FEYNN LABS

TASK – 2 REPORT

NAME – ADITYA CHOUDHARY

STEP – 01:

The key implication is that the organisation needs to commit to the segmentation strategy on the long term. it: Segmenting a market is not free. There are costs of performing the research, fielding surveys, and focus groups, designing multiple packages, and designing multiple advertisements and communication messages. Potentially required changes include the development of new products, the modification of existing products, changes in pricing and distribution channels used to sell the product, as well as all communications with the market.

These changes, in turn, are likely to influence the internal structure of the organization, which may need to be adjusted in view of, for example, targeting a handful of different market segments. It is recommended that – to maximize the benefits of market segmentation – organizations need to organize around market segments, rather than organising around products. Strategic business units in charge of segments offer a suitable organisational structure to ensure ongoing focus on the (changing) needs of market segments.

Implementation Barriers:

The first group of barriers relates to senior management. Lack of leadership, pro-active championing, commitment and involvement in the market segmentation

process by senior leadership undermines the success of market segmentation. Senior management can also prevent market segmentation to be successfully implemented by not making enough resources available, either for the initial market segmentation analysis itself, or for the long-term implementation of a market segmentation strategy. A second group of barriers relates to organizational culture. Lack of market or consumer orientation, resistance to change and new ideas, lack of creative thinking, bad communication and lack of sharing of information and insights across organisational units, short-term thinking, unwillingness to make changes and office politics have been identified as preventing the successful implementation of market segmentation. Another potential problem is lack of training. If senior management and the team tasked with segmentation do not understand the very foundations of market segmentation, or if they are unaware of the consequences of pursuing such a strategy, the attempt of introducing market segmentation is likely to fail. Closely linked to these barriers is the lack of a formal marketing function or at least a qualified marketing expert in the organization. The higher the market diversity and the larger the organizations, the more important is a high degree of formalization. The lack of a qualified data manager and analyst in the organisation can also represent major stumbling blocks. Another obstacle may be objective restrictions faced by the organisation, including lack of financial resources, or the inability to make the structural changes required. Process-related barriers include not having clarified the objectives of the market segmentation exercise, lack of planning or bad planning, a lack of structured processes to guide the team through all steps of the market segmentation process, a lack of allocation of responsibilities, and time pressure that stands in the way of trying to find the best possible segmentation outcome. At a more operational level, Doyle and Saunders (1985) note that management science has had a disappointing level of acceptance in industry because management will not use techniques it does not understand . One way of counteracting this challenge is to make market segmentation analysis easy to understand, and to present results in a way that facilitates interpretation by managers. This can be achieved by using graphical visualisations.

STEP – 2:

Segment Evaluation Criteria:

The third layer of market segmentation analysis depends primarily on user input. the user needs to be involved in most stages, literally wrapping around the technical aspects of market segmentation analysis. In Step 2 the organization must determine two sets of segment evaluation criteria. One set of evaluation criteria can be referred to as knock-out criteria. These criteria are the essential, non-negotiable features of segments that the organization would consider targeting. The second set of evaluation criteria can be referred to as attractiveness criteria. These criteria are used to evaluate the relative attractiveness of the remaining market segments – those in compliance with the knock-out criteria.

Knock-out Criteria:

Knock-out criteria are used to determine if market segments resulting from the market segmentation analysis qualify to be assessed using segment attractiveness criteria. The first set of such criteria was suggested by Kotler (1994) and includes substantiality, measurability and accessibility.

1. The segment must be homogeneous; members of the segment must be similar to one another.
2. The segment must be distinct; members of the segment must be distinctly different from members of other segments.
3. The segment must be large enough; the segment must contain enough consumers to make it worthwhile to spend extra money on customising the marketing mix for them.
4. The segment must be matching the strengths of the organisation; the organisation must have the capability to satisfy segment members' needs.
5. Members of the segment must be identifiable; it must be possible to spot them in the marketplace.
6. The segment must be reachable; there has to be a way to get in touch with members of the segment in order to make the customised marketing mix accessible to them.

Knock-out criteria must be understood by senior management, the segmentation team, and the advisory committee.

Attractiveness Criteria:

Attractiveness criteria are not binary in nature. Segments are not assessed as either complying or not complying with attractiveness criteria. Rather, each market segment is rated; it can be more or less attractive with respect to a specific criterion.

