Fundamentals of Market Segmentation Analysis and McDonalds Case Study

Team-Sonu

Sonu Kumar (Team Lead)
Akash Pratap Singh
Habiba Anjum
Tanmay Sandeep Pawar
Chandru J
C S Ayush Kumar

28th April 2022

Abstract

In this report, we are going through the Steps of Market segmentation and covering the basic idea of market segmentation. We will be covering all the instruction or procedure you have to keep in mind during the market segmentation.

Introduction

What is marketing?

The purpose of marketing is to match the genuine needs and desires of consumers with the offers of suppliers particularly suited to satisfy those needs and desires. This matching process benefits consumers and suppliers, and drives an organization's marketing planning process.

What is marketing planning?

Marketing planning is a logical sequence and a series of activities leading to the setting of marketing objectives and the formulation of plans to achieving them.

- A marketing plan consists of two components: a strategic and a tactical marketing plan.

  - The **strategic plan** outlines the long-term direction of an organization, but does not provide much detail on short-term marketing action required to move in this long-term direction.
  - The **tactical marketing plan** does the opposite. It translates the long-term strategic plan into detailed instructions for short-term marketing action.
- The strategic marketing plan states where the organization wants to go and why. The tactical marketing plan contains instructions on what needs to be done to get there.
- The strategic marketing plan typically identifies consumer needs and desires, strengths and weaknesses internal to the organization, and external opportunities and threats the organization may face.
- A SWOT analysis explicitly states an organization's strengths (S), weaknesses (W), opportunities (O), and threats (T). As such, the SWOT analysis outlines one side of the matching process: what the supplier is particularly suitable to offer consumers.
- The other side of the matching process – consumer needs and desires – is typically investigated using market research.
- Tactical marketing covers how the organization is going to present itself in front of public to look most attractive.
- Tactical marketing planning usually covers a period of up to one year.
- It is traditionally seen to cover four areas: the development and modification of the product in view of needs and desires of the target segment (Product), the determination of the price in view of cost, competition, and the willingness to pay of the target segment (Price), the selection of the most suitable distribution channels to reach the target segment (Place), and the communication and promotion of the offer in a way that is most appealing to the target segment (Promotion).
- The tactical marketing plan depends entirely on the strategic marketing plan, but the strategic marketing plan does not depend on the tactical marketing plan.

What is market segmentation?

Market segmentation is a decision-making tool for the marketing manager in the crucial task of selecting a target market for a given product and designing an appropriate marketing mix.

Steps of Market Segmentation analysis:

STEP 1 : Deciding (not) to Segment

Implications of Committing to Market Segmentation

- To do market segmentation it is a key marketing strategy that many organizations use but not always a good decision to purse this strategy. Before doing investment one should know the implication of the strategy. Market Segmentation is a marriage not date, you have to commit to the strategy on the long term.

Implementation Barriers:

- The first group of barrier is senior management
- The second group of Barrier is Organizational culture.
- The third group of barrier is lack of training
- The fourth group of barrier is lack of formal marketing function.
- The fifth barrier is objective restrictions faced by the organization.

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 visualizations. Another way is that, if going ahead with the market segmentation analysis, McDonald and Dunbar recommend: Above all, a resolute sense of purpose and dedication is required, tempered by patience and a willingness to appreciate the inevitable problems which will be encountered in implementing the conclusions.

Step 1 Checklist

This first checklist includes not only tasks, but also a series of questions which, if not answered in the affirmative, serve as knock-out criteria. For example: if an organization is not market-oriented, even the finest of market segmentation analyses cannot be successfully implemented.

STEP 2: Specifying the ideal target segment

Segment Evaluation criteria

- This layer of market segmentation analysis is generally depending upon user input.
- 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.

Knock-out criterion category:

- The segment must be homogeneous
    - The segment must be distinct
    - The segment must be large enough
    - The segment must be matching the strengths of the organization
    - Members of the segment must be identifiable
    - The segment must be reachable

Attractiveness Criteria

- Attractiveness criteria available to the segmentation team to consider when deciding which attractiveness criteria are most useful to their specific situation.
- The attractiveness across all criteria determines whether a market segment is selected as a target segment in market segmentation analysis.

Implementing a structured process

- Follow a structured process when assessing market segments are beneficial.

Step 3: Collecting data

An understanding of types of variables is essential to collect data and perform analysis on the same. In data analysis, the variables can be broadly classified into three categories:

- Nominal variables – These are categorical variables with no particular order between them. E.g., for location - “India”, “USA”, “Russia”

- Ordinal variables – These are also categorical variables but with undefined borders between them. E.g. For size small<medium<large. Categorical variables generally have to be converted to numerical ones in numerous methods (e.g., one-hot encoding) to feed into evaluation criteria.
- Numerical variables – Numerical variables are numbers. Order between such variables is very strictly defined and additional derivative variables can be created such as distance, sum of squares etc.

For data collection, the organization has to choose segmentation criteria before it decides to start collecting data. Segmentation criteria is best defined as the basis which will be the foundation of our market segmentation. Some very popular segmentation criteria are:

- Geographic segmentation – this segmentation puts the basis in location of the customer.
- Socio-demographic segmentation – here, social demographic data such as age, gender etc. is used for segmentation criteria.
- Psychographic segmentation – This is the most useful form of segmentation as it directly uses the behavior of the customers to perform segmentation however it is hard to collect and incurs additional problems such as response styles.
- Behavorial segmentation – This type of segmentation looks for similarity of behavior or reported behavior in customers to classify them into segments. This also suffers from the problem of response styles.

Choice of variables is very important in fielding surveys since too many questions can tire the responder and result in tired answers of low quality. Also, irrelevant data when present can noise the relevant data and produce poor segments after analysis. Hence, the variables should be chosen such that we can extract maximum information from minimum number of variables.

Response options present in the surveys determine the scale of information that we have available about our variables. However, making your options too expanse will again result in low quality data. Typical response options include binary data (0 or 1) or a metric response (generally 1-5 or 1-7) with the middle value being neutral response.

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. If a bias is displayed by a respondent consistently over time, and independently of the survey questions asked, it represents a response style. An example of response style will include someone mostly selecting extreme options (strongly agree or disagree) or selecting neutral options.

Sample size is another vital feature of data collection. Although there is no sample size too big, collecting big sample sizes is often a very tedious task. A good approximation of a good sample size is 10.p.k; where p is the number of segmentation variables and k is the number of segments. It is also essential to ensure the use of uncorrelated variables as correlation decreases their efficiency.

Hence, to summarize, important checkpoints of data collection are as follows:

- 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
- include a sufficient sample size given the number of segmentation variables

There can two types of data sources for our purpose. Internal data comprises of data generated generally through market transactions such as grocery store sales, ticket booking systems et cetera. This type of data is very accurate as it represents directly the market behavior itself however the downside is that it only represents existing customers, hence might not prove to be good if the goal is market expansion.

