**Report By: Parth Prabhune ( Team Vivek )**
**Date: 24 April 2024**

# Market Segmentation

## Step 1: Deciding(not) to Segment

### Consequences of Adopting Market Segmentation

It is really important to think about what might happen if a company decides to divide its market into different segments. This means that they segment their customers according to different characteristics or needs. But it's not just a short-term thing; it's a long-term commitment. Before starting segmentation, a company needs to make sure it's worth it. Does it really increase sales? If not, it might not be the right move. But if it's worth it, be ready for a change. New products may need to be made, old ones replaced, prices adjusted, and even the way things are sold and talked about may need shaking up. It's not just about sharing customers; it is about making big changes in the company's business.

### Challenges in Executing Market Segmentation:

There can be many things in the company's operations that can prevent the company from dividing its market into different segments. First, if the big bosses don't accept the idea and drive it, it will be difficult. Then you have to think about the culture of the company. When people don't think about what customers want, don't like new ideas, or can't talk to each other, it becomes a struggle. It is also another obstacle if someone does not know what they are doing because they are not trained or there are no marketing experts around. Sometimes it's just old obstacles, like lack of money or inability to change the order of things. If these obstacles cannot be removed, it may be best to divide the market.

## Step 2: Specifying the Ideal Target Segmentation

### Evaluating Market Segments

When a company wants to segment its market, the third important step relies heavily on asking for feedback directly from customers who use its products or services. The evaluation process involves two main approaches. For starters, there are "elimination criteria" that are prerequisites. Segments that do not meet these basic standards are immediately rejected. After that, there are "attractiveness" criteria that measure the desirability of each segment. A segment must pass both criteria to be considered for further processing.

**Removal criteria**

These criteria form the main filter to select the most promising customer groups. For a segment to be worth considering, it must meet some basic requirements. First, individuals within a segment should be remarkably similar to each other while maintaining distinctiveness from individuals in other segments. In addition, the segment should cover a significant amount to justify the allocation of company resources. In addition, it should be possible to identify these people and reach them in the market. Any segment that does not meet these criteria will be immediately removed from further processing.

**Winning Criteria**

Winning Criteria is a more comprehensive evaluation process. Each segment is classified based on various factors to determine its suitability for targeting. Parameters such as segment size, potential purchasing power and accessibility are carefully considered. A cumulative score of all criteria determines segment profitability as a target. Segments that score high enough on all criteria are identified as target groups that deserve focused attention.

**Structured implementation process**

Involving representatives from different organisational units in this process is extremely important for two reasons. First, diverse perspectives enrich the understanding of market dynamics and promote a holistic assessment. Second, because the segmentation strategy affects all aspects of the business, inclusive participation ensures that all stakeholders have a say in shaping the strategy. Thus, each department becomes an important stakeholder in this process.

**Early identification**

Identifying critical aspects of market segments early is critical to streamlining the subsequent data collection process. At this point, the team should have identified about six criteria that describe the attractiveness of the segment, each assigned a weighted score that reflects its relative importance. This ensures consistency and clarity in the prioritisation of criteria during the evaluation process.

## Step 3: Collecting Data

### Segmentation Variables

The quality of empirical data plays a critical role in the development of valid market segmentation solutions, as shown by examples where segmentation variables based on targeted benefits produce characteristic segments. Data quality is important not only for accurately segmenting individuals, but also for effectively describing those segments, enabling customised product development, pricing strategy, distribution channel selection, and advertising. This applies to both ordinary and data-driven segmentation methods, emphasising the importance of high-quality empirical data. Such data can come from surveys, observations such as scanner data, or experimental studies, preferring sources that reflect actual consumer behaviour rather than just survey-based data..

### Segmentation criteria

Before diving into data collection, organisations must choose a segmentation criterion, crucial for effective segmentation. Common criteria include geographic, sociodemographic, psychographic, and behavioural factors. Despite numerous options, simplicity is often preferred, with the recommendation to use the most straightforward approach aligned with product or service needs. Ultimately, effectiveness trumps sophistication in segmentation strategy.

