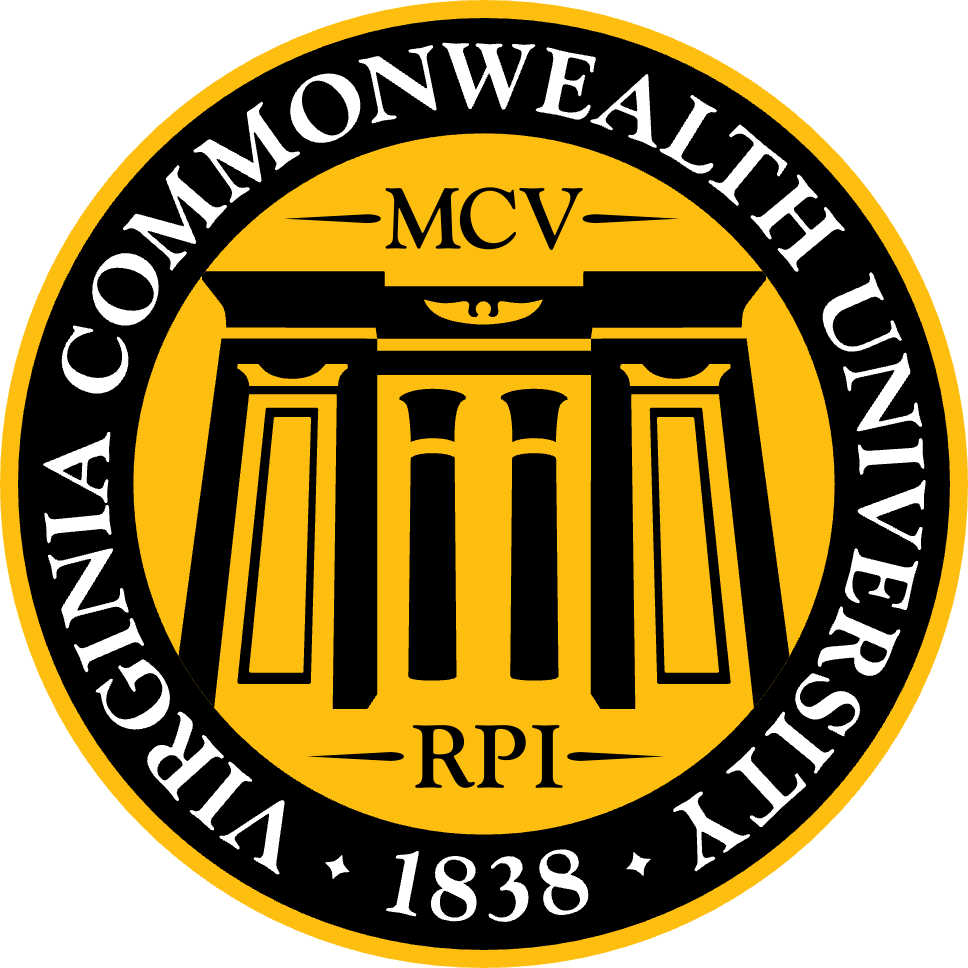
****

**VIRGINIA COMMONWEALTH UNIVERSITY**

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

# **A5- Multivariate Analysis and Business Analytics Application**

**ROSHAN RAJKUMAR SIVAKUMAR**

**V01151141**

**Date of Submission: 05-07-2025**

**CONTENTS**

|  |  |  |
| --- | --- | --- |
| **Sl. No.** | **Title** | **Page No.** |
| **1.** | Introduction | **1** |
| **2.** | Results | **2** |
| **3.** | Interpretations | **2** |
| **4.** | Recommendations | **5** |
| **5.** | Codes | **6** |
| **6.** | References | **13** |

**1. Introduction**

This report is completely based on three separate datasets I worked with. Each one had a different focus one was about housing preferences, another looked at how people feel about ice cream brands, and the last one was related to pizza features.

I used multivariate analysis methods to break down these datasets and try to find useful patterns. The idea was to take a bunch of survey responses and figure out what actually influences people’s choices in each case.

Each method I used helped break down the data in a different awesome showed patterns in responses, others helped group people or rank product features. The main idea was to take complex survey data and pull out insights that could help with real-world business decisions.

**2. Objectives**

* Find underlying patterns in large data using PCA and factor analysis
* Group people with similar preferences using cluster analysis
* Understand how brands are perceived using MDS
* Identify what matters most in a pizza using conjoint analysis

**3. Significance of the Study**

By studying and learning the consumer preferences is at the core of effective business strategy. Through this study, businesses can:

Identify the most critical factors influencing customer choices.

