Customer Segmentation Analysis Report

BUSA 3020

Minh Chau Nguyen - 46233334

Introduction

Nowadays, customer segmentation analysis is fundamental to businesses, since one business need to fully understand their customers to make any marketing or business decision. For this report, a data set that contain information of 2000 customer has been collected and analysed. Python has been used to performed k means clusters and hierarchy clusters, thus, indicating the customer segmentation of the business. Moreover, both clustering techniques will be compared to identify the differences. From the results, customers profile of each cluster will be documented and further be used as resources for marketing techniques.

Exploratory Data Analysis

	Sex	Marital status	Age	Education	Income	Occupation	Settlement size
0	0	0	67	2	124670	1	2
1	1	1	22	1	150773	1	2
2	0	0	49	1	89210	0	0
3	0	0	45	1	171565	1	1
4	0	0	53	1	149031	1	1
...	...	...	...	...	...	...	...
1995	1	0	47	1	123525	0	0
1996	1	1	27	1	117744	1	0
1997	0	0	31	0	86400	0	0
1998	1	1	24	1	97968	0	0
1999	0	0	25	0	68416	0	0

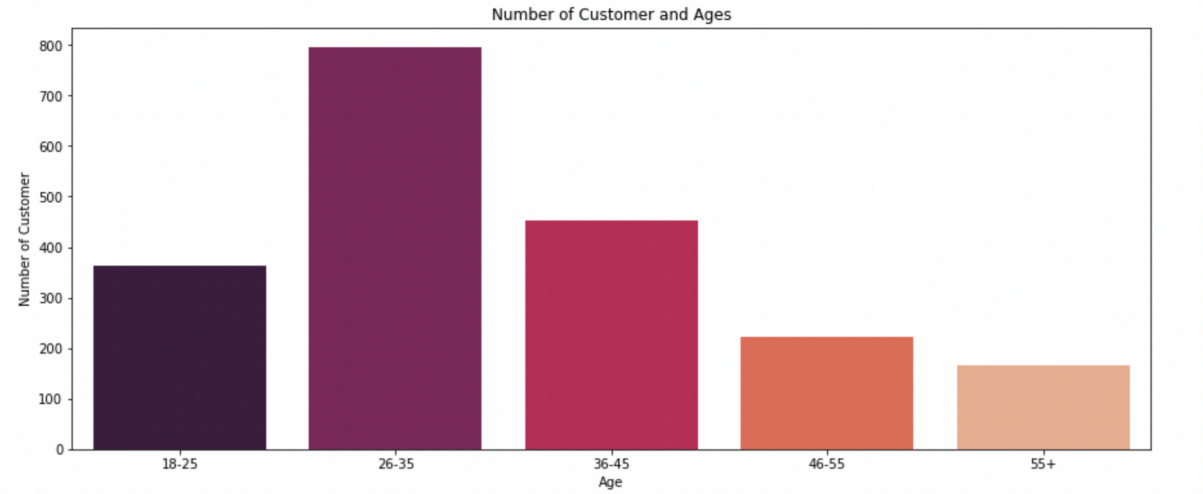
2000 rows × 7 columns

The table above indicates that there are 7 columns represents for 7 information that we have on the customer.

