

2024

Customer Segmentation

BUSA 8001 ASSIGNMENT 2
PRIYAM SINHA 47750731

Table of Contents

<i>Customer Segmentation Analysis Report</i>	2
Exploratory Data Analysis	2
<i>Customer Segmentation</i>	6
Standardization of Numeric Variables	6
<i>Determining the Optimal Number of Clusters</i>	7
Elbow Method	7
Silhouette Analysis	7
<i>Clustering Techniques</i>	9
K-means++ Clustering	9
Agglomerative Clustering	11
<i>Comparison of Clustering Techniques</i>	12
<i>Recommendations</i>	13
<i>Conclusion</i>	13

Customer Segmentation Analysis Report

Word Count: 994

Introduction

In the competitive travel industry, this report identifies distinct customer segments among 2,000 customers using K-means++ and Agglomerative Clustering. By analyzing variables like age, income, and occupation, we aim to inform targeted marketing strategies, enhance customer engagement, and boost revenue through personalized service delivery.

Exploratory Data Analysis

1. Age Distribution:

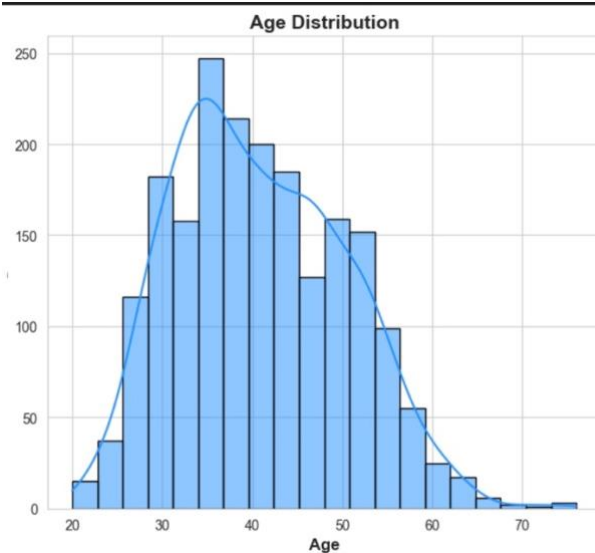


Figure 1: Histogram showing the distribution of customer ages.

Analysis: The age distribution is relatively uniform, with a slight concentration around the 30-40 year range, indicating a diverse age group among our customers.

2. Income Distribution:

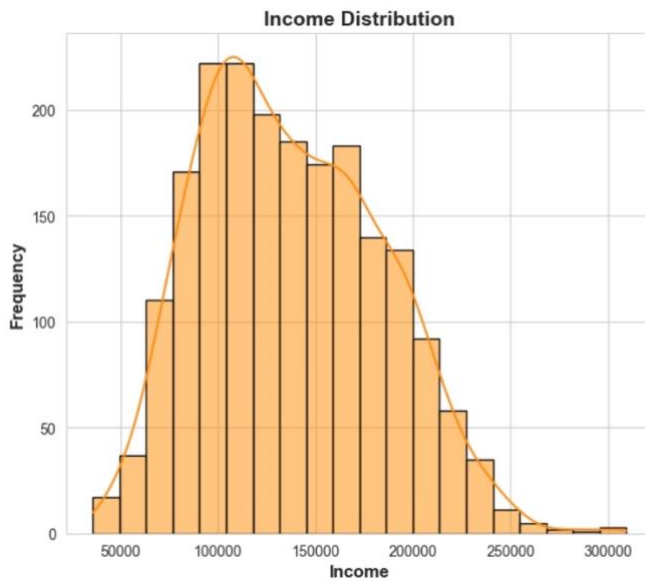


Figure 2: Histogram illustrating the spread of annual incomes among customers.

Analysis: The income distribution is right-skewed, suggesting a larger number of customers with moderate incomes and fewer customers with high incomes.

3. Gender Distribution:

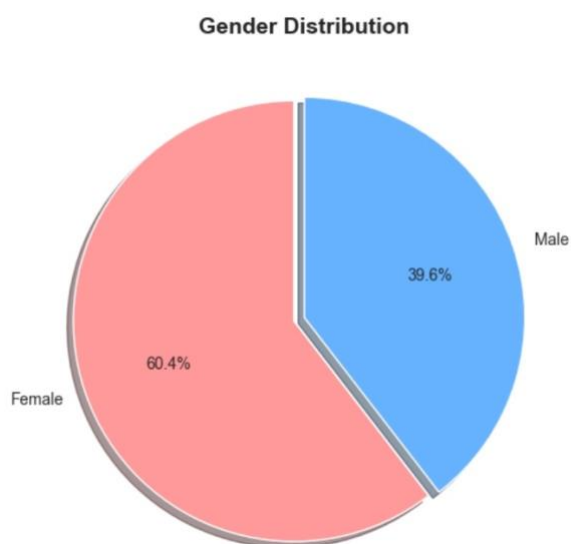


Figure 3: Pie chart representing the gender breakdown of customers.

