model groups
- Match the Levels of Scores
  - Household
  - Individual
  - Company
  - Email
  - Product

Personalization

Road to Holistic Personalization

About Personalization

Personalization is the big buzzword now, but what does it mean to you?

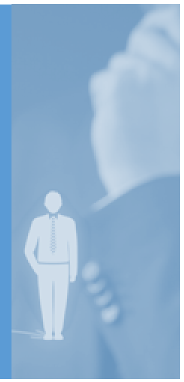
- Addressing your customers by their first names?
- Collecting explicitly expressed preferences and treating them accordingly?
- Customizing emails and landing pages based on customer profiles and behaviors?
- Knowing when to contact, and through what channel?
- Keeping in touch with your customers constantly?
- Suggesting more of the same products that they just purchased via collaborative filtering?

Some efforts are better than mindless batching and blasting, but…

- Maybe you are:
  - "Personally" annoying your customers and prospects
  - Personalizing a fraction of your target base, while completely ignoring others
  - Personalizing sporadically only when obvious trigger data becomes available
  - Mismatching offers to targets customers more often than not

Personalization is about the Person

Look at it from the customer's point of view…



- No one is just an "online" or an "offline" person
- "Personalization Engines" are often overrated (especially when product level collaborative filtering is on "auto-drive" mode)
- Raw SKU level data sets are utterly inadequate for personalization

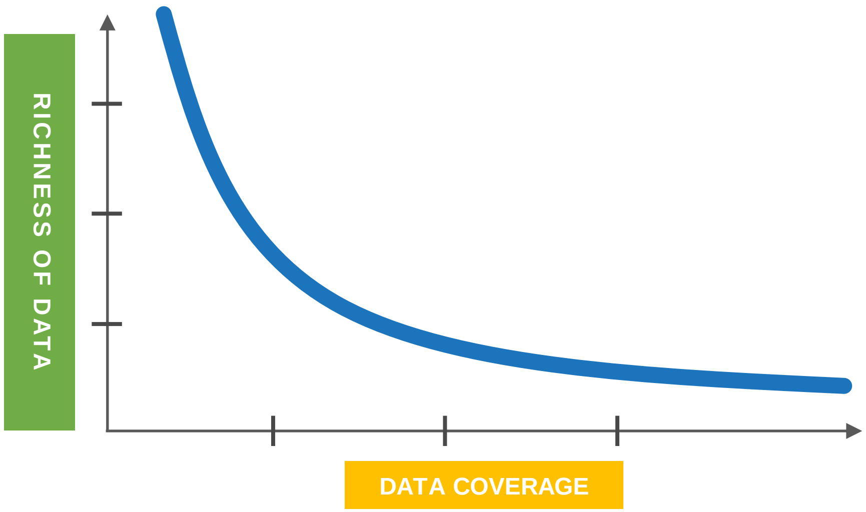
"Personalization is about the Person"

- Transform transaction, event, or product level data to:
  - Describe people, not products
  - Create a 360-degree Single Customer View ("Analytics Sandbox")
- <u>Develop "Personas", then match products to them</u>
  - Not the other way around
  - Fill in the gaps with modeling

Even now, real data are hard to come by

Verified "known" explicit data can be scarce, even in this age.

- Data sets are often missing for targets who are:
  - New to the business
  - Dormant
  - New to specific channels
  - Hiding their tracks