Target different market segments with more tailored strategies.

* Improve product features and pricing based on actual customer value.
* Visualize complex relationships in a way that supports strategic thinking

Each technique applied in this project brings a unique strength —simplifying variables, visualizing relationships, or uncovering what truly matters to customers.

**4. Analysis and Results**

**4.1 Principal Component Analysis – Survey.csv**

The PCA reduced over 20 housing-related variables to 2 main components.

A graph of variables with text

AI-generated content may be incorrect.

* **Component 1 (31.8%)**: Focused on financial concerns loan availability, EMI, booking amount
* **Component 2 (9.5%)**: Focused on lifestyle aspects view, design, shopping proximity

Together they explained 41.35% of the total variation. This shows two distinct thinking patterns among buyers: money-based vs lifestyle-based.

**4.2 Factor Analysis – Survey.csv**

**A graph showing a number of numbers

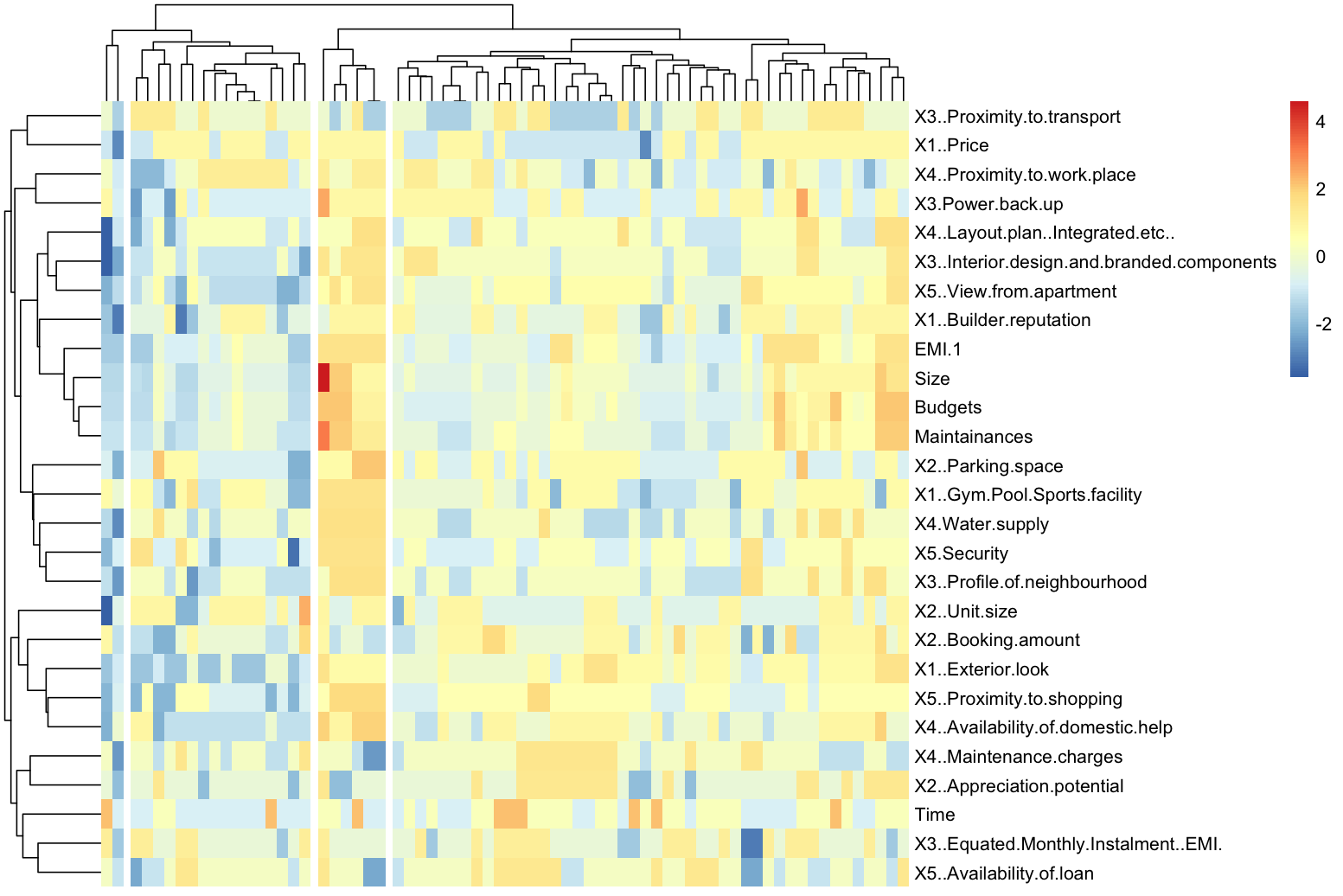
AI-generated content may be incorrect.**