Analysis: There is a higher proportion of female customers compared to males, which could influence the design of targeted marketing campaigns.

4. Marital Status Distribution:

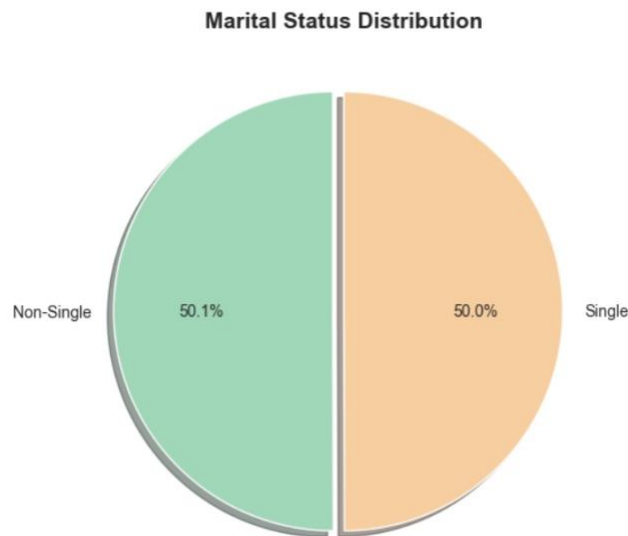


Figure 4: Pie chart depicting the marital status of customers.

Analysis: The near-equal split between single and non-single customers provides an opportunity to design diverse marketing strategies catering to both groups.

5. Occupation Distribution:

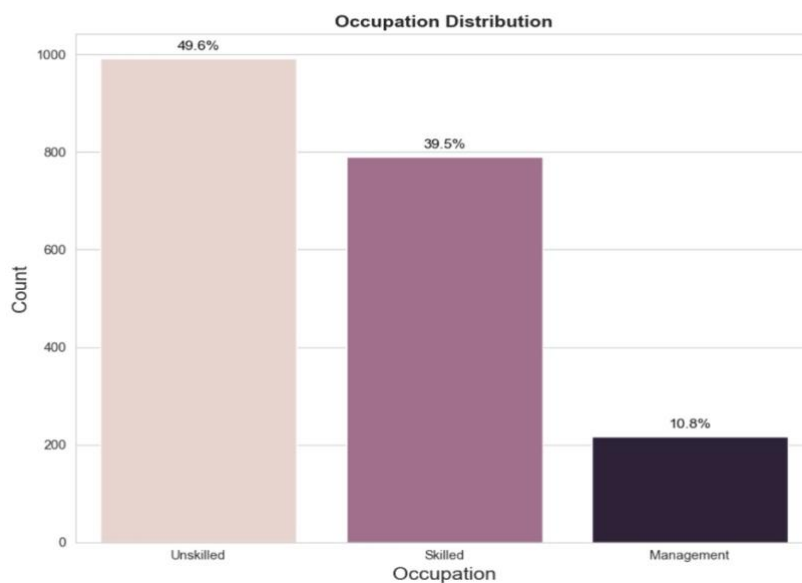


Figure 5: Bar chart showing the distribution of customer occupations.

Analysis: A majority of customers are unemployed or unskilled, indicating potential areas for engagement through affordable travel packages.

