

“ Market Segmentation Analysis ”

Team Lead : Mr. Pranav Pimpalkar

Team Member 1: Mr. Amey Sachin Darokar

Team Member 2: Mr. Harshal Ingale

Team Member 3: Ms. Prachi Kale

Team Member 4: Ms. Priya

Team Member 5: Mr. Shashikumar Kulkarni

Task Division

Team Lead : Mr. Pranav Pimpalkar – Step 9 : Customising the Marketing Mix

Team Member 1: Mr. Amey Sachin Darokar - Step 6: Profiling Segments

Team Member 2: Mr. Harshal Ingale - Step 5: Extracting Segments

Team Member 3: Ms. Prachi Kale - Step 4: Exploring Data

Team Member 4: Ms. Priya - Step 8: Selecting the Target Segment(s)

Team Member 5: Mr. Shashikumar Kulkarni - Step 7: Describing Segments

{ Step 1, Step 2 and Step 3 Were Common for all . }*

Introduction

Market segmentation is the practice of dividing your target market into approachable groups. Market segmentation creates subsets of a market based on demographics, needs, priorities, common interests, and other psychographic or behavioural criteria used to better understand the target audience.

Market Segmentation involves :

- **Purpose of Marketing** - Marketing aims to align consumer needs with supplier offerings for mutual benefit, guiding an organization's marketing planning. This ensures that products and services meet customer demands, fostering a strong market presence and sustainable growth.
- **Marketing Planning** - Marketing planning consists of strategic and tactical plans. The strategic plan sets long-term goals, akin to choosing a mountain in a hiking expedition, providing direction and purpose. The tactical plan outlines short-term actions, similar to selecting the right gear for the hike, ensuring immediate tasks align with the broader strategy.
- **SWOT Analysis** - A SWOT analysis identifies internal strengths and weaknesses and external opportunities and threats. This foundational tool supports strategic marketing planning by highlighting areas for improvement and potential growth, guiding the formulation of effective strategies.
- **Market Research** - Market research is essential for understanding consumer needs and desires. It employs methods such as surveys, focus groups, and observations to gather insights, forming the basis for informed marketing decisions and strategic planning.
- **Segmentation and Positioning** - Segmentation and positioning are critical in strategic planning. By selecting target consumer segments and positioning the organization to appeal to these segments, companies can tailor their offerings to meet specific needs, enhancing market effectiveness and customer satisfaction.
- **Tactical Marketing Plan** - The tactical marketing plan, based on the strategic plan, focuses on the 4 Ps (Product, Price, Place, Promotion) and typically covers a one-year period. It translates strategic objectives into actionable steps, ensuring the organization stays on course to achieve its goals.
- **Importance of Strategic Over Tactical** - Good tactical marketing cannot compensate for poor strategic decisions. Effective strategic marketing is crucial for success, providing a clear roadmap that guides tactical actions and ensures long-term viability.
- **Market Segmentation Definition** - Market segmentation divides a heterogeneous market into smaller, homogeneous segments to better target specific consumer needs. This approach allows for more personalized marketing efforts, increasing relevance and engagement with target audiences.

- **Benefits of Market Segmentation** - Market segmentation leads to better consumer understanding, long-term competitive advantage, higher ROI, and effective resource allocation. However, it requires significant investment and, if poorly implemented, can waste resources.
- **Basics of Market Segmentation** - Market segmentation involves grouping consumers based on similar product preferences or characteristics. This facilitates understanding and targeting different consumer segments effectively.
- **Role of Data Analysts and Users** - Data analysts extract segments from consumer data, but marketers must understand these segments to develop effective marketing strategies. Collaboration between analysts and marketers ensures that segmentation insights are actionable.
- **Importance of Good Data** - Quality data is crucial for effective market segmentation. Poor data quality leads to inaccurate segmentation, undermining marketing efforts.
- **Exploring Data** - Before segment extraction, data must be explored to understand potential segmentation strategies. This exploration helps identify meaningful patterns and relationships within the data.
- **Profiling Segments** - After segmentation, each segment should be profiled and described to understand its characteristics better. Profiling enables marketers to tailor their strategies to specific segment needs.
- **Strategic and Tactical Implementation** - Market segmentation must lead to actionable marketing strategies and plans. Good data alone is insufficient; it must be translated into effective marketing actions.

Three Layers of Market Segmentation

- **Core Layer:** Extracting segments from data.
- **Second Layer:** Ensuring high-quality segmentation by collecting and exploring data, and profiling segments.
- **Third Layer:** Implementing segmentation by targeting selected segments and developing marketing strategies.

User Involvement - Users need to be involved in data collection to ensure relevant information is captured. They also play a key role in selecting target segments and developing marketing plans.

Approaches to Segmentation

- **Organizational Constraints:** Approaches vary depending on the organization's willingness and ability to change its current marketing strategy.
- **Segmentation Variables:** Using different consumer characteristics (e.g., age, gender) to create segments.

- **Three Segmentation Approaches by Dibb and Simkin:**
 - **Segment Revolution:** Starting from scratch with new segments.
 - **Segment Evolution:** Refining existing segments.
 - **Segment Mutation:** Discovering new segments through exploratory research.
- **Unidimensional vs. Multidimensional Segmentation:** Creating segments based on single or multiple pieces of customer information.

Commonsense vs. Data-Driven Segmentation

- **Commonsense Segmentation:** Using pre-existing knowledge to create segments.
- **Data-Driven Segmentation:** Using data analysis to discover segments. Combining both approaches can lead to more effective segmentation strategies.

Data Structure and Segmentation

Understanding the structure of data is crucial before segmentation. Market segments can be natural, reproducible, or constructed based on data characteristics. Conducting a data structure analysis helps avoid mistakes in the segmentation process.

Step 1: Deciding (not) to Segment

Market segmentation is a strategic approach where an organization divides its broad consumer or business market into sub-groups of consumers based on shared characteristics.

While widely adopted, evaluating whether to pursue this strategy is crucial as it involves significant long-term commitments and investments. Based on the "*Market Segmentation Analysis*" PDF, here is the detailed evaluation for deciding whether to proceed with segmentation for McDonald's:

1. Data Collection and Preparation Gather Relevant Data: Collect data on customer demographics, purchase behavior, preferences, and any other relevant variables.

Data Cleaning: Ensure the data is clean by handling missing values, removing duplicates, and correcting inconsistencies.

2. Exploratory Data Analysis (EDA) :

- **Descriptive Statistics:** Compute measures like mean, median, mode, and standard deviation for different variables.
- **Visualization:** Use visualizations (e.g., histograms, bar charts, box plots) to understand the distribution of data and identify patterns or anomalies.
- **Correlation Analysis:** Check for correlations between variables to understand relationships within the data.

3. Segmentation Analysis :

- **Clustering Techniques:** Apply clustering algorithms (e.g., K-means, hierarchical clustering) to see if distinct groups emerge within the data.
- **Evaluation of Clusters:** Assess the quality of the clusters using metrics like silhouette score, within-cluster sum of squares (WCSS), and between-cluster variation.
- **Segmentation Criteria:** Determine if the clusters have practical and meaningful differences in terms of demographics, behavior, or preferences.

