MARKET SEGMENTATION ANALYSIS

Step-1. Deciding (Not) To Segment - Anandhu S

Purpose of Marketing: To align consumer needs and desires with offerings that satisfy them, benefiting both consumers and suppliers.

Marketing Planning: A logical sequence that leads to setting objectives and formulating plans to achieve them.

Two Components of a Marketing Plan:

- **Strategic Plan**: Defines the long-term direction and "where and why" the organization is heading.
- **Tactical Plan**: Translates strategy into detailed short-term actions ("what to do and how").

Hiking Analogy:

- Choosing the mountain = **Strategic decision**
- Packing gear, timing, etc. = **Tactical decisions**

SWOT Analysis: A core part of strategic planning that identifies:

- Strengths
- Weaknesses
- Opportunities
- Threats

Understanding Consumer Needs: Achieved through:

- Surveys
- Focus groups and interviews
- Observational and experimental research

Key Strategic Decisions:

- **Segmentation and Targeting**: Which consumers to focus on
- **Positioning**: How the organization wants to be perceived.

Tactical Marketing (4Ps):

- **Product**: Designing offerings that meet target segment needs
- **Price**: Based on cost, competition, and willingness to pay
- **Place**: Selecting distribution channels
- **Promotion**: Crafting effective communications for the target segment

Asymmetry Between Strategy and Tactics:

- Good **strategy** + good **tactics** = Success
- Good **strategy** + bad **tactics** = Survival
- Bad **strategy** + good **tactics** = Fast failure

- Bad **strategy** + bad **tactics** = Slow failure

Conclusion: Strategic marketing is the foundation; tactical marketing enhances but cannot fix poor strategy.

Definitions of Market Segmentation

- **Market segmentation** is a decision-making tool for identifying the target market and designing a suitable marketing mix.
- It is a **core element of strategic marketing**, critical to business success.
- **Smith (1956)** introduced the concept, defining it as splitting a heterogeneous market into smaller, more homogeneous submarkets.
- Segmentation lies between the extremes of:
  - Total uniqueness of every customer.
  - Homogeneous treatment of all customers.
- A simple definition: **"cutting markets into slices"** – grouping similar consumers together and separating distinct ones.
- **Segmentation criteria**: Consumer characteristics critical to marketing decisions (e.g., age, gender, behavior, preferences).
- Criteria can be:
  - **Single variable** (e.g., age, income)
  - **Multi-variable** (e.g., values, benefits sought, activities, expenditure)
- **Example**: Mobile phone market segmented by:
  - High-end users (many features, high price)
  - Budget users (few features, low price)
  - Mid-range users (moderate features and price)
- Trying to appeal to **everyone with one product** is ineffective and fails to build a strong brand image.
- Concentrated strategy:
  - Focuses on one market segment.
  - Good for resource-limited firms in competitive markets.
  - Risky due to reliance on a single segment.
- Differentiated strategy:
  - Targets multiple segments with tailored products.
  - Suited for mature markets with experienced consumers.
- Undifferentiated strategy:
  - One product, one message for the whole market.
  - Suitable for basic commodities (e.g., petrol, bread).
  - Often used when resources are high or consumer preferences are undeveloped.

Benefits of Market Segmentation

- Encourages organizations t reflect on:
  - Current market position.
  - Strategic future direction.
  - Competitive advantages.
  - Consumer desires and expectations.
- Leads to new **insights** and **perspectives**.
- Enhances understanding of **consumer differences**, allowing better alignment

between:

    - Organizational strengths
    - **Consumer needs**
- Improves targeting, potentially resulting in:
    - Long-term **competitive advantage**
    - Even **market dominance** within niche segments

- Supports **niche marketing**:
    - Focused on segments that are profitable, growing, and unattractive to competitors.

- Enables hyper-segmentation / micro marketing:
    - Customizing products/services for very small groups or even individual consumers.
- eCommerce & big data have made **finer segmentation** more feasible.
- Higher **return on investment (ROI)** from a tailored marketing mix:
    - Less wasted effort on consumers unlikely to buy.
- Essential for **small organizations**:
    - Limited resources mean they must focus on tightly defined segments to survive.

- Enhances **sales management**:
    - Sales efforts can be focused on groups rather than individuals.
- Encourages **teamwork and collaboration** across departments:
    - Improves communication and data sharing within the organization.

Costs of Market Segmentation

- Requires **significant investment** in:
    - Time
    - Human resources
    - Financial resources
- Development and execution of a **customized marketing mix** increase complexity and cost.
- Ongoing effort is needed for:
    - Evaluating strategy success
    - Continuously monitoring market changes

- Risk of failure:
    - Poorly implemented segmentation may result in wasted resources with no ROI.
    - Can **demotivate staff** if the project fails to deliver value.
- Organizations must make an **informed decision** before committing to segmentation:
    - It is a **long-term strategic effort**, not a quick fix.

The Layers of Market Segmentation Analysis – Key Points

Core Layer: Extracting Market Segments

- Involves statistically grouping consumers with similar preferences or characteristics.
- The process is **exploratory** and influenced by decisions made by the data analyst.
- Requires collaboration between:

- A competent data analyst
- A **user** (e.g., marketing team or department using the insights)

Second Layer: Enabling High-Quality Segmentation

- **Good data collection** is critical — poor data leads to poor segmentation results.
- Key technical tasks:
  - Collecting good data
  - **Exploring data** before analysis to gain preliminary insights
  - **Profiling** segments: identifying defining traits of each segment
  - **Describing** segments in detail to aid decision-making and targeting

Third Layer: Making It Happen in Practice (Implementation Layer)

- Involves **non-technical, strategic and organizational tasks**:
  - **Deciding to segment** (based on organizational fit and opportunity)
  - **Defining the ideal segment**
  - **Selecting target segment(s)**
  - **Developing a customized marketing mix**
  - **Assessing effectiveness** and **monitoring changes**
- This layer wraps around the technical tasks — it's **not sequential**, but critical for real-world success.
- User (organization's decision-makers) must:
  - Decide whether segmentation is a worthwhile strategy.
  - Ensure data collection captures relevant consumer information.
  - Evaluate segment extraction results and select target segments.
  - Design a targeted marketing plan and implement it.