### Geographic Segmentation

Originally used to segment markets, geographic segmentation simplifies targeting by categorising consumers based on where they live. An example of such an approach is the Austrian National Tourism Organization, which adapts its activities to neighbouring countries, taking language differences into account. Despite offering similar products worldwide, IKEA adjusts its offerings, pricing, and online purchasing options based on the customer's geographic location. This strategy demonstrates the practical application of geographic segmentation in addressing regional preferences and market dynamics, enhancing the effectiveness of IKEA's marketing efforts across diverse geographical regions.

### Socio-Demographic Segmentation

Socio Demographic segmentation, which includes criteria such as age, gender, income and education, provides practical insights into consumer behaviour across industries. For example, it will prove invaluable in sectors such as luxury goods, cosmetics, baby products, retirement villages and tourism products, where consumer preferences are closely linked to socio-demographic characteristics. Although socio-demographic criteria are easily determined for each consumer, they may not always be directly correlated with product

preferences, since demographic factors account for only a small part of the variation in consumer behaviour.

**Choice of variables**

When using both common sense and data segmentation, careful selection of variables is critical to achieving quality market segmentation solutions. Data-driven segmentation must include all relevant variables according to the segmentation criteria, but unnecessary ones should be avoided to prevent respondent fatigue and maintain data quality. Unnecessary variables not only complicate data analysis, but also confuse algorithms in identifying optimal segmentation solutions by adding noise or masking effects. Thus, data collection and variable selection play a key role in ensuring the accuracy and effectiveness of segmentation efforts. Strategies such as asking only the necessary questions and exploratory studies to identify critical variables help mitigate the effects of noisy variables and improve the segmentation process..

**Response Options**

Question answer choices are crucial in determining the extent of data available for subsequent analysis, which is particularly important in segmentation research. Binary or metric response options, which allow respondents to answer in a clearly defined manner, are well suited for segmentation analysis because they facilitate direct measurement of the distance between responses. However, the additional information arising from responses with ordered but unspecified alternative distances presents challenges for standard distance measurements unless strong assumptions are made. Although ordinal scales are common in survey research, providing binary or metric response options when appropriate helps avoid the complications of data-driven segmentation analysis, especially given the limitations of ordinal data.

**Response Styles**

Response styles, including trends such as extremes, are embedded. in survey data. responses, using a midpoint, or agreeing with all statements can significantly affect segmentation results. Such biases can lead to misinterpretation of segments, making segments very attractive response patterns rather than actual beliefs or behaviours. Minimising the risk of capturing response styles during data collection is critical for accurate segmentation. Additional analyzes may be needed to identify and address response style biases to ensure the reliability and validity of the segmentation results..

**Data from internal sources**

Increasingly, organisations have access to significant amounts of internal data that can be collected for market segmentation analysis.The strength of such data is that it reflects actual consumer behaviour rather than consumer statements. about their behaviour or intentions.Another advantage is that such data is usually generated automatically and – if

organisations can store the data in an easily accessible form – data collection requires no additional effort.

**Experimental research data**

Another potential source of data that can serve as a basis for market segment analysis, there are experimental data. Test data can come from field or lab tests.

Ad response can then be used as segmentation criteria.

