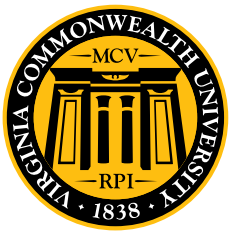


VIRGINIA COMMONWEALTH UNIVERSITY

Statistical analysis and modelling (SCMA 632)

A4b: Multivariate Analysis and Business Analytics Applications: Conjoint Analysis

FERAH SHAN SHANAVAS RABIYA

V01101398

Date of Submission: 08-07-2024

CONTENTS

Introduction

Cluster analysis is a statistical method used to identify groups, or clusters, within a dataset where observations share similar characteristics. In this report, cluster analysis is employed to categorize respondents based on their background variables. These variables encompass a range of demographic, socioeconomic, and possibly psychographic factors that collectively provide insights into the composition and diversity within the respondent pool.

By identifying clusters, we aim to uncover meaningful patterns and associations that may not be apparent through traditional analytical methods. This approach allows for a more nuanced understanding of the respondent population, facilitating targeted strategies and tailored interventions based on distinct group profiles.

This report outlines the methodology used for cluster analysis, presents the findings regarding the identified clusters, and discusses the implications of these findings for decision-making and policy formulation. By delving into the demographic and other background variables, we seek to provide a comprehensive characterization of the respondent base, offering valuable insights that can inform future research directions and operational strategies.

Objectives

• Identify Distinct Respondent Clusters: Use cluster analysis to identify homogeneous groups among respondents.

• Understand Diversity and Distribution: Explore the diversity and distribution of respondents across different clusters.

• Inform Strategic Decision-Making: Provide insights into respondent segmentation to guide organizations in developing engagement, communication, and service delivery approaches.

• Offer Insights for Policy Formulation: Understand the distinct needs and preferences of different respondent clusters to inform policies aimed at improving service delivery and satisfaction.

• Guide Future Research Directions: Identify gaps or opportunities in the respondent base to advance knowledge in relevant fields.

Business Significance

The report focuses on the importance of understanding respondent clusters for businesses to tailor their marketing strategies, enhance customer experience, optimize operational efficiency, and make strategic decisions. By identifying distinct clusters based on demographic and psychographic profiles, businesses can tailor their messaging and product/service customization, potentially increasing customer engagement and loyalty.

The report also highlights the importance of optimizing service delivery, product features, and support mechanisms based on the specific needs and preferences of each cluster, enhancing overall satisfaction and retention. It also highlights the value of resource allocation, as understanding the composition and distribution of respondent clusters helps businesses allocate resources more efficiently, prioritizing investments in areas relevant to each cluster's preferences and demands.

The report also provides valuable insights for strategic decision-making across various business functions, such as product development, market expansion, and resource planning. By leveraging these insights, businesses can gain a competitive advantage by attracting a diverse customer base and standing out in crowded markets.

The diversity within the respondent base also aids in risk management, as businesses can proactively mitigate risks and address issues before they escalate. Additionally, the report can inform compliance strategies for industries subject to regulatory requirements or policy frameworks.

In summary, the report's significance lies in its ability to transform raw respondent data into actionable insights that drive strategic decisions, improve customer relationships, and foster sustainable growth in competitive markets.

Results and Interpretation using R

- **Carry our cluster analysis and characterize the respondents based on their backg round variables.**

```
#Carry our cluster analysis and characterize the respondents based on
their background variables.
> library(cluster)
> library(factoextra)
> show(sur_int)
   X3..Proximity.to.transport X4..Proximity.to.work.place X5..Proximi
ty.to.shopping
1                           5                           2
1
2                           5                           3
1
3                           5                           2
1
4                           3                           5
4
5                           3                           4
3
6                           4                           4
2
7                           4                           4
3
8                           4                           3
1
9                           5                           5
1
10                          4                           2
2
11                          4                           3
3
12                          4                           5
2
13                          4                           4
3
14                          3                           4
3
15                          4                           4
2
16                          4                           4
1
17                          4                           5
2
18                          4                           5
2
19                          5                           4
2
20                          5                           4
2
21                          4                           5
2
22                          4                           5
3
23                          4                           5
2
```

24	5	4	4	
25	3	4	4	
26	4	5	2	
27	4	4	3	
28	4	5	2	
29	5	5	3	
30	3	3	3	
31	4	4	2	
32	4	3	2	
33	4	3	3	
34	5	4	3	
35	4	4	2	
36	5	3	3	
37	5	4	3	

	X1..Gym.Pool.Sports.facility	X2..Parking.space	X3.Power.back.up	X4.Water.supply	X5.Security
1	2	5	3	5	3
2	1	4	2	4	3
3	4	3	2	4	5
4	5	5	4	5	5
5	2	4	3	4	4
6	3	4	4	4	3
7	4	5	5	5	4
8	1	2	3	4	1
9	3	3	3	4	3
10	4	4	3	3	3
11	3	4	3	4	4
12	3	3	3	3	4
13	4	4	4	4	5
14	1	4	3	4	4
15	2	3	3	4	4

```
16                          4                   3                   4
3           2
17                          2                   3                   3
3           2
18                          4                   3                   3
4           3
19                          4                   3                   4
4           4
20                          4                   4                   4
4           5
21                          4                   3                   3
4           3
22                          3                   4                   4
4           3
23                          4                   3                   3
4           3
24                          4                   4                   4
4           3
25                          5                   4                   4
5           5
26                          3                   4                   4
4           4
27                          5                   4                   3
3           4
28                          3                   3                   4
4           4
29                          2                   3                   4
4           3
30                          4                   4                   4
4           4
31                          3                   4                   4
4           3
32                          1                   4                   4
3           3
33                          2                   3                   3
4           3
34                          4                   4                   4
4           5
35                          1                   2                   3
4           3
36                          4                   3                   3
5           4
37                          3                   3                   4
4           4
   X1..Exterior.look X2..Unit.size X3..Interior.design.and.branded.co
mponents
1                 2             4
4
2                 1             4
4
3                 1             4
3
4                 4             4
5
5                 4             3
4
6                 3             2
4
7                 4             3
5
```