Using Varimax rotation, four factors were extracted:

* **Factor 1**: Lifestyle – view, security, look, branded interiors
* **Factor 2**: Structure & finance – unit size, booking amount
* **Factor 3**: Value – price, water, future value
* **Factor 4**: Budget concerns – EMI, maintenance, budget

Each factor groups related thoughts, helping simplify the way decisions are made.

**4.3 Cluster Analysis – Survey.csv**



Hierarchical clustering (Ward’s method) grouped people into 3 clusters:

* **Cluster 1**: Focused on cost and access
* **Cluster 2**: Focused on amenities and looks
* **Cluster 3**: Balanced across both ends

The heatmap and dendrogram confirmed this grouping.

**4.4 Multidimensional Scaling – Icecream.csv**

MDS created a map showing how customers see different ice cream brands.

* Brands close to each other are seen as similar
* One brand stood out far from others possibly seen as unique or very different

A graph with black dots

AI-generated content may be incorrect.

A graph showing different types of ice cream brands

AI-generated content may be incorrect.

This map helps businesses see overlap and space for change.

**4.5 Conjoint Analysis – Pizza\_data.csv**

This analysis found which pizza features are most important.

* **Most important**: Weight (51.2%), then crust (16.8%)
* **Medium**: Toppings, spice, price
* **Least**: Cheese, size, brand

The best combination according to customers:

* Pizza Hut
* $1.00
* 100g
* Thick crust
* Mozzarella cheese
* Regular size
* Mushroom topping
* Extra spicy

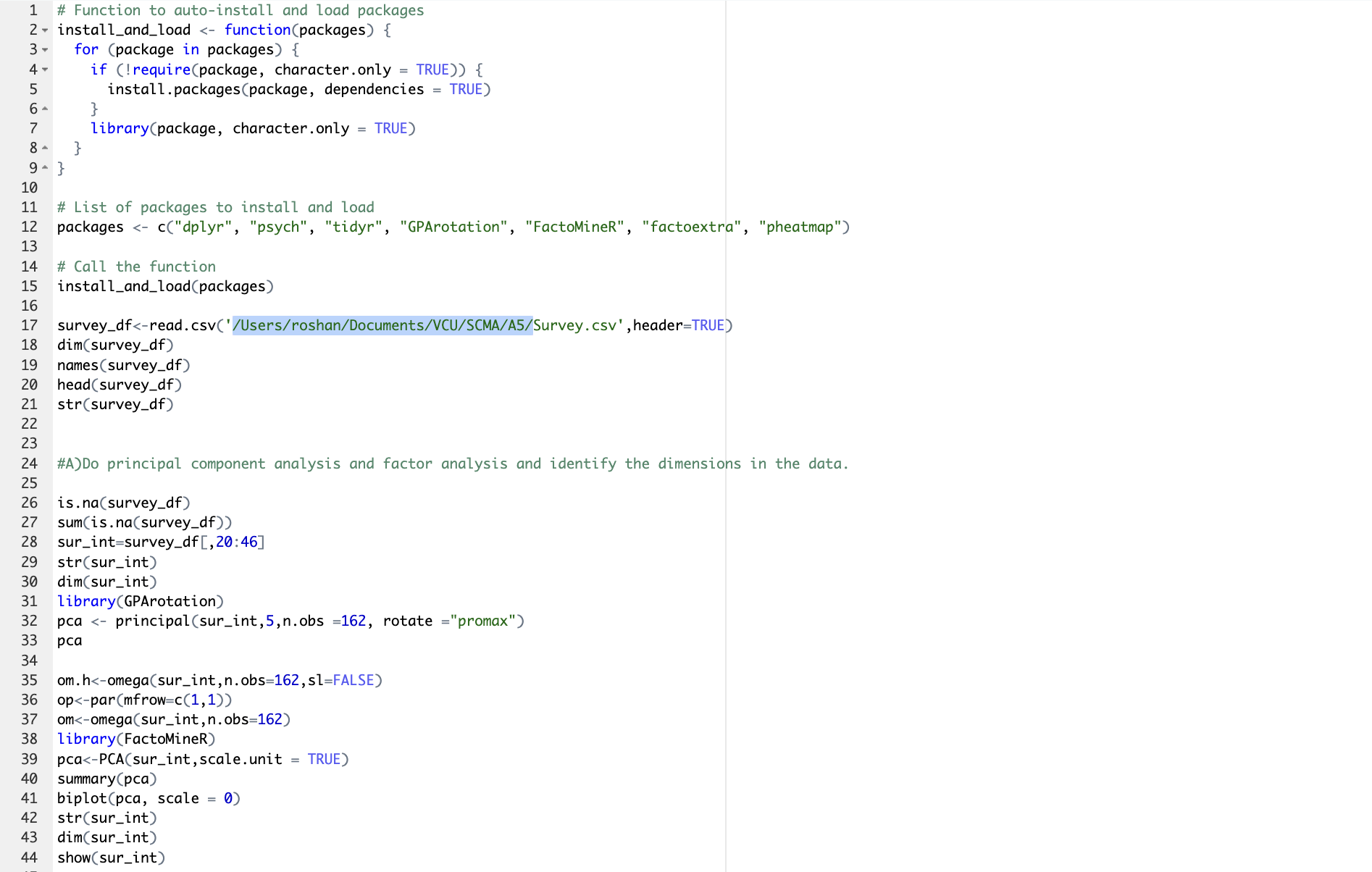
Customers care most about quantity and texture not brand or size.

