# BUSA 8001: Applied Predictive Analysis

Quang Huy Vo



**Assignment 2: Segmentation Report**

Master of Business Analytics

Macquarie University

Date of Submission: 18 October 2024

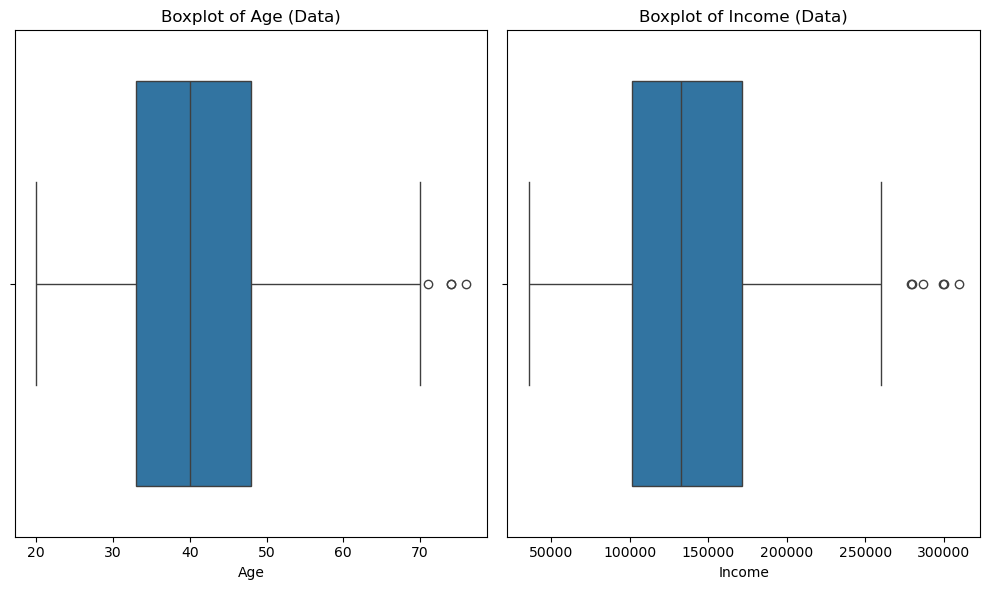
WordCount: 1031

1. **Introduction**

Customer segmentation is an essential strategy for any business, especially in the travel industry, where personalization and tailored experiences drive customer satisfaction and loyalty. By dividing a large customer base into distinct groups based on shared characteristics, travel agencies can refine their marketing approaches, optimize offerings, and ultimately enhance customer engagement. This report presents an analysis of a dataset containing 2,000 customers of a large travel agency. Key attributes such as age, income, gender, education, occupation, marital status, and settlement size are considered in this study. By using advanced clustering techniques: K-means++ and Agglomerative Clustering, the analysis identifies distinct customer segments, providing insights that can guide more focused and effective marketing strategies. The objective of this report is to present actionable recommendations for marketing efforts tailored to each customer segment based on their demographic and economic characteristics.

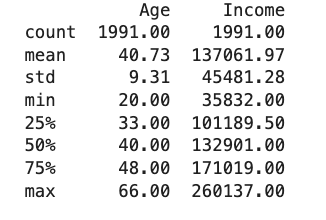
**2. Exploratory Data Analysis**

The final dataset comprises 1,991 customers, after removing 3 duplicate entries due to the absence of a primary key. Additionally, outliers were identified in the Age and Income columns, and these values were cleaned to ensure the accuracy and reliability of the subsequent analysis.



