MACQUARIE
University
SYDNEY·AUSTRALIA

BUSA3020 ASSIGNMENT 3
CUSTOMER REPORT

PHUC LINH NGUYEN - 47542470

03 November, 2023

Contents

Introduction

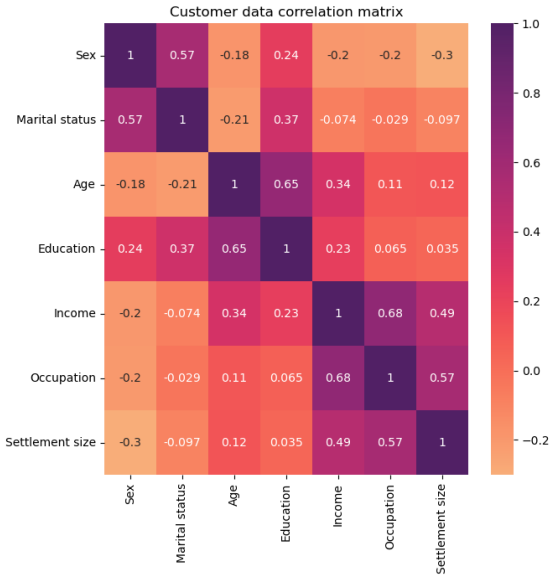
This analysis, performed on a dataset collected through loyalty cards utilized by 2000 customers, aims to establish a relationship between different information of the clients. In addition, this report also sheds light on the distinct customer segments (by k-means and hierarchical clustering) that comprise our clientele. I seek to identify common characteristics and behaviours that define these segments to provide valuable information and actionable insights that can be leveraged in tailoring marketing strategies, product offerings, and services for each customer segment. This report will serve as a guide to understanding the intricacies of our customer base and provide clear recommendations for optimizing our interactions with them.

Data summary

FEATURES	MEAN	MEDIAN	MODE
Sex	-	-	Male
Marital status	-	-	Single
Age	35.91	33	-
Education	-	-	High school
Income	120,954.42	115,548.5	-
Occupation	-	-	Skilled employee /official
Settlement size	-	-	Small city

1 Exploratory Data Analysis

1.1 Correlation Matrix



Age has a moderate positive linear relationship with Education (0.65). The same goes for Occupation - Income (0.68) and Occupation - Settlement size (0.57). Besides these correlations, a slight positive correlation exists between Settlement size and Income (0.49).

Figure 1: Customer data correlation matrix

1.2 Customers' Age and Marital status

Customers of the supermarket range from 18 to 88 years old, with people between 20 and 40 accounting for the majority. For single consumers, their median age is approximately 35, with the density focusing within the scope of 30 to 45. On the contrary, the non-single group has a lower median age of about 27. Most of them are in their 20s to 40s.

