THESIS DEFENSE

Presentation by Thanh Hien

Customer Segmentation Analysis | 2024

National Economics University

INTRODUCTION

Customer segmentation offers a potential means for marketers to efficiently target individual customers. This study, in particular, concentrates on the use of K – means method to cluster customers, thereby offering valuable insights that can help stores implement effective business strategies to boost revenue and retain customers.

THEORETICAL FRAMEWORK

OVERVIEW

A process where customers are categorized based on shared characteristics, such as demographics or behaviors, with the aim of improving the effectiveness of marketing or sales efforts.

IMPORTANT

- Improve brand loyalty and customer lifetime value
- Deliver 1:1 experiences at scale
- Stay on top of changing customer needs

BACKGROUND

Dataset has 2240 rows and 29 columns and it has attributes in 4 areas: people, products, promotions, place

- **People contains personal information of each customer**
- **Products contains the amount that a customer spent on each type of products in the last 2 years**
- **Promotion features contain the information about the number of order with discount and the number of discount that a customer accepted**
- **Place features describe where the customer make their purchase, from website or at the store**

PROBLEM

Categorize customers into distinct groups based on various characteristics and behaviors, aiming to uncover unique needs and pain points within each segment

METHODOLOGY

PRINCIPAL COMPONENT ANALYSIS (PCA)

A statistical procedure that allows you to summarize the information content in large data tables by means of a smaller set of “summary indices” that can be more easily visualized and analyzed

K – MEANS CLUSTERING

A centroid-based algorithm or a distance-based algorithm, where we calculate the distances to assign a point to a cluster

