Analytix

Custom Predictive Modeling

ANALYTICS RESULTS FOR MEDICAL SUPPLEMENT COMPANY

Linx Think Ahead.™

Summary

The results presented herein reflect our current knowledge of marketplace demographics and sales, and provides a way to predict sales in a given market and define what marketing investment should be made to maximize ROI. Sales data from 08/2016 to 07/2017 are combined with 2015 data from census.gov to learn meaningful insights. The data is transformed into a new dataset that is "market-area" centric (DMA or designated market area). Specifically, for each market area we obtain:

1. Number of boxes sold.
2. Number of recommending physicians. For this product, physicians would recommend this product to their patients.
3. Total population size measured as:
    i. **the number women aged 25 years or older**
4. Average income per household.
5. Average age.

Here are the summary points for this work:

1. We develop a predictive model that takes as input the number of recommending physicians, target population size, average age, and average household income and returns the number of boxes sold per year per DMA.

2. At constant population size, boxes sold increases linearly as the number of recommending physicians increases.

3. The value of each recommending physician (where value is defined as the number of boxes sold per recommending physician) is higher in more populated DMAs, and is proportional to a DMA's target population size.

4. DMAs with average ages over 50 years old sold approximately 3x as many boxes for every 1,000 people. So, for each physician added in this DMA we would add another 3X return in this market over a younger DMA.

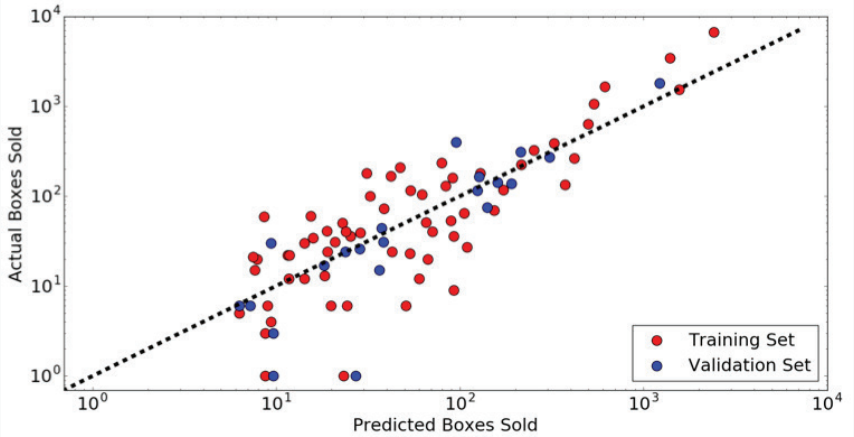
5. DMAs with average household incomes above 60,000 and under 90,000 sold between 2x and 4x as many boxes for every 1,000 households.

6. We identify a list of DMAs which lie in the optimal age, income, and population size ranges as prime candidates for more marketing efforts.

7. The biggest predictors of boxes sold was target population size and the number of recommending physicians, but there are some DMAs were older age and higher incomes compensate for having a smaller population size and fewer numbers of recommending physicians.

Key Insights

1 We developed a predictive model summarized by equation (1) below, which well-approximates the number of boxes sold in a given DMA. First, the data was split into 2 sets, one used to train the model (red dots) and the other used to test the model's predictions (blue dots). In the accompanying figure, we see that both the red and blue dots closely cluster around the black dotted line. Since the black dotted line indicates perfect agreement, we see that our model is near-perfect. In equation (1) we can see that the number of boxes sold per year, B, is proportional to the numberof recommending physicians R, the average income I, the target-population size P (number of women aged 25 years or older), and the average age A in each DMA. The learning-parameters are the α_k values, and each feature is raised to one these learning-parameters to determine the overall strength of the feature. The exact values are shown in the Table below.



$$B = a_0 R^{a_1} I^{a_2} P^{a_3} A^{a_4}$$

Parameter	a_0	a_1	a_2	a_3	a_4
Value	$8.29 \cdot 10^{-6}$	1.03	0.94	0.18	0.52

Using this model, we can accurately predict yearly sales in new markets and how sales will change when aspects of the DMA change. Since α_1≈1, the model predicts that at constant target population size, doubling the number of recommending physicians will double the number of boxes sold. As an example, consider the New York market area in which 3447 boxes were sold in total and the number of recommending physicians is 53. If we double the number of recommenders to 106 while keeping the target population size constant, the equation predicts that the number of boxes sold would double to 6894 for the year. This equation can also be used to shift marketing efforts across DMAs such that we focus on recruiting more physicians in areas predicted to have higher sales, and focus less on DMAs which are predicted to produce fewer recommending physicians.

