PROGRAMMING TASK 3 – CUSTOMER SEGMENTATION

Cam Tu Pham

Student ID: 46864598

Master of Business Analytics

Macquarie University

1. Introduction:

Customer segmentation is a critical process for business to understand each group of customers' needs, preferences, and behaviors. Therefore, business can develop the effective marketing strategy and campaign to each customer segment.

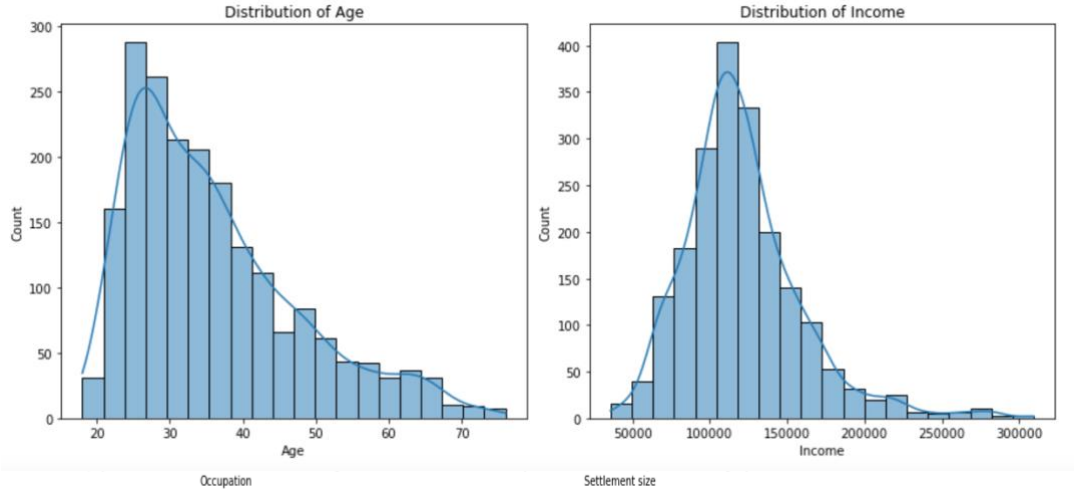
In this report, we will explore a dataset of 2,000 customers collected through loyalty cards. Using this dataset, we will perform customer segmentation analysis. The goal is to identify distinct customer segments based on their attributes such as gender, age, marital status, income, education, occupation and the size of the city that the customer lives in. This customer segementation analysis is expected to provide marketing strategy recommendations to the business management on how to target these customer segments effectively.

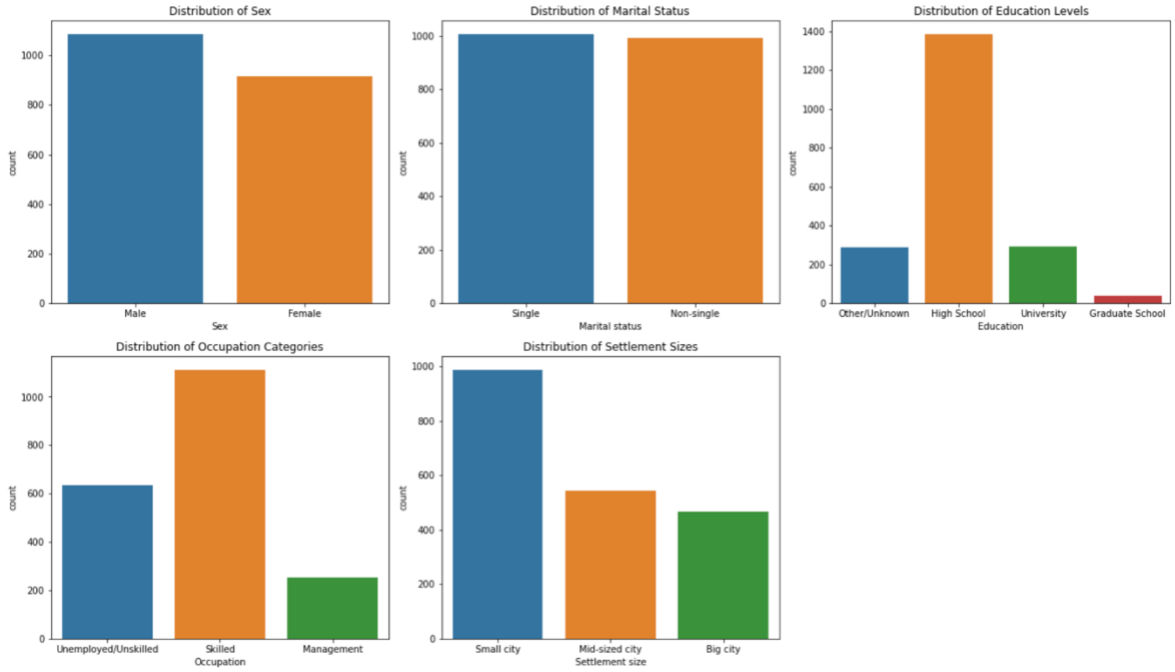
2. Exploratory Data Analysis:

