**Strategies for Increasing Sales Through Data Analysis**

David Reese

Colorado State University Global Campus

MIS 543 Enterprise Performance Management

Dr. Kimberly Ford

12/20/20

**Strategies for Increasing Sales Through Data Analysis**

Every business owner wants to increase sales, but not at the expense of additional profits. Maintaining margins is vital to any operation. Finding new ways to achieve the goal increased sales at minimum costs is where data analytics comes in to play.

For this project, we have a dataset of over 28,000 records to work with a new wrinkle. The data includes information based on the MicroVision Market Segmentation System. The system includes 50 clusters that describe various demographic factors. The clusters have titles which describe the market you might be looking for. By combining data from various sources, the company is able to create accurate descriptions which assist in describing potential markets.

This segmentation will help in our search for new clients, and increased profits. The logical place to start the search would be in the existing data and customer base. While it is common knowledge that gaining a new client is more than five times the cost of retaining an existing client, the reason may not be clear. “The success rate of selling to a customer you already have is sixty to seventy percent, while the success rate of selling to a new customer is five to twenty percent” (Marr, 2019 ¶ 4). By using the existing datasets, the chances of success go up dramatically while keeping costs to a minimum.

Using the data and the segmentation information, the top six clusters were identified by the variable Sales\_per\_visit. This variable is used to track the total amount spent by that customer. By using the demographic information from the clusters, a picture of a typical customer appears.

The picture that emerges from the top six clusters is one of a customer who owns a home, is between 35 and 60 years of age, has above average income, and a college degree. The have executive level positions in most cases. Children are present but older.

**Figure 1**

*Chart of Top 6 Clusters*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **The Top Six** | | | | | | | |
| Demographic | Cluster One - Upper Crust | Cluster 4- Mid Life Success | Cluster 8 - Movers and Shakers | Cluster 10- Home Sweet Home | Cluster 15 - Great Beginnings | Cluster 16 - Country Home Families | **Totals** |
| Observation Rank and Frequency | 2 ( 2716 ) | 3 ( 2284) | 5 (1430 ) | 1 (3488) | 6 (1327) | 4(1893) | **13138** |
| Household Income | Highest | High | High | Upper Middle | Middle | Lower Middle |  |
| Age | 45-59 | 40-54 | 35-49 | 50-65 | 25-34 | 40-54 |  |
| Homeowner | Yes | Yes | Renter | Yes | Renter | Yes |  |
| Children present | Yes | Yes | No | Yes | No | Yes |  |
| Education ranking | Post Graduate | Post Graduate | College | College | College | High School |  |
| Job Level | Executive | Executive | Mid Executive | Retired | Support | Blue Collar precision |  |
| Total Spend | $313,887.35 | $270,440.70 | $164,892.59 | $388,647.44 | $148,914.38 | $217,629.60 | **$1,504,412.06** |

The challenge was to see which clusters matched the existing top six and at the same time offered a large enough group to expect a meaningful level of success. The clusters selected were chosen based on the Sales\_per\_visit variable. This is the average amount spent per visit.

Each of the demographic clusters shares a significant amount of the same characteristics. The client is in the same age range, owns the home, and has a college degree. The volume of contacts is roughly half the size of the top six clusters so there is room for growth both in the number of customers as well as the sales volume of this group.

**Figure 2**

*Chart of the next 6 clusters*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **The Next Six** | | | | | | | |
| Demographic | Cluster 5- Prosperous Metro Mix | Cluster 11- Family Ties | Cluster 12 - A good step forward | Cluster 18- Home Sweet Home | Cluster 23 - Settled In | Cluster 38 - Rustic Homesteaders | **Totals** |
| Observation Rank and Frequency | 5 (1219) | 7 (1301) | 12 ( 893) | 8 (1224) | 10 (1158) | 11 (1155) | **6950** |
| Household Income | High | Middle | High | Middle | Lower Middle | Lower |  |
| Age | 30-44 | 35-49 | 22-34 | 25-34 | 60-69 | 55-64 |  |
| Homeowner | Yes | Yes | Renter | Yes | Yes | Yes |  |
| Children present | Yes | Yes | No | Yes | No | Yes |  |
| Education ranking | Post Graduate | Associate | Post Graduate | High School | College | High School |  |
| Job Level | Admin Support | Technical support | Executive | Precision Blue Collar | Retired | Blue Collar |  |
| Total Spend | $133,878.42 | $139,610.31 | $102,326.63 | $137,392.42 | $133,015.16 | $ 131,559.53 | **$777,782.47** |

Given this information some interesting questions can be asked. First, what is the best strategy to reach these customers? I can see a strategy where the clusters that match up, such as cluster 5 and cluster 4, receive the same promotional offers. This would be verified via A/B testing and linear regression analysis.