4. Decision Making Assess Homogeneity: If the customer base shows significant homogeneity with little variation, segmentation may not be necessary.

- **Assess Business Goals:** Align the segmentation findings with business objectives. If distinct segments align with targeted marketing strategies or product offerings, segmentation can be beneficial.
- **Cost-Benefit Analysis:** Consider the costs and benefits of implementing segmentation.

If the benefits outweigh the costs, segmentation is justified.

5. Leadership Support: Without active support and involvement from senior management, segmentation initiatives may struggle to gain traction and resources, making it difficult to implement effectively.

6. Cultural Resistance: Organizational cultures resistant to change, poor communication practices, and internal politics can impede the acceptance and adoption of segmentation strategies, creating barriers to successful implementation.

7. Training Needs: Adequate training and education on market segmentation are essential for employees to understand its principles and methodologies. Without sufficient knowledge and awareness, segmentation efforts may falter or yield suboptimal results.

8. Marketing Expertise: Having skilled marketing professionals is crucial for conducting effective segmentation analysis and developing tailored marketing strategies for different market segments. Lacking expertise in marketing can hinder the identification of target segments and the formulation of targeted marketing campaigns .

9. Resource Constraints: Limited financial resources or organizational structures that aren't conducive to segmentation efforts can pose significant challenges. Overcoming these constraints may require innovative approaches or reallocation of existing resources to support segmentation activities .

10. Effective Planning: Thorough planning is essential for the success of segmentation initiatives. This involves setting clear objectives, establishing robust processes and methodologies, and ensuring effective project management to guide segmentation projects from inception to completion.

11. Simplifying Insights: Segmentation research often generates complex data and insights that may be difficult for non-specialists to understand. Using simple, visually appealing methods such as charts, graphs, and visualizations can help simplify complex insights and facilitate better understanding and decision-making across the organization.

Step 2: Specifying the Ideal Target Segment

2.1 Segment Evaluation Criteria

- The third layer of market segmentation analysis relies heavily on user input, involving the user in most stages of the process.
- After investigating the value of a segmentation strategy, the organization must make a significant contribution to market segmentation analysis in Step 2.
- In Step 2, the organization must determine two sets of segment evaluation criteria: knock-out criteria, which are essential features of segments considered, and attractiveness criteria, which evaluate the relative attractiveness of remaining market segments.
- The segmentation team must select which criteria to use to determine the attractiveness of potential target segments and assess the relative importance of each criterion to the organization.

2.2 Knock-Out Criteria

- Knock-out criteria are used to assess market segments based on market segmentation analysis.
- This criterion includes substantiality, measurability, accessibility, homogeneity and distinctness of the segment.
- The ideal target segment should be large enough, matching the organization's strengths, identifiable, and reachable.
- These criteria must be understood by senior management, the segmentation team, and the advisory committee.

2.3 Implementing a Structured Process

- The segmentation literature suggests that a structured process is beneficial for assessing market segments.
- The most popular approach is a segment evaluation plot, which shows segment attractiveness and organisational competitiveness are determined by segmentation team.
- Factors determining segment attractiveness and organisational competitiveness need to be negotiated and agreed upon.
- McDonald and Dunbar recommend using no more than six factors as the basis for calculating these criteria.

- This is important because each unit has different perspectives on the organization's business and if the segmentation strategy is implemented, it will affect every unit of the organization. At the end of this step, the market segmentation team should have a list of approximately six segment attractiveness criteria, each with a weight attached to indicate its importance to the organization.
- Structured Process Implementation: It's agreed that following a structured process for segment evaluation is beneficial. One common approach is using a segment evaluation plot, which compares segment attractiveness with organizational competitiveness. Criteria for this evaluation are agreed upon by the team, ideally with input from all parts of the organization.
- Importance of Early Criteria Selection: Even though the segment evaluation plot isn't finished yet, picking attractiveness criteria early is helpful. It ensures that relevant information is collected during data gathering and makes it easier to choose target segments later. The team should agree on about six criteria, giving each a weight to show how important it is. Approval from the advisory committee, representing different perspectives, is a good idea.

To define segment evaluation criteria, organizations should focus on size, growth potential, profitability, accessibility, and compatibility. Establishing knock-out criteria involves setting minimum thresholds for segment size and revenue potential to ensure segments are viable and worthwhile. Identifying attractiveness criteria includes targeting segments with high spending, frequent visits, strong loyalty, and low competition. Implementing a structured evaluation process involves developing a scoring system, assigning weights to each criterion, and calculating aggregate scores for segments to systematically prioritize and select the most attractive market segments.

Step 3: Collecting Data:

For this project, we need to gather data from relevant sources. The primary data source for the McDonald's case study is the provided dataset. This dataset includes various customer information that can be used for segmentation analysis.

Data-Set: [McDonalds Case Study - Google Drive](#)

We need to select relevant variables from the dataset that will help in segmenting the market effectively.

Key variables include:

- Demographic information like age and gender.
- Behavioural data such as purchase history and frequency of visits.
- Psychographic data including preferences and interests.

3.1 Segmentation Variables

Market segmentation can be based on commonsense or data-driven approaches, both relying on empirical data to identify or create segments. In commonsense segmentation, a single characteristic, such as gender, splits the sample, with descriptor variables like age and benefits sought detailing each segment. This helps marketers develop targeted marketing mixes. Data-driven segmentation uses multiple variables to identify natural or artificial segments, enhancing targeting accuracy. High-quality empirical data is crucial, sourced from surveys, observations, scanner data, or experiments, ideally reflecting actual consumer behavior. Reliable data supports the development of tailored products, pricing, distribution, and promotional strategies.

3.2 Segmentation Criteria

Before segment extraction and data collection, organizations must decide on the segmentation criteria, which refer to the nature of the information used for segmentation, such as benefits sought. Common criteria include geographic, socio-demographic, psychographic, and behavioral factors. Key considerations are profitability, bargaining power, benefit preferences, choice barriers, and consumer interactions. The best approach is to use what works for your project or service at the least possible cost.

3.2.1 Geographic Segmentation

- The only criterion is the consumer's location of residence
- Advantage is that each consumer can easily be assigned a geographic unit
- 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.

3.2.2 Socio-Demographic Segmentation

- Socio-demographic segmentation criteria include age, gender, income and education
- Advantage is segment membership can easily be determined for every consumer
- Disadvantage is that in many instances, the socio-demographic criterion is not the cause for product preferences, thus not providing sufficient market insight for optimal segmentation decisions.

3.2.3 Psychographic Segmentation

- Psychographic Segmentation includes consumers' beliefs, interests, preferences, aspirations or benefits sought when purchasing a product
- 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.
- Advantage is that it is generally more reflective of the underlying reasons for differences in consumer behaviour.

3.2.4 Behavioural Segmentation

- Behavioural Segmentation is knowing the similarities in behaviour or reported behaviour.
- Advantage of behaviour approach is that the behaviour of interest is used as the basis of segment extraction.
- Disadvantage is that behavioural data is not always readily available especially if the aim is to include segmentation analysis of potential customers who have not previously purchased the product, rather than limiting oneself to the study of the existing customers.

