

Customer Segmentation Analysis for Travel Agency

Word count: 984

Table of Contents

1. Introduction.....	2
2. Exploratory Data Analysis	2
3. Customer Segmentation	4
3.1. Optimal Number of Clusters	4
3.2. K-means++ Clustering Results	5
3.3. Agglomerative Clustering Results	6
3.4. Comparison of Clustering Techniques.....	6
4. Recommendations.....	7
5. Conclusion.....	7
Figure 1: Distribution of Age and Income.....	3
Figure 2: Age vs Income by Gender	3
Figure 3: Income Distribution by Education and Occupation	4
Figure 4: Elbow plot for determining optimal number of clusters.....	4
Figure 5: Silhouette plot for different numbers of clusters	5
Figure 6: Comparison of K-means++ and Agglomerative clustering results	6
Table 1: K-means++ cluster centers and sizes	5
Table 2: Agglomerative cluster centers and sizes	6

1. Introduction

Our travel agency aims to better understand and serve our diverse customer base. This report presents the results of a customer segmentation analysis performed on a dataset of 2,000 customers. By identifying distinct customer groups, we can tailor our marketing strategies and improve customer satisfaction.

Our approach involves:

1. Exploratory data analysis to understand customer characteristics
2. Application of clustering techniques to identify customer segments
3. Profiling of identified segments
4. Recommendations for targeted marketing strategies

The dataset includes variables such as customer age, gender, annual income, marital status, education, occupation, and settlement size collected through bookings and interactions with our agency.

2. Exploratory Data Analysis

Key observations from our exploratory analysis:

- The average age of our customers is about 41 years, ranging from 20 to 66 years.
- The average annual income is \$137,101, ranging from \$36,035 to \$260,137.
- 60.4% of our customers are female, 39.6% are male.
- Marital status is almost evenly split between single (50.1%) and non-single (49.9%).
- Education levels: 44% high school, 37.7% university, 8.7% graduate school, 9.6% other/unknown.
- Occupations: 49.7% unemployed/unskilled, 39.5% skilled employees/officials, 10.8% management/self-employed.
- Settlement sizes: 56.5% small city, 39.8% big city, 3.6% mid-sized city.

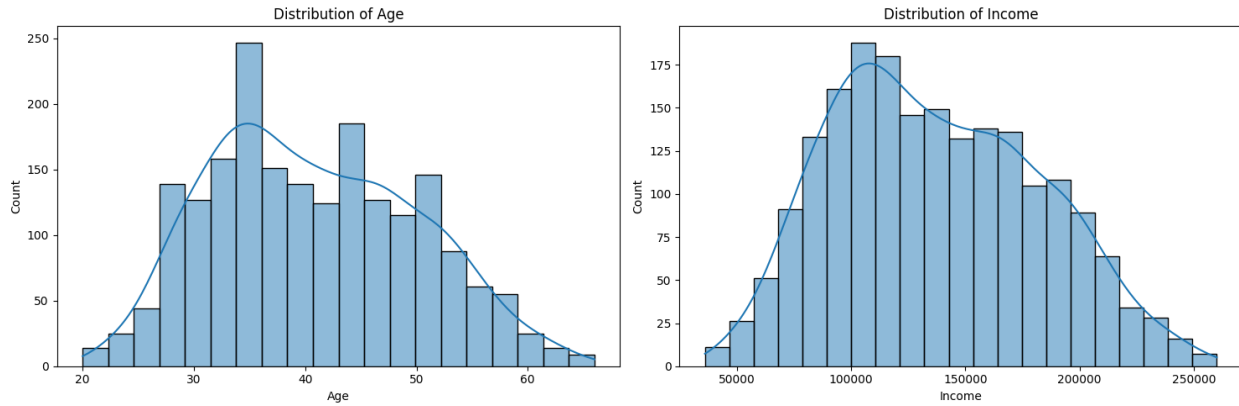


Figure 1: Distribution of Age and Income

The histograms show that both age and income follow approximately normal distributions. The age distribution is slightly right-skewed, peaking around 35-40 years. The income distribution is more notably right-skewed, with a peak around 100,000-125,000 units.

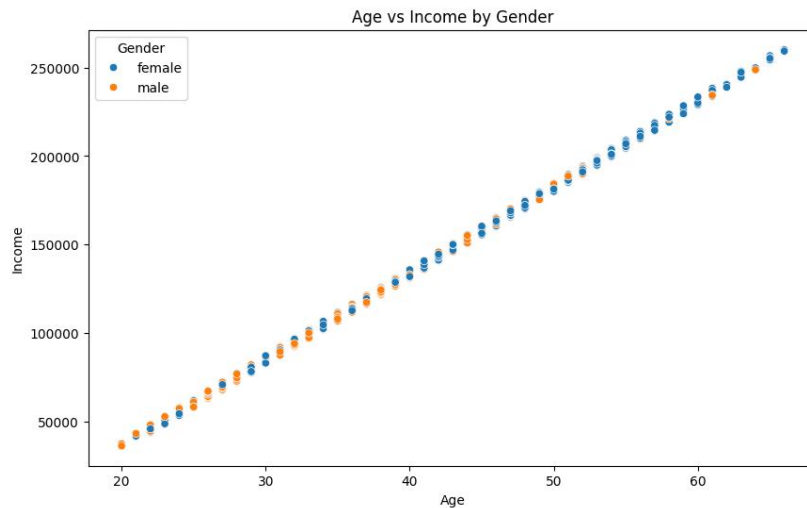


Figure 2: Age vs Income by Gender

This scatter plot reveals a remarkably strong positive linear relationship between age and income for both genders. There appears to be little gender-based income disparity, as the data points for males and females overlap significantly across all ages.