*Figure 1: Boxplot of Age and Income*

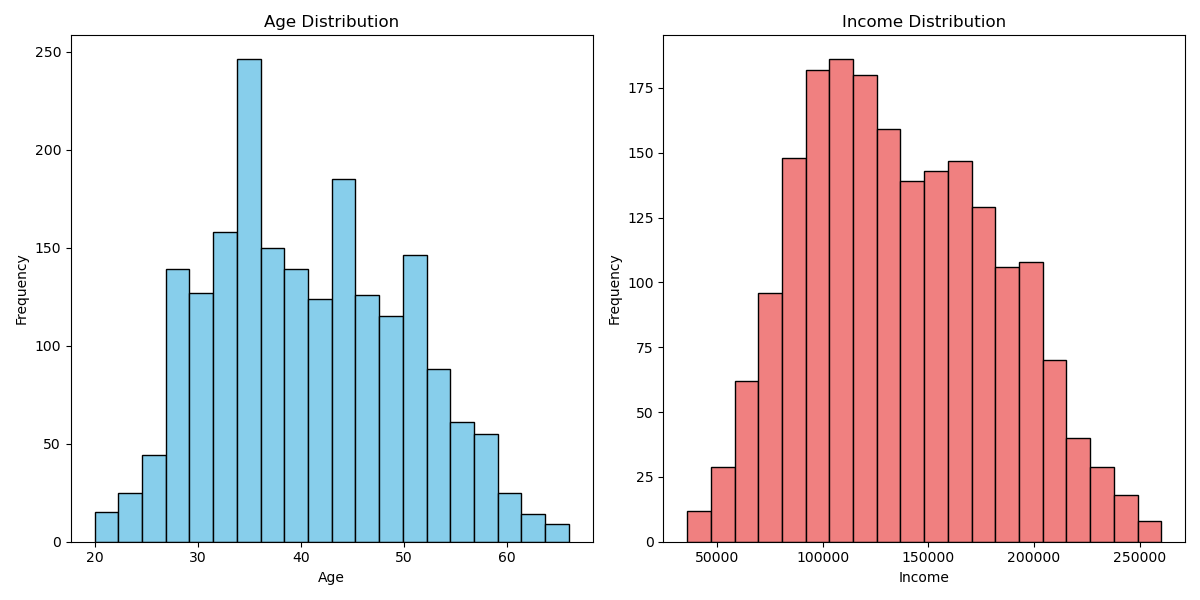
Summary Statistics:



*Figure 2: Summary Statistics of Numerical Variables*

* Age ranges from 20 to 66 with a mean of 40.73 years old.
* Income ranges from 35,832 to 260,137 USD, with an average income of approximately 137,061.97 USD.