PCA: CHOOSE SCALING METHOD

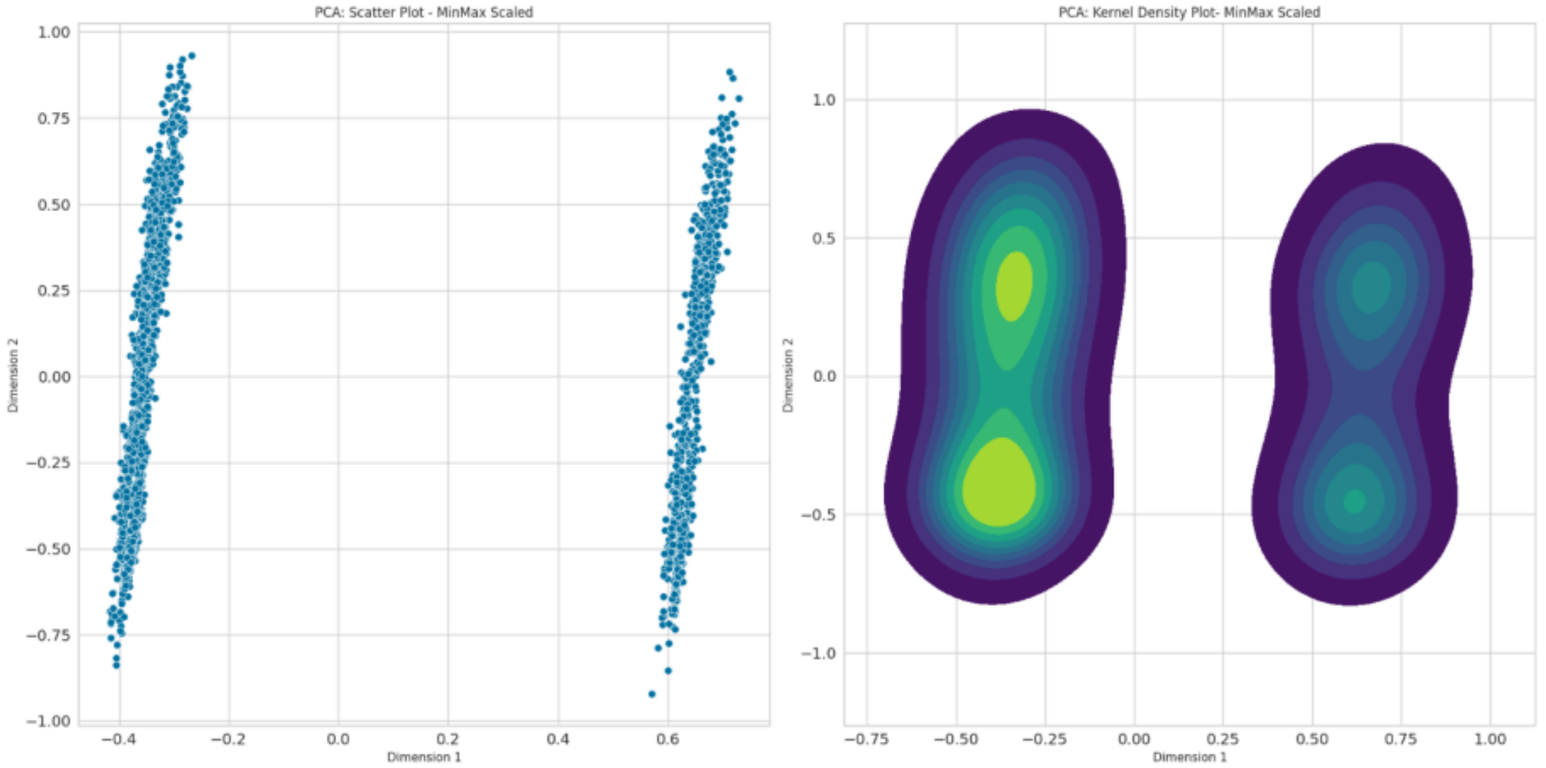


Figure III.1.1: Scatter plot and KDE plot for MinMax scaling

PCA: CHOOSE SCALING METHOD

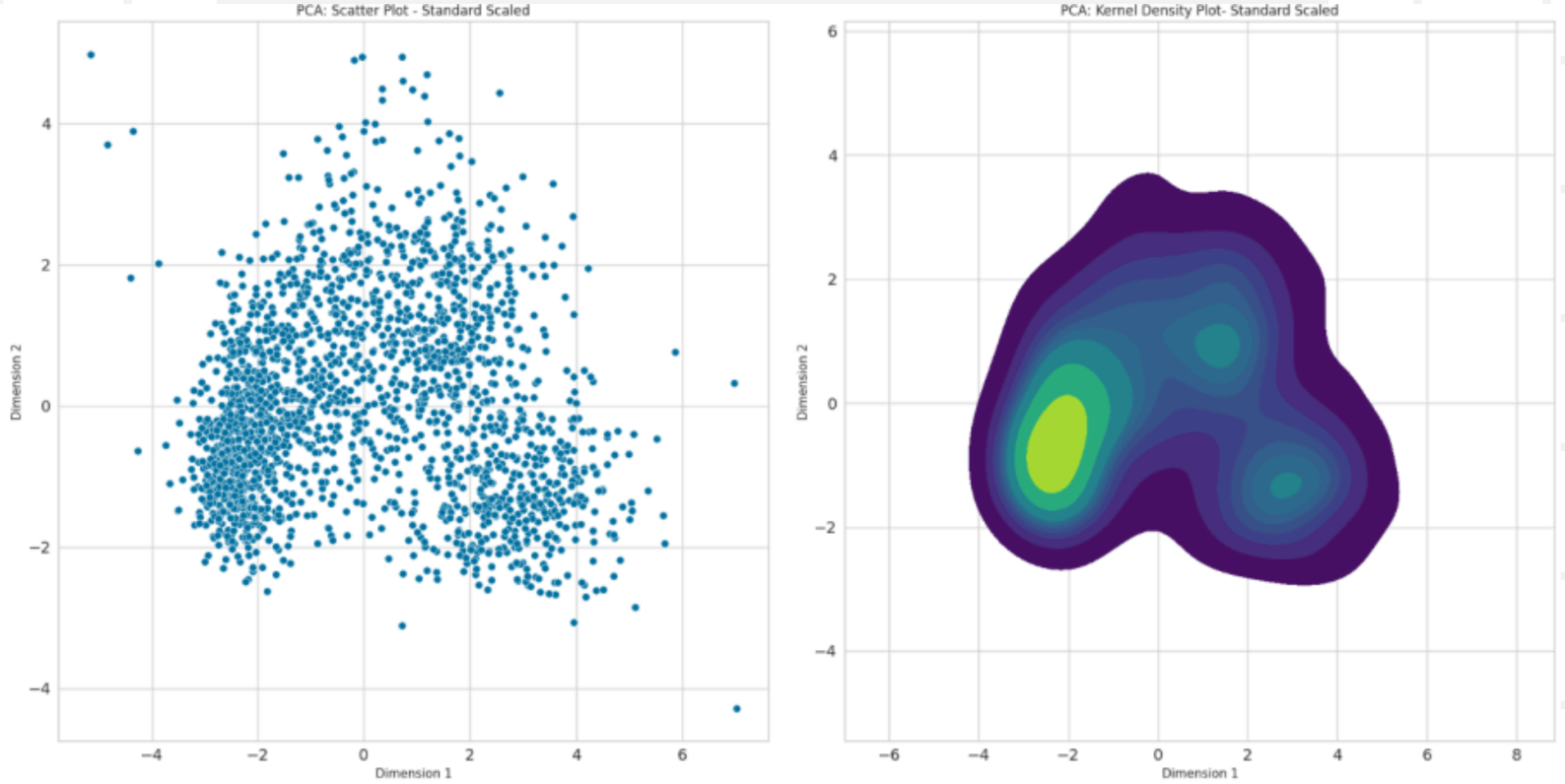


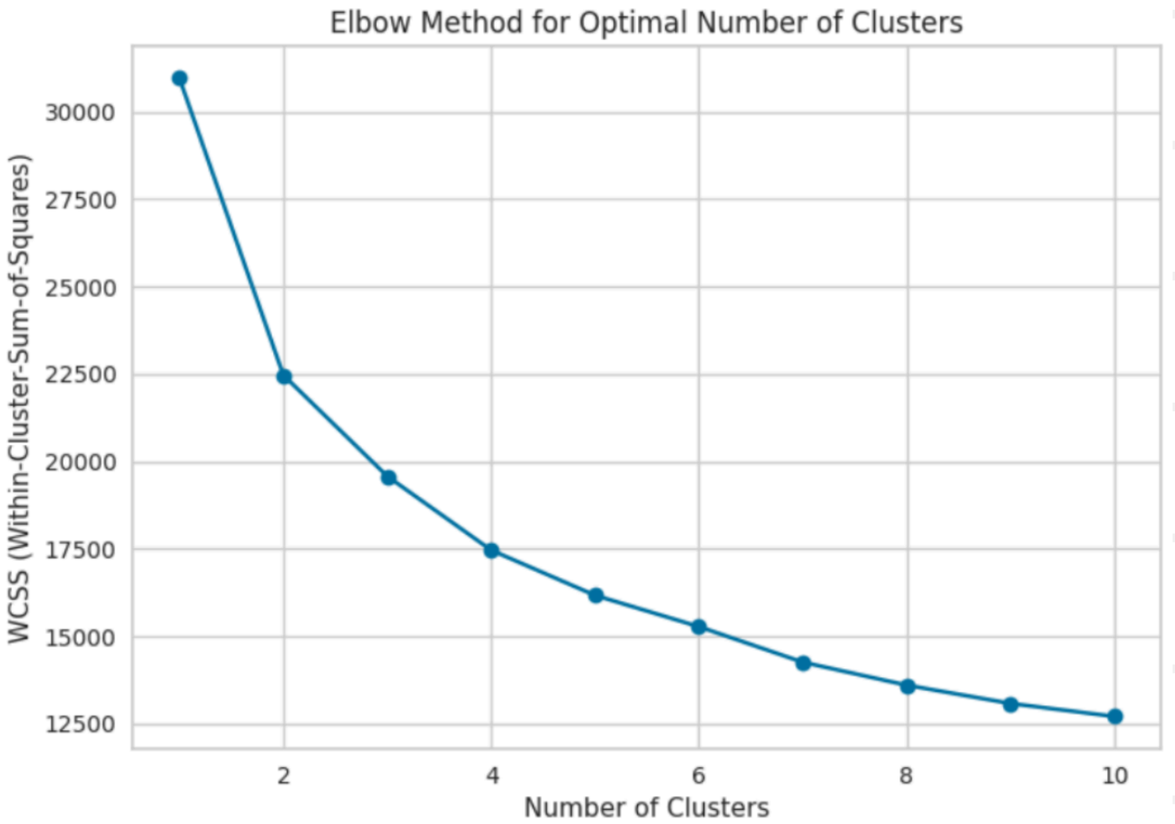
Figure III.1.2: Scatter plot and KDE plot for Standard scaling

PCA: CHOOSE SCALING METHOD

- **It seems MinMax scaling is unsuitable for this dataset, leading to two distinct clusters with a substantial distance of around 1. This discrepancy is attributed to the way MinMax scaling treats binary features, particularly by maintaining a constant distance of 1 between individuals in different categories.**
- **Therefore, I will use Standard Scaling because it provides a more balanced representation of the importance of features.**

