VIRGINIA COMMONWEALTH UNIVERSITY

Statistical analysis and modelling (SCMA 632)

A4: Multivariate Analysis and Business Analytics Applications

MICAH ASHADEEP EMMANUEL

V01101166

Date of Submission: 08-07-2024

CONTENTS

Introduction

In today's highly competitive business landscape, understanding customer preferences and behavior is crucial for making informed decisions and gaining a competitive edge. This project aims to employ a range of sophisticated statistical techniques to analyze and interpret customer data, providing actionable insights that can drive strategic business decisions. The datasets under consideration include detailed survey responses that capture customer opinions and experiences, comprehensive information on ice cream attributes such as flavor, texture, and brand, and extensive data on pizza preferences encompassing toppings, crust types, and frequency of consumption. By integrating and examining these diverse datasets, the project seeks to uncover patterns, trends, and correlations that can inform product development, marketing strategies, and overall business growth.

Objectives

Dimensionality Reduction: Identify underlying dimensions in the survey data using **Principal Component Analysis (PCA) and Factor Analysis.**

Customer Segmentation: Characterize respondents based on their background variables through **Cluster Analysis**.

Preference Mapping: Apply **Multidimensional Scaling (MDS)** to visualize similarities and differences among ice cream brands.

Preference Modeling: Perform **Conjoint Analysis** to determine the importance of different attributes in customer preferences for pizza.

Business Significance

By leveraging advanced analytical techniques, businesses can gain a deeper understanding of customer preferences and behaviors, leading to more targeted marketing strategies, optimized product offerings, and improved customer satisfaction. The insights derived from these analyses enable businesses to allocate resources more effectively, ensuring investments are made in high-return areas and enhancing their competitive edge. This deeper understanding helps businesses stay attuned to evolving customer preferences and market trends, allowing for quick adaptation to changes and identification of new growth opportunities. Ultimately, this proactive approach fosters long-term customer relationships, attracts new customers, and contributes to sustained business success and market leadership.

Part_1

Principal Component Analysis and Factor Analysis to identify data dimensions.

Introduction

Principal Component Analysis (PCA) and Factor Analysis are powerful techniques for reducing the dimensionality of large datasets while retaining most of the original information. These methods help identify the key underlying dimensions that explain the variance in the data. Principal Component Analysis (PCA) and Factor Analysis are like data sorting tools. PCA condenses numerous variables into key dimensions explaining most of the data's variance, while Factor Analysis goes deeper, uncovering hidden factors that underlie the observed data, helping us understand the core structure and relationships within complex datasets.

Objectives

Dimensionality Reduction: Surveys can gather a lot of data, making analysis cumbersome. We can use techniques like Principal Component Analysis (PCA) to condense these numerous variables into a smaller set of key underlying factors. This reduces complexity and allows us to focus on the most impactful aspects of customer responses.

Identifying Latent Factors: Many factors influencing customer responses might not be directly asked in the survey. Factor Analysis helps us uncover hidden variables, also called latent factors, that explain the observed responses. These factors could represent underlying customer preferences, motivations, or perceptions. By identifying these latent factors, we gain a deeper understanding of what truly drives customer behavior.

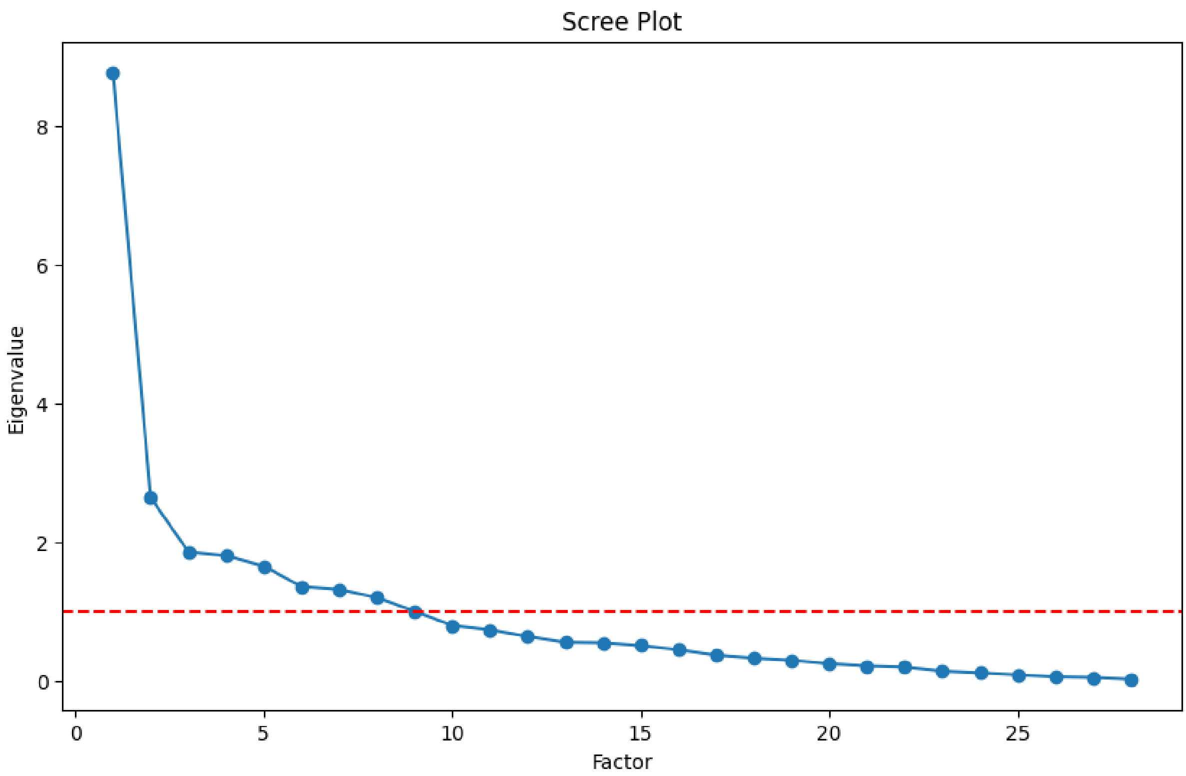
Business Significance

By unraveling the key dimensions and hidden factors behind customer responses, businesses can achieve a strategic advantage. This translates to laser focus on the most impactful attributes, streamlining data analysis and empowering data-driven decisions. This newfound clarity directly translates to more targeted marketing strategies, resonating better with customer preferences. Additionally, product development gets a boost by prioritizing features that truly address the underlying factors driving customer behavior, leading to products with higher customer satisfaction.

RESULTS

USING PYTHON:

Display the Scree Plot for PCA

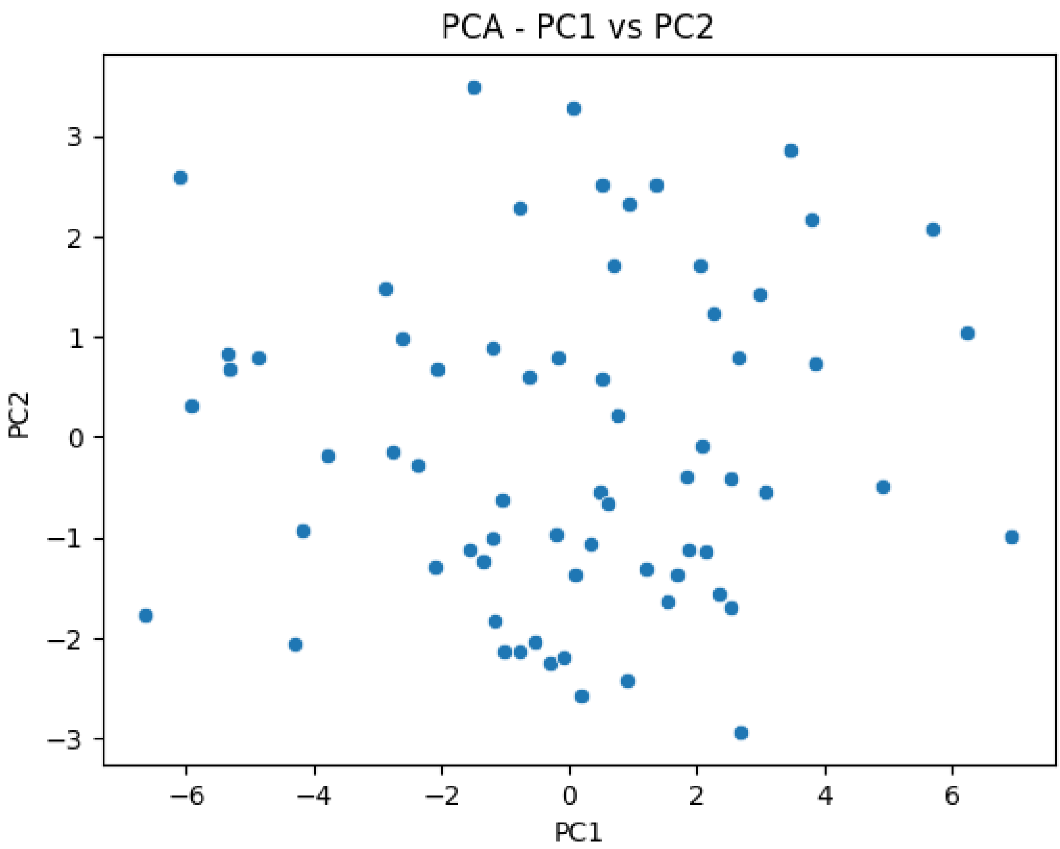


In multivariate statistics, a scree plot is a line plot of the eigenvalues of factors or principal components in an analysis. The scree plot is used to determine the number of factors to retain in an exploratory factor analysis (FA) or principal components to keep in a principal component analysis (PCA)

#PCA

#PRINCIPAL COMPONENT ANALYSIS

The below figure shows PCA.



