****

**VIRGINIA COMMONWEALTH UNIVERSITY**

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

**A4: Multivariate Analysis and Business Analytics Applications**

**Lakshmy Cherussery Jayaprakash**

**V01110023**

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

**CONTENTS**

|  |  |  |
| --- | --- | --- |
| **Sl. No.** | **Title** | **Page No.** |
| **1.** | Introduction | **1** |
| **2.** | Objective | **1** |
| **3.** | PART 1-Principal Component Analysis & Factor Analysis  Introduction  Objective  Results & Interpretations | **2-18** |
| **4.** | PART 2- Cluster Analysis  Introduction  Objective  Results & Interpretations | **19-23** |
| **5.** | PART 3-Multidimensional Scaling  Introduction  Objective  Results & Interpretations | **24-25** |
| **6.** | PART 4- Conjoint Analysis  Introduction  Objective  Results & Interpretations | **26-27** |
| **7.** | Business Significance | **28** |

**Introduction**

Factor Analysis (FA) is a statistical technique used to identify patterns among observed variables and reduce them to a smaller number of unobserved variables called factors.

Principal Component Analysis (PCA) is a dimensionality reduction technique that transforms a large set of variables into a smaller set of orthogonal (uncorrelated) components.

Multidimensional Scaling (MDS) is a technique used to visualize the relative similarity or dissimilarity of individual cases in a dataset.

Conjoint Analysis is a statistical technique used to understand how consumers make trade-offs between different attributes of a product or service.

**Objective**

The objective of Factor Analysis in the survey dataset is to uncover latent factors (like preferences or decision-making criteria) that explain the correlations among multiple observed variables related to housing preferences.

The objective of PCA on the survey dataset was to identify patterns and underlying structures within the data, potentially reducing the dimensionality for easier interpretation.

The objective of MDS on the ice cream dataset was to visualize how different ice cream brands are perceived based on multiple attributes (e.g., taste, price, availability).

In the pizza dataset, Conjoint Analysis aimed to determine the relative importance of various pizza attributes (like price, crust type, cheese amount) and their impact on consumer preferences.

**PART 1: Principal Component Analysis & Factor Analysis**

**INTRODUCTION**

Here, we conduct a comprehensive analysis using Factor Analysis (FA) and Principal Component Analysis (PCA) on a dataset “Survey.csv” related to housing preferences. The dataset contains various factors influencing housing decisions, such as proximity to amenities, security, pricing, and neighborhood characteristics.

**OBJECTIVE**

1. **Dimensionality Reduction:**
   * Utilize Principal Component Analysis (PCA) to reduce the number of variables while retaining as much variance as possible.
   * By identifying principal components, we aimed to simplify the dataset without losing critical information, thereby facilitating easier interpretation and analysis.
2. **Factor Analysis:**
   * Employ Factor Analysis (FA) to uncover latent variables or factors that underlie observed variables related to housing preferences.
   * By examining loadings, we sought to understand which variables cluster together and contribute most significantly to each factor.

**RESULTS & INTERPRETATIONS**

**R**

**1.PRINCIPAL COMPONENT ANALYSIS**



> pca

Principal Components Analysis

Call: principal(r = sur\_int, nfactors = 5, rotate = "promax", n.obs = 162)

Standardized loadings (pattern matrix) based upon correlation matrix

RC1 RC3 RC4 RC2 RC5 h2 u2 com

X1.Proximity.to.city 0.10 0.71 -0.26 0.18 0.14 0.65 0.35 1.5

X2.Proximity.to.schools -0.08 0.57 0.24 -0.11 0.20 0.51 0.49 1.8

X3..Proximity.to.transport -0.17 -0.26 0.71 0.12 0.01 0.51 0.49 1.5

X4..Proximity.to.work.place -0.38 0.98 0.03 0.07 -0.10 0.73 0.27 1.3

X5..Proximity.to.shopping 0.78 0.12 -0.01 0.27 -0.22 0.69 0.31 1.5

X1..Gym.Pool.Sports.facility 0.42 0.20 0.30 -0.08 -0.19 0.46 0.54 2.9

X2..Parking.space 0.53 0.19 0.04 -0.18 -0.06 0.47 0.53 1.6

X3.Power.back.up 0.35 0.53 0.05 0.09 -0.43 0.59 0.41 2.8

X4.Water.supply 0.21 0.07 0.75 0.01 0.07 0.73 0.27 1.2

X5.Security 0.76 -0.28 0.34 -0.11 -0.22 0.68 0.32 1.9

X1..Exterior.look 0.85 -0.08 -0.32 0.20 0.09 0.76 0.24 1.5

X2..Unit.size -0.06 -0.12 -0.04 -0.13 0.74 0.52 0.48 1.1

X3..Interior.design.and.branded.components 0.60 0.22 0.01 -0.07 0.18 0.66 0.34 1.5

X4..Layout.plan..Integrated.etc.. 0.34 0.39 -0.05 -0.05 0.32 0.60 0.40 3.0

X5..View.from.apartment 0.84 -0.08 0.00 -0.07 0.13 0.74 0.26 1.1

X1..Price -0.13 0.16 0.66 0.13 0.32 0.53 0.47 1.8

X2..Booking.amount 0.22 -0.11 -0.08 0.66 0.04 0.51 0.49 1.3

X3..Equated.Monthly.Instalment..EMI. -0.09 -0.06 0.44 0.73 -0.10 0.58 0.42 1.8

X4..Maintenance.charges 0.05 -0.14 0.00 0.52 0.01 0.30 0.70 1.2

X5..Availability.of.loan -0.15 0.30 -0.01 0.83 -0.14 0.71 0.29 1.4

X1..Builder.reputation 0.08 0.08 0.34 -0.05 0.68 0.69 0.31 1.5

X2..Appreciation.potential 0.23 -0.02 0.13 0.43 0.41 0.44 0.56 2.7

X3..Profile.of.neighbourhood 0.58 -0.13 0.35 -0.16 0.26 0.68 0.32 2.4

X4..Availability.of.domestic.help 0.94 -0.40 -0.09 -0.05 0.02 0.69 0.31 1.4

RC1 RC3 RC4 RC2 RC5

SS loadings 5.17 2.58 2.39 2.32 1.96

Proportion Var 0.22 0.11 0.10 0.10 0.08

Cumulative Var 0.22 0.32 0.42 0.52 0.60

Proportion Explained 0.36 0.18 0.17 0.16 0.14

Cumulative Proportion 0.36 0.54 0.70 0.86 1.00

With component correlations of

RC1 RC3 RC4 RC2 RC5

RC1 1.00 0.47 0.22 -0.06 0.27

RC3 0.47 1.00 0.18 -0.15 0.21

RC4 0.22 0.18 1.00 -0.27 0.01

RC2 -0.06 -0.15 -0.27 1.00 0.00

RC5 0.27 0.21 0.01 0.00 1.00

Mean item complexity = 1.7

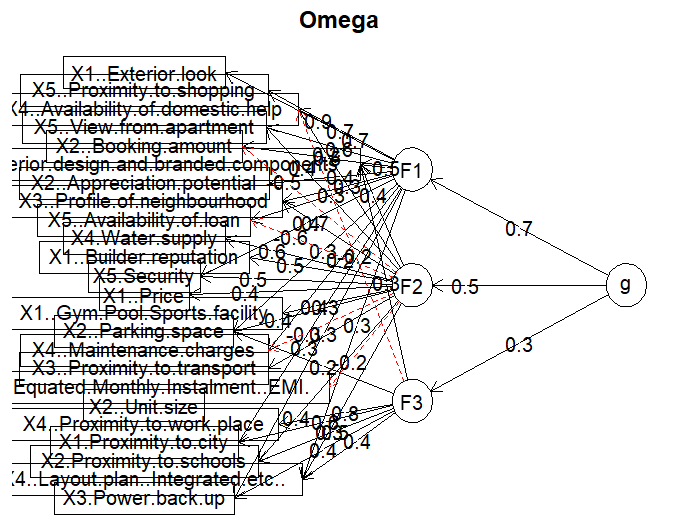
Test of the hypothesis that 5 components are sufficient.