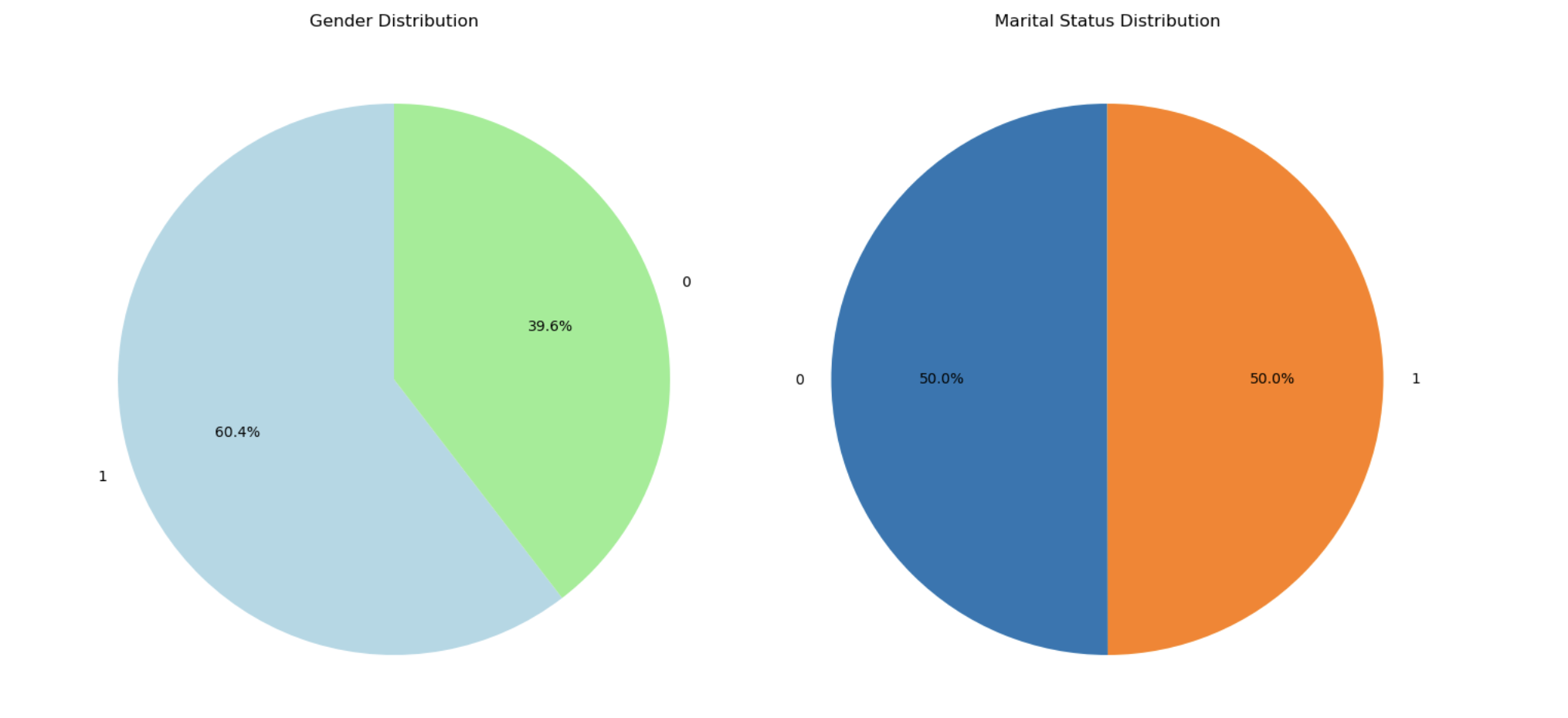
**Age and Income Distribution:**



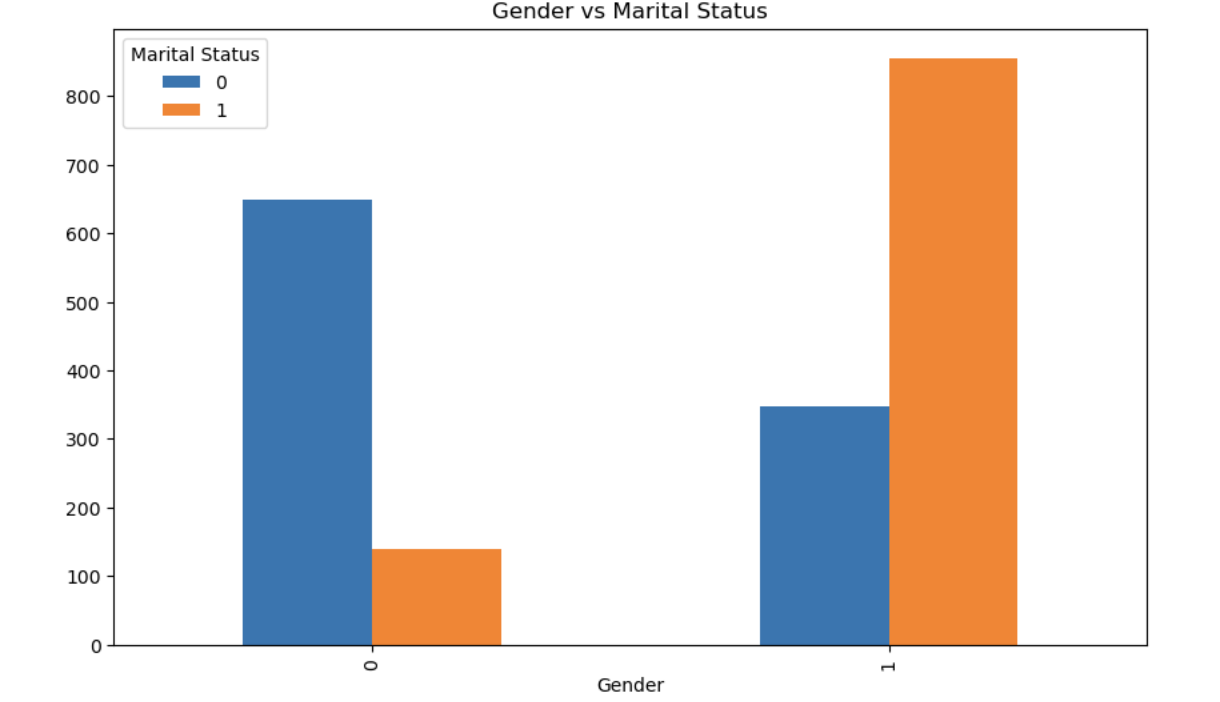
*Figure 3: Age and Income Distribution Histograms*