6. Settlement Size Distribution:

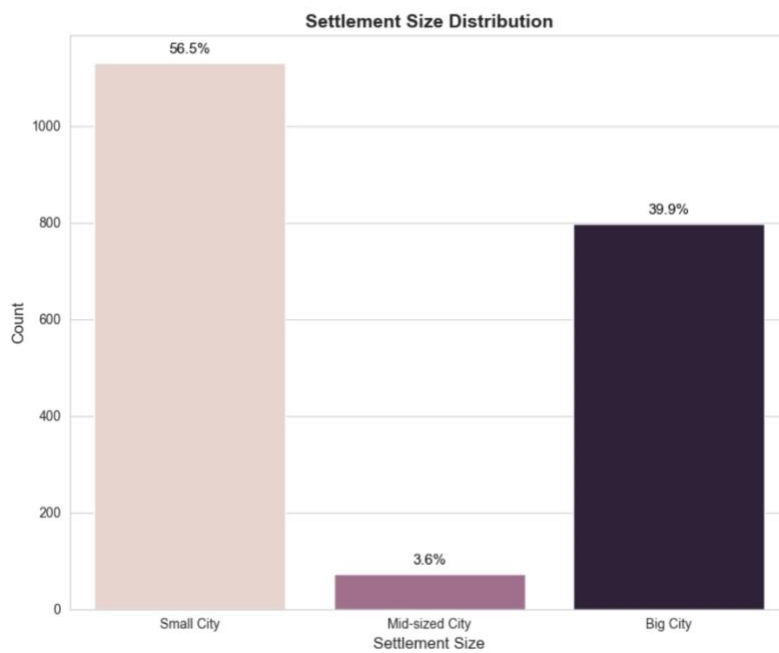


Figure 6: Bar chart illustrating the settlement sizes of customer residences.

Analysis: The overwhelming majority reside in big cities, suggesting that our marketing efforts should focus primarily on urban areas.

7. Education Level Distribution:

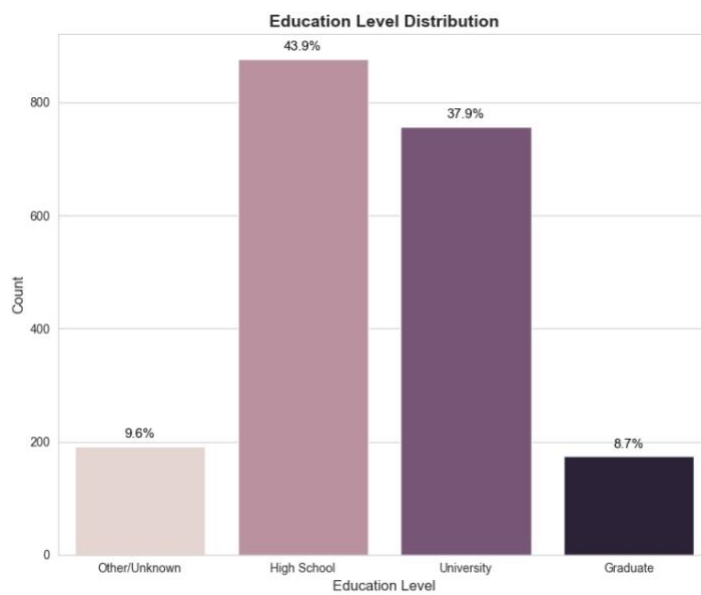


Figure 7: Bar chart illustrating the Education Level of customers

Analysis: Most customers have completed at least university-level education, with a significant portion holding graduate degrees.

Customer Segmentation

Standardization of Numeric Variables

```
# Feature scaling (for Age and Income)
scaler = StandardScaler()
df[['Age', 'Income']] = scaler.fit_transform(df[['Age', 'Income']])
✓ 0.0s
```

Determining the Optimal Number of Clusters

Elbow Method

The point where the rate of decrease sharply slows, resembling an elbow, suggests the ideal number of clusters.

