Ten Steps of Market Segmentation Analysis

Name: Rohit Ingole

Step 1: Deciding (not) to Segment

The market segmentation strategy and its execution issues are addressed in this paragraph. The firm must make a sustained commitment to market segmentation, which includes making investments in new products, changing prices, altering distribution, and improving communication. Support from senior management is essential for segmentation to be successful. The process may be hampered by a lack of resources, market focus, or leadership. Before beginning the study, organizations should carefully assess the ramifications of segmentation. If the problems don't seem to be solved, it might be better to give up on segmentation.

Step 2: Specifying the Ideal Target Segment

Segment Evaluation Criteria

It underscores that user involvement should extend beyond initial briefings or final marketing mix development, spanning various stages of the analysis. In Step 2, the organization plays a vital role in defining segment evaluation criteria, including

1. Knock-out criteria
2. attractiveness criteria.

These criteria guide subsequent steps, such as data collection and target segment selection. While the literature offers numerous segment evaluation criteria, it often doesn't distinguish between essential knock-out criteria and relative attractiveness criteria.

1. knock-out criteria

The concept of knock-out criteria is introduced in the work. These criteria are used to evaluate market segments that come from segmentation analysis and determine if they are eligible for additional evaluation based on segment attractiveness.

- Homogeneous: Members of the segment should be similar to each other.
- Distinct: Members of the segment should be different from those in other segments.
- Large enough: The segment should be big enough to justify targeting it.
- Matching strengths: The organization's capabilities should align with the segment's needs.
- Identifiable: Segment members must be easy to identify.
- Reachable: There must be a way to reach them with marketing messages.

2. Attractiveness Criteria

- Attractiveness criteria are not binary but involve rating each segment based on various factors like size, growth potential, profitability, and compatibility with the organization's capabilities.
- These ratings help determine the overall attractiveness of a segment, guiding the selection of target segments.
- The text emphasizes that attractiveness criteria are diverse and can be tailored to suit specific circumstances, aiding in the strategic decision-making process of market segmentation.

Step 3: Collecting Data

3.1 Segmentation Variables:-

- Empirical data is fundamental for both commonsense and data-driven market segmentation, used to identify and describe market segments.
- In commonsense segmentation, a single characteristic (such as gender) is typically used as the segmentation variable to divide the sample into segments.
- Descriptor variables, including socio-demographics and consumer behavior information, are other personal characteristics used to describe segments in detail and develop effective marketing strategies.
- Commonsense segmentation example: using gender as the segmentation variable to create segments of men and women.
- Data-driven segmentation employs multiple segmentation variables to identify naturally existing or artificially created market segments.
- Descriptor variables are essential for understanding and targeting specific segments with tailored marketing mixes and communication messages.
- Market segmentation based on data analysis is focused on identifying segments of consumers who share common benefits sought rather than traditional demographic characteristics like gender.
- Quality empirical data is crucial for developing accurate segmentation solutions.
- The benefits sought represent segmentation variables, while socio-demographic variables like gender, age, and frequency of vacations serve as descriptor variables.
- Good data quality ensures accurate assignment of individuals to segments and correct description of segments.
- Accurate segment descriptions enable customized product development, appropriate pricing strategies, optimal distribution and communication channels for advertising.
- Data for segmentation studies can be sourced from surveys, observations (e.g., scanner data), or experimental studies, with preference for data reflecting actual consumer behavior.
- Survey data may not always accurately reflect behavior, especially for socially desirable actions, so other data sources should be considered.
- The most suitable data source is one that closely mirrors real consumer behavior.

3.2 Segmentation Criteria

- Organizations must choose a segmentation criterion which is more comprehensive than a segmentation variable before extracting segments.
- The term "segmentation criterion" describes the type of data geographic, sociodemographic, psychographic, or behavioral that is utilized in segmentation.
- It is difficult to outsource this choice and requires previous understanding of the market.
- Profitability, bargaining power, preferences for features or goods, obstacles to choice, and the consequences of customer contact are examples of common segmentation criteria.
- The marketing environment determines which segmentation criterion is best, and it is advised to choose the most straightforward strategy that is effective for the given good or service.
- Using demographic or geographic segmentation is advisable if they suffice for the product or service, as the goal is to find what works at the least cost.