* Travel agency's core customer base is middle-aged and moderately affluent. The income distribution was positively skewed, with most customers earning between 50,000 to 150,000 USD, while a small proportion of customers earned over 200,000 USD. Therefore, marketing efforts should focus on this 30-50 age group with mid-range income travel packages.

**Gender and Marital Status Distribution:**



*Figure 4: Gender Distribution Pie Chart*



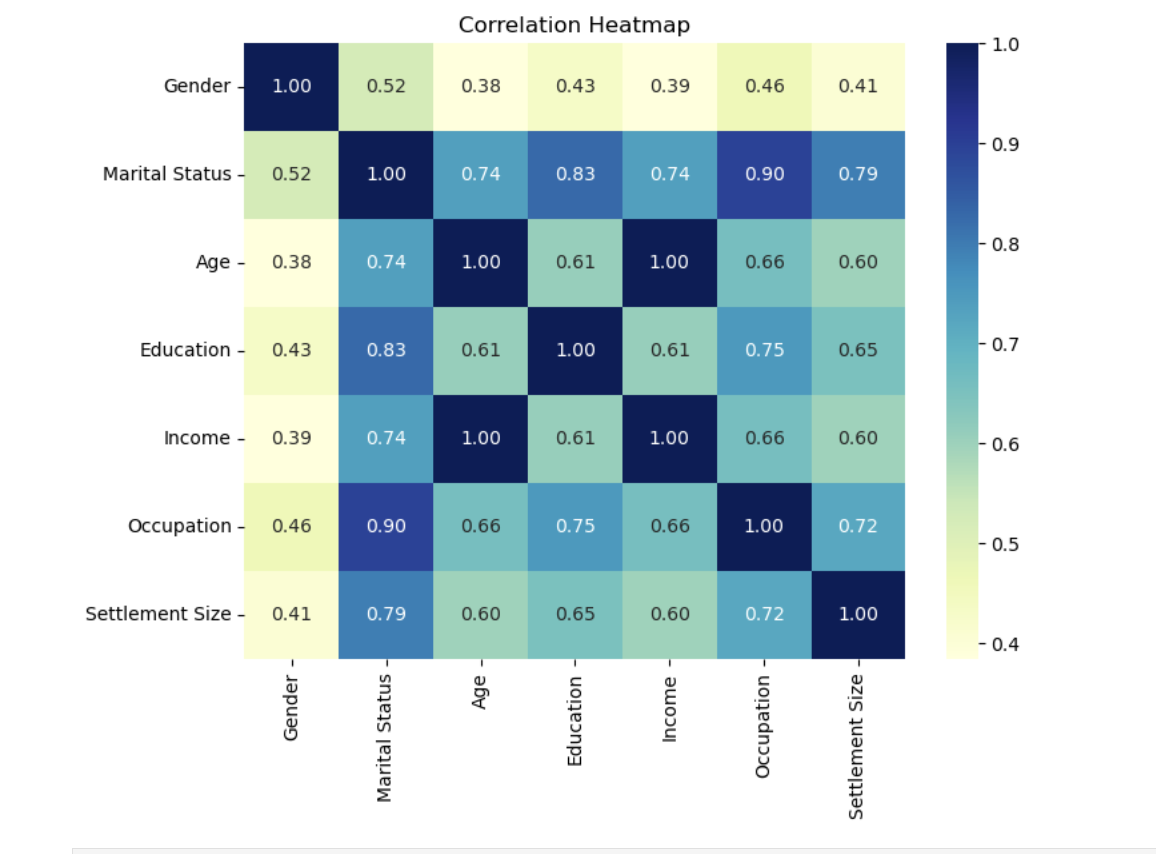
*Figure 5: Gender vs Martial Status Charts*