Step 4: Exploring Data

Exploring data generally means performing exploratory data analysis (EDA) to

```
In [14]: # Step 3: Exploratory Data Analysis (EDA)
```

```
print(data.describe())
```

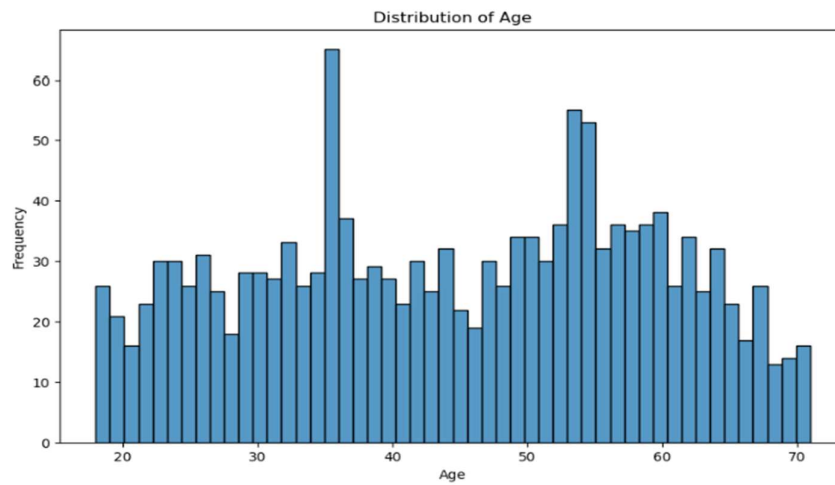
	Age
count	1431.000000
mean	44.656184
std	14.199400
min	18.000000
25%	33.000000
50%	45.000000
75%	57.000000
max	71.000000

```
In [71]: print(data.shape) # number of rows and columns
```

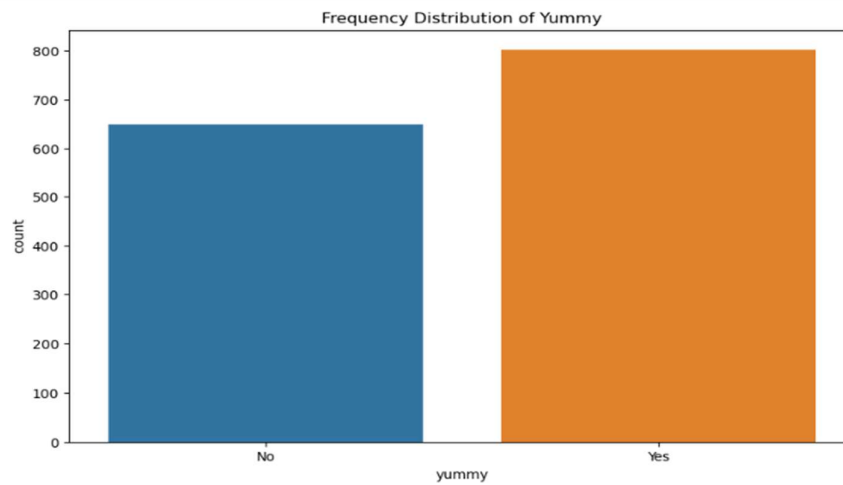
```
(1453, 15)
```

analyze and investigate data sets and summarize their main characteristics, often employing data visualization methods. EDA helps determine how best to manipulate data sources to get the answers you need, making it easier for data scientists to discover patterns, spot anomalies, test a hypothesis, or check assumptions. It can also help in understanding data structures, identifying patterns and relationships, detecting anomalies and outliers, testing assumptions, and handling missing values.

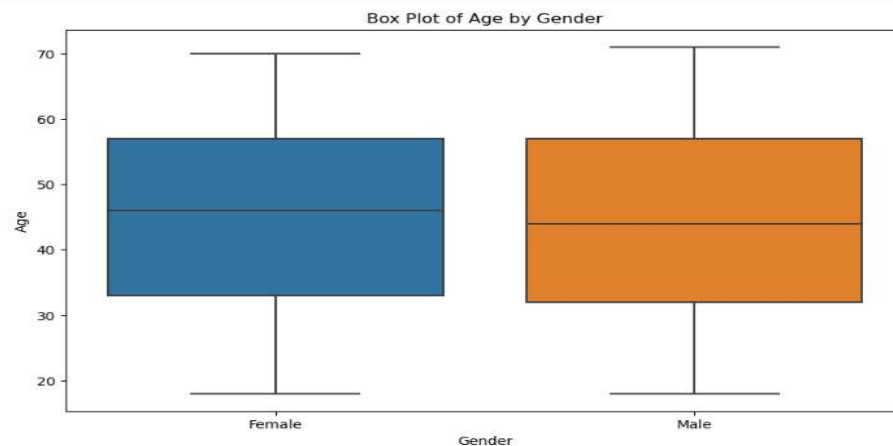
```
In [75]: # Histogram of Age
plt.figure(figsize=(10, 6))
sns.histplot(data['Age'], bins=50)
plt.title('Distribution of Age')
plt.xlabel('Age')
plt.ylabel('Frequency')
plt.show()
```



```
In [77]: # Bar chart of frequency distribution of categorical variables
plt.figure(figsize=(10, 6))
sns.countplot(x='yummy', data=data)
plt.title('Frequency Distribution of Yummy')
plt.show()
```



```
In [80]: # Box plot of Age by Gender
plt.figure(figsize=(10, 6))
sns.boxplot(x='Gender', y='Age', data=data)
plt.title('Box Plot of Age by Gender')
plt.xlabel('Gender')
plt.ylabel('Age')
plt.show()
```



```
In [6]: #step 02 cleaning
# Check for missing values
print(data.isnull().sum())
```

```
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
```

```
In [7]: # Fill or drop missing values as appropriate
data = data.fillna(method='ffill')
```

```
In [8]: # Check for duplicates
print(data.duplicated().sum())
```

```
22
```

```
In [9]: # Remove duplicates
data = data.drop_duplicates()
```

```
In [8]: # Check for duplicates
print(data.duplicated().sum())
```

```
22
```

```
In [9]: # Remove duplicates
data = data.drop_duplicates()
```

```
In [10]: # Display the cleaned data
print(data.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
```

```
In [11]: # Convert the first 11 columns to a matrix
D_x = data.iloc[:, 0:11].to_numpy()
```

```
In [12]: # Convert "Yes" values to 1 and "No" values to 0
D_x = (D_x == "Yes") + 0
```

```
In [13]: # Calculate the column means and round to 2 decimal places
col_means = np.round(np.mean(D_x, axis=0), 2)

print(col_means)
```