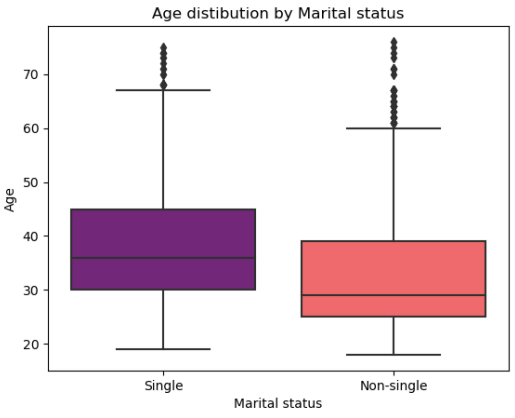


Figure 2: Customers' Age distribution by Marital status

1.3 Customers' Income, Education, and Occupation

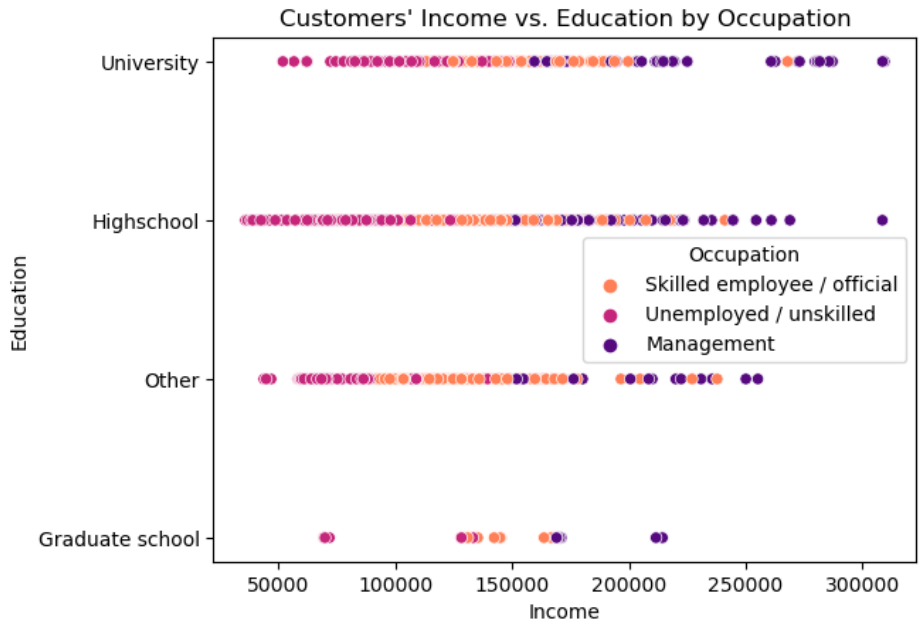


Figure 3: Customers' Income, Education and Occupation

The customers exhibit diverse occupations, subordinately from graduate schools, with a balanced spread. Notably, high school students are also part of your clientele. The majority earns between $50,000 and $200,000 annually, indicating accessibility across a middle to upper-middle-income range.

1.4 Customers' Gender and Settlement size

About 50% of the clients reside in small cities, and 35% of them are women. Small cities have a comparable amount of male customers—400 individuals—to mid-sized and large cities. However, there are only 300 clients overall in major and middle-sized cities, which is just half of it in small cities.

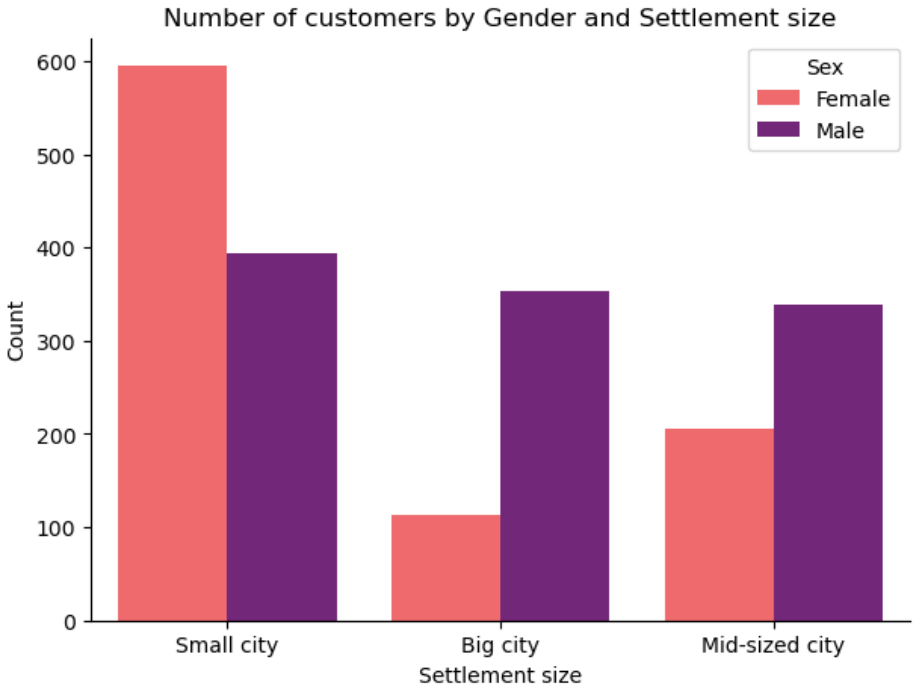


Figure 4: Customers' number by Gender and Settlement size

2 Customer Segmentation

I use two different methods in this section in order to classify the customers. They are k-means clustering and hierarchical clustering respectively.