2 Equation 1 also gives us a way to shift marketing spend and efforts such that we focus on recruiting more physicians in locations where they are more valuable. The value V of each DMA can be quantified by the ratio V=B/R which gives the average number of boxes sold per recommending physician in that DMA. DMAs with a higher V value are locations in which individual recommending physicians will have a higher payout. From equation 1, we see that the value of each DMA is approximately:

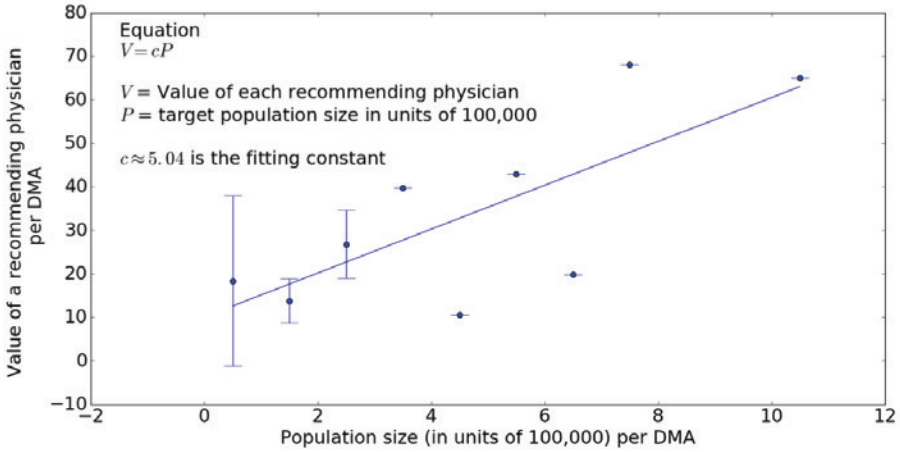
$$V = a_0 I^{a_2} P^{a_3} A^{a_4}$$

This shows that the value of each DMA is determined by the target-population size, the average household income, and the average household age. We can then calculate this quantity for all DMAs and prioritize our marketing efforts by recruiting more recommending physicians in areas with the highest V value. We recommend that this become a new KPI which is continuously tracked and used to focus market efforts.

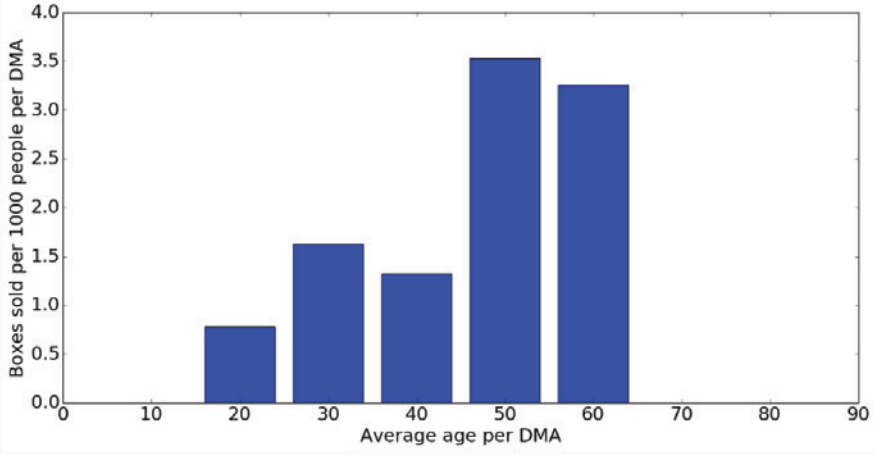
3 To see the marginal effect of population size on the value of each recommending physician, we analyzed the correlations between the value V and the size of the target-population per DMA. We fit the KPI value V=B/R against the target-population size P according to equation 3 in which the value V is modeled as being proportional to the target-population size P.