**Case Study: Fast Food**

Code implementation :

```
//Import libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.decomposition import PCA
from sklearn.cluster import KMeans

// Load Dataset
mcdonalds_path = https://homepage.boku.ac.at/leisch/MSA/datasets/mcdonalds.csv
mcdonalds = pd.read_csv(mcdonalds_path)
print(mcdonalds.columns.tolist())
print(mcdonalds.shape)
# Display first 3 rows
print(mcdonalds.head(3))

// Selecting relevant columns and converting categorical responses into binary format
MD_x = mcdonalds.iloc[:, 0:11].copy()
# Convert "Yes" to 1 and "No" to 0
MD_x = (MD_x == "Yes").astype(int)
# Calculate column means
```

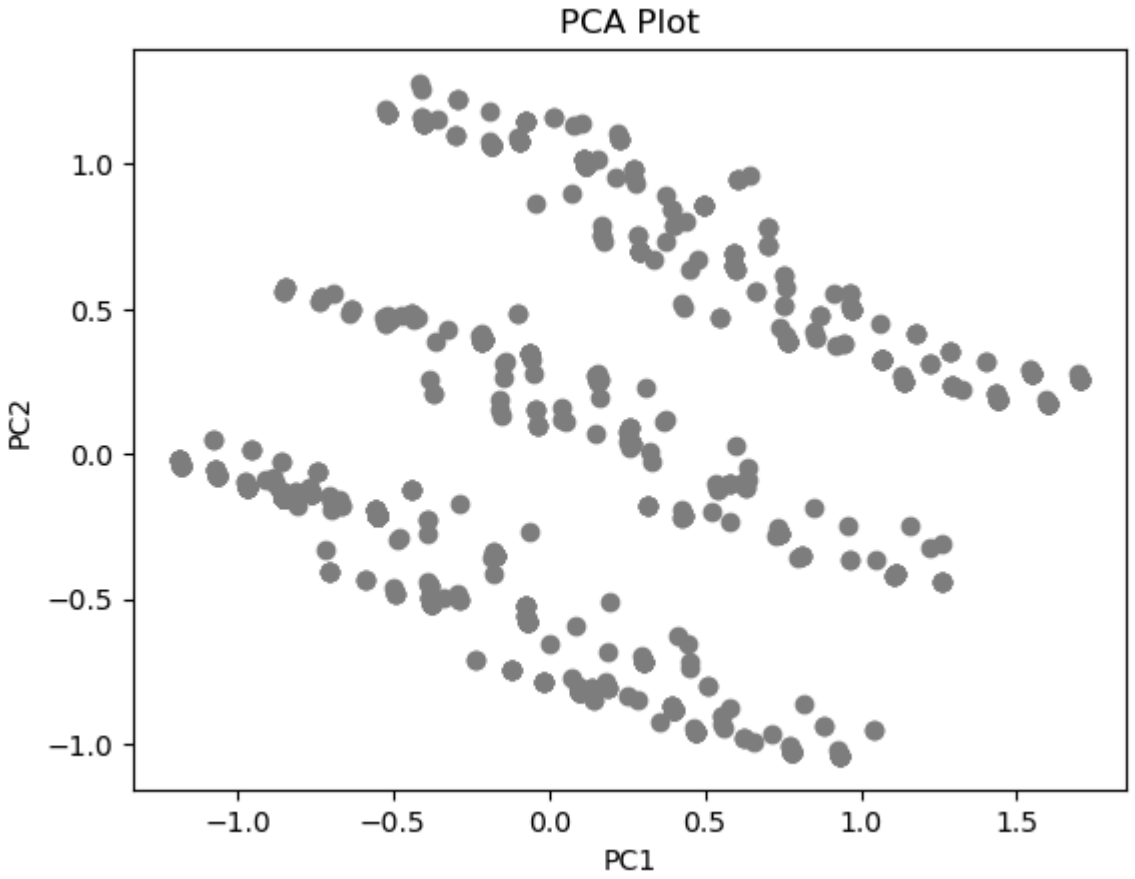
```
column_means = np.round(MD_x.mean(), 2)
print(column_means)

// PCA
# Perform PCA
MD_pca = PCA()
MD_pca.fit(MD_x)
print("Importance of components:")
print(pd.DataFrame({
    "Standard deviation": np.round(MD_pca.explained_variance_, 4),
    "Proportion of Variance": np.round(MD_pca.explained_variance_ratio_, 4),
    "Cumulative Proportion": np.round(np.cumsum(MD_pca.explained_variance_ratio_), 4)
}))

// Printing standard deviation and rotation matrix
def print_pca(pca_obj, digits):
    print("Standard deviations (1, .., p={}):".format(pca_obj.n_components_))
    print(np.round(pca_obj.explained_variance_, digits))
    print("Rotation (n x k) = ({} x {}):".format(pca_obj.components_.shape[1],
pca_obj.components_.shape[0]))
    print(np.round(pca_obj.components_, digits))
# Print PCA object with specified number of digits
print_pca(MD_pca, digits=1)

//Plot PCA
# Transform data using PCA
transformed_data = MD_pca.transform(MD_x)
plt.scatter(transformed_data[:, 0], transformed_data[:, 1], color='grey')
plt.xlabel("PC1")
plt.ylabel("PC2")
plt.title("PCA Plot")
```