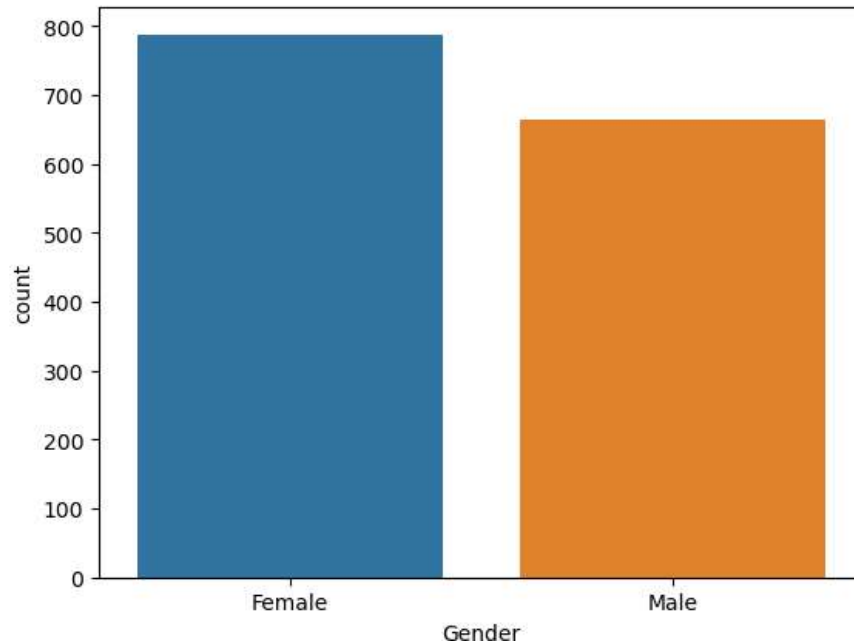
```
[0.55 0.91 0.1  0.87 0.53 0.9  0.6  0.64 0.36 0.2  0.24]
```

Data cleaning, also known as data preprocessing, is an essential step in the data science workflow. Here we perform data cleaning:

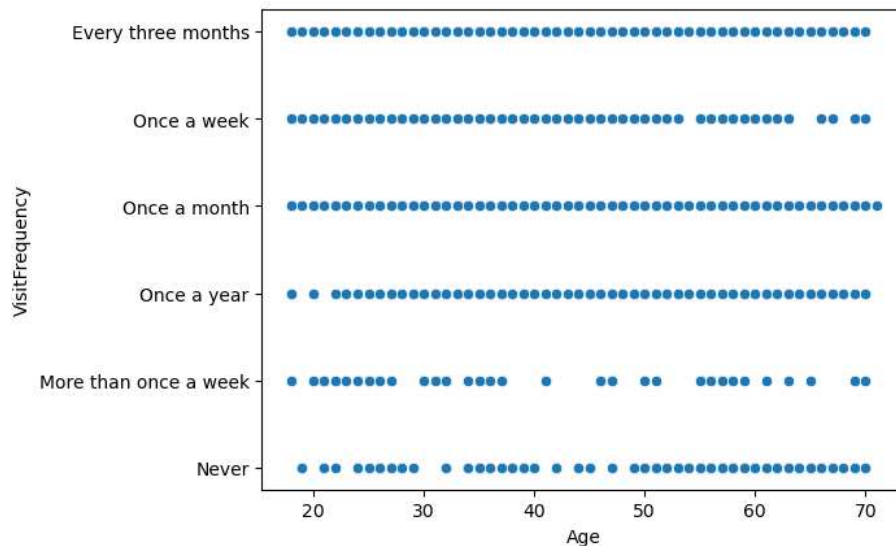
Descriptive Analysis: descriptive analysis is a fundamental tool for transforming raw data into clear and concise information, making it easier to draw meaningful conclusions.

```
In [85]: #Descriptive Analysis on data
# Frequency distribution
print(data['Gender'].value_counts())
sns.countplot(x='Gender', data=data)
plt.show()
```

```
Female    788
Male      665
Name: Gender, dtype: int64
```



```
In [87]: # Data visualization
sns.scatterplot(x='Age', y='VisitFrequency', data=data)
plt.show()
```



Pre-Processing : Two pre-processing procedures are often used for categorical variables. One is merging levels of categorical variables before further analysis, the other one is converting categorical variables to numeric ones, if it makes sense to do so.

```
In [95]: # Pre-Processing
# sort the frequency table of 'VisitFrequency' column
# Get the categorical columns
categorical_cols = mcdonalds.select_dtypes(include=['object']).columns

# Apply one-hot encoding to each categorical column
mcdonalds_ohc = pd.get_dummies(mcdonalds, columns=categorical_cols)
```

```
In [94]: print(mcdonalds_ohc.head())
```

```
   Age  yummy_No  yummy_Yes  convenient_No  convenient_Yes  spicy_No \
0    61         1         0             0             1         1
1    51         0         1             0             1         1
2    62         1         0             0             1         0
3    69         0         1             0             1         1
4    49         1         0             0             1         1

   spicy_Yes  fattening_No  fattening_Yes  greasy_No  ...  Like_I hate it!-5 \
0          0             0             1          1  ...             0
1          0             0             1          0  ...             0
2          1             0             1          0  ...             0
3          0             0             1          0  ...             0
4          0             0             1          0  ...             0

   Like_I love it!+5  VisitFrequency_Every three months \
0                   0                                 1
1                   0                                 1
2                   0                                 1
3                   0                                 0
4                   0                                 0

   VisitFrequency_More than once a week  VisitFrequency_Never \
0                                       0                       0
1                                       0                       0
2                                       0                       0
3                                       0                       0
4                                       0                       0

   VisitFrequency_Once a month  VisitFrequency_Once a week \
0                               0                           0
1                               0                           0
2                               0                           0
3                               0                           1
4                               1                           0

   VisitFrequency_Once a year  Gender_Female  Gender_Male
0                             0              1             0
1                             0              1             0
2                             0              1             0

1          0          1          0
2          0          1          0
3          0          1          0
4          0          0          1
```

[5 rows x 42 columns]

```
In [97]: # Select the categorical columns (yummy to disgusting)
categorical_cols = mcdonalds[['yummy', 'convenient', 'fattening', 'greasy', 'fast', 'cheap', 'tasty', 'expensive', 'healthy', 'disgusting']]

# Convert the categorical columns to binary indicators (0 or 1)
vacmot = (categorical_cols == 'yes').astype(int)

print(vacmot.head())
```

```
   yummy  convenient  fattening  greasy  fast  cheap  tasty  expensive \
0      0           0           0      0      0      0      0           0
1      0           0           0      0      0      0      0           0
2      0           0           0      0      0      0      0           0
3      0           0           0      0      0      0      0           0
4      0           0           0      0      0      0      0           0

   healthy  disgusting
0         0           0
1         0           0
2         0           0
3         0           0
4         0           0
```

```

In [100]: #pre-processing on numeric value
data['Age'].fillna(data['Age'].mean(), inplace=True)

In [101]: # Data Normalization
from sklearn.preprocessing import MinMaxScaler
scaler = MinMaxScaler()
data['Age'] = scaler.fit_transform(data[['Age']])

In [102]: #Outlier Detection and Handling
from scipy import stats

z_scores = np.abs(stats.zscore(data['Age']))
data = data[(z_scores < 3)]

In [104]: #Binning
from sklearn.preprocessing import KBinsDiscretizer

discretizer = KBinsDiscretizer(n_bins=5, encode='ordinal')
data['Age'] = discretizer.fit_transform(data[['Age']])

```

PCA (Principal component analysis):

```

In [22]: # Fit PCA to the data
D_pca = PCA().fit(D_x)

# Print the summary of the PCA results
print("PCA Summary:")
print("-----")
print("Proportion of variance explained by each PC:")
print(D_pca.explained_variance_ratio_)
print("\nCummulative proportion of variance explained:")
print(np.cumsum(D_pca.explained_variance_ratio_))

