mosaic
DATA SCIENCE
http://mosaicdatascience.com/
info@mosaicdatascience.com

Data Science Case Study

Customer Lifetime Value & Segmentation

Developing a custom, forward-looking customer lifetime value metric is a great way businesses can improve customer relations.

Industry

Health Insurance

Use Case

Customer Lifetime Value, Segmentation

Techniques

Bayesian Statistics, Classification, Unsupervised Learning

Outcome

Ranked list of high and low value customers.

UNDERSTANDING CUSTOMER LIFETIME VALUE

One of the largest healthcare insurance companies in North America wanted to optimize their marketing efforts specifically around retention. This organization serves over 28 million members, with one in eight people in the United States relying on them for eye-care health coverage. The company had made an investment in a consumer insights team. This team believed if they could start segmenting their members into different groups, they could begin to improve their marketing efforts, ultimately identifying high revenue generating and segments at risk for leaving. They identified utilizing machine learning to develop a customer lifetime value measure.

The company has one of the largest healthcare databases in the world consisting of vision care claims data, lens and frame wholesale data, patient data, and lab data.

The insights team had already started digging into the problem, but they required a seasoned data scientist to come in and help guide their customer segmentation approach. The team wanted to segment on age, tenure, company size, and geographic location.

Mosaic was able to provide a data scientist with over 20 years' experience to not only guide the team in innovative approaches, but also get their hands dirty in the analysis.

SETTING UP A CUSTOMER LIFETIME VALUE ANALYSIS

In looking at the membership, Mosaic needed to compute a customer lifetime value metric[1]. In the initial phase of the study the Mosaic data scientist used data from two preexisting sources. Both sources contain consumer data for a three-year-period. The data set included both the consumer membership earliest start and latest termination dates and the consumer's membership tenure. Mosaic computed a tenure metric by subtracting the membership earliest start and latest termination dates. This represents a maximum tenure, as gaps in consumer membership (e.g. due to changing employment) are ignored. It is identified as tenure1 throughout the rest of the analyses.

The data set is evenly split between males and females (50.42% males), three-in-five consumers have an active membership (60.4%) and the average and median ages are 45.3 and 44 years. 14,479 members were selected for the analysis.

The consumer insights team was operating under the assumption that younger eyes tend to be healthier eyes. Mosaic's data scientist believed the data would support this assumption. They analyzed age, tenure, and the out-of-pocket (OOP) revenue collected from members in a three-year time horizon.