Conclusion

- A theoretically excellent segmentation is **useless** unless it is translated into actionable business decisions and strategies.
- **Full collaboration** between data analysts and organizational users is essential for success.
- Approaches to Market Segmentation Analysis – Key Points

Overview

- No single "best" approach to segmentation.
- Two main systematics presented:
  1. **Based on organizational constraints** (Dibb & Simkin, 2008)
  2. **Based on the nature of segmentation variables** (covered in later sections)

Based on Organisational Constraints

Three Main Approaches (Dibb & Simkin, 2008):

1. **Segment Revolution**
   - **Most radical** approach – starts from scratch.
   - Based on **quantitative, survey-based segmentation**.
   - Organization must be willing to completely rethink marketing.
   - Likely to yield the **greatest benefits**, but often **not viable** due to internal resistance or resource limitations.
2. Segment Evolution

    - **Moderate change** – refines existing segmentation strategies.
    - Based on **existing segments**, improved through workshops or internal insights.
    - Focuses on **enhancing what's already in place**.
    - Suitable when existing segments are performing reasonably well.

3. Segment Mutation
    - **Least radical** – accidental discovery of new segments.
    - Arises through **exploratory or qualitative research**, or **big data mining**.
    - Segments emerge as a **by-product** of other research or data processes.
    - Can lead to significant insights if noticed and acted upon.

Sandcastle Analogy:

- **Revolution**: Destroying the old castle and building a new one.
- **Evolution**: Improving and refining the current sandcastle.
- **Mutation**: Discovering a random pile of sand perfect for building.

Practical Considerations:

- Organizations may **not be free to choose** any method due to **resource, cultural, or structural constraints**.
- All approaches can be enhanced using **data analytics** and **ongoing tracking** (e.g., in big data environments).
- Continuously monitoring segment structure helps in **keeping segmentation strategies relevant** over time.

Based on the Choice of Segmentation Variable(s)

Unidimensional vs. Multidimensional Segmentation

- **Unidimensional segmentation** uses **one variable** (e.g., age, gender).
    - Example: Age groups (e.g., targeting seniors).
- **Multidimensional segmentation** uses **multiple variables** (e.g., expenditure patterns, preferences).
    - Example: Tourists' spending on different vacation activities to target high-value customers.

◈ Commonsense (A Priori) Segmentation

- Also called:
    - Convenience-group segmentation
    - **A priori segmentation**
- Based on **predefined characteristics** or **intuition** before data analysis.
- Commonly relies on:
    - Managerial intuition
    - Secondary data sources
    - Internal databases or previously known segments
- **Not exploratory** – assumes segments are already known.
- Example: Segmenting based on brand preference.
- Strengths:
    - Simple and efficient
    - Useful when a **strong, known variable** relates to buying behavior
- Limitations:

  - May lack depth
  - Risk of missing emerging or hidden segments
- Considered a **reactive approach**.

Data-Driven (A Posteriori) Segmentation

- Also called:
  1. **Cluster-based segmentation**
  2. **Post hoc segmentation**
  3. **Data-driven segmentation**
- Conducted **after analyzing data** – segments are discovered, not predefined.
- Based on **primary research** into preferences and behavior.
- Purpose:
  1. Discover meaningful market segments.
  2. Develop detailed profiles of these segments.
- Considered a **proactive, exploratory approach**.
- Allows flexibility and uncovers hidden insights in the data.

Combining Segmentation Approaches (Hybrid/Multistage)

- In real-world applications, segmentation often combines both approaches.
- Multistage segmentation:
  - Start with commonsense (e.g., age group),
  - Then apply data-driven analysis to refine within the group.
- Example:
  - Boksberger & Laesser (2009): First segment senior travelers (commonsense), then apply motive-based segmentation (data-driven).
- Benefits:
  - More precision
  - Better-targeted marketing strategies

Step – 2. Specifying the Ideal Target Segment - Anandhu S

Market Segmentation Analysis

Ten Steps of Market Segmentation
(Applicable to Both Commonsense & Data-Driven Approaches)

1. **Deciding (Not) to Segment**
    - Assess if segmentation is appropriate.
    - Consider market suitability and ability to commit long-term.
2. Specifying the Ideal Target Segment
    - Define characteristics of the "ideal" segment before analyzing data.
3. Collecting Data
    - Gather data on **segmentation variables** and **descriptor variables**.
    - Use surveys, internal records, or existing datasets.
4. Exploring Data
    - Preprocess and analyze data for patterns and quality.
    - Ensure the dataset is ready for segment extraction.
5. Extracting Segments
    - **Commonsense**: Split consumers based on predefined criteria (e.g., age).
    - **Data-driven**: Use clustering methods (e.g., distance-based, model-based, hybrid).
6. Profiling Segments
    - Identify key features that define each segment.
    - Understand what makes each group unique.
7. Describing Segments
    - Provide detailed descriptions using statistics, charts, and behavior insights.
8. Selecting Target Segment(s)
    - Evaluate and choose which segments to focus on.
    - Consider profitability, accessibility, and strategic fit.
9. Customizing the Marketing Mix
    - Tailor the 4Ps (Product, Price, Place, Promotion) for selected segments.
10. Evaluation and Monitoring

- Assess performance of the segmentation strategy.
- Continuously monitor segment changes over time.

Deciding (Not) to Segment

Implications of Committing to Market Segmentation

- Market segmentation requires **long-term commitment**—it's more like a marriage than a casual decision.
- Significant **resource investment** is needed (e.g., surveys, packaging, advertising).
- Segmentation is only worthwhile if **increased profitability** outweighs the implementation cost.
- May involve:
    - Developing **new products**
    - Changing **pricing and distribution**

- Redesigning **marketing communications**

- Often requires **organizational restructuring** to focus on market segments rather than products.
- Strategic decision must be made at the **executive level** and supported across all departments.

Implementation Barriers

A. Leadership & Senior Management Barriers

- Lack of leadership support and understanding can derail segmentation efforts.
- Insufficient allocation of resources for research or implementation is a major obstacle.

B. Organisational Culture Barriers

- Resistance to change and poor market orientation can hinder success.
- Lack of communication and cross-functional collaboration weakens execution.
- Office politics and short-term thinking reduce commitment to long-term strategies.

C. Knowledge & Skill Barriers

- Lack of training in segmentation principles among leadership and team members.
- Absence of a formal marketing function or skilled marketing/data experts.
- Lack of a qualified data analyst or data manager can block progress.

D. Resource Constraints

- Financial and structural limitations may prevent full execution.
- Smaller companies must be selective in pursuing segmentation due to limited capacity.

E. Process & Planning Barriers

- Unclear objectives, lack of planning, or poor structuring of the process hampers success.
- Time pressure can reduce the quality and effectiveness of segmentation outcomes.