Data from experimental studies, field surveys is another source. Although it hard to generate in large amounts and can have various obstacles such as response styles, misrepresentation, it is viable to target all kinds of customer, existing and non-existing. Unless a bias is introduced at the survey end itself, such data is supposed to be a good option for overall market capturing.

Step 3 checklist:

Task	Who is responsible?	Completed?
Convene a market segmentation team meeting.		☐
Discuss which consumer characteristics could serve as promising segmentation variables. These variables will be used to extract groups of consumers from the data.		☐
Discuss which other consumer characteristics are required to develop a good understanding of market segments. These variables will later be used to describe the segments in detail.		☐
Determine how you can collect data to most validly capture both the segmentation variables and the descriptor variables.		☐
Design data collection carefully to keep data contamination through biases and other sources of systematic error to a minimum.		☐
Collect data.		☐

Step 4: Exploring Data

Before performing any analysis, preprocessing, or cleaning, we take a general view of the data. We do this by printing the dimensions of the data available, the columns that it contains and reading their description if available. Number of missing values (nan values in python) are also printed. The most commonly used library for such processes (in Python) is Pandas, using structures called DataFrames. It facilitates a lot of tools involved in exploration of the data.

Data Cleaning

The next step, Data cleaning, is a simple process where we clean small variations and mistakes from the data to make it easily understandable (or less confusing). We generally check the data for invalid values (e.g., age=2000), standardize the strings (“make”, “male”, “Male” all might mean the same thing when put in gender column). We also deal with nan values, however the methods for doing so vary in number. E.g., we might choose to remove the rows containing nan values, or fill the nan values with interpolated values or average values of the total dataset or replace them simply with a 0, the method we choose here depends on the context and is the analyst’s call.

Descriptive Analysis

Up next comes is Descriptive Analysis, where we describe the statistical features of data. These features generally include count, mean, variance, percentiles, modality, distribution for numerical data and total count, frequency count, modes for categorical data. Obviously, we might choose to analyze more components depending on the immediate needs and these instructions are only general in nature. Histograms, kernel density estimation plots, and boxplots are some common plots used for this purpose.

Pre-processing

Categorical variables

Categorical levels can be merged to form lesser number of more relevant levels

Categorical data can be converted to numerical data by methods such as one-hot encoding or Likert scale.

Likert scale is a popular scale where the response varies from strongly disagree, disagree, neither disagree nor agree, agree, strongly agree to 1,2,3,4,5 on certain basic assumptions.

Numerical variables

Numerical features are standardized so that simple magnitude of variables do not affect the significance of other variables.

A common way to do it is to subtract the features by their mean and divide by their standard deviation.

$$\frac{x-\bar{x}}{\sigma}$$

Principal Component Analysis (PCA)

PCA has an underlying assumption that the data available is linear in some nature, that is the variance in the data can be explained by linear structures.

Principal components are coefficients for linear combinations of features which explain the variation in decreasing order of relevance. So, first principal components explain maximum variance in the system.

In terms of linear algebra, PCs are the eigenvectors of the correlation matrix wrt decreasing magnitude of eigenvalues.

Step 5: Extracting Segments

Grouping Consumers

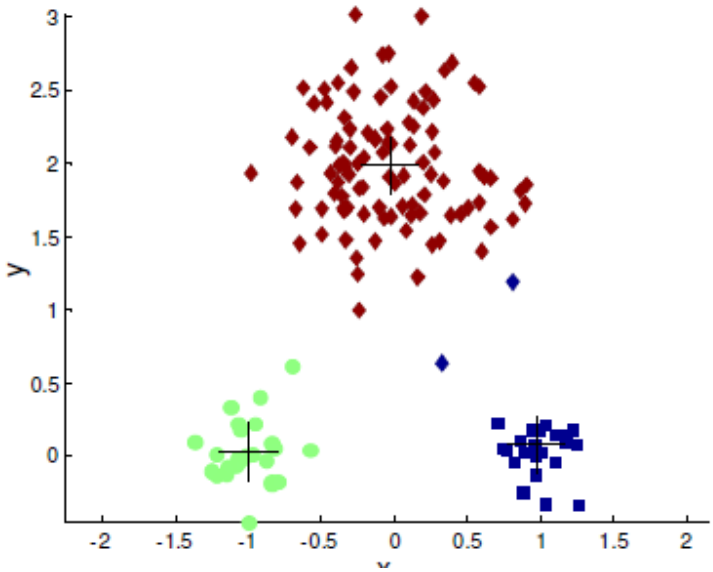
This step is the result of market segmentation analysis. Many segmentation methods used to extract market segments are taken from the field of cluster analysis. selecting a suitable clustering method requires matching the data analytic features of the resulting clustering with the context-dependent requirements that are desired.

Types of clustering techniques :

1. K-Means
2. Hierarchical clustering techniques
3. Density-based spatial clustering of applications with noise (DBSCAN)

K-Means :

- It is popular, simple centroid based clustering technique.



- You have to give k=3 (number of clusters) , it will group them into 3 clusters and assign every cluster a centroid which is a central point . When you say I want k- clusters, you get k centroids (C1, C2, C3… Ck) and k set of points (S1, S2, S3 …Sk).
- Each point is assigned to the cluster to the closest centroid.
- The objective is to minimize the sum of the distances of the points to their respective centroid.
- The centroid depends on the distance function.
- Closeness is measured by Euclidean distance, cosine similarity.
- Finding a centroid is not always easy.

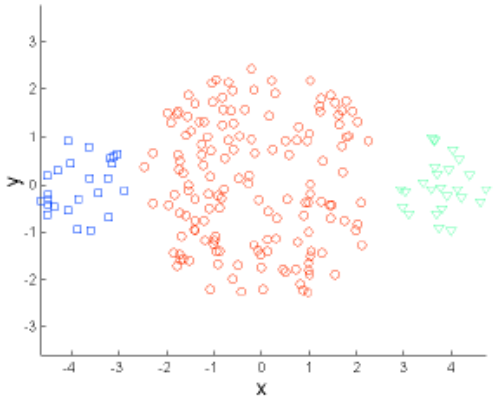
Limitations of K-Means:

K-Means has a problem when data is of different

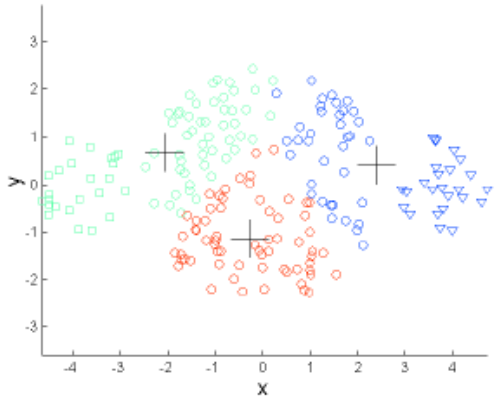
- Sizes
- Shapes
- Densities

Also has a problem with data contains outliers.

