Paper 028-2011

Predictive Analysis of Auto Insurance Purchasing Behavior on the Internet
Roosevelt C. Mosley, Jr., FCAS, MAAA, Pinnacle Actuarial Resources Inc.,
Bloomington, IL

ABSTRACT

Today, there is much information available about current and potential auto insurance customers. This information can be used to perform more detailed analyses to predict the purchasing behavior of certain types of customers and to determine what drives that behavior. This paper will discuss the analysis of two stages in the customer purchase decision.

1. Submitting a Quote – for those consumers that visit an insurance company website, what is the likelihood that they complete the quote process and submit the quote to the company
2. Purchasing a Policy – of those consumers that submit a quote, what is the likelihood that they ultimately purchase a policy from the company

The results of these types of analyses can be instrumental in helping to set marketing, pricing and customer service strategy. Potential insureds with a higher likelihood of purchasing can be identified, and marketing strategies can be put into place that optimize the marketing dollars spent. Prices can be established that optimize profits relative to customer conversion and retention. Also, customers with low conversion and retention likelihoods can be identified, and customer service strategies can be implemented to improve these likelihoods.

INTRODUCTION

In the last several years, auto insurance has become a very competitive business. Commercials flood televisions and radios about saving money, customer service, and coverage options. Insurers are sponsoring major sporting events. Companies are attempting to connect to potential customers through affinity groups and by greater use of direct mailing. And just like every other business, insurers are making greater use of the explosion of the internet to connect with potential customers. This includes sponsoring searches on search engines, email marketing, the use of social media, and the promotion of websites that can be used to purchase insurance.

The increase in the amount of competition in auto insurance and the increased focus on marketing has led to an increase in the amount of information that is collected regarding the insurance purchase decision. Insurance companies can benefit from a predictive analysis of this information that is being collected. This type of analysis will allow insurers to identify potential customers that are more likely to respond to marketing efforts and ultimately purchase a policy with the company. This can help lead to more efficient spending of marketing dollars and a more effective marketing program. Ultimately, when combined with a holistic company analysis, this can be a part of the equation that allows insurers to increase the value of their customers.

This paper presents the results of an analysis of internet auto insurance purchasing decisions. The analysis looks at two separate steps in the insurance purchasing decision. First, of those consumers that visited an insurance company website, which of those visitors were more likely to submit an auto insurance quote to the company. The second step involves customers that submitted a quote to the company. Of these consumers, which were more likely to actually purchase a policy from the company.

HISTORICAL CONTEXT

Insurance companies have historically been concerned about marketing and attracting customers, but there have not been many detailed statistical analyses around marketing and customer attraction. Insurance companies have always been involved in marketing, but it has been targeted mainly at building name recognition for direct companies. Independent agent companies focused their marketing activities on the independent agent and usually not the end customer. The judgment of most insurers was that customer decisions were driven mainly by price, and so if a company wanted to attract more customers, they would simply decrease the premiums. Insurers are beginning to find out, though, that while price is an important factor, price is not the only factor that drives a customer's purchase decision. There are a number of other considerations that go into purchase choices, including company reputation, customer service, and perceived value. In addition, different customers can place a different level of value on different things. For example, price may be more important for a younger customer, while company reputation may be more important for an older customer. These and many other considerations are now being taken into account in understanding the purchase decisions of a customer.

Customer Response Analyses

These realizations have led insurance companies to a series of studies called customer response analyses. Insurance companies undertake a series of efforts in marketing, quoting, selling, and servicing current and potential policyholders. A diagram of these steps is shown below.

Figure 1: Insurance Life Cycle

Insurers undertake many different marketing efforts. This marketing includes television, internet and radio advertisements, direct mail, and sponsorship of sporting events. As a result of these marketing efforts, insurers are hoping that potential customers will contact insurance companies to obtain more information and a quote for insurance coverage. Once the potential insured has received a quote for insurance, they have the option of either accepting the quote and purchasing coverage or rejecting the quote. If the customer purchases the coverage, they can at any point during the policy period choose to cancel their coverage or simply stop paying the premiums. If the customer continues the coverage through the entire policy period, at the end of the policy period the insurance company will present the insured with a renewal option. At this point, the insured will have the option of either continuing coverage with the company or discontinuing the relationship.

