



**KS&R**  
Data to Knowledge

*International Workshop on*  
BUSINESS DATA MINING

**CR RAO AIMSCS & DCIS**  
University of Hyderabad

**December 24, 2008**

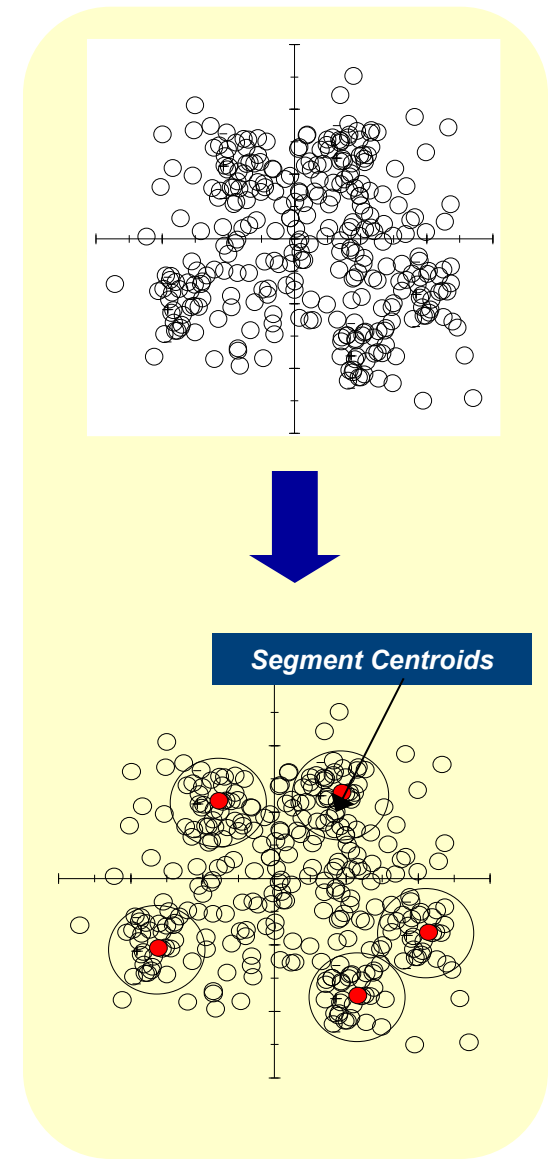
# Segmentation Methods

Thomas T. John, Ph.D.  
Lead Statistician, KS&R, Inc., Syracuse, USA  
Adjunct Faculty, Syracuse University, Syracuse, USA

Contact:  
tjohn@ksrinc.com (Office)  
thomas\_t\_john@yahoo.com (Personal)

# Segmentation – Definition & Purpose

- Segmentation divides a defined market into groups of customers/targets
  - who are likely to exhibit similar behaviors related to purchasing behavior, loyalty, retention, etc.
  - These groups are likely to have manageable sizes (versus the entire market)
- Its purpose is to drive tailored approaches to these smaller groups (segments/clusters), in a number of key areas:
  - Marketing Communication – deliver the right message
  - Value Propositions and Offerings – tailored to a segment's specific needs
  - Delivery – deliver products/services in unique ways that meets/exceeds customer needs



# Segmentation – Definition & Purpose (continued)

*“To be valid, a segmentation must identify groups that matter to a company’s financial performance. To start, companies can rank their own customers by profitability so as to concentrate the right amount of attention on them. But to grow revenues, a company should understand what makes its best customers as profitable as they are and then seek new customers who share at least a couple of those characteristics.”*

Rediscovering Market Segmentation,  
Harvard Business Review, 2006

- Market segmentation should also be operational (“implementable”)
  - Need to enable target marketing/selling (or de-selection)
  - Attained by classifying customers/prospects into defined segments
  - Classification based on certain known/knowable characteristics
- The end objective of market segmentation is business performance driven
  - Tailoring communications, value propositions, offerings and delivery to specific segment needs to improve win rates and thus increase market share
  - and/or, to reduce customer loss (“churn”) and thus retain market share

# Step 1 – Proper Planning

- Understand the specific business objectives
  - A kick-off meeting between researchers & business managers highly valuable for ultimate success
  - Specify & clarify the goals: what the managers expect to achieve and what researchers expect to deliver
- Determine the “*basis variables*”
  - Basis variables are dimensions in which the segments are to be defined
  - Must be closely tied to business objectives
  - Basis variables could be of different types:
    - Needs of individuals/businesses
    - Behavior/Attitudes of individuals/businesses
    - Purchase/Buying-Decision-Making patterns
    - Others – demographic/firmographic, etc.