Implementing a Structured Process:

The most popular structured approach for evaluating market segments in view of selecting them as target markets is the use of a segment evaluation plot showing segment attractiveness along one axis, and organisational competitiveness on the other axis. The segment attractiveness and organisational competitiveness values are determined by the segmentation team. This is necessary because there is no standard set of criteria that could be used by all organisations. Factors which constitute both segment attractiveness and organisational competitiveness need to be negotiated and agreed upon. To achieve this, a large number of possible criteria has to be investigated before agreement is reached on which criteria are most important for the organisation. If a core team of two to three people is primarily in charge of market segmentation analysis, this team could propose an initial solution and report their choices to the advisory committee – which consists of representatives of all organisational units – for discussion and possible modification.

There are at least two good reasons to include in this process representatives from a wide range of organisational units.

First, each organisational unit has a different perspective on the business of the organisation. As a consequence, members of these units bring different positions to the deliberations. Secondly, if the segmentation strategy is implemented, it will affect every single unit of the organization. Consequently, all units are key stakeholders of market segmentation analysis. At the end of this step, the market segmentation team should have a list of approximately six segment attractiveness criteria. Each of these criteria should have a weight attached to it to indicate how important it is to the organisation compared to the other criteria.

STEP – 3:

Segmentation Variables:

Empirical data forms the basis of both commonsense and data-driven market segmentation. Empirical data is used to identify or create market segments and – later in the process – describe these segments in detail. detail. Describing segments is critical to being able to develop an effective marketing mix targeting the segment. Typical descriptor variables include socio-demographics, but also information about media behaviour, allowing marketers to reach their target segment with communication messages.

The difference between commonsense and data-driven market segmentation is that data-driven market segmentation is based not on one, but on multiple segmentation variables. These segmentation variables serve as the starting point for identifying naturally existing, or artificially creating market segments useful to the organisation. Empirical data for segmentation studies can come from a range of sources:

from survey studies; from observations such as scanner data where purchases are recorded and, frequently, are linked to an individual customer's long-term purchase history via loyalty programs; or from experimental studies. Optimally, data used in segmentation studies should reflect consumer behaviour. Survey data – although it arguably represents the most common source of data for market segmentation studies – can be unreliable in reflecting behaviour, especially when the behaviour of interest is socially desirable, such as donating money to a charity or behaving in an environmentally friendly way.

Segmentation Criteria:

The term segmentation criterion is used here in a broader sense than the term segmentation variable. The term segmentation variable refers to one measured value, for example, one item in a survey, or one observed expenditure category. The term segmentation criterion relates to the nature of the information used for market segmentation. It can also relate to one specific construct, such as benefits sought. The most common segmentation criteria are geographic, socio-demographic, psychographic and behavioural.

Geographic Segmentation:

Typically – when geographic segmentation is used – the consumer's location of residence serves as the only criterion to form market segments. While simple, the geographic segmentation approach is often the most appropriate. The key advantage of geographic segmentation is that each consumer can easily be assigned to a geographic unit. As a consequence, it is easy to target communication messages, and select communication channels (such as local newspapers, local radio and TV stations) to reach the selected geographic segments. The key disadvantage is that living in the same country or area does not necessarily mean that people share other characteristics relevant to marketers, such as benefits they seek when purchasing a product. While, for example, people residing in luxury suburbs may all be a good target market for luxury cars, location is rarely the reason for differences in product preference. Despite the potential shortcomings of using geographic information as the segmentation variable, the location aspect has experienced a revival in international market segmentation studies aiming to extract market segments across geographic boundaries. Such an approach is challenging because the segmentation variable(s) must be meaningful across all the included geographic regions, and because of the known biases that can occur if surveys are completed by respondents from different cultural backgrounds.

Socio-Demographic Segmentation:

Typical socio-demographic segmentation criteria include age, gender, income and education. Socio-demographic segments can be very useful in some industries. For example: luxury goods (associated with high income), cosmetics (associated with gender; even in times where men are targeted, the female and male segments are treated distinctly differently), baby products (associated with gender), retirement villages (associated with age), tourism resort products (associated with having small children or not).

As is the case with geographic segmentation, socio-demographic segmentation criteria have the advantage that segment membership can easily be determined for every consumer. In some instances, the socio-demographic criterion may also offer an explanation for specific product preferences (having children, for example, is the actual reason that families choose a family vacation village where previously, as a couple, their vacation choice may have been entirely different).