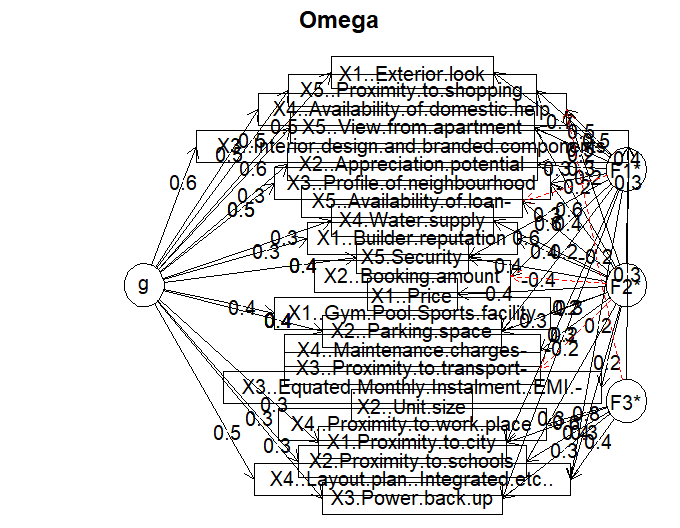
The root mean square of the residuals (RMSR) is 0.07

with the empirical chi square 206.42 with prob < 0.018

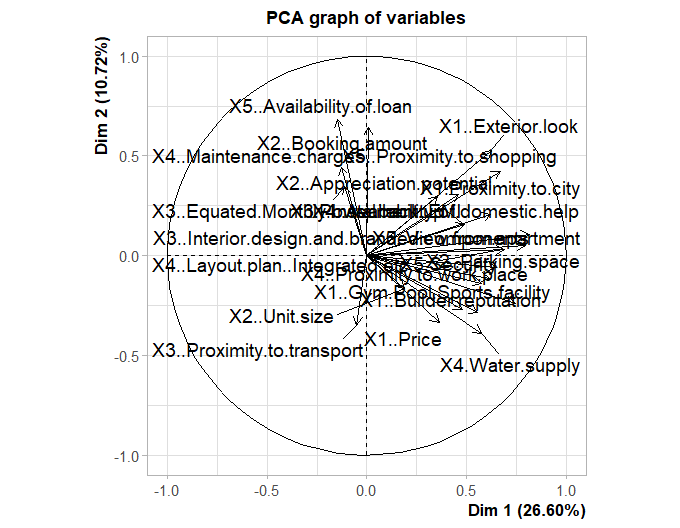
Fit based upon off diagonal values = 0.93

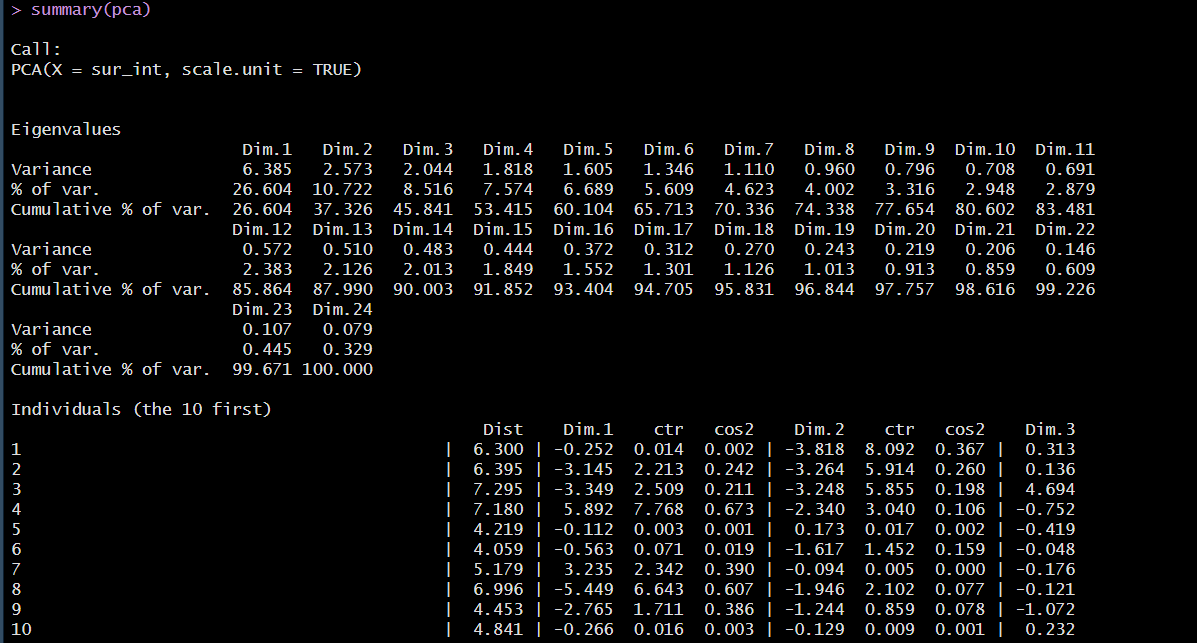
The values in the matrix indicate the loading of each variable on each principal component. Higher absolute values indicate a stronger association between the variable and the component.



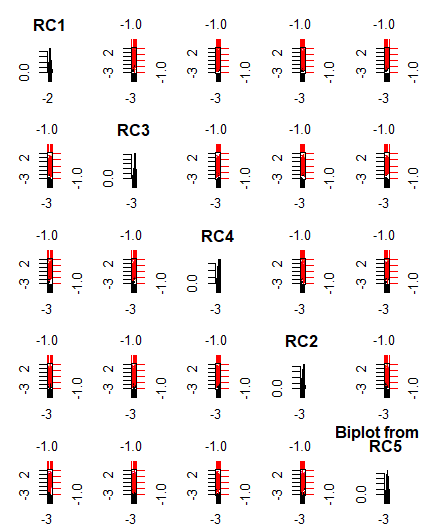
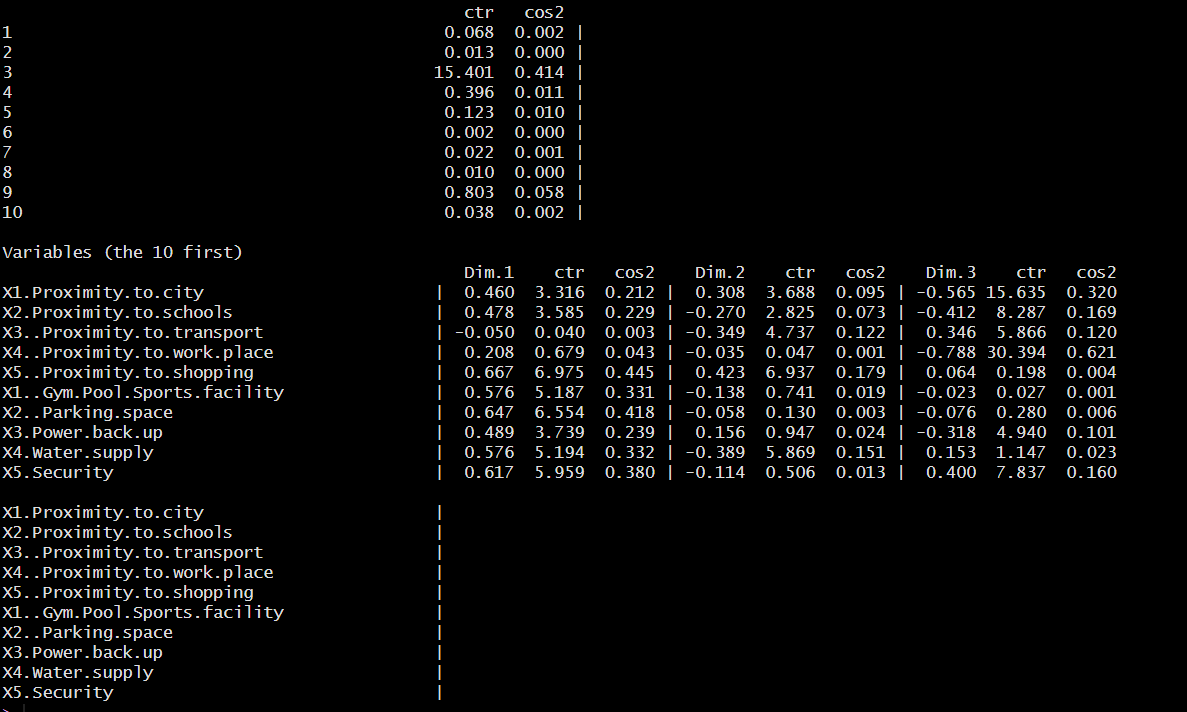


This function(omega) is particularly useful in the context of factor analysis to understand the structure of the data in terms of general and specific factors.