Limitation of K-Means : differing sizes

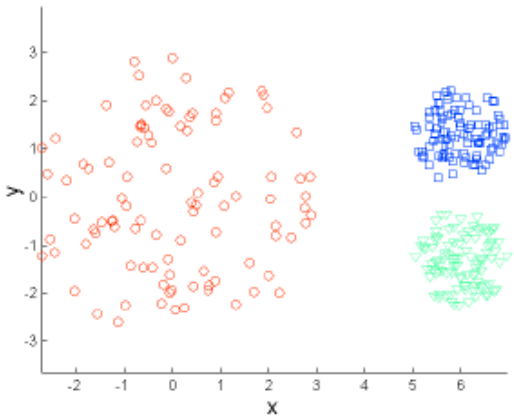


Original Points

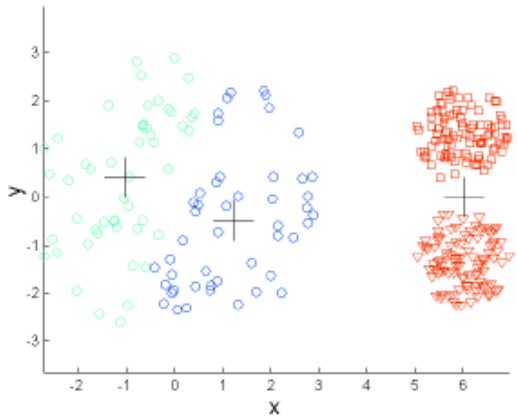


K-means (3 Clusters)

Limitation of K-Means : differing densities.

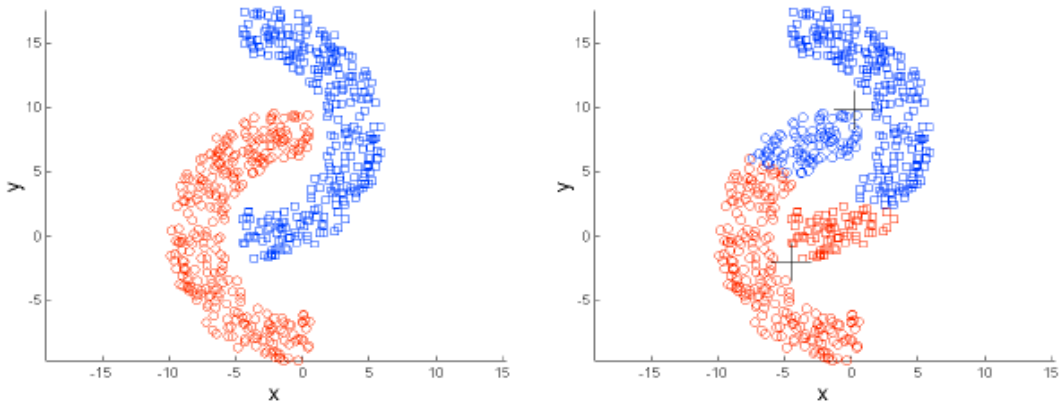


Original Points



K-means (3 Clusters)

Limitation of K-Means : Non-globular shapes



Variation of K-Means:

K-Medoids - the centroid of the cluster is defined to be one of the points in the cluster unlike in k-means the centroid are mean value of the data points.so k-medoids are interpretable.

Hierarchical clustering:

It is a similarity measure or distance measure between clusters (not points).

Two main types of Hierarchical clustering are as follows:

Agglomerative:

- It is more popular.
- Starts with point as individual clusters.
- At each step, merges the closest pair of clusters until only one cluster left.

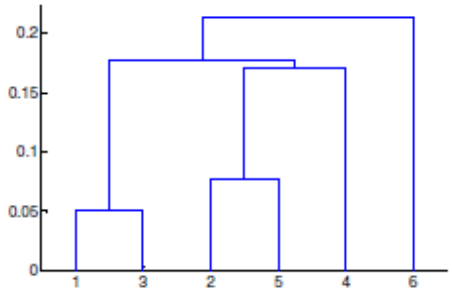
Divisive:

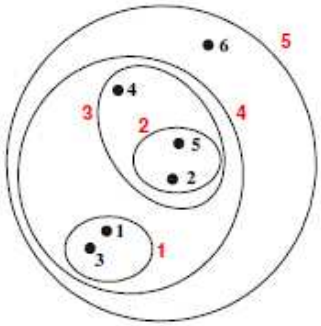
- Start with one, all- inclusive cluster.
- At each step , split a cluster until each cluster contains a point.

Hierarchical clustering uses a similarity matrix or distance matrix

- Merge or split one cluster at a time.
- Produces a set of nested cluster organised as hierarchical tree.
- Can be visualised as a dendrogram

- Dendrogram – a tree like diagram that records the sequence of merges and split.





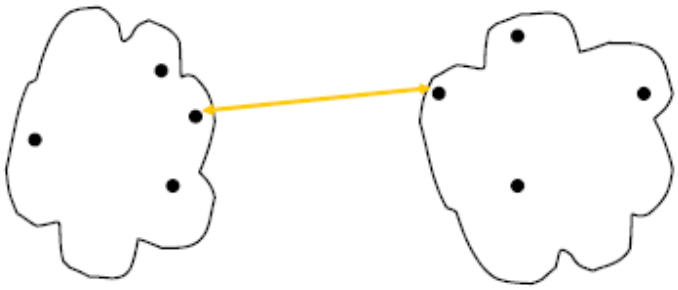
- Here do not have to assume any particular number of clusters
- Any desired number of clusters can be obtained by cutting the dendrogram at the proper level
- Basic algorithm
    1. Compute the similarity matrix.
    2. Let each data point be a cluster
    3. Repeat step 4 and 5
    4. Merge the two closest clusters
    5. Update the proximity matrix
    6. Until a only single cluster remains.

How to define inter- cluster similarity:

- MIN
- MAX
- Group Average
- Ward’s method

MIN :

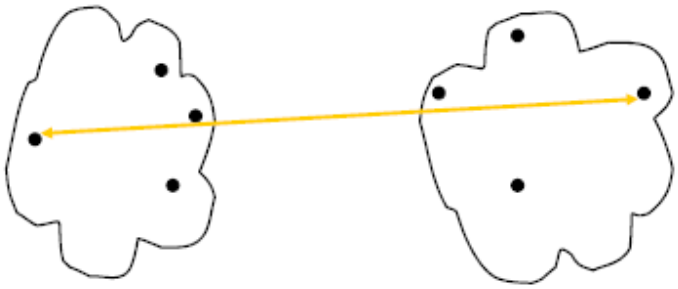
- Take all pair distance between all points of cluster C1 and C2 and take minimum of it or picks smallest distance between points C1 and C2.