8	1	3
3		
9	3	3
3		
10	4	3
4		
11	4	4
3		
12	3	4
4		
13	1	2
3		
14	2	3
3		
15	3	3
3		
16	1	1
1		
17	1	4
3		
18	1	4
3		
19	2	3
5		
20	2	3
4		
21	1	4
3		
22	3	4
4		
23	1	4
3		
24	4	3
4		
25	4	3
4		
26	3	3
5		
27	3	2
4		
28	3	4
5		
29	1	3
3		
30	4	3
4		
31	4	3
4		
32	3	4
3		
33	3	3
4		
34	4	4
4		
35	2	5
2		
36	4	3
4		
37	4	3
4		

	X4..Layout.plan..Integrated.etc..	X5..View.from.apartment	X1..Price	X2..Booking.amount
1	4	4	5	1
2	2	2	5	1
3	2	2	4	2
4	5	5	5	2
5	4	4	4	2
6	3	3	5	2
7	5	4	5	2
8	4	1	5	3
9	3	2	4	2
10	4	4	5	1
11	3	3	4	4
12	3	1	4	2
13	3	1	5	3
14	4	2	4	3
15	4	2	5	3
16	1	1	4	4
17	4	2	4	3
18	4	2	5	3
19	5	4	5	2
20	4	5	5	1
21	4	2	5	3
22	5	4	5	4
23	4	2	5	3
24	4	4	4	3
25	4	5	5	2
26	4	3	4	3
27	4	4	5	4
28	4	4	4	2
29	4	3	5	3

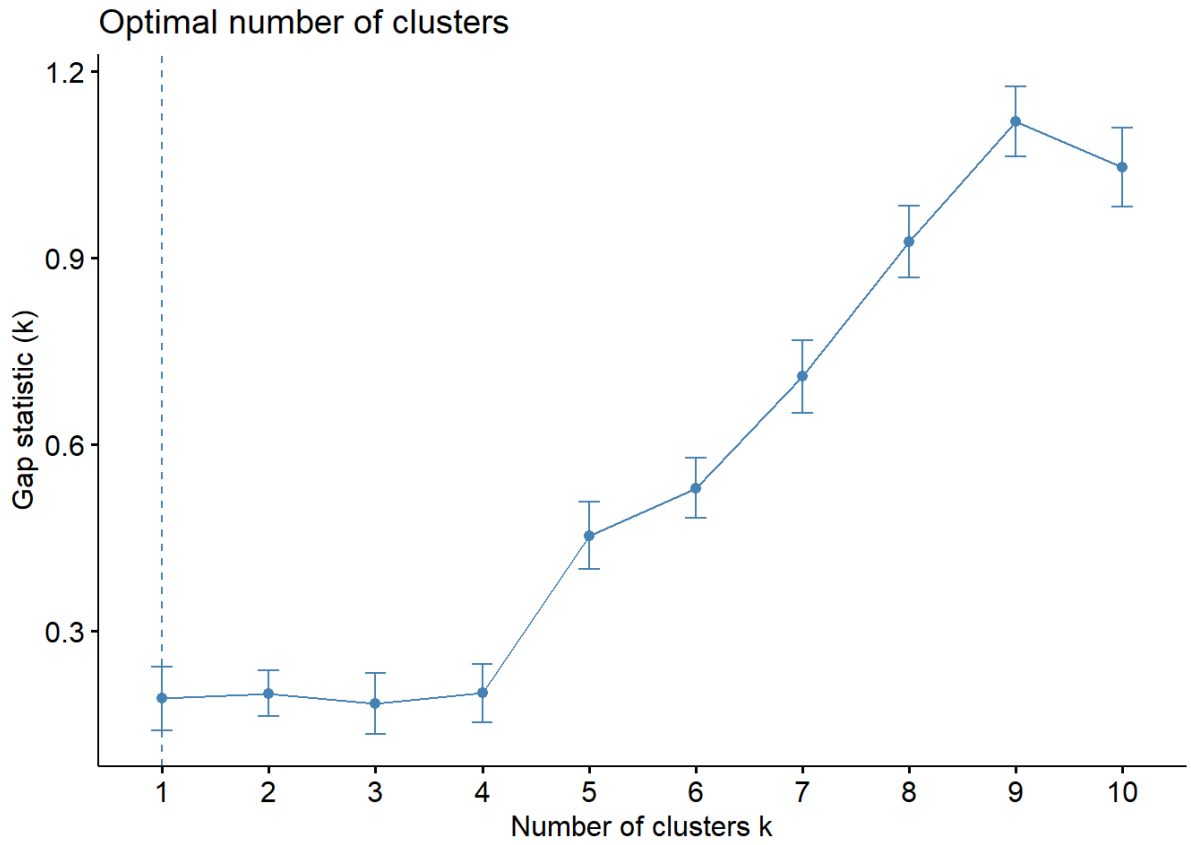
```
30                                    4                                    4
4                    3
31                                    4                                    4
5                    3
32                                    3                                    3
4                    3
33                                    3                                    3
4                    3
34                                    3                                    4
5                    4
35                                    3                                    2
5                    5
36                                    3                                    3
5                    4
37                                    3                                    3
5                    4
   X3..Equated.Monthly.Instalment..EMI. X4..Maintenance.charges X5..A
vailability.of.loan
1                                        4                                    3
3
2                                        4                                    4
4
3                                        5                                    4
2
4                                        4                                    2
2
5                                        3                                    4
4
6                                        4                                    3
3
7                                        5                                    4
4
8                                        4                                    4
3
9                                        4                                    3
4
10                                       5                                    4
4
11                                       5                                    4
4
12                                       3                                    3
4
13                                       5                                    5
5
14                                       4                                    5
5
15                                       5                                    4
4
16                                       5                                    4
4
17                                       4                                    4
4
18                                       4                                    4
4
19                                       5                                    4
3
20                                       2                                    5
2
21                                       4                                    4
4
```