F. Operational & Acceptance Barriers

- Management may resist data-driven techniques they don't understand.
- Visualization tools and clear presentations can increase managerial acceptance.

Checklist

Knock-Out Criteria Questions (Must be Answered "Yes" to Proceed)

- Is the **organization market-oriented**?
- Is the organization genuinely willing to change?
- Does the organization take a **long-term perspective**?
- Is the organization **open to new ideas**?
- Is there **good cross-department communication**?
- Can the organization make **significant structural changes**?
- Are there **sufficient financial resources** available for segmentation?

If any of the above are answered "No," the organization should **reconsider proceeding** with segmentation.

Foundational Setup Tasks

- Secure **visible and active support from senior management**.
- Obtain **financial commitment** from leadership.
- Ensure **full understanding** of market segmentation concepts and implications through training if necessary.
- Form a **segmentation team** of 2–3 people.

Team & Structure Requirements

- Include a **marketing expert**, **data expert**, and **data analysis expert** on the team.
- Establish an **advisory committee** with reps from all affected departments.
- Clearly define the **objectives** of the segmentation analysis.
- Develop and follow a **structured segmentation process**.
- Assign **specific responsibilities** to each team member.

Step – 3. Collecting Data – Segmentation Variables - Anandhu S

- **Empirical data** is essential for both **commonsense** and **data-driven** segmentation.
- The **segmentation variable** is used to divide consumers into segments.
  - In **commonsense segmentation**, this is usually a **single characteristic** (e.g., gender).
  - In **data-driven segmentation**, **multiple variables** are used to form segments.

Commonsense Segmentation Example (Table 5.1)

- Uses **gender** as the segmentation variable.
- Splits the sample into two groups: **men and women**.
- Other characteristics (e.g., age, vacations taken, travel motives) are used as **descriptor variables** to describe each group.

Descriptor Variables

- Not used to form the segments but to **describe them**.
- Help in developing **tailored marketing strategies**.
- Common descriptors include:
  - **Socio-demographics** (e.g., age, gender)
  - Media behavior
  - **Benefits sought** (e.g., relaxation, action, culture, explore, meet people)
- Uses **multiple segmentation variables** like:
  - Travel motives (relaxation, action, culture, explore)
  - Vacation behavior (e.g., number of vacations)
- Allows for **more complex and accurate grouping** based on natural patterns in the data.

Importance of Data Quality

- Poor data → inaccurate segmentation and misinformed marketing strategies.
- Empirical data should ideally reflect **actual consumer behavior**.
- Data sources include:
  - **Surveys** (most common, but may lack behavioral accuracy)
  - **Scanner data / Loyalty programs** (more reliable behavioral insights)
  - Observational or experimental studies
- Surveys can misrepresent **socially desirable behaviors** (e.g., charity, sustainability).

Segmentation Criteria – What to Segment On?

1. Geographic Segmentation

- Based on **location** (e.g., country, city, region).
- Advantages:
  - Simple and cost-effective.
  - Easy to target via local media (radio, newspapers, etc.).
- Disadvantages:
  - Location may not explain actual product preferences.
  - People in the same area can have widely different interests and behaviors.
- **Example**: Austrian tourism targeting neighboring countries with different languages.

2. Socio-Demographic Segmentation

- Based on **age, gender, income, education, etc.**

- **Advantages**:
  - Easy to collect and assign.
  - Useful in industries like cosmetics, luxury goods, baby products.
- Disadvantages:
  - Often not the cause of buying behavior.
  - Limited insight into **motivations** or **preferences**.
  - Explains only ~5% of variance in behavior (Haley, 1985).

3. Psychographic Segmentation

- Based on **beliefs, values, interests, lifestyle, benefits sought**.
- Includes:
  - **Benefit segmentation** (e.g., relaxation, adventure)
  - **Lifestyle segmentation** (activities, opinions, interests)
- Advantages:
  - Reflects **true drivers** of behavior.
  - Better for understanding **motivations**.
- Disadvantages:
  - More complex to measure.
  - Requires **valid and reliable survey design**.

4. Behavioral Segmentation

- Based on **past actions or actual consumer behavior**.
  - E.g., frequency of purchase, amount spent, loyalty, product usage.
- Advantages:
  - Most accurate when based on **real behavior**.
  - Directly ties to consumer actions.
- Disadvantages:
  - Requires access to behavioral tracking data (e.g., purchase logs).

Survey Data Overview

General Points

- **Surveys are the most common data source** for market segmentation due to low cost and ease of collection.
- However, **surveys are prone to biases**, which can distort segmentation results.
- Success in segmentation heavily depends on:
  - The choice of variables
  - The **type of response options**
  - Managing **response styles** and biases

Choice of Variables

- **Careful variable selection** is critical—irrelevant variables hurt segmentation quality.
- Include only variables **relevant to the segmentation criterion**.
- Unnecessary or noisy variables:
  - Increase respondent fatigue → lower data quality
  - Add irrelevant dimensions → confuse algorithms
- **Masking variables** divert attention from relevant patterns, reducing segmentation accuracy.
- Avoid **redundant questions** that inflate the data without adding value.
- Combine **qualitative (exploratory)** and **quantitative (survey)** research for effective questionnaire design.

Response Options

- **Binary (0/1)** and **metric (numeric)** responses are ideal for segmentation analysis:
  - Easy to interpret
  - Compatible with distance-based algorithms
- **Ordinal data** (e.g., Likert scales) is harder to interpret and may distort distance measures.
- **Visual analogue scales** (e.g., sliders) help create metric-like data from subjective opinions.

- Best practice:
  - Use binary or metric when possible.
  - Avoid unnecessary use of ordinal scales unless nuances are critical.

Response Styles

- **Response styles = systematic biases** in how respondents answer (regardless of question content).
  - Examples:
    - Acquiescence bias (tendency to agree with everything)
    - Extreme responding (always choosing strong options)
    - Midpoint bias (always selecting neutral)
- These styles can **distort segmentation** by creating false segments.
  - E.g., a segment that always says "yes" may look like high spenders, but could be due to bias.

- Prevention & Mitigation:
  - Design surveys to **minimize response style risk**.
  - **Investigate unusual segment patterns** that may be caused by response styles.
  - Consider **removing biased respondents** or conducting **additional analyses** to validate findings.