```

PCA Summary:

Proportion of variance explained by each PC:
[0.29899056 0.19156392 0.13267983 0.08290307 0.05969759 0.05069322
 0.04429957 0.03985029 0.03715547 0.03260161 0.02956487]

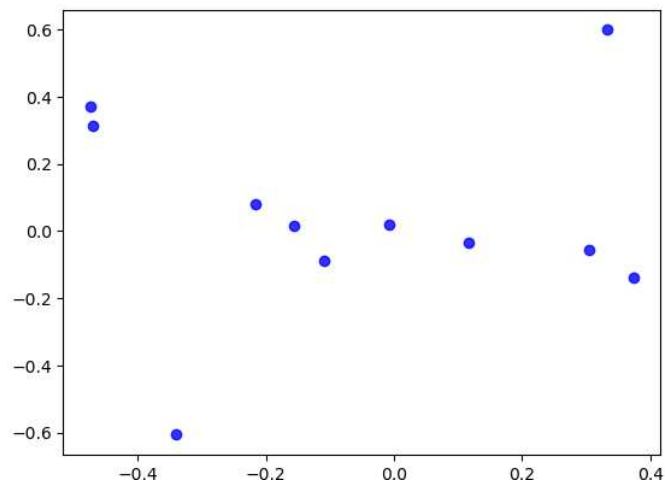
Cummulative proportion of variance explained:
[0.29899056 0.49055448 0.62323431 0.70613738 0.76583497 0.81652819
 0.86082776 0.90067805 0.93783352 0.97043513 1.]

```

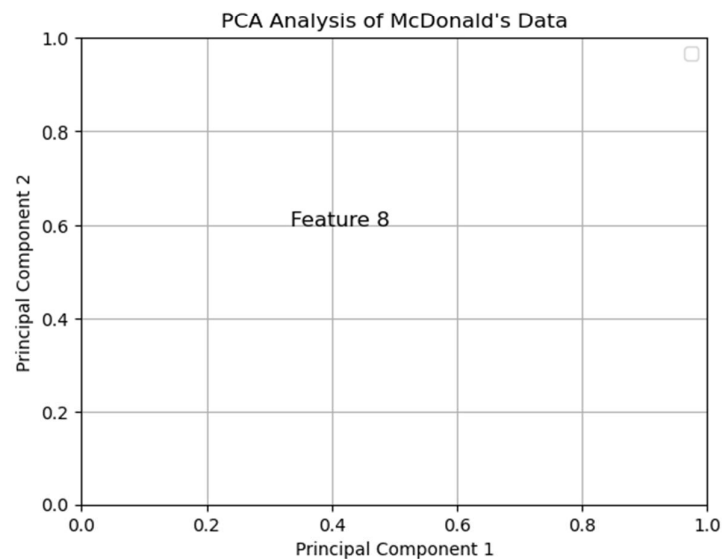
In [23]: # Scatter plot of the first two principal components
plt.scatter(D_pca.components_[0], D_pca.components_[1], alpha=0.8, c='blue', marker='o', label='PC1 vs PC2')

```

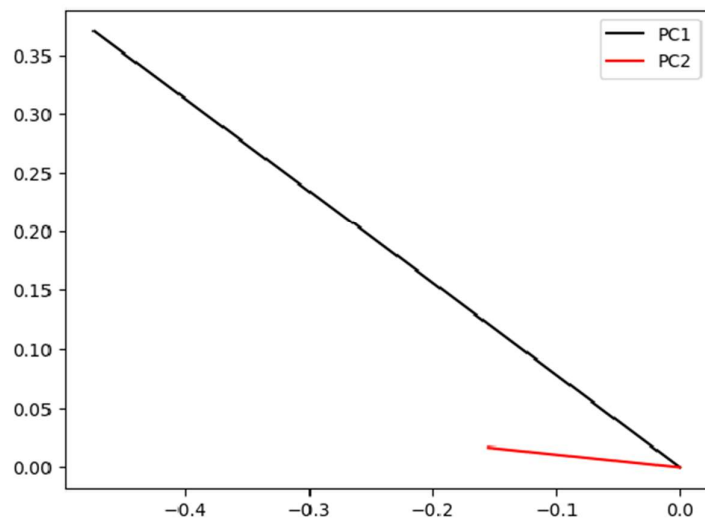
Out[23]: <matplotlib.collections.PathCollection at 0x20308d0e510>




```
In [24]: # Annotation for each feature
for i, feature in enumerate(range(11)): # Replace range(11) with your actual feature names or indices
    plt.annotate(f'Feature {feature}', (D_pca.components_[0, i], D_pca.components_[1, i]), fontsize=12)
# Labels and title
plt.xlabel('Principal Component 1')
plt.ylabel('Principal Component 2')
plt.title('PCA Analysis of McDonald's Data')
# Add a legend with a single entry (dummy entry)
plt.legend(['PC1 vs PC2'], loc='upper right')
# Show plot
plt.grid()
plt.show()
```



```
In [25]: # plot the projection axes of the first two principal components
plt.plot([0, D_pca.components_[0, 0]], [0, D_pca.components_[1, 0]], 'k-', label='PC1')
plt.plot([0, D_pca.components_[0, 1]], [0, D_pca.components_[1, 1]], 'r-', label='PC2')
plt.legend()
plt.show()
```



Step 5: Extracting Segments

Feature Selection –

- K-Means Clustering : K-Means is an unsupervised learning algorithm that groups data into clusters based on shared characteristics. This can be a useful technique for market segmentation, as it can help identify distinct customer groups.
- The Elbow Method : The Elbow Method is a technique for determining the optimal number of clusters in a K-Means analysis. It involves calculating the sum of squares for each cluster size and identifying the "elbow" point, where the marginal gain of adding another cluster starts to diminish.

Model Building –

- Market Segmentation: Market segmentation is the process of grouping a target market into subgroups based on shared characteristics. This can be done using a variety of methods, including geographic, demographic, psychographic, and behavioral segmentation.

Types of Market Segmentation

- Geographic Segmentation: Dividing the market based on location, such as country, region, or city.
- Demographic Segmentation: Dividing the market based on characteristics like age, gender, income, or education level.
- Psychographic Segmentation: Dividing the market based on lifestyle, personality, or values.
- Behavioral Segmentation: Dividing the market based on purchase behavior, usage patterns, or loyalty.

Hyperparameter Tuning - Importance of Market Segmentation Market segmentation is important for enhancing profits, targeting specific customers, and informing product development. By understanding the unique needs and preferences of different customer groups, businesses can tailor their offerings

and marketing strategies to better meet their customers' needs.

Step 6: Profiling Segments

6.1 Identifying Key Characteristics of Market Segments

- The profiling step is crucial in data-driven market segmentation to understand the resulting market segments.
- It is only necessary when the profiles of the segments are predefined, unlike commonsense segmentation where the profiles are predefined.
- Profiling aims to identify the defining characteristics of market segments with respect to the segmentation variables.
- It involves characterizing market segments individually and comparing them to other market segments.
- Good profiling is the basis for correct interpretation of the resulting segments, which is critical for making strategic marketing decisions.
- However, data-driven market segmentation solutions can be difficult to interpret, with many managers having difficulties understanding the results.
- Traditional and graphical statistics approaches to segment profiling can help make profiling less tedious and less prone to misinterpretation.