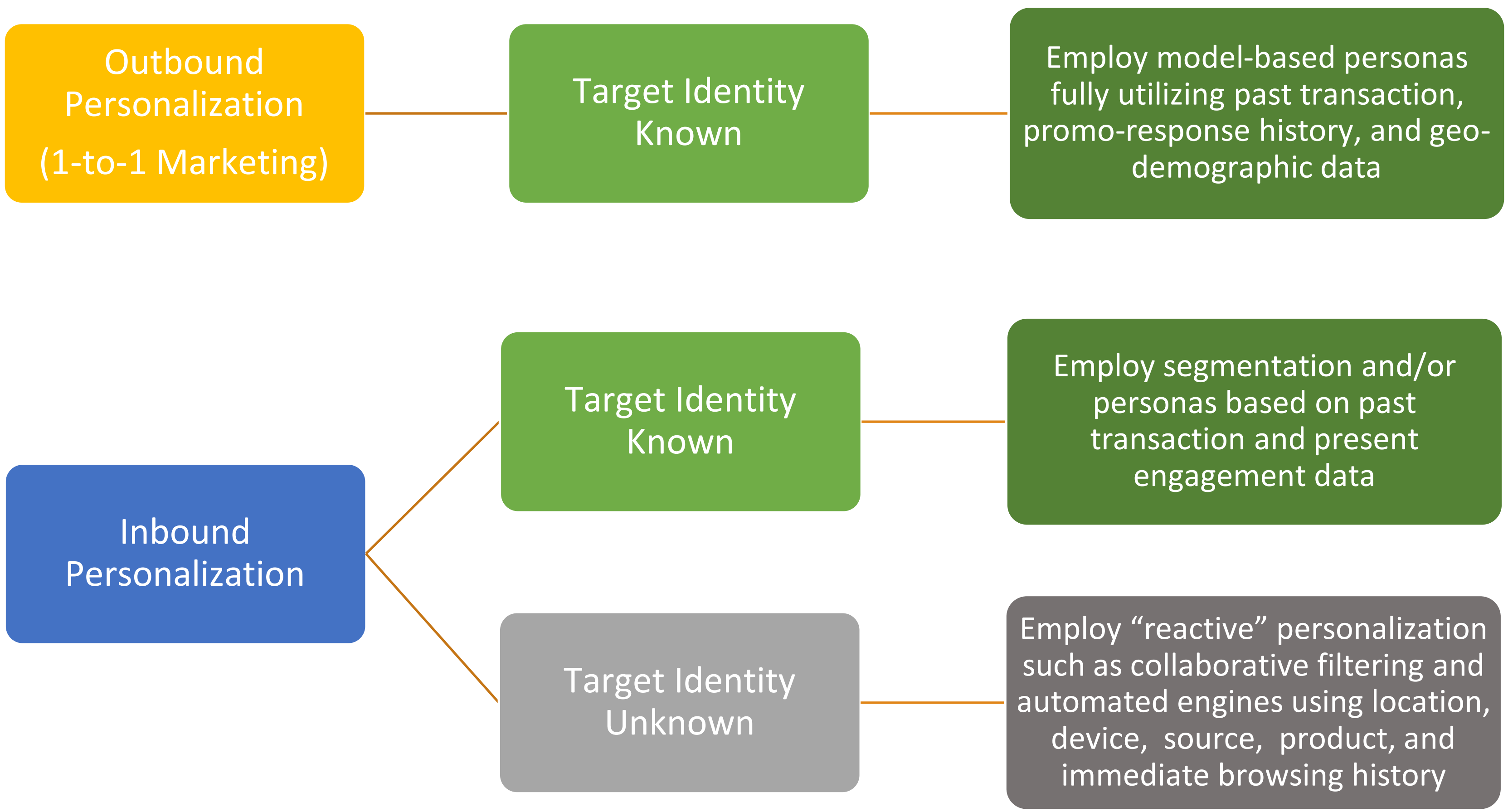


But, most personalization efforts are done based only on "known" explicit data!