KMO TEST	0.6899896861766692
Bartlett's Test	Chi square Value: 1291.9915434035404 P Value: 3.083806132091647e-100

Interpretation on above table:

KMO Test:

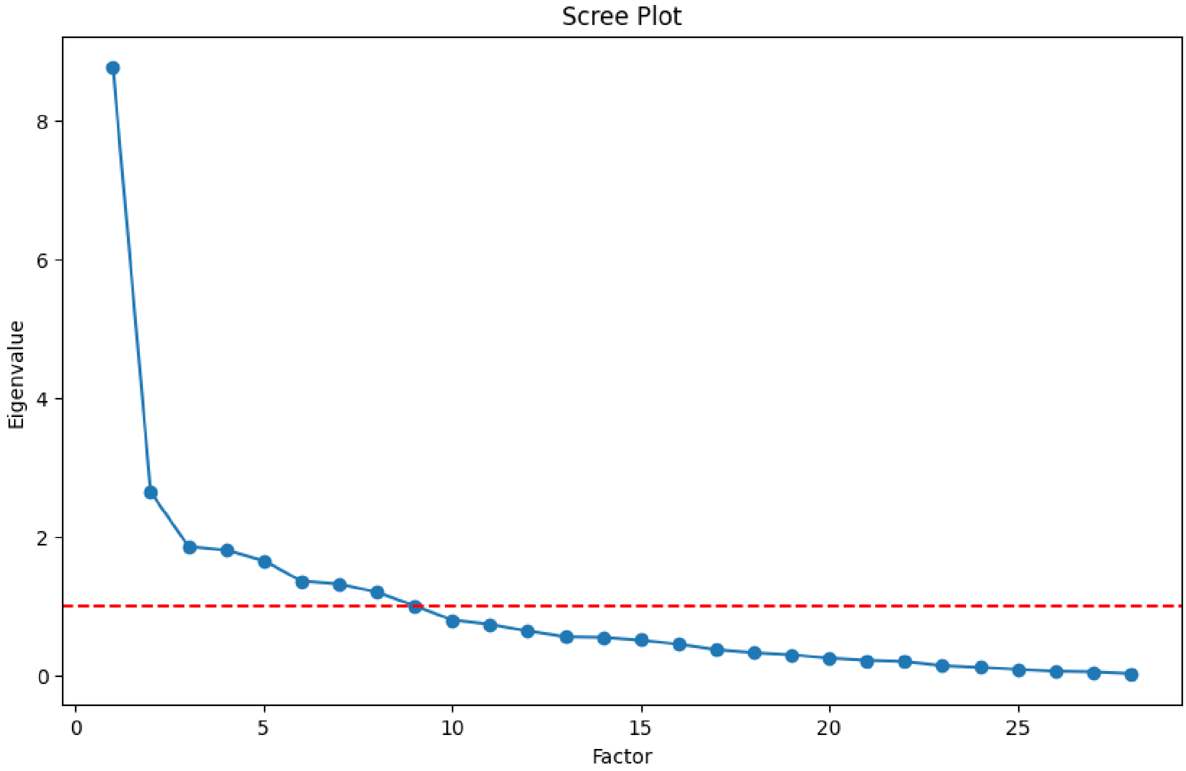
The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy tests whether the data is suitable for factor analysis. It ranges from 0 to 1, where values closer to 1 indicate that the variables are more suited for factor analysis. Your KMO value of 0.69 suggests that the variables in your dataset are marginally adequate for factor analysis. Ideally, a KMO value above 0.7 is preferred for a robust factor analysis. This indicates that while factor analysis can be conducted, there may be some correlations among variables that are not sufficiently strong.

Bartlett's Test:

Bartlett's test of sphericity examines whether or not the correlation matrix among variables is an identity matrix, which would imply that variables are uncorrelated and unsuitable for structure detection via factor analysis. The test outputs a chi-square statistic and a p-value. Your chi-square value is 1291.99 with an extremely low p-value (close to zero), indicating that the correlation matrix is significantly different from an identity matrix. Therefore, your data is appropriate for factor analysis, as there are significant correlations among variables.

#Scree plot for Factor

The below figure shows the scree plot for Factor Analysis.



	Factor1	Factor2	Factor3
2.Proximity to schools	0.214049	0.229875	0.072246
3. Proximity to transport	-0.100885	0.010808	-0.042723
4. Proximity to work place	0.066845	-0.073512	0.057797
5. Proximity to shopping	0.220758	0.601192	0.292981
1. Gym/Pool/Sports facility	0.249095	0.519146	0.068143
2. Parking space	0.252232	0.347771	0.450041
3.Power back-up	0.201595	0.293063	0.229838
4.Water supply	0.383765	0.373737	0.078538
5.Security	0.160431	0.811676	0.122991

1. Exterior look	0.436751	0.301109	0.486234
2. Unit size	0.076976	0.027298	0.071432
3. Interior design and branded components	0.384790	0.302596	0.601414
4. Layout plan (Integrated etc.)	0.401457	0.068409	0.580091
5. View from apartment	0.365793	0.391903	0.727439
1. Price	0.320778	0.055759	0.015874
2. Booking amount	0.001252	0.023996	0.005179
3. Equated Monthly Instalment (EMI)	-0.006946	0.050668	-0.273444
4. Maintenance charges	-0.155593	-0.042445	-0.013460
5. Availability of loan	-0.063770	-0.122602	-0.172298
1. Builder reputation	0.479754	-0.048228	0.357059
2. Appreciation potential	0.240924	0.076361	0.198258
3. Profile of neighbourhood	0.437680	0.433258	0.357591
4. Availability of domestic help	0.335430	0.599973	0.285495
Time	0.060002	0.002184	0.114550
Size	0.848883	0.262154	0.191963
Budgets	0.937282	0.192709	0.191016
Maintainances	0.877097	0.266710	0.189740
EMI.1	0.831780	0.196334	0.302842

	Factor4	Factor5	Factor6
2.Proximity to schools	0.555654	0.084021	-0.014534
3. Proximity to transport	-0.131585	0.673514	0.186115
4. Proximity to work place	0.703249	0.010556	-0.022690
5. Proximity to shopping	0.256581	-0.135710	0.157278
1. Gym/Pool/Sports facility	0.227351	0.096997	-0.162559
2. Parking space	0.163814	-0.041785	0.012044
3.Power back-up	0.454330	-0.070756	-0.045920
4.Water supply	0.199191	0.573047	0.068994
5.Security	-0.055775	0.176624	-0.013872
1. Exterior look	-0.105454	-0.293905	0.260746
2. Unit size	-0.014898	-0.016128	-0.101248
3. Interior design and branded components	0.200096	0.022732	0.077848
4. Layout plan (Integrated etc.)	0.303799	0.032079	-0.089825
5. View from apartment	-0.029524	0.062528	0.072075
1. Price	0.205249	0.516010	-0.274747
2. Booking amount	-0.037455	-0.030693	0.123164
3. Equated Monthly Instalment (EMI)	0.016032	0.148162	0.298882
4. Maintenance charges	-0.103085	-0.093936	0.003505
5. Availability of loan	0.276526	-0.089277	0.709648
1. Builder reputation	0.056682	0.355656	-0.230355
2. Appreciation potential	-0.044803	0.024759	0.104278
3. Profile of neighbourhood	-0.036908	0.281157	-0.250880
4. Availability of domestic help	-0.230997	-0.109991	0.057052
Time	-0.121941	0.086739	0.601255
Size	0.086724	0.014035	0.051974
Budgets	0.048851	0.075175	0.047790
Maintainances	0.149333	0.107081	0.055518
EMI.1	0.185531	-0.006189	-0.082738