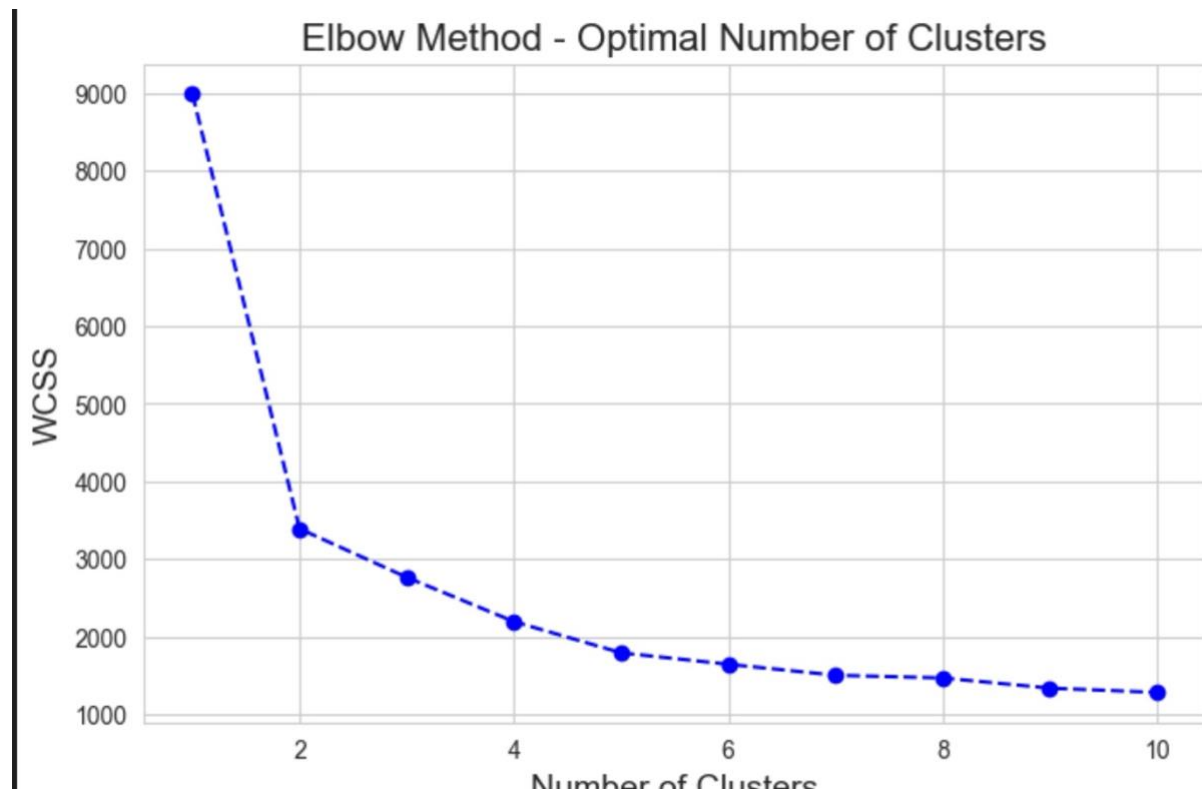


Figure 7: Elbow Method plot indicating the optimal number of clusters.

Analysis: The elbow point appears at **3 clusters**, where adding more clusters results in diminishing returns regarding WCSS reduction.

Silhouette Analysis

Silhouette Scores were calculated for 2,3 and 4 clusters to assess the cohesion and separation of the clusters.

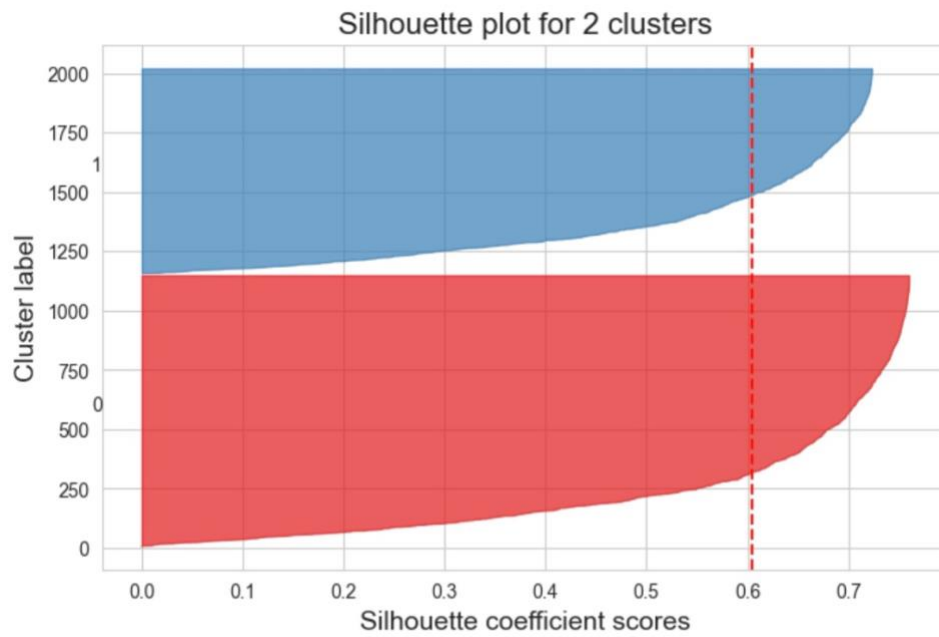


Figure 8: Silhouette plot for 2 clusters.