Psychographic Segmentation:

When people are grouped according to psychological criteria, such as their beliefs, interests, preferences, aspirations, or benefits sought when purchasing a product, the term psychographic segmentation is used. Psychographic criteria are, by nature, more complex than geographic or socio-demographic criteria because it is difficult to find a single characteristic of a person that will provide insight into the psychographic dimension of interest. As a consequence, most psychographic segmentation studies use a number of segmentation variables, for example: a number of different travel motives, a number of perceived risks when going on vacation. The psychographic approach has the advantage that it is generally more reflective of the underlying reasons for differences in consumer behaviour. For example, tourists whose primary motivation to go on vacation is to learn about other cultures, have a high likelihood of undertaking a cultural holiday at a destination that has ample cultural treasures for them to explore.

tourism (Bieger and Laesser 2002; Laesser et al. 2006; Boksberger and Laesser 2009). The disadvantage of the psychographic approach is the increased complexity of determining segment memberships for consumers. Also, the power of the psychographic approach depends heavily on the reliability and validity of the empirical measures used to capture the psychographic dimensions of interest.

Behavioural Segmentation:

Another approach to segment extraction is to search directly for similarities in behaviour or reported behaviour. A wide range of possible behaviours can be used for this purpose, including prior experience with the product, frequency of purchase, amount spent on purchasing the product on each occasion (or across multiple purchase occasions), and information search behaviour. In a comparison of different segmentation criteria used as segmentation variables, behaviours reported by tourists emerged as superior to geographic variables.

The key advantage of behavioural approaches is that – if based on actual behaviour rather than stated behaviour or stated intended behaviour – the very behaviour of interest is used as the basis of segment extraction. But behavioural data is not always readily available, especially if the aim is to include in the segmentation analysis potential customers who have not previously purchased the product, rather than limiting oneself to the study of existing customers of the organisation.

Data from Survey Studies:

Most market segmentation analyses are based on survey data. Survey data is cheap and easy to collect, making it a feasible approach for any organisation. But survey data – as opposed to data obtained from observing actual behaviour – can be contaminated by a wide range of biases.

Choice of Variables:

In data-driven segmentation, all variables relevant to the construct captured by the segmentation criterion need to be included. At the same time, unnecessary variables must be avoided. Including unnecessary variables can make questionnaires long and tedious for respondents, which, in turn, causes respondent fatigue. Fatigued respondents tend to provide responses of lower quality. The issue of the appropriate ratio of the number of variables and the available sample is discussed later in this chapter. Unnecessary variables included as segmentation variables divert the attention of the segment extraction algorithm away from information critical to the extraction of optimal market segments. Such variables are referred to as noisy variables or masking variables and have been repeatedly shown to prevent algorithms from identifying the correct segmentation solution. Noisy variables do not contribute any information necessary for the identification of the correct market segments. Instead, their presence makes it more difficult for the algorithm to extract the correct solution. Noisy variables can result from not carefully developing survey questions, or from not carefully selecting segmentation variables from among the available survey items. The problem of noisy variables negatively affecting the segmentation solution can be avoided at the data collection and the variable selection stage of market segmentation analysis.

The recommendation is to ask all necessary and unique questions, while resisting the temptation to include unnecessary or redundant questions. Redundant questions are common in survey research when scale development follows traditional psycho-metric principles. Redundant items are particularly problematic in the context of market segmentation analysis because they interfere substantially with most segment extraction algorithms' ability to identify correct market segmentation solutions.

Developing a good questionnaire typically requires conducting exploratory or qualitative research. Exploratory research offers insights about people's beliefs that survey research cannot offer. These insights can then be categorised and included in a questionnaire as a list of answer options. Such a two-stage process involving both qualitative, exploratory and quantitative survey research ensures that no critically important variables are omitted.

Response Options:

Answer options provided to respondents in surveys determine the scale of the data available for subsequent analyses. Because many data analytic techniques are based on distance measures, not all survey response options are equally suitable for segmentation analysis. Options allowing respondents to indicate a number, such as age or nights stayed at a hotel, generate metric data. Metric data allow any statistical procedure to be performed (including the measurement of distance), and are therefore well suited for segmentation analysis. The most commonly used response option in survey research, however, is a limited number of ordered answer options larger than two. This answer format generates ordinal data, meaning that the options are ordered. But the distance between adjacent answer options is not clearly defined. As a consequence, it is not possible to apply standard distance measures to such data, unless strong assumptions are made. Using binary or metric response options prevents subsequent complications relating to the distance measure in the process of data-driven segmentation analysis. Although ordinal scales dominate both market research and academic survey research, using binary or metric response options instead is usually not a compromise.