Sample Size

- **Unlike other statistical analyses**, market segmentation lacks standard sample size guidelines.
- **Sample size directly affects** the ability to:
  - Detect the correct number of segments.
  - Accurately define the nature of each segment.
- Even with just **two segmentation variables**, insufficient sample size makes it difficult to extract valid segments (as shown in Fig. 5.1).

Sample Size Recommendations

- Formann (1984):
  - Minimum = **2^p** (where *p* is number of segmentation variables).
  - Preferably = **5 × 2^p**.
  - Best for **latent class analysis** using binary variables (not generalizable).

- Qiu and Joe (2015):
  - Recommended sample size = **10 × p × k**
    - *p* = number of segmentation variables

        - k = number of expected segments
    - For **unequal segments**, smallest segment should have **at least 10 × p** observations.
    - Used realistic artificial datasets (e.g., from tourism research).
    - Segment recovery measured with Adjusted Rand Index:
        - ARI = 1 → perfect match
        - ARI = 0 → match is no better than random
    - **Larger sample sizes improve accuracy** of segment recovery.

Key Insight

- **Segment recovery accuracy improves** consistently with larger sample sizes.
- Ensuring an adequate number of observations is **crucial for valid segmentation**—especially as the number of variables or segments increases.

Factors Affecting Segment Recover

Market Characteristics That Increase Difficulty

- **Unequal segment sizes** → harder to detect correct segments.
- **Overlapping segments** → lowers accuracy of segment extraction.

Data Characteristics That Influence Accuracy

- **Correlated segmentation variables** severely impair recovery.
    - Even large samples can't fully compensate.

- **Noisy variables** (irrelevant or redundant) slightly degrade performance.
- **Measurement type** matters:
    - **Binary or metric** data is best.
- **Noisy respondents** (with inconsistent answers) and **random replacements** reduce quality.
- **Uncorrelated data** yields better results even with smaller samples.
- **Correlated data** performs poorly regardless of sample size.
- Quality and structure of the data often **matter more than quantity alone**.

Best Practice Checklist for Survey-Based Segmentation Data

To ensure optimal segmentation results, your dataset should:

- Include **all necessary** segmentation variables
- Exclude **unnecessary or redundant** variables
- Avoid correlated segmentation variables
- Have **high-quality**, consistent responses
- Use **binary or metric** measurement scales
- Be free from **response styles** or biases
- Come from a **relevant and well-defined sample**
- Include **at least 100× the number of segmentation variables** in sample siz

Step 4: Exploring Data- Bandaru Chandra Mouli

What is Exploratory Data Analysis (EDA) in Market Segmentation?

After you collect your data, the **first step** is to understand and prepare the data for segmentation analysis. This involves:

1. **Identifying Measurement Levels:**
    - Understanding whet`her variables are nominal, ordinal, interval, or ratio scales.
    - This matters because different algorithms require data in specific formats.
2. **Investigating Univariate Distributions:**
    - Checking the distribution of each variable individually (e.g., mean, median, spread, skewness).
    - Identifying outliers or strange values that might distort the segmentation.
3. **Assessing Dependency Structures:**
    - Exploring relationships between variables, like correlations or dependencies.
    - This helps decide if some variables are redundant or if they group naturally, guiding variable selection.
4. **Preprocessing Data:**
    - Cleaning (handling missing values, removing errors).
    - Transforming variables if necessary (e.g., scaling, encoding categorical variables).
    - Preparing the data in a way that fits the input requirements of segmentation algorithms.

Example Used: Travel Motives Data Set

- The example involves 1000 Australian residents.
- Each reported on **20 travel motives** related to their last vacation.
- An example motive: *"I AM INTERESTED IN THE LIFESTYLE OF LOCAL PEOPLE."*
- The data is stored as a CSV file and can be loaded through an R package called **MSA**.
- Appendix C.4 provides detailed information about this data.

Data Cleaning: The Essential First Step

1. **Check for Correct Values:**
    - Verify that every recorded value makes sense.
    - For example, **numeric variables** like age should be within a plausible range — like 0 to 110 years.
    - Any value outside this range likely indicates errors during data collection or entry and needs correction.
2. **Check Consistency of Categorical Variables:**
    - Ensure categories (labels) are consistent and only contain allowed values.
    - For instance, **gender** often has two categories: *male* and *female*.
    - If the survey did not provide a third gender option, any other recorded values should be identified and corrected.

Key Points:

1. Reproducible Data Cleaning Using Code

- Variables like **Income** can be reordered or transformed programmatically.
- **Always keep your cleaning and transformation steps in code** rather than manual spreadsheet edits.
- Coding your cleaning steps:

    - Ensures full **documentation** of all transformations.
    - Makes your work **reproducible** by others or yourself later.
    - Facilitates reusing the same steps when new data arrives (ongoing monitoring or updating).
- In R, save cleaned data frames with save() and reload them anytime using load().

2. Descriptive Analysis

- **Know your data** before diving into complex analysis to avoid misinterpretation.
- Use numeric summaries and graphical methods to understand your variables.
- In R:
    - summary() provides:
        - For numeric variables: min, max, quartiles, mean.
        - For categorical variables: frequency counts.
        - Also reports missing values.
- **Useful graphs for numeric data**:
    - Histograms (show distribution and shape),
    - Boxplots (show spread and outliers),
    - Scatter plots (show relationships between variables).
- **Useful graphs for categorical data**:
    - Bar plots (frequency of categories),
    - Mosaic plots (show association between categorical variables).
- **Histograms** involve binning data into intervals (bins) of equal length and plotting frequency.
- The R package **lattice** is good for creating segmented histograms, which helps in comparing data subsets or segments.

Histograms and Bin Width

- Finer bins reveal more detailed patterns in data distribution (e.g., **bi-modal** age distribution with peaks around 35–40 and 60 years).
- Using type = "density" in histogram plotting rescales the y-axis to show **density estimates** (area under bars sums to 1).
- This allows overlaying parametric distributions (e.g., normal curve).
- Example insight: About **three-quarters of respondents are younger than 57**.
- Extreme values (like a respondent aged 105) could indicate outliers or data entry errors.

2. Box-and-Whisker Plot (Boxplot)

- Visualizes five-number summary: minimum, first quartile (Q1), median, third quartile (Q3), and maximum.
- The **box** covers Q1 to Q3 (middle 50% of data), with a line inside showing the median.
- The **whiskers** extend to the smallest and largest values within 1.5 times the interquartile range (IQR = Q3 - Q1).
- Values outside whiskers (e.g., 105-year-old) are plotted as **outliers** (shown as circles).
- Boxplot reveals skewness; if the median is closer to Q1 or Q3, distribution is skewed.
- In this case, the median is closer to Q1, so **age distribution is right-skewed** (more younger respondents).
- Boxplots are good for showing **unimodal distributions** and avoiding manual bin selection like histograms require.
- Despite their usefulness, boxplots are underused in business and social sciences compared to natural sciences.