2.1 K-means Clustering

Before applying the k-means clustering technique, I utilize the elbow method to choose the optimal k centroids for the process:

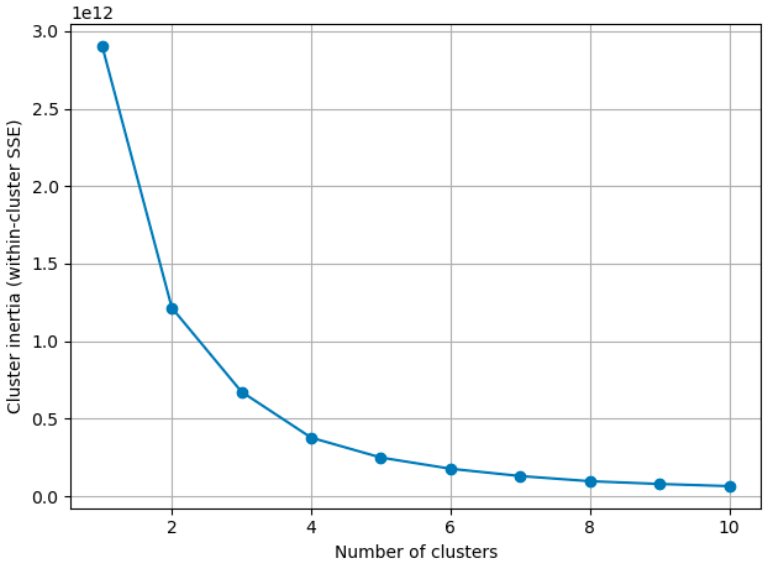


Figure 5: Elbow method

According to Figure 5, the optimal number of clusters is 4 as the inertia stops deteriorating rapidly from there. The customer clusters group by their Income and Age after employing k-means can be visualised as corresponding:

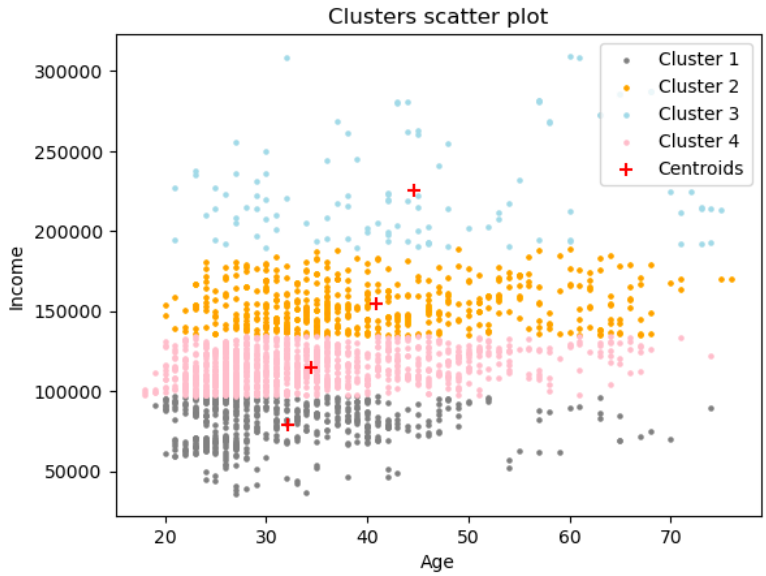


Figure 6: Customers' Clusters by Age and Income

2.1.1 Cluster 1

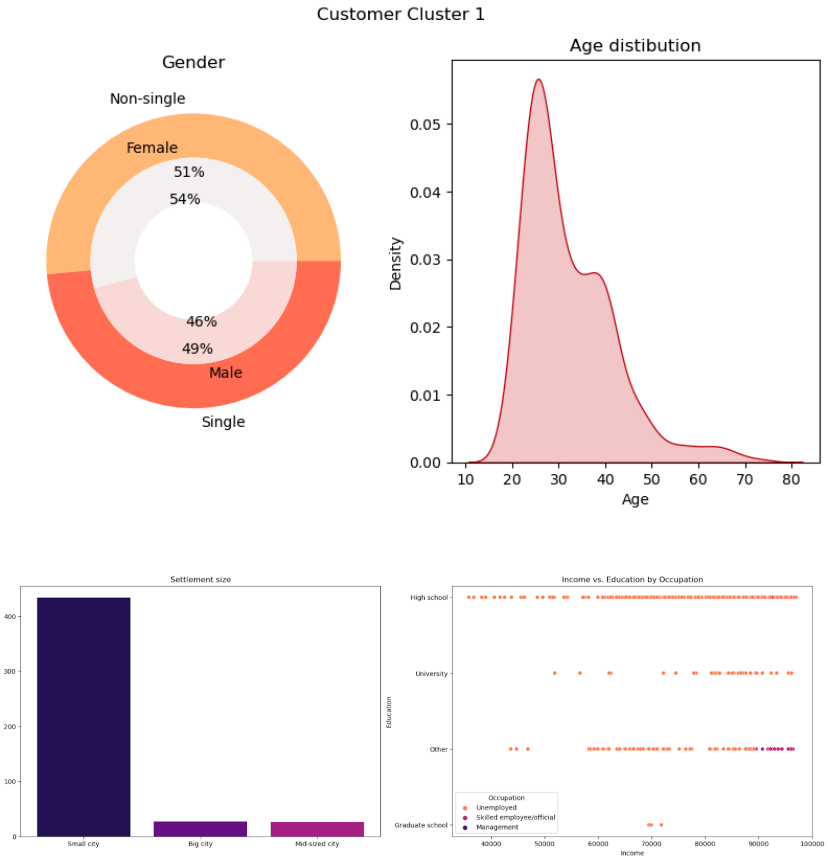


Figure 7: Customer Cluster 1

In the first cluster, the difference between females and males, single and non-single apportion is not considerable. Meanwhile, there is an apparent disparity between patrons' Settlement size and Occupation. There are 486 observations in Cluster 1, and more than 80% are out-of-work and staying in small cities.

2.1.2 Cluster 2

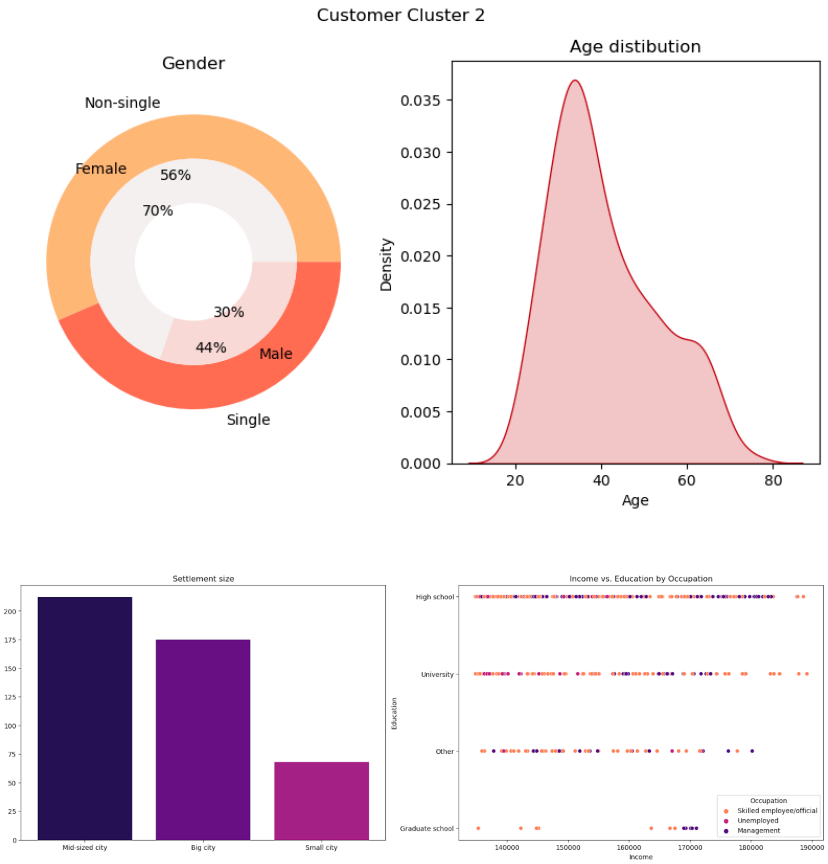


Figure 8: Customer Cluster 2

Customers in Cluster 2 are mostly non-single females in their adulthood. Unlike Cluster 1, the habitué distribution in Cluster 2 does not have significant dissimilarity in terms of Occupation and Settlement.