At each point in this process, the consumer has the opportunity to respond to the insurance company efforts. This response can either be negative or positive. When a customer sees an ad on television or a banner on the internet, they can choose to contact the company for a quote or not. When the quote is received, the consumer chooses whether or not to purchase the policy. Once the policy is purchased, the insured has the option to either continue coverage or cancel coverage. Ultimately, the insured has the option to respond to the company long term and renew the relationship.

Each of these opportunities to respond can be translated into a customer response analysis.

1. Quoting analysis: analysis of the likelihood of a prospective insured obtaining an insurance quote from a company
2. Conversion analysis: analysis of the likelihood of a prospective insured that has received a quote purchasing insurance from a company
3. Lapse/Cancelation analysis: likelihood of an insured not lapsing in making their premium payments or canceling the policy mid-term
4. Retention analysis: analysis of the likelihood of a current insured renewing with a company

While these analyses are gaining more popularity in the insurance industry, they are not without their challenges.

- **Process Challenges**
  First, there are challenges related to the modeling process. Insurance companies have become adept at incorporating Generalized Linear Models in the development of pricing analyses. Customer response analyses are a different type of analysis, and require the application of different model structures. Lack of familiarity with different model types, and the lack of software to implement these analyses can present a challenge.

- **Data Challenges**
  Another technical challenge is the availability and applicability of data. For the marketing and quoting analysis, information may either not be available, or it may not be retained for a long period of time for analysis. For retention analyses, if an existing policyholder does not accept the renewal quote, the data related to the renewal offer may not be retained.

- **Price Data**
  Price data is a challenge to obtain. For new business, the new price relative to the old price will be a significant factor in making a purchase decision; however, most insurers do not capture the new insured's previous premium, and if they do, it is often unreliable. For existing policyholders, a factor in determining whether they renew with you will be how competitive your premiums are relative to your competitor's premiums, but obtaining competitor quotes is more difficult in today's world. For policyholders that leave at renewal, oftentimes the renewal premium is not retained if the insured does not accept the renewal offer.

- **Market Coverage Challenges**
  Lastly, the new business and renewal information that an insurance company has is only related to their specific niche, and may not be representative of the overall market. This can particularly be a challenge for independent agency companies, since often agents may only quote certain types of business with certain companies.

These and a number of other considerations make customer responses analyses a challenge. To address some of these challenges, we have undertaken an analysis of internet quote and purchase behavior based on the insurance quotes for five insurance company websites. These analyses are described in more detail below.

Population	Target	Description
Auto Insurance Website Visitor	Quote Submitted	Of those that visit the website, people with what characteristics are more likely to submit a quote?
Quote Submitted	Policy Bind	Of those that actually submit the quote, how many complete the purchase?

Figure 2: Customer Response Analyses Performed

DATA

This analysis is based on a subset of data from comScore, Inc., which is a digital market research firm. comScore leverages a panel of 1 million internet users in the US that have given explicit permission to allow comScore to passively observe their online behavior and see everything they do online. In the insurance industry, comScore is able to identify consumers that visit an insurance company site and see what they do, including online quoting and binding transactions. There are several categories of information that are collected related to insurance company websites. For the purposes of this analysis, data points related to website visitors and auto insurance quotes were used.

- Visitors: for consumers that visit insurance company sites, information that is collected includes the date and time of the visit, what sites the customer viewed prior to coming to the insurance company site, how the customer got to the site, if the customer used a search phrase, what search phrase the customer used to get to the site, how long they spent at the site, and what they viewed while they were there.
- Quote: for consumers that actually complete the process to obtain and quote for auto insurance, information is collected on the details of the policy (limits, deductibles, etc.), information about the customer, details of the vehicles being insured, details on the drivers of those vehicles, and accident and violation history.

