DATA MINING PROJECT

Master in Data Science and Advanced Analytics

NOVA Information Management School

Universidade Nova de Lisboa

Amazing International Airlines Inc.

Bonus Delivery 2 - Fuzzy Clustering

Group 4

Pedro Santos, 20250399

Miguel Correia, 20250381

Pedro Fernandes, 20250418

Tiago Duarte, 20250360

1. INTRODUCTION

A fuzzy clustering approach makes sense in this application because airline **customers may exhibit characteristics of multiple groups**. This method allows customers to belong to more than one segment. We defined two different perspectives: the **behavioural-based perspective** and the **value-based perspective.** Fuzzy clustering was applied for both segmentation tasks, testing different values of *k* (number of clusters) and *m* (fuzziness), to ensure better generalisation.

In the end, we merged both clustering solutions, to obtain a final, combined profiling system:

Cluster	Amount	Description	Strategy
0	7355 (44.1%)	**Mixed Active Travelers**: Stable travelling, indifferent to point rewards	Service-focused rather than loyalty-focused schemes
1	4210 (25.2%)	**Inactive, Low-Activity Travelers**: Never activated, low engagement	Reactivation campaigns with clear value framing
2	733 (4.4%)	**Ultra Value Maximizers**: Low activity, huge seasonality + point redeemers	Points-based engagement, seasonal offers, gamification
3	4389 (26.3%)	**Loyalty Champions**: Active and consistent loyalty users	Point multipliers, companion-focused rewards

2. IMPORTANT INSIGHTS

2.1. Membership degree analysis

By carefully analysing the **assignment confidence distribution**, we can obtain important insights:

- Almost **half of the customer base does not fit cleanly into a cluster** into a single segment (only about half is considered a strong assignment (FCM>0.8). This confirms the presence of gradual transitions between profiles, validating the usage of the fuzzy clustering over hard segmentation.

- **Value-based segmentation shows clearer separation than behavioural segmentation** (54.4% vs. 49.7% strong assignments). Economic variables generate sharper boundaries than behaviour alone.

- Between **17% and 19% of customers exhibit weak cluster assignments**, indicating mixed or evolving profiles. These customers are likely transitioning between segments and represent a key opportunity for targeted, adaptive interventions.

Based on the above, we propose different strategies for the different assignment strength. For **strong assignments** (50–54%), we propose standard segment-level campaigns. For the **moderate ones** (26-33%), A/B testing, refinement of messaging and monitoring. For the **weak ones**, personalized strategies and predictive models to anticipate transition direction.

2.2. Hard vs Soft Clustering

Some key insights can be extracted if we don’t impose a single cluster to each client:

- There is **significant overlap**, since 25.3% of customers in behavioural clustering and 18.7% in value clustering show meaningful membership in two or more clusters.

- There are **truly ambiguous customers that consistently exhibit balanced memberships** (around 33–37%) **across clusters**, indicating genuinely hybrid profiles rather than noise or outliers.

- Behavioural clustering shows 6.6% more secondary memberships than value clustering, confirming that **usage patterns are more transitional and fluid than underlying economic** characteristics.

2.3. Customer Overlap between Segments

To access the quality of the clustering solution, we can check how many clients were segmented into multiple clusters. If many clients are mapped into multiple clusters, that indicates a poor solution.

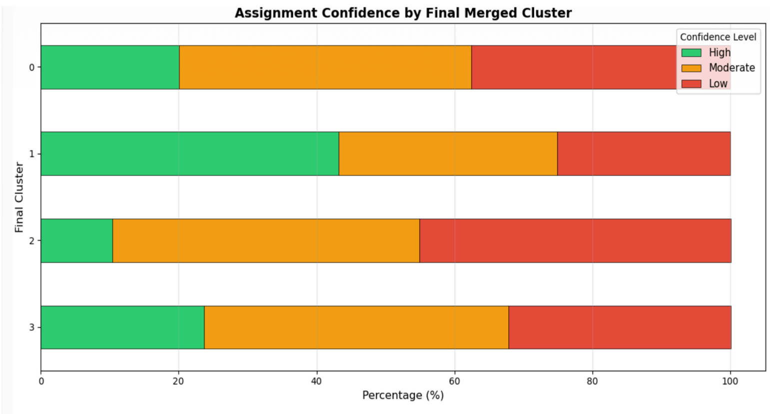
The percentage of **clients mapped to all 3 clusters is close to 0** in both segmentation strategies, proving that the 3 clusters solution produces well-defined and distinct segment boundaries.

We also noticed that **dual-cluster membership is higher for behavioural segmentation** (25.3%) **than for value segmentation** (18.5%), reinforcing once more that behavioural patterns are more fluid and transitional than underlying economic characteristics. As a strategy to capture them we propose hybrid strategies, ensuring periodic monitoring.

Finally, a small (4.8%) but critical group is s**imultaneously transitional in both behavioural and value dimensions**. They require predictive models and manual review for high-value or high-risk cases.

2.4. Uncertainty quantification in business recommendations

For each customer, maximum membership value in each perspective is used as a proxy for certainty.



When quantifying the uncertainty, we obtained as overall confidence the following: 26.4% of clients were mapped with high confidence; 40.2% with moderate confidence and 33.4% with low confidence. Most customers (73.6%) show moderate or low confidence. **Cluster 1 is the most actionable cluster** we obtained (43% of high confidence), and **Cluster 3 requires the most caution**.

3. CONCLUSION

We selected this enhancement - implementing an alternative clustering approach - because assigning each client to a single, fixed cluster is overly simplistic. Acknowledging that the cluster solutions that force clients into one single cluster as **careless** is fundamental when creating strategies to target specific segmentation.

Looking at the plot, it is easy to understand why. Given the visible overlap in the distribution, a fuzzy solution provides a more actionable representation of client heterogeneity.