- Need to maximize the value of available data, even implicit or anonymous data

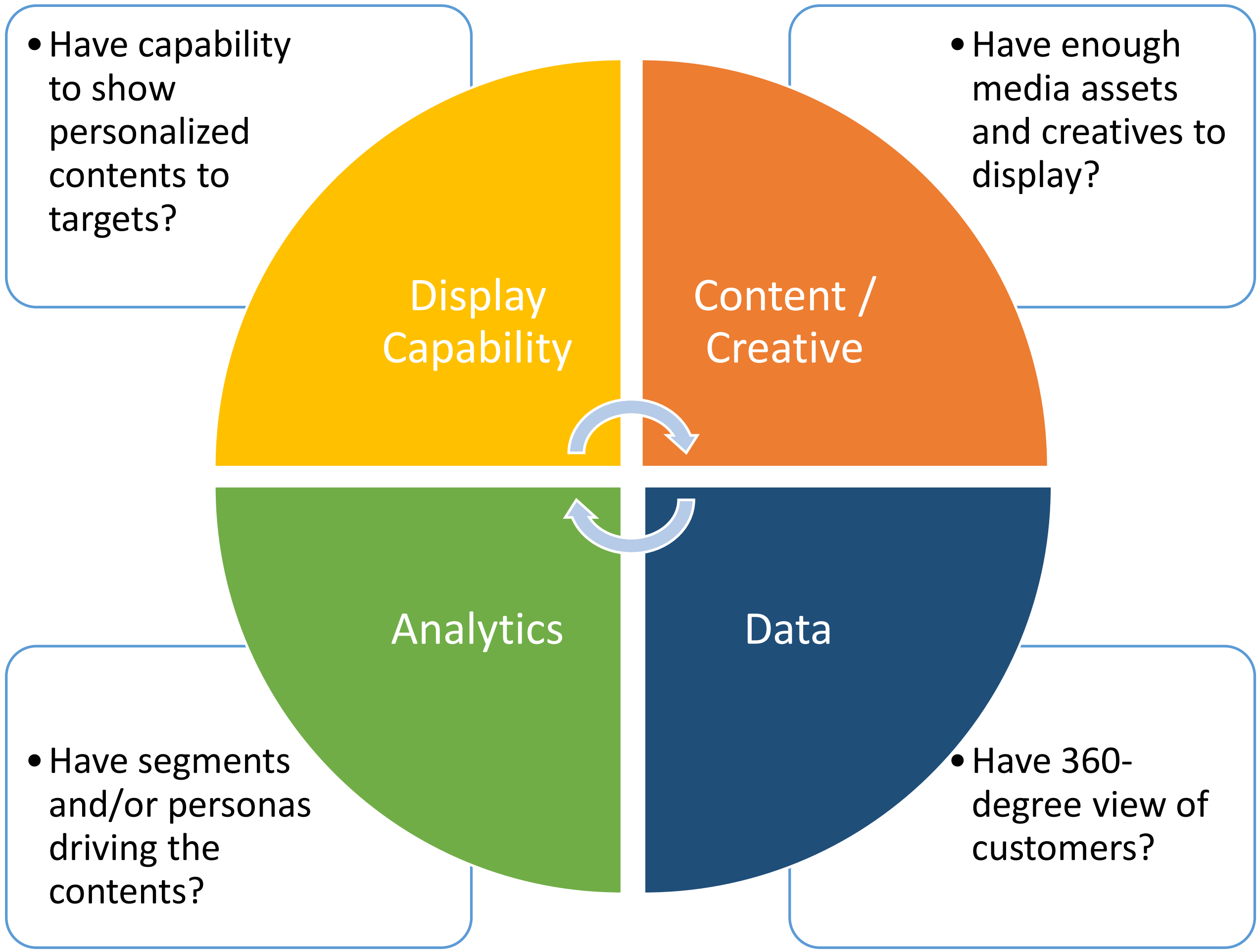
Personalization Framework

Differentiate "Planned" and "Reactionary" Personalization Efforts



Elements of Holistic Personalization

<u>Personalization is about connecting the dots among these elements</u>



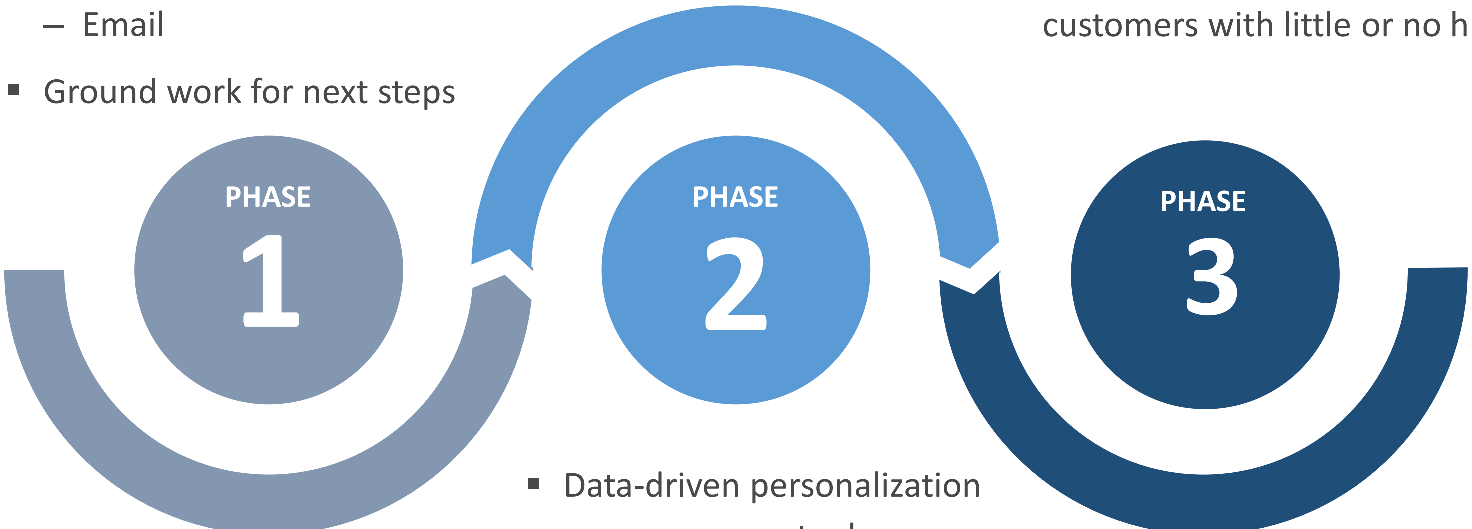
WILLOW DATA STRATEGY
making sense of data

Step-by-Step Approach to Personalization

- Deployment of interactive display capabilities
  - Web
  - Email
- Ground work for next steps

TEST THE ENGINE WITH SIMPLE SEGMENTS

- Employ model-based “personas”
- Convert “unknown” to “potentials”
- Extend personalization to customers with little or no history



- Data-driven personalization - secure access to deeper data
- Employ all available “known” explicit data