3. Outliers

- The 105-year-old respondent skews the whisker length.
- Most software (including R) limits whiskers to 1.5 times the box length to reduce outlier impact.

- Outliers are important to detect but shown separately so the main data shape is clear.

Key Points about PCA:

- **What is PCA?**
  PCA is a statistical technique that transforms a set of possibly correlated metric variables into a new set of uncorrelated variables called **principal components** (PCs).
  - The PCs are ordered by the amount of variability they explain in the data.
  - The first PC explains the most variance, the second PC explains the second most, and so forth.
  - The number of PCs equals the original number of variables.
  - PCA essentially changes the "angle" or perspective on the data but keeps the relative positions of observations unchanged.
- **How PCA Works:**
  PCA is done by analyzing the **covariance matrix** or the **correlation matrix** of the numeric variables.
  - Use the **covariance matrix** if all variables are on the same scale with similar ranges.
  - Use the **correlation matrix** (equivalent to standardizing the data) if variables have different scales/ranges.
- **Purpose of PCA:**
  - Mainly used to **reduce dimensionality** for visualization or analysis.
  - Instead of using all variables, often only the first few PCs are used because they capture most of the information (variance).\
  - For example, the first two PCs can be plotted in a scatter plot to visualize data structure.
- **Example in R:**

r

CopyEdit

```
vacmot.pca <- prcomp(vacmot)
```

- Here, vacmot is the travel motives dataset.
- prcomp() centers the data by default but does not standardize unless specified. Since variables are binary here, standardization is not necessary.
- **Interpretation of PCs:**
- The **rotation matrix** shows how original variables contribute (load) to each PC.
- In the example, **PC1** contrasts two answer patterns:
    - "Almost no motives apply" vs. "All motives apply."
    - This PC may not be very useful for managerial decisions.
- **PC2** loads highly on variables related to FUN, ENTERTAINMENT, LUXURY, and ENVIRONMENTAL CARE.
- **PC3** is associated with budget constraints, cultural offers, and interest in the local lifestyle.
- **Next Steps:**
- You can get more details about the PCA results with functions like summary() or by examining the rotation and sdev components of the PCA object.

Summary

PCA helps simplify complex datasets by creating new variables (principal components) that summarize the main patterns in data. It’s particularly useful for visualization and for reducing noise before applying segmentation algorithms.

Step 5 - Extracting Segments - Eranti Venkata Sai Dheeraj

Market segmentation is a cornerstone of modern marketing strategy, enabling businesses to divide a diverse consumer base into smaller, more homogeneous groups based on shared characteristics, preferences, or behaviors. This process allows organizations to tailor their products, services, and marketing efforts to specific segments, thereby enhancing efficiency and effectiveness. Chapter 7 of the document, titled "Step 5: Extracting Segments," delves into the technical aspects of identifying and extracting these segments from consumer data, particularly in the context of data-driven market segmentation. This approach is inherently exploratory due to the unstructured nature of consumer data, which often lacks clear, predefined groups. Instead, consumer preferences are typically scattered across a multidimensional space, resembling a continuous distribution rather than distinct clusters.

The chapter emphasizes that the outcome of a market segmentation analysis is influenced by two primary factors: the underlying consumer data and the algorithm used to extract segments. Different algorithms impose distinct structural assumptions on the data, shaping the resulting segments in unique ways. For instance, some algorithms favor compact, spherical clusters, while others can identify elongated or irregular shapes. This interaction between data and algorithm underscores the importance of carefully selecting a segmentation method that aligns with the data's characteristics and the researcher's objectives. The document provides a framework for this selection through Table 7.1, which outlines key data set characteristics (e.g., size, scale level, special structures) and segment characteristics (e.g., within-segment similarities, between-segment differences, number, and size of segments) that inform algorithm choice.

Exploratory Nature of Data-Driven Segmentation

Data-driven market segmentation relies on exploratory methods to uncover patterns in consumer data. Unlike commonsense segmentation, where segments are predefined (e.g., based on age or geographic location), data-driven segmentation requires the analyst to discover segments based on patterns in the data, such as preferences, behaviors, or needs. The document illustrates this challenge with an example from Fig. 7.1, where a dataset containing two spiraling segments is analyzed using two different clustering algorithms: k-means and single linkage hierarchical clustering. The k-means algorithm, which seeks compact, round clusters, fails to identify the spiral structure, instead grouping consumers based on proximity in Euclidean space. In contrast, single linkage hierarchical clustering, which constructs snake-shaped clusters, accurately captures the spiral segments. This example highlights a critical point: the choice of algorithm significantly influences the segmentation solution, and no single method is universally superior. The effectiveness of an algorithm depends on how well its assumptions match the data's underlying structure.

The document stresses that consumer data is often not well-structured, meaning that natural, distinct segments may not exist. In such cases, the algorithm imposes a structure that aligns with its objective function, potentially creating artificial segments. This makes it essential to explore multiple segmentation methods and compare their results to identify the most robust and meaningful solution. The exploratory nature of the process also necessitates careful consideration of the data's characteristics, such as the number of consumers, the number and type of segmentation variables, and any special structures (e.g., longitudinal data).

Clustering Methods for Segment Extraction

The document categorizes segmentation methods into two main groups: distance-based methods and model-based methods, with additional specialized algorithms that address multiple objectives, such as simultaneous variable selection and segment extraction. Below, we explore each category in detail, focusing on their methodologies, applications, and limitations.

Distance-Based Methods

Distance-based methods are the most commonly used approaches in market segmentation, as they group consumers based on a measure of similarity or dissimilarity, typically defined by a distance metric. The goal is to form segments where consumers within the same segment are as similar as possible, while those in different segments are as dissimilar as possible. The document discusses several key distance measures:

- Euclidean Distance: This is the most widely used distance metric, representing the straight-line distance between two points in multidimensional space. It is intuitive and effective for continuous data but assumes that all dimensions contribute equally to the distance calculation.
- Manhattan Distance: Also known as absolute distance, this metric sums the absolute differences along each dimension, resembling travel along a grid (e.g., city streets). It is less sensitive to outliers than Euclidean distance and can be more appropriate for data with different scales.
- Asymmetric Binary Distance: Designed for binary data (e.g., 0s and 1s indicating the presence or absence of an attribute), this measure focuses on shared positive attributes (e.g., common vacation activities). It is particularly useful when the absence of an attribute is less informative than its presence, as in the case of rare activities like horseback riding.