First of all, we has performed the visualisation to have a overview of the group of 2,000 customers.

- **Gender**: There is a fairly balanced representation of males and females.
- **Marital Status**: A number of customers are single is quite similar to those who are non-single.
- **Age**: The age of customers ranges are widely from 18 and 76 years old, with a average age of 36 years old.
- **Income**: Most customers have income range from $35K *to* $309K. The average income is approximately $120K.
- **Education**: Majority of customers complete high school. 15% out of 2,0000 customers have university-level education, and 14% out of 2,0000 customers fall under the "Other/Unknown" category.
- **Occupation**: The largest group are skilled employees or officials. The second largest group is unemployed or unskilled. The dataset also includes customers who work in management level, self-employed and high quality employees
- **Settlement** (size of city where customer lives in): The majority of customers come from small cities, followed by mid-sized cities, and then big cities.

Chart 1: Overview of the customer dataset:



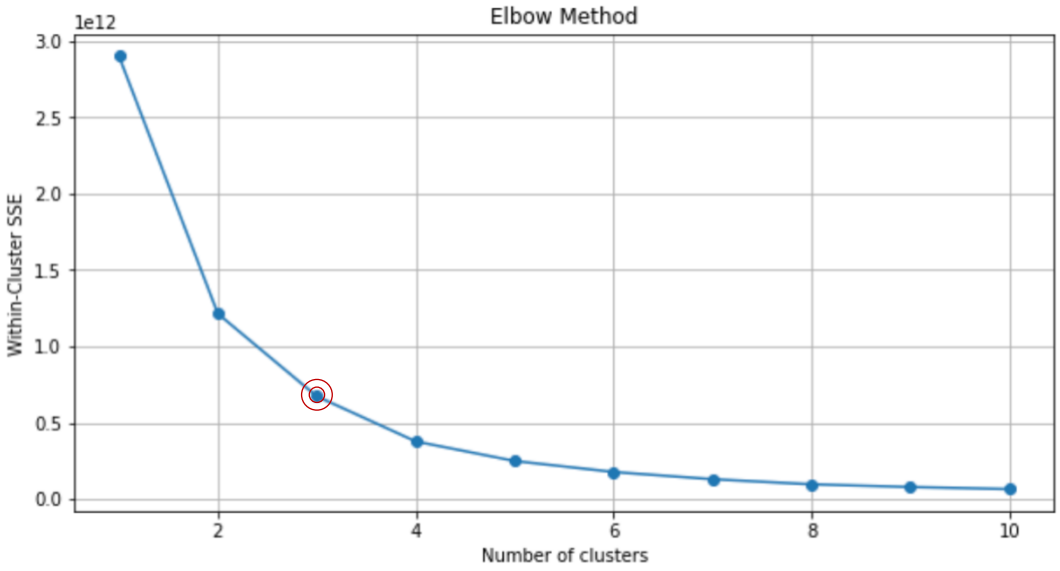


3. Customer Segmentation:

3.1. Find the optimal number of clusters (customer segments):

We will use the Elbow Method to determine the optimal number of clusters (customer segments). From the Elbow method chart below, the potential optimal number of clusters is 3.