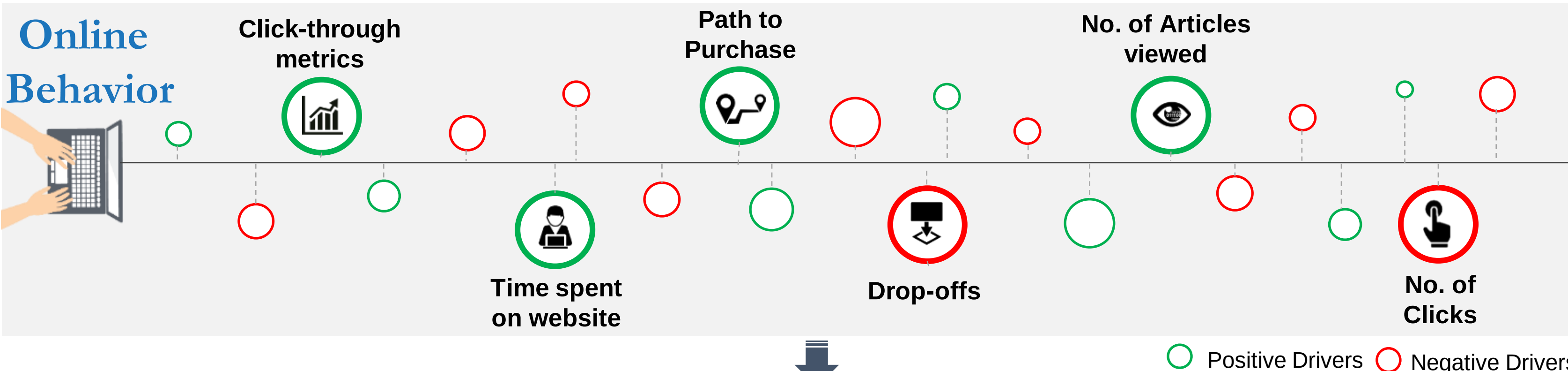
TARGETED CONTENT DISPLAY

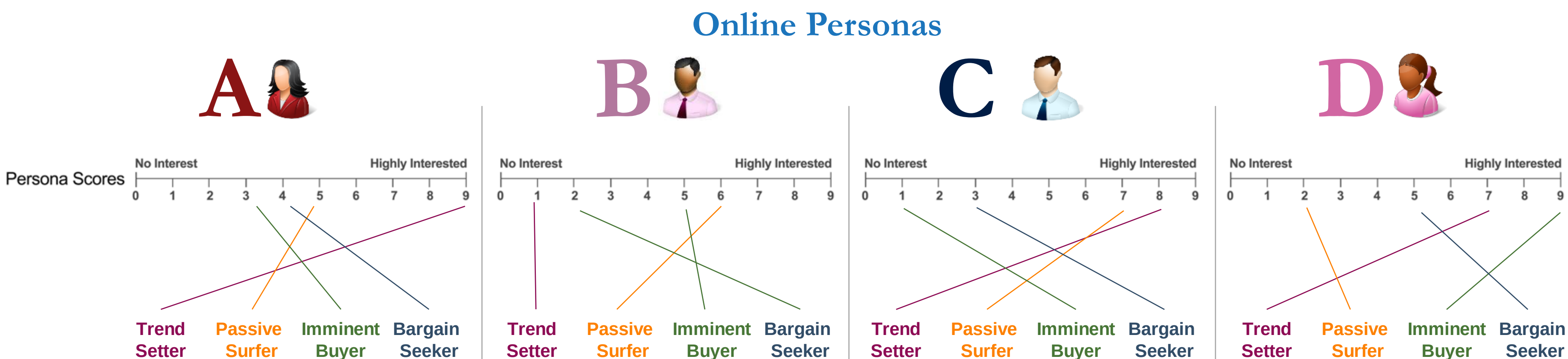
PERSONALIZE EVERY MESSAGE

Data, analytics, content management and delivery working in conjunction

WILLOW DATA STRATEGY
making sense of data

Reactive Personalization w/ Online Behavior





Identify Personas

Heuristic or statistical personas based on visitor's online behavior, converting "unknowns" to "potentials"