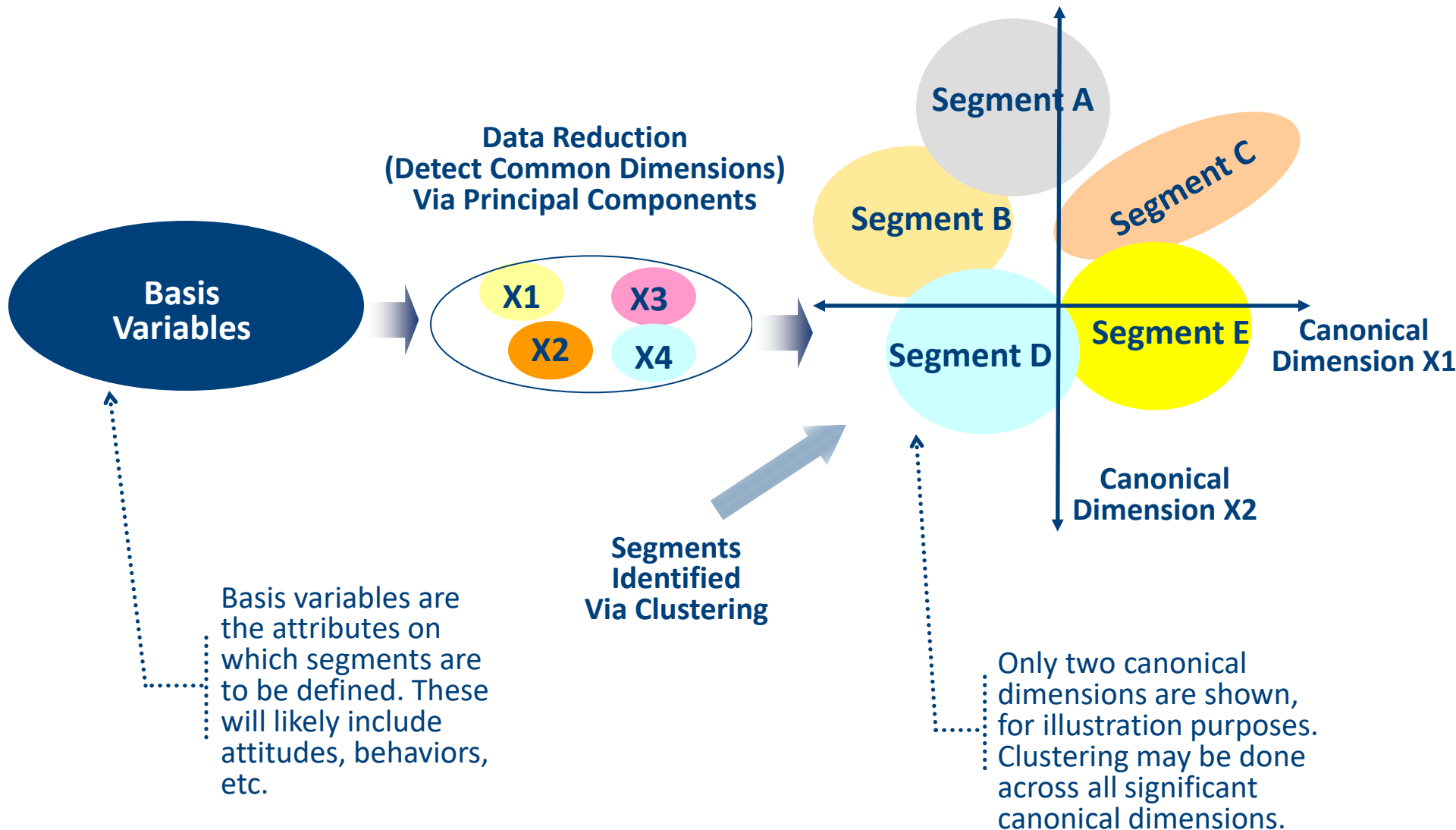
## Step 2 – Data Preparation

- If data not pre-existing, design a proper data collection method
  - Most data mining scenarios have data readily available
- Preliminary Data Review
  - Data cleaning
  - Outlier detection
  - Missing value specifications (required for certain software)
- Exploratory Data Analyses
  - Distributional properties
  - Assess the need for variable scaling/standardization/transformations
- Data Reduction
  - Principal components methods: extract factors (canonical dimensions) and identify subject specific meanings of those
  - Subject-specific variable combinations: e.g., P/E (price-to-earnings) ratio.