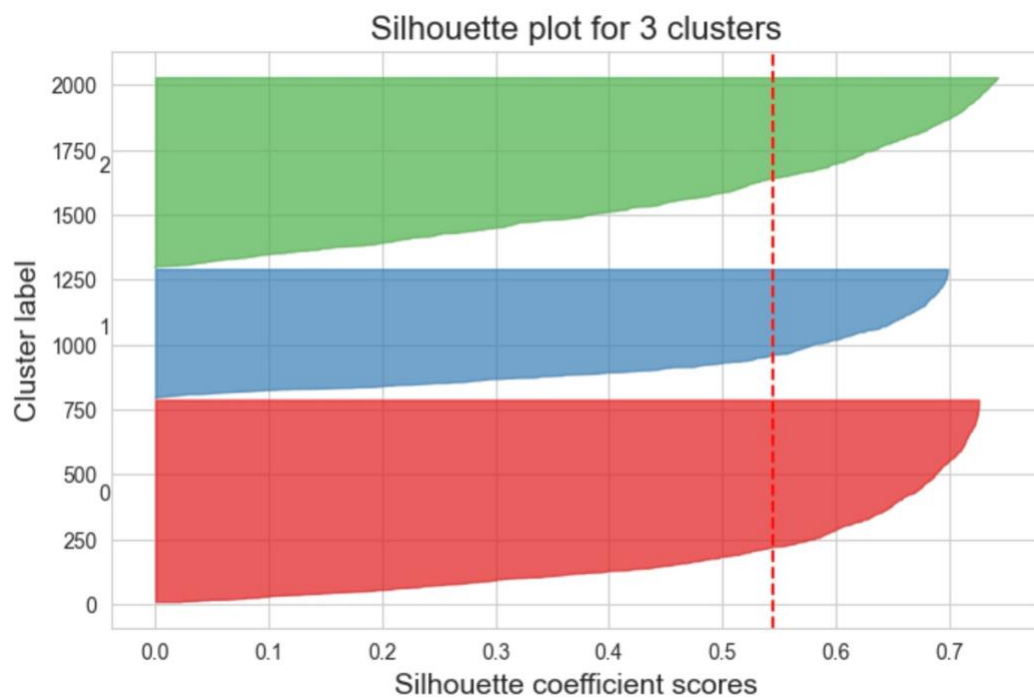


Figure 9: Silhouette plot for 3 clusters.

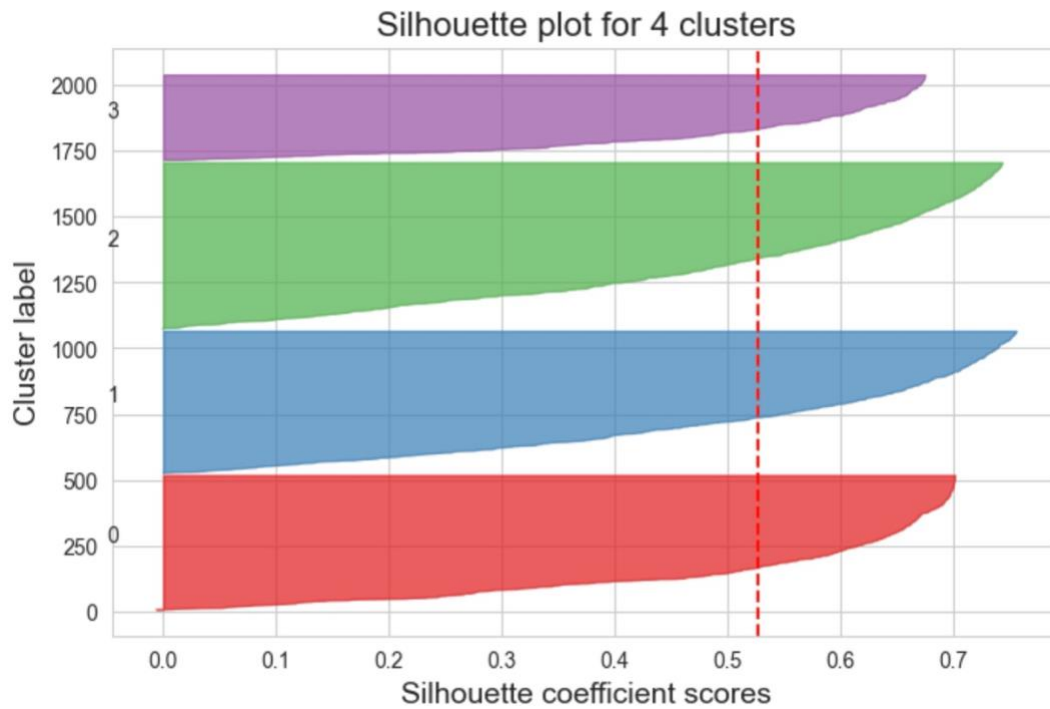


Figure 10: Silhouette plot for 4 clusters.