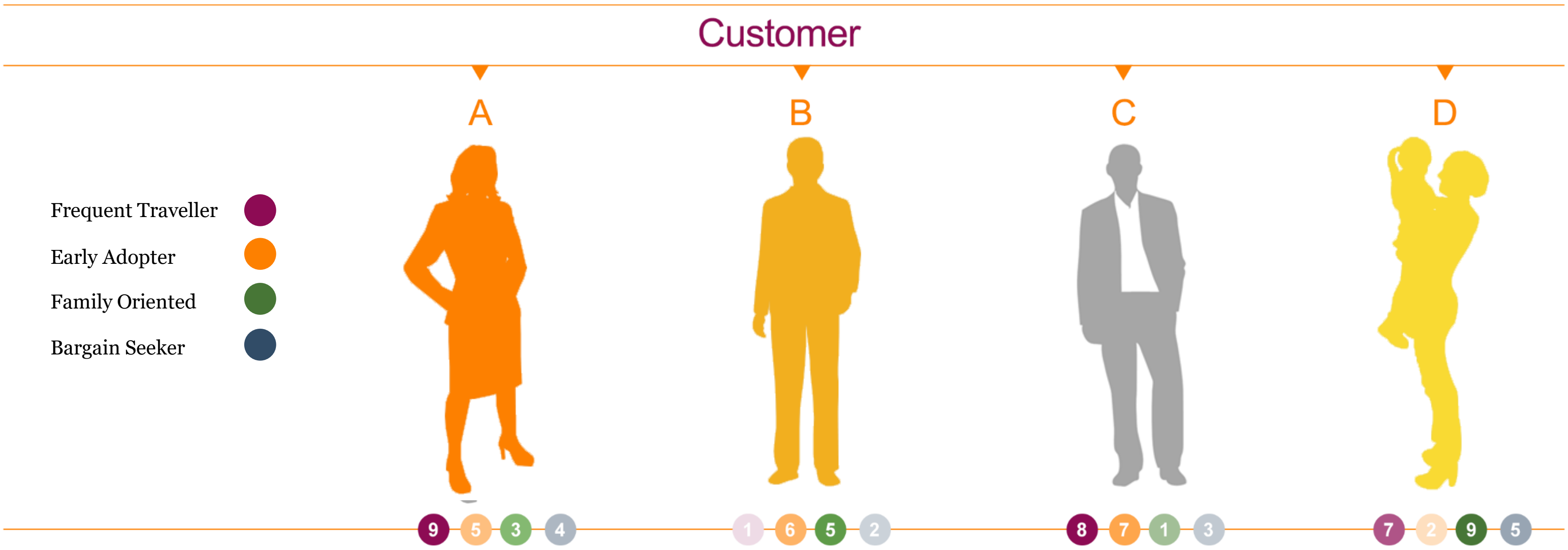
Identify Levers

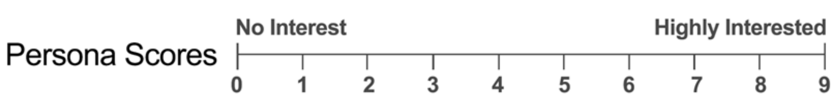
Connect the dots between the visitor behavior and desired business outcome

Personalize

Extend personalization to new visitors in real-time, driving them towards desired outcome

Personas Based-on Full Set of Data





Extend personalization to customers with limited data using model-based personas

- Convert “unknown” to “potentials”
- Go beyond rudimentary segments
- Match best products/offers to each target

WILLOW DATA STRATEGY
making sense of data

Examples of Personas

For Hospitality Industry

- Frequent Flyer
- Foreign Traveler
- Luxury Hotel
- Gourmet
- Wine Enthusiast
- Adventure Seeker
- Young Family
- Budget Conscious

For Gift Industry

- Family Oriented
- Romantic
- High-end / Luxury
- Seasonal
- Frequent Small Gifts
- Pre-packages
- Bargain Seekers
- Specialty Items

For Business Solutions

- Corporate Purchase
- Home Office
- Consumables / Repeat Purchase
- Big Ticket Items
- Technology Buyers
- Early Adopters
- Trend Followers

Examples of Personas

For Telco Industry

- Home office
- Family with kids
- Heavy download
- Gamer
- Mobile phone only
- VOIP user
- International Caller
- Pay-per-view Movie
- Pay-per-view Sports
- Season ticket holder

For Financial Industry

- Frequent trader
- Do-it-yourself investor
- High value investor
- Institutional investor
- Annuity investor
- Risk-averse investor
- Safety conscience
- Credit revolver
- Credit card transactor

For Women's Interest

- Fashion enthusiast
- High-end accessories
- Beauty / Skincare
- Children's interest
- Gardening enthusiast
- Organic food
- Weight-loss / Diet
- Gourmet cooking
- Spa / Day-spa
- Family entertainment
- Foreign vacation
- Bargain seeker / Value shopper
- Coupon collector

Segments vs. Personas

Clustering/Segmentation/Cohorts

- More about messaging than targeting
- Pin a target individual into one segment at a time
- Hard to update algorithms with reliable consistency – Most just keep rescoring new data using dated formula
- Group them first, describe them later
  - End up calling everyone in a segment the same way

Personas

- Built for 1 attribute at a time
- Describe an individual with multiple attributes
- Identifies dominant characteristics of a person via side-by-side comparison
- Each persona represents diverse array of data
- Easy to update, 1 persona (i.e., one model) at a time with consistency
- Ready for multi-channel marketing

Personas for Personalization

"Personas" built for specific attributes

- Project small "known" attributes to large universes of "unknowns" in form of model scores
- Fill in data gaps leaving *no missing value* – Scores for every record using all available data
- Enable side-by-side comparison of attributes – *Quickly find dominant characteristics of an individual*
- Simplify matching process between *individuals and best suitable products/services*
- Support *message "rotation"* for an individual customer using multiple personas
- Lead to *marketing automation* – Simple scores are no burden to personalization engines

Post-Campaign Analysis and Attribution

Close the Loop Properly

Closed-Loop Marketing

- 1-to-1 Marketing is a Continuous Loop
- Analytics “before” and “after” campaigns → What really worked?
- Last-click attribution is seriously limited
    - Digital channels only
    - Limited view of customer journey

→ Again, it is about commitment, not technology

WILLOW DATA STRATEGY
making sense of data

What worked, and what didn't?