C1 and C2

MAX:

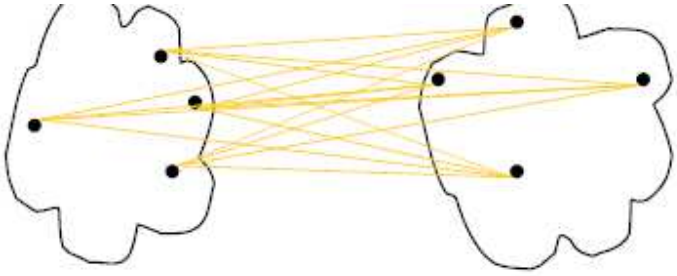
- Take maximum distance and use their similarity between cluster C1 and C2 (takes farthest point).



C1 and C2

Group Average:

- Takes every pair of points , computes their similarity and take average.

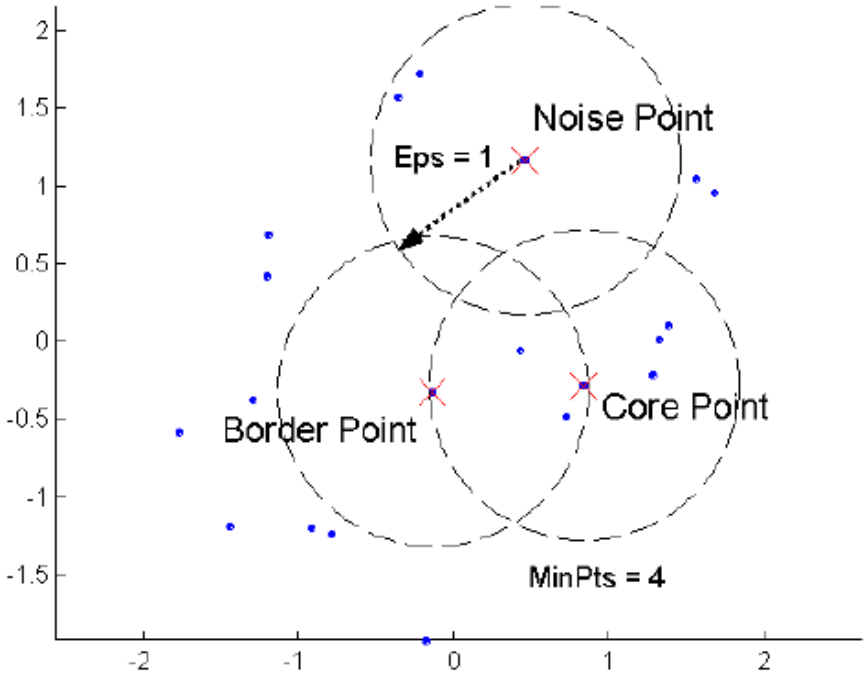


Ward's Method:

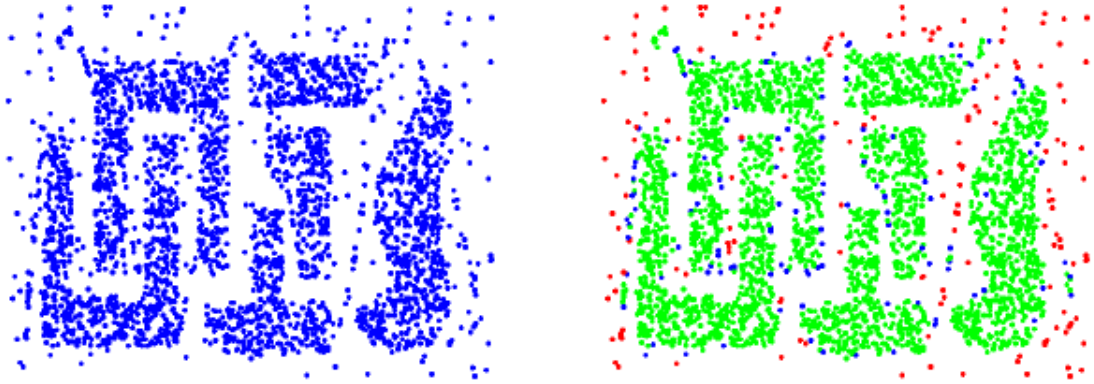
- The distance between clusters is the sum of squared differences within all clusters.

Density-based spatial clustering of applications with noise (DBSCAN):

- In density-based clustering we partition points into dense regions separated by sparser regions.
- Density at point p: number of points within a circle of radius Eps.
- Dense Region: A circle of radius Eps that contains at least MinPts points.
- Characterization of points
- A point is a core point if it has more than as specified number of points (MinPts) within Eps.
- These points belong in a dense region and are at the interior of a cluster.
- A border point has fewer than MinPts within Eps, but is in the neighborhood of a core point.
- A noise point is any point that is not a core point or a border point.



DBSCAN : Core Points , Border Points and Noise points



Original Points

Point types: core, border and noise

Eps = 10, MinPts = 4

DBSCAN : Core Points , Border Points and Noise points

Advantages:

- Resistant to noise.
- Can handle clusters of different sizes

Disadvantages:

- Does not work well with varying densities.
- Does not work well with high dimensions data.

STEP 6: Profiling Segments

This segment is used to define the key features of extended market segments.

Identifying Key Characteristics of Market Segments

The aim of profiling segment is to know the market segment resulted from extraction step and it is only required when data driven market segmentation is used, it is predefined in commonsense segmentation.

This stage we inspect number of different profiling market solution and it is necessary if no natural segment exist in data, either reproducible or a constructive market segmentation approach has to be taken. Good profiling leads to correct interpretation of resulting segments and correct interpretation result to making good strategic marketing decisions. Data-driven market segmentation solutions are not easy to interpret.

Traditional Approaches to Profiling Market Segments

Data-driven segmentation solutions are usually presented to users (clients, managers) in one of two ways: (1) as high level summaries simplifying segment characteristics to a point where they are misleadingly trivial, or (2) as large tables that provide, for each segment, exact percentages for each segmentation variable. Such tables are hard to interpret, and it is virtually impossible to get a quick overview of the key insights.

Segment Profiling with Visualizations

Data visualization using graphics is an integral part of statistical data analysis. Graphics are particularly important in exploratory statistical analysis because they provide insights into the complex relationships between variables.

Identifying Defining Characteristics of Market Segments

A good way to understand the defining characteristics of each segment is to produce a segment profile plot.

The segment profile plot shows – for all segmentation variables – how each market segment differs from the overall sample.

Assessing Segment Separation

Segment separation can be visualized in a segment separation plot.