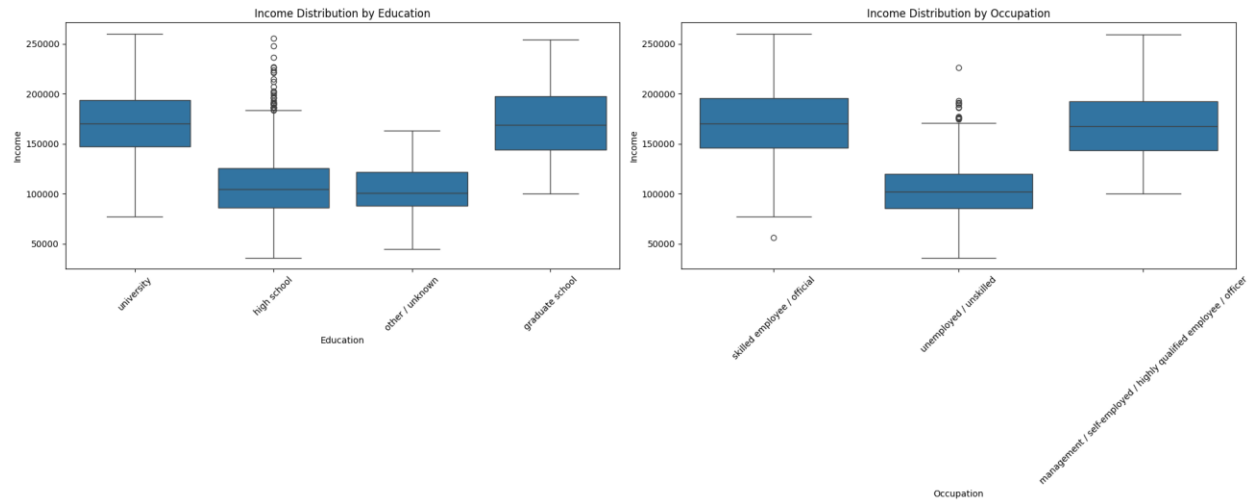


Figure 3: Income Distribution by Education and Occupation

These box plots illustrate significant income variations across education levels and occupations. University and graduate school graduates tend to have higher median incomes and wider income ranges. In terms of occupation, the management/self-employed/highly qualified group shows the highest median income and widest range, followed by skilled employees/officials.

3. Customer Segmentation

We applied two clustering techniques - K-means++ and Agglomerative Clustering - to identify distinct customer segments.

3.1. Optimal Number of Clusters

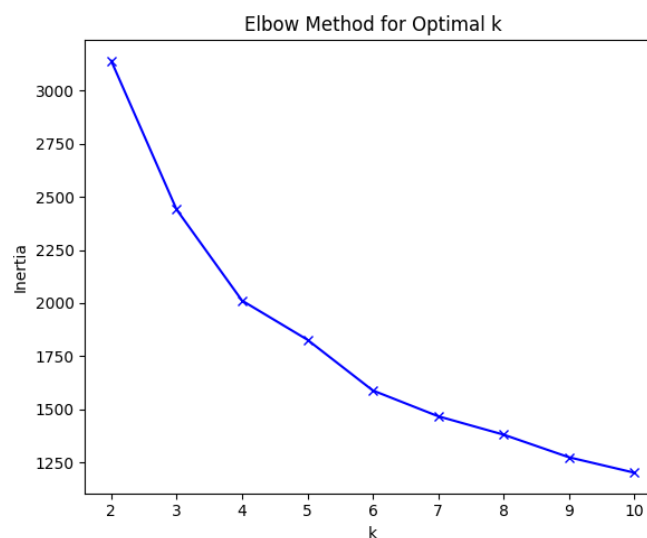


Figure 4: Elbow plot for determining optimal number of clusters

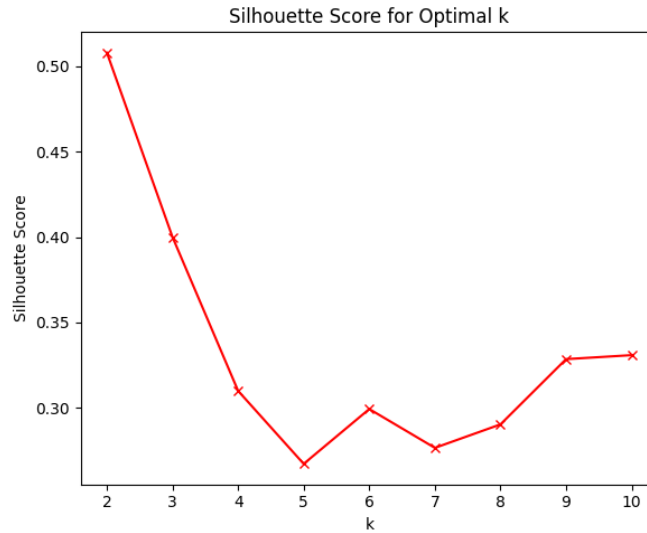


Figure 5: Silhouette plot for different numbers of clusters

Based on these analyses, we determined that 2 clusters provide the optimal segmentation of our customer base.