Table 8.1 Six segments computed with the neural gas algorithm for the Australian travel motives data set. All numbers are percentages of people in the segment or in the total sample agreeing to the motives

	Seg. 1	Seg. 2	Seg. 3	Seg. 4	Seg. 5	Seg. 6	Total
Rest and relax	83	96	89	82	98	96	90
Change of surroundings	27	82	73	82	87	77	67
Fun and entertainment	7	71	81	60	95	37	53
Free-and-easy-going	12	65	58	45	87	75	52
Not exceed planned budget	23	100	2	49	84	73	51
Life style of the local people	9	29	30	90	75	80	46
Good company	14	59	40	58	77	55	46
Excitement, a challenge	9	17	39	57	76	36	33
Maintain unspoilt surroundings	9	10	16	7	67	95	30
Cultural offers	4	2	5	96	62	38	28
Luxury / be spoilt	19	24	39	13	89	6	28
Unspoilt nature/natural landscape	10	10	13	15	69	64	26
Intense experience of nature	6	8	9	21	50	58	22
Cosiness/familiar atmosphere	11	24	12	7	49	25	19
Entertainment facilities	5	25	30	14	53	6	19
Not care about prices	8	7	43	19	29	10	18
Everything organised	7	21	15	12	46	9	16
Do sports	8	12	13	10	46	7	14
Health and beauty	5	8	10	8	49	16	12
Realise creativity	2	2	3	8	29	14	8

Traditional Approaches to Profiling Market Segments

- Traditional approaches to profiling market segments often involve high-level summaries or large tables providing exact percentages for each segmentation variable.
- Table 8.1 illustrates this issue, showing the mean values of segmentation variables by segment and the overall mean values.
- To identify the defining characteristics of market segments, the percentage value of each segment for each segmentation variable needs to be compared with the values of other segments or the total value provided in the far-right column.
- Profiling all six market segments based on Table 8.1 requires comparing 120 numbers if each segment's value is only compared to the total, or 300 pairs of numbers if each segment's value is compared to the values of other segments which is enormously tedious task, even for the most astute user.
- Sometimes, information is provided about the statistical significance of the difference between segments for each of the segmentation variables, but this approach is not statistically correct as segment membership is directly derived from the segmentation variables and segments are created in a way that makes them maximally different, not allowing for standard statistical tests to assess the significance of differences.

6.2 Segment Profiling with Visualisations

- Graphics are essential in statistical data analysis, particularly in exploratory statistical analysis like cluster analysis, as they provide insights into complex relationships between variables.
- Visualisation offers a simple way to monitor developments over time and makes the results of market segmentation analysis easier to interpret.
- Visualisation techniques are recommended for cluster analysis and mixture models, and examples of prior use of segmentation solutions are provided in various studies.
- Statistical graphs facilitate the interpretation of segment profiles and make it easier to assess the usefulness of a market segmentation solution, as selecting one of the possible solutions is a critical decision.

6.2.1 Identifying Defining Characteristics of Market Segments

- A segment profile plot is a useful tool for understanding the defining characteristics of market segments.
- It visually represents the differences between each market segment and the overall sample.
- Variables can be displayed in the order of appearance in the data set, or they can be rearranged to improve visualization.
- Different thresholds for defining a marker variable need to be specified if the segmentation variables are not binary.
- An eye tracking study by Nazila Babakhani investigated differences in people's ability to interpret complex data analysis results from market segmentation studies presented in tabular versus graphical statistics format.
- Good visualizations facilitate interpretation by managers who make long-term strategic decisions based on segmentation results, offering an excellent return on investment.

6.2.2 Assessing Segment Separation

- Segment separation plots visually represent the overlap of segments in a data space.
- They are simple for low-dimensional data sets but become complex as the number of segmentation variables increases.
- They provide a quick overview of the data situation and the segmentation solution.
- For 20-dimensional travel motives data, the 20-dimensional space needs to be projected onto a small number of dimensions to create a segment separation plot.
- Various projection techniques can be used, including principal components analysis and perceptual maps.

Step 7: Describing Segments

- ***Segment Profiling:*** Segment profiling is a critical step in market segmentation analysis. It involves examining the differences in segmentation variables across various market segments. These

segmentation variables are selected early in the analysis process and are fundamental in identifying and extracting distinct market segments from empirical data. Understanding these differences helps in delineating the unique characteristics and preferences of each segment.

- ***Describing Segments*** : This step of market segmentation analysis focuses on describing the identified market segments in detail. Unlike segment profiling, which primarily focuses on segmentation variables, segment description involves incorporating additional information about segment members, such as demographics, behavior patterns, and preferences. This step provides deeper insights into the nature of each segment, which is essential for developing tailored marketing strategies.
- ***Visualizing Differences***: Visualizations play a crucial role in segment description by simplifying the interpretation of differences between market segments. By visually representing data, such as through charts and graphs, visualizations make it easier for analysts and decision-makers to identify patterns and trends across segments. They also help integrate information on statistical significance, ensuring that meaningful insights are derived from the data.
- ***Using Visualizations***: Various visualization techniques are available for describing market segments. Nominal and ordinal descriptor variables can be effectively visualized using techniques like mosaic plots, which provide a clear visual representation of differences in categorical data. On the other hand, metric descriptor variables can be depicted using conditional plots and parallel box-and-whisker plots, allowing for the visualization of distributional differences between segments.
- ***Nominal and Ordinal Descriptor Variables***: Cross-tabulation and mosaic plots are powerful tools for visualizing differences in nominal or ordinal descriptor variables. These techniques allow analysts to observe disparities in observed frequencies across segments and identify any significant variations in segment characteristics.
- ***Metric Descriptor Variables***: Conditional plots and parallel box-

and-whisker plots are particularly useful for visualizing differences using metric descriptor variables. These techniques enable analysts to gain insights into the distributional differences between segments, providing a more detailed understanding of segment characteristics.

- ***Insights from Visualizations:*** Visualizations offer valuable insights into the characteristics and preferences of different market segments. By examining patterns in data, such as gender distribution, income levels, and behavioral tendencies, analysts can develop a deeper understanding of each segment's unique needs and preferences. These insights are crucial for designing targeted marketing strategies that resonate with specific segments.
- ***Segment Level Stability Across Solutions (SLSA) Plot:*** The SLSA plot is a powerful tool for identifying stable segments across multiple segmentation solutions. By tracing the value of a metric descriptor variable across various solutions, analysts can identify segments that consistently exhibit certain characteristics. This helps in identifying the most promising market segments for targeted marketing efforts, ensuring that resources are allocated effectively.
- ***Introduction to Testing for Segment Differences:*** The text begins by introducing the concept of testing for differences in descriptor variables across market segments. It emphasizes the importance of segment membership as a nominal variable and discusses the relevance of statistical tests for association between segment membership and other variables.
- ***Chi-Square Test for Nominal Variables:*** The chi-square test is introduced as a method for testing the association between nominal variables, such as segment membership and gender. An example R command for conducting a chi-square test is provided, along with an explanation of the output and interpretation of the p-value.
- ***Visualizing Associations:*** The text mentions using mosaic plots to visualize associations between segment membership and other nominal or ordinal variables.
- ***ANOVA for Metric Variables:*** Analysis of Variance (ANOVA)

is introduced as a method for testing differences in means across multiple market segments for metric variables. An example R command for conducting ANOVA is provided, along with interpretation of the output.