The second question that this group of clusters can answer is the best strategy for growing the business in cluster groups. Marketing to smaller clusters with similar demographics is a way to grow the business.

In the pie chart, we can see a large chunk called other. By sorting out the clusters that are similar in that group, we may find growth areas that were ignored because the size was not seen as viable.

**Figure 3**

*Pie chart of clusters, based on observations*.

The two groups are statistically similar. The charts show how similar the two groups are, and how the correlations in the data are nearly the same. The correlations are important for predicting changes in the data due in part to the lack of a time frame. Predictive analytics is more powerful when the data can be viewed over a time frame.

**Figure 4**

*Descriptive Statistics and Correlations for the Top Six Clusters*

| **Simple Statistics** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **N** | **Mean** | **Std Dev** | **Sum** | **Minimum** | **Maximum** | **Label** |
| **CC\_CARD** | 13138 | 0.38537 | 0.48670 | 5063 | 0 | 1.00000 | CC\_CARD |
| **Number\_of\_Visits** | 13138 | 5.14614 | 6.63068 | 67610 | 1.00000 | 115.00000 | Number of Visits |
| **Sales\_per\_Visit** | 13138 | 114.50845 | 86.63438 | 1504412 | 2.50000 | 1726 | Sales per Visit |
| **Tot\_Spend** | 13138 | 487.26451 | 687.85948 | 6401681 | 4.00000 | 24140 | Tot\_Spend |
| **GMPCNT** | 13138 | 0.52525 | 0.15707 | 6901 | -3.36000 | 0.83000 | GMPCNT |

| **Pearson Correlation Coefficients, N = 13138 Prob > |r| under H0: Rho=0** | | | | | |
| --- | --- | --- | --- | --- | --- |
|  | **CC\_CARD** | **Number\_of\_Visits** | **Sales\_per\_Visit** | **Tot\_Spend** | **GMPCNT** |
| **CC\_CARD CC\_CARD** | 1.00000 | 0.34787 <.0001 | 0.07344 <.0001 | 0.34084 <.0001 | -0.04059 <.0001 |
| **Number\_of\_Visits Number of Visits** | 0.34787 <.0001 | 1.00000 | -0.17759 <.0001 | 0.70826 <.0001 | -0.11968 <.0001 |
| **Sales\_per\_Visit Sales per Visit** | 0.07344 <.0001 | -0.17759 <.0001 | 1.00000 | 0.23033 <.0001 | 0.27556 <.0001 |
| **Tot\_Spend Tot\_Spend** | 0.34084 <.0001 | 0.70826 <.0001 | 0.23033 <.0001 | 1.00000 | 0.03138 0.0003 |
| **GMPCNT GMPCNT** | -0.04059 <.0001 | -0.11968 <.0001 | 0.27556 <.0001 | 0.03138 0.0003 | 1.00000 |

In our top six clusters the average sale is $114.50. This is expressed by the variable Sales\_Per\_Visit. The $114.50 figure has a standard deviation of $86.63. The Tot\_spend variable represents the total of the all the sales by this customer. This is an important variable in that it shows that for this data set, customer lifetime value, or CLV. “Knowing the CLV of individual customers enables the decision maker to improve the customer segmentation and marketing resource allocation efforts and this in turn will lead to higher retention rates and profits for the firm” (Kareh et al. 2014 pg 591).

In reviewing the correlations, there are some clear connections that would be expected, and some that are worth reviewing. The correlation for credit cards and increased sales (Tot\_Spend) as well as credit cards and Number\_of\_Visits is strong in both the top 6 and the next six clusters.

