Marketing – Customer Insights

Customer Groups

Prepared by: Aishetu Amara, Carla Jackson, Delphine Mason, Zijiang Zhang

Delphine Ma

TABLE OF CONTENTS

Executive Summary

The objective of our analysis was to provide management with a targeted list of customer zip codes in Texas to market the new electric BMW i4 Model, which is scheduled for production in 2021. Furthermore, the underlying purpose of this analysis is to increase our campaign efficiency for a higher ROI on our marketing efforts. In order to do this, we needed to fine-tune our customer profile to uncover market opportunities in zip codes where our target demographic resided. Our criteria were based on four variables which we used to segment our target demographic: Household Income, Age, Gender, and Households with 2 or more Cars.

We performed a Hierarchical Cluster Analysis (HCA) and compared the results from two cluster methods, Wards and Furthest Neighbor. As a result, we were able to identify clusters of zip codes representing wealthier customer groups which met the variable criteria above. The implications of this is that these identified customer groups are closely related so the marketing campaigns can be coordinated and designed with similar messaging to cater to the common needs of both segments. Other considerations used in our analysis was proximity to urban areas and BMW dealerships.

Finally, our geo-demographic analysis led us to conclude that our greatest marketing opportunities were in zip codes where household incomes averaged between \$100K – \$184K. Our report identifies 151 zip codes for our marketing campaign, representing a population total of 4.5M potential customer leads and projected gross revenues of ~\$28.9M. Our recommendation is to market through two distribution channels – digital and mail. In addition to this, since our targeted consumers value technology, and luxury; we also recommend that the company launch an app-based campaign. For example, BMW can develop a VR app that would allow customers to build their own i4 and mail targeted consumers disposable VR headsets to simulate a virtual driving experience.

TECHNICAL REPORT

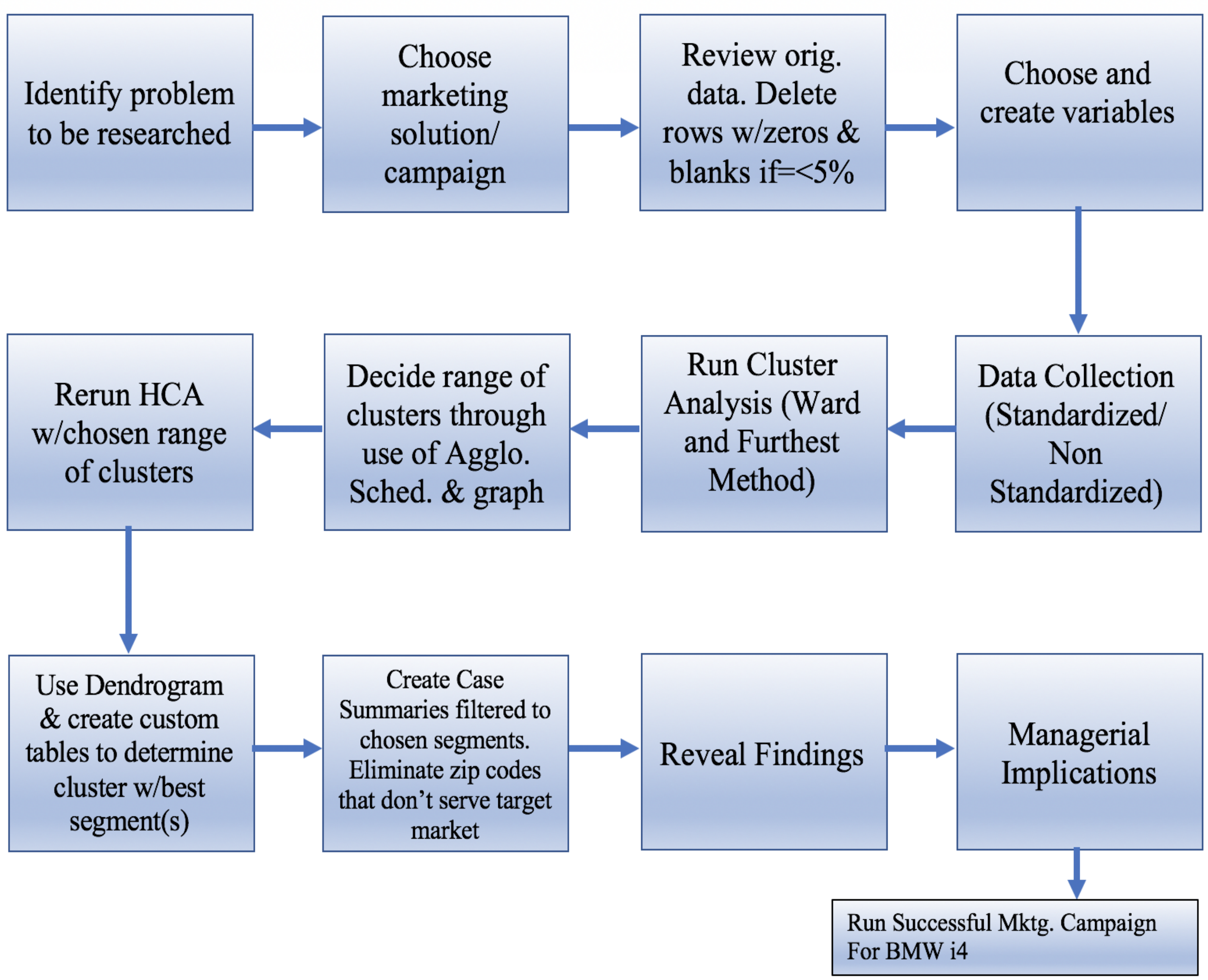
Overview:

The data that was used for this analysis is census data of Texas zip codes and some proprietary information. It was prepared by the marketing company Nejad and friends.

After manipulating the data (i.e. eliminating rows with multiple zeros, and adding a new variable) to serve our target market, we performed a hierarchical cluster analysis using the Ward Method with Euclidean Distance Non-Standardized and decided to evaluate a range of three to six clusters. We reran our Hierarchical Cluster Analysis with this range of solutions. Upon review of the cluster segment results, we determined segments two and three in Ward Clusters five best fit our target. These segments had the highest incomes within our desired range of income, as well as a significant and manageable amount of zip codes. In compiling our final list of target zip codes, we analyzed the case summaries and eliminated any zip codes that did not meet our requirements. Details about the zip codes that were excluded are included below in the data analysis section. Our final target includes 151 zip codes with a total population of 4,533,161. The details of our analysis are presented in Figure 1.

Process for Geo-Demographic Analysis

Figure 1. Analysis Flow Chart



Data Analysis:

Table 1. Variable Explanation

Variable Name	Definition	Reasons	Target
Income	The mean of the household incomes in the zip code	We can choose the income range that we need	Incomes between 100k-1804k
Age	% of occupied housing units w/householders between 35-44 in the zip code	We can choose the age range that we need	Gen X's & Y's purchase comparable Tesla Model 3
Gender	% of males living in the zip code	Males are more likely to buy a BMW i4.	Seeking high % of males
Households w/2 or more vehicles	% of residents who own 2 or 3 or more cars	Electric cars are generally bought by HH who already own other cars.	Car owners who own 2 or more cars.

Income:

As you can see from the Variable Explanation in Table 1, we chose the Income variable because research shows that our target market is males with incomes that average $128,000 annually[1]. Furthermore, according to the U.S. Energy Information Administration, 67% of households that own an electric vehicle make more than $100,000 per year[2]. However, this income is based on major metropolitan areas like Orange County or Los Angeles, California. Our target owners live in Texas where the cost of living[3] is at least 25% lower, not including home purchases. Therefore, around $100k in Texas is at least equal to $128k in California where most Tesla owners are located. On the higher end of the income spectrum, customers earning above $200k, purchase higher end luxury vehicles[4] like Ferrari or Porsche. This change in customer preferences driven by income differentiation, led us to create income thresholds between $100 to $184k. Another consideration for these income limits was the affordability factor. The estimated price of our BMW i4 will be ~$55K, which is similar to its major competitor Tesla model 3's average price of $59,300[5]; as a result, customers earning an income of at least $100K would reasonably be able to afford the new model.

Age:

The BMW i4 owner[5a] is the most comparable to the Tesla Model 3 owner. These owners skew younger than higher end luxury car owners, and their higher desire for a flashy car is often more accessible than homeownership (56% of Tesla Model 3 owners do not own their homes)[6]. Therefore, we included the Age variable in our analysis, and decided to focus on a specific range of ages that are part of the Generation X & Y[7] (35-44). Generation X tends to spend more carefully and look for well-made products, making a BMW an optimal product for them. Generation Y is "confident, self-expressive, liberal, upbeat and receptive to new ideas and ways of living" [7a], therefore, purchasing the BMW i4 fits in perfectly with this model of thought. Overall, customers within these two generations are drawn to our electric car models for the technology and innovation, luxury product experience, and eco-friendliness (supportive of sustainability).