- ***Pairwise Comparisons:*** The text discusses the use of pairwise t-tests to compare means between individual segments after ANOVA, along with adjustments for multiple testing using methods like the Holm procedure.
- ***Tukey's Honest Significant Differences:*** Tukey's HSD test is introduced as an alternative to pairwise t-tests for identifying significant differences between segment means. The text explains how to interpret the results of Tukey's HSD test.
- ***Regression Analysis for Predicting Segment Membership:*** The text transitions to discussing regression analysis as a method for predicting segment membership based on descriptor variables. It covers linear regression for continuous variables and binary logistic regression for categorical variables.
- ***Interpretation of Regression Coefficients:*** The text explains how to interpret regression coefficients in the context of binary logistic regression, including the log odds ratio and odds ratio interpretation.
- ***Model Selection:*** Model selection methods, such as stepwise regression, are introduced to refine regression models by selecting relevant independent variables based on criteria like the AIC.
- ***Introduction to Multinomial Logistic Regression:*** Unlike binary logistic regression, which predicts two categories, multinomial logistic regression predicts multiple categories of a dependent variable. In this case, the dependent variable represents different market segments, which are assumed to follow a multinomial distribution.
- ***Using the multinom() Function:*** In R, the multinom() function from the nnet package is used to fit a multinomial logistic regression model. The model is specified using a formula and a data frame containing the variables.
- ***Interpreting Coefficients:*** The coefficients of the fitted model represent the change in log odds of belonging to each segment

relative to a baseline category (usually the first segment). The output provides a matrix of regression coefficients for each category of the dependent variable.

- ***Assessing Model Fit:*** The residual deviance and AIC (Akaike Information Criterion) are used to assess the fit of the model. Lower values indicate better fit.
- ***Model Selection:*** The Anova() function is used to assess if dropping any variable significantly reduces the model fit. This helps in model selection by identifying variables that contribute significantly to predicting segment membership.
- ***Visualizing Predictive Performance:*** Predicted segment memberships can be compared to observed segment memberships using mosaic plots and boxplots of predicted probabilities for each segment.
- ***Effect Visualization:*** The effects of independent variables on predicted segment membership probabilities can be visualized using plots. This helps in interpreting the estimated effects of each variable on segment membership.
- ***Tree-Based Methods as an Alternative:*** The passage also discusses the use of classification and regression trees (CARTs) as an alternative method for predicting segment membership. CARTs offer advantages such as variable selection, ease of interpretation, and the ability to handle large numbers of independent variables.

Step 8: Selecting the Target Segment

Selecting target segments involves evaluating and choosing the most promising segments to focus marketing efforts on. This process typically involves assessing the attractiveness and fit of each segment with the company's strategic goals.

Steps to Perform

- Evaluate Segment Attractiveness
- Create a Scoring System
- Prioritize Segments
- Select Target Segments

1. Evaluate Segment Attractiveness:

```
from scipy.cluster.hierarchy import dendrogram, linkage
from statsmodels.graphics.mosaicplot import mosaic

In [ ]: from rpy2.robjects.packages import importr
        from rpy2.robjects import Formula, pandas2ri
        import rpy2.robjects as ro
        import plotly.express as px
        import warnings
        warnings.filterwarnings('ignore')

In [3]: # Load the McDonald's dataset from a CSV file
mcdonalds = pd.read_csv('https://homepage.boku.ac.at/leisch/MSA/datasets/mcdonalds.csv')
print(mcdonalds.columns)

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

In [4]: # Print the dimensions of the dataset
print(mcdonalds.shape)

(1453, 15)

In [5]: # Display the first 3 rows of the dataset
print(mcdonalds.head(3))
```

	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	

	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

2. Create a Scoring System

We created a scoring system based on the columns available. For simplicity, we focussed on VisitFrequency, Like, and Age.

Steps to Encode and Normalize Data:

- Extract Numerical Values from Like Column: Remove non-numeric characters.
- Encode Categorical Variables: Convert VisitFrequency and Gender into numerical values.
- Normalize Numerical Variables: Normalize the Like and Age columns to bring all features to a comparable scale.

```
In [73]: mcdonalds['Like'] = LabelEncoder().fit_transform(mcdonalds['Like'])
Like = mcdonalds.groupby('cluster_num')['Like'].mean()
Like = Like.to_frame().reset_index()
Like
```

```
Out[73]:
```

	cluster_num	Like
0	0	5.109890
1	1	6.741667
2	2	3.868668
3	3	3.787975

```
In [74]: #Gender
from sklearn.preprocessing import LabelEncoder
mcdonalds['Gender'] = LabelEncoder().fit_transform(mcdonalds['Gender'])
Gender = mcdonalds.groupby('cluster_num')['Gender'].mean()
Gender = Gender.to_frame().reset_index()
Gender
```

```
Out[74]:
```

	cluster_num	Gender
0	0	0.414835
1	1	0.579167
2	2	0.392120
3	3	0.525316

```
In [75]: #Calculating the mean
#Visit frequency
mcdonalds['VisitFrequency'] = LabelEncoder().fit_transform(mcdonalds['VisitFrequency'])
visit = mcdonalds.groupby('cluster_num')['VisitFrequency'].mean()
visit = visit.to_frame().reset_index()
visit
```

```
Out[75]:
```

	cluster_num	VisitFrequency
0	0	2.730769
1	1	2.691667
2	2	2.602251

3. Prioritize Segments

I have normalized the scores and create a composite score to rank the segments. Steps to Calculate Composite Score:

- Assign Weights to Each Criterion: Define the importance of each criterion.
- Calculate Composite Score: Use the weighted sum of the criteria.

