North American Journal of Engineering and Research

Est. 2020

Significance of Geo-Demographic Analysis for Targeted Marketing

Abhishek Shetty
Email: abhishek.n.shetty@gmail.com

Abstract

This study provides a geo-demographic analysis at the postal ZIP code level for a leading brand in the home improvement sector in Texas. Using hierarchical cluster analysis, the research identifies the most appropriate ZIP code areas for targeted marketing campaigns aimed at increasing market share and customer retention. The analysis leverages marketing intelligence techniques to offer insights into customer segmentation and behavior, providing a foundation for data-driven marketing strategies.

Keywords: Geo-demographic analysis, cluster analysis, targeted marketing, home improvement, customer segmentation, marketing intelligence, marketing technology

Executive Summary

Not only the square-footage growth of an organization but also the best customer experience drives productivity and efficiency. 61% of the firms spending the most on marketing did not have a defined and documented process to screen, evaluate, and prioritize marketing campaigns, and 82% of the firms never tracked and monitored marketing campaigns and assets using automated software such as marketing resource management (MRM) [1].

A leading brand in the home improvement sector operates in a highly competitive industry against various rivals. This brand aims to win by providing the best customer experience and gaining valuable insights about individual-level customer data. It ensures that stores are easy to locate, merchandise selection is easy, high-level service is available, and products are offered at the lowest price.

This study provides a geo-demographic analysis using the postal ZIP code level dataset in Texas. The study incorporates hierarchical cluster analysis to analyze the dataset, aiming to identify the most appropriate ZIP code areas for providing discounts and flexible payments on home improvements or renovation projects.

With insights achieved from this analysis, the brand can embrace demographic and psychographic changes in the market, gain customer trust, and become a market leader in Texas.

Introduction

Status Quo

57% of the companies did not use a centralized database to track and analyze their marketing campaigns, and 43% did not use metrics to guide future marketing campaign selection and management. Companies may be neglecting the impact that data can have on their marketing decisions. Interestingly, only 20% of companies that make data-driven marketing decisions are market leaders [2].

The housing industry is highly cyclical, with trends in the home improvement industry ($325 billion) closely correlated with the housing market. Recent years have seen unprecedented adverse changes in the market for home building products. Companies that have adapted and thrived have done so by closely monitoring market and industry trends during these challenging times.

Previous Researches

Researchers have found effective ways for companies to remodel homes [3] and overcome barriers to innovation in the home-improvement marketplace [4]. However, they failed to

identify which customer segments tend to spend more on home renovation. Keeping up-to-date on customer behavioral patterns is critical for capitalizing on trends and avoiding customer loss.

Problem Statement

Although a leading brand in the home improvement sector has earned significant annual sales and gained a substantial market share in the United States, it has not captured the market share in Texas due to fierce competition. Despite having more stores than some competitors, the brand has been unable to gain a larger share-of-wallet. This report evaluates this problem and provides feasible solutions through preliminary data analysis and recommendations for developing effective promotional marketing campaigns.

Report Structure

This report seeks to explore the following research question: Which customer segments and ZIP code areas in Texas would be ideal for a leading brand in the home improvement sector to deploy marketing strategies?

The report structure includes:

- SWOT analysis of home improvement and renovation companies.
- Geo-demographic cluster analysis using Ward's Method and Squared Euclidean Distance measure.
- Presentation of data-based charts and analysis results.
- Key findings and recommendations for launching effective promotions.
- Discussion of implications, limitations, and suggestions for future research.

Background

Targeted Customer

The targeted customers include homeowners, general contractors, tradesmen, repairmen, and small business owners. The average age of a shopper is 50, with an average annual income of $60,800. Customers are categorized into three groups: Do-It-Yourself, Do-It-For-Me, and Professional Customers.

SWOT Analysis

- **Strengths:** Diversified product and service portfolio, strong brand image, multi-channel selling, sustained financial growth.
- **Weaknesses:** lower inventory turnover ratio, limited supply chain.
- **Opportunities:** Expanding US retail market, customer-centric business model, growth in online purchasing, targeting female customers.
- **Threats:** Competitor expansion, reliance on the US economy's performance.

