****

**VIRGINIA COMMONWEALTH UNIVERSITY**

**Statistical analysis and modelling (SCMA 632)**

**A4C-Apply Multidimensional Scaling and interpret the results**

**N V SATYANARAYAN**

**V01108247**

**Date of Submission: 08-07-2024**

**CONTENTS**

|  |  |  |
| --- | --- | --- |
| **Sl. No.** | **Title** | **Page No.** |
| **1.** | Introduction | 1 |
| **2.** | Objectives | **1** |
| **3.** | Business Significance | **1-2** |
| **4.** | Results and analysis | **2-9** |
| **5.** | Conclusion | **10** |
| **6.** | References |  |

**\*NOTE- PYTHON AND R CODES WTH RESULT ADDED IN GITHUB-** [Satyanaldiga (github.com)](https://github.com/Satyanaldiga)

**Introduction to Multidimensional Scaling (MDS) in Ice Cream Market Analysis**

Multidimensional Scaling (MDS) is a statistical technique used for analyzing similarity or dissimilarity data. It helps in visualizing the level of similarity of individual cases of a dataset. In this analysis, MDS is applied to a dataset of various ice cream brands with different attributes such as Price, Availability, Taste, Flavour, Consistency, and Shelf life.

**Objective**

The primary objective of this analysis is to:

1. **Identify Similarities and Differences**: Understand how different ice cream brands compare to one another based on the selected attributes.
2. **Visualize Relationships**: Create a visual representation of the relationships between the ice cream brands to identify clusters of similar brands and outliers.
3. **Inform Business Decisions**: Use the insights gained from the analysis to inform strategic business decisions such as product development, marketing strategies, and competitive positioning.

**Business Significance**

The business significance of performing MDS in this context includes:

1. **Competitive Analysis**:
   * **Benchmarking**: Companies can benchmark their products against competitors, identifying where they stand in terms of key attributes like taste and price.
   * **Positioning**: Understanding the market positioning of different brands helps in devising strategies to differentiate products.
2. **Market Segmentation**:
   * **Targeting**: By identifying clusters of similar brands, businesses can target specific segments of the market more effectively.
   * **Customization**: Tailoring products to meet the specific needs and preferences of different market segments.
3. **Product Development**:
   * **Identifying Gaps**: Discover gaps in the market where new products could be introduced.
   * **Enhancing Features**: Improve or modify existing products based on the strengths and weaknesses identified through MDS.
4. **Strategic Planning**:
   * **Resource Allocation**: Allocate resources more efficiently by focusing on attributes that significantly impact consumer preferences.
   * **Market Entry**: Make informed decisions about entering new markets or launching new products based on similarity to successful brands.

The dataset has 10 rows and 7 columns. The columns are:

* Brand
* Price
* Availability
* Taste
* Flavour
* Consistency
* Shelflife

**R CODES AND INTERPRETATION**

|  |
| --- |
| icecream\_df<-read.csv('C:\\Users\\SPURGE\\Desktop\\SCMA\\A4\\Icecream.csv')  > dim(icecream\_df)  [1] 10 7 |
|  |
| |  | | --- | | > | |

* read.csv reads the CSV file into a data frame called icecream\_df.
* dim(icecream\_df) returns the dimensions of the data frame, i.e., the number of rows and columns. This helps verify that the data has been read correctly.

names(icecream\_df)

[1] "Brand" "Price" "Availability" "Taste" "Flavour"

[6] "Consistency" "Shelflife"

names(icecream\_df) lists the column names in the data frame. This is useful to understand what features are available in the dataset.

Subsetting the Data to Exclude the Brand Column:

|  |
| --- |
| ice<-subset(icecream\_df,select = -c(Brand)) |
|  |
| |  | | --- | | > | |

subset creates a new data frame ice by excluding the Brand column from icecream\_df. This is because the Brand column is a categorical variable and is not needed for calculating the distance matrix.

Calculating the Distance Matrix:

|  |
| --- |
| distance\_matrix<-dist(ice) |
|  |
| |  | | --- | | > | |

dist computes the distance matrix for the data frame ice. This matrix represents the pairwise distances between the rows (ice cream brands) based on their attributes (Price, Availability, Taste, Flavour, Consistency, Shelf life).

Applying Classical Multidimensional Scaling (MDS):

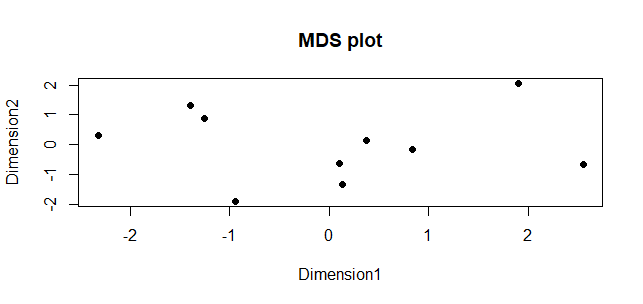
mds\_result<-cmdscale(distance\_matrix,k=2)

cmdscale performs Classical MDS on the distance matrix and reduces the data to 2 dimensions (k=2). The result is stored in mds\_result, which contains the coordinates of the brands in the new 2D space.