Chart 2: Optimal number of cluster / customer segments:



3.2. Estimate and analyse customer segments:

We will use two clustering techniques: KMeans and Agglomerative Hierarchical Clustering to estimate the clusters.

3.2.1. Kmeans Clustering technique:

First of all, we exclude the ID column and group the data into 3 clusters (customer segment) using KMeans clustering technique.

Secondly, we check the size of each cluster and compute the overall means of the dataset, which will help us to compare with the mean value of each cluster.

Thirdly, we obtain the result of mean value of each cluster as below. From these information, we can make the summary of key charateristics of each customer segment in

terms of gender, age, income, marital status, education, occupation as well as the city where customers live in.

Table 1: Customer segments by Kmeans Clustering technique:

KMeans_Cluster	Sex	Marital status	Age	Education	Income	Occupation	Settlement size
0	0.384187	0.447661	37.927617	1.079065	132162.346325	0.998886	0.954343
1	0.569179	0.554556	31.947132	0.920135	90836.787402	0.433071	0.368954
2	0.295775	0.460094	43.934272	1.356808	199404.446009	1.591549	1.375587

The key features of three customer segments are below:

- **Cluster 0 - "Middle-aged Urban Females, Upper Income":**
  - Gender: comprising slightly more females than males.
  - Age: averaging around 38, which places them in the mid-career stage.
  - Marital Status: significant portion of individuals who are not in marital or formal relationships.
  - Education: slightly above high school.
  - Income: Earning an average income of approximately $132,162.
  - Occupation: tend to be skilled or official roles.
  - Settlement: Predominantly residing in bigger cities, suggesting an urban lifestyle.

- **Cluster 1 - "Young Economical Males, Average Income":**
  - Gender: A majority of males
  - Age: averaging around 32 years.
  - Marital Status: majority are either married or in some form of relationship.
  - Education: Mostly holding up to a high school level education.
  - Income: Earning below the overall average, at around $90,837, suggesting early career stages or roles that are not high-paying.
  - Occupation: lean more towards unskilled or entry-level roles.
  - Settlement: Reside mainly in smaller to mid-sized cities.

- **Cluster 2 - "Middle-Age Affluent Females, Highest Income":**
  - Gender: Predominantly females
  - Age: in their early 40s.
  - Marital Status: Significant portion of individuals who are not in marital or formal relationships.
  - Education: Highly educated, potentially holding advanced degrees or specialized training.
  - Income: With an income significantly above average at $199,404, they're likely in senior roles or specialized professions.
  - Settlement: Predominantly based in bigger cities, reflecting a cosmopolitan lifestyle.

3.2.2. Agglomerative Hierarchical Clustering:

We proceed similar steps for applying Agglomerative Hierarchical Clustering technique to estimate and evaluate clusters (customer segments) for the dataset.

Table 2: Customer segments by Agglomerative Hierarchical Clustering technique:

Mean value of each feature in every customer segment:

Agg_Cluster	Sex	Marital status	Age	Education	Income	Occupation	Settlement size
0	0.521086	0.521801	33.533238	0.959257	101944.944246	0.590422	0.515368
1	0.290909	0.400000	44.727273	1.345455	223444.563636	1.745455	1.454545
2	0.311609	0.446029	40.702648	1.193483	152156.739308	1.228106	1.215886

Median value of each feature in every customer segment:

Agg_Cluster	Sex	Marital status	Age	Education	Income	Occupation	Settlement size
0	1.0	1.0	31.0	1.0	106043.0	1.0	0.0
1	0.0	0.0	43.0	1.0	214364.0	2.0	2.0
2	0.0	0.0	37.0	1.0	149653.0	1.0	1.0

Mode value of each feature in every customer segment:

Agg_Cluster	Sex	Marital status	Age	Education	Income	Occupation	Settlement size
0	1	1	26	1	69487	1	0
1	0	0	42	2	214364	2	2
2	0	0	30	1	133328	1	1

The key features of three customer segments are below:

- **Cluster 0 - "Young Economical Males, Average Income":**
  - Gender: Predominantly male.
  - Age: Younger group, averaging around 33 years old.
  - Marital Status: Around half are married or in a relationship.
  - Education: Primarily high school level education.
  - Income: Earning slightly above $100,000
  - Occupation: Mainly entry-level or unskilled positions.
  - Settlement: Living mainly in small cities or suburbs.