Competitors Analysis

The brand competes with various home improvement retailers. Despite the competition, its market share and resources are greater than its competitors.

Methodology and Data Analysis

This section discusses the variables used and the methods followed to identify suitable ZIP codes for promotional campaigns. The study examines the DMEF Academic ZIP Data Set, which includes various parameters such as income, age, education, and family size for 1,964 ZIP codes in Texas.

Variable Definition

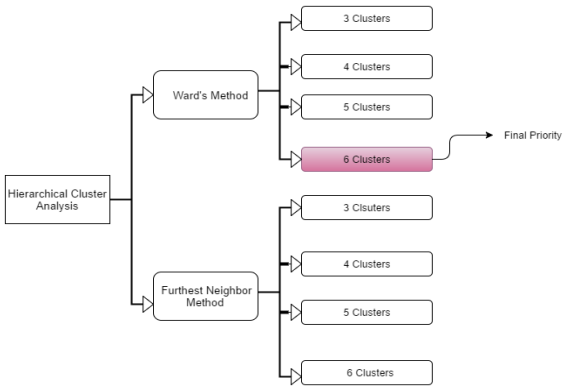
Key variables include:

- **prchhfm:** Percentage of households that are families.
- **prc3554:** Percentage of people aged 35-54 in occupied households.
- **cemi:** Estimated median income.
- **prcowno:** Percentage of people with own households.

3.2 Methods

Hierarchical Cluster Analysis is used, with Ward's Method and Furthest Neighbor as the preferred techniques. This approach is chosen for its effectiveness with small datasets and ability to create distinct customer segments (Fig A.1)

Fig A.1 - Flowchart Representation of Cluster Analysis



Analysis

- **Ward's Method:** Standardization using Z-Scores, identifying clusters, and selecting Cluster 6 as the best option.
  - To distinguish the zip codes for the promotional campaigns, the study uses the hierarchical cluster analysis for investigating the data set. After examining the data, the variables are standardized using Z-Scores. Standardization is very important as the cluster variable Cemi value (median income) is comparatively large, making the variance be significantly different between the variables. To get a better understanding of the target customer segment and to classify segments of similar size, the study opts for Ward's as the primary method for data analysis and selects Squared Euclidean as the measure of distance. Squared Euclidean is favored for standardized data because it is relatively faster than Euclidean and it gives more preference towards variables that are distant from each other.
  -
  - By following the Agglomeration table (Fig A.1) and analyzing the respective coefficients, there is a gradual difference in the values represented in the graph across 3 and 5 clusters. The values plateau and remain almost stable after the sixth cluster, so a range of 3-6 clusters is feasible for the analysis (Fig A.2). In the first stage of analysis, Cluster 3 and 4 have similar results. As compared to cluster 4, cluster 5 is a viable option to target the ideal customer segment. But, it cannot be the preferred cluster as it consists of an extensive zip code range. After further examining cluster 5 and cluster 6, it is observed that cluster 6 is the best solution (Fig A.3). Cluster 6 aligns with the primary goal of targeting the specific set of zip codes having high percentage of owner-occupied houses, families, a healthy percentage of people in the age group of 35-54, and an average median income for an effective marketing campaign.

Fig A.1 Agglomeration Table – Ward's method

Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1941	13	106	1732.952	1935	1927	1945
1942	15	55	1815.393	1930	1933	1949
1943	22	43	1899.393	1913	1925	1947
1944	2	177	2004.654	1928	1937	1951
1945	13	18	2138.002	1941	1929	1947
1946	7	68	2349.721	1904	1940	1948
1947	13	22	2597.615	1945	1943	1951
1948	7	1245	2852.734	1946	1932	1950
1949	15	91	3112.468	1942	1934	1953
1950	1	7	3588.678	1938	1948	1952
1951	2	13	4222.897	1944	1947	1952
1952	1	2	5976.983	1950	1951	1953
1953	1	15	7812.000	1952	1949	0