* + 60.4% female, 39.6% male.
  + Males (0): More single males than non-single.
  + Females (1): Majority is non-single.
  + Non-single females form a large segment, while single males dominate the male group.
  + Females represent the majority of the customer base.

Key Insights:

* Focus on non-single females: Marketing campaigns targeting families or couples could be effective.
* Target single males: Adventure or solo travel promotions may appeal to this group.

**Correlation Heatmap:**

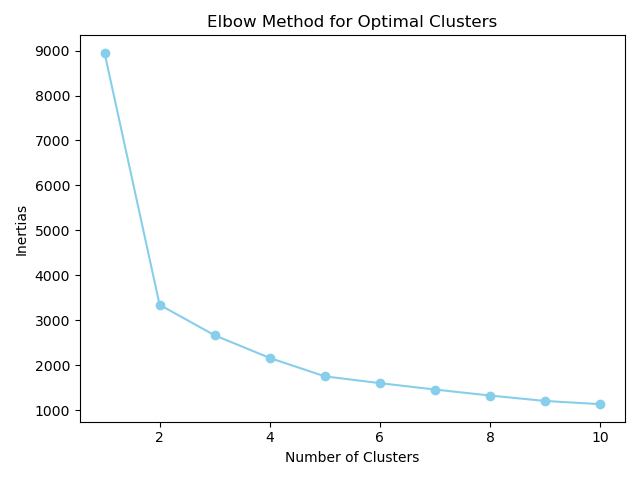


*Figure 6: Correlation Heatmap*

The heatmap indicated strong correlations between age, income, and education, with income being highly correlated with occupation and education. This insight suggests that more educated and experienced customers tend to have higher incomes, which is useful for segment profiling.