**Figure 5**

*Descriptive statistics and Correlations for the Next Six Clusters.*

| **Simple Statistics** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **N** | **Mean** | **Std Dev** | **Sum** | **Minimum** | **Maximum** | **Label** |
| **CC\_CARD** | 6950 | 0.39698 | 0.48931 | 2759 | 0 | 1.00000 | CC\_CARD |
| **Number\_of\_Visits** | 6950 | 5.03424 | 6.14880 | 34988 | 1.00000 | 81.00000 | Number of Visits |
| **Sales\_per\_Visit** | 6950 | 111.91115 | 88.91329 | 777782 | 1.99000 | 1920 | Sales per Visit |
| **Tot\_Spend** | 6950 | 473.37151 | 689.59682 | 3289932 | 1.99000 | 15784 | Tot\_Spend |
| **GMPCNT** | 6950 | 0.50996 | 0.19522 | 3544 | -6.46000 | 0.99000 | GMPCNT |

| **Pearson Correlation Coefficients, N = 6950 Prob > |r| under H0: Rho=0** | | | | | |
| --- | --- | --- | --- | --- | --- |
|  | **CC\_CARD** | **Number\_of\_Visits** | **Sales\_per\_Visit** | **Tot\_Spend** | **GMPCNT** |
| **CC\_CARD CC\_CARD** | 1.00000 | 0.36507 <.0001 | 0.05962 <.0001 | 0.33199 <.0001 | -0.04710 <.0001 |
| **Number\_of\_Visits Number of Visits** | 0.36507 <.0001 | 1.00000 | -0.16467 <.0001 | 0.70244 <.0001 | -0.09340 <.0001 |
| **Sales\_per\_Visit Sales per Visit** | 0.05962 <.0001 | -0.16467 <.0001 | 1.00000 | 0.25535 <.0001 | 0.25095 <.0001 |
| **Tot\_Spend Tot\_Spend** | 0.33199 <.0001 | 0.70244 <.0001 | 0.25535 <.0001 | 1.00000 | 0.05115 <.0001 |
| **GMPCNT GMPCNT** | -0.04710 <.0001 | -0.09340 <.0001 | 0.25095 <.0001 | 0.05115 <.0001 | 1.00000 |

The average sale is $111.91 with a standard deviation of $88.91. The CLV of the customers in this group is nearly the same as the top six clusters, while being half the sample size. This reflects the similarity on the two groups.

The data is pointing towards a real correlation between credit card use and increased sales.

In terms of total frequency, the percentage of customers using a credit card is roughly half of the non-credit card group.

**Figure 6**

*Breakdown of Payment type. 0= non credit card 1= Credit card*



When the descriptive statistics are included, the picture changes dramatically. In terms of average sale, CLV, and number of visits all go up substantially. When the data set for payments was split into the two component parts, descriptive statistics were run as well as correlations. There is a positive relationship between Sales\_per\_Visit as well as gross margin percentage (GMPCNT).

**Figure 7**

*Descriptive and Correlation Statistics when Credit Cards are used for payment*

| **Simple Statistics** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **N** | **Mean** | **Std Dev** | **Sum** | **Minimum** | **Maximum** | **Label** |
| **Number\_of\_Visits** | 11031 | 7.94987 | 8.38154 | 87695 | 1.00000 | 115.00000 | Number of Visits |
| **Sales\_per\_Visit** | 11031 | 120.37546 | 91.53042 | 1327862 | 3.31000 | 1565 | Sales per Visit |
| **Tot\_Spend** | 11031 | 763.07399 | 893.52942 | 8417469 | 5.00000 | 24140 | Tot\_Spend |
| **GMPCNT** | 11031 | 0.50768 | 0.14695 | 5600 | -1.99000 | 0.99000 | GMPCNT |

| **Pearson Correlation Coefficients, N = 11031 Prob > |r| under H0: Rho=0** | | | | |
| --- | --- | --- | --- | --- |
|  | **Number\_of\_Visits** | **Sales\_per\_Visit** | **Tot\_Spend** | **GMPCNT** |
| **Number\_of\_Visits Number of Visits** | 1.00000 | -0.25276 <.0001 | 0.66513 <.0001 | -0.14948 <.0001 |
| **Sales\_per\_Visit Sales per Visit** | -0.25276 <.0001 | 1.00000 | 0.21248 <.0001 | 0.29587 <.0001 |
| **Tot\_Spend Tot\_Spend** | 0.66513 <.0001 | 0.21248 <.0001 | 1.00000 | 0.05236 <.0001 |
| **GMPCNT GMPCNT** | -0.14948 <.0001 | 0.29587 <.0001 | 0.05236 <.0001 | 1.00000 |

The non-credit card payments are lower in two key descriptive statistics: sales per visit and total sales. This is important to understanding the picture of the client drawn earlier in the paper. They use credit cards for purchases.