3.2. K-means++ Clustering Results

Table 1: K-means++ cluster centers and sizes

Gender	Marital Status	Age	Education	Income	Occupation	Settlement Size
33.69724	102689.2	male	single	high school	unemployed / unskilled	small city
48.03166	172742.6	male	single	university	skilled employee / official	small city

Interpretation of K-means++ clusters:

1. Cluster 0 (Established Suburban Families - ESFs): Older, higher-income customers. Average age: 48 years, average income: \$172,743. Predominantly female, non-single, university educated, skilled employees/officials, living in big cities.
2. Cluster 1 (Young Urban Professionals - YUPs): Younger, lower-income customers. Average age: 34 years, average income: \$102,689. Predominantly male, single, high school educated, unemployed/unskilled, living in small cities.

3.3. Agglomerative Clustering Results

Table 2: Agglomerative cluster centers and sizes

Agglomerative Clustering	Customers per cluster
0	999
1	994

Interpretation of Agglomerative clusters:

The Agglomerative clustering results are very similar to the K-means++ results, with two distinct clusters mirroring the ESFs and YUPs profiles.

3.4. Comparison of Clustering Techniques

The two clustering techniques show a high level of agreement, with an Adjusted Rand Index of 0.9098. This indicates that both methods produced nearly identical customer segmentations, reinforcing the robustness of our analysis.

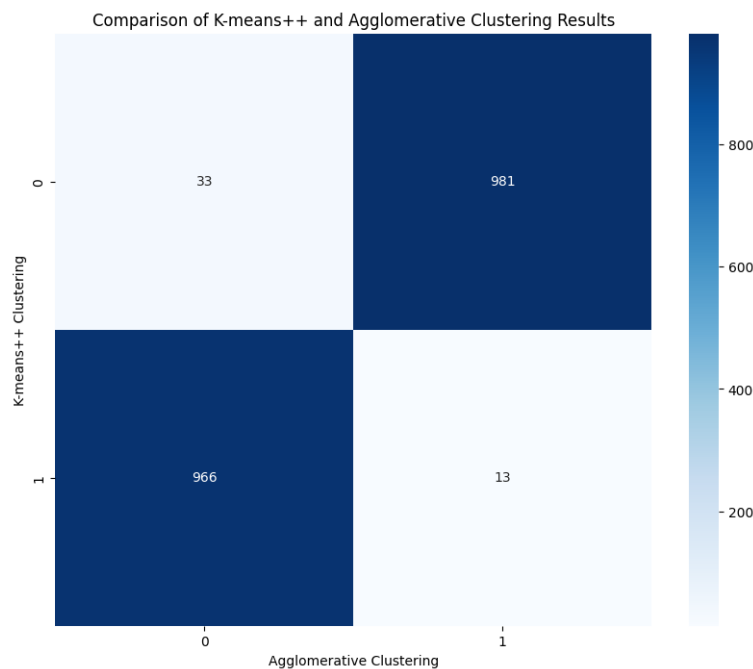


Figure 6: Comparison of K-means++ and Agglomerative clustering results

97.7% of customers were assigned to corresponding clusters by both methods, with only 2.3% assigned differently.

4. Recommendations

Based on the K-means++ clustering results, we recommend the following marketing strategies:

1. For Young Urban Professionals (YUPs):

- Launch a "Career Launchpad" program offering online courses and certifications.
- Introduce compact, multifunctional furniture for small city apartments.
- Create a "Future You" AR app for visualizing long-term financial goals.
- Develop a "Singles in the City" event series for networking and socializing.

2. For Established Suburban Families (ESFs):

- Introduce a "Family Wellness" subscription box with age-appropriate activities.
- Offer executive coaching programs for mid-career professionals.
- Launch a "Smart Suburb" initiative for home automation and security.
- Develop a "Midlife Wellness" program combining health checks and stress management.

5. Conclusion

This analysis has revealed two distinct customer segments within our customer base: Young Urban Professionals (YUPs) and Established Suburban Families (ESFs). By understanding the unique characteristics and preferences of each segment, we can tailor our services and marketing efforts to better meet their needs.

Key takeaways:

- Age and income are the primary differentiators between segments.
- Education, occupation, marital status, and settlement size also play significant roles in defining the segments.
- Both clustering techniques produced highly consistent results, reinforcing the validity of our segmentation.

By implementing the recommended strategies, we expect to see improved customer satisfaction, increased booking rates, and higher customer retention across both segments.

Limitations and future work include incorporating psychographic data, conducting qualitative research, and exploring more granular segmentation within each major segment.