2.1.3 Cluster 3

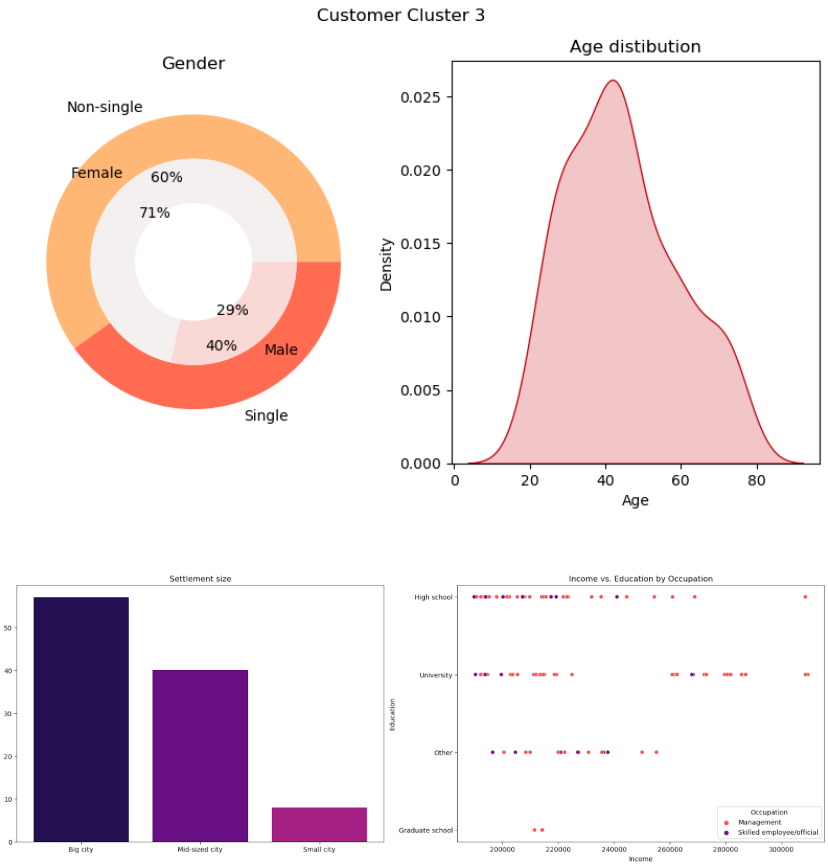


Figure 9: Customer Cluster 3

Cluster 3 shares many common traits with Cluster 2, which is called middle-aged non-single females. Nonetheless, Cluster 3 only has 105 subjects, while Cluster 2 has 455. Moreover, shoppers in this cluster do not dwell much in small cities.

2.1.4 Cluster 4

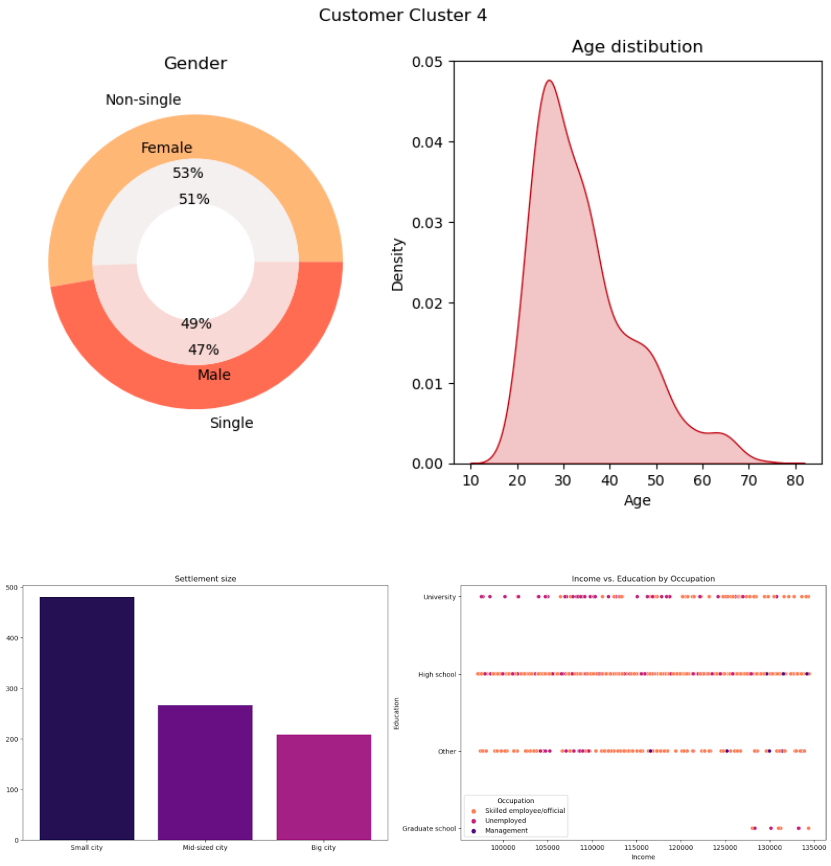


Figure 10: Customer Cluster 4

Cluster 4 mostly consists of skilled workers. Out of 4 groups, this class has the most number of individuals, 954. In addition, about half of them are 30 years old and settling in small cities. This collection seems to have the most balanced ratio of customers concerning Income, and Educational level among High school, University and Other.

