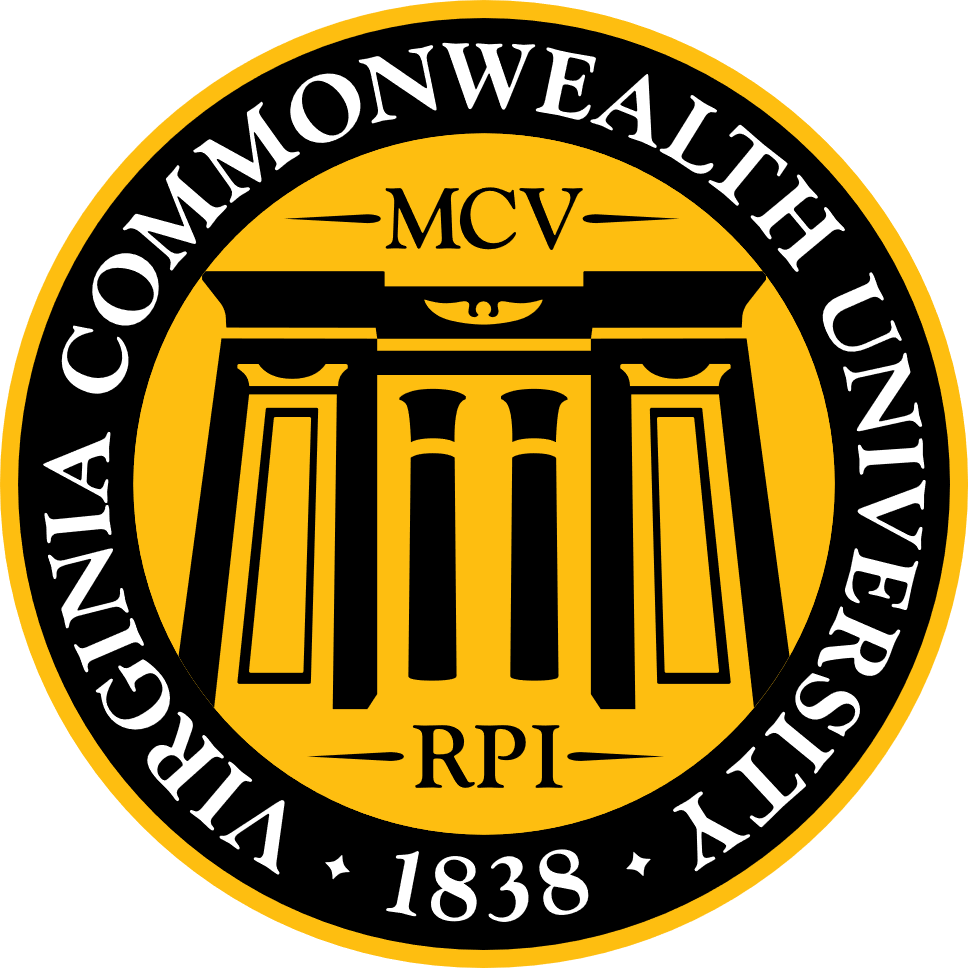
****

**VIRGINIA COMMONWEALTH UNIVERSITY**

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

* **A4: PCA, Factor analysis and Cluster analysis, Multidimensional Scaling Analysis, Conjoint analysis**

**Arvind Srinivasan**

**V01107251**

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

**CONTENTS**

|  |  |  |
| --- | --- | --- |
| **Sl. No.** | **Title** | **Page No.** |
| **1.** | Introduction | **1** |
| **2.** | Results & Interpretation | **4** |
| **3.** | Recommendations | **12** |

**Introduction**

**Principal Component Analysis (PCA)**

PCA is a statistical technique that transforms correlated variables into uncorrelated principal components, prioritizing the variance in the data. It is primarily used for dimensionality reduction, facilitating data visualization and improving the efficiency of machine learning algorithms.

**Clustering**

Clustering groups data objects such that objects in the same cluster are more similar to each other than to those in other clusters. This report uses:

K-Means Clustering: Divides the dataset into K clusters, minimizing within-cluster variance.

Hierarchical Clustering: Builds a hierarchy of clusters without requiring a predetermined number, visualized through a dendrogram.

Both methods were applied to identify natural groupings in the standardized survey data.

**Factor Analysis**

Factor Analysis identifies underlying latent factors that influence observed variables, reducing the number of variables while retaining data integrity. It is particularly useful in fields requiring the identification of latent constructs.