The segment separation plot depicts – for all relevant dimensions of the data space – the overlap of segments.

Segment separation plots are very simple if the number of segmentation variables is low, but become complex as the number of segmentation variables increases.

Check list

Use the selected segments from Step 5.

Visualize segment profiles to learn about what makes each segment distinct.

Use knock-out criteria to check if any of the segments currently under consideration should already be eliminated because they do not comply with the knock-out criteria.

Pass on the remaining segments to Step 7 for describing.

Step 7: Describing Segments

Segment profiling is about understanding differences in segmentation variables across market segments. In Step 7 market segments are described using additional information available about segment members. These additional variables are referred to as descriptor variables.

Using Visualisations to Describe Market Segments:

Mosaic Plots: - For Categorical Variables

The mosaic plot also visualises cross-tabulations. The width of the bars indicates the absolute segment size. The height of the rectangles represents the proportion of men or women in each segment.

Cell colours are based on the standardised difference between the expected and observed frequencies. Negative differences mean that observed are lower than expected. They are coloured in red. Positive differences mean that observed are higher than expected frequencies. They are coloured in blue. The saturation of the colour indicates the absolute value of the standardised difference.

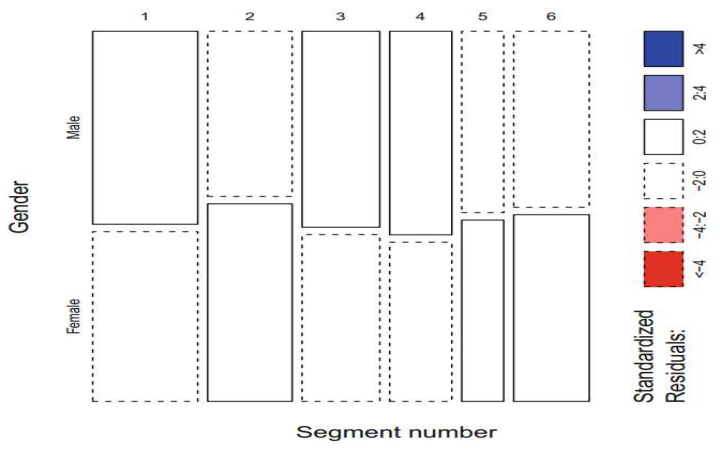


Fig. 9.2 Shaded mosaic plot for cross-tabulation of segment membership and gender for the Australian travel motives data set

All cells are white, indicating that the six market segments extracted from the Australian travel motives data set do not significantly differ in gender distribution. The proportion of female and male tourists is approximately the same across segments. The dashed and solid borders of the rectangles indicate that the number of respondents in those cells are either lower than expected (dashed borders), or higher than expected (solid black borders). But, irrespective of the borders, white rectangles mean differences are statistically insignificant.

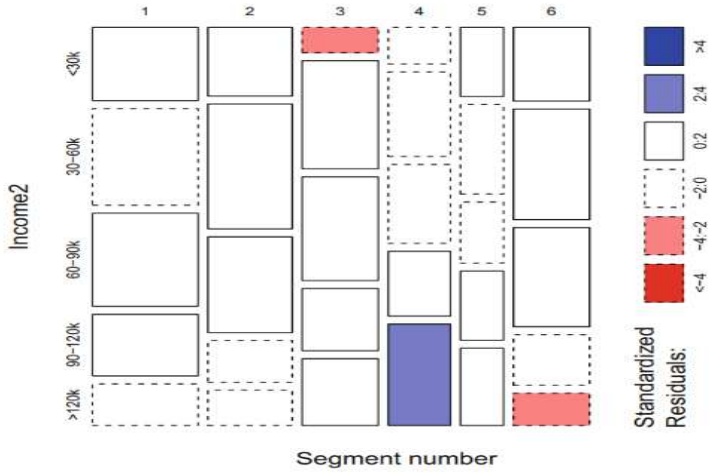


Fig. 9.3 Shaded mosaic plot for cross-tabulation of segment membership and income for the Australian travel motives data set

Histograms and Box Plots: - For Metric variables

Conditional in this context means that the plots are divided in sections (panels, facets), each presenting the results for a subset of the data (for example, different market segments). Conditional plots are well-suited for visualising differences between market segments using metric descriptor variables.

In a histogram, the differences between market segments are difficult to assess just by looking at the plots.

Predicting segments from descriptor variables

In this we predict segment membership from descriptor variables. To achieve this, we use a regression model with the segment membership as categorical dependent variable, and descriptor variables as independent variables. The prediction performance indicates how well members of a market segment can be identified given the descriptor variables.

Binary Logistic Regression

Here, we fit the model to predict the likelihood of a consumer to belong to a particular segment given their descriptor variables. The dependent variable is a binary indicator of being in the segment. The output contains the regression coefficients, and information on the model fit, including the degrees of freedom, the null deviance, the residual deviance, and the AIC.

Multinomial Logistic Regression

Multinomial logistic regression can fit a model that predicts each segment simultaneously. Because segment extraction typically results in more than two market segments, the dependent variable y is not binary. Rather, it is categorical and assumed to follow a multinomial distribution with the logistic function as link function.

The fitted model contains regression coefficients for each segment. We assess the predictive performance of the fitted model by comparing the predicted segment membership to the observed segment membership. Starting with the full model containing all available independent variables, the procedure returns the best-fitting model, the model which deteriorates in AIC if an independent variable is either dropped or additionally included. So, from this we can identify the important descriptor variables which help us in describing a segment.

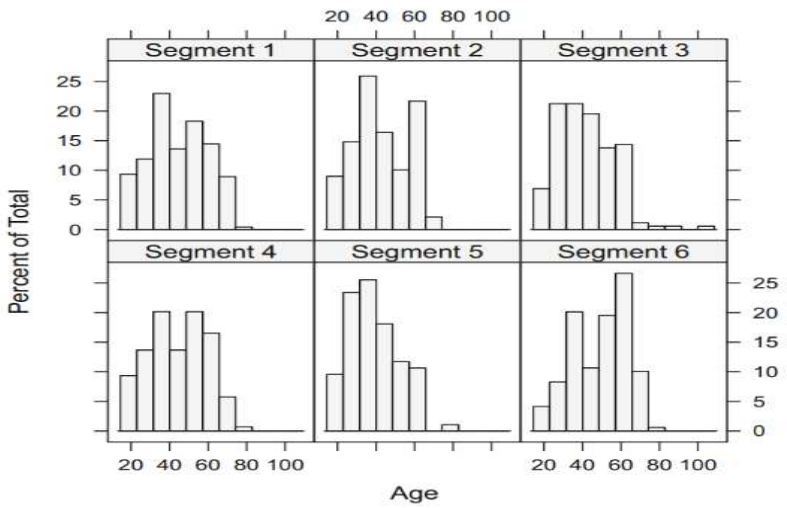


Fig. 9.5 Histograms of age by segment for the Australian travel motives data set

Tree-Based Methods

At each step, consumers are split into groups based on one independent variable. The aim of the split is for the resulting groups to be as pure as possible with respect to the dependent variable. This means that consumers in the resulting groups have similar values for the dependent variable

We predict segment membership by moving down the tree. At each node, we move down the branch reflecting the consumer's independent variable. When we reach the terminal node, segment membership can be predicted based on the segment memberships of consumers contained in the terminal node.