2.2 Hierarchical Clustering

In the pursuit of understanding the intricate relationships and patterns within our customer base, we turn our attention to an agglomerative approach. This section delves into the world of hierarchical clustering and how it was employed to partition our 2000 customers into distinct segments based on shared characteristics. After implementing hierarchical clustering using ward linkage and Euclidean norm, we also obtained 4 clusters, which are demonstrated by the following dendrogram:

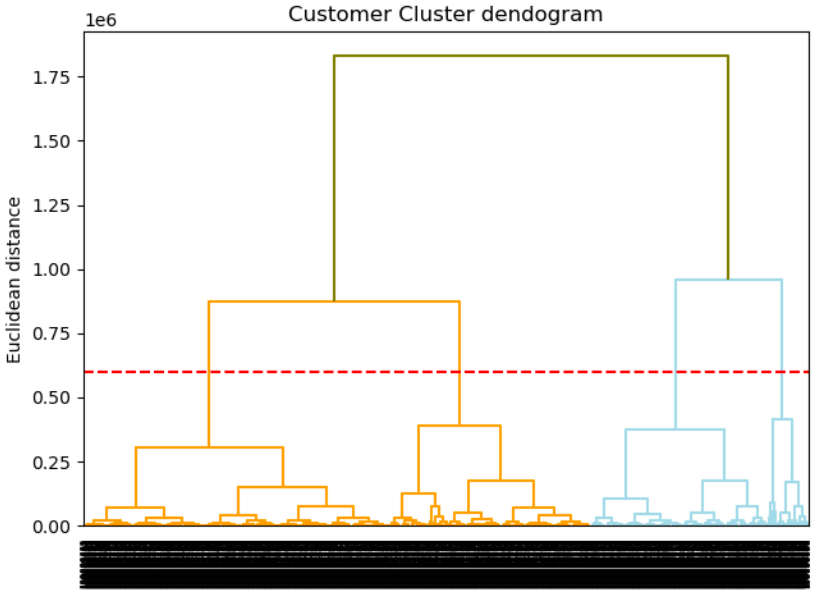
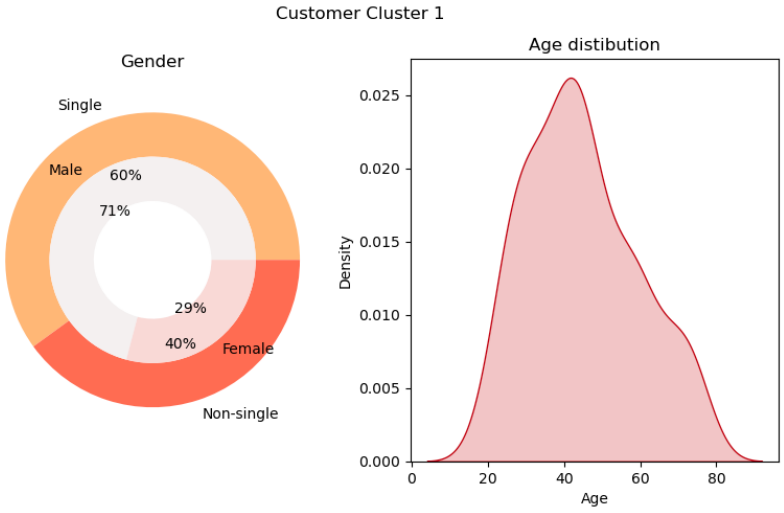


Figure 11: Customer Cluster Dendogram

2.2.1 Cluster 1



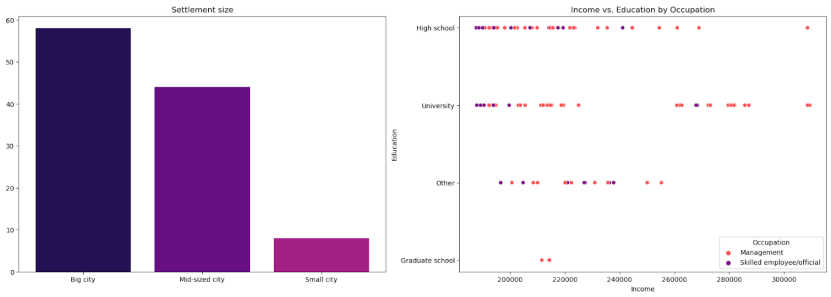


Figure 12: Customer Cluster 1

Punters in Cluster 1 are mostly single males in mid-sized or bigger cities. Remarkably, there is no unemployed customers in this group. This is comprehensible since there are only 110 entries in this group, which is equivalent to 5.5% of the total.