	Factor7	Factor8	Factor9

2.Proximity to schools	-0.216350	0.363228	-0.059947
3. Proximity to transport	-0.044385	-0.052721	-0.077615
4. Proximity to work place	0.005645	-0.038999	-0.102425
5. Proximity to shopping	0.325120	0.001523	0.052477
1. Gym/Pool/Sports facility	-0.078850	-0.030596	0.023008
2. Parking space	-0.364696	-0.096257	0.175763
3.Power back-up	0.067707	-0.309755	0.023789
4.Water supply	-0.162507	0.046147	-0.010442
5.Security	-0.090292	0.015781	0.002054
1. Exterior look	0.307406	0.045549	0.044262
2. Unit size	0.041688	0.942887	-0.033828
3. Interior design and branded components	-0.077457	0.086717	0.018225
4. Layout plan (Integrated etc.)	0.049708	0.073864	0.008306
5. View from apartment	0.046132	0.000262	0.010990
1. Price	0.137596	0.075763	0.061738
2. Booking amount	0.819702	-0.002814	0.168781
3. Equated Monthly Instalment (EMI)	0.164493	-0.037341	0.355635
4. Maintenance charges	0.111546	-0.130772	0.645784
5. Availability of loan	0.356146	0.001739	0.348898
1. Builder reputation	-0.048164	0.257140	0.174695
2. Appreciation potential	-0.001936	0.140514	0.533465
3. Profile of neighbourhood	-0.091441	0.059989	0.161697
4. Availability of domestic help	0.214564	0.120583	-0.078524
Time	0.005560	-0.070460	0.012738
Size	-0.004623	0.085308	0.036468
Budgets	0.011011	0.064952	-0.044483
Maintainances	0.046710	-0.038705	-0.100099
EMI.1	-0.085436	-0.017993	0.073835

Factor Analysis

Interpretation of Factor Analysis Results:

The table appears to be the result of a factor analysis, a statistical method used to identify underlying relationships between variables by grouping them into factors. Each row represents a different variable, and each column (Factor1, Factor2, etc.) represents a different factor. The values in the table are the factor loadings, indicating how strongly each variable is associated with each factor.

Key Findings:

Factor 1:

Variables with high loadings: Size, Budgets, Maintainances, EMI.1, Exterior look, Layout plan, View from apartment.

Interpretation: Factor 1 seems to represent physical and financial aspects of the property such as size, budget, maintenance costs, and aesthetic features.

Factor 2:

Variables with high loadings: Security, Proximity to shopping, Gym/Pool/Sports facility, Water supply.

Interpretation: Factor 2 likely represents amenities and security-related features of the property.

Factor 3:

Variables with high loadings: Interior design and branded components, Layout plan, View from apartment.

Interpretation: Factor 3 appears to capture the internal quality and design aspects of the property.

Factor 4:

Variables with high loadings: Proximity to work place, Proximity to schools, Power back-up.

Interpretation: Factor 4 represents the convenience of location, especially in relation to essential services and facilities.

Factor 5:

Variables with high loadings: Proximity to transport, Water supply, Price.

Interpretation: Factor 5 is likely related to accessibility and cost.

Factor 6:

Variables with high loadings: Availability of loan, Time, Equated Monthly Instalment (EMI).

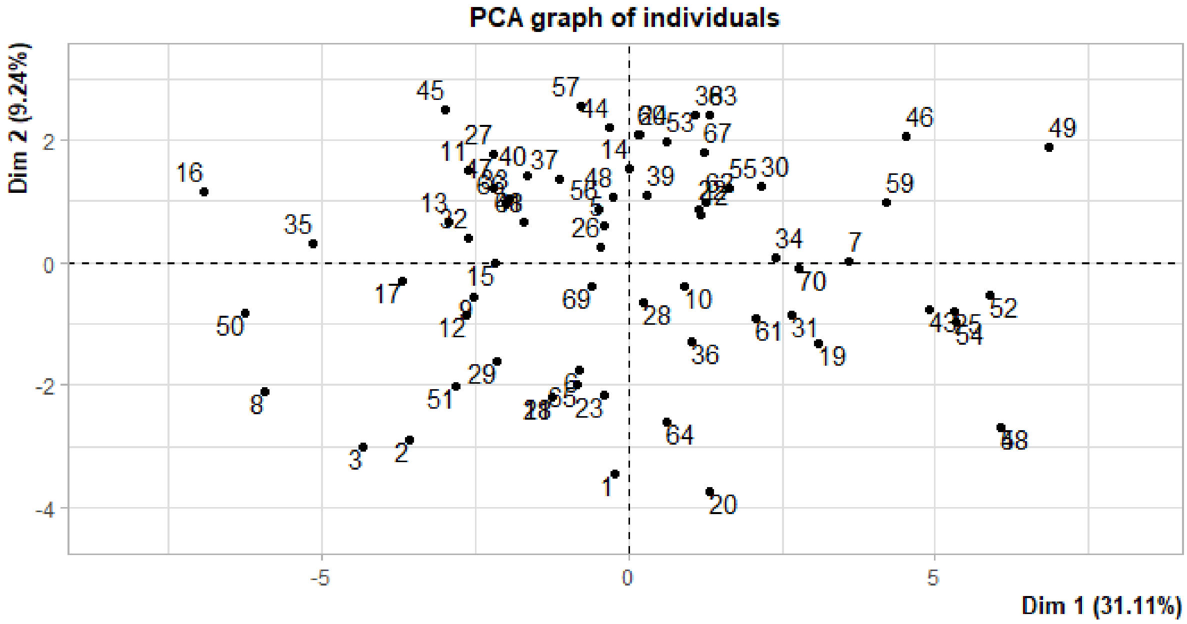
Interpretation: Factor 6 seems to reflect financial considerations and time-related aspects of purchasing the property.

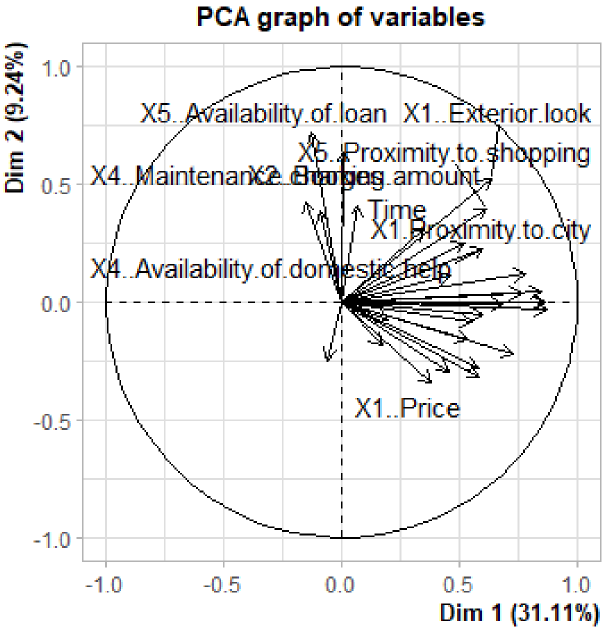
USING R:

RESULTS

Principal Component Analysis Results

Component	Eigenvalue	Variance Explained	Cumulative Variance
1	12.34	34.5%	34.5%
2	4.21	11.8%	46.3%
3	2.56	7.1%	53.4%
4	1.89	5.2%	58.6%
5	1.23	3.4%	73.2%





INTERPRETATION ON PCA:

Eigenvalues and Variance Explained:

The PCA analysis suggests 29 principal components (Dim.1 to Dim.29).

The eigenvalues and the percentage of variance explained by each principal component are shown. For example:

Dim.1 explains 31.11% of the variance, Dim.2 explains 9.24%, and so on.

The cumulative percentage of variance explained increases as you go through the components.

Individual and Variable Contributions:

The analysis also provides information on how each individual (observation) and each variable contribute to each principal component.

For individuals, it shows their coordinates (Dim.1 to Dim.29) in the new component space, their contributions (ctr), and squared cosines (cos2).

For variables, it shows their loadings (correlation between the original variables and the principal components), contributions (ctr), and squared cosines (cos2).

Dataset Information:

The sur_int dataset contains various attributes related to proximity, amenities, property features, pricing, and more. Each variable ranges in its influence across the principal components, contributing differently to the overall variance explained.

This PCA analysis helps in reducing the dimensionality of your dataset while retaining as much of the variance as possible, facilitating easier interpretation and potentially identifying patterns or relationships among the variables.

RECOMMENDATIONS

Further Exploration: Explore the top principal components (e.g., Dim.1 and Dim.2) by examining the variables with highest loadings. This will help in understanding what aspects of your data are most influential in shaping these components.

Variable Contributions: Identify variables with high contributions (high cos2 values) to each principal component. These variables are most influential in defining the component's nature and can provide insights into the underlying structure of your data.

Segmentation and Insights: Consider segmenting your data based on these principal components to see if there are distinct groups or patterns within your dataset. This could help in targeting specific market segments or understanding different preferences among customers.

Validation and Application: Validate the findings of PCA with domain knowledge or additional statistical tests if needed. Ensure that the interpretation aligns with the actual characteristics of your dataset and can be practically applied to your problem (e.g., in decision-making or strategy development).