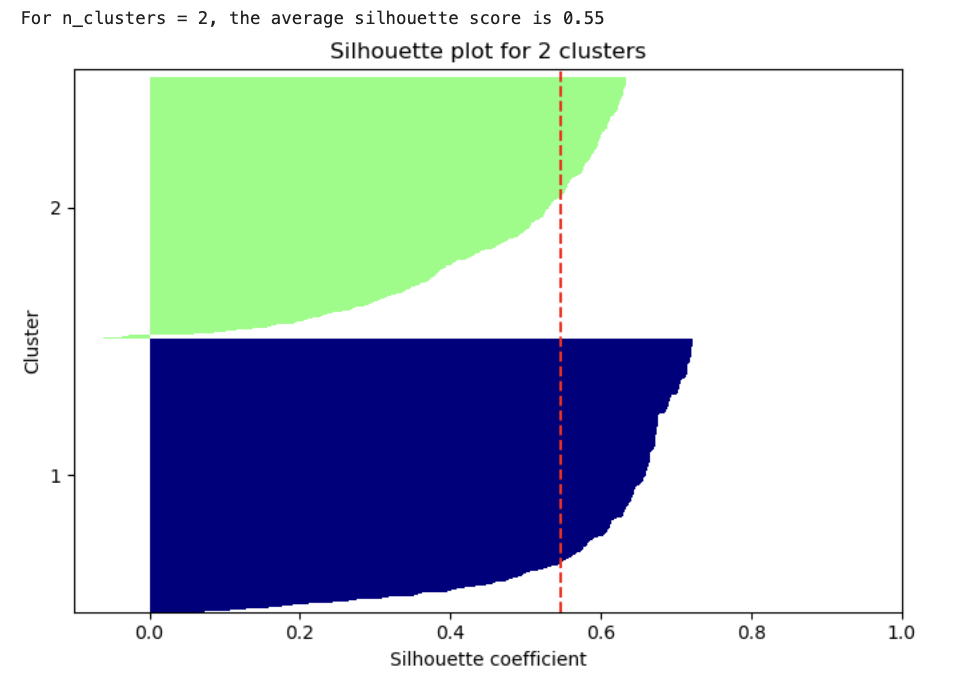
**3. Optimal Cluster**

* The Elbow Method plot suggested that 2 clusters provided the best trade-off between variance and number of clusters.