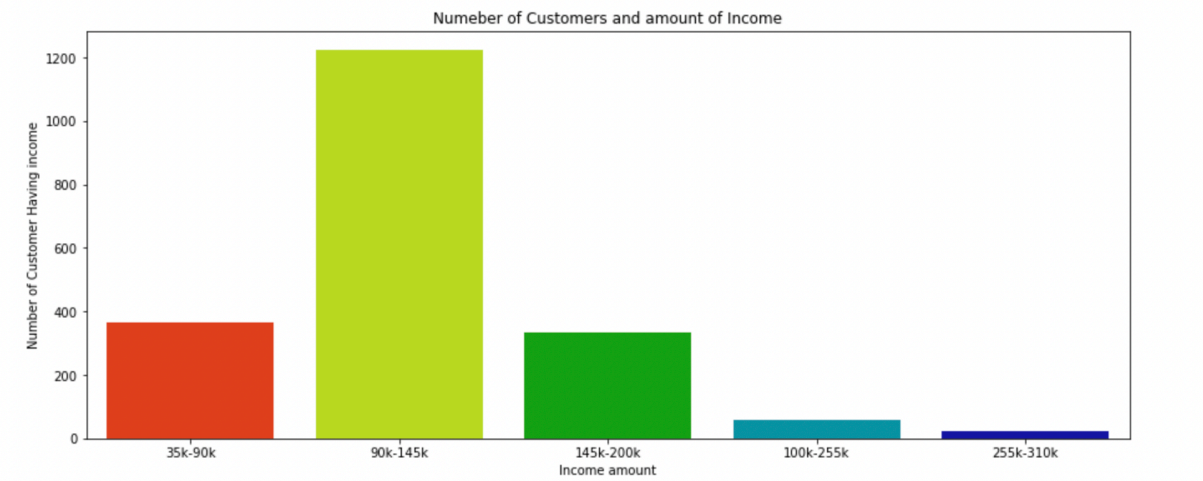
```
Number of Male Customers: 914 customers
Number of Female Customers: 1086 customers
-----------------------------------
Number of Unemployed Customers: 633 customers
Number of Skilled Employed Customers: 1113 customers
Number of Managament Customers: 254 customers
-----------------------------------
Number of Single Customers: 1007 customers
Number of Non-single Customers: 993 customers
-----------------------------------
Unknown education : 287 customers
Highschool level : 1386 customers
University level: 291 customers
Graduate: 291 customers
-----------------------------------
Customers live in small city: 989 customers
Customers live in Mid-size city: 544 customers
Customers live in Big city: 544 customers
```

As statistics have shown, the major customer of the supermarket chain is Female with around 54%, and Male around 46%. Moreover, of 2000 customers, over half of them are skilled employed customers and only 254 customers are management or self-employed, whereas 633 customers are unemployed. Another significant feature of the data set is that nearly 70

percent (1386 customers) of the chain have a high school as their level of education and the number of customers has other level of education ranking from unknown to university and graduate are roughly equal. From the last session of the statistics above, most of the customers live in a small city.



From the graph above, it is obvious that most of the current customers belong in the age group of 26 to 35 years old which is adult age. This group make up to nearly 40 percent of the group. The second largest age group is 36 to 45 years old which make up around 25 percent of customer. The smallest group is old age from 55 above, this customers group only contribute around 10%.



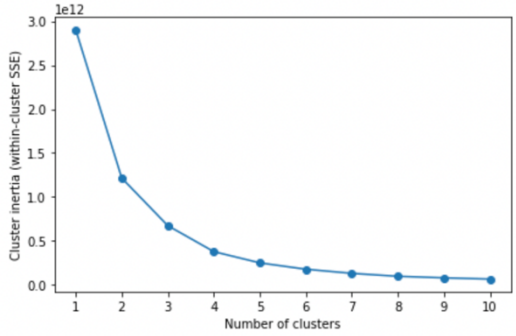
From the income graph, the largest group of customers which is 60 % of the group earn around 95k to 145k per year. Follow by group of income from 35K to 90k per year and 145k to 200k per year which are around 20% and the least amount of customer belong to the highest income group which is 255k to 310k annually.

Customer Segmentation

There are multiple features of the customer in the data. In this part, cluster analysis which means grouping a set of customers in such a way that customers in the same group (cluster) are more like each other than to those in other groups (clusters).

K-means Clustering

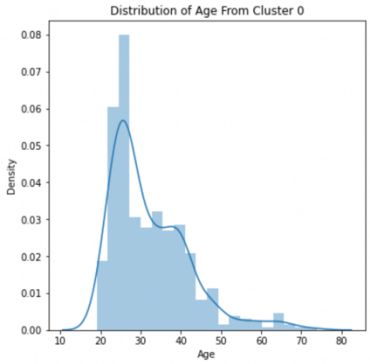
Firstly, K-means clustering is applied on the dataset, this method divider data into K clusters in a way that data in the same cluster has close distance together and data in different clusters are further apart.