Analysis:

As we can see this Silhouette score for optimal cluster 2 is high but for cluster 3 is also close. Therefore, optimal cluster can be 2 or 3. As we see from the graphs we can see that when we make 3 clusters there's a good distribution of the customers and now we will have the balanced customer segmentation.

Clustering Techniques

With the optimal number of clusters determined as **3**, we applied both **K-means++** and **Agglomerative Clustering** techniques.

K-means++ Clustering

Cluster Centres and Sizes:

	Age	Income	Gender	Marital Status	Education	Occupation	Settlement Size	Number of Customers
0	31.473008	91819.532134	0.390746	0.069409	0.902314	0.093830	0.164524	778
1	53.617409	200089.060729	0.842105	0.973684	2.076923	1.176113	1.578947	494
2	42.134615	143891.210165	0.671703	0.640110	1.627747	0.784341	1.043956	728

Table 1: K-means++ Cluster Centres and Sizes.

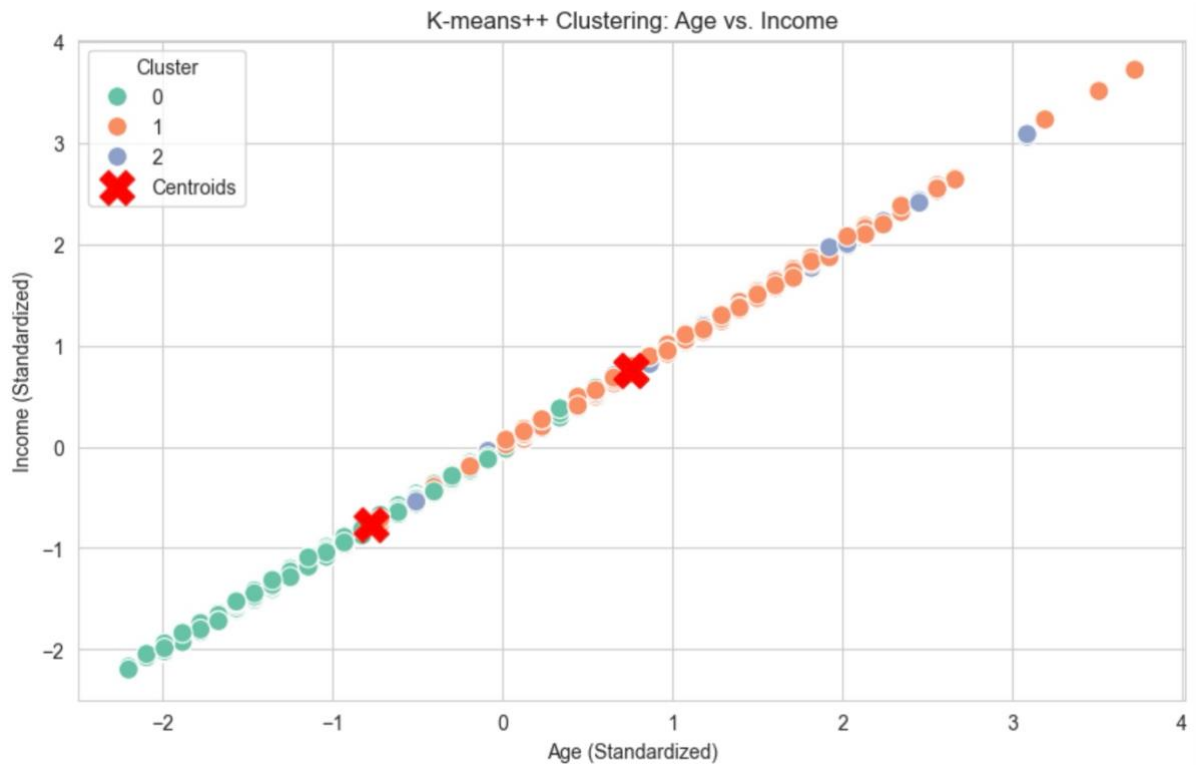


Figure : K-means++ Clustering Age vs Income with Centroids

Analysis:

- **Cluster 0:** Represents a large segment of customers with average age and income, balanced across gender and marital status. These customers likely form the core customer base.
- **Cluster 1:** Comprises moderately aged customers with higher income levels. This segment may include professionals seeking premium travel experiences.
- **Cluster 2:** A smaller group characterized by younger age and lower income, possibly consisting of budget-conscious travellers.