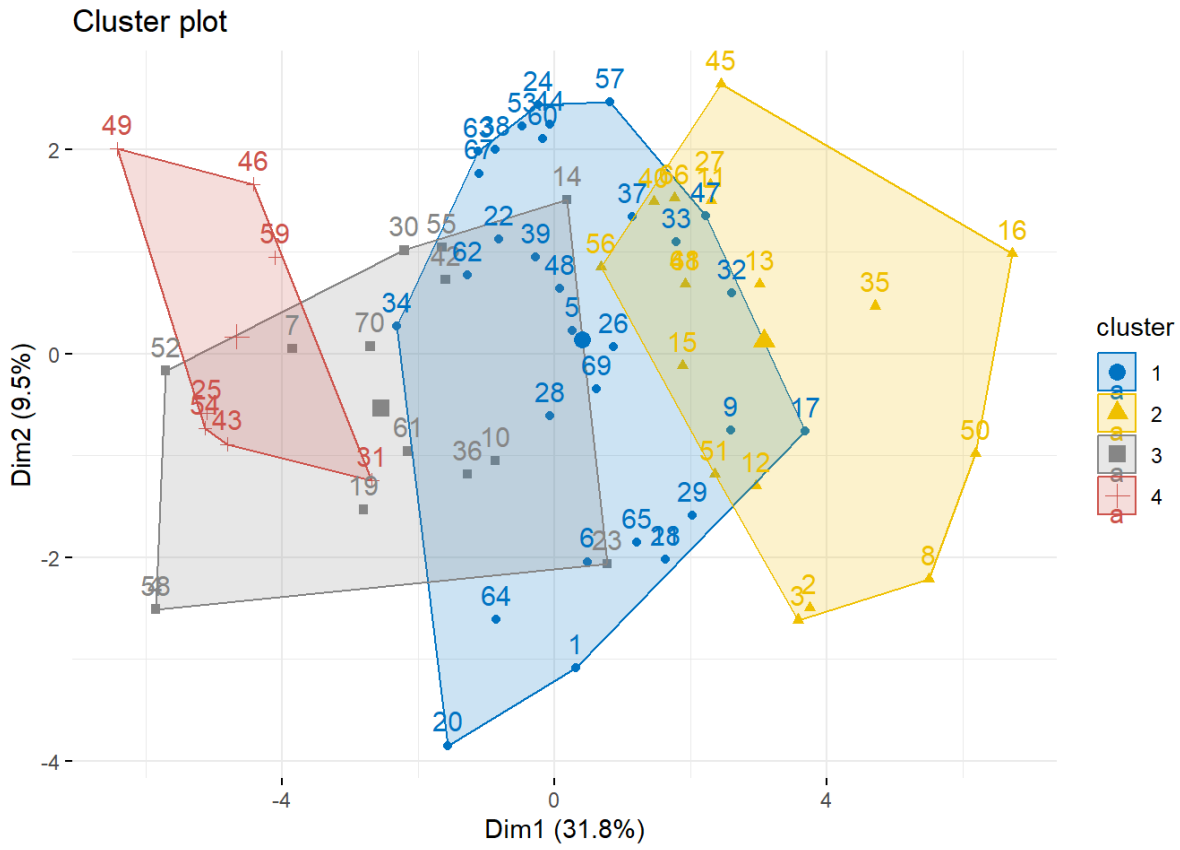
```
22                                5                         4
5
23                                4                         4
4
24                                5                         5
5
25                                4                         4
4
26                                5                         4
4
27                                5                         4
5
28                                4                         4
4
29                                4                         3
4
30                                4                         5
3
31                                4                         4
3
32                                4                         4
4
33                                5                         5
5
34                                4                         3
4
35                                5                         5
4
36                                4                         3
3
37                                4                         4
5
   X1..Builder.reputation X2..Appreciation.potential X3..Profile.of.n
eighbourhood
1                         4                         5
4
2                         5                         4
3
3                         4                         4
4
4                         5                         4
5
5                         4                         3
4
6                         5                         4
4
7                         5                         5
4
8                         4                         3
3
9                         4                         4
3
10                        5                         4
4
11                        3                         4
3
12                        3                         4
3
13                        2                         4
4
```