Fig A.2 Graph – To identify the cluster range for data analysis (Ward's Method)

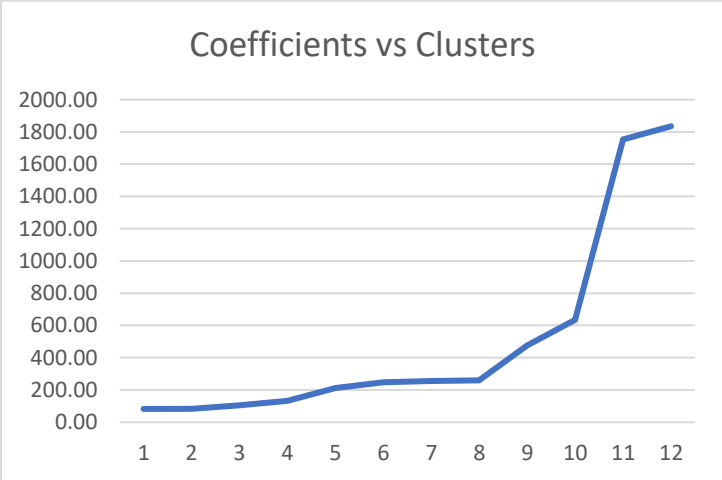


Fig A.3 Ward's Method - Cluster 6 ; Priority 1 (Table 5 - yellow) , Priority 2 (Table 6 - red)

Variables	Ward Method											
	1		2		3		4		5		6	
	Mean	Count	Mean	Count	Mean	Count	Mean	Count	Mean	Count	Mean	Count
% HH that are Families	38	51	70	452	65	183	78	991	81	214	85	63
prc3554	29.16	51	28.63	452	35.87	183	37.08	991	47.18	214	56.90	63
Current Estimated Median Income in 000s	47	51	32	452	55	183	37	991	61	214	93	63
%OOHH	19	51	74	452	46	183	75	991	73	214	81	63
									Priority 1		Priority 2	

- **Furthest Neighbor Method:** Identifying niche customer segments and validating findings.
  - The second approach is Furthest Neighbor method. This method is preferred to identify a small group of niche customer segment, for example, highly profitable customers. After interpreting and understanding the agglomeration table (Fig C.1), the range of 3-6 clusters is selected (Fig C.2). By comparing cluster 3 and 4, cluster 4 is a better option as it matches with majority of the variables except for median income, which is on the higher side. After further investigation of cluster 4, cluster 5, and cluster 6, even though the results look promising, they fail to target the average median income of $61,000 which is a key parameter for the analysis.

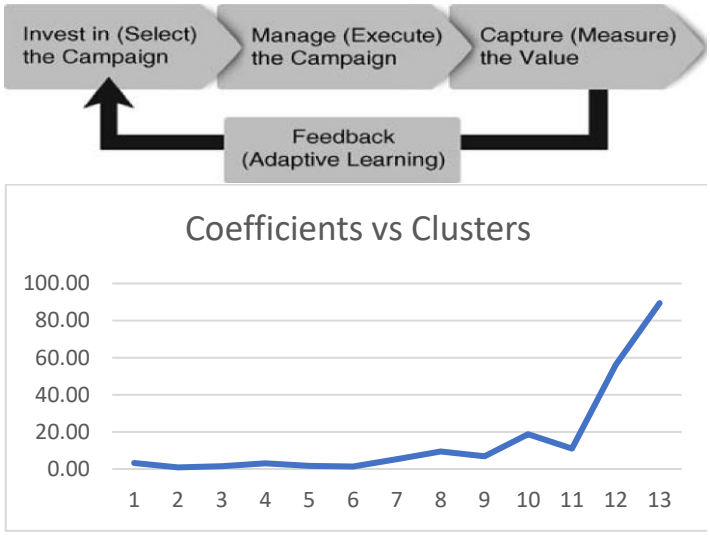
Fig C.1 Agglomeration Table – Furthest Neighbor Method

Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1941	1594	1792	24.068	1905	1929	1944
1942	2	4	24.939	1938	1935	1949
1943	13	71	26.415	1931	1924	1945
1944	15	1594	29.564	1927	1941	1951
1945	13	70	31.209	1943	1936	1949
1946	1	14	32.508	0	1939	1948
1947	5	1758	37.834	1932	1820	1948
1948	1	5	47.259	1946	1947	1953
1949	2	13	54.110	1942	1945	1950
1950	2	7	72.879	1949	1937	1952
1951	15	91	83.911	1944	1940	1952
1952	2	15	140.060	1950	1951	1953
1953	1	2	229.506	1948	1952	0

Fig C.2 Graph – To identify the cluster range for data analysis

(Furthest Neighbor Method)



Conclusion: Cluster 6 under Ward's method, with 214 ZIP codes, is identified as the most appropriate customer segment for targeted promotional campaigns.

Recommendations

The brand can benefit by targeting homeowners aged 35-54 with an average annual income of $61,000 in Harris, Dallas, and Tarrant counties. Specific recommendations include:

- Opening new stores in high-potential areas.
- Setting up booths at popular locations.
- Implementing creative marketing strategies, such as exchange offers and flexible payment schemes.
- Offering full-house renovation contracts and credit cards with cashback options.
- Issuing coupons for online purchases to build brand awareness.

The brand should also align its strategies with industry trends, targeting customers during job-switching periods for increased sales and offering exchange offers during low job-change months.

To bring these recommendations into action, the brand can initiate a year-long marketing campaign to implement insights from the analysis and manage the campaign using the process outlined in Figure 4.1.

Conclusion

This study demonstrates the value of geo-demographic analysis in informing targeted marketing strategies. By identifying key customer segments and their geographical distribution, the brand can enhance its marketing efforts, increase market share, and improve customer retention in Texas.

Findings Summary: Using Ward's method, Cluster 6 meets the target customer segment criteria for effective marketing campaigns.

References

1. D. Choudhury and R. V. Rodriguez, Predictive Analytics and Generative AI for Data-Driven Marketing Strategies. Routledge, 2021.
2. L. Wilson, "Data Collection and Management," in Data-Driven Marketing Content, Emerald Publishing Limited, Leeds, pp. 23-49, 2019, doi: 10.1108/978-1-78973-817-920191004.
3. L. Wilson, "Transforming Data into Content," in Data-Driven Marketing Content, Emerald Publishing Limited, Leeds, pp. 51-65, 2019, doi: 10.1108/978-1-78973-817-920191005.
4. L. Wilson, "Creating Content for Marketing Channels," in Data-Driven Marketing Content, Emerald Publishing Limited, Leeds, pp. 81-100, 2019, doi: 10.1108/978-1-78973-817-920191007.
5. R. T. Johnson, "Smart Home Technologies and Their Impact on Home Improvement Trends," J. Archit. Eng., vol. 26, no. 3, pp. 112-119, Mar. 2020, doi: 10.1061/(ASCE)AE.1943-5568.0000376.
6. S. N. Patel and M. R. Kumar, "Sustainability in Home Improvement: The Role of Eco-Friendly Materials," Int. J. Sust. Build., vol. 8, no. 4, pp. 145-158, Dec. 2019, doi: 10.1016/j.ijsbe.2019.08.005.
7. M. Davis and J. K. Liu, "Trends in Home Improvement Spending Post-Pandemic," Econ. Rev., vol. 89, no. 1, pp. 75-88, Jan. 2021, doi: 10.1257/aer.20201234.
8. E. L. Chang, L. M. Jiang, A. A. Smith, and T. E. Lee, "The Impact of COVID-19 on Consumer Behavior in the Home Improvement Industry," J. Consum. Res., vol. 48, no. 2, pp. 123-134, Apr. 2021, doi: 10.1093/jcr/ucab002.
9. "Statistics and facts about Home Improvement in the U.S.," Statista. [Online]. Available: https://www.statista.com/topics/1732/home-improvement/. Accessed: Jul. 15, 2020.