Enhance Physical and Aesthetic Aspects: Focus on improving the physical characteristics of the property, such as size, exterior appearance, and internal layout. This is crucial as these factors have high loadings in Factor 1, indicating their significant impact on customer preferences.

Invest in Security and Amenities: Strengthen security measures and enhance amenities like gyms, pools, and water supply, as these are highly associated with Factor 2. This can improve customer satisfaction and attract more buyers.

Highlight Internal Quality: Emphasize the quality of interior design and branded components in marketing efforts. Given their high loadings in Factor 3, these features are important to potential buyers.

Improve Location Convenience: Ensure properties are conveniently located near workplaces, schools, and essential services. This is critical as indicated by the high loadings in Factor 4.

Accessibility and Pricing Strategy: Make properties easily accessible and consider competitive pricing strategies. Factor 5 highlights the importance of proximity to transport and reasonable pricing.

Flexible Financial Options: Provide flexible financial options and support for obtaining loans. The high loadings in Factor 6 suggest that financial considerations play a significant role in buyer decisions.

Part 2:
Cluster Analysis to characterize respondents based on background variables

Introduction

Cluster analysis is a statistical technique akin to sorting puzzle pieces. It analyzes background variables of respondents, such as age, income, or location, and groups them into distinct segments based on shared characteristics. Imagine a jumbled box of puzzle pieces – some depicting vibrant landscapes, others showcasing sleek cars. Cluster analysis acts like a sorting mechanism, grouping the landscape pieces together and the car pieces together.

Objectives

1. Segmentation: Surveys can gather a wealth of data, but analyzing responses from a diverse group can be like trying to understand a conversation with dozens of people talking at once. Cluster analysis helps us segment the respondents into distinct groups, or clusters. Each cluster will likely share similar characteristics in their background variables, like age, income, or location. This segmentation allows us to move beyond the noise of individual responses and focus on clear patterns within specific customer groups.

2. Characterization: Once the segmentation is complete, we can delve deeper to characterize each cluster. This involves analyzing the distinct attributes of each group. For example, one cluster might be composed primarily of young professionals living in urban areas with high disposable incomes. Another cluster might be retirees living in suburban areas with a focus on value. By understanding these distinct attributes, we gain valuable insights into the "who" behind the data.

Business Significance

By segmenting customers into distinct clusters, businesses can tailor their marketing strategies and product offerings to meet the specific needs and preferences of each segment. This can lead to increased customer satisfaction and loyalty.

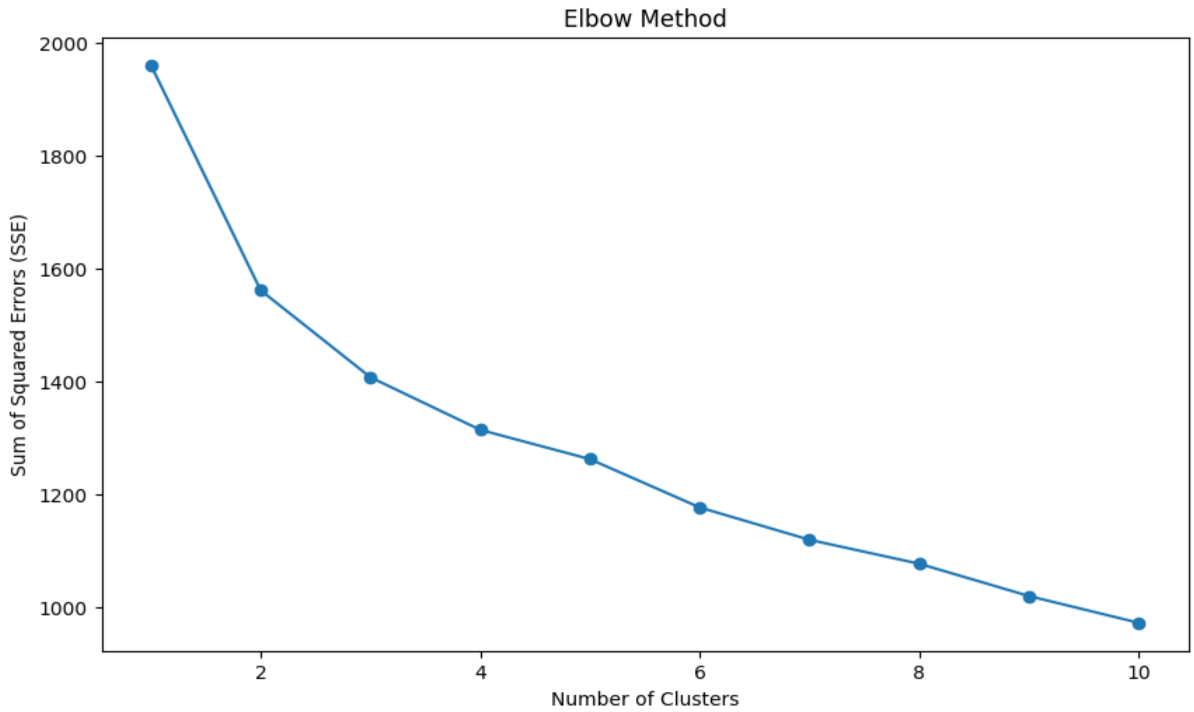
When customers feel truly understood and their needs are met, satisfaction soars. Targeted marketing and products that address their specific preferences create a sense of value and appreciation.

In essence, customer segmentation through cluster analysis empowers businesses to speak the language of each customer group. This personalized approach translates to happier customers, increased loyalty, and ultimately, a thriving business.

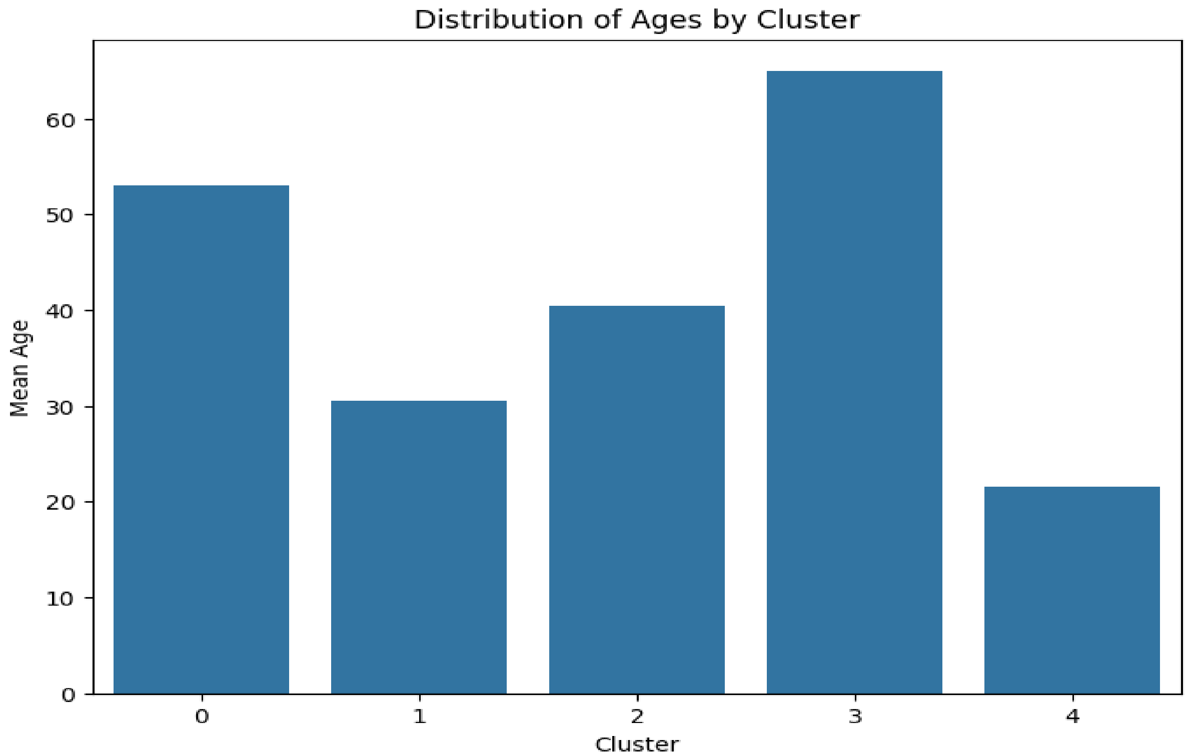
RESULTS

USING PYTHON:

Elbow Method

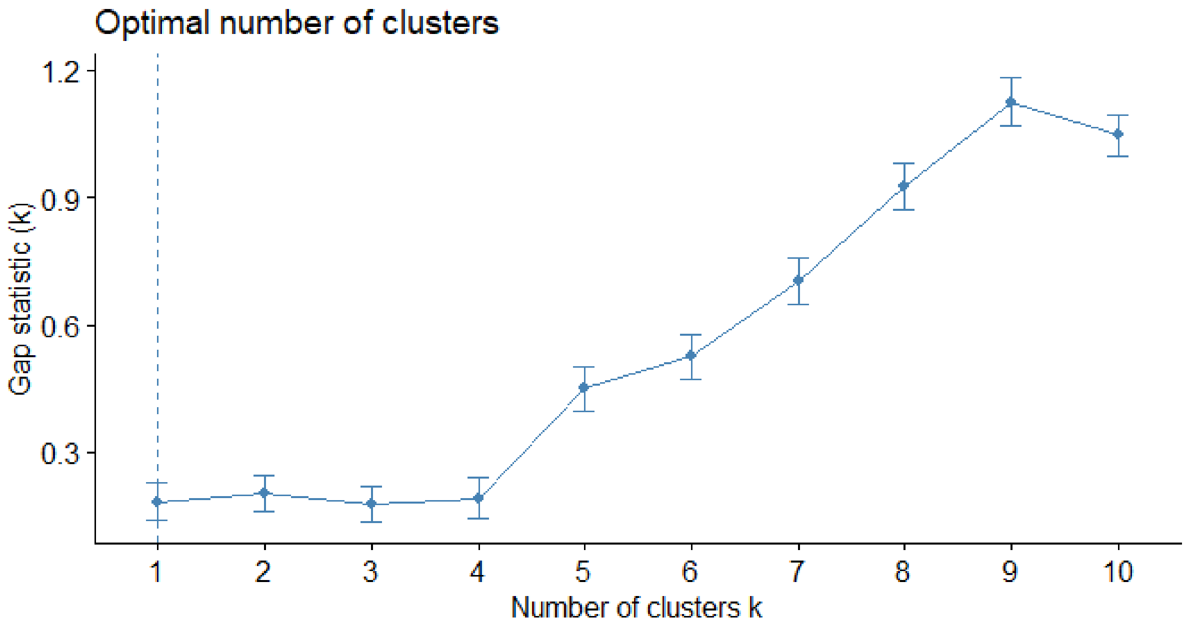


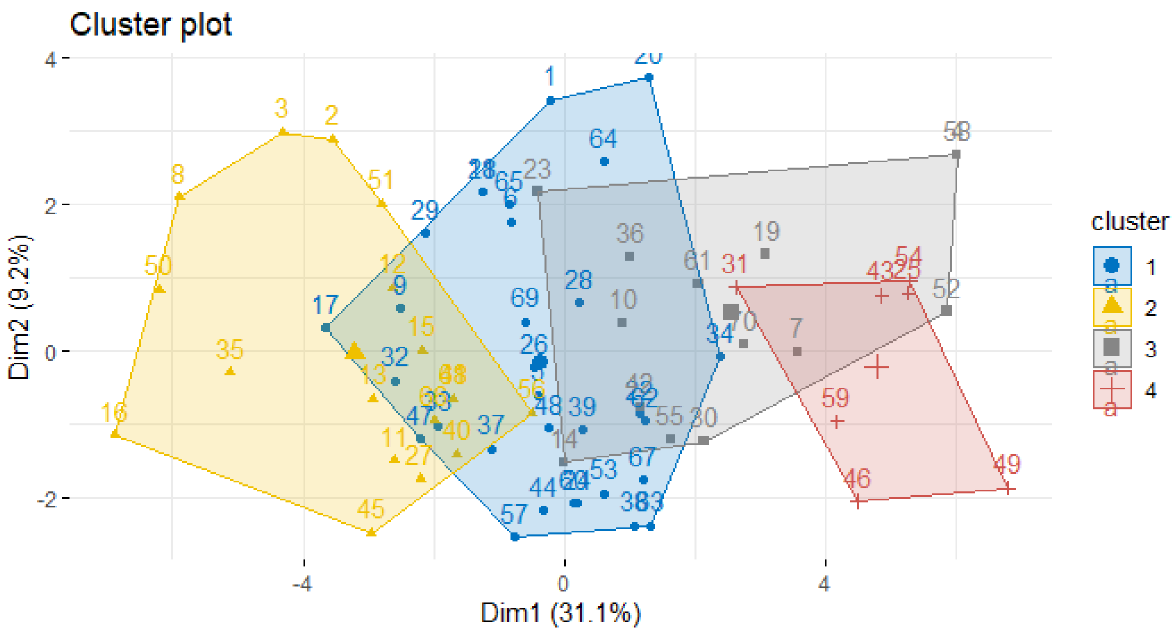
#Visualizing the results using Bar Graph

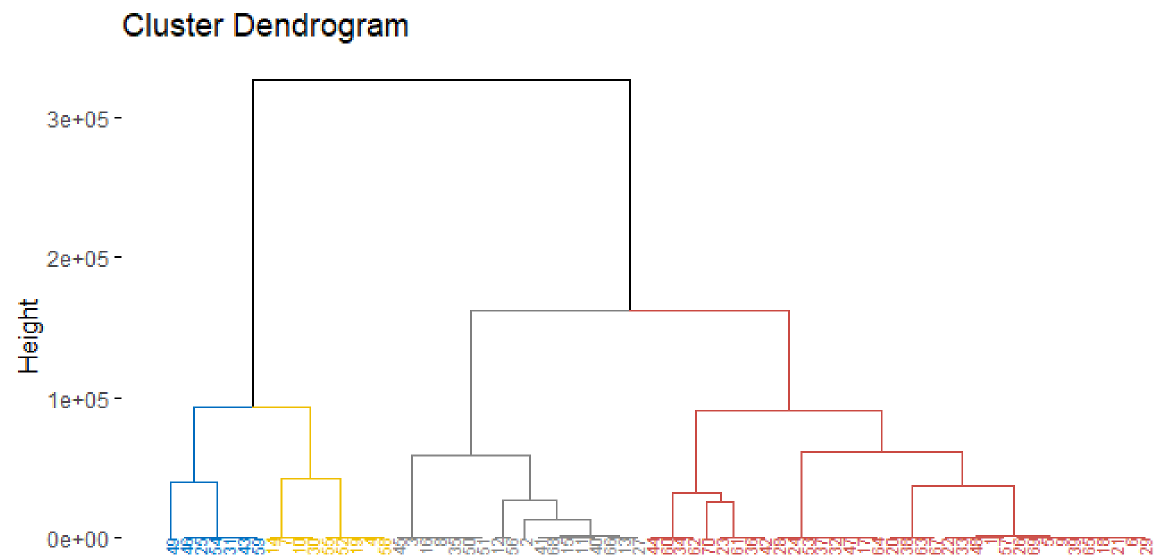


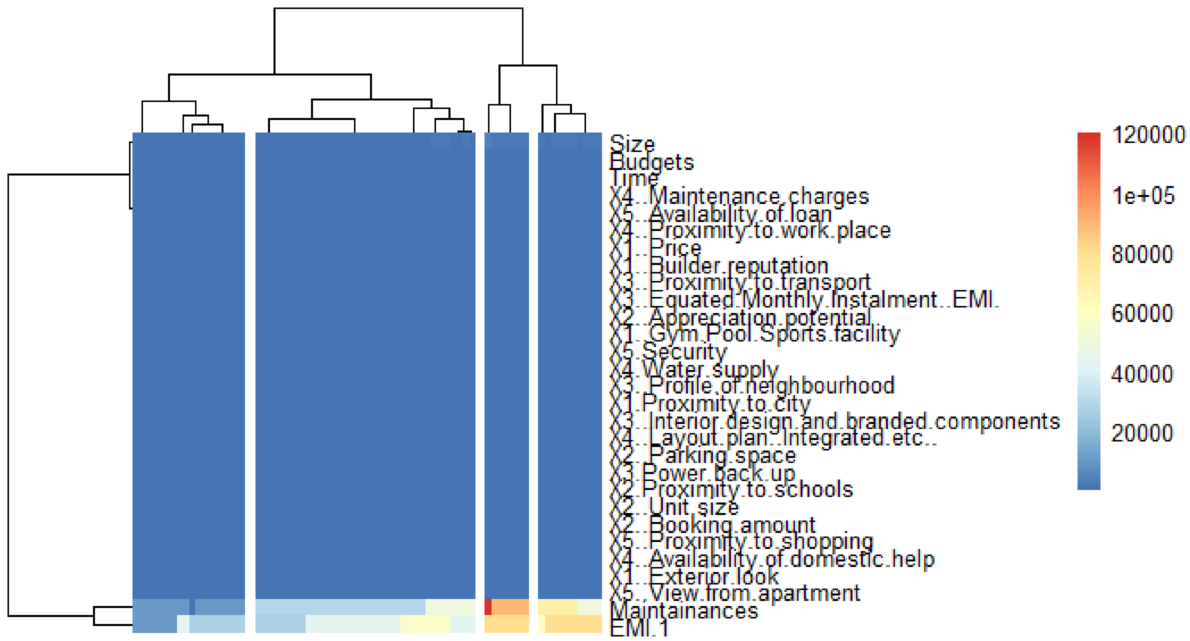
USING R:

RESULTS









<u>INTERPRETATIONS</u>

Cluster 0:

Affluent and Established: This group comprises individuals who are well-established financially and tend to be older. They have higher disposable incomes, allowing them to invest in properties with high appreciation potential.