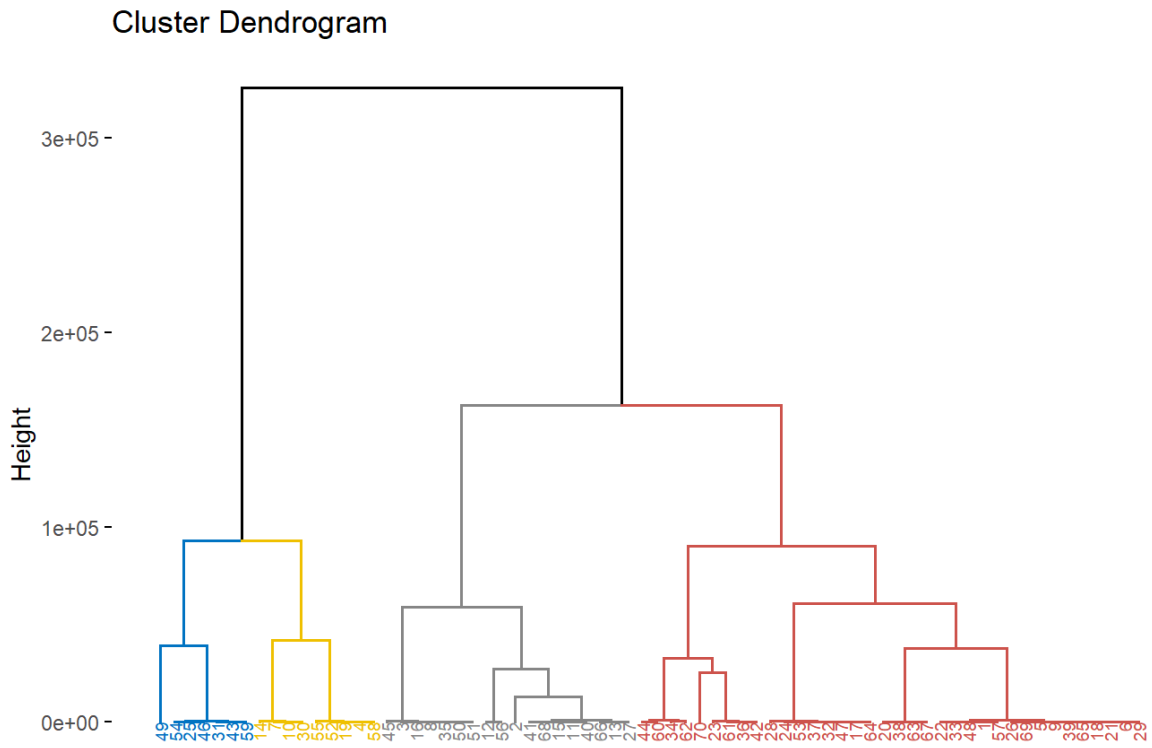
```
14                          4                          5
4
15                          5                          5
4
16                          3                          4
3
17                          4                          4
3
18                          5                          4
4
19                          5                          5
5
20                          5                          4
5
21                          5                          4
4
22                          5                          5
4
23                          5                          4
4
24                          4                          5
4
25                          5                          3
5
26                          4                          4
3
27                          3                          4
2
28                          5                          4
4
29                          4                          3
3
30                          4                          5
4
31                          5                          4
4
32                          4                          4
3
33                          5                          3
3
34                          4                          5
5
35                          4                          4
3
36                          5                          4
4
37                          4                          4
3
   X4..Availability.of.domestic.help Time Size Budgets Maintainances
EMI.1
1                                    1    9 1200    72.5          30000
42500
2                                    2    9  800    32.5            120
27500
3                                    4    3  400    12.5          10000
10000
4                                    5    3 1600   102.5          70000
80000
5                                    3   18  800    52.5          30000
42500
```

6	3	3	800	52.5	30000
42500					
7	3	9	1600	87.5	50000
80000					
8	2	3	300	12.5	10000
10000					
9	3	18	800	52.5	30000
42500					
10	2	3	1600	102.5	50000
80000					
11	2	9	800	32.5	10000
27500					
12	2	3	800	52.5	10000
42500					
13	2	3	300	12.5	10000
27500					
14	4	9	1200	72.5	50000
80000					
15	2	9	800	32.5	10000
27500					
16	1	18	300	12.5	10000
10000					
17	2	3	800	32.5	30000
27500					
18	2	3	800	52.5	30000
42500					
19	3	9	1600	102.5	70000
80000					
20	3	3	1200	52.5	30000
57500					
21	2	3	800	52.5	30000
42500					
22	2	9	1200	52.5	30000
42500					
23	2	3	1200	87.5	50000
57500					
24	3	9	800	52.5	30000
27500					
25	4	9	2400	150.0	90000
80000					
26	2	3	800	32.5	30000
42500					
27	2	3	300	12.5	10000
27500					
28	3	9	1200	32.5	30000
27500					
29	2	3	800	52.5	30000
42500					
30	4	9	1600	102.5	50000
80000					
31	3	3	2000	150.0	90000
80000					
32	3	9	800	32.5	30000
27500					
33	3	3	1200	52.5	30000
42500					
34	4	9	1600	87.5	50000
42500					
35	2	3	300	12.5	10000
10000					

```
36                                    4     3 1600     72.5          50000
57500
37                                    3     9  800     32.5          30000
27500
 [ reached 'max' / getOption("max.print") -- omitted 33 rows ]
> fviz_nbclust(sur_int,kmeans,method = "gap_stat")
Clustering k = 1,2,..., K.max (= 10): .. done
Bootstrapping, b = 1,2,..., B (= 100)  [one "." per sample]:
.................................................. 50
.................................................. 100
> set.seed(123)
> km.res<-kmeans(sur_int,4,nstart = 25)
> fviz_cluster(km.res,data=sur_int,palette="jco",
+              ggtheme = theme_minimal())
> res.hc <- hclust(dist(sur_int), method = "ward.D2")
> fviz_dend(res.hc,cex=0.5,k=4,palette = "jco")
> library(pheatmap)
> pheatmap(t(sur_int),cutree_cols = 4)
```