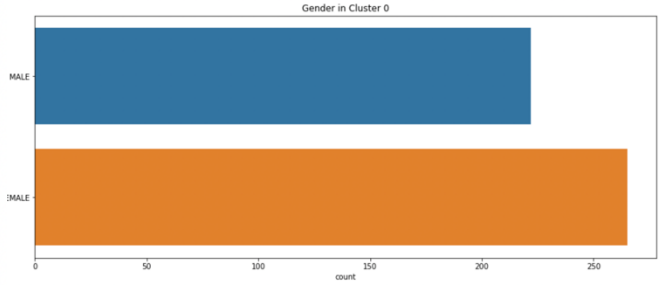


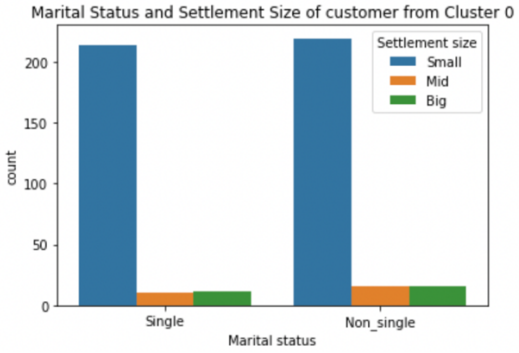
From the graph above, which shows number of clusters and their cluster inertia, the optimal number of cluster K is 4 clusters. In other word, 4 clusters are the amount that maximize the average silhouette from all value of k ranging from 1 to 10.

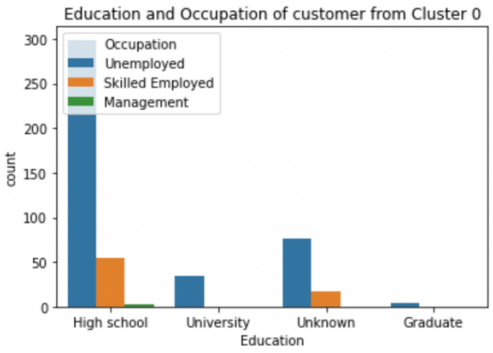
Interpret identified clusters in terms customer profiles

Cluster 0



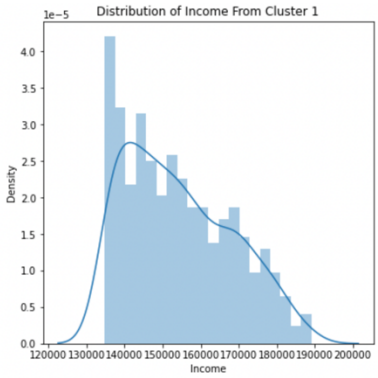


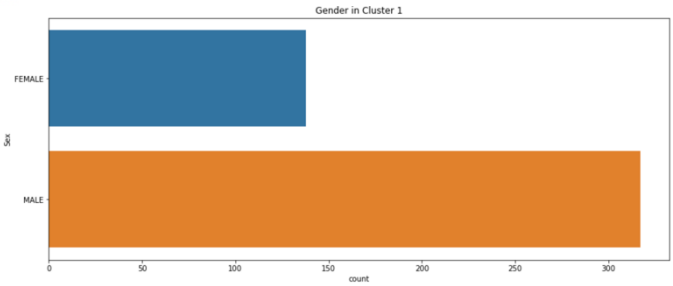


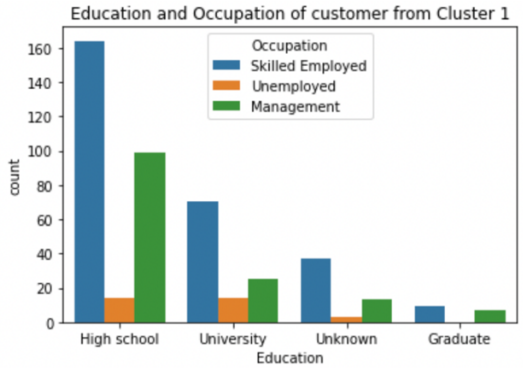


From all the figures we can see that cluster 0 has mostly young adult around the age of 20 to 25 years old who are mainly high school and unemployed. The income of these customers is below 100000 annually which is on the low side compared to the rest of the data set. Also, this customer group live commonly in small city, and it is male dominant.