Gender:

There is a gender divide when it comes to electric vehicles (EVs) purchases. Women tend to be more practical, often preferring cars that can accommodate family needs with lower price points, while men are usually more into technology and more willing to spend their money on automobiles. According to a 2018 survey by AAA, women are more likely to cite the reason for buying an electric vehicle was for environmental concerns than men (90% vs. 68%)[8], however most women still question the security of EV's. Furthermore, knowing that males purchase 84% of Tesla Model 3s[8a] it was essential to have the Gender variable in our analysis. Given these facts, we strongly believe that the high percentage of males who like the Tesla Model 3 will cross over into buying our BMW i4.

Households with 2 or More Vehicles:

Our models also resonate with customers who own multiple cars for different usages. The BMW i4 is a premium product suitable for selective occasions (i.e. short trips, weekend driving, etc.) as a result it is typically not the primary source of transportation for our customers.[9] This variable also served as an indicator of wealth - the idea being that high-end customers are likely to have more variability in their cars. Therefore, in order to identify segments that met our criteria, we needed to have a multiple-car ownership variable that had zip codes with owners of two or three or more cars. To increase our chances of success, we created a new variable, which is the sum of the variables indicating ownership of two or more and three or more cars: prc2vehicle + prc3vehicle = prc2_3vehicle (Households w/2 or more vehicles.

Data Preparation & Analysis:

Before performing hierarchical cluster analysis, we reviewed the Excel data. We saw that there was a large number of rows that had zeros or blanks instead of larger numeric data. In order not to have skewed data, we deleted many of these rows by sorting for zeros using column F (NoOccupiedHousingUnits had the greatest number of zeros). This added up to only 1.7% of the total data, an amount that is less than the accepted 5% deletion rate of the total data (Of the 1936 rows, 33 were deleted). Next, we determined that we would need to combine two car ownership variables in order to make one variable that includes all owners of two or more cars in that zip code (see Table 1). Then, due to the large differences in the variances (see Table 2), we thought that standardizing the data would bring us the best segments for our final count. Interestingly, our hypothesis was incorrect.

Table 2. Differences in Variances

Descriptive Statistics						
	N	Minimum	Maximum	Mean	Std. Deviation	Variance
Income	1,849.00	16,875.00	312,420.00	69,634.71	28,669.80	821,957,438.46
Age	1,902.00	-	100.00	16.89	9.68	93.65
Male	1,896.00	-	100.00	56.77	10.42	108.64
Households w/2 or more vehicles	1,894.00	-	100.10	77.80	13.89	192.83

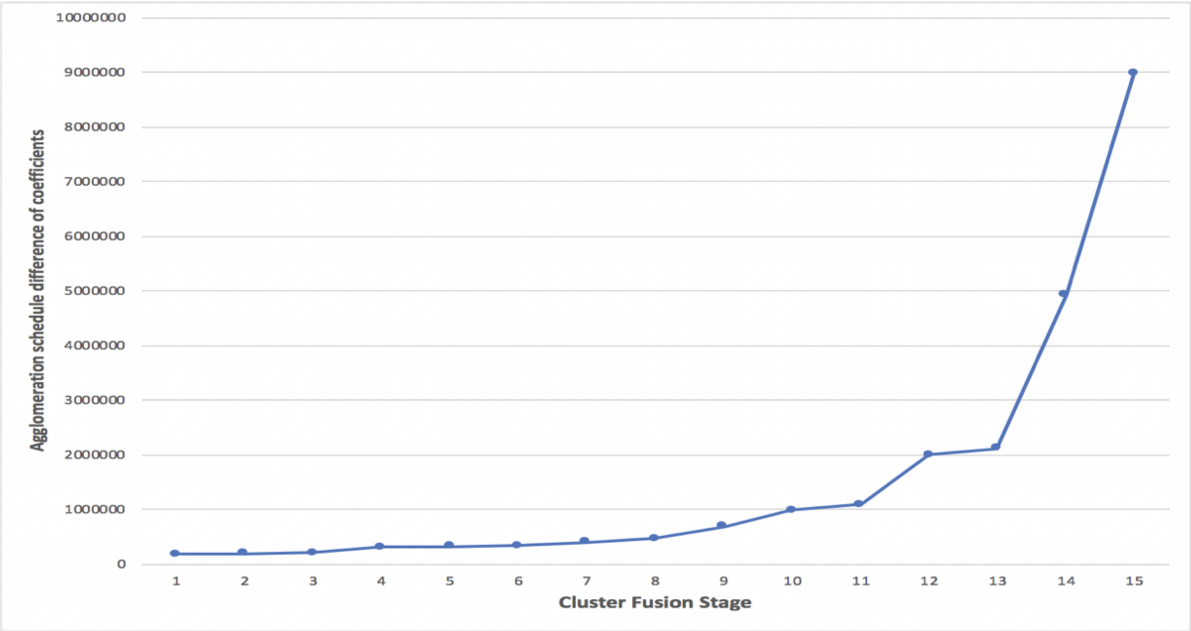
The surprising results came once we performed a hierarchical cluster analysis using the Ward Method with Euclidean Distance Non-Standardized. Following this, we compared the results with all of the previous HCAs (see Table X near the end of the report). Our best segments were found with the former method.