Agglomerative Clustering

Agglo_Cluster	Age	Income	Gender	Marital Status	Education	Occupation	Settlement Size
0	42.50	145633.39	0.67	0.64	1.63	0.78	1.04
1	30.70	88084.87	0.38	0.05	0.87	0.07	0.13
2	55.45	209084.89	0.85	0.98	2.10	1.18	1.59

Table 2: Agglomerative Clustering Cluster Sizes.

Cluster Sizes:

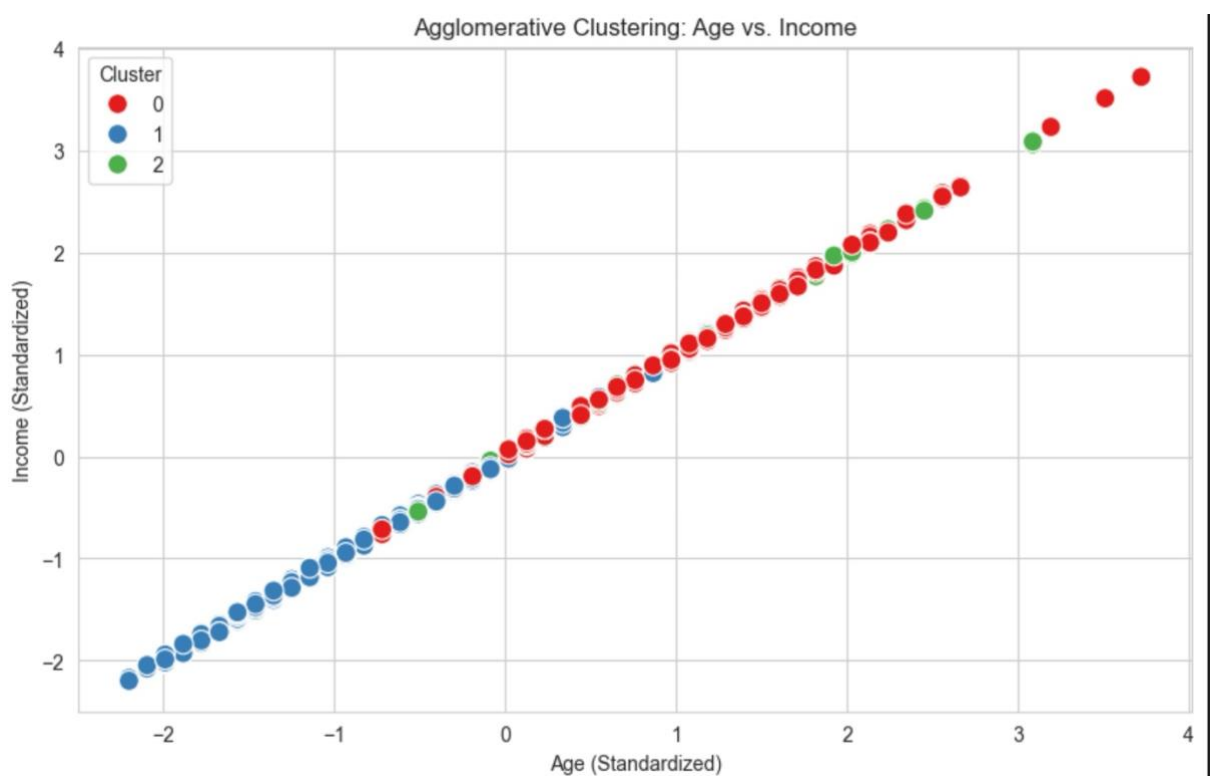


Figure : Agglomerative Clustering Age vs Income with Centroids

Analysis:

- **Cluster 0:** Similar in size to K-means Cluster 1, indicating a significant segment of higher-income customers.
- **Cluster 1:** The largest cluster, analogous to K-means Cluster 0, comprising the main customer base with average demographics.
- **Cluster 2:** Comparable to K-means Cluster 2, representing a smaller, younger, and lower-income segment.