Response Styles:

Survey data is prone to capturing biases. A response bias is a systematic tendency to respond to a range of questionnaire items on some basis other than the specific item content (i.e., what the items were designed to measure). If a bias is displayed by a respondent consistently over time, and independently of the survey questions asked, it represents a response style. A wide range of response styles manifest in survey answers, including respondents' tendencies to use extreme answer options (STRONGLY AGREE, STRONGLY, DISAGREE), to use the midpoint (NEITHER AGREE NOR DISAGREE), and to agree with all statements. Response styles affect segmentation results because Commonly used segment extraction algorithms cannot differentiate between a data entry reflecting the respondent's belief from a data entry reflecting both a respondent's belief and a response style. Such a segment could be misinterpreted. Imagine a market segmentation based on responses to a series of questions asking tourists to indicate whether or not they spent money on certain aspects of their vacation, including DINING OUT, VISITING THEME PARKS, USING PUBLIC TRANSPORT, etc. It is critical, therefore, to minimise the risk of capturing response styles when data is collected for the purpose of market segmentation. In cases where attractive market segments emerge with response patterns potentially caused by a response style, additional analyses are required to exclude this possibility.

Alternatively, respondents affected by such a response style must be removed before choosing to target such a market segment.

Sample Size:

Many statistical analyses are accompanied by sample size recommendations. Not so market segmentation analysis. The market segmentation problem in this figure is extremely simple because only two segmentation variables are used. Viennese psychologist Formann (1984) recommends that the sample size should be at least 2p (better five times 2p), where p is the number of segmentation variables. This rule of thumb relates to the specific purpose of goodness-of-fit testing in the context of latent class analysis when using binary variables. Not surprisingly, increasing the sample size improves the correctness of the extracted segments. Interestingly, however, the biggest improvement is achieved by increasing very small samples. As the sample size increases, the marginal benefit of further increasing the sample size decreases. For a more difficult artificial data scenario Dolnicar et al. (2014) recommend using a sample size of at least $70 \cdot p$; no substantial improvements in identifying the correct segments were identified beyond this point. Market characteristics studied included: the number of market segments present in the data, whether those market segments are equal or unequal in size, and the extent to which market segments overlap. In addition, some of the characteristics of survey data discussed above have been shown to affect segment recovery, specifically: sampling error, response biases and response styles, low data quality, different response options, the inclusion of irrelevant items, and correlation between blocks of items. Overall, this study demonstrates the importance of having a sample size sufficiently large to enable an algorithm to extract the correct segments (if segments naturally exist in the data).

It can be concluded from the body of work studying the effects of survey data quality on the quality of market segmentation results based on such data that, optimally, data used in market segmentation analyses should:

• contain all necessary items;
• contain no unnecessary items;
• contain no correlated items;
• contain high-quality responses;
• be binary or metric;
• be free of response styles;
• include responses from a suitable sample given the aim of the segmentation
study; and
• include a sufficient sample size given the number of segmentation variables (100
times the number of segmentation variables).