To illustrate these measures, the document presents a fictitious dataset of seven tourists (Table 7.2), with each tourist's preferences for three vacation activities: beach, action, and culture, expressed as percentages of time spent. For example, Anna and Bill spend 100% of their time on the beach, while Michael focuses on action (90%) and culture (10%). Using R, the document demonstrates how to compute Euclidean and Manhattan distances between these tourists. The results show that Anna and Bill have a distance of zero (identical profiles), while Michael is significantly distant from others due to his unique preferences. This example underscores the importance of choosing an appropriate distance measure based on the data's scale and structure.

Hierarchical Clustering

Hierarchical clustering is an intuitive approach for smaller datasets (up to a few hundred observations), as it mimics how a human might manually group consumers. It can be divisive (starting with one cluster and splitting) or agglomerative (starting with individual consumers and merging). Both approaches produce a sequence of nested partitions, visualized as a dendrogram, which shows the hierarchy of mergers or splits.

The document discusses several linkage methods that define how distances between clusters are calculated:

- Single Linkage: Measures the distance between the closest pair of observations in two clusters. It excels at identifying non-convex structures (e.g., spirals) but can suffer from chaining effects, where clusters are merged based on a single close pair, leading to undesirable groupings in poorly separated data.
- Complete Linkage: Uses the distance between the farthest pair of observations, producing compact, well-separated clusters.
- Average Linkage: Calculates the average distance between all pairs of observations in two clusters, balancing single and complete linkage.

- Ward's Method: Minimizes the weighted squared Euclidean distance between cluster centroids, favoring compact clusters. It requires careful specification of whether Euclidean or squared Euclidean distance is used.

The tourist activities dataset is used to illustrate hierarchical clustering with single and complete linkage, producing dendrograms (Fig. 7.4). The results show that Anna and Bill are grouped together first due to their identical profiles, while Michael is merged last due to his distinct preferences. Another example involves a larger dataset on tourist risk-taking (563 respondents rating six risk categories: recreational, health, career, financial, safety, and social). Using Manhattan distance and complete linkage, the document demonstrates how to extract segments and visualize them through a dendrogram (Fig. 7.5) and a bar chart of cluster means (Fig. 7.6). The bar chart highlights segment characteristics, such as risk-averse segments and those with specific risk proclivities (e.g., social or career risks).

Partitioning Methods

For larger datasets (>1000 observations), hierarchical clustering becomes computationally intensive due to the need to calculate all pairwise distances. Partitioning methods, such as k-means, are more efficient, as they compute distances between each consumer and cluster centroids. The k-means algorithm is the most popular partitioning method and involves five steps:

1. Specify the number of segments (k).
2. Randomly select k observations as initial centroids.
3. Assign each observation to the nearest centroid based on a distance measure (typically squared Euclidean).
4. Recompute centroids as the mean of assigned observations.
5. Repeat until convergence or a maximum number of iterations is reached.

The document emphasizes that k-means is iterative and sensitive to initial centroid selection, which is random. This randomness necessitates repeated calculations to assess solution stability. An example using an artificial mobile phone dataset (500 consumers, two variables: number of features desired and price willingness to pay) illustrates k-means clustering. The data, generated with three distinct segments (low-end, mid-range, high-end), is visualized in a scatter plot (Fig. 7.9). Using the cclust() function from the R package flexclust, the document extracts three segments, confirming their distinct characteristics (e.g., high-end consumers want many features and are willing to pay more).

The choice of distance measure significantly impacts the segmentation solution. The document demonstrates this with a Gaussian dataset clustered using squared Euclidean, Manhattan, and angle distances, resulting in different cluster shapes (Fig. 7.8). Squared Euclidean distance produces diagonal borders, Manhattan distance yields axis-parallel borders, and angle distance creates cake-piece-shaped segments. This highlights the need to test multiple distance measures to understand their impact on the segmentation solution.

Model-Based Methods

Model-based methods assume that the data follows a specific statistical distribution (e.g., normal mixtures) and formulate a stochastic model for the segments. These methods are particularly useful for complex data structures, such as longitudinal data or when segments are defined by latent variables (e.g., price sensitivity inferred from purchase histories). For example, regression models can be used to segment consumers based on price sensitivity if purchase histories and price data are available. The document briefly mentions finite mixture models and regression-based approaches but focuses primarily on distance-based methods, as they are more commonly used in market segmentation.

Specialized Algorithms

Specialized algorithms, such as biclustering, address multiple objectives simultaneously, such as variable selection and segment extraction. Biclustering is particularly effective for binary data, focusing on shared positive attributes (e.g., common vacation activities). This is useful when the dataset contains many variables, but only a subset is relevant for segmentation. For instance, in the tourist activities example, biclustering could identify segments based on shared activities like horseback riding, ignoring the absence of such activities, which is less informative.

Data and Segment Characteristics

Table 7.1 provides a comprehensive guide for selecting segmentation algorithms based on data and segment characteristics:

- Data Set Characteristics:
  - Size: The number of consumers and segmentation variables affects algorithm choice. Larger samples allow finer segmentation but require sufficient observations to support the expected number and size of segments. The document notes that minimum segment size is a knock-out criterion defined in Step 2 of the segmentation process.
  - Scale Level: Variables can be nominal, ordinal, metric, or mixed, influencing the choice of distance measure or model-based approach. For example, binary data may require asymmetric distance measures, while metric data suits Euclidean or Manhattan distances.
  - Special Structure: Longitudinal data or repeated measurements require algorithms that account for temporal dependencies, often necessitating model-based approaches.
- Segment Characteristics:
  - Similarities Within Segments: Consumers in the same segment should share key characteristics, such as benefits sought or behavioral patterns. These characteristics are defined in Step 2 and must align with the algorithm's output.
  - Differences Between Segments: Segments should be sufficiently distinct to justify targeted marketing strategies. For example, in the tourist activities dataset, Michael's unique preference for action and culture differentiates him from beach-focused tourists.
  - Number and Size: The desired number and size of segments depend on managerial objectives (e.g., targeting a niche vs. a mass market). Larger samples are needed for niche segments to ensure statistical reliability.