**4. Recommendations**

* Real estate: Separate plans for cost-focused and lifestyle-focused buyers
* Brands: Use MDS to check if you’re lost in the crowd
* Pizza chains: Focus on value, crust, toppings not brand name

**Codes:**

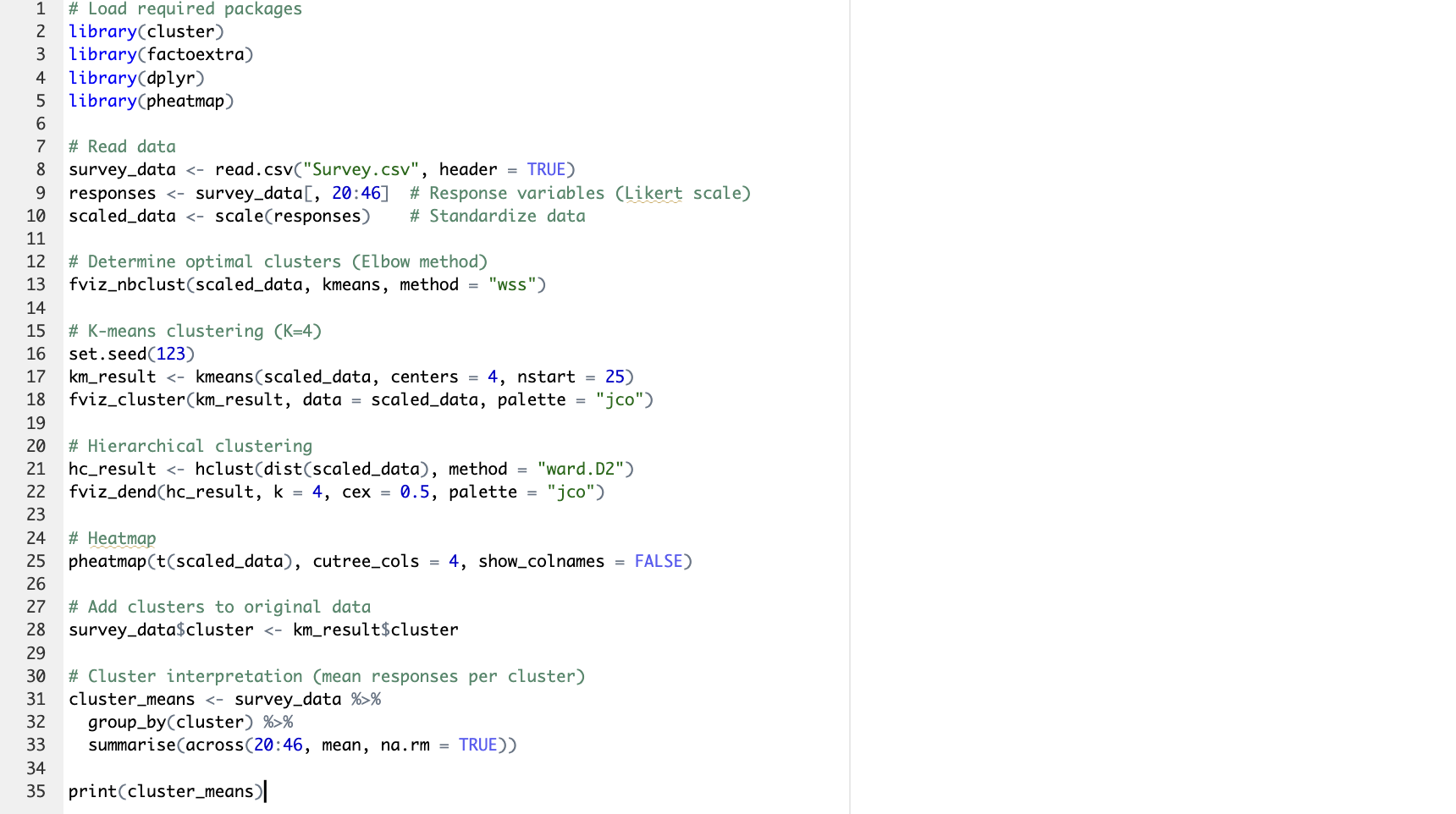
* 1. **R language:**
  2. ***Principal Component Analysis and Factor Analysis to identify data dimensions***