- **Cluster 1 - "Middle-Aged Affluent Females, Highest Income":**
  - Gender: Dominated by females.
  - Age: Middle-aged group, averaging around 45 years old.
  - Marital Status: About 40% are married or in a relationship.
  - Education: More educated, with many holding a university degree or higher.
  - Income: High earners, with incomes averaging around $223,444, indicating senior roles or specialized professions.
  - Occupation: Primarily in management or highly specialized roles.
  - Settlement: Mainly living in big cities, indicating an urban lifestyle.

- **Cluster 2 - "Middle-aged Urban Females, Upper Income":**
  - Gender: Predominantly female.
  - Age: Mature group, averaging around 41 years old.
  - Marital Status: Less than half are married or in a relationship.

- Education: Above high school education, indicating some form of higher education.
- Income: Comfortable earners, with an average income of around $152,156.
- Occupation: Skilled jobs or lower-tier management roles.
- Settlement: Predominantly in big cities.

3.3. Compare the clusters identified by the two techniques:

KMeans Cluster 0 appears to align closely with Agglomerative Cluster 2.
Both clusters represent a group that is mature with a balanced gender ratio and high income, primarily living in larger settlements.

KMeans Cluster 1 seems to match Agglomerative Cluster 0.
Both represent a younger group, with slightly more females, living primarily in smaller settlements with an average income.

KMeans Cluster 2 is closely aligned with Agglomerative Cluster 1.
Both clusters cater to an older, predominantly female group with the highest income, living in big cities.

4. Recommendations of marketing techniques for the target customer segments:

Middle-Aged Affluent Females, Highest Income *(KMeans Cluster 0 / Agglomerative Cluster 2)*
- Promotional Offers: Given the average income level, consider offering discounts or bundle deals to entice purchases.
- Digital Marketing: Focus on online advertising, especially on social media platforms popular among younger demographics.
- Loyalty Programs: Introduce or promote a loyalty program tailored to frequent, smaller purchases.

Young Economical Males, Average Income *(KMeans Cluster 1 / Agglomerative Cluster 0)*
- Localized Advertising: Since they reside mainly in smaller settlements, use local newspapers, radio stations, and community events for advertising.
- Online Shopping: Offer convenient online shopping experiences with perks like free shipping or returns, given the potential distance from major shopping hubs.
- Partnership: Partner with local influencers or community leaders for product endorsements or joint events.

Middle-Aged Affluent Females, Highest Income *(KMeans Cluster 2 / Agglomerative Cluster 1)*
- Exclusive Events: Host invite-only events or early access sales for new product launches.
- Personalized Services: Offer personalized shopping experiences, concierge services, or bespoke product options.
- Health & Wellness: Given the older age group, focus on products and services that promote health, wellness, and anti-aging benefits.

5. Conclusion:

In this analysis report, we focused on indentify the most appropriate customer segments to understand distinct groups for the business marketing's strategy development. We applied two clustering techniques, KMeans and Agglomerative Hierarchical Clustering, we identified three significant segments: "Middle-Aged Affluent Females, Highest Income", "Young Economical Males, Average Income", and "Middle-Aged Affluent Females, Highest Income". These segments

represent unique characteristics in terms of gender, age, marital status, income, education, occupation and settlement size. We also proposed the marketing strategies for each segment to enhance sales effectiveness. The consistency in segment profiles across both clustering methods has supported for data-driven marketing decisions for the business in the future.

End.