$$V=cP$$

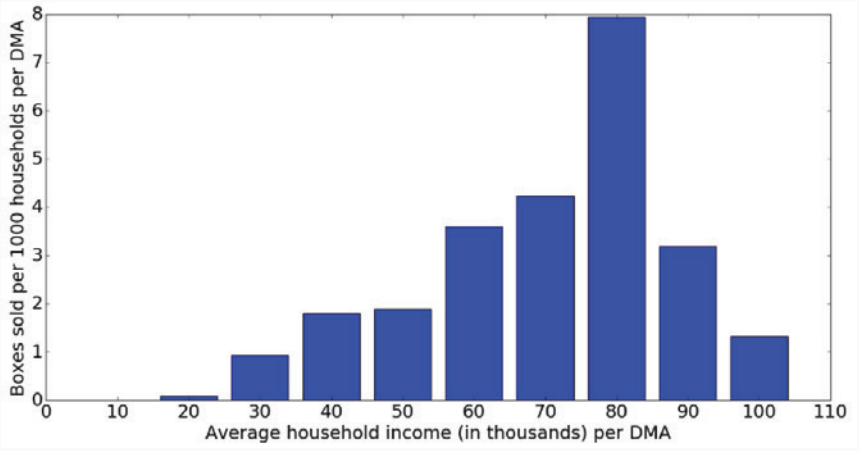
The figure below shows that the value of each recommending physician is not constant across all DMAs, but instead the value of each additional recommending physician recruited is directly proportional to the target population size. This is confirmed in the figure, in which the average and standard deviation (error bars) of boxes sold per DMA for every 100,000 women aged 25 years or older are shown. Equation 2 captures the trend suggesting that recruiting physicians in larger markets translates into higher payoffs.



4 To see the marginal effect of age on the number of boxes sold, we analyzed the correlations between boxes sold and average age per DMA separately. The number of boxes sold per 1000 households in a given DMA was found to be approximately 3 times as large for market areas with average ages 50 years or older. Since we normalize the boxes sold by the number of households, the results in this figure are not biased by the fact that DMAs have different population sizes and allows us to see the true effect of age. This suggests that we should preferentially target market areas with higher average ages, and this a reflection of the fact that older people have higher incidences of UTIs. This is confirmed by a study by Rowe et al in the Journal of Aging Health (Rowe, Theresa A., and Manisha Juthani-Mehta. "Urinary tract infection in older adults." Aging health 9.5 (2013): 519-528.). Interestingly, the study also indicates that institutionalized adults have even higher rates of UTIs due to the presence of comorbidities. Therefore, it would be worthwhile to determine which DMAs have higher concentrations of elderly-care facilities and market the product preferentially there. It would also be interesting to consider marketing the product to these elderly-care facilities directly, or to the physicians who interact with these facilities.



5 To see the marginal effect of household income on the number of boxes sold, we analyzed the correlations between boxes sold and average household income per DMA separately. Analysis of the average household income in a given DMA revealed that sales of our product per 1000 households peaks (is bell-shaped) around market areas with average household incomes in the range of over 60 to under 90 thousand dollars. Since we normalize the boxes sold by the number of households, the results in this figure are not biased by the fact that DMAs have different population sizes and allows us to see the true effect of income. Sales per 1000 households peaks significantly in DMAs with average household incomes in the range of 80 to 90 thousand dollars, in which sales were at most 4x and at least 2x that of all other market areas. This strongly suggests we should target market areas with higher incomes, but not too high since sales in market areas with average household incomes above 100 thousand were just as low as sales in market areas with incomes in the range of 30 to 40 thousand. The optimal market areas appear to be those in which average incomes are in the range of over 60 to under 90 thousand dollars.