Cluster 1

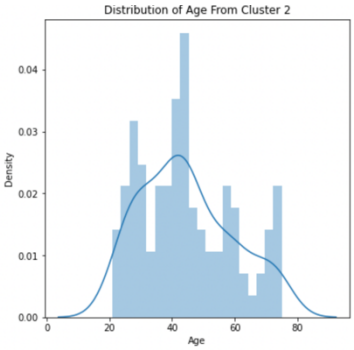


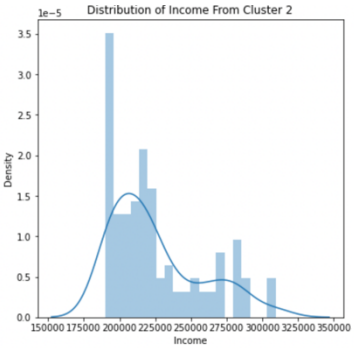


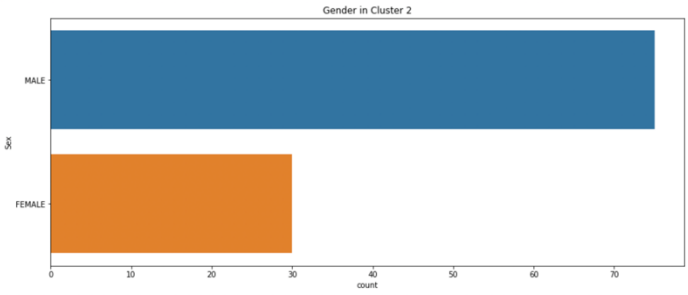


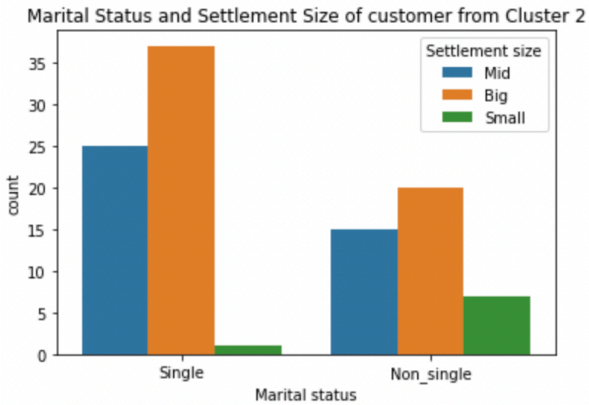
This cluster is the biggest clusters. Also, it's male dominated and they generally in the late adults age (30-40 years old). Most of them are skilled employed that live in middle size city or big size if they are single. Their income is ranging from 140000 to 190000.