## Step 3 – Determine the segments

- Select an appropriate Clustering algorithm:
  - k-Means, k-Medoid, Hierarchical, Agglomerative hierarchical, Two-Step, etc.
- Select a proper similarity metric (distance measure):
  - Euclidean, Minkowski, Mahalanobis, Log-Likelihood Ratio, etc.
- Select an appropriate stopping rule:
  - No changes in cluster membership (k-Means/k-Medoids)
  - Bayesian Info Criterion (BIC), Akaike Info Criterion (Hierarchical/Two-Step)
- Number of clusters
  - To be specified for k-Means. Rules of thumb available.
  - BIC/AIC can be used for Hierarchical.
- Decision Tree Type Algorithms may also be used for segmenting
  - CHAID, Classification & Regression Trees (CART), etc.

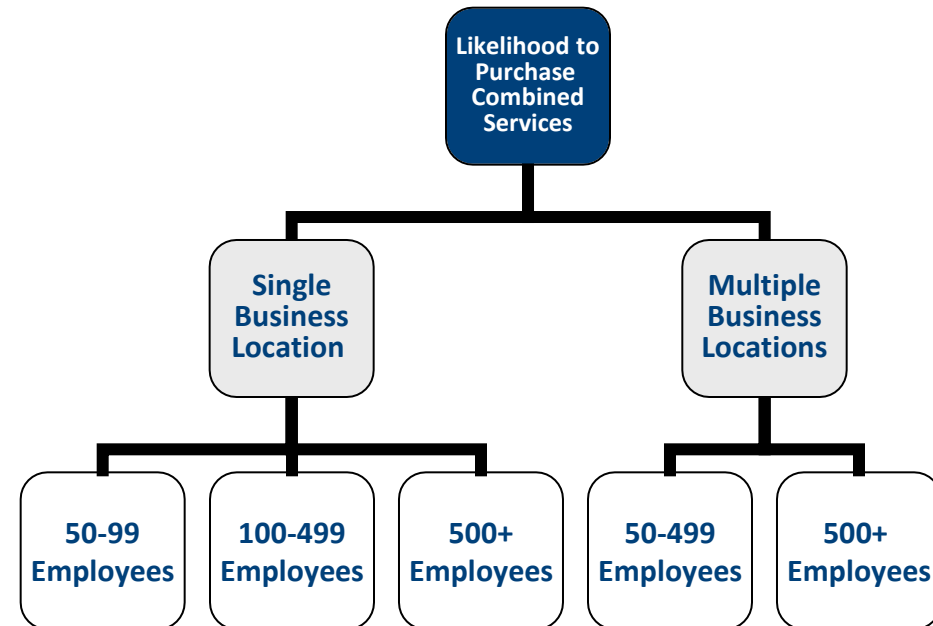
# Segmentation via Data Reduction & Clustering



# Segmentation via Decision Tree Type Algorithms

- CHAID (Chi-Square Automatic Interaction Detector) illustrated:
  - Identifies segments based on the differences in the dependent variable (e.g., likelihood to purchase bundled services)
  - Segments defined by the order of the explanatory variables (e.g. industry, company size, etc.) that best predict bundle purchase likelihood.
  - There are related methods called AID, THAID, and SICHAID for other data types

Sample CHAID Output  
(Tree-Type Chart)