2.2.2 Cluster 2

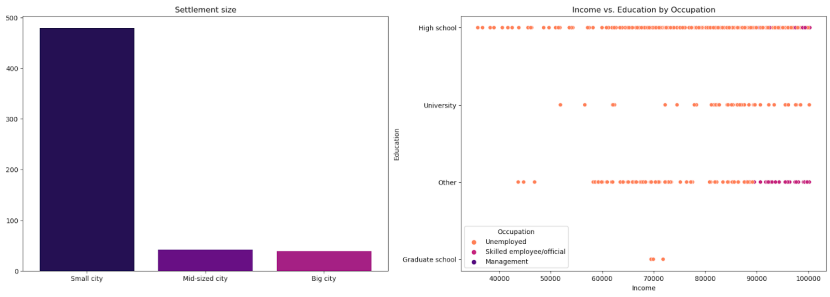
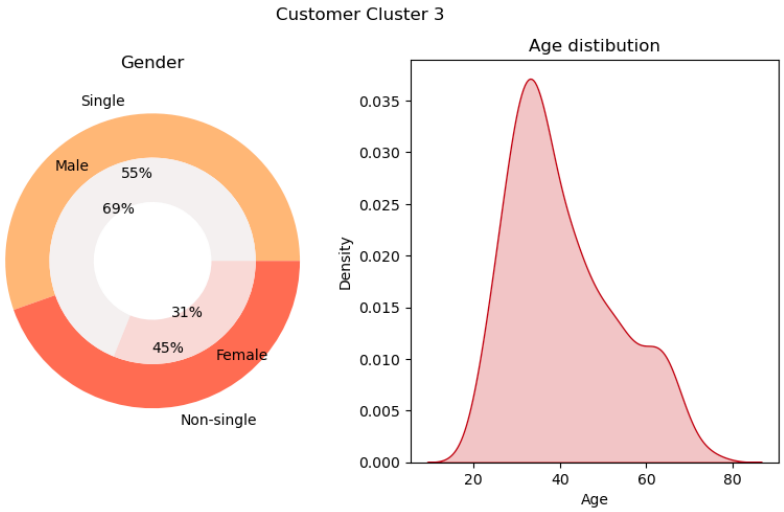


Figure 13: Customer Cluster 2

Cluster 2 is dominated by unemployed non-single females living in small cities. There is an enormous discrepancy between buyers' Settlement size and Occupation in Cluster 2, which is similar to Cluster 1 of the k-means Clustering method. Nearly 90% of them are jobless and staying in small cities.

2.2.3 Cluster 3



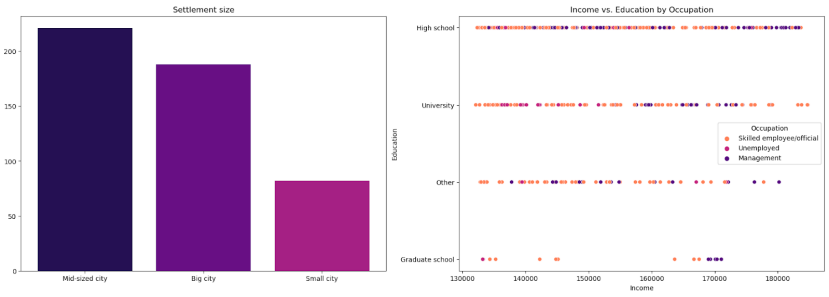
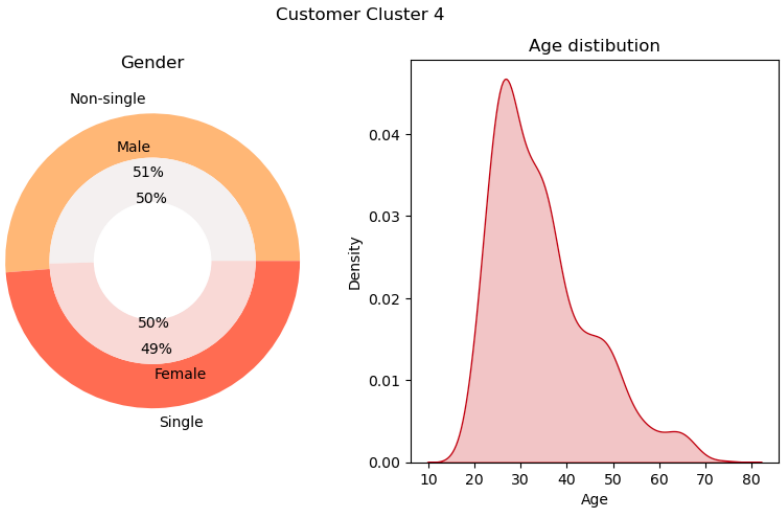


Figure 14: Customer Cluster 3

Cluster 3 bears resemblance to Cluster 1 to a certain extent. The majority of them are single males who are in their 30s. These people favor dwelling in middle-sized cities or bigger.