3.1.1 Geographic Segmentation

- Geographic segmentation, using the consumer's location of residence, was one of the earliest segmentation criteria used for market segmentation.
- This approach is often simple and appropriate, especially for industries like tourism and global retail.
- For instance, language differences among neighboring countries make it practical to treat tourists from different countries as distinct segments.
- Companies like Amazon and IKEA tailor their offerings based on customers' geographic locations.
- Advantages include easy assignment of consumers to geographic units and targeted communication messages through local media channels.
- However, a key disadvantage is that living in the same area doesn't necessarily mean consumers share other relevant characteristics for marketers, such as product preferences.
- Despite limitations, geographic segmentation has seen a resurgence in international market segmentation studies.
- Extracting market segments across geographic boundaries presents challenges, including ensuring meaningful segmentation variables and addressing biases from different cultural backgrounds.

3.1.2 Socio-Demographic Segmentation

- The age, gender, income, and education categories are used in socio demographic segmentation.
- In sectors like luxury goods, cosmetics, infant items, retirement communities, and tourism, socio-demographic categories are helpful.
- Easy segment membership determination and possible explanations for certain product preferences (e.g., having children impacting vacation selections) are two benefits of socio-demographic segmentation.
- However, the value of socio-demographic parameters in segmentation decisions may be limited as they may not always be the reason for product preferences.
- Research indicates that a little portion about 5% of the variation in consumer behavior may be attributed to demographic factors.
- According to some analysts, values, tastes, and preferences have a greater impact on customers' purchasing decisions than sociodemographics, hence efficient market segmentation should prioritize these aspects.

3.1.3 Psychographic Segmentation

- Psychographic segmentation groups people based on psychological criteria such as beliefs, interests, preferences, aspirations, or benefits sought when purchasing a product.
- Benefit segmentation and lifestyle segmentation are popular types of psychographic segmentation.
- Psychographic criteria are complex because a single characteristic may not provide sufficient insight into the psychographic dimension of interest.
- Therefore, psychographic segmentation studies often use multiple segmentation variables, such as different travel motives or perceived risks when going on vacation.
- The psychographic approach is advantageous because it reflects the underlying reasons for differences in consumer behavior.
- For example, tourists motivated by cultural exploration are likely to prefer destinations rich in cultural treasures.
- However, the psychographic approach is more complex in determining segment memberships and relies heavily on the reliability and validity of empirical measures used to capture psychographic dimensions.

3.1.4 Behavioural Segmentation

- Behavioral segmentation involves grouping individuals based on their observed behaviors or reported behavior.
- Behaviors such as prior product experience, purchase frequency, amount spent, and information search behavior can be used for segmentation.
- Compared to geographic variables, behaviors reported by tourists have been found to be superior as segmentation variables.
- The key advantage of behavioral approaches is that they use actual behavior, which is often more predictive than stated or intended behavior.
- Examples include using actual expenses or purchase data as segmentation variables.
- Behavioral segmentation eliminates the need to develop measures for psychological constructs.
- However, behavioral data may not always be readily available, especially for potential customers who have not previously purchased the product.

3.3 Data from Survey Studies

- Survey data is often used in market segmentation analysis since it is easy to gather and reasonably priced.
- On the other hand, information gathered from surveys is more biased than information gathered from seeing real behavior.
- Survey data biases can have a negative effect on the caliber of solutions produced by market segmentation analysis.
- To ensure the quality and dependability of the results, businesses employing survey data for segmentation analysis must identify and deal with any biases.

3.3.1 Choice of Variables

- Selecting factors with care is necessary for both data-driven and commonsense segmentation strategies.
- All factors related to the segmentation criterion should be included in data-driven segmentation, whereas unnecessary variables must be avoided.
- Adding extraneous variables can make respondents tired and result in less accurate responses.
- Inadequate variable selection or badly written survey questions might result in noisy variables, which don't help identify the right market categories.
- The presence of noisy variables complicates the segmentation process and can lead to incorrect solutions.
- It is recommended to ask only necessary and unique questions, avoiding redundant items that interfere with segmentation algorithms.
- Developing a good questionnaire often involves a two-stage process of exploratory qualitative research followed by quantitative survey research to ensure no critical variables are omitted.