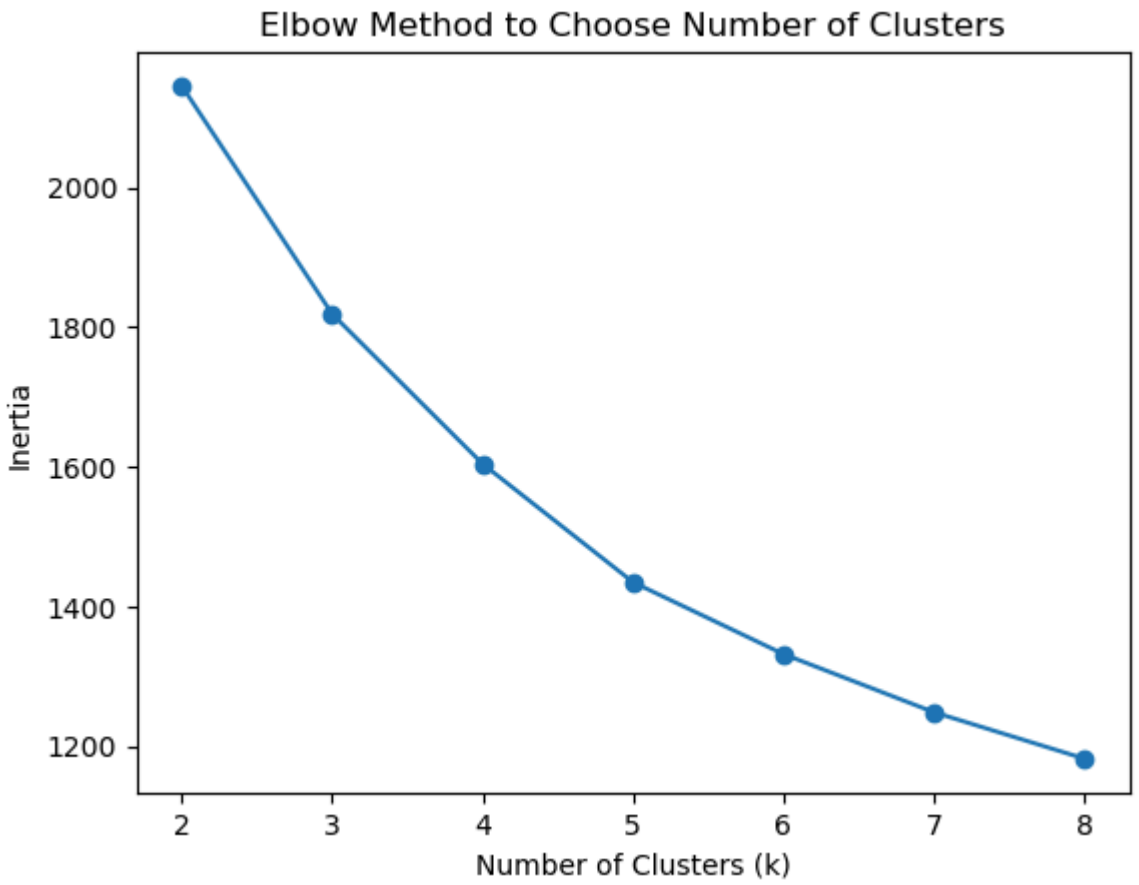
plt.show()