Comparison of Clustering Techniques

The cross-tabulation of cluster memberships between K-means++ and Agglomerative Clustering reveals substantial alignment between the two methods:

Cluster Sizes Comparison		
	K-means++ Cluster Size	Agglomerative Cluster Size
0	778	982
1	494	668
2	728	350

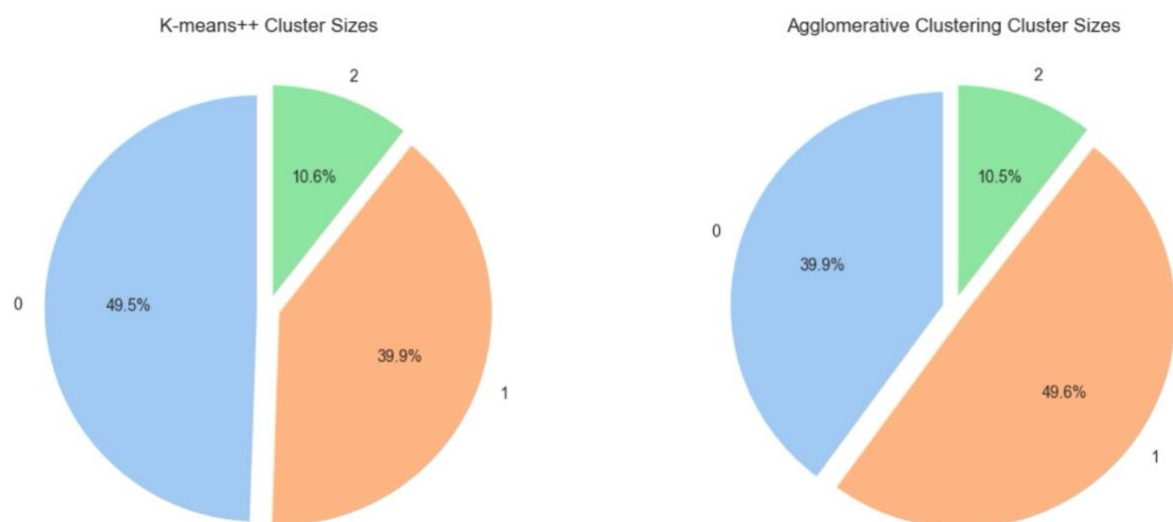


Figure 11: Pie chart comparison of K-means++ and Agglomerative Clustering results.

Analysis:

- **Cluster 0 (K-means++)** predominantly aligns with **Cluster 0 (Agglo)**, indicating consistency in identifying the core customer base.
- **Cluster 1 (K-means++)** exclusively maps to **Cluster 0 (Agglo)**, reinforcing its identification as a distinct high-income segment.
- **Cluster 2 (K-means++)** largely corresponds to **Cluster 2 (Agglo)**, validating the identification of the younger, budget-conscious segment.
- There is minimal overlap, suggesting that both clustering techniques reliably identify similar customer segments with slight variations.

Recommendations

1. Core Customer Base (Cluster 0):

- **Personalized Communication:** Implement targeted email campaigns highlighting travel packages that cater to their average income and diverse demographics.
- **Loyalty Programs:** Introduce loyalty rewards to enhance retention and encourage repeat bookings.

2. High-Income Professionals (Cluster 1):

- **Premium Packages:** Develop exclusive, high-end travel packages that include luxury accommodations, personalized services, and unique experiences.
- **Upselling Opportunities:** Promote add-ons such as private tours, first-class flights, and premium insurance plans.

3. Budget-Conscious Young Travelers (Cluster 2):

- **Affordable Travel Options:** Create budget-friendly packages with flexible dates, economical accommodations, and group discounts.
- **Social Media Campaigns:** Leverage platforms like Instagram and TikTok to reach younger audiences with engaging and visually appealing content.

4. Urban Focus:

- **Localized Marketing:** Given that the majority reside in big cities, focus marketing efforts on urban centers through digital advertising, local events, and partnerships with city-based businesses.
 - **Convenient Booking Platforms:** Enhance mobile and online booking experiences to cater to tech-savvy urban customers.
-

Conclusion

This customer segmentation identified three distinct groups among 2,000 customers using K-means++ and Agglomerative Clustering: a core average segment, high-income professionals, and budget-conscious young travellers. These insights support targeted marketing strategies, enhancing customer satisfaction and business growth through data-driven decision-making and periodic re-segmentation.