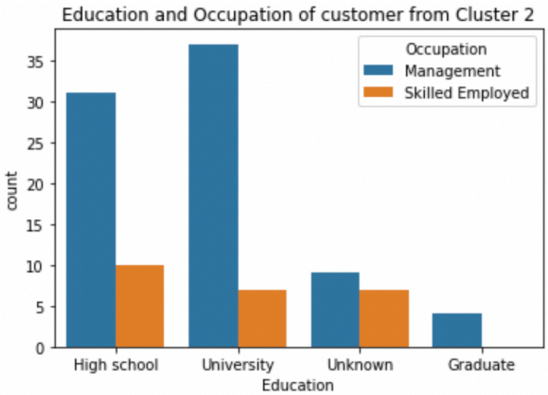
Cluster 2





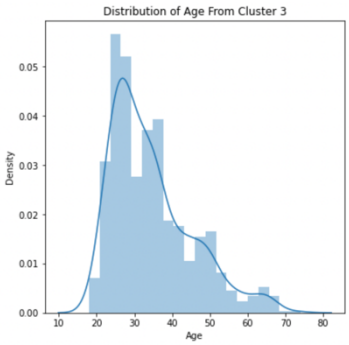


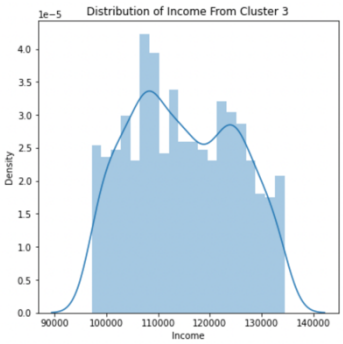


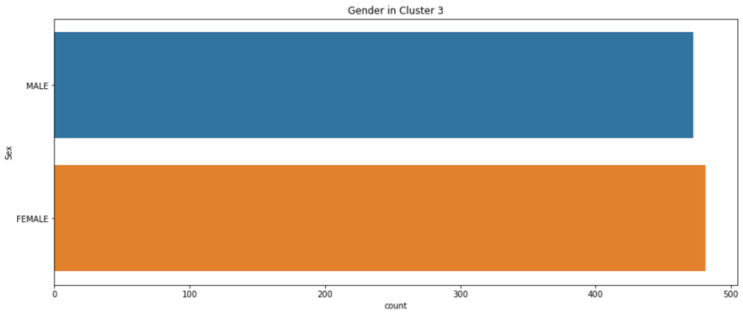


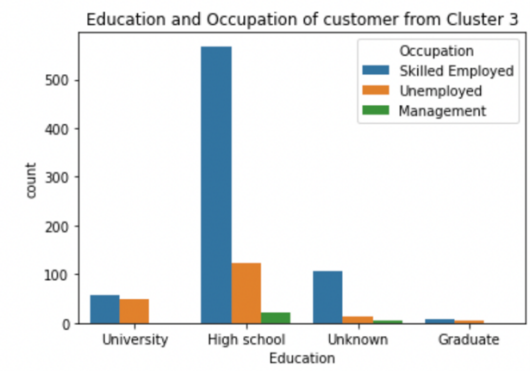
Cluster 2 is mostly male customer from and over the age of 40. They earn high income, live in big city and their occupation level is mainly management.

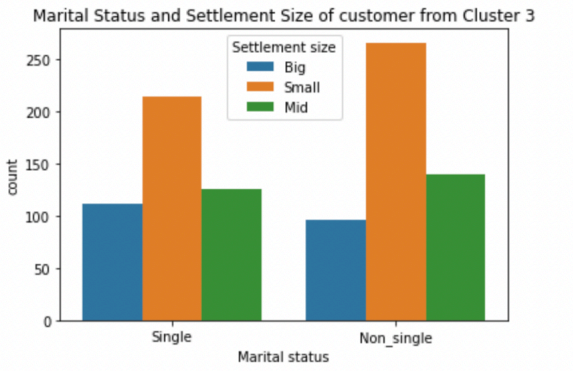
Cluster 3





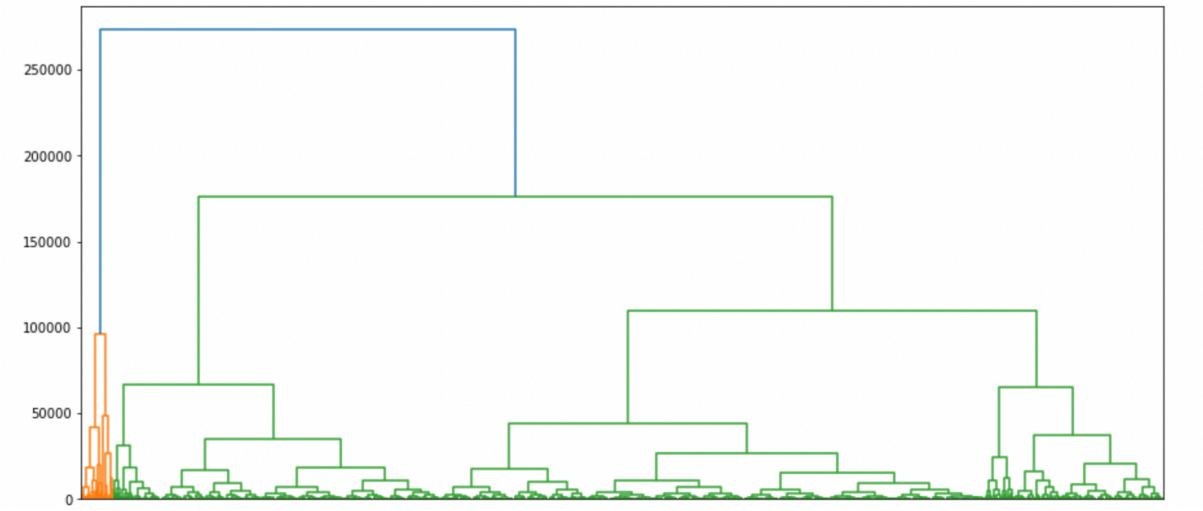






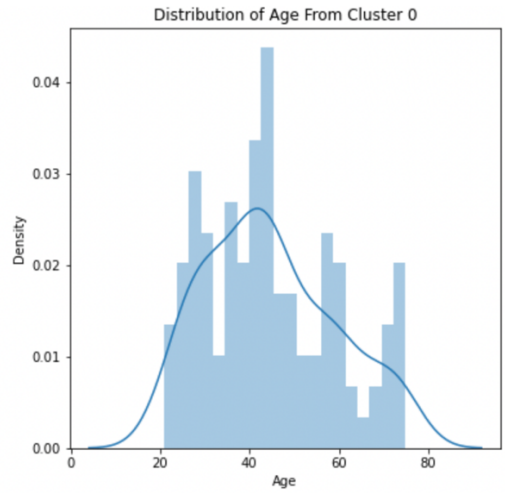
This biggest group (800 customers) is adult customers with average income compared to the dataset. Most of them are high school, skill employed and live in small city.