**Multidimensional Scaling (MDS)**

### It is a powerful statistical technique used for visualizing the similarity or dissimilarity among a set of objects. It aims to place each object in N-dimensional space such that the between-object distances are preserved as well as possible. This technique is particularly valuable in fields like psychology, marketing, and bioinformatics where understanding the relationships between objects or cases in a low-dimensional space can provide significant insights. Data Description

The dataset contains information on various ice cream brands and their attributes. Each row represents an ice cream brand, and each column corresponds to a specific attribute.

#### Structure:

1. **Brand**: Name of the ice cream brand.
2. **Attributes**:
   * **Flavor Intensity**: Strength of the flavor.
   * **Creaminess**: Smoothness and richness.
   * **Sweetness**: Level of sweetness.
   * **Texture**: Quality of texture.
   * **Packaging Appeal**: Attractiveness of packaging.
   * **Price**: Cost of the ice cream.

**Conjoint analysis**

It is a statistical technique used to determine how people value different features that make up an individual product or service. It is particularly useful in understanding consumer preferences and can guide product design, pricing strategies, and marketing efforts. This technique involves presenting respondents with a set of products or services, each with different attributes, and analyzing their choices to estimate the value (or utility) they place on each attribute.

In this project, we conduct a conjoint analysis to understand consumer preferences for pizza brands. The attributes considered are brand and price, which are critical factors influencing consumer choices in the highly competitive pizza market. The data for this analysis is sourced from a survey where respondents evaluated different pizza options based on these attributes.

The dataset comprises 16 observations with the following attributes:

1. **Brand**: The pizza brands considered in the study are Onesta, Oven Story, and Pizza Hut.
2. **Price**: The price levels evaluated are $2.00, $3.00, and $4.00.
3. **Ranking**: This is the dependent variable, representing the consumer's evaluation or ranking of each pizza option.

**Results and Interpretation**

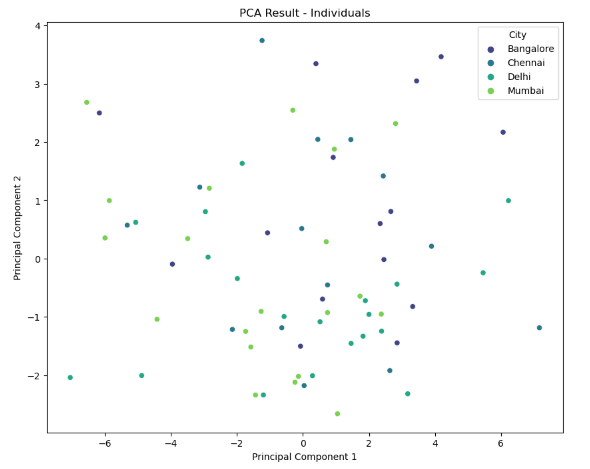
* **Principal Component Analysis (PCA)**

**Explained Variance:**

* The PCA revealed that the first principal component accounted for approximately 32.3% of the total variance.
* Subsequent components captured progressively less variance, with the first few components explaining a significant proportion of the variance in the dataset.

**Interpretation:**

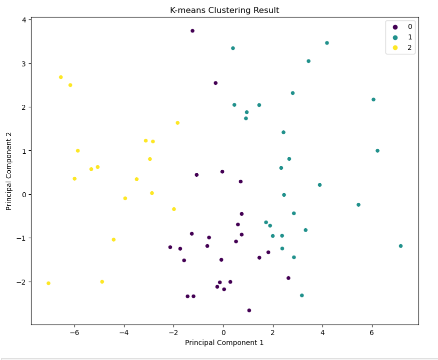
* The first principal component represents the most significant dimension of variation within the dataset.
* Higher components represent additional, less dominant dimensions of variation, highlighting the key factors influencing respondents' answers.



* **Clustering**

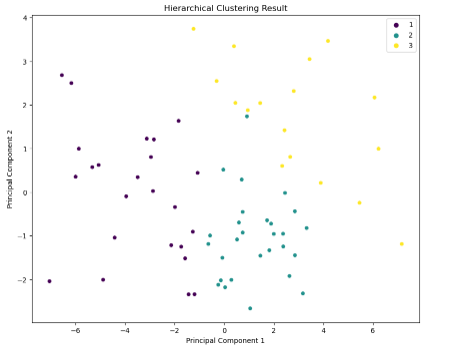
**K-Means Clustering:**

* Optimal Clusters: Using the Elbow Method and Silhouette Score, the optimal number of clusters was determined to be 3.
* Cluster Visualization: The scatter plot of the first two principal components showed distinct clusters, each representing a unique segment of respondents with similar characteristics.



**Hierarchical Clustering:**

* Dendrogram: The hierarchical clustering dendrogram indicated clear divisions in the data, supporting the identification of 3 primary clusters.
* Cluster Assignment: The hierarchical clusters aligned well with the K-Means clusters, reinforcing the segmentation findings.