Plotting the MDS Results:

|  |
| --- |
| > plot(mds\_result[,1],mds\_result[,2],pch=16,xlab="Dimension1",ylab="Dimension2",main="MDS plot") |
|  |
| |  | | --- | | > | |

* plot creates a scatter plot of the MDS results.
* mds\_result[, 1] and mds\_result[, 2] are the coordinates of the brands in the first and second dimensions, respectively.
* pch=16 specifies the plotting character (a solid dot).
* xlab and ylab label the axes as "Dimension 1" and "Dimension 2".
* main sets the title of the plot to "MDS plot".



### Interpretation of the MDS Plot

The MDS plot shown represents the similarities and dissimilarities between different ice cream brands based on their attributes (Price, Availability, Taste, Flavour, Consistency, Shelf life). Here's how to interpret it:

1. **Proximity of Points**:
   * **Closer Points**: Brands that are plotted closer to each other have more similar attribute values. For instance, if two points are near each other, it means the corresponding brands are similar in terms of price, taste, availability, etc.
   * **Farther Points**: Brands that are farther apart are more dissimilar. If a point is isolated, it means that brand has unique attributes compared to others.
2. **Clusters**:
   * Look for clusters of points. A cluster indicates a group of brands that have similar profiles. If several points are grouped together, those brands share similar attributes and might compete in the same market segment.
3. **Dimensions**:
   * **Dimension 1 (x-axis)**: Represents one major source of variation in the data. Brands on the left side of the plot differ from those on the right side along this dimension. The specific attributes contributing to this dimension are not directly interpretable from the plot, but it indicates one way the brands are different.
   * **Dimension 2 (y-axis)**: Represents another major source of variation. Brands higher up differ from those lower down along this dimension.

### Specific Observations:

1. **Left Cluster**: There is a cluster of points towards the left side of the plot, slightly scattered vertically. These brands have similar characteristics, distinguishing them from brands on the right side.
2. **Right Cluster**: Another group is spread out more horizontally towards the right side. This indicates that while they are more similar to each other compared to the left cluster, there is more variability among them in the attributes.
3. **Middle Points**: Some points are more centered, suggesting these brands have attributes that are somewhat in between the extremes represented by the left and right clusters.

### Business Implications:

1. **Market Segmentation**: The clusters represent different segments in the ice cream market. Understanding these segments can help tailor marketing strategies and product offerings.
2. **Product Development**: Brands that are outliers (isolated points) might indicate unique selling propositions or areas for improvement. Brands in clusters could indicate standard market expectations.
3. **Competitive Positioning**: By knowing which brands are similar, businesses can identify direct competitors and strategize accordingly.

**PYTHON CODES**



**Lines 1-2: Import libraries**

The first two lines import necessary libraries for data manipulation and visualization in Python.

* pandas as pd: This line imports the pandas library, a popular library for data analysis in Python. It provides data structures and data manipulation tools.
* import numpy as np: This line imports the NumPy library, which is used for scientific computing and array manipulation.
* from sklearn.preprocessing import StandardScaler: This line imports the StandardScaler class from scikit-learn, a machine learning library in Python. This class is used to standardize data (i.e. transform features to have a mean of 0 and standard deviation of 1).
* from sklearn.manifold import MDS: This line imports the MDS class from scikit-learn, which is used for dimensionality reduction. In this context, it might be used to reduce the number of features for visualization purposes.
* import matplotlib.pyplot as plt: This line imports the matplotlib library, a popular Python library for creating visualizations.

**Lines 3-4: Read the data**

These lines read the data from a CSV file into a pandas DataFrame.

* icecream\_df = pd.read\_csv("C:\\Users\\SPURGE\\Desktop\\SCMA\\A4\\Icecream.csv"): This line reads the CSV file located at the specified path and stores it in a pandas DataFrame named icecream\_df.

**Lines 5-6: Inspect the data**

These lines inspect the data in the DataFrame to get a better understanding of its contents.

* print(icecream\_df.shape): This line prints the shape of the DataFrame, which indicates the number of rows and columns in the data.
* print(icecream\_df.columns): This line prints the column names of the DataFrame.
* print(icecream\_df.head()): This line prints the first few rows of the DataFrame to get a glimpse of the data.