In the Multi-Channel Marketing Environment, pinpointing "What Worked" is difficult:

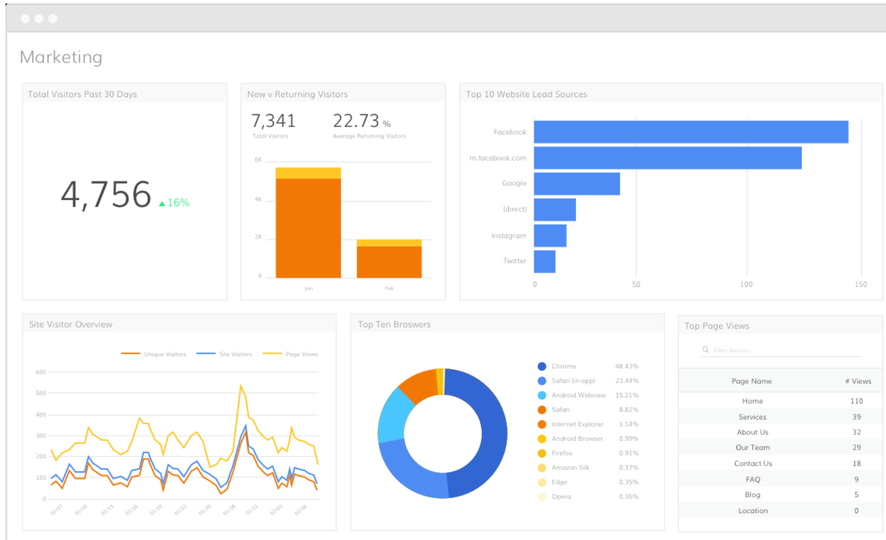
- Escape channel-centric view, and consider other elements
- Don't settle with "Last Touch" attribution
- Retain Customer-Centric view
- Continuously apply learnings to the next campaign

What did *you* do?

Target	Offer/Message
• Target List Source • Primary Segments • Selection Rules • Model Score Group / Clusters	• Promotion Channel • Campaign • Drop • Primary Offer(s) • Feature Product(s) • Creative Version(s)

Always maintain promotion and response history on a transaction and an individual customer level

Response Reports & BI Analytics



- Start with “Canned” Reports and BI Dashboards from vendors
- Don’t insist real-time for every report
- Get ready to create “Custom” reports and dashboard views
  - Prioritize what you want
- Format and Delivery
- Timing and Interval
- Timeline to be covered and pulled (YTD, 12-mo, 3-mo, 2-weeks, daily, etc.)

Key Performance Indicators

Set KPIs upfront:

- Open, Click-through, Conversion Rates
  - “Denominator” for each rate?
- Revenue
  - Per 1,000 Mailed/Contacts
  - Per Order/Transaction
  - Per Display, Email, Click-through, Conversion
- By
  - Source, Campaign, Drop, Season, Day-part, Segment, Model Group, Offer, Creative, Featured Products, Channels (in & outbound), Ad server, Publisher, Affiliate, Key word, Script/Copy, etc.
- Key variables must reside in the database
  - Most digital tags are NOT in consistent format

  → Keep them in “report-ready” forms

1:1 Channels and Customer Journey

Allocation Methods:

- Last Touch
- First Touch
- Double Credit
- Equal Credit
- Proportional Split (based on touches)
- Weighted Value
- Modeled Weighted Value

Also Consider:

- Direct Click and/or Look-back Attribution
- Go beyond Google UTM tags and consider non-digital Channels
- Time Duration – depending on the channel in question

Touch Channels Regardless of Order

Outbound	Inbound
• Email • Direct Mail	• Direct • Organic Search • Paid Search • Social Media • Display Ads • Referrals • Affiliates

Assume every target is exposed to multiple channels.