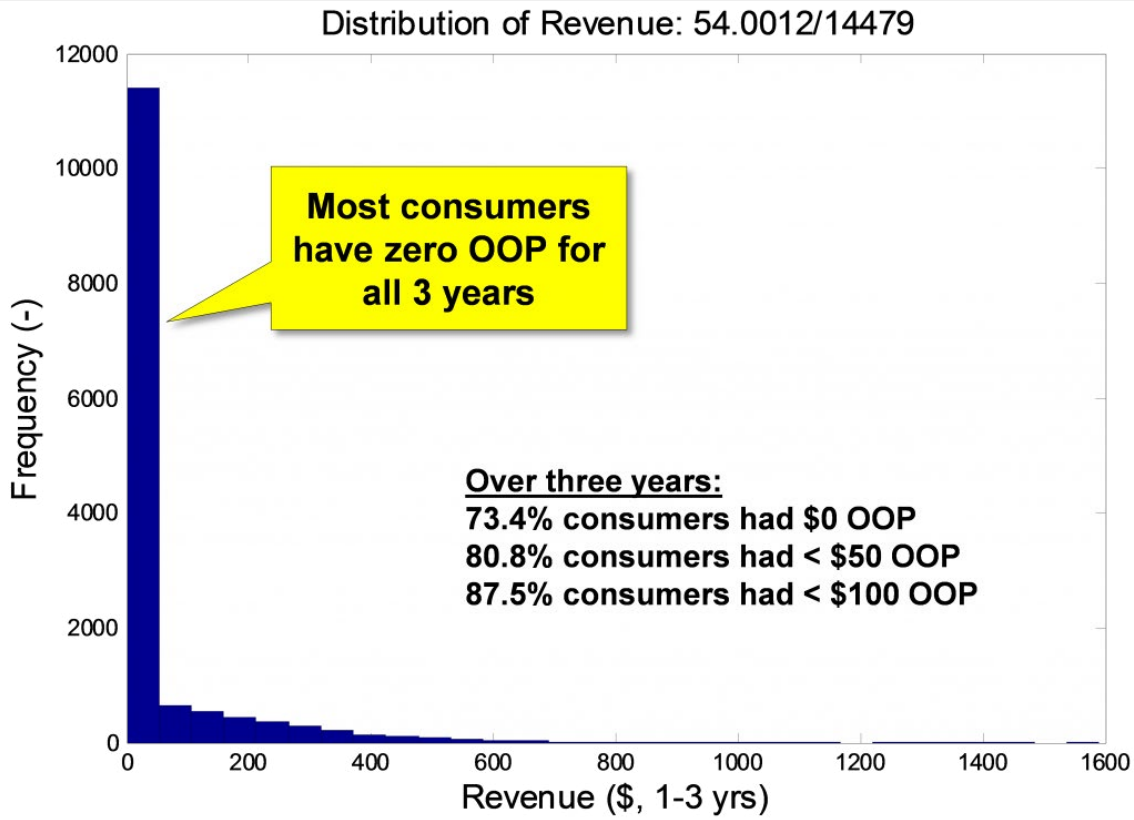


Figure 1. Distribution of consumer OOP revenue summed over three year period.

As figure 1 shows, the majority of consumers have zero OOP revenue over the three year period. The average OOP revenue is $54.

Next we categorized the OOP by consumer age and the tenure metric each into 12 categories. The age categories range from 20 to 75 years of age, in 5 year increments.

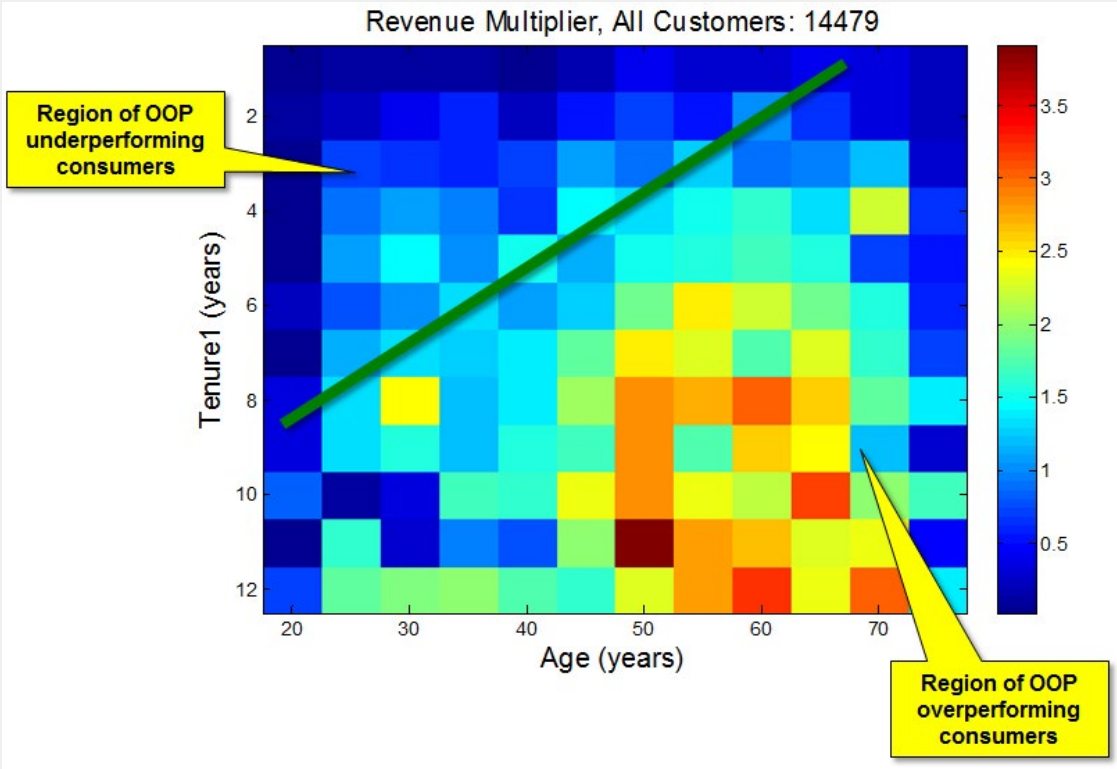


Figure 2. OOP relative revenue. Revenue by age-tenure categories, summed over three year period and divided by the overall mean value across all ages and tenures.

