Identifying higher-potential target markets

By Anna De Sanctis

OBJECTIVES

We aim to optimize our client's marketing budget by identifying the most favorable consumer segments and reducing spend on low probability audiences. This project compares two approaches, unsupervised clustering and supervised CHAID segmentation, to determine which method more effectively identifies high likelihood customers. Clustering groups consumers based on natural similarities in financial and demographic traits, while CHAID directly uses customer behavior to predict purchase likelihood. By evaluating segments using Index for performance and Size for confidence, we determine which approach provides clearer targeting guidance and stronger predictive accuracy.

BEHAVIOR- BASED SEGMENTS

CORE

01

Description: Estimated home value ($349,999 - $799,99), male, and invests in real estate
Index: 208.78
Size: 1.6

02

Description: Estimated home value ($249,999 - $399,999), male, and invests in real estate
Index: 199.25
Size: 1.3

TEST

03

Description: Estimated home value ($249,999 - $399,999), female, and invests in real estate
Index: 147.87
Size: 0.7

04

Description: Estimated home value (<= 124.999), length of residency (5000, 7000), and male
Index: 125.31
Size: 0.8

AVOID

05

Description: Estimated home value (<=124.999),length of residency (>13,000), and female
Index: 42.35
Size: 4.4

06

Description: Estimated home value (<=124.999),length of residency (7000, 8980), owns a home, and male
Index: 20.55
Size: 1.2

QUESTION

07

Description: Estimated home value (>794.999), male, and length of residency (> 13,000)
Index: 72.43
Size: 0.5

08

Description: Estimated home value (<=124,999),length of residency (5000, 7000), and female
Index: 60.65
Size: 1.2

Findings

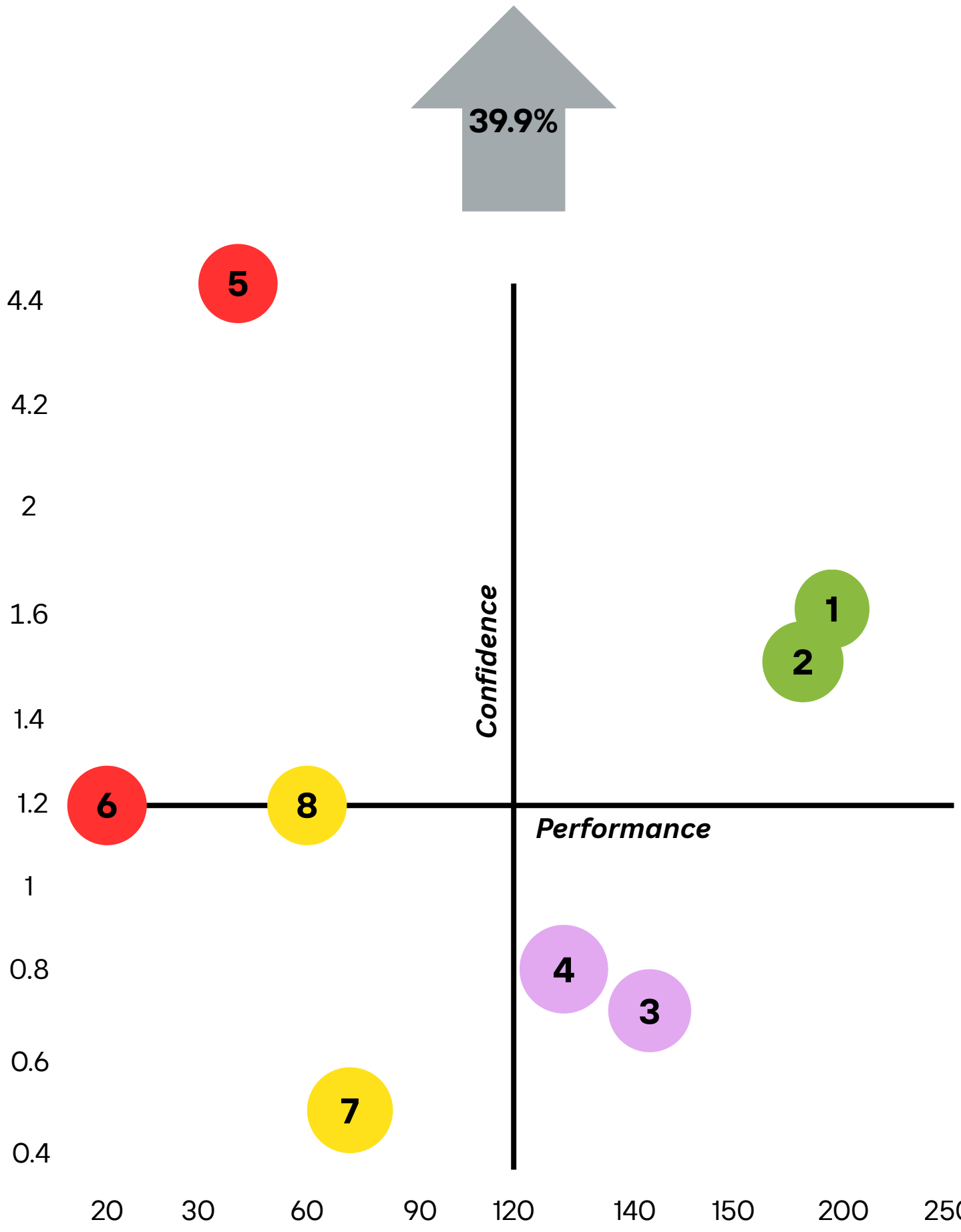
CORE
Segments 01 and 02 include male real estate investors in mid to higher home value ranges. These groups show the strongest alignment with customer behavior and represent the most efficient targets for marketing spend.