Table 3. Agglomeration Schedule

Agglomeration Schedule									
Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage			
	Cluster 1	Cluster 2		Cluster 1	Cluster 2				
1839.00	7.00	116.00	4,217,577.02	1835.00	1817.00	1842.00		396,713.95	
1840.00	60.00	64.00	4,614,290.97	1823.00	1833.00	1846.00		466,507.39	
1841.00	2.00	4.00	5,080,798.36	1832.00	1827.00	1845.00		690,182.80	
1842.00	7.00	8.00	5,770,981.16	1839.00	1830.00	1845.00		986,953.39	
1843.00	9.00	22.00	6,757,934.54	1838.00	1834.00	1846.00		1,085,284.72	6.00
1844.00	1.00	21.00	7,843,219.27	1837.00	1836.00	1847.00		1,996,206.06	5.00
1845.00	2.00	7.00	9,839,425.33	1841.00	1842.00	1848.00		2,107,392.36	4.00
1846.00	9.00	60.00	11,946,817.69	1843.00	1840.00	1847.00		4,912,500.25	3.00
1847.00	1.00	9.00	16,859,317.94	1844.00	1846.00	1848.00		8,961,736.37	
1848.00	1.00	2.00	25,821,054.31	1847.00	1845.00	0.00			

Included in the process was a review of our Agglomeration Schedule (see Table 3). Which after graphing the difference of coefficients (see Graph 1) led us to evaluate a range of 3-6 solutions.

Graph 1. Differences of Coefficients for Graphing Using the Ward Method



We reran our hierarchical cluster analysis with this range of solutions, eventually determining by reviewing our Dendrogram and custom tables, that Segments 2 and 3 amongst Ward Clusters 5 best fit our target numbers of all the Ward HCAs run. The segments have the best incomes in the range of incomes we wanted, and a significant amount of zip codes in their counts. This was unexpected because we assumed that one of the standardized clusters would have the best results due to the 'weight' of the Income variable. Additionally, we chose Ward Cluster 5 and segments 2 & 3 because when we ran Furthest Squared Euclidean Non-Standardized, Cluster 8 with segments 2 & 3, the results had similar counts but lower incomes. The Furthest Euclidean Non-Standardized had the same results as the Furthest Squared Euclidean Non-Standardized. So, again, we considered Segments 2 and 3 amongst Ward Clusters 5 to be our best result.

Final Zip Codes Results:

There were 291 zip codes in our chosen segments. However, the total population of those zip codes equaled more than six million people (6,760,442.00). Since our digital, online, and mailing campaign is already comprehensive, we felt that a $3 million dollar budget could not cover all the marketing costs for our target zip codes. Also, not everyone in those zip codes had an income between $100-184k. Therefore, in order to stay within our marketing budget and to target the correct demographic, we applied the following methods:

Deleted all zip codes with households earning less than 100k because the cost of the car would be outside of the owner's means.

Deleted all zip codes over with households earning over $184k because these drivers would probably buy a more expensive luxury car.

Then, we cross referenced the remaining zip codes with a zip mapping application[10] to identify which ones were located in wealthier areas.