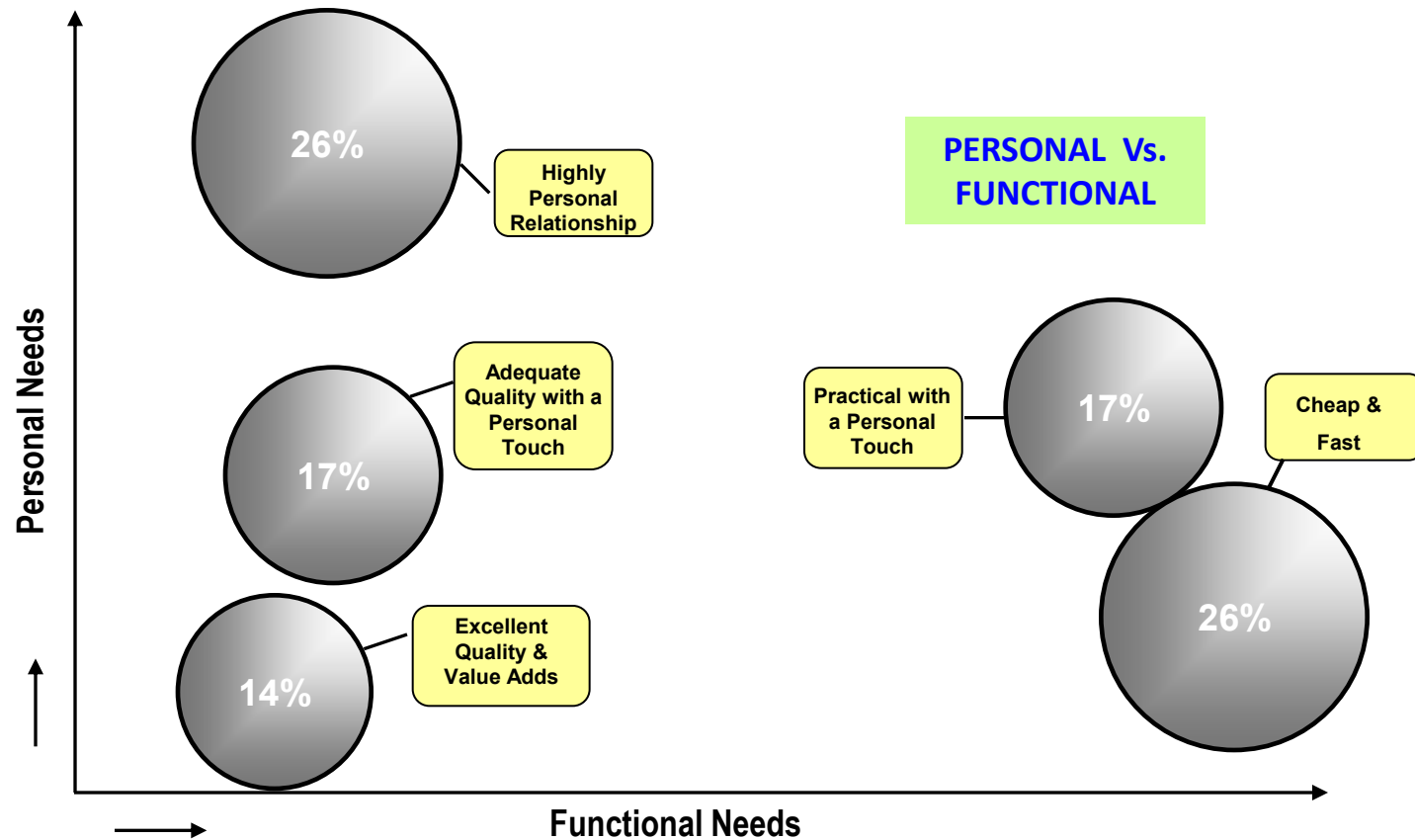


## Step 4 – “Identify” the Segments

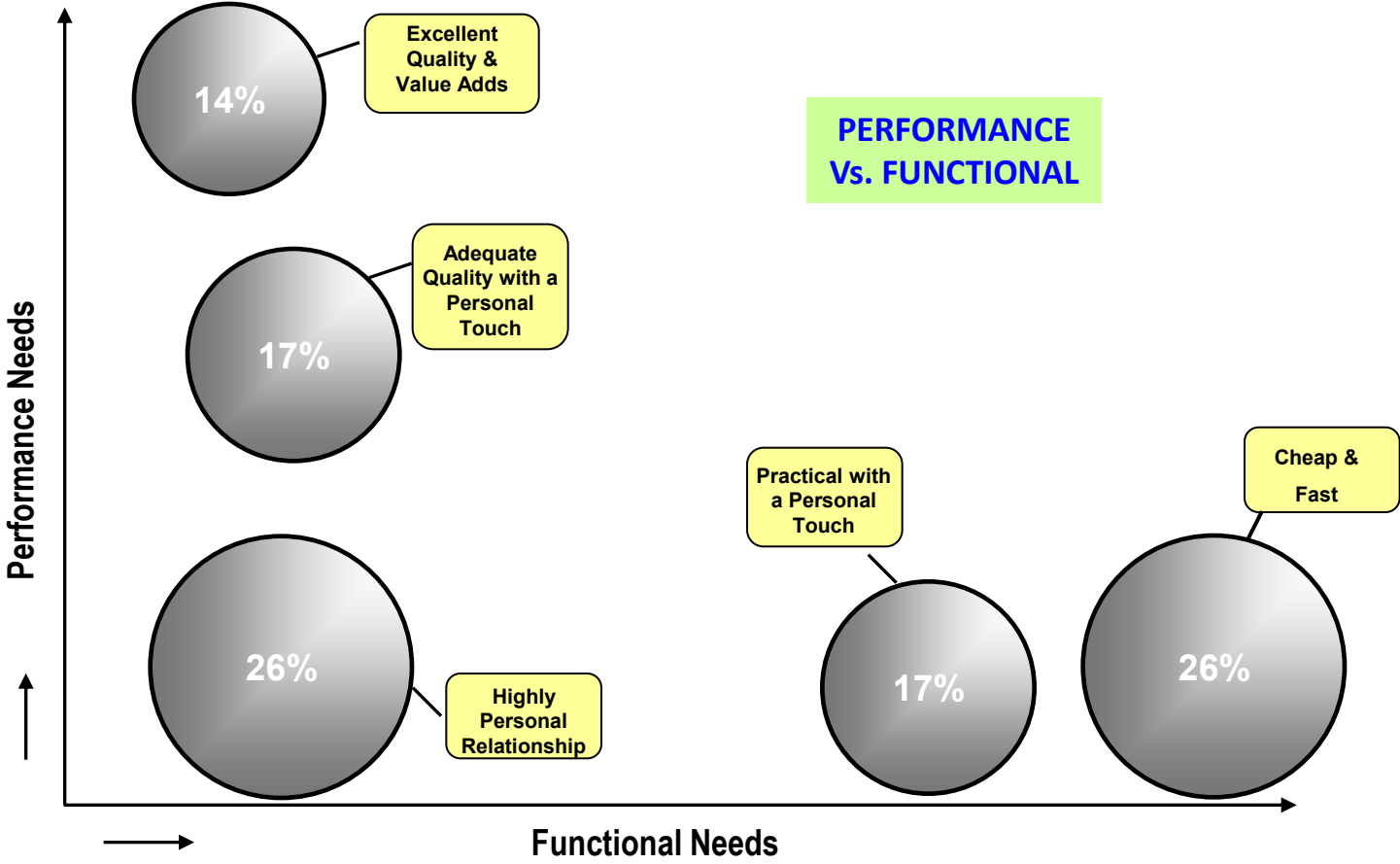
- Provide a detailed profile of each segment
  - Understand “WHO” are the segment members
  - Understand the size of the segment: propensity-based and opportunity-based
- Analyze the relationship of the newly-developed segments to
  - Existing segmentation schemes
  - Demographic/Firmographic delineations (e.g., income/revenue level, region, etc)
- Wholly understand, capture, and analyze the underlying dimensions (needs/attitudes/buying patterns)
  - Across the market and within the segments
- Descriptive Data Analyses are usually sufficient:
  - Cross tabulations of other variables, with segments as columns