Figure 2 shows a significant trend in the consumer OOP relative revenue data, increasing with both age and tenure. The green line illustrates the frontier of this relative revenue gradient. So, the older someone is and the longer they have been a member, the more revenue they generate, therefore a higher customer lifetime value score. This is a good target segment for the marketing team.

Figure 3 shows the same data using a three-dimensional plot for perspective.

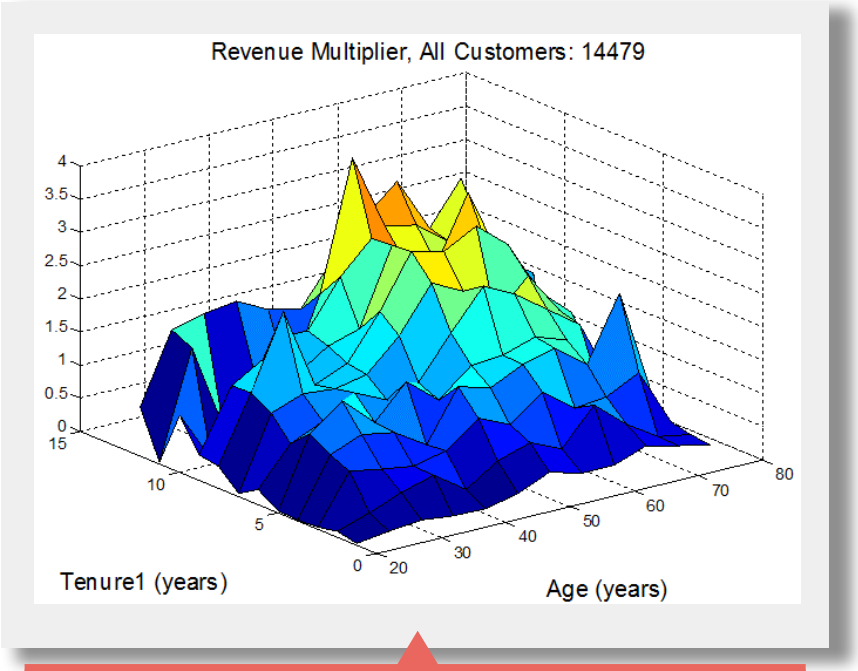


Figure 3. OOP relative revenue. Revenue by age-tenure categories, summed over three year period and divided by the overall mean value across all ages and tenures.

The trend of increasing OOP with age and tenure is coincident with a decreasing trend in membership population. The majority of members have 1-2 years tenure and are of younger age. While the OOP revenue increases with a combination of age and tenure, the membership populations creases in this direction.

The OOP revenue under-performing groups tended to be younger age-shorter tenure combination, which matches up with the assumption that young eyes are healthy eyes.

The majority of consumers in the 22.5-37.5 age region had higher revenue when tenure is longer, which suggests that age is not the only factor. Longer-tenure consumers in this age region may have more awareness. Potential targets for the organization's marketing department would be consumers in high performing age-tenure regions with zero OOP revenue.

One interesting find was around client size. Client size was measured by the number of people in each company using the vision insurance. The OOP revenue was fairly strong related to the client size. The large clients were under-performers.

There was an exception, one of the large clients had much better OOP revenue than the other large clients. This client actually turned out to be a collection of small companies, a nice verification of the finding. The marketing team could now target based on client size as well.