For binary data, the document discusses the importance of symmetric vs. asymmetric treatment. Symmetric treatment (e.g., using Manhattan distance) considers both shared 0s and 1s, while asymmetric treatment (e.g., in biclustering) focuses on shared 1s, which is more relevant for sparse data. For example, in the tourist activities context, it is more informative that two tourists both engage in horseback riding than that they both do not.

Stability Analysis

Given the exploratory nature of data-driven segmentation and the sensitivity of algorithms like k-means to random initialization, stability analysis is crucial for validating segmentation solutions. The document introduces two stability measures proposed by Dolnicar and Leisch (2017):

- Segment Level Stability Within Solutions (SLS_W): This measures how consistently a segment's characteristics are identified across repeated calculations with the same number of segments. It uses bootstrap sampling to create multiple datasets, applies the clustering algorithm to each, and computes the maximum agreement with the original segments using the Jaccard index. The Jaccard index is the ratio of the intersection to the union of two segments, with higher values indicating greater stability. The process involves:
    1. Computing a k-segment solution for the original data.
    2. Drawing b bootstrap samples (e.g., b=100).
    3. Clustering each bootstrap sample into k segments.
    4. Computing the Jaccard index for each bootstrap segment against the original segments.
    5. Visualizing the results as boxplots to assess stability.
- Segment Level Stability Across Solutions (SLS_A): This evaluates whether a segment reappears across solutions with different numbers of segments (e.g., from 3 to 8 segments). Stable segments across multiple solutions are more likely to be natural rather than algorithm-imposed. The document uses entropy to quantify SLS_A, where lower entropy indicates higher stability. The process involves:
    6. Computing a series of segmentation solutions with k_min to k_max segments.
    7. Renumbering segments to ensure consistent labeling across solutions using the relabel() function in flexclust.
    8. Calculating the proportion of consumers each segment recruits from segments in the previous solution.
    9. Computing entropy as a measure of stability, with SLS_A = 1 - (entropy / maximum entropy).

The document illustrates these measures with two examples:

- Artificial Mobile Phone Data: This dataset, with 500 consumers and three distinct segments (low-end, mid-range, high-end), is clustered into 3 to 8 segments. The three-segment solution is highly stable (SLS_W = 1), as shown in Fig. 7.42, reflecting the natural clusters. The six-segment solution shows lower stability for some segments, indicating artificial splits. The SLS_A plot (Fig. 7.44) confirms that the high-end segment remains stable across solutions, while others are subdivided, suggesting it is a natural segment.
- Australian Travel Motives Data: This dataset, with 1000 respondents rating 20 travel motives, is clustered into 3 to 8 segments using the neural gas algorithm. The six-segment solution reveals that segments 1, 5, and 6 are stable (SLS_W, Fig. 7.43), but segment 3 is unstable, as it recruits members from multiple segments in the five-segment solution and disappears in the seven-segment solution (SLS_A, Fig. 7.45). Segment 6, interested in local lifestyles and nature, is identified as a potential target segment due to its high stability and meaningful profile.

Practical Implementation in R

The document provides detailed R code examples for implementing clustering and stability analysis, primarily using the flexclust and cluster packages. Key functions include:

- dist(): Computes pairwise distances (e.g., Euclidean, Manhattan) between observations.
- hclust(): Performs hierarchical clustering with specified linkage methods.

- cclust(): Implements k-means and other partitioning methods, offering richer output for visualization.
- slswFlexclust(): Calculates SLS_W by comparing bootstrap samples to the original solution.
- slsaplot(): Generates SLS_A plots to visualize segment stability across different numbers of segments.

For example, the tourist risk-taking dataset is analyzed using dist(risk, method = "manhattan") and hclust(risk.dist, method = "complete"), with results visualized as a dendrogram and bar chart. The mobile phone dataset uses cclust(PF3, k = 3) to extract three segments, with stability assessed using slswFlexclust() and slsaplot().

Conclusion

Market segmentation is a powerful strategy for enhancing customer relevance, optimizing resources, and achieving competitive advantage, but it requires strategic commitment, robust data, and meticulous execution. Step 1 ensures organizational readiness through audits and stakeholder alignment, addressing barriers like resource constraints and cultural resistance. Step 2 defines viable segments using knock-out and attractiveness criteria, prioritizing those with high potential and organizational fit. Step 3 collects high-quality data, balancing simplicity and depth to support segmentation. Step 5 extracts meaningful segments through clustering techniques, with algorithm choice and stability analysis critical to reliability. By leveraging practical tools, addressing challenges, and following a structured timeline, organizations can implement effective segmentation strategies that drive growth and market success.

Step 6: Profiling Segments – G.K.S Jyoteesh

- Profiling begins by characterizing clusters on the very variables used to form them, reinforcing the validity of each segment's defining features.
- Initial cross-tabulations and univariate statistics offer quick sanity checks but risk oversimplifying multidimensional differences.
- Segment profile plots, which juxtapose segment means against overall means, vividly reveal the variables that most sharply differentiate groups.
- Gorge plots condense variable importance comparisons across segments into an intuitive graphical format.
- Separation plots—built on PCA or other dimensionality-reduction techniques—overlay segment hulls to visualize overlap and distinctness.
- Quantitative separation metrics (silhouette score, Davies–Bouldin index) complement visual tools and provide objective validation.
- Highlighting marker variables directly on visuals focuses stakeholder attention on the most salient segment differentiators.
- Interactive dashboards or drill-down storyboards enhance stakeholder engagement and facilitate ad-hoc exploration of segment profiles.
- Testing profile stability on hold-out samples or via alternate clustering methods guards against overfitting.
- Supplementing profiles with secondary market data—such as industry reports—enriches the narrative and external validity.
- Using significance tests sparingly—since segments were optimized for maximal difference—avoids inflated Type I error risks.
- Documenting profiling methodologies and subjective choices (e.g. visualization thresholds) preserves reproducibility and audit trails.
- Crafting concise segment "snapshots" that summarize quantitative and qualitative attributes facilitates senior-level decision making.
- Leading stakeholder workshops to walk through profiles builds shared understanding and consensus around segment characteristics.
- Aligning output formats with organizational reporting standards—whether slide decks, dashboards or written reports—streamlines communication.
- Translating profiling insights into preliminary targeting hypotheses (e.g. messaging themes, channel preferences) accelerates downstream activation.
- Iterating on visuals based on stakeholder feedback ensures clarity without sacrificing analytical depth.
- Conducting a final coherence check confirms each profile tells a distinct, actionable story.
- Archiving all analytic code, transformation logs and datasets underpins future updates and governance requirements.
- Presenting profiled segments formally to executive leadership secures endorsement before investing in descriptive enrichment.