# Example: *Business Needs Expected From Providers*

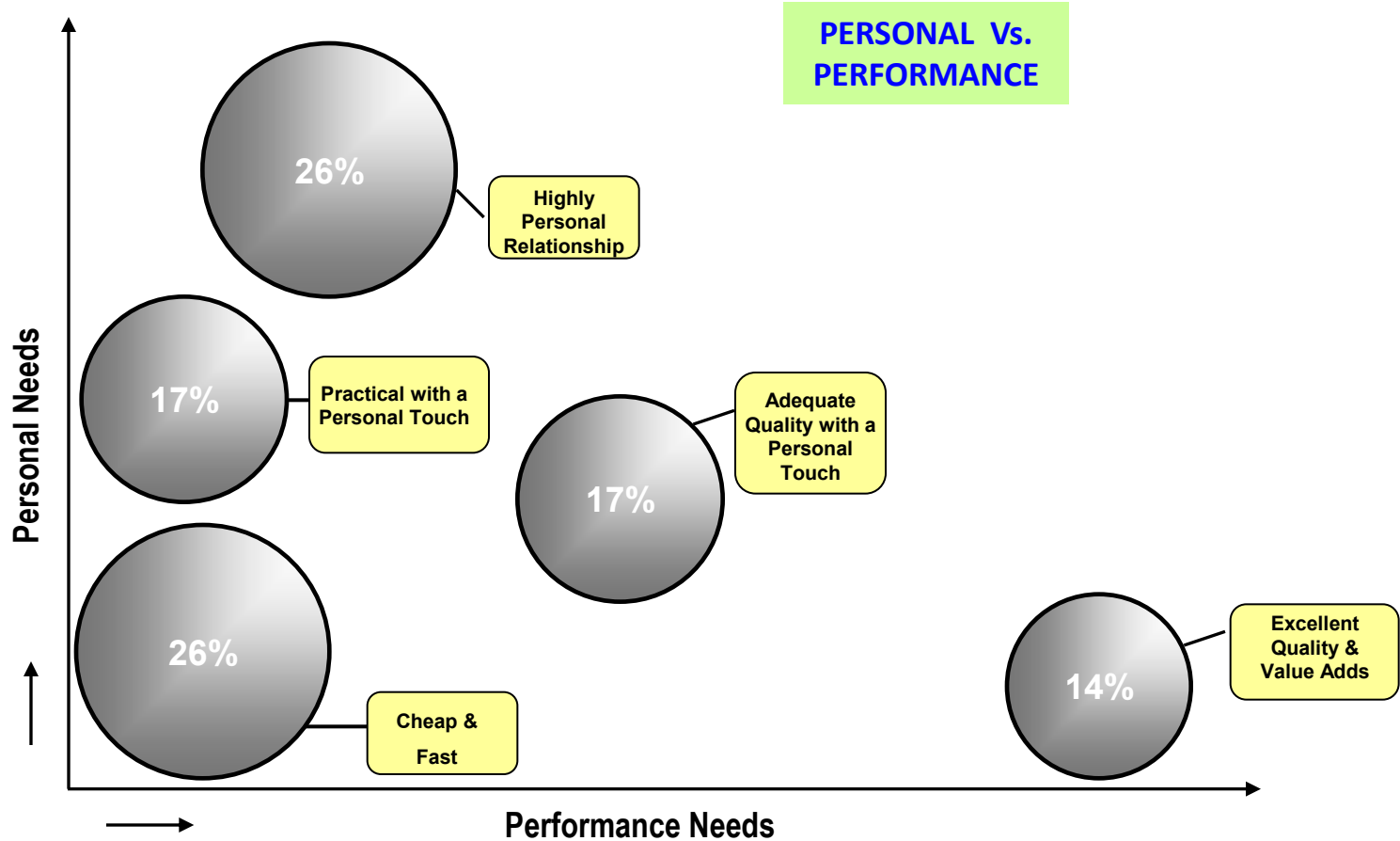
- Three key dimensions were identified:
  - Functional, Personal, Performance
- Locations of segments relative to these axes help identify “actionable meanings” to each customer group



# Example: Business Needs Expected From Providers (cont'd)



# Example: Business Needs Expected From Providers (cont'd)



# Step 5 – Towards “Operationalising” Segmentation

- Analyze & identify the implications of the segments
  - How to communicate (targeted messaging, value proposition, etc.)
  - Find the best media for messaging
  - How to increase satisfaction through proper product/service delivery
  - Find the best channels for product delivery
- Determine Market Positioning
  - Identify segment alignments to specific brands/products
  - Recognize gaps/white space for new product entries or product updates
- Develop a classification model to map customers/prospects to the newly developed segments
  - Supervised classification schemes, or logistic regression/discriminant
  - Business managers can then implement the scheme in their databases
  - Knowing the likely segment membership and associated behaviors can extremely improve the customer interactions