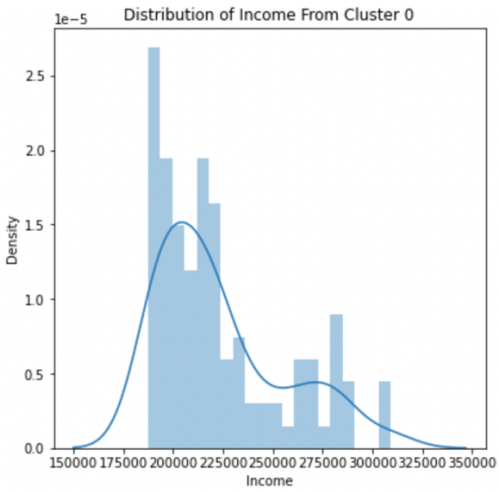
Clustering Data using hierarchical clustering



Hierarchical clustering suggest that the optimal number of clusters is 4

Cluster 0

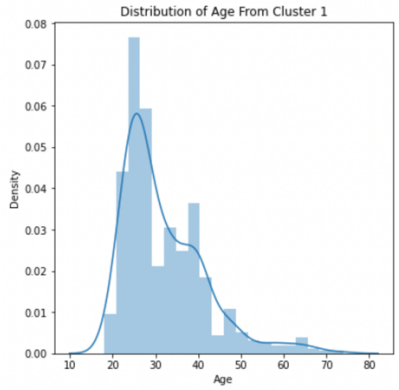





```
Number of Male Customers: 32 customers
Number of Female Customers: 78 customers
-----------------------------------
Number of Unemployed Customers: 0 customers
Number of Skilled Employed Customers: 28 customers
Number of Managament Customers: 82 customers
-----------------------------------
Number of Single Customers: 66 customers
Number of Non-single Customers: 44 customers
-----------------------------------
Unknown education : 16 customers
Highschool level : 44 customers
University level: 46 customers
Graduate: 46 customers
-----------------------------------
Customers live in small city: 8 customers
Customers live in Mid-size city: 44 customers
Customers live in Big city: 44 customers
```

This group is mostly female customers around 40-60 years old who have high income and education level. They mainly live in middle size or big size city.