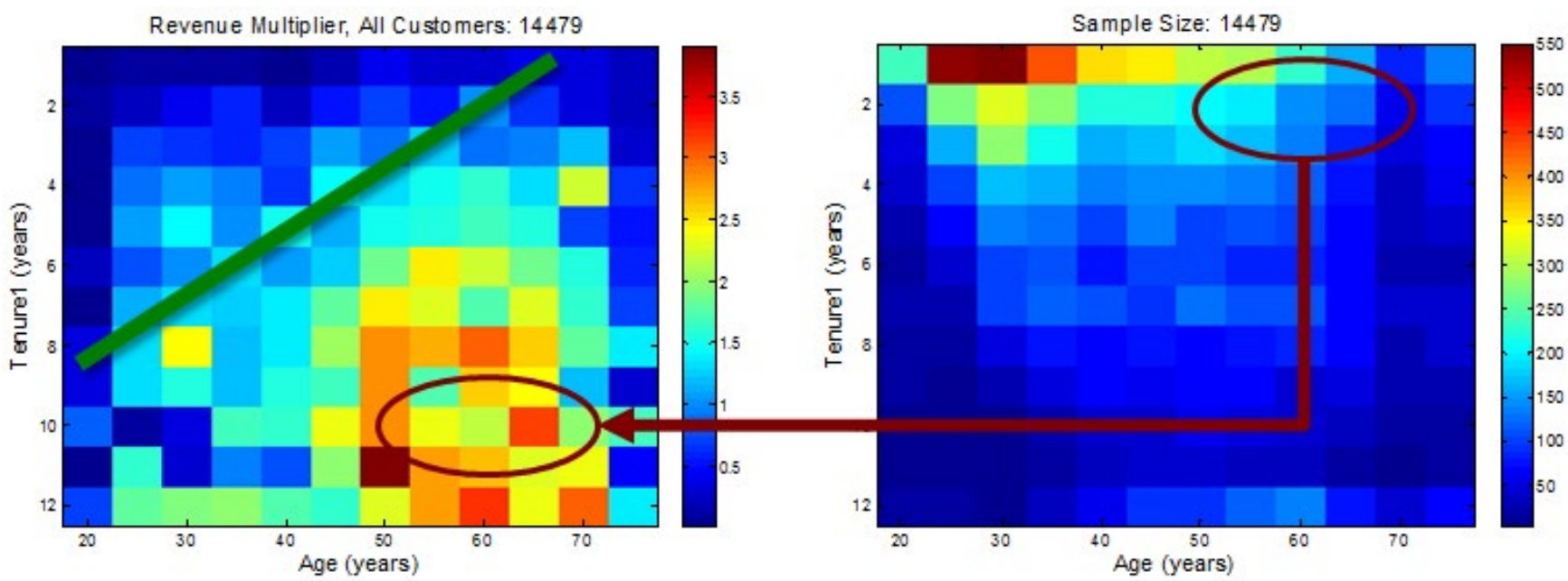
When age and tenure are controlled, there remains a large difference in OOP revenues between the large and small client consumers. Clearly there are other factors influencing the large-client member's OOP revenue aside from merely the age and tenure demographic. When the age and tenure biases are removed, the large-client consumers continue to have significantly lower revenue.

To summarize our OOP revenue findings, the underperforming OOP groups are the younger age-shorter tenure members and the members in the large-clients. Likewise, the overperforming OOP groups are the older age-longer tenure members and the members in the small-clients. Guardian Life is the apparent exception. It is a large client with small client OOP behavior. This could be because Guardian Life consists of smaller groups.

These results suggest possible OOP marketing themes. For instance, the majority of the consumers are in the 22.5-37.5 age bracket. It may seem that their OOP underperformance is due to their age. As Figure 4 shows, however, this age bracket has improved performance when tenure is longer. So the OOP underperformance of the vast majority of the members is not due solely to age. Years of age is not the only factor, and perhaps the longer-tenure consumers in this age bracket simply have more awareness. This suggests there may be a large demographic that is untapped.

A similar reasoning may apply to the older consumers of shorter tenure. As Figure 5 shows, this age bracket has high OOP revenue when the tenure is longer. So again, this raises the question: What is the difference between the shorter and longer tenure members? It may be that the shorter-tenure members lack the plan awareness that longer-tenure members have.

Figure 4. Older age, low tenure consumers have low OOP revenue. Older age and longer tenure consumers have higher OOP revenue.