2.2.4 Cluster 4



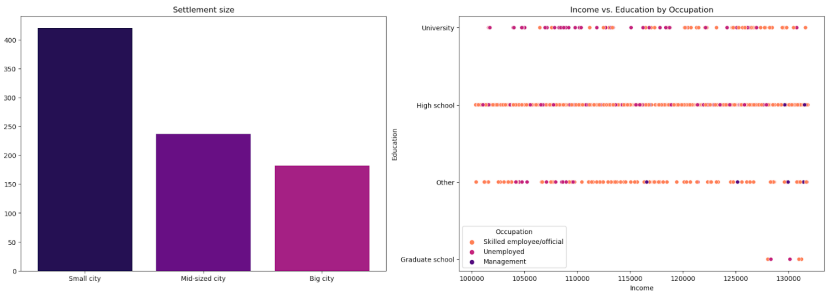


Figure 15: Customer Cluster 4

Last but not least, Cluster 4 is full of adults working as officials or skilled employees. There is little to no difference between females and males as well as single and non-single allocation. Moreover, these individuals prefer small cities. It is apparent that despite using two different methods, both Cluster 4 are surprisingly identical.

2.3 Clusters Summary*

	K-MEANS	HIERARCHICAL
Cluster 1	The unemployed in small cities	Single males in mid-sized & big cities (100% employed)
Cluster 2	Non-single females in their 30s	Unemployed, non-single females in small cities
Cluster 3	Middle-aged, non-single females in middle to big cities	Single males in their 30s
Cluster 4	Skilled workers	Skilled workers

* *Regarding customers' Education, any disproportion between graduate and other backgrounds is left out due to the fact that there is merely 1.8% of graduates in the original data.*

Marketing Recommendations

CLUSTER	STRATEGIES
Cluster 1 (k-means) & Cluster 2 (hierarchical)	- Offer affordable pricing and discounts on essential goods. - Enhance your customer loyalty program, rewarding regular shoppers with discounts, exclusive offers, or points that can be redeemed for groceries. - Create a community board within the supermarket where local businesses can post job opportunities. This not only supports the unemployed but also positions your supermarket as a hub for community information.
Cluster 2 (k-means & hierarchical)	- Introduce special promotions or discounts on family-oriented products. - Highlight time-saving features such as online ordering, curbside pickup, or home delivery services. - Tailor promotions around family-oriented holidays and seasons. Offer special deals on holiday meals, decorations, and gifts.
Cluster 3 (k-means)	- Highlight premium and gourmet product selections. Create dedicated sections for specialty items, organic produce, and high-quality brands. - Collaborate with local fashion or lifestyle brands to offer discounts on fashion items or exclusive access to events in partnership with these brands. - Integrate wellness and beauty sections. Offer a curated selection of health foods, supplements, and beauty products.

CLUSTER	STRATEGIES
Cluster 4 (k-means & hierarchical)	- Create grab-and-go meals, salads, and snacks for those with limited time for meal preparation. - Integrate technology to enhance the shopping experience. Implement features such as mobile apps for easy ordering, self-checkout options, and loyalty program management through digital platforms. - Allow them to submit grocery lists in advance, and have their orders ready for pickup or delivery at a designated time.
Cluster 1 & 3 (hierarchical)	- Design promotions and sections in the supermarket that cater to health-conscious single males. Offer a variety of fresh produce, lean proteins, and gym-friendly snacks. - Designate specific nights as "Singles Night" with exclusive discounts, promotions, or social events. This can create a sense of community and attract singles looking for a social atmosphere. - Create promotions or discounts that pair snacks with popular beverages like beer. This can attract single males who are looking for convenient and enjoyable options for solo dining. - Collaborate with tech brands or gadget retailers to offer promotions or exclusive discounts on electronics and gadgets. Appeal to the tech-savvy interests of single males.

Conclusion

Through exploratory data analysis and the application of clustering techniques, we have unveiled the rich tapestry of our customer segments, each with its unique characteristics and preferences. By categorising our cus-

tomers into meaningful segments, we have gained a profound understanding of their needs and expectations. This, in turn, equips us to tailor our marketing strategies, product offerings, and services to cater to these distinct groups.