Cluster plot
Dim2 (9.5%)
Dim1 (31.8%)
cluster
1
2
3
4


Cluster Dendrogram
Height

Interpretation:

The text describes a cluster analysis using R, which involves several steps. The first step involves determining the optimal number of clusters using the gap statistic method. The second step involves applying k-means clustering to partition the data into k clusters based on similarity. The third step involves visualizing the clustering results using fviz_cluster, which displays the clusters in a scatterplot. The fourth step involves performing hierarchical clustering using Ward's method on the distance matrix, creating a tree-like structure called a dendrogram. The fifth step involves visualizing the hierarchical clustering dendrogram using fviz_dend, which helps understand the relationships between clusters and determines the appropriate number of clusters. The sixth step involves plotting a heatmap of the clustered data using `pheatmap`, which visualizes how different variables behave across the identified clusters.

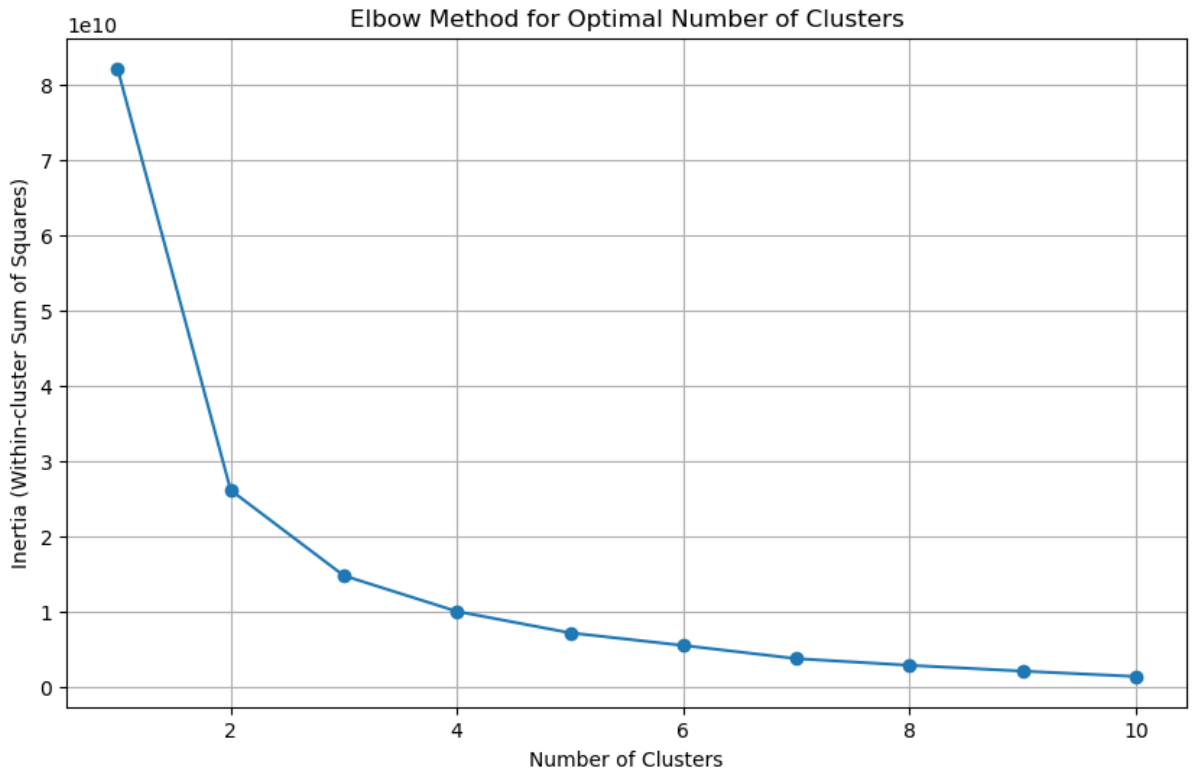
The interpretation of the results involves identifying distinct clusters, analyzing the dendrogram's insights, and examining the heatmap patterns. To fully interpret the results, focus on cluster profiles, identifying significant differences between clusters, and relating cluster characteristics to business objectives. This process should provide actionable insights for further business decisions or strategic planning.