PYTHON CODE FOR CASE STUDY: FAS FOOD:

```
import pandas as pd
```

```
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
!pip install bioinfokit
```

```
df = pd.read_csv("../input/mcdonalds/mcdonalds.csv")
df.shape
df.head()
df.dtypes

df.info()
df.isnull().sum()
```

```
df['Gender'].value_counts()
df['VisitFrequency'].value_counts()
df['Like'].value_counts()
```

```
labels = ['Female', 'Male']
size = df['Gender'].value_counts()
colors = ['pink', 'cyan']
explode = [0, 0.1]
plt.rcParams['figure.figsize'] = (7, 7)
plt.pie(size, colors = colors, explode = explode, labels = labels,
shadow = True, autopct = '%.2f%%')
plt.title('Gender', fontsize = 20)
plt.axis('off')
plt.legend()
plt.show()
```

```
plt.rcParams['figure.figsize'] = (25, 8)
f = sns.countplot(x=df['Age'],palette = 'hsv')
f.bar_label(f.containers[0])
plt.title('Age distribution of customers')
plt.show()
```

```
df['Like']= df['Like'].replace({'I hate it!-5': '-5','I love
it!+5':'+5'})
```

```
sns.catplot(x="Like", y="Age",data=df,
            orient="v", height=5, aspect=2,
palette="Set2",kind="swarm")
plt.title('Likelyness of McDonald w.r.t Age')
plt.show()
```

```
from sklearn.preprocessing import LabelEncoder
def labelling(x):
```

```
    df[x] = LabelEncoder().fit_transform(df[x])
    return df

cat = ['yummy', 'convenient', 'spicy', 'fattening', 'greasy', 'fast',
'cheap',
       'tasty', 'expensive', 'healthy', 'disgusting']

for i in cat:
    labelling(i)
df
```

```
plt.rcParams['figure.figsize'] = (12,14)
df.hist()
plt.show()
```

```
df_eleven = df.loc[:,cat]
df_eleven
```

```
x = df.loc[:,cat].values
x
```

```
from sklearn.decomposition import PCA
from sklearn import preprocessing

pca_data = preprocessing.scale(x)

pca = PCA(n_components=11)
pc = pca.fit_transform(x)
names =
['pc1','pc2','pc3','pc4','pc5','pc6','pc7','pc8','pc9','pc10','pc11']
pf = pd.DataFrame(data = pc, columns = names)
pf
```

```
pca.explained_variance_ratio_
```

```
np.cumsum(pca.explained_variance_ratio_)
```

```
loadings = pca.components_
num_pc = pca.n_features_
pc_list = ["PC"+str(i) for i in list(range(1, num_pc+1))]
loadings_df = pd.DataFrame.from_dict(dict(zip(pc_list, loadings)))
loadings_df['variable'] = df_eleven.columns.values
loadings_df = loadings_df.set_index('variable')
loadings_df
```

```
plt.rcParams['figure.figsize'] = (20,15)
```

```
ax = sns.heatmap(loadings_df, annot=True, cmap='Spectral')
plt.show()
```

```
from bioinfokit.visuz import cluster
cluster.screeplot(obj=[pc_list,
pca.explained_variance_ratio_],show=True,dim=(10,5))
```

```
pca_scores = PCA().fit_transform(x)

# get 2D biplot
cluster.biplot(cscore=pca_scores, loadings=loadings,
labels=df.columns.values,
var1=round(pca.explained_variance_ratio_[0]*100, 2),
    var2=round(pca.explained_variance_ratio_[1]*100,
2),show=True,dim=(10,5))
```

```
from sklearn.cluster import KMeans
from yellowbrick.cluster import KElbowVisualizer
model = KMeans()
visualizer = KElbowVisualizer(model, k=(1,12)).fit(df_eleven)
visualizer.show()
```

```
kmeans = KMeans(n_clusters=4, init='k-means++',
random_state=0).fit(df_eleven)
df['cluster_num'] = kmeans.labels_
```

```
print (kmeans.labels_)
print (kmeans.inertia_)
print(kmeans.n_iter_)
print(kmeans.cluster_centers_)
```

```
from collections import Counter
Counter(kmeans.labels_)
```

```
sns.scatterplot(data=pf, x="pc1", y="pc2", hue=kmeans.labels_)
plt.scatter(kmeans.cluster_centers_[:,0], kmeans.cluster_centers_[:,1],
            marker="X", c="r", s=80, label="centroids")
plt.legend()
plt.show()
```

```
from statsmodels.graphics.mosaicplot import mosaic
from itertools import product
```

```
crosstab =pd.crosstab(df['cluster_num'],df['Like'])
```

```
crosstab = crosstab[['-5','-4','-3','-2','-
1','0','+1','+2','+3','+4','+5']]
crosstab
```

```python
plt.rcParams['figure.figsize'] = (7,5)
mosaic(crosstab.stack())
plt.show()
```

```python
crosstab_gender =pd.crosstab(df['cluster_num'],df['Gender'])
crosstab_gender
```

```python
plt.rcParams['figure.figsize'] = (7,5)
mosaic(crosstab_gender.stack())
plt.show()
```

```python
sns.boxplot(x="cluster_num", y="Age", data=df)
```

```python
df['VisitFrequency'] =
LabelEncoder().fit_transform(df['VisitFrequency'])
visit = df.groupby('cluster_num')['VisitFrequency'].mean()
visit = visit.to_frame().reset_index()
visit
```

```python
df['Like'] = LabelEncoder().fit_transform(df['Like'])
Like = df.groupby('cluster_num')['Like'].mean()
Like = Like.to_frame().reset_index()
Like
```

```python
df['Gender'] = LabelEncoder().fit_transform(df['Gender'])
Gender = df.groupby('cluster_num')['Gender'].mean()
Gender = Gender.to_frame().reset_index()
Gender
```

```python
segment = Gender.merge(Like, on='cluster_num', how='left').merge(visit,
on='cluster_num', how='left')
segment
```

```python
plt.figure(figsize = (9,4))
sns.scatterplot(x = "VisitFrequency", y = "Like",data=segment,s=400,
color="r")
plt.title("Simple segment evaluation plot for the fast food data set",
          fontsize = 15)
plt.xlabel("Visit", fontsize = 12)
plt.ylabel("Like", fontsize = 12)
plt.show()
```