Step 7 checklist:

Bring across from Step 6 (profiling) one or a small number of market segmentation solutions selected on the basis of attractive profiles.

Select descriptor variables. Descriptor variables are additional pieces of information about each consumer included in the market segmentation analysis. Descriptor variables have not been used to extract the market segments.

Use visualisation techniques to gain insight into the differences between market segments with respect to descriptor variables. Make sure you use appropriate plots, for example, mosaic plots for categorical and ordinal descriptor variables, and box-and-whisker plots for metric descriptor variables.

Test for statistical significance of descriptor variables. If you used separate statistical tests for each descriptor variable, correct for multiple testing to avoid overestimating significance.

"Introduce" each market segment to the other team members to check how much you know about these market segments.

Ask if additional insight into some segments is required to develop a full picture of them.

Step 8: selecting the target segments

The targeting decision

Which market segments are selected for targeting? M.S is a marketing tool. Selection of target segments affects the performance of the organization. The market segments are inspected earlier and selected for targeting. All the market segments that are considered to be selected as target market segments should have passed the knock out criteria test.

Checklist:

- Convene a segmentation team meeting
- Determine which of the market segments profiled in step6 and described in step7 are being considered as potential target markets.
- Double check that all of those remaining segments comply with the knock out criteria of homogeneity, distinctness, size, match, identifiability and reachability. If a segment does not comply, eliminate it from further consideration.
- Discuss and agree on values for each market segment for each segment attractiveness criterion.
- Discuss and agree on values for each relative organizational competitiveness criterion for each of the market segments.
- Calculate each segment's overall attractiveness by multiplying the segment value with the weight for each criterion and then summing up all these values for each segment.
- Calculate each segment's overall relative organizational competitiveness by multiplying the segment value with the weight for each criterion and then summing up all these values for each segment.
- Plot the values into a segment evaluation plot.
- Make a preliminary selection.
- If you intend to target more than one segment, make sure that the selected target segments are compatible with one another.
- Present the selected segments to the advisory committee for discussion and reconsideration.

Step 9: Customizing the Market Mix

One of the most difficult tasks for organisations is gaining a marketing edge in a competitive environment. The purpose of this article is to define market segmentation and to examine the role of the 4Ps (product, price, location, and promotion) in getting a competitive edge over competitors by providing customers with higher-value products or services.

Implication for Marketing Mix Decision

Marketing was originally seen as a toolbox to assist in selling products, with marketers that marketers have at their disposal 12 ingredients: product planning, packaging, Market segmentation does not stand independently as a marketing strategy. In fact, the segmentation process is frequently seen as part of what is referred to as the segmentation-targeting-positioning

(STP) The segmentation-targeting-positioning approach postulates a sequential process. It entails segmenting a large market into recognizable and separate groups or segments, each with shared features and demands, as well as similar responses to marketing efforts. Data is used to segment a market based on different consumer, user, organizational, and market attributes. Market segmentation is used to deal with the economic realities of resource scarcity. Companies can't create all conceivable items or services for all individuals all of the time since they have limited resources. Companies will also guarantee that the marketing mix's aspects (product, price, location, and promotion) are planned and adjusted to fit the requirements and wants of various consumer groups. While this is frequently the case, the primary goal of segmenting is to identify the race (segment) that the company's automobile can win.

The process starts with market segmentation positioning (the measures an organization can take to ensure that their product is approach is useful because it ensures that segmentation is not seen adhere too strictly to the sequential nature of the segmentation-targeting-positioning to the targeting step, before being in the position of making a long-term commitment to one or a small number of target segments.

The above figure illustrates how the target segment decision – which has to be model of the marketing mix including Product, Price, Place and Promotion serves To best ensure maximizing on the benefits of a market segmentation strategy, it is important to customize the marketing mix to the target segment. Discount structures (Price), the selection of suitable distribution channels (Place), are attractive to the target segment (Promotion).

Segmentation analysis around one of the 4Ps, affects the choice of segmentation, for example,

If the segmentation analysis is undertaken to inform pricing segmentation variables.

If the market segmentation analysis is conducted to inform advertising decisions.

If the market segmentation analysis is conducted for the purpose of informing typically,

However, market segmentation analysis is not conducted how to develop or adjust the marketing mix to best cater for the target segment.

Looking into the 4Ps

A sector is also worth pursuing if a firm can better address the segment's requirements and wants than its competitors while still remaining profitable. If no solutions currently meet these requirements, the firm has discovered a market with no competitors. If a corporation can identify a need that isn't being addressed but that it can fill, it may have stumbled onto innovation and therefore be the first in a product category. Consider a corporation situated in Rwanda that wants to sell automobiles as an example of market segmentation.

If the company's business is to sell vehicles in Rwanda, the target market will be quite wide, with many various types of clients with varying requirements, wants, expectations, buying patterns, and occasions. Let's say that after conducting market research, the corporation discovers that there are several groups of potential clients that require automobiles for various reasons, but each group has a lot of characteristics in terms of need, desire, expected outcomes, purchasing patterns, and occasions.

Product & Price

One of the key decisions an organization needs to make when developing the product dimension of the marketing mix, is to specify the product in view of customer needs. Other marketing mix decisions that fall under the product dimension are: naming the product, packaging it, offering or not offering warranties, and after sales support services.

For example the Australian vacation activities data set for illustrating how product design or modification is driven by target segment and imagine having chosen to target segment 3 members in terms of vacation activities are that they engage much more than the average tourist in visiting museums, monuments and gardens. In terms of the product targeted at this market segment, possible product MONUMENTS & MUCH, MUCH MORE product (accompanied by an activities pass) points to the existence of these offers at the destination during the vacation planning another opportunity for targeting this segment is that of proactively making. Setting the price for a product and selecting on to be supplied are two common decisions an organization must make when building the pricing dimension of the marketing mix.