3.3.2 Response Options

- The size of data accessible for analysis is determined by survey response options, which influences the results of subsequent segmentation analyses.
- Data that are binary, or dichotomous, can be used for segmentation analysis since they are produced from options that provide respondents a choice between two alternatives and are represented as 0s and 1s.
- Binary data can be created from nominal variables, which are responses to a survey where participants choose one item from an ordered list.
- Metric data are appropriate for all statistical operations, including distance measurements, since they are produced from replies that contain numbers (such as age or the number of nights spent at a hotel).

- Because ordinal data lack clearly defined distances between options, segmentation analysis must carefully analyze these data, which are derived from replies with ordered but indeterminate distances between options (e.g Likert scales with five or seven possibilities).
- To prevent issues with distance measures in segmentation analysis, respondents should ideally be given binary or metric response options.
- Visual analogue scales are becoming more and more common, particularly in online surveys, as they enable respondents to identify a position along a continuous line and can offer metric data.
- It has been demonstrated that binary response options perform better than ordinal alternatives, especially when they are expressed in a level-free manner.

3.2.3 Psychographic Segmentation

- Psychographic segmentation is the process of grouping people according to psychological standards such as attitudes, hobbies, tastes, goals, and advantages desired while making a purchase.
- Haley is credited with developing benefit segmentation in 1968. Lifestyle segmentation, which takes into account an individual's interests, beliefs, and activities, is another popular psychographic segmentation technique.
- Compared to geographic or socio demographic criteria, psychographic criteria are more complicated and frequently need many variables to capture the various aspects of an individual's psychographics.
- Various reasons for traveling or anticipated hazards whilst on vacation are examples of common segmentation factors.
- By examining the underlying causes of variations in customer behavior, the psychographic technique offers deeper insights into consumer behavior.
- For example, in the tourism industry, knowing why people go, such as for cultural immersion, is useful for product creation and market segmentation.

3.2.4 Behavioural Segmentation

- Behavioral segmentation, which emphasizes similarities in claimed or actual behavior, is an alternative to psychographic segmentation.
- Behavioral segmentation makes use of a variety of activities, including past product experience, frequency of purchases, spending volume, and information search activity.
- Research has indicated that when it comes to segmentation, behavioral criteria—especially those reported by tourists—may be more effective than geographic characteristics.
- The primary benefit of behavioral segmentation is that it allows for more accurate segmentation since it is based on actual behavior rather than stated or anticipated behavior.

- Using customer spending actuals or purchase data across product categories as segmentation variables are two examples of behavioral segmentation.
- Numerous writers have also used brand choice behavior over time as a segmentation characteristic.
- It is no longer necessary to create metrics for psychological constructs thanks to behavioral segmentation.
- But getting behavioral data might not always be simple, particularly if you include in the segmentation study potential consumers who haven't bought the product yet.

3.4 Data from Survey Studies

- Market segmentation analyses are based on survey data.survey data is cheap and easy to collect, making it a feasible approach for any organization.
- Survey data is popular because it's affordable and easy to collect. However, people's responses in surveys can be biased, leading to inaccurate segmentation results.

3.4.1 Choice of Variables

- Both data-driven and commonsense segmentation depend on the careful selection of factors.
- All factors pertinent to the segmentation criterion must be included in data-driven segmentation, with superfluous variables being excluded.
- Adding extraneous factors might make segmentation analysis more difficult, cause respondent fatigue, and reduce answer quality.
- Noisy variables might make it more difficult for segmentation algorithms to get the right answer because they don't provide any useful information.
- Poorly crafted survey questions or a poor choice of segmentation factors can lead to noisy variables.
- It is crucial to ask pertinent and original questions and to avoid including unnecessary ones in order to prevent noisy variables.
- In survey research that adheres to standard psychometric concepts, redundant questions are frequently present and might cause issues for segmentation analysis.
- Insights that quantitative survey research alone cannot offer can be obtained through qualitative exploratory research, which also helps to guarantee that the questionnaire has all necessary factors.
- Using both qualitative and quantitative research methods in two stages makes it easier to create a thorough questionnaire that doesn't leave out any vitally crucial elements.