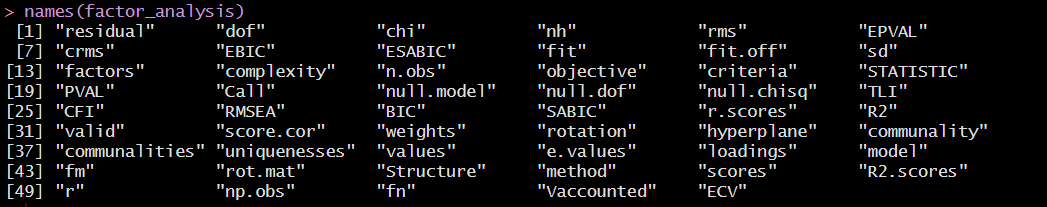
biplot(pca, scale = 0)

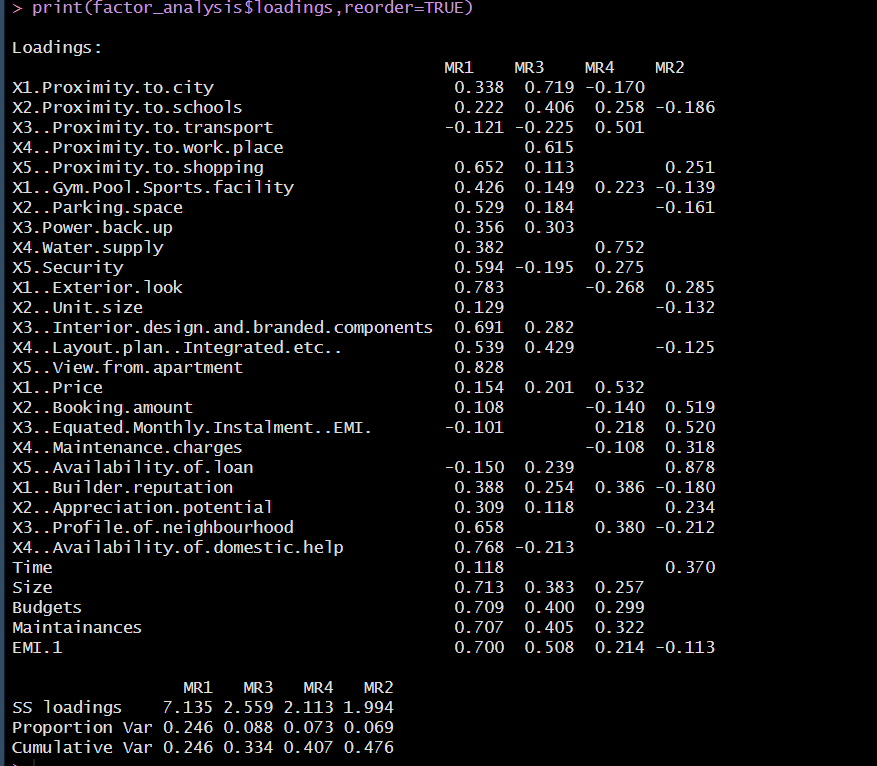


**Interpretation:**

* **First Component (RC1):**
  + High loadings on variables related to proximity to shopping, security, exterior look, view from the apartment, and profile of the neighborhood. This suggests that RC1 captures aspects related to location convenience and aesthetic appeal.
* **Second Component (RC2):**
  + High loadings on variables related to booking amount, EMI, availability of loans, and maintenance charges. This component appears to capture financial aspects and affordability.
* **Third Component (RC3):**
  + High loadings on variables related to proximity to the city and schools. This component captures aspects related to accessibility and proximity to essential services.
* **Fourth Component (RC4):**
  + High loadings on proximity to transport, water supply, and price. This component captures aspects related to basic amenities and cost.
* **Fifth Component (RC5):**
  + High loadings on unit size and builder reputation. This component captures aspects related to the size and reputation of the property.

**2.FACTOR ANALYSIS**

****

****

The loadings indicate how much each variable contributes to the identified factors (MR1, MR2, MR3, MR4). The factors can be interpreted based on the variables that load highly on them.

* **MR1 (Factor 1):**

This factor seems to capture aspects related to the aesthetic appeal and comprehensive features of the property, including proximity to shopping, design, neighborhood profile, and overall size and budget considerations.

* **MR2 (Factor 2):**

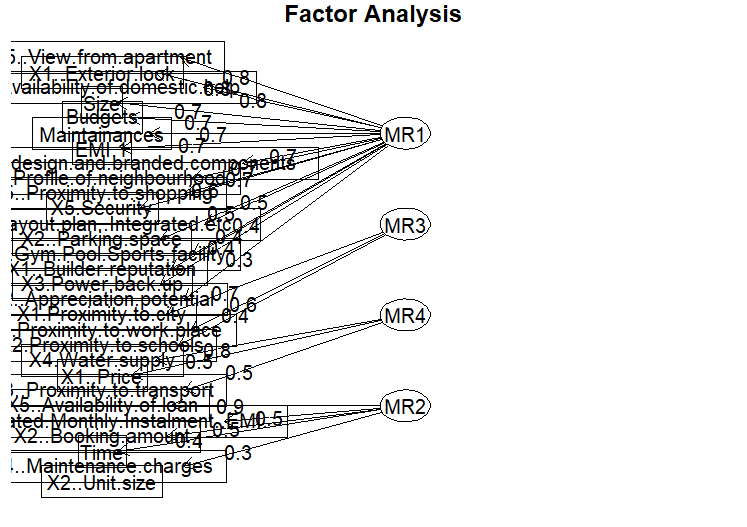
This factor captures financial and logistical aspects related to acquiring the property, such as loan availability, booking amount, and EMI.

* **MR3 (Factor 3):**

This factor seems to capture accessibility aspects such as proximity to city, schools, shopping, and availability of amenities like parking and power backup.

* **MR4 (Factor 4):**

This factor appears to capture essential services and affordability aspects, such as proximity to transport, water supply, and price considerations.



print(factor\_analysis$communality)

X1.Proximity.to.city X2.Proximity.to.schools

0.66641636 0.31505092

X3..Proximity.to.transport X4..Proximity.to.work.place

0.32255808 0.38704235

X5..Proximity.to.shopping X1..Gym.Pool.Sports.facility

0.50168188 0.27313215

X2..Parking.space X3.Power.back.up

0.34504278 0.21911638

X4.Water.supply X5.Security

0.72104641 0.47189943

X1..Exterior.look X2..Unit.size

0.77466504 0.04040037

X3..Interior.design.and.branded.components X4..Layout.plan..Integrated.etc..

0.57000009 0.49408411

X5..View.from.apartment X1..Price

0.69182220 0.35455614

X2..Booking.amount X3..Equated.Monthly.Instalment..EMI.