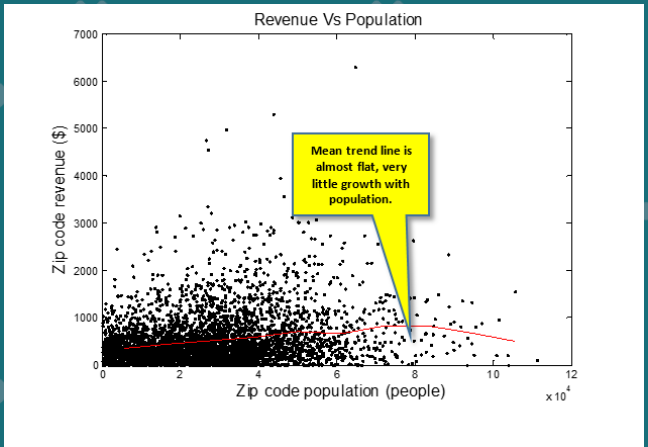
Even the high-performing older age-longer tenure cells have many zero OOP members. This raises the question: Why are some members making significant OOP expenditures while many others in the same age-tenure demographic are not? Is this significant difference merely due to differences in need, or can it be due to a lack of convenience or awareness?

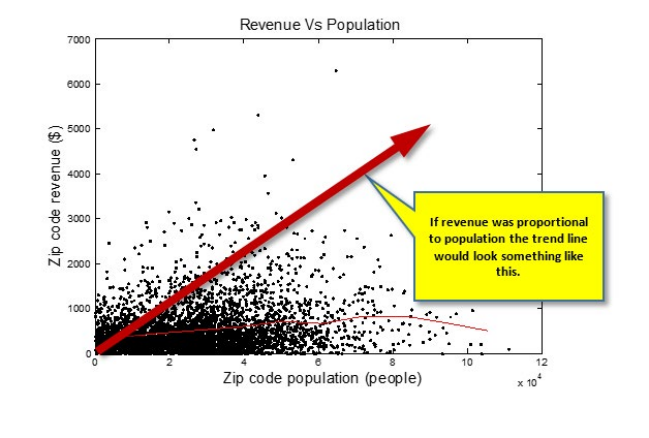
As noted before the client also has a successful frame and lens wholesale business. This is a big revenue generator, and the trends were very similar to that of the OOP trends. The frame and lens revenue results also show similar trends with client size, as were found for the OOP revenues.

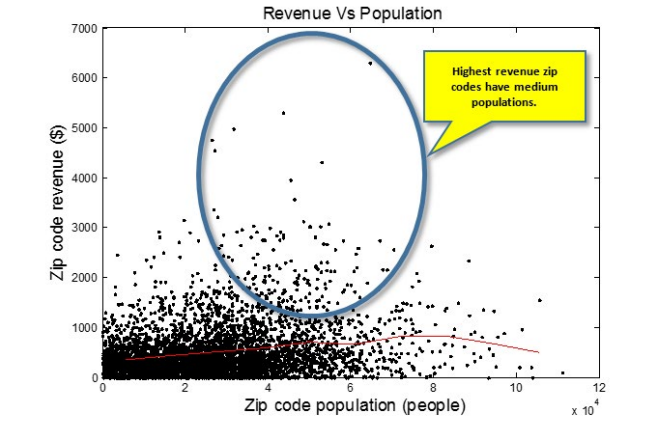
Age and tenure were big factors in the lifetime value of a member. Would geographic location prove to be the same? Mosaic analyzed to find out.

Mosaic's data scientist created a first-order revenue metric by summing lens, wholesale frame and out of pocket (OOP) purchases over all four years to obtain a single metric for each member. In the data there were zip codes associated with these purchase. The team had access to a zip code geographic data as well, location (a latitude / longitude point) and population from 2010 census. Both data sets were merged to analyze the geographic trends.

Revenue does not increase very significantly with population. Perhaps higher population zip codes have greater proportion of young people. Perhaps the insurance's membership is not highly correlated with population.







RESULTS

Mosaic was able to effectively use predictive data analysis to build out high target and high risk customer segments for the consumer insights team. Now that the marketing team understands their customers better, they can increase revenue with improved marketing campaigns. Focusing in on age, tenure, and small company size instead of geographic location will only help the vision insurance company improve their customer experience. The company can now optimize marketing response models to customers falling in the 'at risk for leaving' category.

Endnotes

1. https://www.crazyegg.com/blog/customer-lifetime-value/