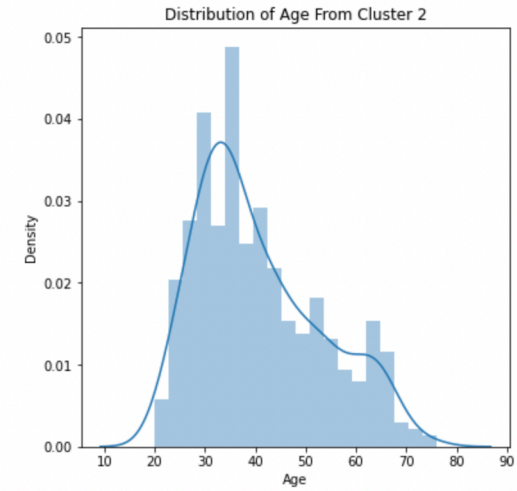
Cluster 1

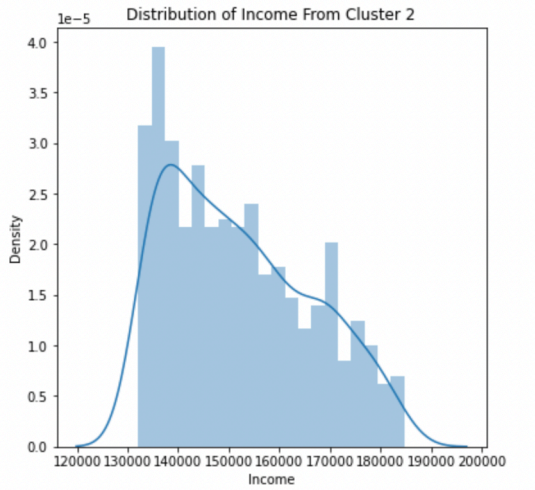



```
Number of Male Customers: 313 customers
Number of Female Customers: 247 customers
-----------------------------------
Number of Unemployed Customers: 432 customers
Number of Skilled Employed Customers: 122 customers
Number of Managament Customers: 6 customers
-----------------------------------
Number of Single Customers: 260 customers
Number of Non-single Customers: 300 customers
-----------------------------------
Unknown education : 103 customers
Highschool level : 415 customers
University level: 38 customers
Graduate: 38 customers
-----------------------------------
Customers live in small city: 479 customers
Customers live in Mid-size city: 42 customers
Customers live in Big city: 42 customers
```

This is the second biggest group which is mainly young adult who earn low income and have high school level of education or unknown. They commonly live in small city.

Cluster 2

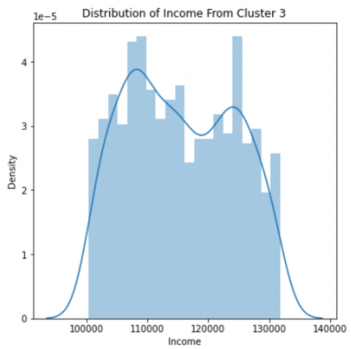





```
Number of Male Customers: 153 customers
Number of Female Customers: 338 customers
-----------------------------------
Number of Unemployed Customers: 35 customers
Number of Skilled Employed Customers: 309 customers
Number of Managament Customers: 147 customers
-----------------------------------
Number of Single Customers: 272 customers
Number of Non-single Customers: 219 customers
-----------------------------------
Unknown education : 61 customers
Highschool level : 296 customers
University level: 112 customers
Graduate: 112 customers
-----------------------------------
Customers live in small city: 82 customers
Customers live in Mid-size city: 221 customers
Customers live in Big city: 221 customers
```

This cluster is mainly female around the age of 30-40 years old who earn middle range income from 140000 to 190000 annually. Most of the have high school education or higher, they likely to be skill employed and live in middle or big size city.

Cluster 3



```
Number of Male Customers: 416 customers
Number of Female Customers: 423 customers
-----------------------------------
Number of Unemployed Customers: 166 customers
Number of Skilled Employed Customers: 654 customers
Number of Managament Customers: 19 customers
-----------------------------------
Number of Single Customers: 409 customers
Number of Non-single Customers: 430 customers
-----------------------------------
Unknown education : 107 customers
Highschool level : 631 customers
University level: 95 customers
Graduate: 95 customers
-----------------------------------
Customers live in small city: 420 customers
Customers live in Mid-size city: 237 customers
Customers live in Big city: 237 customers
```

This is the biggest young adult group (800 customers) with middle range income who is likely to be skilled employed and have high school education level. They tend to live in small city

How do the clusters identified by the two techniques compare?

Firstly, both techniques suggests that 4 is optimal number of clusters. With K-means cluster, the group is usually male dominated while with hierarchical clustering, the group is likely to be female dominated. Other features are quite similar between two techniques. For example, if the group is young age, then they likely to have lower level of education and income, thus living in a smaller settlement size or vice versa.

Recommendations

Both techniques suggest that the biggest group of customers that the chain has is young customers living in small cities with average income. Therefore, the most suitable marketing for this group should be social media advertisement or affiliate marketing. This is because young people heavily interact with social media daily so it would be effective to attract attention to the product or service of the chain.

With the other group, which is older, earning a higher income, the company should have other strategies to apply to them such as email advertising or phone call. This is because they have limited time to interact with online advertisements and the suggested strategies could directly give them necessary information about the product. Moreover, if the group is more male, they tend to care more about sports or other male-related product, whereas female clients would have their preferred interests. Based on that, the company can determine what product or service to introduce.

Conclusion

Overall, in this report, K-means clustering, and hierarchical clustering were utilized to determine different customer groups of the blockchain. The groups that were suggested by both techniques are only slightly different in the profile features of each customer group, especially in their gender. From the results, different marketing strategies such as online marketing or telephone marketing have been suggested for the suitable customer group. By developing marketing strategies based on customer profiles from customer segmentation analysis, it's easier able to make accurate and effective decisions.