Market Segmentation Summary

--NETHRANAND P.S

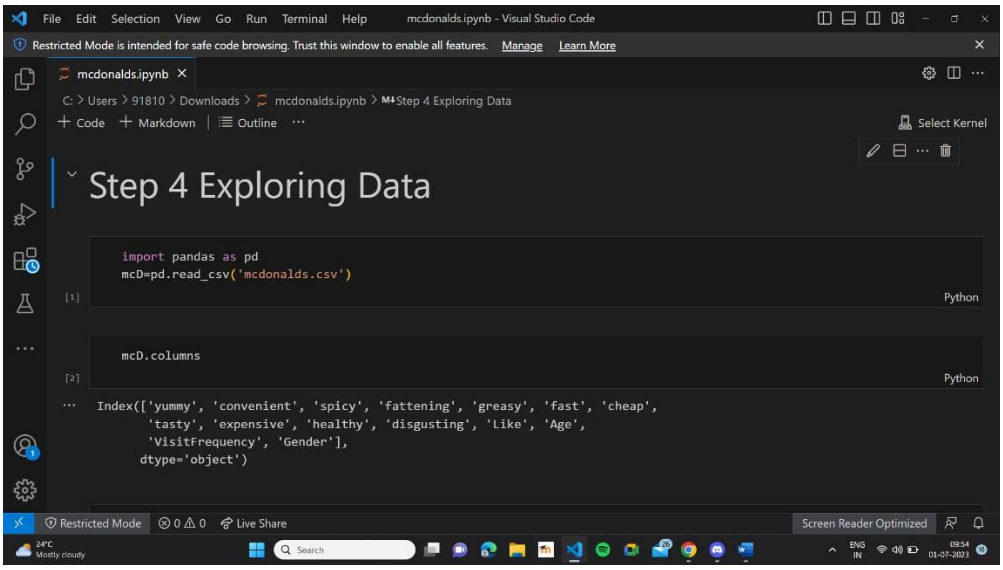
- ## McDonalds_Market_Segmentation_Analysis:

GitHub Repository Link - https://github.com/nethranand01/feynn_intern

Step 4: Exploring Data

4.1 A first Glimpse at Data

- Exploratory data analysis purges and, if necessary, preprocesses the data after data collection. The most appropriate algorithm for extracting valuable market segments is also provided guidance during this exploratory stage. Technically speaking, data exploration aids in determining the measurement levels of the variables, investigating their univariate distributions, and evaluating the dependence structures among them. Additionally, preprocessing and preparation of the data may be necessary before it can be utilised as input for various segmentation techniques. Results from the data exploration step shed light on whether certain segmentation techniques are effective for isolating market segments.

- 

4.2 Data Cleaning

- Cleaning the data is the first step before starting data analysis. This entails verifying that all values have been recorded accurately and that the levels of categorical variables have been given consistent labels. The majority of metric variables' The range of conceivable values is known beforehand. Similar to categorical variables, levels can be verified to make sure they only contain acceptable values. For instance, the gender variable in surveys normally contains two values: female and male. Only those two should show up in the statistics, excluding the possibility that the questionnaire provided a third choice. Any other numbers must be adjusted as part of the data cleaning process because they are not acceptable. Here are some key steps involved in data cleaning:
- Handling Missing Values
- Removing Duplicates
- Correcting Inaccurate or inconsistent data
- Validating Data
- Handling Outliers

4.3 Descriptive Analysis

- Information about the data is provided by descriptive numerical and graphical representations. Descriptive analysis tools are available in a wide range of statistical software packages. With the command summary in R, we can get a numerical summary of the data (). For numerical variables, this command returns the mean, the quartiles, and the range. Frequency counts are returned by the command for categorical variables. The number of missing values for each variable is also returned by the command.
- Histograms, boxplots, and scatter plots are useful graphical techniques for numerical data. Frequency count bar charts are effective for visualising categorical variables. The relationship of numerous categorical variables is shown through mosaic plots. In Step 7, where we use them to compare market segments, we define mosaic plots. Histograms show how numerical variables are distributed. They display the frequency of observations falling inside a given value range. Histograms show whether a variable's distribution is symmetrical and unimodal, or if it is skewed. We must first define categories of values before we can obtain a histogram. This is known as binning. The bins must be close to one another and must include the whole range of observations. They typically have equal lengths. After establishing the bins, we use one bar for each bin to plot the proportion of observations that fall into each bin. We plot the frequency of observations in each bin on the y-axis and the bin range on the x-axis.