3.4.2 Response Options

- The options provided to respondents in surveys determine the scale of the data available for subsequent analyses, which affects segmentation analysis.
- Binary or dichotomous response options generate binary data (0s and 1s), suitable for segmentation analysis based on distance measures.
- Nominal variables, such as selecting an occupation from a list, correspond to unordered categories and can be transformed into binary data.
- Metric data, such as age or nights stayed at a hotel, allow for any statistical procedure, including distance measurement, making them well-suited for segmentation analysis.
- Ordinal data, generated by responses to statements with multiple ordered answer options, pose challenges for distance measures due to undefined distances between adjacent options.

- Preferably, binary or metric response options should be provided to respondents to avoid complications in segmentation analysis related to distance measures.
- Visual analogue scales, such as slider scales in online surveys, allow respondents to indicate a position along a continuous line and generate metric data.
- Binary response options often outperform ordinal ones, especially when formulated in a level-free way, without sacrificing capturing fine nuances in responses.

3.4.3 Response Styles

- Response bias in survey data refers to a systematic tendency to respond to questionnaire items based on factors other than the intended content. biases displayed by respondents over time, independent of specific survey questions.
- Various response styles exist, including tendencies to use extreme answer options, select the midpoint, or agree with all statements.
- Response styles can impact segmentation results because algorithms may not distinguish between genuine beliefs and response styles.
- For example, an acquiescence bias (tendency to agree with all questions) can inflate agreement levels in a segment, potentially misrepresenting it.
- It's crucial to minimize the risk of capturing response styles in data collected for market segmentation.
- Additional analyses may be needed to identify and exclude response styles from segmentation results or remove affected respondents before targeting a segment.

3.4.4 Sample Size

- Market segmentation analysis does not typically come with specific sample size recommendations, but insufficient sample sizes can impede accurate segmentation.
- Studies have shown that a sample size of at least 2p (better five times 2p), where p is the number of segmentation variables, is recommended by Formann (1984) for latent class analysis using binary variables.
- Qiu and Joe (2015) suggest a sample size of at least 10 times the number of segmentation variables times the number of segments in the data for clustering algorithms.
- Dolnicar et al. (2014, 2016) conducted extensive simulation studies to determine sample size requirements for correctly identifying segments in various scenarios.
- The studies found that larger sample sizes improve the algorithm's ability to identify correct segmentation solutions, but the extent of improvement varies depending on market and data characteristics.
- Factors such as unequal segment sizes, overlapping segments, and correlation between variables can affect segment recovery.

- Recommendations suggest a sample size of at least 60p or 70p, and ideally at least 100 respondents per segmentation variable to ensure accurate segmentation results.
- High-quality, unbiased data collection is crucial for effective market segmentation analysis.

3.5 Data from Internal Sources

- These data are strong because they show real-world consumer behavior, as opposed to self-reported behavior or intentions, which are subject to biases in reaction and recall.
- Internal data don't need to be collected additional; they are typically routinely created and easily accessible if saved in the right format.
- Internal data, however, carries the risk of systematic bias since it can overrepresent current clients. Due to this bias, information on possible future clients who might have different consumption habits than present clients may be lacking.

3.6 Data from Experimental Studies

- Experimental data, derived from field or laboratory studies, can be a useful source for market segmentation analysis.
- Conjoint analysis, choice experiments, and advertisement reactions are a few types of experimental data.
- Experimental data can reveal customer behavior in reaction to specific adverts, which can serve as segmentation criteria.
- Consumers are given stimuli comprising particular product qualities and levels in choice experiments and conjoint analysis, which enable them to express their preferences.
- These trials yielded data that can be utilized as segmentation criterion since it shows how each attribute and attribute level affects customer choice.

Step 4: Exploring Data

4.1 A First Glimpse at the Data

- Data exploration helps clean, pre-process data, and choose segmentation methods.
- In that matter in particular we see behavior of data and predict which model is suitable for it.
- It helps identify measurement levels, investigate variable distributions, and assess dependencies between variables.