**Lines 7-9: Data Preprocessing**

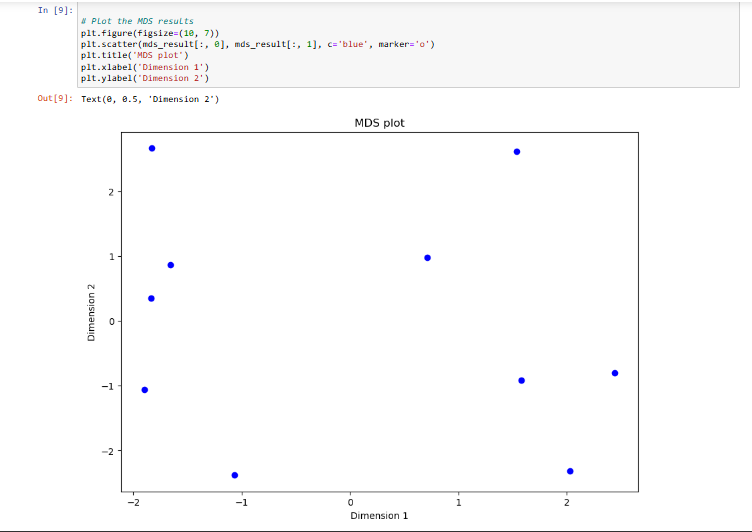
These lines appear to be pre-processing the data for further analysis.

* ice = icecream\_df.drop(columns=['Brand']): This line creates a new DataFrame named ice by dropping the 'Brand' column from the original DataFrame.
* scaler = StandardScaler(): This line creates an instance of the StandardScaler class to standardize the data.
* ice\_scaled = scaler.fit\_transform(ice): This line transforms the data in the ice DataFrame using the standard scaler. This likely centers the data around a mean of 0 and scales it to have a standard deviation of 1.

**Line 10: Compute distance matrix**

This line calculates the distance matrix between data points, likely for use in the MDS algorithm in the next step.

* distance\_matrix = np.linalg.norm(ice\_scaled[:, np.newaxis] - ice\_scaled, axis=2): This line calculates the distance matrix using the NumPy library. It appears to be computing the pairwise Euclidean distance between the data points in the scaled data (ice\_scaled).



* plt.scatter(mds\_result[:, 0], mds\_result[:, 1], co='blue', marker='o'): This line creates a scatter plot using the matplotlib library. It appears to be plotting the first two dimensions (0 and 1) of the MDS results. The data points are colored blue and represented by circles (marker='o').
* plt.title("MDS plot"): This line sets the title of the plot to "MDS plot".
* plt.xlabel("Dimension 1"): This line labels the x-axis of the plot as "Dimension 1".
* plt.ylabel("Dimension 2"): This line labels the y-axis of the plot as "Dimension 2".

**Plot:**

The plot itself is a 2-dimensional scatter plot, which is a common way to visualize higher-dimensional data.

* **Axes:** The plot has two axes, labeled "Dimension 1" and "Dimension 2". These axes represent the first two dimensions extracted from the data using MDS. MDS is a technique used to reduce the dimensionality of data while preserving the relationships between data points. In this case, it has taken higher-dimensional data and projected it onto a 2-dimensional space for visualization.
* **Data Points:** The blue circles in the plot represent the data points after dimensionality reduction. The position of each circle on the axes corresponds to its values in the first two dimensions.

**CONCLUSION**

The Multidimensional Scaling (MDS) analysis of the ice cream dataset provides a valuable visualization of the similarities and dissimilarities among various ice cream brands based on their attributes (Price, Availability, Taste, Flavour, Consistency, Shelf life).

#### Key Findings:

1. **Clustering of Brands**:
   * **Left Cluster**: A group of brands that are similar to each other but distinct from brands on the right side. This suggests these brands share common attributes that set them apart from others.
   * **Right Cluster**: Another group of brands that are similar to each other, though more spread out, indicating a slightly higher variability in their attributes within this cluster.
   * **Middle Points**: Brands in the central region have attributes that are not extreme in any dimension but rather balanced, indicating moderate levels of all attributes.
2. **Distinct Groups**:
   * The clear separation between clusters suggests the presence of distinct market segments within the ice cream market. Each segment represents brands with similar characteristics appealing to specific consumer preferences.
3. **Outliers**:
   * Any isolated points or outliers represent brands with unique attributes. These could either be niche products with specialized appeal or brands that need to align more closely with market expectations to improve their competitive position.

The MDS plot effectively illustrates the competitive landscape of the ice cream market, revealing distinct clusters of brands and their relative similarities and differences. By leveraging these insights, businesses can enhance their market strategies, improve product offerings, and better meet consumer needs, ultimately leading to increased market share and customer satisfaction. This analysis serves as a foundation for making data-driven decisions that align with the dynamic preferences of the ice cream market.