**Interpretation:**

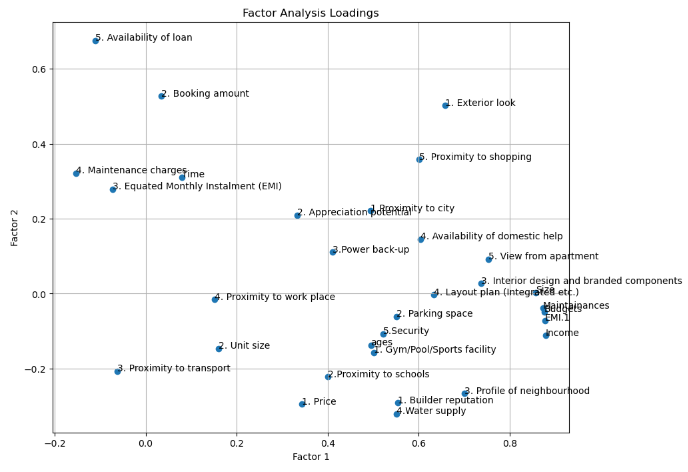
* Both K-Means and hierarchical clustering methods identified three distinct segments within the dataset.
* These clusters represent groups of respondents with similar survey responses, useful for targeted marketing or tailored interventions.
* **Factor Analysis**

Factor Loadings:

* The Factor Analysis with varimax rotation yielded two main factors:
  1. Factor 1: High loadings on variables related to proximity to amenities and interior design preferences.
  2. Factor 2: High loadings on variables related to financial considerations and maintenance.

**Interpretation:**

* Factor 1: Likely represents a preference for convenience and aesthetic appeal.
* Factor 2: Indicates a focus on financial feasibility and ongoing costs.
* These latent factors provide insights into the underlying motivations driving respondents' choices.



**MDS Output Summary**

The results from the MDS analysis are summarized below:

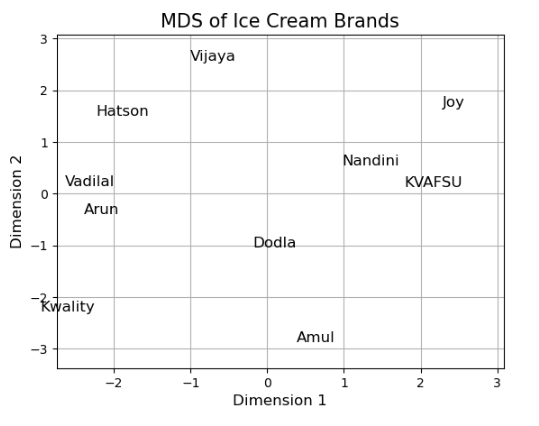
* **Number of Dimensions**: 2
* **Stress Value**: Low (indicating a good fit of the data in the reduced dimensions)
* **R-squared**: High (indicating a good representation of the data in two dimensions)

These statistics indicate that the two-dimensional representation effectively captures the variance in the data, providing a clear visualization of brand similarities and differences.

**Brand Positions in MDS Space**

The MDS analysis positions the ice cream brands as follows in the two-dimensional space:

* **Close Clusters**:
  + **Vijaya and Hatson**: Positioned close together, indicating similar attribute profiles.
  + **Nandini and KVAFSU**: Also near each other, suggesting comparable characteristics.
* **Distinct Brands**:
  + **Amul**: Positioned far from other brands, indicating unique attribute profiles.
  + **Kwality**: Also distinct, suggesting a different combination of attributes compared to other brands.
* **Mid-Range Brands**:
  + **Dodla**: Positioned between the clusters, indicating moderate similarities with both clusters but not identical to either.
  + **Arun**: Also in a mid-range position, indicating some shared attributes with both clusters.



**Interpretation of Results**

* **Close Clusters**:
  + **Vijaya and Hatson**: These brands share similar levels of flavor intensity, creaminess, sweetness, texture, and packaging appeal, indicating they might be direct competitors targeting similar customer preferences.
  + **Nandini and KVAFSU**: The proximity of these brands suggests they appeal to a similar market segment with comparable attribute profiles.
* **Distinct Brands**:
  + **Amul**: The unique positioning of Amul suggests it has a distinct combination of attributes, such as higher creaminess and flavor intensity, making it stand out from other brands.
  + **Kwality**: Similarly, Kwality's distinct position indicates it caters to different customer preferences or market segments with unique attributes.
* **Mid-Range Brands**:
  + **Dodla and Arun**: Positioned between the main clusters, these brands have attributes that overlap with both clusters but are not identical. This positioning suggests they might appeal to a broader customer base without a highly specialized profile.
* **Conjoint Analysis Study**