4.4 Pre-Processing

- ### 4.4.1 Categorical Variables
- For categorical variables, two pre-processing techniques are frequently utilised. Before further analysis, one method combines levels of category variables; the other, if appropriate, converts categorical variables to numeric ones. Merging levels of categorical variables is useful if the original categories are too differentiated (too many).

- 4.4.2 Numeric Variables
- In distance-based methods of segment extraction, a segmentation variable's relative influence depends on the range of values it can take. Factors can be standardised to balance how segmentation variables affect segmentation outcomes. Standardizing variables entails putting them on a common scale by converting them.
- The default standardisation method in statistics subtracts the empirical mean x¯ and divides by the empirical standard deviation s:

$$z_i = \frac{x_i - \bar{x}}{s},$$

with

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2,$$

- for the n observations of a variable x = {x1,...,xn}. This implies that the empirical mean and the empirical standard deviation of z are 0 and 1, respectively. Standardisation can be done using function scale().

Step 5: Extracting Segments

- In market segmentation analysis, consumers are grouped based on their similarities and differences in preferences and behaviours. However, consumer data sets are often unstructured, and consumer preferences are spread across the entire range of possibilities, making it difficult to identify clear groups of consumers.

- Segmentation methods, often borrowed from cluster analysis, are used to extract market segments from consumer data. Different clustering algorithms impose different structures on the extracted segments. For example, the k-means cluster analysis aims to find compact clusters covering a similar range in all dimensions, while single linkage hierarchical clustering constructs snake-shaped clusters. The choice of the clustering method and the number of segments specified upfront can greatly impact the segmentation solution.

- It is important to explore market segmentation solutions derived from a range of different clustering methods and understand how different algorithms impose structure on the segments. No single algorithm is best for all datasets, and the interaction between the data and the algorithm is critical. Well-structured and well-separated data may require less consideration of the algorithm's tendencies, while less structured data will be more influenced by the algorithm's structure-imposing tendencies.

- When selecting an algorithm for market segmentation, researchers should consider the characteristics of the data set (size, scale level of variables, special structure) and the desired segment characteristics (similarities within segments, differences between segments, number and size of segments). It is also important to compare and investigate alternative segmentation solutions to arrive at a good final solution.

- The chapter discusses different extraction methods for market segmentation, including distance-based methods and model-based methods. Distance-based methods use similarity or dissimilarity measures to find groups of similar observations, while model-based methods formulate stochastic models for the segments. Some methods combine multiple aims, such as variable selection during the extraction process.

- The selection of an appropriate extraction algorithm depends on the specific characteristics of the data and the desired segment characteristics. There is no single best algorithm, and each method has its own advantages and disadvantages. Therefore, it is crucial to carefully choose and compare different algorithms to find the most suitable solution for a given market segmentation task.

- The example shown in Figure 7.1 illustrates how different algorithms can impose different structures on the data. The top row shows the market segments obtained using k-means cluster analysis, which fails to identify the naturally existing spiral-shaped segments. On the other hand, the bottom row shows the segments obtained using single linkage hierarchical clustering, which correctly identifies the spirals even when an incorrect number of segments is specified.

- This example highlights the importance of the interaction between data and algorithm in shaping the segmentation solution. There is no single best algorithm for all datasets, and the choice of algorithm depends on the characteristics of the data and the desired segment structure.

- To select an appropriate algorithm for market segmentation, it is important to consider the characteristics of the data set, such as its size, the scale level of segmentation variables, and any special structures or additional information. Additionally, the desired segment characteristics, including similarities within segments and differences between segments, should align with the structure imposed by the algorithm.