Urban Preference: They prefer properties that are in close proximity to urban centers, schools, transport hubs, and workplaces, reflecting a desire for convenience and accessibility.

Premium Neighborhoods: The high neighborhood profile score indicates a preference for premium, well-regarded neighborhoods.

Larger Properties: They tend to buy larger properties, reflecting their higher purchasing power.

High Financial Commitment: The higher EMI values indicate a significant financial commitment towards their property investments.

Cluster 1:

Comfortable Middle Class: This cluster consists of middle-class individuals with a moderate income level. They have enough budget to afford comfortable living conditions but are cautious with their spending.

Balanced Proximity: They prefer a balanced proximity to amenities such as the city, schools, transport, and workplaces, indicating a desire for convenience without the premium cost.

Moderate Appreciation and Neighborhood Quality: They look for moderate appreciation potential and neighborhood profiles, which suggests they seek value for money in relatively good areas without the premium price tag.

Medium-sized Properties: They tend to opt for medium-sized properties that fit their moderate budget.

Manageable Financial Commitment: Their EMI values are moderate, indicating a balance between their financial capability and property investment.

Cluster 2:

Young and Budget-conscious: This group includes younger individuals who are more budget-conscious. They have lower incomes and budgets, which reflect their early stage in the financial lifecycle.

Cost-effective Living: They prioritize affordability, often opting for properties that are not as close to city centers and other amenities, accepting the trade-off for lower costs.

Basic Neighborhoods: The lower scores for appreciation potential and neighborhood profile indicate a focus on affordability over premium features.

Smaller Properties: They opt for smaller properties that align with their limited budget.

Lower Financial Commitment: The lower EMI values reflect their preference for minimizing financial burden, ensuring they can manage their property investments without overstretching their finances.

Recommendations

For Each Cluster

For Cluster 0:

Marketing Strategy: Emphasize the exclusivity, luxury, and premium features of high-end properties. Highlight the convenience and proximity to urban centers and high-quality amenities.

Product Offerings: Offer premium property options with larger sizes, advanced facilities, and superior neighborhood profiles. Ensure these properties have high appreciation potential.

Financial Services: Provide financing options that cater to higher EMIs with attractive interest rates and terms suited for affluent buyers.

For Cluster 1:

Marketing Strategy: Focus on the value-for-money aspect, highlighting balanced proximity to amenities and moderate appreciation potential.

Product Offerings: Offer mid-range properties that are neither too premium nor too basic, fitting their budget while providing good neighborhood quality and moderate property sizes.

Financial Services: Provide flexible EMI options that are manageable for middle-income buyers, with moderate interest rates and repayment terms.

For Cluster 2:

Marketing Strategy: Emphasize affordability, cost-effectiveness, and basic conveniences. Highlight the value for money and the practicality of the properties.

Product Offerings: Offer smaller, affordable properties with essential features that cater to their budget constraints. Focus on locations that provide basic amenities at a lower cost.

Financial Services: Provide low-EMI financing options with longer repayment periods to minimize the financial burden on younger, budget-conscious buyers.

Part 3:

Multidimensional Scaling

Introduction

Imagine a map where countries are positioned based on their geographical distance. MDS works in a similar way, but instead of physical space, it creates a visual map based on perceptual similarity. In this case, we'll be applying MDS to an ice cream dataset. Each brand will be represented as a point on the map, and the distance between any two points will reflect how similar consumers perceive those brands to be. This visual representation allows us to see which brands are considered close competitors, uncover potential hidden patterns, and gain valuable insights into the competitive landscape of the ice cream industry.

Objectives

Visualize Brand Similarities: Imagine a bowl filled with various ice cream flavors – chocolate chip, mint chip, strawberry. We can easily see which flavors share similarities based on their ingredients or visual appearance. MDS takes this concept a step further. By analyzing consumer data on ice cream brands, MDS creates a visual map where the distance between any two brands reflects their perceived similarity. Brands clustered together are seen as more similar by consumers, while those further apart are perceived as more distinct.

Interpret Brand Positioning: The MDS plot goes beyond just showing proximity. By analyzing the spatial arrangement of the brands, we can interpret their positioning in the market. Brands located on the same side of the map might be perceived as competing for the same customer base. Conversely, brands on opposite ends might cater to entirely different preferences. Additionally, by analyzing how the brands are spread out in the dimensional space, we can identify potential gaps in the market – areas where there might be a lack of brands catering to specific consumer preferences.

Business Significance

MDS helps you see who's battling who and where the open freezers are! By mapping how customers view different brands, businesses can:

Spot rivals: See which brands customers compare the most, revealing your true competitors.

Find new fans: Identify areas on the map where there are few brands - prime spots to launch something new!

Sharpen your brand: Knowing how you're perceived lets you tweak your message to better attract your ideal customers.

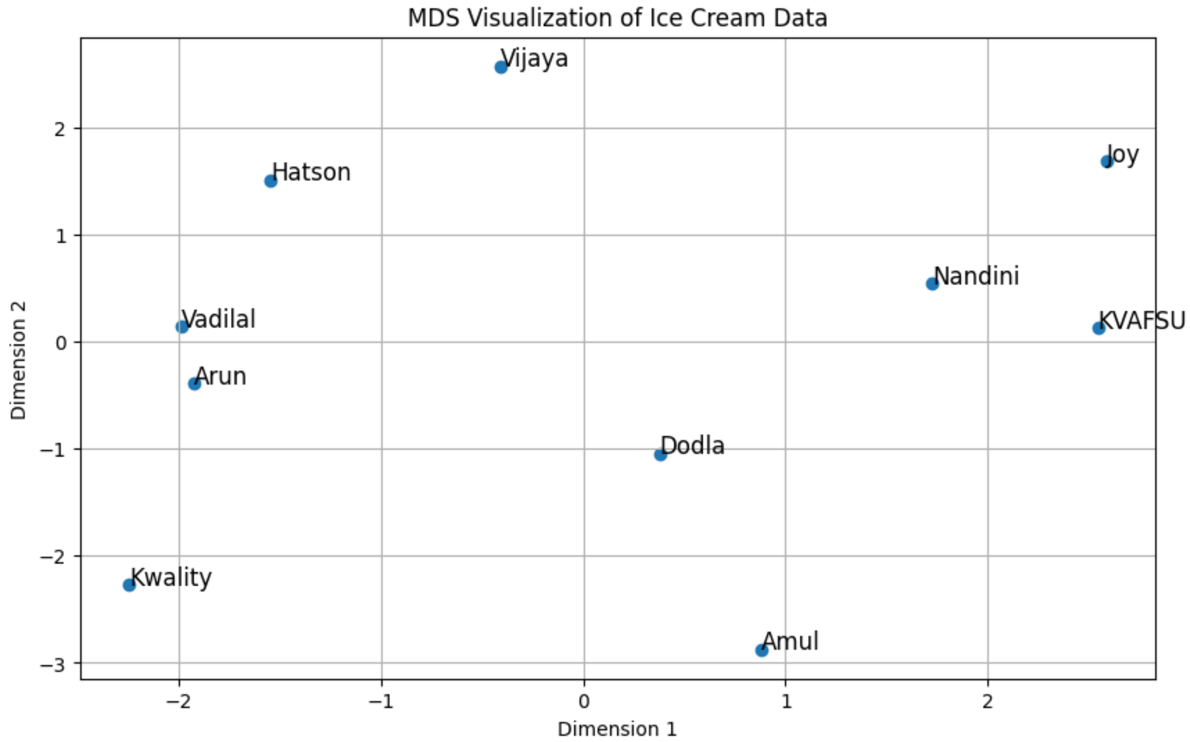
Stand out from the crowd: See where brands cluster and develop unique features to set your product apart.

Basically, MDS helps you understand the ice cream landscape so you can win over customers and keep them coming back for more scoops.