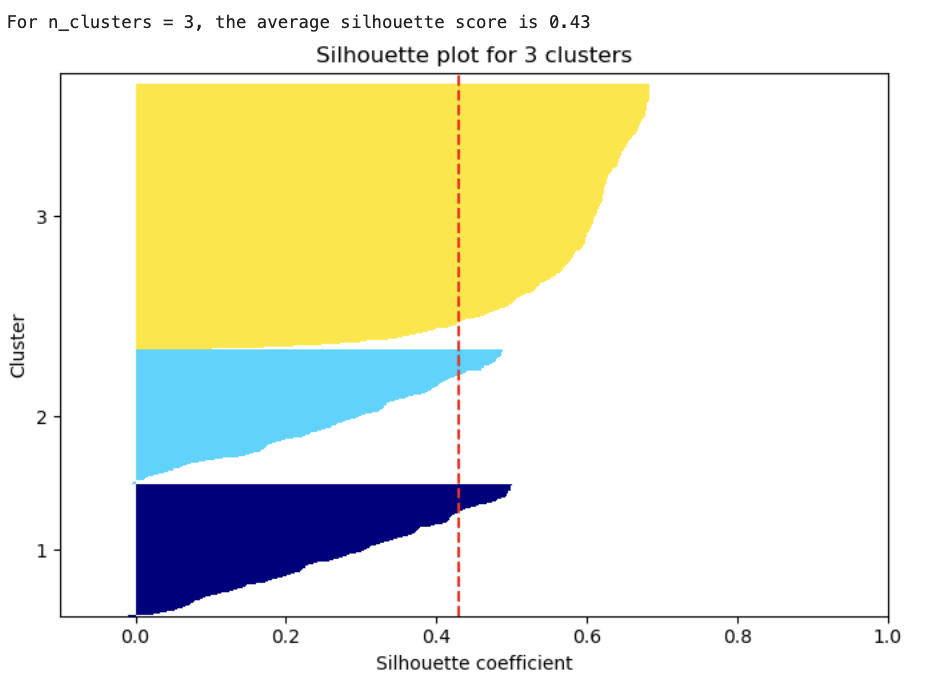


*Figure 7: Elbow Graph*

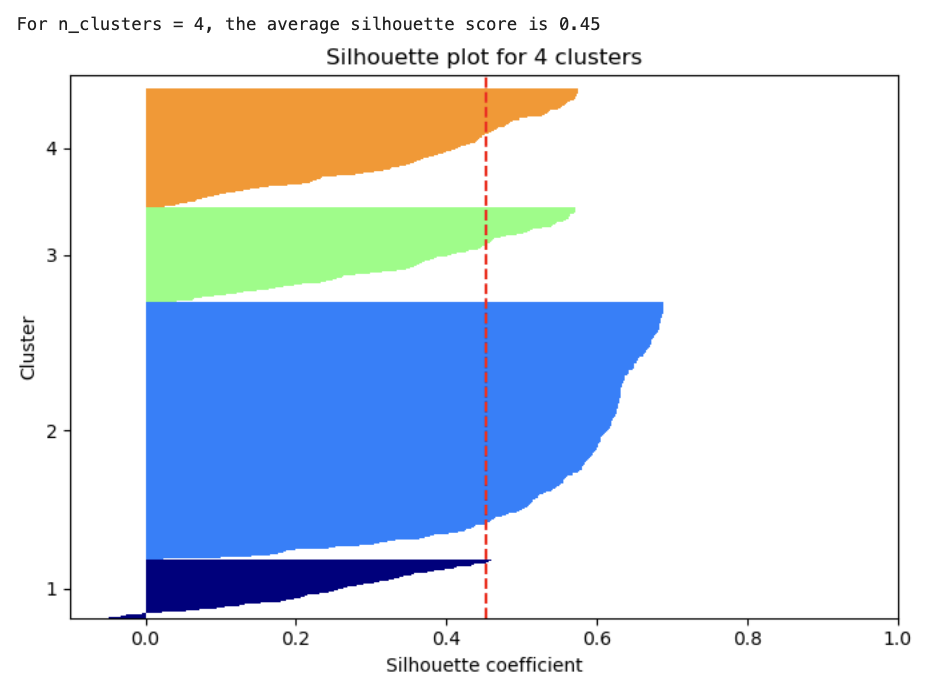
* The Silhouette Scores for 2, 3, and 4 clusters were also calculated.
* 2 Clusters: The average silhouette score was 0.55, indicating well-separated and cohesive clusters.
* 3 and 4 Clusters: Lower silhouette scores for 3 and 4 clusters (0.43 and 0.45, respectively) suggest less distinct clustering.



*Figure 8: 2 Clusters - Silhouette*



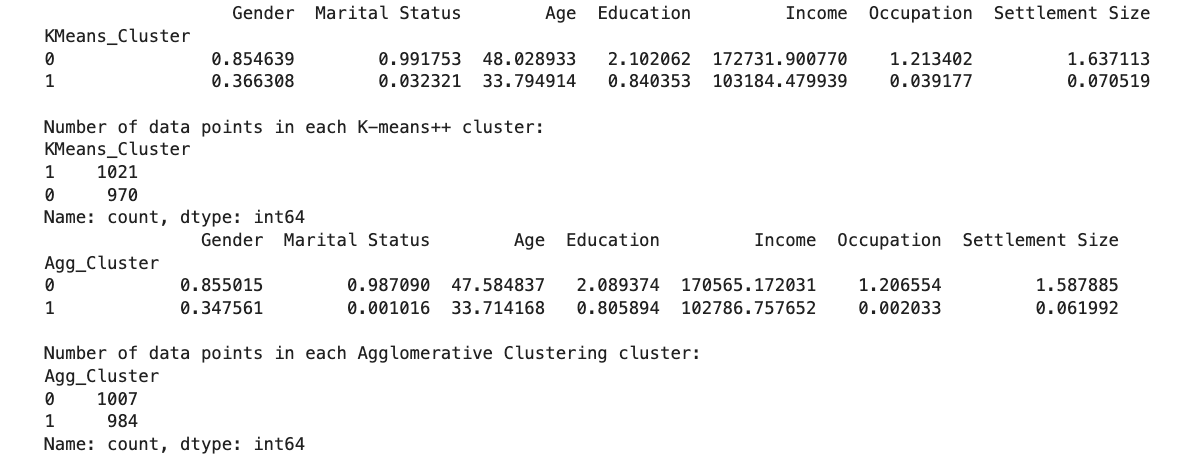
*Figure 9: 3 Clusters - Silhouette*



*Figure 10: 4 Clusters - Silhouette*

**4. Clustering**

After determining the optimal number of clusters (2), both K-means++ and Agglomerative Clustering methods are applied to segment the customers into two distinct groups.



1. K-means++ Clustering:
   * Cluster 0:
     + Predominantly female (85.5%).
     + Older customers (average age ~48).
     + Higher income (~$172,731).
     + Mostly non-single (99.2%).
     + More educated (average education level ~2.1, closer to graduate school) and living in larger cities.
   * Cluster 1:
     + Mostly male (63.7%).
     + Younger customers (average age ~34).
     + Lower income (~$103,184).
     + Primarily single (97%).
     + Less educated (average education ~0.84, closer to high school) and residing in smaller cities.