Results and Interpretation using Python

- **Determining the optimal number of clusters using the Elbow Method and plottin g it and visualizing clustering results.**

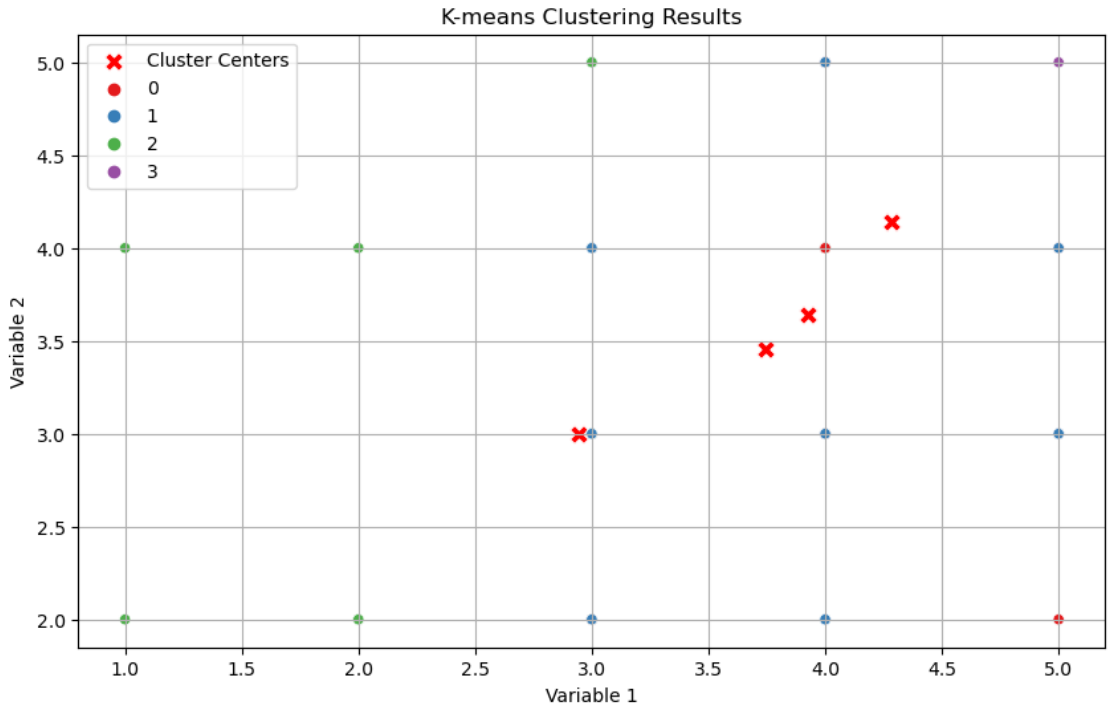
```
# Determine Optimal Number of Clusters using the Elbow Method
inertia = []
for k in range(1, 11):
    kmeans = KMeans(n_clusters=k, random_state=123, n_init=25)
    kmeans.fit(sur_int)
    inertia.append(kmeans.inertia_)

# Plotting the Elbow Method
plt.figure(figsize=(10, 6))
plt.plot(range(1, 11), inertia, marker='o')
plt.xlabel('Number of Clusters')
plt.ylabel('Inertia (Within-cluster Sum of Squares)')
plt.title('Elbow Method for Optimal Number of Clusters')
plt.grid(True)
plt.show()
```




```
# Perform k-means clustering
kmeans = KMeans(n_clusters=4, random_state=123, n_init=25)
km_res = kmeans.fit(sur_int)
# Visualizing k-means Clustering Results
plt.figure(figsize=(10, 6))
sns.scatterplot(x=km_res.cluster_centers_[:, 0], y=km_res.cluster_centers_
[:, 1], color='red', marker='X', s=100, label='Cluster Centers')
sns.scatterplot(x=sur_int.iloc[:, 0], y=sur_int.iloc[:, 1], hue=km_res.lab
els_, palette='Set1', legend='full')
plt.title('K-means Clustering Results')
```

```
plt.xlabel('Variable 1')
plt.ylabel('Variable 2')
plt.grid(True)
plt.legend()
plt.show()
```



Interpretation:

The plot generated shows the number of clusters on the x-axis and the inertia on the y-axis. T he "elbow" point in the plot suggests the optimal number of clusters. After determining the op timal number of clusters, k-means clustering is performed using `KMeans` from scikit-learn. The final results are the best output of `n_init` consecutive runs in terms of inertia.