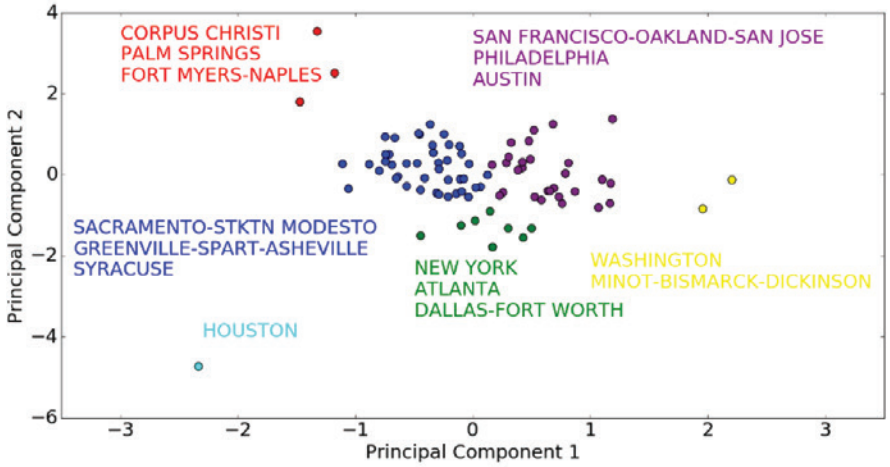


6 Our analysis thus far has focused on correlating market area features with the number of boxes sold. We found that i) market areas with larger population sizes correlated with more boxes sold ii) market areas with average household incomes between 60 and 90 thousand correlated with more boxes sold per household and iii) market areas with average ages above 40 years old correlated with more boxes sold per person. We can create two lists using this information. The first list (left column in following table) shows the DMAs that have average ages above 40 years old and average household incomes between 60 and 90 thousand. The second list (right column in following table) shows the DMAs that have incomes and ages in the optimal ranges as before, but which additionally have population sizes greater than the total average DMA population size (these are bigger markets). We get the same list of DMAs if we use total population size or target population size (number of women aged 25 years or older).

DMAs with: **Average household incomes between 60K and 90K** **Average age above 40 years old**	**DMAs with:** **Average household incomes between 60K and 90K** **Average age above 40 years old** **Population size above average DMA size**
SAINT LOUIS, MIAMI-FT LAUDERDALE, SANTA BARBARA-SAN MAR-SAN LUOB, CHARLESTON, TOLEDO, KANSAS CITY, WEST PALM BEACH-FT PIERCE FORT MYERS-NAPLES, CLEVELAND, PALM SPRINGS, BINGHAMTON, MONTEREY-SALINAS, MILWAUKEE SESATTLE-TACOMA, TUSCALOOSA	MIAMI-FT LAUDERDALE WEST PALM BEACH-FT PIERCE MILWAUKEE

These features (income, age, and population size) are all things we cannot control within each DMA, but we can control the number of recommending physicians. In particular, the DMAs in the right column are predicted to be prime candidates for recruiting additional recommending physicians. These market areas simultaneously had the highest sales per capita (per individual or per household) and also had above average (larger) population sizes. Thus, recruiting additional physicians in these areas would be predicted to result in the highest payoffs.

We implemented a machine learning technique called Principal Component Analysis (PCA) on the market areas in our data set. This technique clusters the market areas in terms of their greatest differences or their greatest variability (these are the principal components). In the accompanying figure, we show these clusters (each data point is an individual DMA) and for each cluster we show the top 1, 2, or 3 market areas measured by the most boxes sold. The 'average' of a cluster describes the average traits of the DMAs within that cluster. Further analysis can be done here, but one key insight from this clustering is the comparison between the market areas labeled red and blue. Both clusters had approximately the same number of recommending physicians, but the blue cluster had a target population size twice as large as the red cluster. Nevertheless, the red cluster sold 3x as many boxes, and further analysis revealed that this was the result of the red cluster having a average age about 20 years older than the blue cluster, and an average household income approximatley 10,000 dollars greater. This demonstrates that age and income can significantly affect boxes sold, and in some cases even overcompensate for a smaller target population size. This can be used in conjunction with the value KPI V to better assess which market areas have higher potential.



Future Directions

This analysis can always be expanded on. For example, we can add rates of urinary tract infections per location, the population sizes of peoples' races and ethnicities, the population sizes of peoples' professions (it could be that certain professions are predisposed to UTIs), etc. We can also try to factor in weather if it's the case that UTIs are more common in warmer/colder climates. Additionally, we can improve the predictive accuracy of the models, and extend them to make predictions on other important marketplace variables. Finally, we can factor in the dollar value of boxes sold and the cost of recruiting additional physicians to predict ROIs quantitatively.