Cluster Size:

* + - Cluster 1 has 1,021 customers, while Cluster 0 has 970.

1. Agglomerative Clustering:
   * Cluster 0:
     + Predominantly female (85.5%).
     + Older customers (average age ~47.6).
     + Higher income (~$170,565).
     + Mostly non-single (98.7%).
     + Higher education and residing in larger cities.
   * Cluster 1:
     + Mostly male (65.2%).
     + Younger customers (average age ~33.7).
     + Lower income (~$102,787).
     + Primarily single (99.9%).
     + Less education and living in smaller cities.

Cluster Size:

* + - Cluster 0 has 1,007 customers, while Cluster 1 has 984 customers.

**Key Insights:**

* Gender plays a significant role in both clustering methods. Cluster 0 is predominantly female, while Cluster 1 is male-dominated.
* Cluster 0 consists of older, wealthier, and more educated females who are mostly non-single and live in larger cities.
* Cluster 1 includes younger, lower-income males, who are primarily single, less educated, and reside in smaller cities.
* The results of the Agglomerative Clustering were consistent with K-means++, identifying similar patterns in terms of all variables.

**Cluster Overlap:**

* The overlap between the two clustering methods was significant, with an overlap percentage of 98.14%, indicating that both methods identified similar customer groups with minimal differences in cluster assignments.

**5. Recommendations**

**Segment 1:** Predominantly Female, Older, Wealthier, Non-Single

* **Luxury and Family-Oriented Packages**: This group is likely to appreciate luxury travel experiences. Offering high-end family vacation packages or couple-centric getaways would appeal to their interests.
* **Exclusive Membership and Loyalty Programs**: Since these customers have high incomes, promoting exclusive travel memberships with benefits like first-class upgrades, access to exclusive events, or personalized concierge services would resonate well with them.
* **Cultural and Relaxation-Focused Trips**: Since this segment consists of older individuals, promoting relaxing retreats such as spa resorts, cultural tours, or wellness vacations could be attractive.
* **Focus on Sustainability and Ethical Travel**: Highlight eco-friendly travel options or sustainable tourism initiatives to appeal to their educated and likely socially conscious mindset.

**Segment 2:** Predominantly Male, Younger, Lower Income, Single

* **Budget-Friendly Travel Options**: This group may be more price-sensitive due to their lower income. Promoting affordable travel deals, discount packages, and group travel offers (e.g., adventure travel, road trips) would be appealing.
* **Adventure and Solo Travel**: This predominantly single, younger group may enjoy adventure tours, backpacking trips, or solo travel packages. Emphasize activities like hiking, scuba diving, or exploring exotic locations.
* **Flexible Payment Plans**: Offering flexible payment options or instalment plans for travel might attract customers who are interested in travel but concerned about upfront costs.
* **Tech-Savvy Promotions**: Leverage social media and online campaigns targeting younger audiences. Promotions through Instagram or YouTube influencers can effectively reach this demographic.
* **Weekend Getaways and Short Trips**: Shorter, more affordable weekend trips could appeal to this group, who may prefer quick escapes rather than long vacations.

**6. Conclusion**

This report successfully applied K-means++ and Agglomerative Clustering techniques to segment the customer base of a large travel agency, identifying two distinct groups: one predominantly consisting of older, wealthier, non-single females, and the other made up of younger, lower-income, single males. The clustering methods demonstrated strong consistency, with an overlap of over 98%, confirming the reliability of the segmentation. By understanding the unique characteristics of each segment, the travel agency can implement targeted marketing strategies—such as luxury family-oriented packages for wealthier females and adventure-based budget travel for younger males—that will enhance customer satisfaction and drive sales. The segmentation insights offer a pathway for the agency to increase profitability by delivering personalized experiences that cater directly to the preferences and needs of each group.