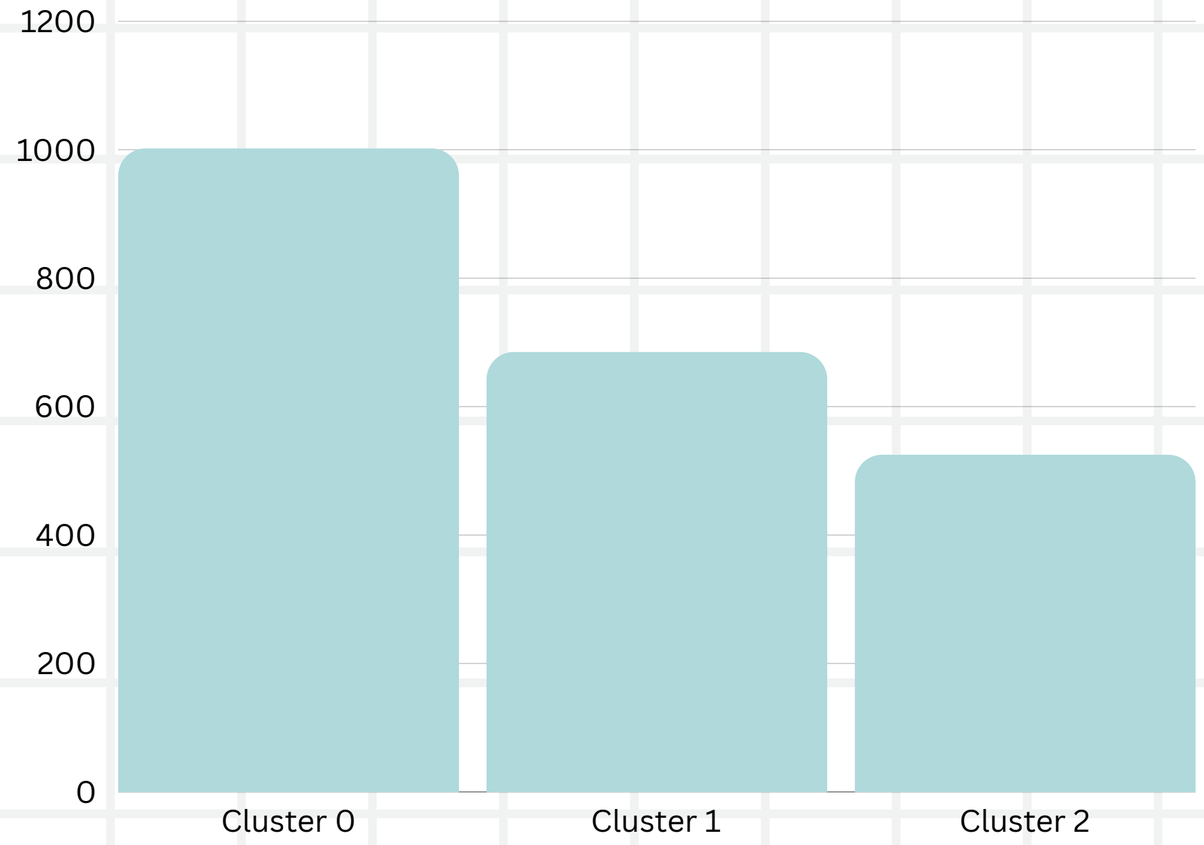
KMEANS: NUMBER OF CLUSTER

The chart suggests that there may be an optimal number of clusters for this dataset, as the Within – Cluster Sum of Squares value begins to plateau around three clusters.



RESULT

- **The bar chart indicates that K-means has successfully segmented the customer base into three groups of different sizes**
- **The cluster 0 has the most customers (around 1000), followed by the cluster 1 (around 685), and the cluster 2 (around 525)**



CONCLUSION

- **Accounting for 45% of the total, belong to cluster 0. They are predominantly younger individuals with lower incomes and have a higher percentage of new customers.**
- **Cluster 1 comprises 31% of customers and has a slightly higher proportion of long–term customers compared to other groups.**
- **Approximately 24% of customers fall into cluster 2, and they are the highest income earners among all customer segments.**

RECOMMENDATIONS

⚙ RECOMMENDATION 1

Focus on offering exclusive catalogs and campaigns through traditional channels

⚙ RECOMMENDATION 2

Emphasize affordability and family–friendly options

⚙ RECOMMENDATION 3

Promote loyalty programs and senior discounts to enhance customer retention and satisfaction

THANK YOU

Presentation by Thanh Hien