A computer screen shot of a code

AI-generated content may be incorrect.

* 1. ***Cluster Analysis to characterize respondents based on background variables***



* 1. ***Multidimensional Scaling***

***A screenshot of a computer program

AI-generated content may be incorrect.***

* 1. ***Conjoint Analysis***

***A screenshot of a computer

AI-generated content may be incorrect.***

***A computer screen with text

AI-generated content may be incorrect.***

1. **Python**
   1. ***PCA, FA and Cluster***

***A screenshot of a computer

AI-generated content may be incorrect.***

***A screenshot of a computer program

AI-generated content may be incorrect.***

***A screenshot of a computer program

AI-generated content may be incorrect.***

* 1. ***MDS***

***A black screen with text

AI-generated content may be incorrect.***

***A screenshot of a computer program

AI-generated content may be incorrect.***

* 1. ***Conjoint***

******

***A screenshot of a computer program

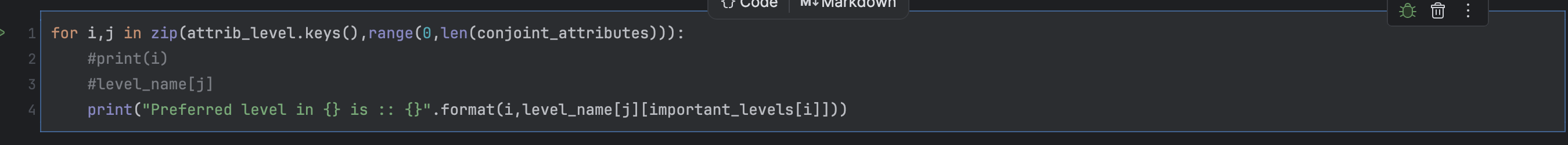
AI-generated content may be incorrect.***

***A screenshot of a computer program

AI-generated content may be incorrect.***

***A screenshot of a computer program

AI-generated content may be incorrect.***

******

**Reference:**

1. SCMA 632 Team. *Survey.csv* [Data file]. GitHub. https://github.com/scma-632/scma632-A5-2025-Multivariate-Analysis/blob/main/Survey.csv
2. SCMA 632 Team. *icecream.csv* [Data file]. GitHub. https://github.com/scma-632/scma632-A5-2025-Multivariate-Analysis/blob/main/icecream.csv
3. SCMA 632 Team. *pizza\_data.csv* [Data file]. GitHub. https://github.com/scma-632/scma632-A5-2025-Multivariate-Analysis/blob/main/pizza\_data.csv
4. JetBrains. (n.d.). *PyCharm* [Computer software]. https://www.jetbrains.com/pycharm/
5. Posit, PBC. (n.d.). *RStudio* [Computer software]. https://posit.co/products/open-source/rstudio/
6. Project Jupyter. (n.d.). *Jupyter Notebook* [Computer software]. https://jupyter.org/
7. Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering, 9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55
8. McKinney, W. (2010). Data structures for statistical computing in Python. In *Proceedings of the 9th Python in Science Conference* (pp. 51–56).
9. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.
10. Virtanen, P., Gommers, R., Oliphant, T. E., et al. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods, 17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2
11. Waskom, M. L. (2021). seaborn: Statistical data visualization. *Journal of Open Source Software, 6*(60), 3021. https://doi.org/10.21105/joss.03021
12. Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis* (2nd ed.). Springer. https://ggplot2.tidyverse.org
13. Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., … & Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software, 4*(43), 1686. https://doi.org/10.21105/joss.01686