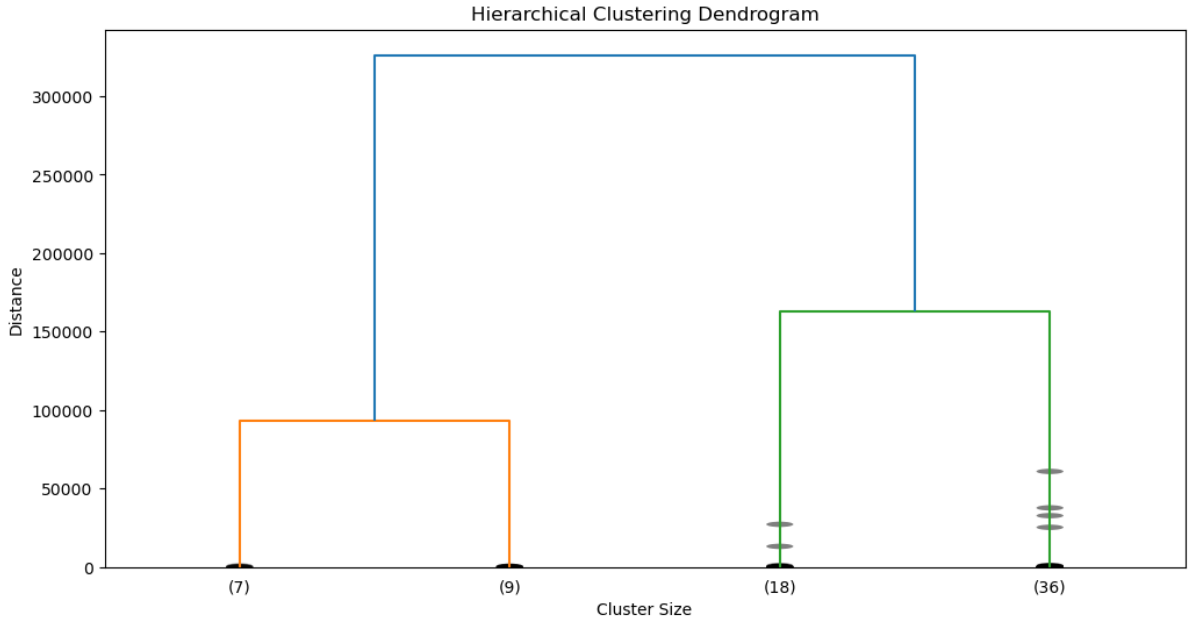
The second plot visualizes the clustering results, showing cluster centers as red 'X' markers an d data points colored according to their assigned cluster. The results are interpreted by looking for the point where adding more clusters does not significantly reduce the inertia, which is the optimal number of clusters in your case.

- Perform Hierarchical Clustering and plot dendrogram and heatmap.

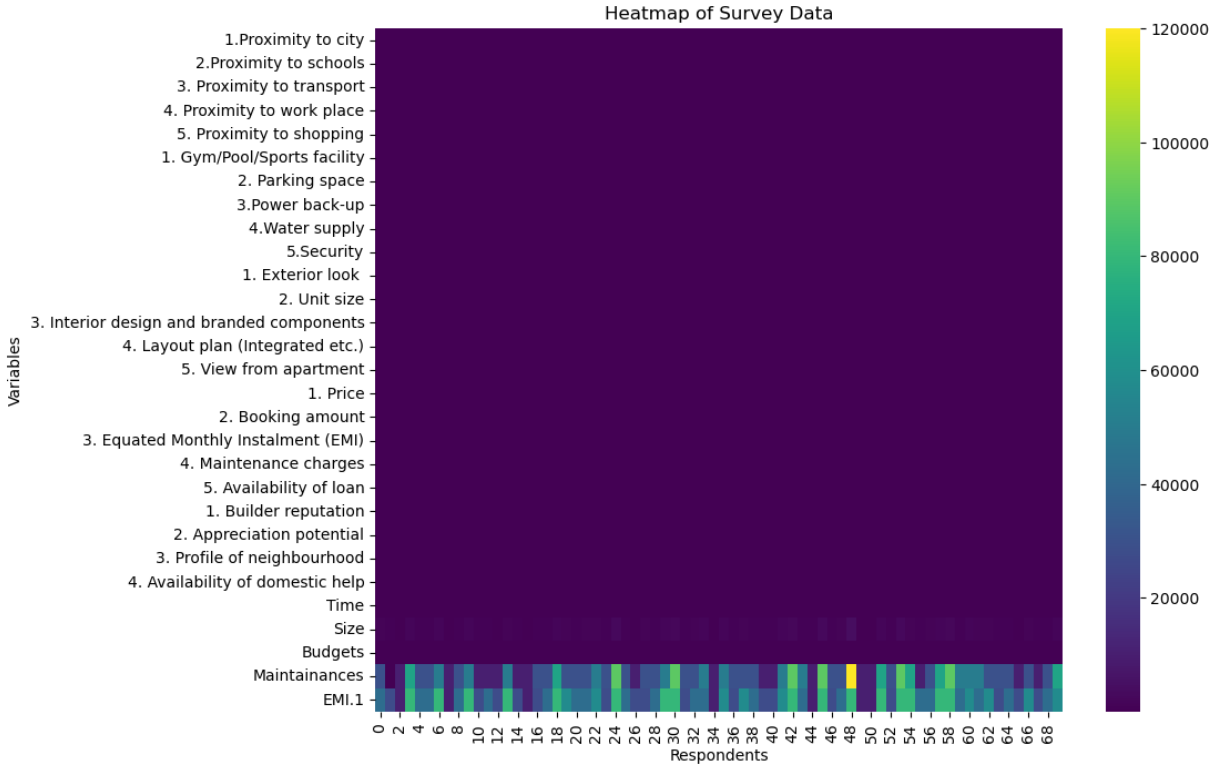
```
# Perform Hierarchical Clustering (Ward's method)
Z = linkage(sur_int, method='ward')

# Plotting the Dendrogram
plt.figure(figsize=(12, 6))
dendrogram(Z, p=4, truncate_mode='lastp', orientation='top', leaf_font_size=10
, show_contracted=True)
```

```
plt.title('Hierarchical Clustering Dendrogram')
plt.xlabel('Cluster Size')
plt.ylabel('Distance')
plt.show()
```




```
# Heatmap of clustered data
plt.figure(figsize=(10, 8))
sns.heatmap(sur_int.T, cmap='viridis', cbar=True)
plt.title('Heatmap of Survey Data')
plt.xlabel('Respondents')
plt.ylabel('Variables')
plt.show()
```



Interpretation:

Dendrogram visualizes the hierarchical clustering process. Points where branches merge verti cally indicate clusters. The height at which branches merge reflects the distance or dissimilari ty between clusters or individual data points. The longer the vertical lines (height), the greater the dissimilarity.