Consider 2-step summary process:
1. Transaction Level Attribution
2. Individual Level Attribution

Future of Data & Analytics

What will we be doing in the future?

Personalized Ad in the Future

WILLOW DATA STRATEGY
making sense of data

The Future Is Here Already

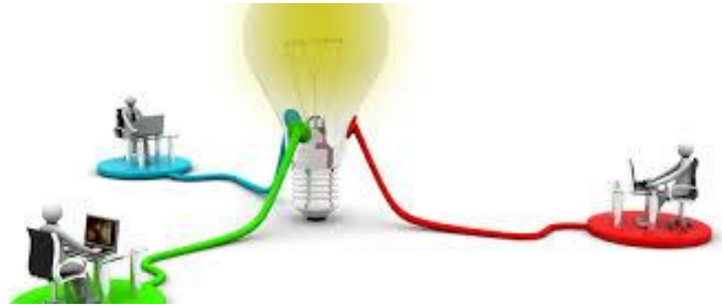
[A Scene from the Movie "Minority Report" - Personal Advertising in the Future]

What just happened in this scene?

1. Machine identified the target individual
2. Retrieved the target's past transaction data
3. Algorithm chose suitable products to offer
4. Displayed images of selected products to the target individual in real-time

- We can do all this now – albeit more slowly
- Cannot skip these steps even with AI & automation – Different algorithms for each step
- It is about the target individuals and their data, not product, channel, or display technology

Information Must Flow



In the world of ubiquitous data and AI:

- It is really about organizational commitment to harness the power of data
- Higher values are created when information is freely exchanged – among divisions, brands, companies and individuals
- Predictive power increases as multiple types of data are joined together
- Speed of data exchange and processing is the only game left in town
- It will require less human intervention, and more AI & Machine Learning

→ *Will most humans even be relevant or necessary in the near future?*

Humanizing the Data

- Data play is people business in the end
  - Customer-centric mindset, not business-, division-, product-, or channel-centric
  - Make data usage easier, not more complicated
- Don't do it just because you can
  - Respect people's privacy
  - Gentle nudges, not hard-sells
  - Don't be creepy – TMI is bad
- Try not to get replaced by machines
  - Command machines with purposes
  - Think logically even as end-users
  - Don't be a Data Plumber and go beyond data and statistical knowledge – Look at things from multiple perspectives

Only human factors and mathematics will remain unchanged.

Consumer Bill of Rights

Consumer Bill of Rights for Online Engagement (1999)

by Lester Wunderman, the Father of Direct Marketing

1. Tell me clearly who you are and why you are contacting me.
2. Tell me clearly what you are—or are not—going to do with the information I give.
3. Don't pretend that you know me personally. You don't know me; you know some things about me.
4. Don't assume that we have a relationship.
5. Don't assume that I want to have a relationship with you.
6. Make it easy for me to say "yes" and "no."
7. When I say "no," accept that I mean *not this, not now*.
8. Help me budget not only my money, but also my TIME.
9. My time is valuable, don't waste it.
10. Make my shopping experience easier.
11. Don't communicate with me just because you can.
12. If you do all of that, maybe we will then have the basis for a relationship!

Key Takeaways

1. Business first; it is not about data or technology
2. Invest in analytics – Models pack large amount of data into simple answers to questions
3. Databases must be optimized for analytics – “Analytics Sandbox” for consistency and speed
4. Start small with a proof of concept – Don’t take a big bite
5. Consider every data source, but don’t wait for a perfect dataset
6. Don’t settle for simple segments – Employ personas for personalization through every channel
7. Close the loop properly – Always conduct post-campaign analyses with proper channel attribution
8. Don’t get lost among machines – Give them purposes
9. Ask for help

Q&A

Further Questions?

Contact:

Stephen H. Yu
President & Chief Consultant
Willow Data Strategy, LLC
stephen.yu@willowdatastrategy.com

For comprehensive data and analytics solutioning, check our website:

www.willowdatastrategy.com

For further details on this presentation, please read my series in Adweek:

Complete list of my articles, webinars and interviews is at:

https://www.willowdatastrategy.com/articles-and-publications