0.30237766 0.33109089

X4..Maintenance.charges X5..Availability.of.loan

0.12587052 0.85702785

X1..Builder.reputation X2..Appreciation.potential

0.39708855 0.17201887

X3..Profile.of.neighbourhood X4..Availability.of.domestic.help

0.62294275 0.63983787

Time Size

0.15522805 0.72147280

Budgets Maintainances

0.75210691 0.76752409

EMI.1

0.80701346

> print(factor\_analysis$scores)

MR1 MR3 MR4 MR2

[1,] -0.5876607 -0.11251823 2.050119670 -0.80909066

[2,] -1.4792705 -0.36503965 0.775643390 -1.16305313

[3,] -0.5478608 -3.35305405 0.945557247 -1.32923887

[4,] 1.8807274 0.13402370 0.266228876 -2.06159338

[5,] 0.3295689 -0.18023376 -1.170876218 -0.36010367

[6,] -0.2866020 -0.47379445 0.367382156 -0.21010883

[7,] 0.5003641 0.65274995 1.058454876 0.43822528

[8,] -1.9088343 -0.71638354 0.125781819 -1.29105206

[9,] -0.7069313 0.10856075 -0.040407655 0.37696557

[10,] 0.1573651 0.97898633 -0.458207894 0.32151704

[11,] -0.3560346 -0.87215158 -0.548893781 0.20034247

[12,] -1.0750688 1.06862183 -1.324450733 -0.87149108

[13,] -1.2632780 0.08817647 0.497275087 0.87321384

[14,] -0.1478539 0.79809053 0.072043274 0.84550179

[15,] -0.8943020 -0.57613016 0.146791486 -0.06867479

[16,] -2.3861008 -0.26161531 -0.224930358 0.28216534

[17,] -1.6756359 0.98678140 -0.553005149 -0.12911862

[18,] -1.3558992 1.21940046 0.959335531 -0.38269031

[19,] 0.3070668 1.15207947 0.709196598 -1.63990272

[20,] 0.6265597 -0.78100996 0.357780784 -2.31140305

[21,] -1.3558992 1.21940046 0.959335531 -0.38269031

[22,] -0.4194756 1.30069137 0.503442662 0.91390513

[23,] -1.1801257 1.51209556 1.086366589 -0.54903222

[24,] 0.5281885 -0.61791805 -0.458074607 0.95748245

[25,] 1.5544448 0.47387769 0.923143557 0.02046889

[26,] -0.6560738 1.28545333 -0.217972896 0.05690345

[27,] -0.5592703 0.09090353 -0.895706580 1.14485399

[28,] -0.2026950 0.43630856 0.148534968 0.28641784

[29,] -1.3721154 0.67538936 0.895056459 -0.08885140

[30,] 1.0626369 -0.05059363 -0.642613792 -0.85437235

[31,] 0.8055042 0.64353368 0.081147945 -0.65568444

[32,] -0.6404405 -0.24517693 -1.178682910 -0.09102854

[33,] -0.5394157 -0.09824719 0.202530097 1.22865673

[34,] 0.8810889 -0.47378272 0.659986298 0.70591671

[35,] -1.4194026 -0.95544542 0.515394129 0.77098383

[36,] 0.4944889 -0.80617264 0.926183237 -0.84615059

[37,] -0.3920544 0.11286019 -0.185478395 1.25609547

[38,] 0.9187362 0.37784351 -2.021924632 0.03073602

[39,] -0.2688068 0.43165531 0.135943047 0.71033819

[40,] 0.3238506 -1.31842216 -0.956078235 -0.04785917

[41,] -0.6013308 0.39092682 -1.170649492 -0.39382993

[42,] 0.7672588 -0.97998184 1.065783179 1.38843255

[43,] 1.2268022 0.92424961 -0.506390529 -1.74468234

[44,] 0.1531872 -0.77891730 0.598245535 1.48113689

[45,] 0.0310595 -1.91707747 -0.829528716 1.67981703

[46,] 1.5259121 0.37706979 -0.106363781 0.67027212

[47,] 0.1309902 -0.37026373 -2.097303080 -0.58011507

[48,] 0.6094975 -0.49059484 -1.347669950 0.20423233

[49,] 1.7828850 1.41737185 0.926221436 1.50375524

[50,] -1.3382877 -0.75161725 -2.343753911 -1.08189200

[51,] -0.4664299 -2.63297854 0.768856976 -0.96279188

[52,] 1.6302221 0.38795119 0.696176766 -0.46016474

[53,] 0.3446095 -0.26406859 0.785862102 1.91000860

[54,] 1.3756285 0.60316568 0.936286214 -0.57622139

[55,] 0.7398498 0.67345244 0.169092222 0.57967423

[56,] 0.3269372 0.06586277 -1.542789689 -0.16210359

[57,] -0.1855857 0.17245157 0.001009407 1.08219628

[58,] 1.8807274 0.13402370 0.266228876 -2.06159338

[59,] 0.9987618 1.15565927 0.692293241 1.45269523

[60,] 0.2569674 -0.23430610 0.186779172 1.47210427

[61,] 0.5694039 -0.35107186 1.166005563 -0.37318649

[62,] 0.8810067 -0.59364313 -0.707466321 -0.20736685

[63,] 0.9555754 0.24108053 -1.713145455 0.32223769

[64,] 0.2601443 -1.07529756 0.502800902 -1.87754200

[65,] -0.8113446 0.98765809 0.277402326 -0.20207944

[66,] -0.1096387 -1.38076313 0.216101775 1.88705736

[67,] 0.8768180 -0.28471376 -1.360191871 0.28202914

[68,] -0.6013308 0.39092682 -1.170649492 -0.39382993

[69,] -0.4169898 0.74655193 -0.353765544 -0.60564613

[70,] 0.5132101 -0.05290097 1.503170655 0.48989636

**Interpretation of Proportions:**

* **MR1** explains the largest portion of the variance (24.6%), indicating that factors related to aesthetic appeal and comprehensive property features are the most significant in this survey.
* **MR2**, **MR3**, and **MR4** explain smaller proportions of the variance but still contribute important dimensions related to financial/logistical aspects, accessibility, and essential services/affordability, respectively.

**PYTHON**

**1.PCA**

Explained variance ratio: [0.31822766 0.09526344 0.06889323 0.06341429 0.06111318]

Cumulative explained variance: [0.31822766 0.4134911 0.48238432 0.54579862 0.6069118 ]

**Explained Variance Ratio**:

* This is a measure of how much variance each principal component explains from the total variance in the dataset.

For example,

* + PC1 explains approximately 31.82% of the total variance in the data.
  + PC2 explains an additional 9.53%.

Together, these ratios provide insight into how much information each principal component carries. A higher ratio indicates that the principal component explains a larger portion of the variance in the original data.

**Cumulative Explained Variance**:

* This shows the cumulative proportion of variance explained by the principal components.
* For example:
  + The first two principal components together explain approximately 41.35% (31.82% + 9.53%) of the total variance.
  + The first three explain 48.24%, and so forth.

This cumulative measure helps determine how many principal components are needed to explain a significant portion of the data's variance. It's useful for deciding how many principal components to retain in the analysis.

PCA Components:

3. Proximity to transport 4. Proximity to work place \

0 0.014928 -0.052653

1 0.126411 0.060153

2 -0.465915 0.052699

3 0.069723 -0.592034

4 0.271472 -0.097866

5. Proximity to shopping 1. Gym/Pool/Sports facility 2. Parking space \

0 -0.210212 -0.180419 -0.203748

1 -0.270909 0.089917 0.037004

2 0.129612 0.015402 0.146867

3 -0.053158 -0.081124 -0.056974

4 0.177992 0.292307 0.182483

3.Power back-up 4.Water supply 5.Security 1. Exterior look \

0 -0.155294 -0.203859 -0.196652 -0.219630

1 -0.076559 0.163703 0.022475 -0.316341

2 0.156561 -0.339980 -0.003218 0.177654

3 -0.426886 -0.075773 0.191186 0.138517

4 0.189252 0.209854 0.454456 -0.093192

2. Unit size ... 5. Availability of loan 1. Builder reputation \

0 -0.054351 ... 0.050746 -0.198477

1 0.088531 ... -0.465934 0.182091

2 -0.023657 ... -0.216143 -0.217037

3 0.330608 ... -0.221261 0.077041

4 -0.365054 ... -0.073568 -0.298374

2. Appreciation potential 3. Profile of neighbourhood \

0 -0.118254 -0.253555

1 -0.190878 0.132598

2 -0.199931 -0.080191

3 0.123135 0.155028

4 -0.148766 0.093081

4. Availability of domestic help Time Size Budgets \

0 -0.213143 -0.026919 -0.288181 -0.292770

1 -0.156897 -0.270334 -0.016512 0.016734

2 0.143303 -0.170545 -0.060315 -0.086750

3 0.325950 0.032492 -0.013361 -0.008785

4 0.142015 0.114240 -0.155017 -0.181846

Maintainances EMI.1

0 -0.294955 -0.294392

1 0.009794 0.046093

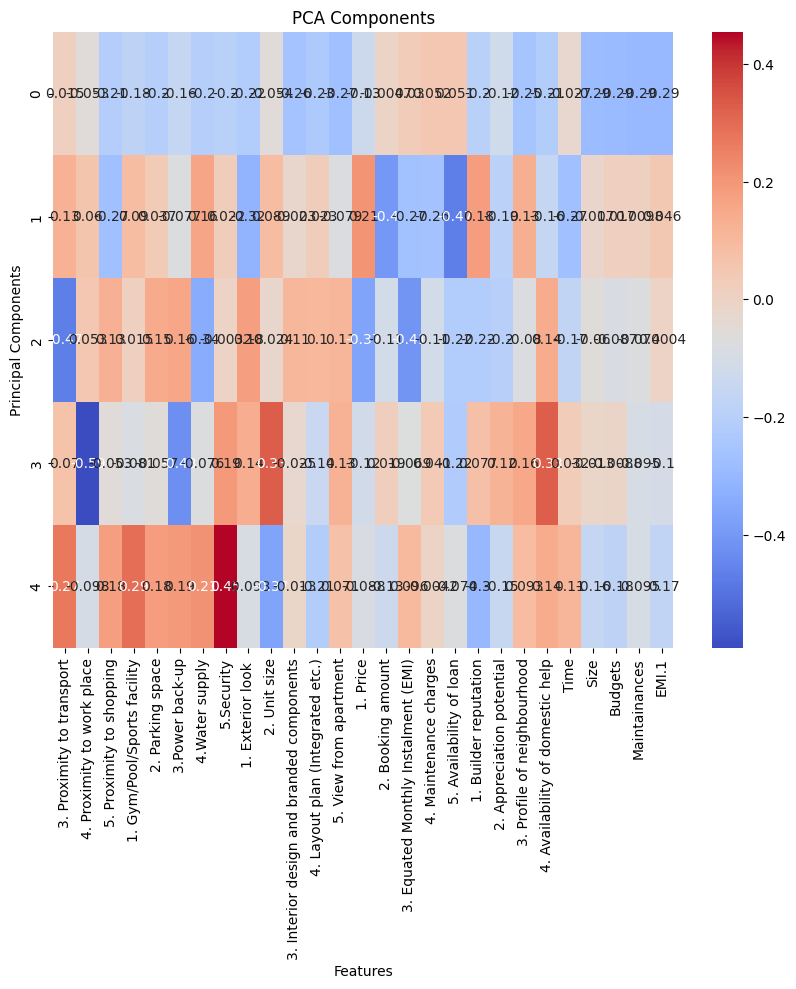
2 -0.074044 0.000399

3 -0.095202 -0.101849

4 -0.095451 -0.167535

[5 rows x 27 columns]

The PCA components matrix represents the contribution of each original variable to the principal components. Each row corresponds to a principal component, and each column represents a feature from the dataset. The values in the matrix indicate the weights (or loadings) of each feature in the corresponding principal component.



**2.FA**

Factor Variance:

(array([4.97928835, 3.6638373 , 1.99739684, 1.46935161, 1.31270193]), array([0.18441809, 0.13569768, 0.07397766, 0.05442043, 0.04861859]), array([0.18441809, 0.32011576, 0.39409343, 0.44851386, 0.49713245]))

**Eigenvalues (Array 1)**:

* These are the eigenvalues associated with each factor extracted from the factor analysis.
* Eigenvalues represent the amount of variance explained by each factor. Higher eigenvalues indicate that the factor explains more variance in the original variables.

**Proportion of Variance (Array 2)**:

* This array shows the proportion of total variance explained by each factor.
* Each value indicates the percentage of total variance in the original variables that is explained by the corresponding factor.

**Cumulative Proportion (Array 3)**:

* This array provides the cumulative proportion of variance explained as you move through the factors.
* It shows how much of the total variance is explained by the first factor, the first two factors together, the first three factors together, and so on.
* The cumulative proportion helps in deciding how many factors to retain for analysis. Typically, factors with eigenvalues greater than 1 are considered significant, but this criterion can vary depending on the context and specific goals of the analysis.

Factor Analysis Loadings:

0 1 2 \

3. Proximity to transport -0.062650 0.053449 0.102691

4. Proximity to work place 0.214466 -0.242324 -0.015612

5. Proximity to shopping 0.023969 0.646406 0.273362

1. Gym/Pool/Sports facility 0.035508 0.497716 -0.123070

2. Parking space 0.141555 0.458574 -0.152466

3.Power back-up 0.033851 0.324156 0.025881

4.Water supply 0.334492 0.295378 0.032303

5.Security -0.247295 0.963297 -0.083131

1. Exterior look 0.357562 0.413706 0.286870

2. Unit size 0.307274 -0.086644 -0.096613

3. Interior design and branded components 0.408981 0.385738 -0.046064

4. Layout plan (Integrated etc.) 0.574897 0.072914 -0.077908

5. View from apartment 0.312396 0.592523 -0.021110

1. Price 0.468356 -0.108705 -0.007293

2. Booking amount 0.055089 -0.005007 0.517348

3. Equated Monthly Instalment (EMI) -0.013245 -0.060950 0.550510

4. Maintenance charges -0.118321 -0.007978 0.300754

5. Availability of loan 0.069242 -0.275819 0.909565

1. Builder reputation 0.770507 -0.151793 -0.101498

2. Appreciation potential 0.333179 0.056012 0.270933

3. Profile of neighbourhood 0.363382 0.482280 -0.173209

4. Availability of domestic help 0.066108 0.733663 0.046809

Time 0.029564 0.083021 0.374954

Size 0.809970 0.100932 0.088567

Budgets 0.914751 0.019184 0.068697

Maintainances 0.792661 0.116358 0.073753

EMI.1 0.843329 0.050710 -0.039784

3 4

3. Proximity to transport 0.566766 -0.108953

4. Proximity to work place -0.059643 0.657572

5. Proximity to shopping -0.118425 0.170156

1. Gym/Pool/Sports facility 0.175301 0.143681

2. Parking space -0.034074 0.144277

3.Power back-up -0.100512 0.558963

4.Water supply 0.617495 0.106283

5.Security 0.319567 -0.138278

1. Exterior look -0.404235 -0.126286

2. Unit size -0.039999 -0.298122

3. Interior design and branded components -0.109582 0.103988

4. Layout plan (Integrated etc.) -0.175612 0.204097

5. View from apartment -0.110427 -0.077293

1. Price 0.403942 0.077952

2. Booking amount -0.102740 -0.032806

3. Equated Monthly Instalment (EMI) 0.308988 0.040785

4. Maintenance charges -0.006142 -0.032559

5. Availability of loan -0.005900 0.293333

1. Builder reputation 0.180912 -0.127029

2. Appreciation potential 0.044642 -0.101233

3. Profile of neighbourhood 0.240515 -0.148348

4. Availability of domestic help -0.101157 -0.323017

Time 0.069180 -0.028749

Size 0.004628 0.025844

Budgets 0.028929 -0.000888

Maintainances 0.063025 0.133921

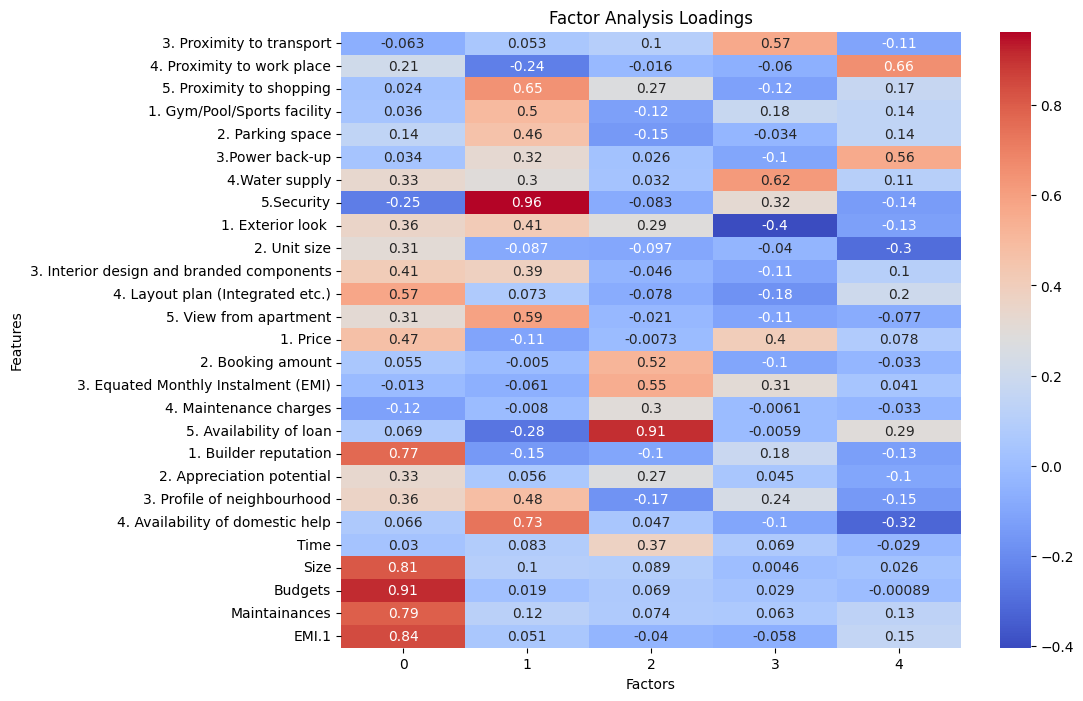
EMI.1 -0.058241 0.154306

 The loading value indicates the strength and direction of the relationship between the variable and the factor.

 Higher absolute values (closer to 1) indicate a stronger relationship between the variable and the factor.

 Positive values indicate a positive correlation, while negative values indicate a negative correlation.

 Variables with higher loading values on a particular factor are more strongly influenced by that factor.



This heatmap visualization helps in visually assessing the strength and direction of relationships between variables and factors extracted from the factor analysis.

**PART 2: Cluster Analysis**

**INTRODUCTION**

In this analysis, we conduct K-means clustering on a dataset related to housing preferences. The dataset includes various factors such as proximity to amenities, property features, financial considerations, and demographic information of potential homebuyers. The aim is to group respondents into distinct clusters based on similarities in their housing preferences and related factors.

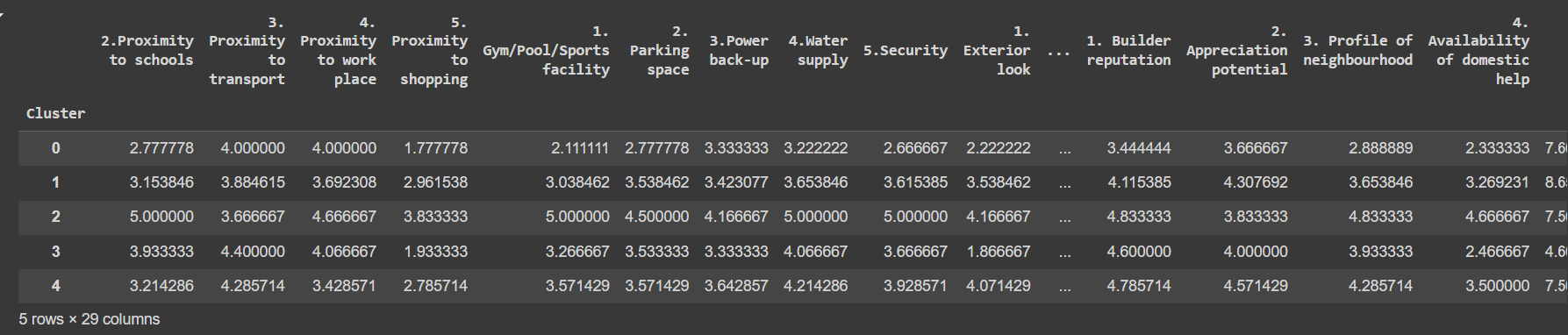
**OBJECTIVE**

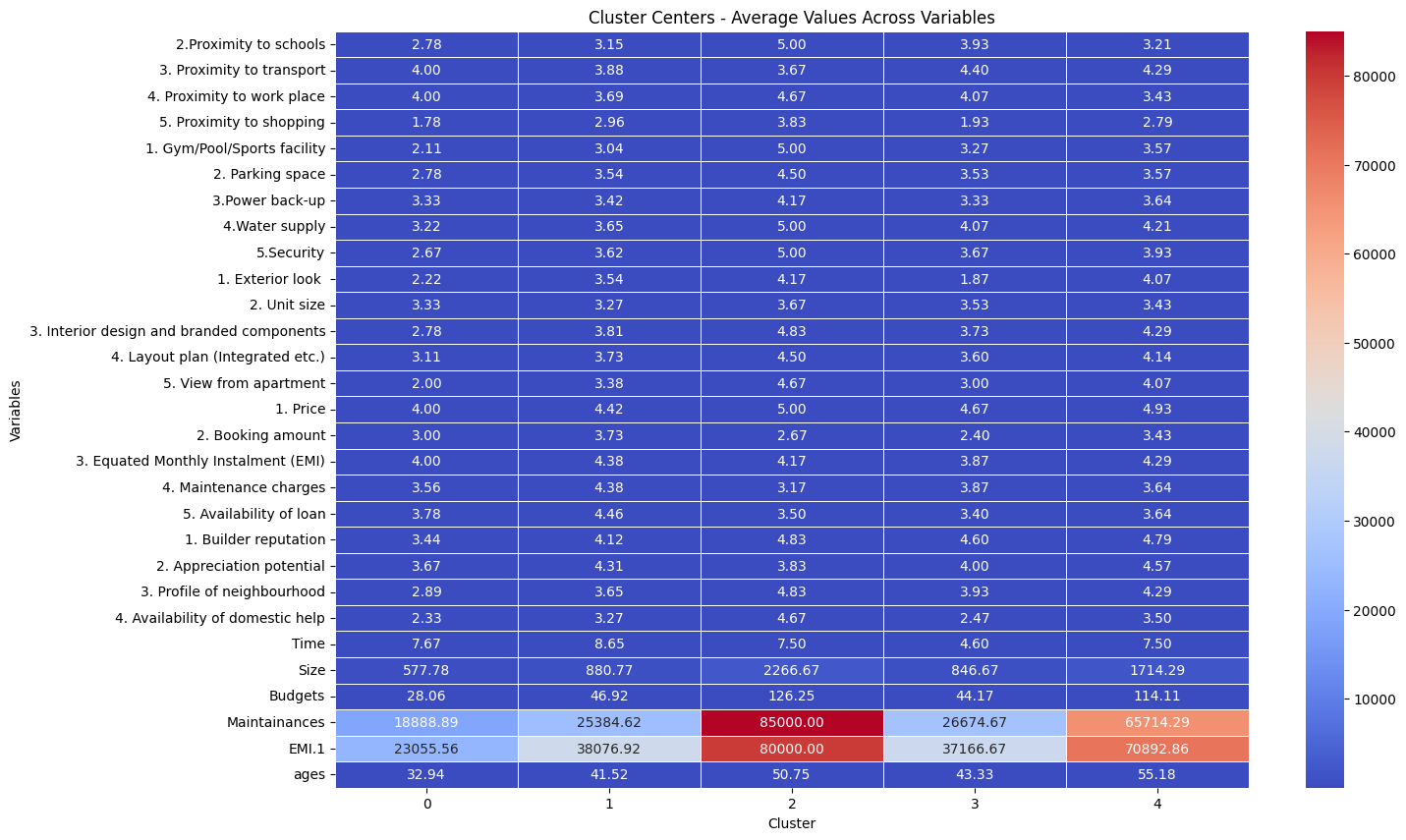
**Identify Consumer Segments**: Utilize K-means clustering to identify distinct segments or clusters of potential homebuyers based on their housing preferences and related factors.