**Figure 8**

*Descriptive statistics and correlations for non-credit card payments.*

| **Simple Statistics** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **N** | **Mean** | **Std Dev** | **Sum** | **Minimum** | **Maximum** | **Label** |
| **Number\_of\_Visits** | 17768 | 3.23188 | 3.63326 | 57424 | 1.00000 | 85.00000 | Number of Visits |
| **Sales\_per\_Visit** | 17768 | 109.37993 | 83.75801 | 1943463 | 0.50000 | 1920 | Sales per Visit |
| **Tot\_Spend** | 17768 | 293.25622 | 352.71343 | 5210577 | 0.99000 | 11230 | Tot\_Spend |
| **GMPCNT** | 17768 | 0.52431 | 0.18595 | 9316 | -6.46000 | 0.89000 | GMPCNT |

**Figure 8 cont.**

| **Pearson Correlation Coefficients, N = 17768 Prob > |r| under H0: Rho=0** | | | | |
| --- | --- | --- | --- | --- |
|  | **Number\_of\_Visits** | **Sales\_per\_Visit** | **Tot\_Spend** | **GMPCNT** |
| **Number\_of\_Visits Number of Visits** | 1.00000 | -0.19798 <.0001 | 0.69251 <.0001 | -0.08172 <.0001 |
| **Sales\_per\_Visit Sales per Visit** | -0.19798 <.0001 | 1.00000 | 0.30263 <.0001 | 0.26543 <.0001 |
| **Tot\_Spend Tot\_Spend** | 0.69251 <.0001 | 0.30263 <.0001 | 1.00000 | 0.09432 <.0001 |
| **GMPCNT GMPCNT** | -0.08172 <.0001 | 0.26543 <.0001 | 0.09432 <.0001 | 1.00000 |

The hypothesis that should be tested is simple. Does the data support the idea that credit cards increase consumer spending?

First question: Do credit cards increase consumer spending?

H0  : The use of credit cards increases the amount spent on purchases

Ha  : The use of credit cards has no impact on consumer spending.

The second hypothesis to test would be the use of them by the highest CLV clients. Do the clients who spend, and visit the most use credit cards to pay for the purchases they are making?

Ho  : Higher value clients use credit cards on average more frequently for purchases

Ha : Higher value clients do not use credit cards more frequently for purchases

The reason for the hypothesis is in large part due to the goal of the project. The increase of sales and profits across the organization. Two items will be required to get the success the company is seeking. First, grow the business in similar clusters as the top six group. The data supports growth in this sector and it will be easier to work with existing customers as opposed to finding new ones.

The second action item is to discover a way to co-brand a credit card with a loyalty program. Rewarding customers for purchases is a great way to build CLV. Having the data from a branded credit card would allow for insights into where else the customer shops for similar products. “Co-branded credit cards remain integral to issuers and retail partners alike: the report estimates that, in 2018, co-branded credit cards generated $990 billion in purchase value, up an average of 7.9% from 2016” (Research and Markets, 2019 ¶ 2).

The ability to see how customers are spending and where they are spending is priceless information. It would open up new markets potentially while at the same time allowing the organization to grows its business.

As the data is currently, there are some challenges to getting truly predictive data out of it. Correlations are good; however, linear regression would allow for accurate predictions for sales, costs, and other variables in the business. For that, the data needs a time component.

It would be good to have any data on promotional efforts as well. What worked, and what did not is important. Certain clusters in the data respond well to coupons and special offers while others do not.

The short-term plan of action would be to go after the next six clusters to increase sales, and begin to work on a co-branded credit card. This way, the company can increase sales and GMPCNT, but increase the understanding of the existing customer base through data analysis.

References

Kahreh, M. S., Tive, M., Babania, A., & Hesan, M. (2014). Analyzing the Applications of Customer Lifetime Value (CLV) based on Benefit Segmentation for the Banking Sector. *Procedia - Social and Behavioral Sciences*, *109*, 590–594. <https://doi.org/10.1016/j.sbspro.2013.12.511>

Marr, C. (2019). *Why You Should Be Focusing on Customer Retention, Not Acquisition | Convince and Convert: Social Media Consulting and Content Marketing Consulting*. Convince and Convert: Social Media Consulting and Content Marketing Consulting. <https://www.convinceandconvert.com/online-customer-experience/customer-retention-not-acquisition/>

Research and Markets. (2019, September 10). *Global Co-Branded and Affinity Cards Markets, 2016-2018 & 2019-2023*. Www.Prnewswire.com. <https://www.prnewswire.com/news-releases/global-co-branded-and-affinity-cards-markets-2016-2018--2019-2023-300915409.html>

‌