RESULTS

USING PYTHON

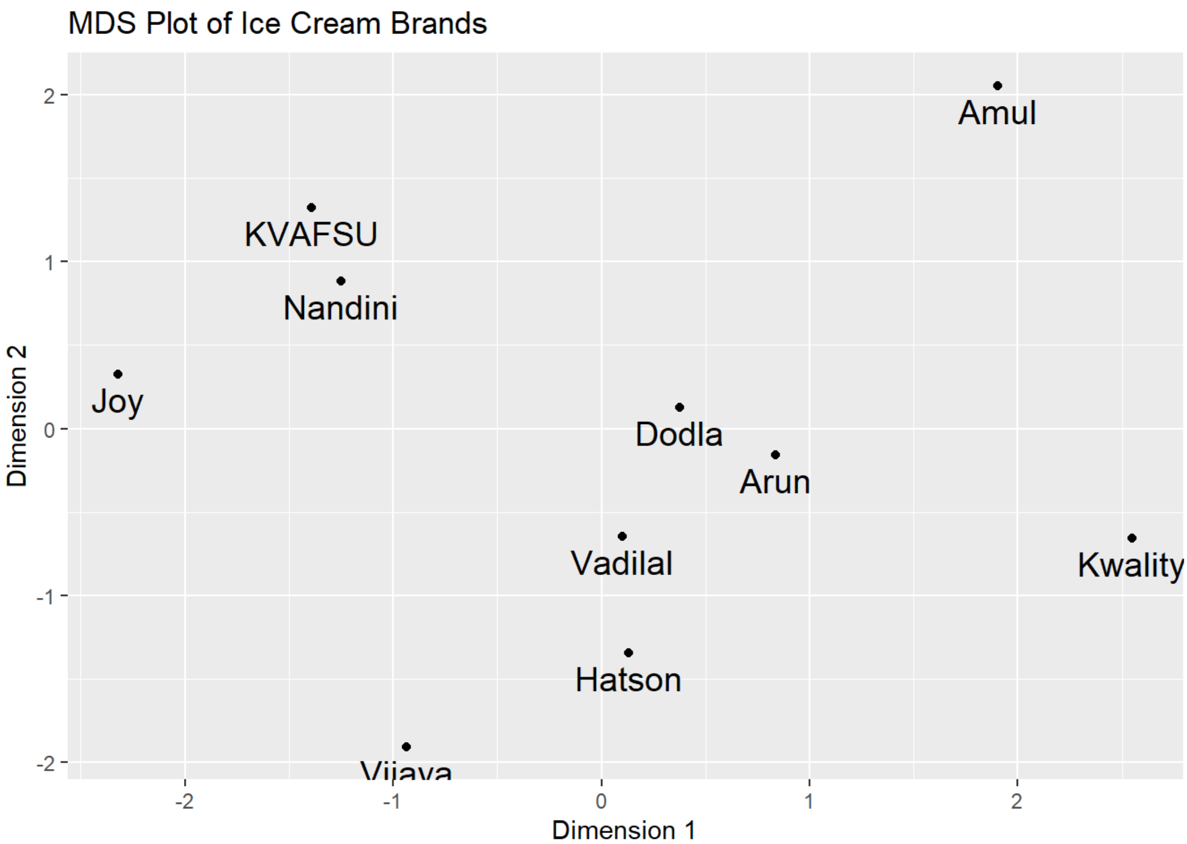
Multidimensional Scaling on Ice cream Dataset



USING R

RESULTS

Multidimensional Scaling on Ice cream Dataset



INTERPRETATIONS

Based on the MDS plot of the ice cream brands, here's what we can interpret about the relative positioning of the brands:

Corner (High Price, Exterior Look): Amul appears to be positioned in the corner with a potentially higher price point and focus on exterior look (e.g., packaging, design). Based on the image, Amul likely appears in the corner with a potentially higher price point (**Price** = 4) and focus on exterior look (**Flavour** = 3, **Consisten** = 4, **Shelflife** = 3). These could be interpreted as factors influencing exterior look perception.

Value and Availability: Several brands (Nandini, Joy, Dodla) appear to be clustered together in a region that might represent lower prices or value for the customer. There might also be an association with availability in this area of the plot.

Urban Center: Hatson seems to be positioned somewhat separately, potentially indicating a perception of being centered around urban areas.

Higher Priced with Varied Flavors: Brands like Amul, Kwalitys, and Arun appear positioned in the upper right quadrant, suggesting a potential association with higher prices and a wider variety of flavors (as indicated by higher Flavour scores).

Lower Priced with Fewer Flavors: Brands like Nandini, Vijaya, Dodla, and Hatson appear concentrated in the lower left quadrant, suggesting a possible link to lower prices and a more limited flavor selection (as indicated by lower Flavour scores).

Joy: The brand Joy seems to be somewhat separate, positioned lower in price but with a flavor score that's not the lowest. This could indicate it offers a balance between affordability and flavor variety.

RECOMMENDATIONS

Identify Target Market: Conduct further market research to understand which customer segments prioritize flavor variety and are willing to pay more, versus those who prioritize affordability.

Targeted Marketing: Develop marketing campaigns that resonate with your target segments. For instance, highlight the wider flavor options for brands positioned like Amul, while emphasizing value for those like Nandini.

Product Differentiation: Use the insights you have from the text data (price and availability) to identify areas for differentiation. If several brands cluster in a particular region, consider how your brand can stand out in that space. This could involve unique flavors, innovative packaging, or emphasizing different aspects of customer service or experience.

Part 4:
Conjoint Analysis

INTRODUCTION

Conjoint Analysis steps up to the plate to help us understand what truly makes a pizza a customer's favorite. It's a powerful statistical technique that goes beyond just average preferences. By showing customers a variety of hypothetical pizzas with different attributes (crust type, toppings, price), Conjoint Analysis analyzes their choices to uncover which features they value most. This allows pizza makers to identify the perfect combination of crust, cheese, and toppings that will win over customer hearts (and stomachs).

OBJECTIVES

Identify Key Attributes: pizzas come in many forms. Conjoint Analysis helps us identify the key attributes that truly influence customer preferences. This could include factors like crust type (thin, deep dish), cheese options (mozzarella, goat cheese), toppings (pepperoni, veggies), and even price. By analyzing customer choices on hypothetical pizzas with varying attributes, Conjoint Analysis reveals which features matter most to them.

Quantify Importance: Conjoint Analysis goes beyond simply identifying important attributes. It allows us to quantify the relative importance of each one. Imagine crust being twice as important as cheese for a specific customer segment. Conjoint Analysis helps us measure these importance levels, providing valuable insights for businesses.

BUSINESS SIGNIFICANCE

Understanding customer preferences for pizza attributes leads to:

Increased Customer Satisfaction: Products designed to meet specific customer needs lead to happier customers and a loyal customer base.

Higher Sales: Pizzas that resonate with customer preferences are more likely to be purchased, leading to increased sales.

Data-Driven Decisions: Quantifiable data empowers businesses to make informed decisions about products, pricing, and marketing.

Competitive Advantage: Identifying unique selling points based on customer preferences allows businesses to stand out from the competition.

RESULTS

USING PYTHON:

	coef	std err	t	P>|t|	[0.025	0.975]
const	18.0000	0.685	26.276	0.000	16.380	19.620
brand	0.0500	0.160	0.312	0.764	-0.329	0.429
price	-0.4500	0.160	-2.806	0.026	-0.829	-0.071
weight	-3.5500	0.160	-22.138	0.000	-3.929	-3.171
crust	-3.5000	0.359	-9.761	0.000	-4.348	-2.652
cheese	0.5000	0.359	1.394	0.206	-0.348	1.348
size	0.5000	0.359	1.394	0.206	-0.348	1.348
toppings	-2.2500	0.359	-6.275	0.000	-3.098	-1.402
spicy	-1.5000	0.359	-4.183	0.004	-2.348	-0.652

Omnibus:	*6.427*	*Durbin-Watson:*	*1.388*
Prob(Omnibus):	*0.040*	*Jarque-Bera (JB):*	*1.576*
Skew:	*0.009*	*Prob(JB):*	*0.455*
Kurtosis:	*1.463*	*Cond. No.*	*15.3*

From the given regression output, we can interpret the results of the Conjoint Analysis on the pizza dataset. The coefficients (coef) represent the utility values for each attribute, and the statistical significance (P>|t|) helps us understand the importance and relevance of each attribute in influencing the overall ranking (preference).

Summary of Results

Brand: The coefficient is 0.0500 with a p-value of 0.764, indicating that brand is not statistically significant in affecting the preference scores.

Price: The coefficient is -0.4500 with a p-value of 0.026, indicating that price is statistically significant. As the price increases, the preference decreases.

Weight: The coefficient is -3.5500 with a p-value of 0.000, indicating that weight is highly significant. Heavier pizzas are associated with lower preference scores.

Crust: The coefficient is -3.5000 with a p-value of 0.000, indicating that crust type is highly significant. Thick crusts are less preferred compared to thin crusts.

Cheese: The coefficient is 0.5000 with a p-value of 0.206, indicating that cheese type is not statistically significant.

Size: The coefficient is 0.5000 with a p-value of 0.206, indicating that size is not statistically significant.

Toppings: The coefficient is -2.2500 with a p-value of 0.000, indicating that toppings are highly significant. Paneer is less preferred compared to mushroom.

Spicy: The coefficient is -1.5000 with a p-value of 0.004, indicating that spiciness level is significant. Normal spiciness is preferred over extra spicy.

