**MARKETING ANALYSIS**

**DUY PHAM – 47522003**

1. **Introduction**

A small clothing store is trying to make a marketing plan, but the marketing team doesn’t really know its customers. To decide which customer segment the store should focus on, they gathered some basic information about their customers and tried to find some insight.

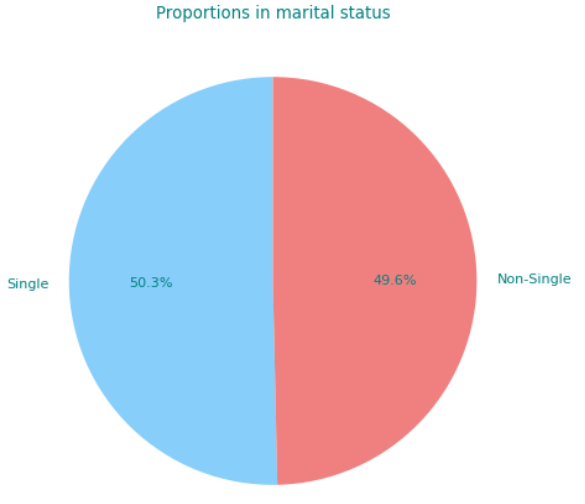
1. **Data exploratory**

* 2000 customer information is gathered, including Gender, Marital status, Age, Educational level, Income, Occupation, and Settlement size (where they live)
* **Gender**: It is not much difference in the number of Male and Female customers. The proportion is 54.3% for Males and 45.7% for Females

Chart, pie chart

Description automatically generated

* **Marital status**: The number of single and non-single customers is equal with 50.3% single and 49.6% non-single customers.



* **Education**: We can see a big difference in customers’ educational levels. 69.3% finished High School, 14.6% are Undergraduate, and 1.8% are from Graduate School
* Chart, pie chart

  Description automatically generated
* **Occupation**: The majority of the customers are Skilled Employees or Officials, accounting for 55.7%. Followed by Unemployed or Unskilled at 31.7% and Managers or Highly qualified employees at 12.7%.

Chart, pie chart

Description automatically generated

* **Age**: Bellow is the Age distribution. We can see that our customers are mostly around 25 to 40 years old

Chart, histogram

Description automatically generated

1. **Customer Segmentation**

In this report, we conduct two clustering method to find out the customer segments, which is k-Mean ++ and Agglomerative clustering. Before clustering, we should find out how many segments we want. The Elbow Method is used to find out the optimal number of clusters we should have from the dataset.

Chart, line chart

Description automatically generated

**The graph above** shows with the increasing number of clusters, the difference between customers in each cluster decreases. We can see that the optimal number of segments is between 3 to 4 as from 5 and so on, the reduction in differential is not that significant. In this campaign, we will divide the dataset into 3 clusters.

Chart, pie chart

Description automatically generated

**The two pie charts above** show the number of customers in each cluster. From the graph, we can obviously see that our target segment is Cluster 0 with the Agglomerative clustering method or Cluster 1 with the k-Mean++ method. To find out who is in cluster 0 and cluster 1 that we are focusing on, let’s look deeper into the characteristics of these segments.

Firstly, we will look at the average income and average age of each cluster using the Agglomerative clustering method

Chart, bar chart

Description automatically generated

**Agglomerative clustering method**

Cluster 0 contains customers in middle age, nearly 40, and with middle income (compared to other customers in the dataset), at around $130000. Cluster 1 contains customers with the highest age, around 45, and with the also highest income, at nearly $250000. The last cluster contains the youngest, at just above 30 with around $70000 income.

The result is quite predictable the higher their age, the higher income they have. We divided them into 3 groups lower, middle, and higher.

Now let’s see the clusters we have using the k-Mean++ method.

Chart, bar chart

Description automatically generated

**k-Mean++ method**

We can see that while the cluster labels are different, the characteristics of the three groups are the same as the other method. Cluster 1 of the k-Mean++ method and Cluster 0 of the Agglomerative clustering method are the same group of customers.

Next step, apart from Age and Income, we will investigate other characteristics of the customers in these groups. To find out this information, we get the most frequent value in each segment.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Clusters | Sex | Marital status | Education | Occupation | Settlement Size | Age | Income |
| 0 | Male | Single | High School | Skilled Employee/Official | Small City | Middle | Middle |
| 1 | Male | Single | University | Management/Self-employed/Highly qualified employee/Officer | Big City | High | High |
| 2 | Female | Non-Single | High School | Unemployed/Unskilled | Small City | Low | Low |

**Agglomerative clustering method Clusters Summary**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Clusters | Sex | Marital status | Education | Occupation | Settlement Size | Age | Income |
| 0 | Male | Single | High School | Skilled Employee/Official | Mid-sized City | Middle | Middle |
| 1 | Male | Single | High School | Management/Self-employed/Highly qualified employee/Officer | Big City | High | High |
| 2 | Female | Non-Single | High School | Unemployed/Unskilled | Small City | Low | Low |

**k-Mean++ method Clusters Summary**

For better readability, I’ve changed the Clusters’ name for k-Mean++ to be similar to the Agglomerative clustering method. We can see that while there are some minor differences, these methods provide quite similar clusters.

The segment we should focus on is the male, single, high school, skilled employee/official at the age of 38, living in Small or Mid-sized cities and with Income at around $130000. The next segment we should focus on is female, non-single, high school, unemployed, at the age of 31, living in small cities with income around $70000.

1. **Recommendation**

In general, we mostly served customers who graduated from high school, the age of under 40, and with $130000 or lower income. The clothing style should be casual, comfortable, and affordable.

These customers with not-so-high incomes are quite sensitive to the product’s price and sales offs.

While the proportion of males and females is quite equal, we should maintain the balance of our products for both of these genders or might even focus on unisex clothing.

1. **Conclusion**

From the dataset, we can see the summary of our customer profiles. Agglomerative clustering and k-Mean++ methods are used to segment the customers. The Elbow method is used to find the optimal number of segments. From the clustering methods, we can portray the target customers and provide some marketing suggestions for the store.