Step 7: Describing Segments – G.K.S Jyoteesh

- Enriching segment profiles with descriptor variables—such as media habits, purchase histories or attitudinal scales—not previously used in extraction deepens understanding.
- Mosaic plots elegantly display categorical descriptors, using cell shading to indicate significant over- or under-representation relative to expectations.
- Recognizing that mosaic plot widths correspond to segment sizes and heights to category shares provides immediate visual intuition.
- Parallel box-and-whisker plots with varwidths and notches compare distributions of continuous descriptors—like expenditure or psychometric scores—across segments.
- Histograms serve as a first pass to verify descriptor distributions prior to more sophisticated visualizations.
- Avoiding overuse of significance tests on descriptors—segments were optimized for maximal difference—prevents misleading p-value interpretations.
- Logistic regression models predict segment membership from descriptors, offering interpretable targeting rules and membership probabilities.
- Tree-based algorithms (CART, random forests) uncover non-linear interactions among descriptors that delineate segments in novel ways.
- Integrating descriptive visuals and predictive models into cohesive narratives illustrates both who the segments are and how they can be reached.
- Annotating each segment description with verbatim customer quotes or illustrative case vignettes humanizes the data.
- Writing narrative summaries that interweave quantitative findings with strategic recommendations makes insights accessible to all stakeholders.
- Ensuring descriptor visualizations include clear captions, legends and methodological notes enhances interpretability for non-technical audiences.
- Validating descriptor-driven targeting rules via small-scale pilots reduces risk before large-scale roll-outs.
- Establishing feedback loops from sales, service and field teams refines segment descriptions based on real-world interactions.
- Periodically refreshing descriptor data—through ongoing tracking studies or CRM updates—captures evolving customer behaviors.
- Embedding enriched segment definitions within CRM and marketing automation platforms enables personalized engagements at scale.
- Incorporating external data sources (social listening, syndicated studies) supplements internal descriptors and surfaces emerging trends.
- Translating segment descriptions into audience personas guides creative ideation, media planning and campaign execution.
- Documenting descriptor-selection rationale, model specifications and caveats ensure transparency for downstream users and future audits.
- Producing a final, comprehensive playbook that prescribes targeting, positioning and messaging strategies for each segment operationalizes the entire segmentation framework

Step 8. Selecting (the) Target Segment(s) - Anandhu S

Selecting the Target Segments

- **Critical decision point:** Selecting one or more market segments for long-term targeting.
- Builds on previous steps where segments have been extracted, profiled, and described.
- Use **knock-out criteria** to eliminate unsuitable segments (e.g., too small, not reachable, needs cannot be met).
- Evaluate remaining segments based on:
    1. **Organizational preference:** Which segment(s) does the company want to serve?
    2. **Competitive position:** How likely is the organization to be chosen by that segment?

Market Segment Evaluation

- Use a **decision matrix** to visualize:
    - **Segment attractiveness** (how attractive the segment is to the organization).
    - **Relative organizational competitiveness** (how attractive the organization is to the segment).
- Common decision matrix types:
    - Boston Matrix
    - General Electric / McKinsey Matrix
    - Directional Policy Matrix
    - Market Attractiveness–Business Strength Matrix
- Segments are plotted as circles; circle size can represent other factors (e.g., revenue potential or loyalty).
- No single best measure of attractiveness or competitiveness — depends on organizational priorities.
- Segment attractiveness criteria and weights defined earlier (in Step 2) guide evaluation.
- Assign values for each criterion per segment to calculate overall attractiveness scores.

Checklist – Segment Evaluation and Selection

- Convene the **segmentation team meeting** to review potential target segments.
- Identify which segments profiled (Step 6) and described (Step 7) are still under consideration.
- **Verify knock-out criteria** for each segment, including:
    - Homogeneity
    - Distinctness

- Size
- Match with organizational capabilities
- Identifiability
- Reachability

→ Eliminate any segment that fails these criteria.

- Discuss and agree on **values for each segment attractiveness criterion** per segment.
- Discuss and agree on **values for each relative organizational competitiveness criterion** per segment.
- Calculate each segment's **overall attractiveness**:
    - Multiply criterion values by their weights.
    - Sum the weighted values for each segment.
- Calculate each segment's **overall relative organizational competitiveness** similarly.
- **Plot segments** on an evaluation chart (attractiveness vs. competitiveness) with bubble sizes representing factors like profit potential.
- Make a **preliminary selection** of target segment(s).
- If targeting multiple segments, ensure the **selected segments are compatible** and do not conflict strategically.
- Present the **selected target segments to the advisory committee** for feedback and possible revision.

Step-9. Customising the Marketing Mix - Deepak Kumar Prajapati

1. **Decide Whether to Segment**
   - Check for readiness, resources, and long-term commitment.
2. **Define Ideal Segments**
   - Identify what characteristics your best customers have.
3. **Collect Data**
   - Use surveys, internal databases, loyalty programs, etc.
4. **Explore the Data**
   - Clean, analyze, and prepare data for segmentation.
5. **Extract Segments**
   - Use criteria to group customers meaningfully.
6. **Profile Segments**
   - Describe what makes each group unique.
7. **Describe Segments**
   - Use charts and metrics to present the segments clearly.
8. **Select Target Segments**
   - Evaluate based on size, fit, reachability, and strategic potential.
9. **Tailor the Marketing Mix**
   - Adjust product, price, place, and promotion to suit target segments.
10. **Evaluate and Monitor**
    - Regularly check performance and adapt as needed.

Target Segment Selection Considerations

- **Organizational Fit**: Is this segment right for your brand?
- **Competitive Position**: Can you serve this segment better than competitors?

Evaluating Segments

Use tools like a **Decision Matrix** (e.g., GE/McKinsey Matrix) to compare:

- **Segment Attractiveness**
- **Organizational Competitiveness**

Score each segment and visualize using charts or matrices.

Knock-Out Criteria

Eliminate segments that:

- Are too small
- Can’t be reached
- Don’t align with your business
- Lack clear distinction from other segments

Using Survey Data for Segmentation

- **Data Quality**: Reliable data is essential. Poor data = misleading segments.
- **Variable Relevance**: Only include useful, meaningful variables.
- **Bias Prevention**: Watch for patterns like "always agreeing" or "neutral-only" responses.
- **Sample Size**: More data = more reliable segments.