4.2 Data Cleaning

- Before analysis, it's crucial to ensure all values are recorded correctly and consistently labeled.Metric variables like age are checked for plausible ranges.Categorical variables are checked for consistent levels.
- To find any irregularities or mistakes in the data entry or collection process, summary statistics are analyzed
- Although time-consuming, cleaning data using code ensures full documentation and reproducibility.
- Clicking in a spreadsheet may be error-prone and less reproducible compared to coding the cleaning steps.
- Reproducibility allows for the exact same procedure to be applied when new data is added or when monitoring segmentation solutions over time.

4.3 Descriptive Analysis

- Descriptive numeric and graphic representations provide insights into the data.
- In that graphical representation we use tools like Excel,power book.
- Histograms: Visualize the distribution of numeric variables.
- Box Plots: Illustrate the five-number summary (min, 1st quartile, median, 3rd quartile, max) and identify outliers.Represents the five-number summary of numeric variables.Provides insight into distributional properties assuming unimodality.elps identify outliers beyond 1.5 times the interquartile range. Outliers are depicted as circles outside the whiskers.
- Scatter plots: Display relationships between two numeric variables.
- Bar plots: Present frequency counts for categorical variables.
- Mosaic plots: Show associations among multiple categorical variables.

4.4 Pre-Processing

- Pre-processing alters the data and should be done with careful consideration of the implications.
- It's important to evaluate whether the transformation preserves the original meaning of the variables and is suitable for the intended analysis.
- Strong arguments or evidence supporting the conversion should be considered before applying pre-processing steps.

4.4.1 Categorical Variables

- Two pre-processing procedures are often used for categorical variables. One is merging levels of categorical variables.

1. Merging levels
    - Useful when original categories are too numerous or differentiated.
    - Simplifies analysis by reducing the number of categories.
    - Often applied to variables with many distinct levels, such as income brackets or education levels.
    - Helps improve interpretability and model performance by reducing noise from small categories.

2. categorical variables
    - Applied when converting categorical variables to numeric ones makes sense and preserves the meaning of the data.
    - Ordinal data can sometimes be converted to numeric if distances between categories are approximately equal.
    - Popular agreement scales (e.g., Likert scales) may be treated as numeric if distances between response options are assumed to be equal.
    - Binary categorical variables can be directly converted to numeric variables with 0 and 1.

4.4.2 Numeric Variables

- In distance-based segment extraction methods, the relative importance of a segmentation variable varies depending on its range of values.
- For example, disparities in scale can result in uneven variable weighting in a situation involving binary and continuous data.
- By putting variables on a uniform scale, standardization aids in balancing their influence.

- A standardized variable having a mean of 0 and a standard deviation of 1 is produced by removing the mean and dividing by the standard deviation during the standardization process.
- This guarantees equitable comparison and allocation of weights among the variables during the segmentation procedure, regardless of their initial scales.
- In order to obtain precise and significant segmentation findings, standardization is essential.
- In the presence of outliers or skewed distributions, alternative standardization methods may be preferable.
- Robust estimates for location and spread, such as the median and interquartile range, are more resistant to outliers and provide a more accurate representation of central tendency and variability.

4.5 Principal Components Analysis

- PCA (Principal Component Analysis) is a mathematical technique that helps in reducing the number of dimensions in a dataset while retaining the maximum amount of information.
- A multivariate dataset with metric variables is transformed into a new dataset with uncorrelated variables called principal components using principal components analysis.
- The significance of these components is arranged as follows: the first component captures the most variability, followed by the second, and so forth.
- PCA adopts a distinct perspective on the data while preserving the relative placements of the observations.
- It operates using the correlation or covariance matrix of numerical variables. It is better to use the correlation matrix if the data ranges are different.
- PCA is frequently used, usually with the first few main components, to project high-dimensional data into lower dimensions for display purposes.
- The principle component standard deviations and a rotation matrix showing the contributions of the original variables to each component are among the outputs of PCA.
- The significance of each primary component is interpreted using summary statistics like cumulative proportion, proportion of explained variance, and standard deviation.
- PCA is helpful for finding strongly correlated variables and for data exploration. By eliminating unnecessary variables from the segmentation base, PCA insights can lower dimensionality without sacrificing the original variables.