We compared each zip codes' distance from the nearest BMW dealership. Since most dealerships are located in urban locations, we believe that if a residence were over an hour away from a dealership, it would not be worth our marketing dollars to promote the BMW i4 in these areas.

This filtering process left us with a total of 151 zip codes, which is a population of 4,533,161. All of these residents will be contacted through purchased email and address lists. Additionally, we will cross reference these lists with an acquired list of age data that specifies exactly who in the household is between 35-44, so that we target the correct age group. Finally, we will also create a special VR campaign for only the top 100 residents in Denton county, due to the mean of their income reaching almost $185k.

Key Findings:

Finding 1

As stated, our team decided that the best choice for our target market of consumers (i.e. Males whose income falls between 100k and 184k, are between the ages of 35-44, and who own more than two cars) were found in segments 2 and 3 of the 5 segment cluster for Ward Method. This segment had a good number of possible zip codes to target without going above our budget (After filtering, we have 151 zip codes).

Finding 2

In addition, our selected segment has the best of the targeted income range we needed. All of the other segments in the clusters had incomes that were too low (see Table 4).

Table 4. Ward Euclidean Non-Standardized

	Segment 1	Segment 2	Segment 3	Segment 4	Segment 5
	Mean	Mean	Mean	Mean	Mean
Income	74,416.00	102,489.00	155,342.00	56,670.00	40,028.00
Age	17.30	19.00	20.80	16.10	14.80
Male	56.10	55.60	58.60	56.30	56.60
Households w/2 or more vehicles	81.14	81.36	83.66	76.65	69.84
Number of zip cdes (Count)	568.00	191.00	100.00	706.00	284.00

Finding 3

We would rather that our chosen segment had a higher percentage of males occupying the housing units compared to the other segments. But the decimal points of difference were negligible.

Finding 4

The five urban Texas counties[11] that were listed the most often were: Harris (34), Dallas (19), Travis (17), Bexar (13), Collin (12). Additionally, to demonstrate a sample of zip codes, below are the top five counties sorted by income (see Table 5):

Table 5. Top 5 Counties Sorted by Income and Age

Zip Codes	Population	County	Post Office Name	Income	Age	Male	HH W/Two Or More Vehicles
75022	24,035.00	DENTON		$ 183,449.00	25.7	57.1	94.3
78732	16,851.00	TRAVIS	AUSTIN	$ 181,374.00	34.3	56.1	91.8
77382	38,138.00	HARRIS	SPRING	$ 173,322.00	23.2	62.3	87.9
77479	85,514.00	FT BEND	SUGARLAND	$ 163,208.00	23.4	57.3	90.1
78739	19,564.00	TRAVIS	AUSTIN	$ 161,724.00	33	54.5	93.5

APPENDICES

Appendix A – Explanation of the Hierarchical Cluster Analyses that was run

Our first steps, before running any HCAs, was to delete the rows in our data set that held blanks or zeros (which amounted to less than 5% of the total), and to create a variable to represent Households with two or more cars. Afterwards, we standardized our numbers and ran Furthest Neighbor and Ward Methods with Squared Euclidean as the distance. We standardized the numbers because we did not want the Income variable outweighing all of the other variables, and we used Squared Euclidean in order to consider outcomes that had highly differentiated segments. Having different segments could lead to one perfect segment with Variable means that would service our marketing campaign easily. We repeated the process for both methods using Euclidean, but still with standardized numbers.

After reviewing our Agglomeration Schedules for all methods and distances, and looking at differences in the coefficients, we decided to use a range of 3-6 solutions for each method. We reran our Hierarchical Cluster Analysis with this range of solutions, named each of our cluster solutions, and created tables for each solution in order to show the mean and count for each of the variables in each of the clusters.

After examining all of these clusters, we repeated the HCA process using the Ward and Furthest Neighbor Methods again, but with Squared Euclidean distance and unstandardized numbers, then with Euclidean distance and unstandardized numbers. We analyzed the results of all of the HCAs. Furthest Neighbor's Method had more distinct clusters, as expected. However, the Ward Method, using unstandardized numbers and Euclidean distance, resulted in outcomes with a plethora of similar actionable segments to choose from. This was better for our objective of targeting a specific customer segment (Males whose income falls between 100k and 184k, are between the ages of 35-44, and who own more than two cars) with similar tastes for a campaign. See comparison of clusters in Table A below:

Table A. Comparison of Final Clusters *(Mean values given)*

	Squared Euclidean Standardized		**Euclidean Standardized**	
	Ward	**Furthest**	**Ward**	**Furthest**
Income	154,853	80,522	151,780	80,522
Age	24	22	24	22
Males	57	55	58	55
Households w/ 2-3 Cars	88	79	88	79
Count	79	369	87	369
Cluster #	6	5	5	
Evaluation	**Ward** - Count too low. **Furthest** - Low income and count		**Ward** -population size too low. **Furthest** - below income target at 8K and population count too low	

	Squared Euclidean Non-Standardized				**Euclidean Non-Standardized**			
	Ward		**Furthest**		**Ward**		**Furthest**	
	2	3	2	3	2	3	2	3
Income	104,458	139,693	100,371	136,766	**102,489**	**155,342**	100,371	136,766
Age	20	22	19	22	**19**	**21**	19	22
Males	55	58	56	58	**56**	**59**	56	58
Households w/ 2-3 Cars	81	83	81	84	**81**	**84**	81	84
Count	165	78	222	69	**191**	**100**	222	69
Cluster #	5				5			

Evaluation	Overall better results, however because we were going to filter further, we preferred a higher count	**Final Choice:** We chose Cluster 5 segments because when we filtered for income and distance from the dealer, the count for the Furthest Method was < 138

Appendix B - Financials

Below are preliminary cost and revenue projections for our recommended marketing campaign over the course of approximately 2 years (2021 - 2023). We used the weighted close rates for the dealerships and digital channels to calculate the potential number of acquired leads from our total target audience. We also estimated the marketing acquisition costs for our targeted customers at $3M based on postage, digital marketing, and the VR headsets costs.