- The chapter discusses various extraction methods used in market segmentation, including distance-based methods and model-based methods. Distance-based methods rely on a notion of similarity or distance between observations to identify groups of similar consumers. Model-based methods formulate stochastic models for market segments. There are also methods that combine multiple aims, such as variable selection during segment extraction.

- Since there is no single best algorithm, it is important to investigate and compare alternative segmentation solutions. Understanding the data characteristics and the expected segment characteristics can guide the selection of suitable algorithms for comparison.

- Distance-based methods, such as Euclidean distance and Manhattan distance, are commonly used in market segmentation analysis. These measures calculate the distance between two vectors based on their values in each dimension. Another distance measure, asymmetric binary distance, is specific to binary vectors and considers the presence or absence of 1s in common between vectors.

- Overall, the choice of algorithm for grouping consumers in market segmentation depends on the data characteristics, desired segment structure, and the interaction between the data and the algorithm.

Step 6: Profiling Segments

- In market segmentation, profiling refers to the process of identifying and characterizing the key characteristics of different market segments. This step is necessary when data-driven market segmentation is used, as the defining characteristics of the segments are unknown until after the data has been analysed. Profiling involves examining the market segments individually and comparing them to each other.

- When conducting commonsense segmentation, where predefined profiles are used (e.g., age groups), Step 6 of profiling is not necessary. However, with data-driven segmentation, the profiling step is crucial for understanding the unique characteristics of each segment based on the segmentation variables.

- Traditional approaches to profiling market segments often involve presenting the results in tables that provide exact percentages for each segmentation variable in each segment. However, these tables can be difficult to interpret and may not provide a quick overview of the key insights. Managers often struggle to understand data-driven segmentation solutions, and the results are sometimes presented in a manner that is confusing or lacks clarity.

- To make the profiling process less tedious and more intuitive, graphical statistics approaches are recommended. Visualizations, such as segment profile plots, can effectively represent the defining characteristics of market segments. These plots show how each segment differs from the overall sample for each segmentation variable. By comparing the values of each segment to the values of other segments or the total sample, insights about the defining characteristics of each segment can be gained.

- Segment profile plots are panel plots that represent each segment individually, with cluster centres (centroids) displayed for each segment. Marker variables, which are particularly characteristic of a segment, can be highlighted in colour, while other variables are greyed out. This graphical representation makes it easier to interpret and compare the characteristics of different segments.

- Overall, graphical approaches to segment profiling provide a more intuitive and informative way of understanding market segmentation results, making them valuable for strategic marketing decision-making.

- "A relative difference of more than 25% is also required. For example, a variable with a total sample mean of 0.05 and a segment mean of 0.08 would not qualify as a marker variable (0.05 + 0.05 = 0.10 < 0.08). This criterion ensures that only variables with substantial differences between segments and the total sample are highlighted.

- In Figure 8.2, each panel represents one segment, labelled as "Segment 1" to "Segment 6." The vertical bars in each panel indicate the segment means for each segmentation variable. The solid-coloured bars represent the marker variables, which are the variables that significantly differentiate the segment from the overall sample. The greyed-out bars represent the non-marker variables.

- By examining the segment profile plot, it becomes easier to understand the defining characteristics of each segment. For example, in Segment 2, tourists are primarily motivated by rest and relaxation and have a preference for not exceeding their planned travel budget. They also show interest in a change of surroundings but are less concerned about cultural offers, intense nature experiences, price sensitivity, health and beauty, and realizing creativity.

- Segment profiling with visualizations provides a clear and intuitive way to interpret market segments. It allows for a quick overview of the key insights and helps assess the usefulness of a segmentation solution. By visually comparing segment means to the overall sample means, marketers can identify the defining characteristics of each segment and make informed strategic marketing decisions based on these insights.

- In summary, traditional approaches to profiling market segments often rely on complex tables that are difficult to interpret and compare. Visualizations, on the other hand, offer a more accessible and insightful way to understand segment characteristics. By using segment profile plots, marketers can easily identify the defining characteristics of each segment and make more informed decisions."