Heatmap provides a visual representation of the dataset after hierarchical clustering. Variables (columns) and respondents (rows) are reordered based on their clustering similarity. Similar p atterns or clusters will appear as blocks of similar colors in the heatmap.

Recommendation

To improve the hierarchical clustering using Ward's method, adjust the `p` parameter in `den drogram()` to display an appropriate number of merged clusters. Pay attention to the heights a t which clusters merge in the dendrogram, as higher merge heights indicate greater dissimilari ty between clusters or data points. Validate cluster interpretations by looking for distinct bran ches that form at various heights, suggesting natural divisions in the data. Analyze the heatma p after hierarchical clustering to identify patterns and relationships between variables and res pondents. Consider data normalization before clustering to ensure equal contribution of varia bles with different scales. Compare the results with other methods like k-means or DBSCAN to reveal different insights or patterns in the data. Refine your clustering approach based on vi sualizations and domain knowledge, and experiment with different parameters or distance me trics to see how they affect results and interpretability. By following these recommendations, you can enhance your understanding and interpretation of hierarchical clustering results, maki ng more informed decisions based on the structure and patterns revealed in your data.

R Codes

```
# Function to auto-install and load packages
install_and_load <- function(packages) {
  for (package in packages) {
    if (!require(package, character.only = TRUE)) {
      install.packages(package, dependencies = TRUE)
    }
    library(package, character.only = TRUE)
  }
}

# List of packages to install and load
packages <- c("cluster", "FactoMineR", "factoextra", "pheatmap")

install_and_load(packages)
survey_df<-read.csv("C:\\Users\\Ferah Shan\\Downloads\\Survey.csv",header=TRUE)
sur_int=survey_df[,20:46]

#Carry our cluster analysis and characterize the respondents based on their background variab
les.
library(cluster)
library(factoextra)
show(sur_int)
fviz_nbclust(sur_int,kmeans,method = "gap_stat")
set.seed(123)
km.res<-kmeans(sur_int,4,nstart = 25)
fviz_cluster(km.res,data=sur_int,palette="jco",
             ggtheme = theme_minimal())
res.hc <- hclust(dist(sur_int), method = "ward.D2")
fviz_dend(res.hc,cex=0.5,k=4,palette = "jco")
library(pheatmap)
pheatmap(t(sur_int),cutree_cols = 4)
```

Python Codes

```
# Import required libraries
import pandas as pd
import numpy as np
from scipy.cluster.hierarchy import dendrogram, linkage
from sklearn.cluster import KMeans
import seaborn as sns
import matplotlib.pyplot as plt

# Load the dataset
df = pd.read_csv("C:\\Users\\Ferah Shan\\Downloads\\Survey.csv")

# Select columns of interest for clustering
sur_int = df.iloc[:, 17:46]

# Determine Optimal Number of Clusters using the Elbow Method
inertia = []
for k in range(1, 11):
    kmeans = KMeans(n_clusters=k, random_state=123, n_init=25)
    kmeans.fit(sur_int)
    inertia.append(kmeans.inertia_)

# Plotting the Elbow Method
plt.figure(figsize=(10, 6))
plt.plot(range(1, 11), inertia, marker='o')
plt.xlabel('Number of Clusters')
plt.ylabel('Inertia (Within-cluster Sum of Squares)')
plt.title('Elbow Method for Optimal Number of Clusters')
plt.grid(True)
plt.show()

# Perform k-means clustering
kmeans = KMeans(n_clusters=4, random_state=123, n_init=25)
```

```
km_res = kmeans.fit(sur_int)

# Visualizing k-means Clustering Results
plt.figure(figsize=(10, 6))
sns.scatterplot(x=km_res.cluster_centers_[:, 0], y=km_res.cluster_centers_[:, 1], color='red', marker='X', s=100, label='Cluster Centers')
sns.scatterplot(x=sur_int.iloc[:, 0], y=sur_int.iloc[:, 1], hue=km_res.labels_, palette='Set1', legend='full')
plt.title('K-means Clustering Results')
plt.xlabel('Variable 1')
plt.ylabel('Variable 2')
plt.grid(True)
plt.legend()
plt.show()

# Perform Hierarchical Clustering (Ward's method)
Z = linkage(sur_int, method='ward')

# Plotting the Dendrogram
plt.figure(figsize=(12, 6))
dendrogram(Z, p=4, truncate_mode='lastp', orientation='top', leaf_font_size=10, show_contracted=True)
plt.title('Hierarchical Clustering Dendrogram')
plt.xlabel('Cluster Size')
plt.ylabel('Distance')
plt.show()

# Heatmap of clustered data
plt.figure(figsize=(10, 8))
sns.heatmap(sur_int.T, cmap='viridis', cbar=True)
plt.title('Heatmap of Survey Data')
plt.xlabel('Respondents')
plt.ylabel('Variables')
plt.show()
```

References

1. www.github.com
2. www.geeksforgeeks.com
3. www.datacamp.com