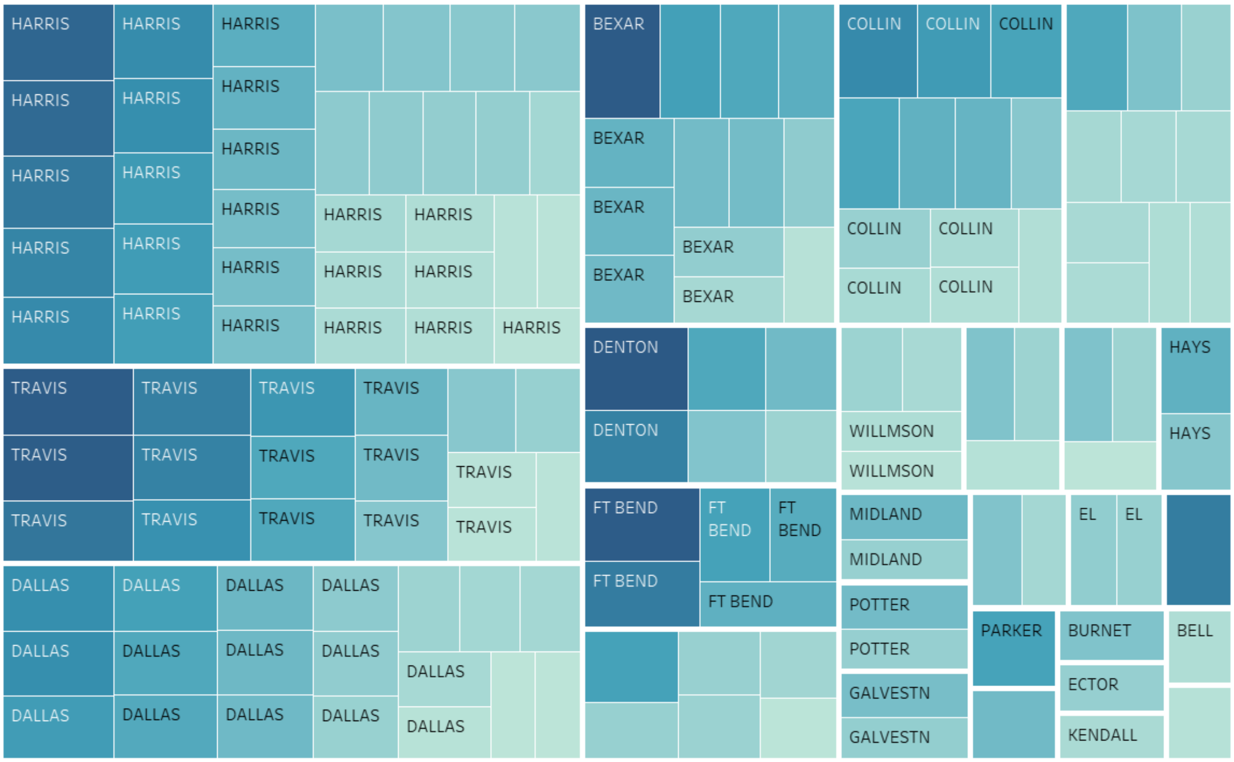
Table B. Estimated Costs & Revenues

Estimated Costs & Revenues *($ in 1000's)*	
Estimated Cost of BMW i4	$55,000
Total Population	**4,533,161**
Close Rate *(car dealership)*	12.50%
Close Rate *(digital marketing)*	3.00%
Weighted Close Rate (*total*)	**11.60%**
Population (Acq. Leads)	**523,580**
Marketing Costs *(digital & mail)*	3,000
Total Revenue	**$28,796,905**

Appendix C - Heatmap of Targeted Counties per Income

Image A. Heatmap

Income per County



Note: Regions vary in color from Darker to lighter based on household incomes. Darker regions represent higher income levels and lighter regions signify lower income levels. Income ranges from ~100K – ~184K. This map was used to get a feel for the counties with the highest wealth concentrations. For example, Harris, Travis and Dallas would be target markets per our managerial objectives.

ENDNOTES

1. Fortuna, Carolyn. “What Do We Know About Tesla Owners?” Clean Technica, 1 June 2019, https://cleantechnica.com/2019/06/01/what-do-we-know-about-tesla-owners/.

2. Orr, Isaac, “Electric Cars are Mostly for Wealthy People, and You’re Subsidizing Their Purchase.” Center of the American Experiment, 22 May 2018, https://www.americanexperiment.org/2018/05/electric-cars-mostly-wealthy-people-youre-subsidizing-purchase/#:~:text=Employee%20Freedom-,Electric%20Cars%20are%20Mostly%20for%20Wealthy%20People,You're%20Subsidizing%20Their%20Purchase&text=According%20to%20data%20from%20the,more%20than%20%24100%2C000%20per%20year.

3. Segal, Troy. “Cost of Living in Texas vs. California: What's the Difference?” Investopedia, 5 February 2020 https://www.investopedia.com/ask/answers/100214/what-cost-living-difference-between-texas-and-california.asp.

4. Liu, Kun. “Car Price vs Income.” Source: carpoint.msn.com, 14 Sept. 1997 https://www.csparks.com/bmw/CarPriceVsIncome.xhtml.

5, 5a. Sachgau, Oliver and Rauwald, Christopher. “BMW Bets on New EV to Reclaim Crown in Entry-Level Luxury Sedans.” Bloomberg News, 3 March 2020, https://www.bloomberg.com/news/articles/2020-03-03/bmw-bets-on-new-ev-to-reclaim-mojo-in-entry-level-luxury-sedans.

6. “Tesla Model 3 Demographics: Income, Age, Gender and More.” Hedges and Company, Hedges and Company, 2019, https://hedgescompany.com/blog/2019/03/tesla-model-3-demographics-income/.

7,7a. Wikipedia Contributors. “Millennials.” Wikipedia, Wikimedia Foundation, 13 Mar. 2019, en.wikipedia.org/wiki/Millennials.

8, 8a. Evarts, Eric C., “Why do Men Seem More Invested in Electric Cars Than Women”, 25 June 2019, https://www.greencarreports.com/news/1123764_why-do-men-seem-more-invested-in-electric-cars-than-women-take-our-twitter-poll12.

9. Orr, Isaac, “Electric Cars are Mostly for Wealthy People, and You’re Subsidizing Their Purchase.” Center of the American Experiment, 22 May 2018, https://www.americanexperiment.org/2018/05/electric-cars-mostly-wealthy-people-youre-subsidizing-purchase/#:~:text=Employee%20Freedom-,Electric%20Cars%20are%20Mostly%20for%20Wealthy%20People,You're%20Subsidizing%20Their%20Purchase&text=According%20to%20data%20from%20the,more%20than%20%24100%2C000%20per%20year.

10. “Key Zip.” USNaviguide LLC., January 2010, http://maps.huge.info/zip.htm

11. White, Steve et al. “Urban Texas” Texas Demographic Center, August 2017, https://demographics.texas.gov/Resources/publications/2017/2017_08_21_UrbanTexas.pdf.