4. Select Target Segments

```
In [75]: weights = {
        'VisitFrequency': 0.4,
        'Like': 0.4,
        'Age': 0.2
    }

    # Calculate composite score
    mcdonalds['Score'] = (
        mcdonalds['VisitFrequency'] * weights['VisitFrequency'] +
        mcdonalds['Like'] * weights['Like'] +
        mcdonalds['Age'] * weights['Age']
    )

    # Sort by score
    mcdonalds = mcdonalds.sort_values(by='Score', ascending=False)

    # Display the top segments
    print(mcdonalds.head(10))
```

	yummy	convenient	spicy	fattening	greasy	fast	cheap	tasty	expensive	\
1409	Yes	Yes	No	Yes	No	Yes	No	Yes	Yes	
1104	No	No	No	Yes	Yes	Yes	No	No	Yes	
1124	No	Yes	No	Yes	Yes	Yes	Yes	No	No	
643	Yes	Yes	No	Yes	No	Yes	Yes	Yes	No	
226	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	
1316	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	
881	No	Yes	No	Yes	Yes	Yes	Yes	No	No	
624	Yes	Yes	No	No	No	Yes	Yes	Yes	No	
1197	No	Yes	No	Yes	Yes	Yes	No	No	Yes	
522	Yes	Yes	No	No	No	Yes	Yes	Yes	No	

	healthy	disgusting	Like	Age	VisitFrequency	Gender	cluster_num	Score
1409	No	No	10	70	4	1	3	19.6
1104	No	Yes	9	70	5	0	1	19.6
1124	No	No	8	70	5	0	0	19.2
643	No	No	8	70	5	0	2	19.2
226	No	No	9	67	5	0	3	19.0
1316	No	Yes	9	67	5	1	1	19.0
881	No	Yes	9	67	5	1	0	19.0
624	Yes	No	10	69	3	0	2	19.0
1197	No	Yes	9	66	5	0	1	18.8
522	Yes	No	10	68	3	1	2	18.8

Summary

- Evaluate Segment Attractiveness: Use VisitFrequency, Like, and Age as criteria.
- Create a Scoring System: Assign weights to each criterion and calculate a compositescore.
- Normalize and Calculate Scores: Ensure all features are on a comparable scale.
- Select Top Segments: Choose segments with the highest scores for targeted marketing efforts.
- This process ensures you systematically evaluate and select the most promising target segments for McDonald's using the dataset provided.

Step 9: Customising the Marketing Mix

The concept of the marketing mix originated as a set of tools designed to enhance sales by strategically combining various marketing elements. Initially proposed by Borden (1964) as twelve elements, the modern understanding has consolidated this into the 4Ps: Product, Price, Promotion, and Place (McCarthy, 1960).

Segmentation, Targeting, and Positioning (STP) Approach

Market segmentation is integral to strategic marketing, working in conjunction with positioning and competition. The STP approach involves:

1. **Segmentation:** Identifying and describing distinct market segments.
2. **Targeting:** Assessing and selecting one or more segments to target.
3. **Positioning:** Differentiating the product in the minds of the target segments.

This approach ensures that segmentation informs other strategic decisions, and the process is often iterative rather than strictly sequential.

Impact on the Marketing Mix

Once target segments are selected, the marketing mix must be tailored to these segments. The 4Ps are adjusted as follows:

- **Product:** May involve developing new products or modifying existing ones to meet the specific needs of the target segment.
- **Price:** Setting appropriate pricing strategies, including discounts or premium pricing based on the segment's price sensitivity and willingness to pay.
- **Place:** Deciding on the distribution channels that best reach the target segment.
- **Promotion:** Crafting messages and choosing media channels that resonate with the target segment.

Customising the 4Ps

1. **Product:**

- Modification of products based on segment needs rather than entirely new product development.
- Includes decisions on product naming, packaging, warranties, and after-sales services.

2. Price:

- Setting prices and discounts according to the target segment's spending patterns and price sensitivity.
- Example: A tourist destination targeting a segment interested in cultural heritage might find that this segment has higher daily expenditures, allowing for premium pricing of tailored packages.

3. Place:

- Determining whether to sell products online, offline, or through specific channels like wholesalers or retailers.
- Example: For segment 3 (culturally inclined tourists), offering online booking options aligns with their higher tendency to book accommodations online.

4. Promotion:

- Developing advertising messages that appeal to the target segment and selecting the most effective communication channels.
- Example: For segment 3, promoting through tourist information centers and preferred TV channels based on their information-seeking behavior.

Practical Applications and Examples

Using the Australian vacation activities data set as a case study:

- **Product Example:** A cultural heritage destination could develop a "MUSEUMS, MONUMENTS & MUCH, MUCH MORE" package, catering specifically to segment 3's interests.

- **Price Example:** Segment 3, with higher daily expenditures, suggests potential for premium pricing without needing discounts.
- **Place Example:** Segment 3's preference for online booking implies the necessity of robust online booking platforms.
- **Promotion Example:** Effective promotion for segment 3 includes leveraging tourist information centers and specific TV channels favored by this segment.

In summary, customising the marketing mix to align with targeted segments enhances the effectiveness of marketing strategies, ensuring products and messages resonate more deeply with the intended audience. This process is dynamic and may require revisiting earlier steps to refine the approach based on ongoing insights and market feedback.

Step 10: Evaluation and Monitoring (Optional)

10.1 Ongoing Tasks in Market Segmentation

Market segmentation analysis is a continuous strategic process rather than a one-time effort. After implementing a segmentation strategy, two ongoing tasks are necessary:

1. **Evaluate Effectiveness:** Regularly assess whether the segmentation strategy increases profit or meets organizational goals. If not, the strategy is deemed unsuccessful.
2. **Monitor Market Changes:** Continuously track market dynamics, including changes in consumer behavior and competitive actions, to adjust the strategy as needed. This can range from periodic reviews to automated data mining systems alerting to changes.

10.2 Evaluating the Success of the Segmentation Strategy

The primary goal is to determine if a customized marketing mix for segments achieved the expected benefits, such as increased profit or other performance metrics in non-profits. Continuous monitoring is essential, and longer-term effectiveness can be gauged through tracking studies to understand market perception and competitive positioning.

10.3 Stability of Segment Membership and Segment Hopping

Research shows that segment membership can change significantly over time, necessitating the re-evaluation of segmentation strategies. For instance, studies have found that only a small percentage of consumers remain in the same segment over a few years. This instability, or segment hopping, is influenced by factors such as situational usage of products and changing consumer preferences. Understanding and addressing segment hopping is crucial for effective market segmentation and tailored marketing actions.

10.4 Segment Evolution

Market segments evolve due to changes in consumer behavior, new product introductions, and disruptive innovations. Continuous monitoring and tracking systems are recommended to detect and respond to these changes promptly. Segment stability analysis is crucial to distinguish between genuine segment evolution and inherent instability in segmentation solutions. Automated procedures, such as the MONIC framework and similar approaches, can help in tracking segment evolution over time, ensuring that marketing strategies remain relevant and competitive.

Example: Winter Vacation Activities

An analysis of winter vacation activities of tourists to Austria between 1991/92 and 1997/98 illustrates the need for ongoing monitoring. Significant changes in activities and segment sizes were observed, highlighting the importance of adapting marketing strategies to evolving tourist preferences. This example underscores the necessity of regular segment evaluation and adjustment to maintain a competitive edge.

Github Repositories link :

- Mr. Pranav Pimpalkar - https://github.com/Pranav-Pimpalkar/Market-Segmentation-on-Mc-Donalds-dataset_Feynn-Labs.git
- Mr. Amey Sachin Darokar - https://github.com/Amey-sd/Feynn_Labs/tree/main/T_1
- Mr. Harshal Ingale - https://github.com/Harshal0886/Market_Segmentation_Analysis-FeynnLab
- Ms. Prachi Kale - https://github.com/prachikale2004/Market_Segmentation_Analysis-FeynnLab.git
- Ms. Priya - <https://github.com/priya99karn/Market-segment-Analysis.git>
- Mr. Shashikumar Kulkarni - <https://github.com/Shashi0170/Market-Segmentation/tree/main>