TEST
Segments 03 and 04 show moderate potential. Segment 03 includes female investors in mid value homes, and Segment 04 consists of lower value, long term male homeowners. Both warrant cautious testing before scaling investment.

AVOID
Segments 05 and 06 are primarily lower-value, long-term residents with weak investment activity. These groups show limited conversion potential and should receive minimal budget.

QUESTION
Segments 07 and 08 fall below average and lack a clear performance pattern. Additional refinement may be needed before determining whether they offer strategic value.

Confidence (Size) v. Performance (Index) Matrix



Summary For Supervised

This segmentation analysis provides a **clear framework for improving marketing efficiency and return on investment**. By mapping segments across confidence (size) and performance (index), we can distinguish high-value prospects from low-return audiences and allocate budget accordingly.

The **CORE segments** demonstrate the **strongest balance of performance and reliability**, making them the **most strategic investment**. These groups show significantly higher likelihood to convert and should receive the majority of marketing spend.

The **TEST segments** present **moderate opportunity** but require **controlled experimentation** before scaling.

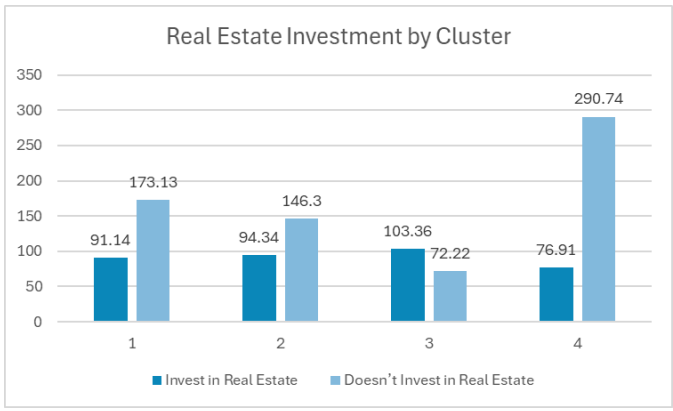
In contrast, the **AVOID segments consistently underperform**, indicating that continued investment would **dilute overall campaign efficiency**.

QUESTION segments require **further refinement or additional data** before meaningful budget allocation decisions can be made.

Overall, this research enables a shift from **broad targeting to precision marketing**. By concentrating resources on **high-performing segments** and reducing exposure to low-yield groups, the client can **increase profitability, improve conversion efficiency, and make data-driven allocation decisions** moving forward.

Unsupervised Clusters

REAL ESTATE INVESTMENT

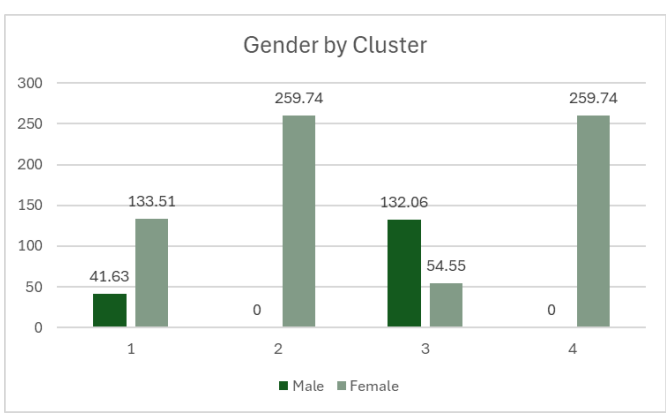


Investment behavior clearly separates the clusters. Cluster 4 is mostly non-investors, Cluster 3 leans more toward investors, and Clusters 1 and 2 are mixed. Real estate participation is a key dividing factor.

NETWORTH

Net worth meaningfully differentiates the groups. Clusters 2 and 4 skew toward higher wealth tiers, Cluster 3 toward lower-to-mid tiers, and Cluster 1 is more balanced. Wealth level is a major structural driver of clustering

GENDER



Gender varies across clusters. Clusters 2 and 4 are predominantly female, Cluster 3 is more male-heavy, and Cluster 1 is balanced. Gender appears to align with financial patterns in shaping segments.

CUSTOMERS

Customer concentration varies, but differences are moderate. Clusters 1 and 4 show stronger customer presence, while Cluster 3 underperforms. However, separation is less distinct than in the CHAID model

SUMMARY FOR UNSUPERVISED

Clustering reveals natural segments primarily driven by real estate investment behavior, net worth, and gender composition. These variables clearly structure the market into distinct financial profiles. Customer concentration is strongest in Clusters 1 and 4, while Cluster 3 shows weaker performance; however, the overall separation between clusters remains moderate. As a result, clustering provides strong insight into who the market is, but it does not sharply isolate the highest likelihood customers for targeting purposes.

FINAL CONCLUSION

While clustering is useful for understanding the structure of the market, CHAID is more accurate for identifying high-probability customers. Because CHAID uses customer status to build segments, it provides stronger predictive power and clearer targeting guidance. Therefore, for marketing allocation decisions, CHAID outperforms clustering in finding the most efficient target markets.