// Cluster labels

```
np.random.seed(1234)
k_values = range(2, 9)
best_model = None
best_score = float('inf')
for k in k_values:
    model = KMeans(n_clusters=k, n_init=10, random_state=1234)
    model.fit(MD_x)
    if model.inertia_ < best_score:
        best_model = model
```

```
        best_score = model.inertia_
cluster_labels = best_model.labels_
print(cluster_labels)
// Elbow method
k_values = range(2, 9)
# Choose the range of k values to try
inertia_values = []
for k in k_values:
    kmeans = KMeans(n_clusters=k, n_init=10, random_state=1234)
    kmeans.fit(MD_x)
    inertia_values.append(kmeans.inertia_)
plt.plot(k_values, inertia_values, marker='o')
plt.xlabel("Number of Clusters (k)")
plt.ylabel("Inertia")
plt.title("Elbow Method to Choose Number of Clusters")
plt.show()
```




// Bootstrapping K-means clustering

```
from sklearn.preprocessing import OneHotEncoder

encoder = OneHotEncoder(sparse_output=False, drop='first')  # Drop first category to avoid multicollinearity

MD_x_encoded = encoder.fit_transform(MD_x)

np.random.seed(1234)

n_bootstraps = 100

k_values = range(2, 9)

n_rep = 10

boot_results = []


# Bootstrap loop

for _ in range(n_bootstraps):
```

```
    # Generate bootstrap sample indices with replacement
    indices = np.random.choice(len(MD_x_encoded), size=len(MD_x_encoded),
replace=True)
    bootstrap_sample = MD_x_encoded[indices]
    # Perform K-means clustering with a chosen number of clusters
    best_model = None
    best_score = float('inf')
    for k in k_values:
        model = KMeans(n_clusters=k, n_init=n_rep, random_state=1234)
        model.fit(bootstrap_sample)
        if model.inertia_ < best_score:
            best_model = model
            best_score = model.inertia_
    boot_results.append(best_model)
for i, model in enumerate(boot_results):
    print(f"Bootstrap {i+1} - Number of Clusters: {model.n_clusters}")
```