Place

How to distribute the product to clients is the most important choice related to the location dimension of the marketing mix. This involves deciding if the product should be sold online alone, offline only, or both; whether the maker should sell directly to customers; and whether a wholesaler, retailer, or both should be employed. Returning to the segment 3 members and the destination with a strong cultural heritage example, the survey on which the market segmentation study was based also asked survey respondents to specify how they booked their previous domestic vacation. Respondents were given the opportunity to select from a variety of

alternatives. Members of segment 3 differ from other tourists in terms of how they booked their hotel on their most recent domestic holiday, as seen by the example above: they book their hotel online significantly more frequently than the ordinary tourist. This knowledge has significant consequences for the marketing mix's location dimension. The hotel must have an online booking option accessible. It would be extremely valuable to gather information on members of segment 3's booking of other products, services, and activities to determine if they book the majority of their items, services, and activities online, or if their online booking behavior is confined to the lodging.

Promotion

When creating a marketing mix, common promotion considerations involve creating an advertising message that will resonate with the target market and determining the most successful manner to communicate this message. Public relations, personal selling, and sponsorship are all tactics in the promotion area of the marketing mix. Returning to segment 3, we must identify the most effective information sources for contacting segment 3 members in order to notify them about the MUSEUMS, MONUMENTS & MUCH, MUCH MORE goods. We investigate their favorite TV stations and compare the information sources they used on their most recent domestic vacation.

Code Implementation:

Data Overview:

Understanding the Dataset

In [8]:
```
df  = pd.read_csv('./mcdonalds.csv')
print('data dimensions ',df.shape)
print (df.columns)
```

```
data dimensions  (1453, 15)
Index(['yummy', 'convenient', 'spicy', 'fattening', 'greasy', 'fast', 'cheap',
       'tasty', 'expensive', 'healthy', 'disgusting', 'Like', 'Age',
       'VisitFrequency', 'Gender'],
      dtype='object')
```

In [9]:
```
#print first 5 rows of the dataset
print(df.head())
```

```
  yummy convenient spicy fattening greasy fast cheap tasty expensive healthy  \
0    No        Yes    No       Yes     No  Yes   Yes    No       Yes      No
1   Yes        Yes    No       Yes    Yes  Yes   Yes   Yes       Yes      No
2    No        Yes   Yes       Yes    Yes  Yes    No   Yes       Yes     Yes
3   Yes        Yes    No       Yes    Yes  Yes   Yes   Yes        No      No
4    No        Yes    No       Yes    Yes  Yes   Yes    No        No     Yes

  disgusting Like  Age      VisitFrequency  Gender
0         No   -3   61  Every three months  Female
1         No   +2   51  Every three months  Female
2         No   +1   62  Every three months  Female
3        Yes   +4   69         Once a week  Female
4         No   +2   49        Once a month    Male
```

Data Cleaning And Pre-Processing:

Data Cleaning and Pre-processing

In [11]:
```
#check NaN Values
num_na = df.isna().sum()
print('nan values',num_na)
```

```
nan values yummy             0
convenient        0
spicy             0
fattening         0
greasy            0
fast              0
cheap             0
tasty             0
expensive         0
healthy           0
disgusting        0
Like              0
Age               0
VisitFrequency    0
Gender            0
dtype: int64
```

Check unique values

[12]:
```
def check_unique():
    for i in df:
        print(i,' ---- ', df[i].unique())

check_unique()
```

```
yummy  ----  ['No' 'Yes']
convenient  ----  ['Yes' 'No']
spicy  ----  ['No' 'Yes']
fattening  ----  ['Yes' 'No']
greasy  ----  ['No' 'Yes']
fast  ----  ['Yes' 'No']
cheap  ----  ['Yes' 'No']
tasty  ----  ['No' 'Yes']
expensive  ----  ['Yes' 'No']
healthy  ----  ['No' 'Yes']
disgusting  ----  ['No' 'Yes']
Like  ----  ['-3' '+2' '+1' '+4' 'I love it!+5' 'I hate it!-5' '-2' '+3' '0' '-4' '-1']
Age  ----  [61 51 62 69 49 55 56 23 58 32 53 28 65 54 67 34 31 47 37 41 36 50 39 35
 20 24 44 40 48 38 57 60 66 42 26 52 29 25 22 45 18 68 43 21 27 33 63 46
 59 19 64 70 30 71]
VisitFrequency  ----  ['Every three months' 'Once a week' 'Once a month' 'Once a year'
 'More than once a week' 'Never']
Gender  ----  ['Female' 'Male']
```

Data Visualization:

Principle Component Analysis

In [17]:
```
pca = PCA(n_components=11)
pca.fit(segment_df)
print('Explained Variance Ratio by PCs ',pca.explained_variance_ratio_)
```

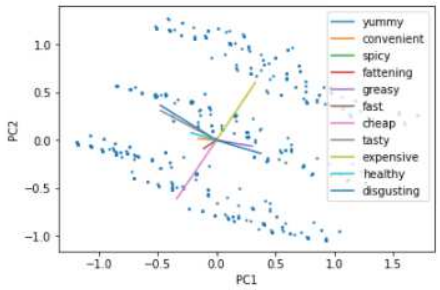
```
Explained Variance Ratio by PCs  [0.29944723 0.19279721 0.13304535 0.08309578 0.05948052 0.05029956
 0.0438491  0.03954779 0.0367609  0.03235329 0.02932326]
```

In [18]:
```
out = pca.transform(segment_df)
plot_data = out[:,0:2]
components = pca.components_
# print('components type is ', type(components))
# print('components shape is ', components.shape)
# print(components)
plot_components = components[0:2,:]
print('plot components shape is ', plot_components.shape)
print('plot component values ',plot_components)
```

```
plot components shape is  (2, 11)
plot component values  [[-0.47693349 -0.15533159 -0.00635636  0.11623168  0.3044427  -0.10849325
  -0.33718593 -0.47151394  0.32904173 -0.21371062  0.37475293]
 [ 0.36378978  0.016414    0.01880869 -0.03409395 -0.06383884 -0.0869722
  -0.61063276  0.3073178   0.60128596  0.07659344 -0.13965633]]
```

In [19]:
```
#plotting Scatter plot
plt.scatter(plot_data[:,0],plot_data[:,1],s=2, marker='o')
for i in range(11):
    plt.plot([0,plot_components[0,i]],[0,plot_components[1,i]], label=binary_columnns[i])
plt.legend()
plt.xlabel('PC1')
plt.ylabel('PC2')
```

Out[19]: Text(0,0.5,'PC2')

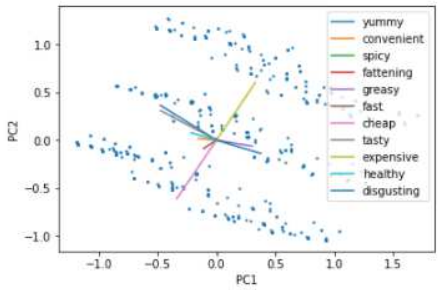


Observations

1. features cheap and expensive are independent of each other , so it plays important role in evaluation.
2. fattening ,disgusting and creasing pointing in the same direction ,so consumers who views's Mc'donalds fattening ,disgusting also likley to views it as creasing.
3. fast ,convenint ,tasty ,yummy and healthy pointing in same direction and it is postive features.