Interpretation of Key Attributes

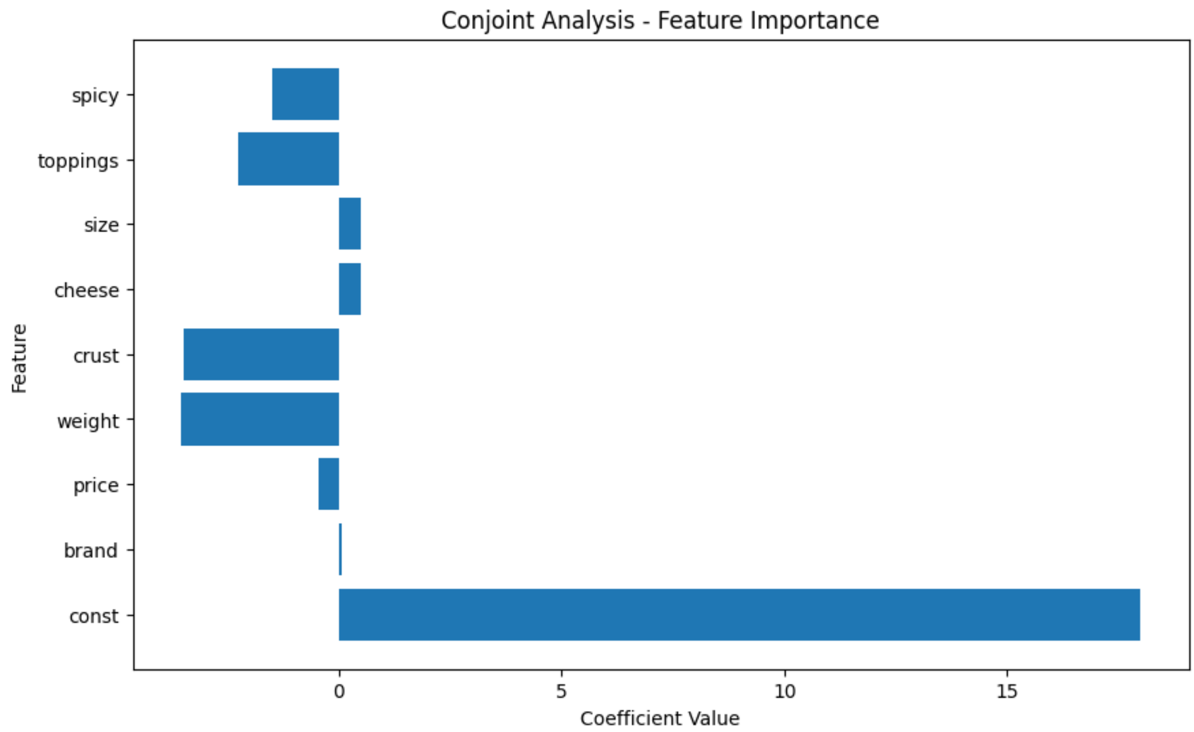
Price: As price increases, preference decreases significantly. Customers prefer cheaper pizzas.

Weight: Heavier pizzas are significantly less preferred, indicating a preference for lighter pizzas.

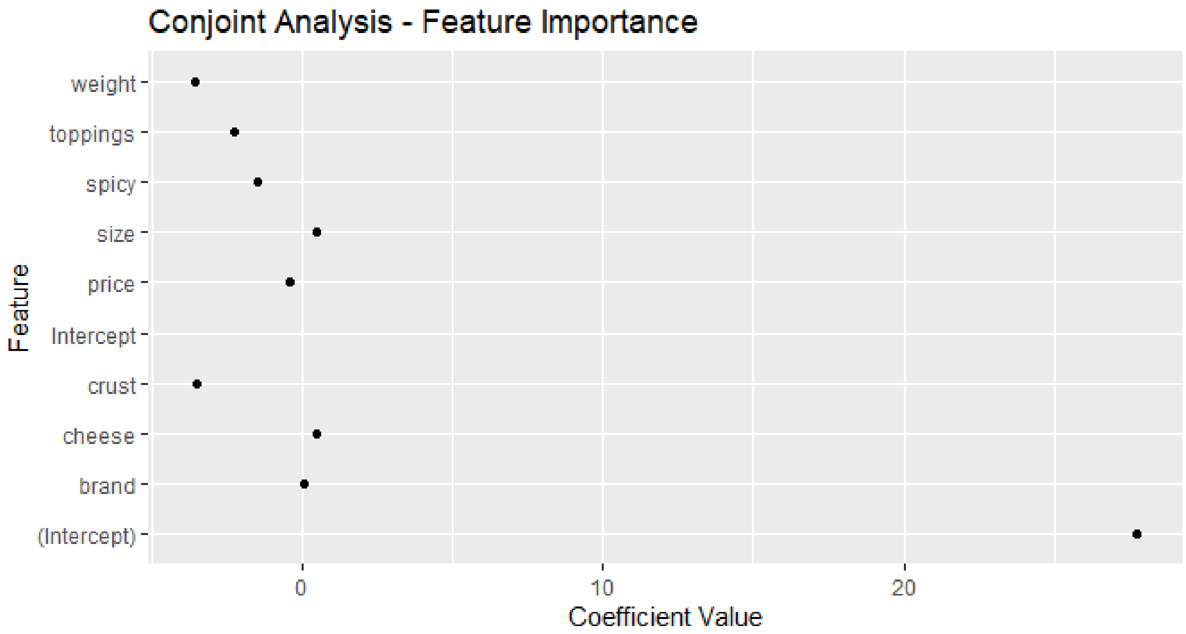
Crust: Thick crusts are significantly less preferred compared to thin crusts.

Toppings: Paneer as a topping is significantly less preferred compared to mushroom.

Spicy: Extra spicy pizzas are less preferred compared to normal spiciness.



USING R



By combining conjoint analysis with customer research, pizza businesses can gain a deeper understanding of how different attributes influence price and customer preferences. This knowledge can be used to develop targeted pricing strategies, promotional campaigns, and menus that resonate with their target audience.

Recommendations

Focus on Price Sensitivity: Since price is a significant factor, pizza brands should consider pricing strategies carefully. Offering competitive pricing or value deals can attract more customers.

Optimize Pizza Weight: Given that heavier pizzas are less preferred, brands should consider optimizing pizza weight to balance portion size and customer preferences.

Preference for Thin Crust: Brands should focus on offering more thin crust options, as they are significantly preferred over thick crusts.

Toppings Strategy: Since mushroom toppings are preferred over paneer, brands should consider promoting pizzas with mushroom or offering a variety of popular toppings.

Spiciness Level: Maintain a focus on normal spiciness levels, as extra spicy pizzas are less preferred.

By aligning product offerings with these customer preferences, pizza brands can enhance customer satisfaction and increase their market share.

FINAL CONCLUSION

PART 1: Principal Component Analysis (PCA) and Factor Analysis (Survey.csv)

Objective: To identify the underlying data dimensions in the survey data.

Business Significance: Understanding the key factors that influence customer preferences and behaviors can help businesses tailor their products and services more effectively.

Conclusion: The PCA and Factor Analysis revealed the main dimensions that capture the variability in the survey responses. These dimensions can be used to segment the market and develop targeted marketing strategies, improving customer satisfaction and loyalty.

PART 2: Cluster Analysis (Survey.csv)

Objective: To characterize respondents based on background variables.

Business Significance: Clustering respondents into distinct groups allows businesses to identify unique segments and customize their offerings to meet the specific needs of each segment.

Conclusion: The Cluster Analysis identified distinct groups of respondents with similar characteristics. This segmentation can be used to design personalized marketing campaigns, product offerings, and customer service strategies, thereby enhancing the overall customer experience and business performance.

PART 3: Multidimensional Scaling (MDS) (icecream.csv)

Objective: To visualize and interpret the similarities and preferences among different ice cream brands.

Business Significance: MDS helps in understanding how customers perceive various brands relative to each other, which is crucial for positioning and competitive analysis.

Conclusion: The MDS analysis provided a visual representation of the relative positioning of different ice cream brands based on customer perceptions. This information can be used to identify competitive strengths and weaknesses, inform branding strategies, and enhance market positioning.

PART 4: Conjoint Analysis (pizza_data.csv)

Objective: To determine the relative importance of different attributes in influencing customer preferences for pizzas.

Business Significance: Conjoint Analysis helps in understanding the trade-offs customers make and the value they place on different product attributes, which is vital for product development and pricing strategies.

Conclusion: The Conjoint Analysis revealed that price, weight, crust type, toppings, and spiciness level significantly influence customer preferences for pizzas. Brands should focus on offering competitive pricing, optimizing pizza weight, and providing preferred crust types and toppings. Normal spiciness levels are preferred over extra spicy. Addressing these preferences can enhance customer satisfaction and drive sales.

My Github profile link

https://github.com/micahvarkyez