**RESULTS & INTERPRETATIONS**

**Python**

Here, we extract the cluster labels, compute the average values for each cluster across the specified columns (from 18 to 46), and then format the data as a matrix suitable for interpretation

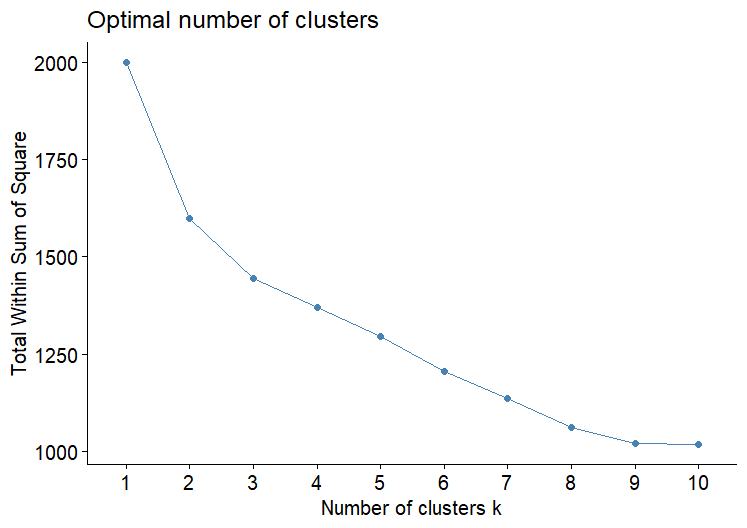




The five clusters represent distinct groups of respondents with similar characteristics and preferences.

This clustering analysis helps to segment the respondents based on their preferences and characteristics. By examining the average values of these clusters, we can identify distinct groups with similar attributes and preferences, which can be used for targeted marketing, policy making, or further analysis.

**R**



K-means clustering with 5 clusters of sizes 20, 12, 26, 6, 6

Cluster means:

X1.Proximity.to.city X2.Proximity.to.schools

1 0.42645297 -0.23895587

2 0.04373877 0.30220889

3 -0.05929967 -0.09514985

4 -1.86983224 -0.92771101

5 0.61781007 1.53212878

X3..Proximity.to.transport

1 0.03920684

2 0.35939603

3 -0.09801710

4 0.13068947

5 -0.55543023

X4..Proximity.to.work.place

1 -0.41689837

2 0.60892428

3 0.04431367

4 -0.89443650

5 0.87422325

X5..Proximity.to.shopping

1 0.1550084

2 -1.0151533

3 0.2777527

4 -1.2279100

5 1.5379269

X1..Gym.Pool.Sports.facility X2..Parking.space

1 0.4028055 0.24623856

2 -0.3608466 -0.28043836

3 -0.4173297 -0.04103976

4 -0.3608466 -1.47743137

5 1.5482837 1.39535185

X3.Power.back.up X4.Water.supply X5.Security

1 0.1644957 0.12688813 0.3133435

2 -0.6853986 0.00352467 -0.7141973

3 0.1265351 -0.32860771 -0.1289855

4 -0.8224783 -0.61329261 -0.6153953

5 1.0966378 1.60724959 1.5582487

X1..Exterior.look X2..Unit.size

1 0.6750830 0.07483749

2 -1.0223240 0.27440412

3 0.1110216 -0.25329611

4 -1.4975979 -0.09978332

5 0.8108756 0.39913326

X3..Interior.design.and.branded.components

1 0.4509889

2 -0.6434109

3 0.1012991

4 -1.9061799

5 1.2507427

X4..Layout.plan..Integrated.etc..

1 0.4194346

2 -0.2401111

3 -0.1010010

4 -1.4102729

5 0.9300507

X5..View.from.apartment X1..Price

1 0.72152166 0.3225928

2 -1.03925597 0.3225928

3 -0.02342272 -0.4069632

4 -1.42539141 -0.7312104

5 1.20032961 0.7742227

X2..Booking.amount

1 0.1941775

2 -0.7605287

3 0.3659500

4 -0.1941775

5 -0.5178068

X3..Equated.Monthly.Instalment..EMI.

1 -0.2623140

2 -0.2623140

3 0.2266145

4 0.4439160

5 -0.0269040

X4..Maintenance.charges

1 0.1794032

2 -0.5476518

3 0.3675223

4 -0.1070124

5 -0.9882911

X5..Availability.of.loan X1..Builder.reputation

1 -0.4883082 0.6235284

2 -0.0813847 0.2267376

3 0.7324623 -0.4854510

4 -0.8952317 -1.0958985

5 -0.4883082 0.6676163

X2..Appreciation.potential

1 0.6989383

2 -0.1436707

3 -0.1541249

4 -0.8231940

5 -0.5513847

X3..Profile.of.neighbourhood

1 0.6393887

2 -0.3629863

3 -0.4257468

4 -0.9457625

5 1.3853422

X4..Availability.of.domestic.help Time

1 0.41450276 -0.11583378

2 -1.07867677 -0.31604031

3 -0.06712595 0.31153016

4 -0.48479855 -0.36609194

5 1.55135536 0.03432112

Size Budgets Maintainances EMI.1

1 0.6695344 0.8396597 0.6873455 0.8241764

2 -0.4038462 -0.2958008 -0.4643225 -0.2718113

3 -0.4549401 -0.5686006 -0.4818407 -0.4857878

4 -1.2806175 -1.2667166 -1.0693715 -1.6070247

5 1.8279353 1.5233888 1.7948410 1.5084733

Clustering vector:

[1] 2 2 4 5 3 2 1 4 2 1 3 2 3 3 2 4 2 2 1 1 2 3 2

[24] 3 5 3 3 3 2 1 1 3 3 1 4 1 3 1 3 3 3 1 1 3 3 1

[47] 3 3 5 4 4 5 3 5 1 3 3 5 1 3 1 1 1 1 2 3 1 3 3

[70] 1

Within cluster sum of squares by cluster:

[1] 368.91294 189.29228 446.08704 171.58257

[5] 83.50496

(between\_SS / total\_SS = 37.1 %)

Available components:

[1] "cluster" "centers" "totss"

[4] "withinss" "tot.withinss" "betweenss"

[7] "size" "iter" "ifault"

**Cluster Sizes**: The sizes of the clusters are given as 6, 20, 18, 20, and 6, indicating the number of data points in each cluster.

**Cluster Means (Centroids)**: These are the mean values of each feature (or dimension) within each cluster.

**Clustering Vector**: This vector shows the cluster assignment for each data point in the dataset. For instance:

* Data point 1 belongs to Cluster 2, data point 2 to Cluster 2, and so on.
* It provides a way to map each data point back to its respective cluster after clustering.

**Within Cluster Sum of Squares**: This metric measures the sum of squares of distances between data points and their cluster centroids within each cluster. Lower values indicate tighter clusters where data points are closer to their centroids.

**Percentage of Variance Explained**: The percentage of variance explained by the clustering solution is also provided. In this case, it's 37%, which tells how well the clusters separate the data points based on the features used in clustering.

**Available Components**: These are the components available from the clustering result, including cluster assignments, centroids, total sum of squares (total variance in the data), within cluster sum of squares, between cluster sum of squares, cluster sizes, iteration count, and potential error codes (ifault).

**PART 3: Multidimensional Scaling**

**INTRODUCTION**

In this analysis, we aim to explore the multidimensional characteristics of various ice cream brands based on their attributes.

**OBJECTIVE**

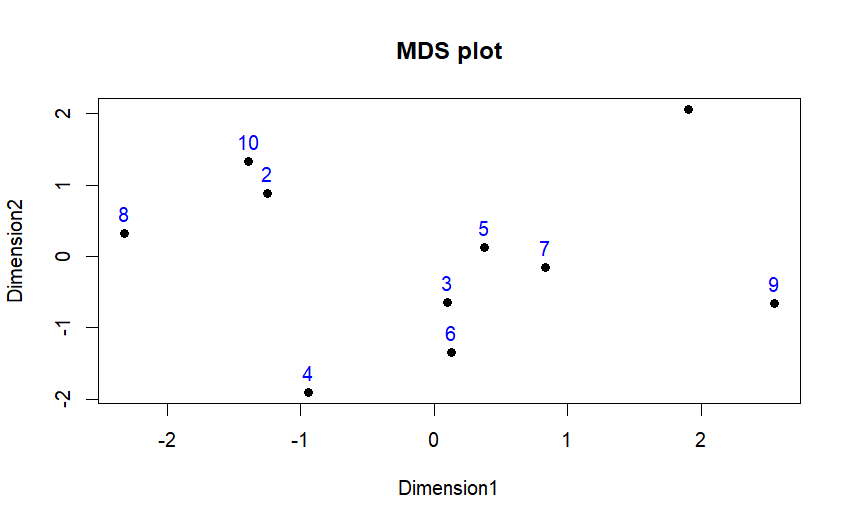
Use clustering techniques to group ice cream brands based on similarities in their attribute profiles, aiming to identify distinct segments of brands with similar characteristics.