In [19]:
```
#plotting Scatter plot
plt.scatter(plot_data[:,0],plot_data[:,1],s=2, marker='o')
for i in range(11):
    plt.plot([0,plot_components[0,i]],[0,plot_components[1,i]], label=binary_columnns[i])
plt.legend()
plt.xlabel('PC1')
plt.ylabel('PC2')
```

Out[19]:
```
Text(0,0.5,'PC2')
```



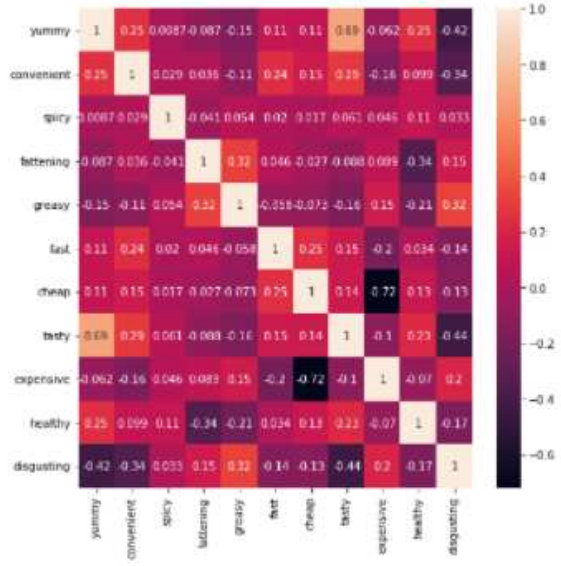
Observations

1. features cheap and expensive are independent of each other , so it plays important role in evaluation.
2. fattening ,disgusting and creasing pointing in the same direction ,so consumers who views's Mc'donalds fattening ,disgusting also likley to views it as creasing.
3. fast ,convenint ,tasty ,yummy and healthy pointing in same direction and it is postive features.

Heat Map:

#Heat Map

In [23]:
```
#Heat Map
import seaborn as sns #for plotting
plt.figure(figsize=(8,8))
sns.heatmap(segment_df.corr(), annot=True)
plt.show()
```



Observations:

1. the attributes cheap and expensive are independent of each other beacuse have value -0.72
2. positive attributes like tasty and yummy have high values 0.69

K-Means:

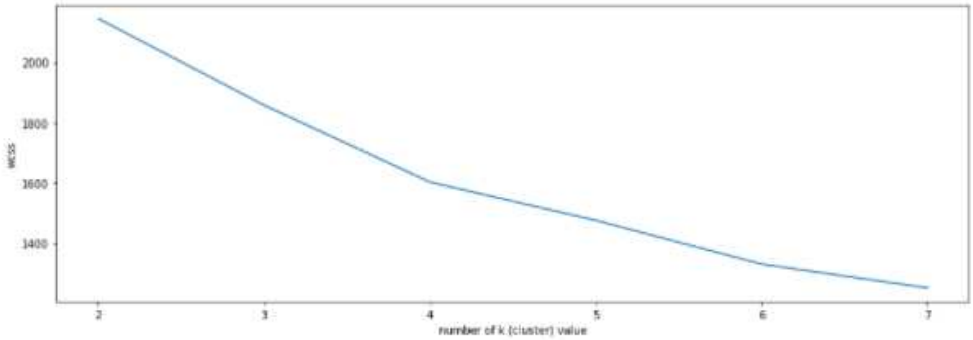
K-Means:

In [24]:
```
#Grouping consumers
#1 Applying K-Means with K=2 to k=8
from sklearn.cluster import KMeans

wcss = []
for k in range(2,8):
    kmeans = KMeans(n_clusters=k)
    kmeans.fit(segment_df)
    wcss.append(kmeans.inertia_)

kmeans.cluster_centers_

# the best value is elbow value.
plt.figure(figsize=(15,5))
plt.plot(range(2,8),wcss)
plt.xlabel("number of k (cluster) value")
plt.ylabel("wcss")
plt.show()
```



Observations :

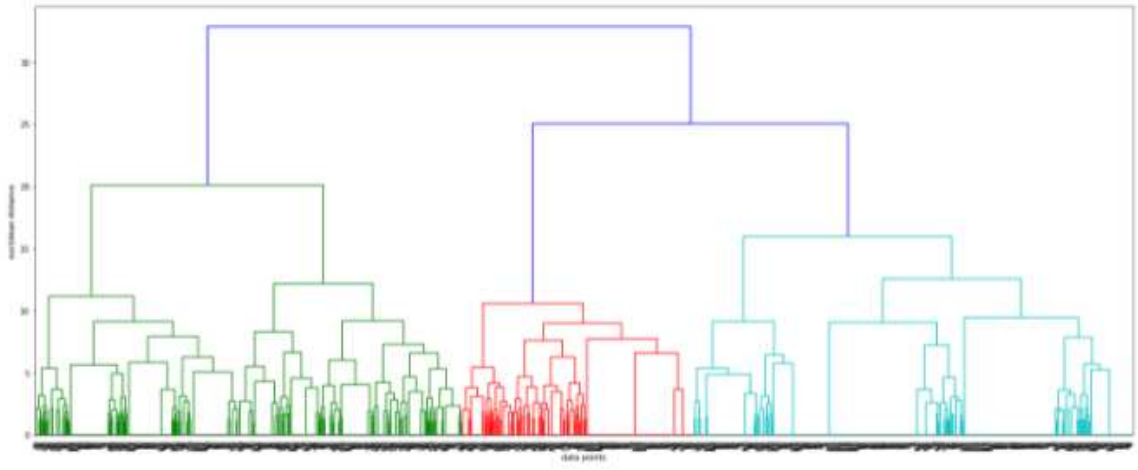
1. Using this Elbow method number of clusters can be determined.
2. we can choose 4 as a number of clusters.

hierarchical clustering

Hierarchical Clustering:

[25]:
```
#Hierarchical Clustering
#create dendrogram and find the best clustering value
# ward distance = The distance between clusters is the sum of squared differences within all clusters
from scipy.cluster.hierarchy import linkage, dendrogram
from sklearn.cluster import AgglomerativeClustering

merg = linkage(segment_df,method="ward")
plt.figure(figsize=(25,10))
dendrogram(merg,leaf_rotation = 90)
plt.xlabel("data points")
plt.ylabel("euclidean distance")
plt.show()
```



Observations :

1. Hierarchy clustering is a similarity measure or dist measure between clusters.
2. Dendogram can be used for visualizations,close clusters are merged together.
3. At y-axis between 4 and 5 corresponding to merging of datapoints in x- axis looks reasonable.
4. so it is good to grouped consumers in 4 clusters.

GitHub Link -