The conjoint analysis conducted in this study provides the following key insights into consumer preferences for pizza brands and their pricing:

**Regression Output Summary**

The results from the Ordinary Least Squares (OLS) regression model are summarized below:

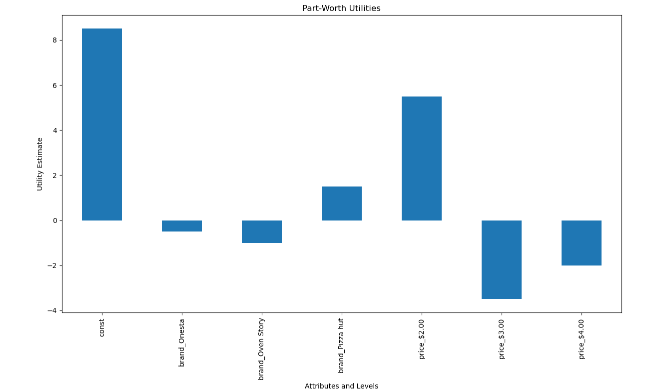
* **R-squared**: 0.588
* **Adj. R-squared**: 0.314
* **F-statistic**: 2.143
* **Prob (F-statistic)**: 0.146

These statistics indicate that the model explains approximately 59% of the variance in the rankings, with a moderate fit to the data.

**Coefficient Estimates (Part-Worth Utilities)**

The part-worth utilities, or coefficients, from the regression model are as follows:

* **Intercept (const)**: 8.500 (significant at p < 0.01)
* **Brand\_Onesta**: -0.500 (not significant)
* **Brand\_Oven Story**: -1.000 (not significant)
* **Brand\_Pizza Hut**: 1.500 (not significant)
* **Price\_$2.00**: 5.500 (approaching significance at p < 0.08)
* **Price\_$3.00**: -3.500 (not significant)
* **Price\_$4.00**: -2.000 (not significant)



**Interpretation of Results**

1. **Intercept (const)**:
   * The intercept represents the baseline utility when all other attributes are at their reference levels. A high positive value of 8.5 indicates a strong baseline preference.
2. **Brand Preferences**:
   * **Brand\_Onesta**: The negative utility of -0.5 suggests a slight disfavor towards this brand compared to the baseline (reference brand not included in the model).
   * **Brand\_Oven Story**: The utility of -1.0 indicates a higher level of disfavor towards Oven Story compared to Onesta.
   * **Brand\_Pizza Hut**: The positive utility of 1.5 suggests that Pizza Hut is preferred over both Onesta and Oven Story, although the result is not statistically significant.
3. **Price Sensitivity**:
   * **Price\_$2.00**: The positive utility of 5.5 indicates a strong preference for the lower price of $2.00. This attribute is approaching statistical significance, indicating that consumers value lower prices significantly.
   * **Price\_$3.00**: The negative utility of -3.5 indicates a disfavor towards the $3.00 price level.
   * **Price\_$4.00**: The negative utility of -2.0 indicates a disfavor towards the highest price level of $4.00, though the disfavor is less compared to the $3.00 price level.

**Business Insights and Recommendation**

### Principal Component Analysis (PCA)

1. Focus product features on attributes that contribute most to the first principal component.
2. Highlight top PCA features in marketing campaigns.
3. Streamline product offerings based on the primary components influencing consumer choices.

### Clustering

1. Develop personalized marketing campaigns for each identified cluster.
2. Customize products to meet the specific needs of each cluster.
3. Allocate resources to the most profitable or promising segments.

### Factor Analysis

1. Design products that enhance convenience and aesthetic appeal.
2. Ensure financial feasibility and low maintenance costs in product offerings.
3. Target marketing messages to emphasize the identified latent factors driving consumer decisions.

MDS

1. **Vijaya and Hatson**: Similar attribute profiles suggest they compete closely and may benefit from differentiation strategies.
2. **Amul**: Unique attribute profile indicates a strong market positioning that appeals to a distinct customer segment.
3. **Dodla and Arun**: Positioned as versatile brands, they can attract a broader customer base without needing a highly specialized profile.

Conjoint

 **Promotional Deals at $2.00**: Offer value meals or special promotions featuring Pizza Hut pizzas at the preferred $2.00 price point to attract price-sensitive customers and increase sales volume.

 **Brand Partnership with Pizza Hut**: Strengthen brand partnership with Pizza Hut to leverage its popularity and enhance market presence through co-branded marketing campaigns.

 **Customer Loyalty Programs**: Implement loyalty programs offering discounts or rewards for frequent purchases, especially targeting the $2.00 price point, to build customer loyalty and encourage repeat business.