**RESULTS & INTERPRETATIONS**

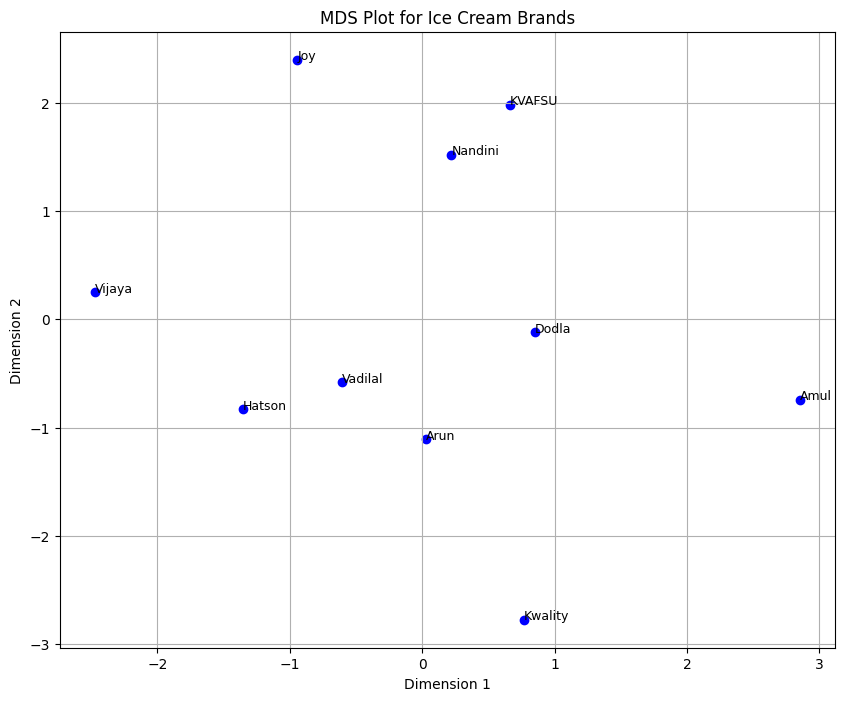
**R**

****

* In the MDS plot, closer points represent ice cream brands that are more similar based on their attributes (Price, Availability, Taste, Flavour, Consistency, Shelflife)
* Brands that are further apart in the plot are more dissimilar in terms of their attribute profiles.
* The plot provides a visual representation of how ice cream brands cluster or spread out in a 2-dimensional space based on their attribute similarities.

****

**Python**

****

**PART 4:** **Conjoint Analysis**

**INTRODUCTION**

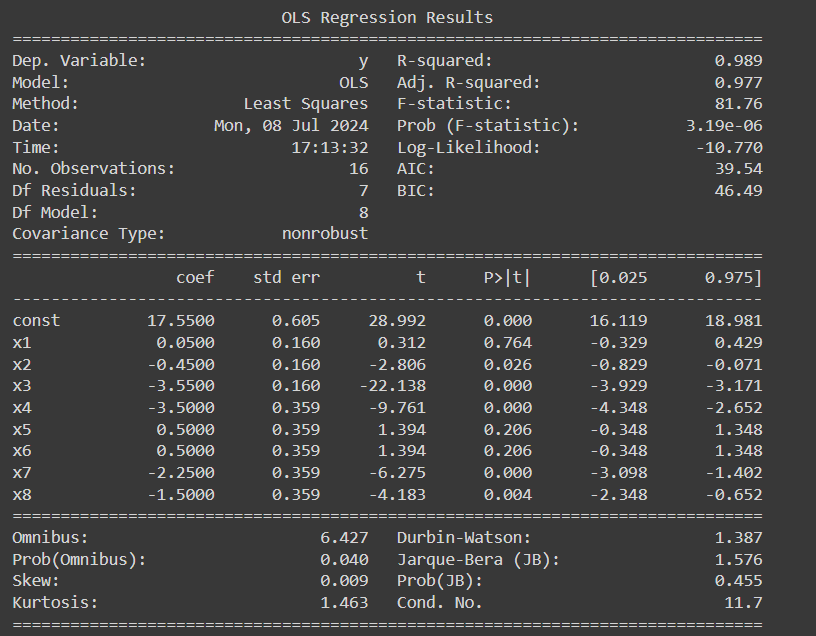
In this Conjoint Analysis study using a pizza dataset, we aim to understand consumer preferences and decision-making processes when selecting pizzas based on various attributes such as brand, price, weight, crust type, cheese, size, toppings, and spiciness. Conjoint Analysis allows us to determine the relative importance of each attribute and how different levels of these attributes influence consumer choices.

**OBJECTIVE**

**Attribute Importance:** Assess the relative importance of each pizza attribute (brand, price, weight, crust, cheese, size, toppings, spiciness) in influencing consumer preferences.

**RESULTS & INTERPRETATIONS**

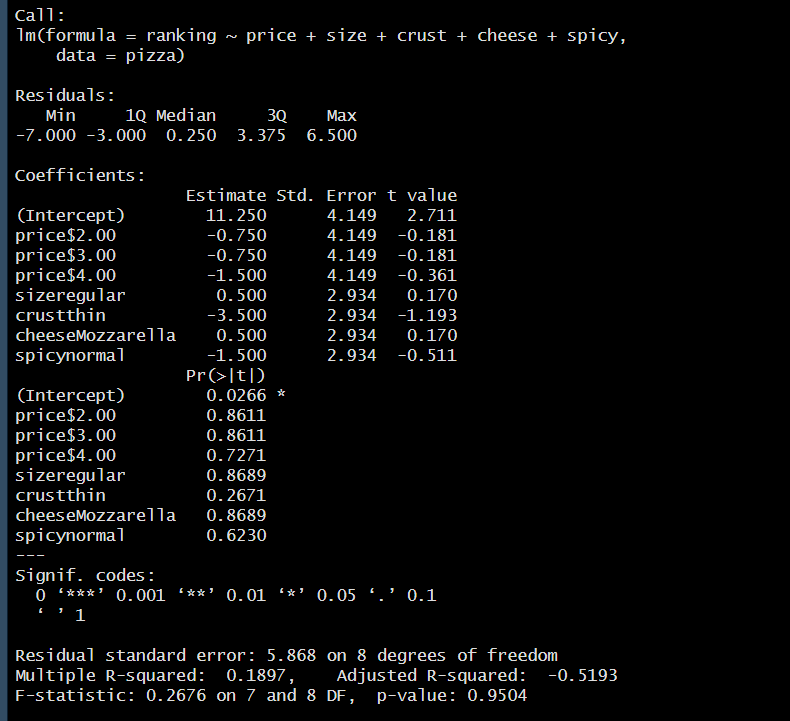
**Python**

****

Based on the coefficients:

* **Negative Coefficients** :These indicate that as the level of the attribute increases from the baseline, the preference or ranking decreases. For example, if x2 represents a higher price level, a negative coefficient suggests that higher prices are less preferred.
* **Positive Coefficients**: These indicate that as the level of the attribute increases from the baseline, the preference or ranking increases. For example, if x1 represents a specific feature that consumers prefer, a positive coefficient suggests higher preference for that feature.

**R**

****

**Business Significance**

FA helps businesses understand underlying factors influencing customer decisions, such as housing preferences. By identifying these factors, businesses can tailor marketing strategies, product offerings, and customer experiences to better meet customer needs and preferences.

PCA is crucial for businesses to simplify complex datasets and understand the most significant attributes driving customer preferences. This insight helps in product development, marketing strategies, and customer segmentation.

MDS helps businesses understand how products or brands are perceived relative to each other by consumers. This insight is valuable for competitive analysis, brand positioning, and identifying market segments based on consumer preferences.

Conjoint Analysis provides insights into product features that drive customer choices and willingness to pay. Businesses can use these insights for pricing strategies, product development, and market positioning to meet customer expectations effectively.

These analyses collectively help businesses gain deep insights into consumer behavior, preferences, and decision-making processes, enabling them to make data-driven decisions that align with customer needs and enhance business performance.