// Bootstrap outputs

Bootstrap 1 - Number of Clusters: 8

Bootstrap 2 - Number of Clusters: 8

Bootstrap 3 - Number of Clusters: 8

Bootstrap 4 - Number of Clusters: 8

Bootstrap 5 - Number of Clusters: 8

Bootstrap 6 - Number of Clusters: 8

Bootstrap 7 - Number of Clusters: 8

Bootstrap 8 - Number of Clusters: 8

Bootstrap 9 - Number of Clusters: 8

Bootstrap 10 - Number of Clusters: 8

Bootstrap 11 - Number of Clusters: 8

Bootstrap 12 - Number of Clusters: 8

Bootstrap 13 - Number of Clusters: 8

Bootstrap 14 - Number of Clusters: 8

Bootstrap 15 - Number of Clusters: 8

Bootstrap 16 - Number of Clusters: 8

Bootstrap 17 - Number of Clusters: 8

Bootstrap 18 - Number of Clusters: 8

Bootstrap 19 - Number of Clusters: 8

Bootstrap 20 - Number of Clusters: 8

Bootstrap 21 - Number of Clusters: 8

Bootstrap 22 - Number of Clusters: 8

Bootstrap 23 - Number of Clusters: 8

Bootstrap 24 - Number of Clusters: 8

Bootstrap 25 - Number of Clusters: 8

Bootstrap 26 - Number of Clusters: 8

Bootstrap 27 - Number of Clusters: 8

Bootstrap 28 - Number of Clusters: 8

Bootstrap 29 - Number of Clusters: 8

Bootstrap 30 - Number of Clusters: 8

Bootstrap 31 - Number of Clusters: 8

Bootstrap 32 - Number of Clusters: 8

Bootstrap 33 - Number of Clusters: 8

Bootstrap 34 - Number of Clusters: 8

Bootstrap 35 - Number of Clusters: 8

Bootstrap 36 - Number of Clusters: 8

Bootstrap 37 - Number of Clusters: 8

Bootstrap 38 - Number of Clusters: 8

Bootstrap 39 - Number of Clusters: 8

Bootstrap 40 - Number of Clusters: 8

Bootstrap 41 - Number of Clusters: 8

Bootstrap 42 - Number of Clusters: 8

Bootstrap 43 - Number of Clusters: 8

Bootstrap 44 - Number of Clusters: 8

Bootstrap 45 - Number of Clusters: 8

Bootstrap 46 - Number of Clusters: 8

Bootstrap 47 - Number of Clusters: 8

Bootstrap 48 - Number of Clusters: 8

Bootstrap 49 - Number of Clusters: 8

Bootstrap 50 - Number of Clusters: 8

Bootstrap 51 - Number of Clusters: 8

Bootstrap 52 - Number of Clusters: 8

Bootstrap 53 - Number of Clusters: 8

Bootstrap 54 - Number of Clusters: 8

Bootstrap 55 - Number of Clusters: 8

Bootstrap 56 - Number of Clusters: 8

Bootstrap 57 - Number of Clusters: 8

Bootstrap 58 - Number of Clusters: 8

Bootstrap 59 - Number of Clusters: 8

Bootstrap 60 - Number of Clusters: 8

Bootstrap 61 - Number of Clusters: 8

Bootstrap 62 - Number of Clusters: 8

Bootstrap 63 - Number of Clusters: 8

Bootstrap 64 - Number of Clusters: 8

Bootstrap 65 - Number of Clusters: 8

Bootstrap 66 - Number of Clusters: 8

Bootstrap 67 - Number of Clusters: 8

Bootstrap 68 - Number of Clusters: 8

Bootstrap 69 - Number of Clusters: 8

Bootstrap 70 - Number of Clusters: 8

Bootstrap 71 - Number of Clusters: 8

Bootstrap 72 - Number of Clusters: 8

Bootstrap 73 - Number of Clusters: 8

Bootstrap 74 - Number of Clusters: 8

Bootstrap 75 - Number of Clusters: 8

Bootstrap 76 - Number of Clusters: 8

Bootstrap 77 - Number of Clusters: 8

Bootstrap 78 - Number of Clusters: 8

Bootstrap 79 - Number of Clusters: 8

Bootstrap 80 - Number of Clusters: 8

Bootstrap 81 - Number of Clusters: 8

Bootstrap 82 - Number of Clusters: 8

Bootstrap 83 - Number of Clusters: 8

Bootstrap 84 - Number of Clusters: 8

Bootstrap 85 - Number of Clusters: 8

Bootstrap 86 - Number of Clusters: 8

Bootstrap 87 - Number of Clusters: 8

Bootstrap 88 - Number of Clusters: 8

Bootstrap 89 - Number of Clusters: 8

Bootstrap 90 - Number of Clusters: 8

Bootstrap 91 - Number of Clusters: 8

Bootstrap 92 - Number of Clusters: 8

Bootstrap 93 - Number of Clusters: 8

Bootstrap 94 - Number of Clusters: 8

Bootstrap 95 - Number of Clusters: 8

Bootstrap 96 - Number of Clusters: 8

Bootstrap 97 - Number of Clusters: 8

Bootstrap 98 - Number of Clusters: 8

Bootstrap 99 - Number of Clusters: 8

Bootstrap 100 - Number of Clusters: 8