Also in the data files are indicators as to whether the quote was ultimately submitted, and also whether a policy was purchased. The time period of the data is 2008 and 2009. The analysis was based on 1.4 million visits to insurance company websites.

Based on a review of the data summarizations, there are some interesting findings that emerge regarding the characteristics of consumers that are shopping for insurance on the internet. These results are simply univariate summaries, yet they provide interesting initial insights into the data. Multivariate model results that take all variables into account at once will be shown later.
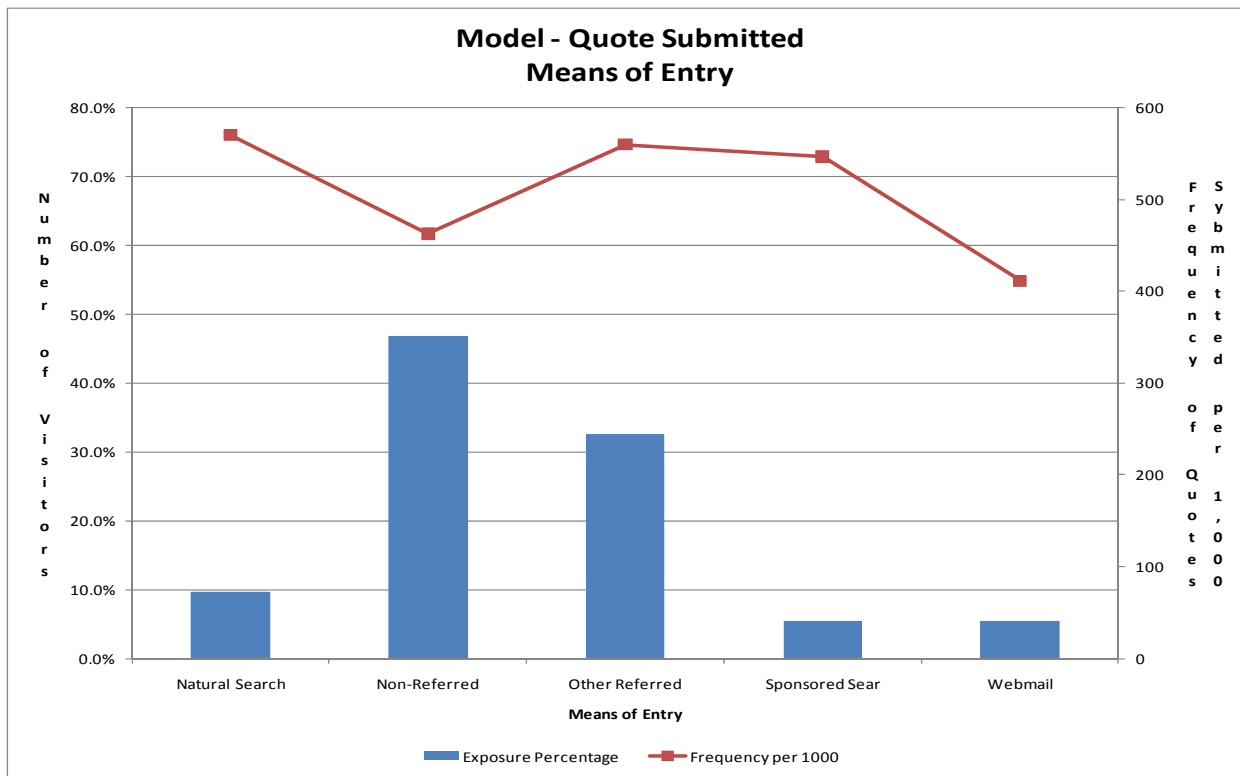
Figure 3: Shopper Characteristics - Means of Entry

The figure above shows the means of entry of a website visitor. The blue horizontal bars represent the distribution of exposure in each category, and the red line represents the frequency of a purchase being made per 1,000 visitors to the website. As can be seen, about 50% of the visitors are not referred to the site. About 10% of the visitors come from natural searches (the consumer not clicking on a paid search but on natural search engine results), and about 5% each from sponsored searches and webmail. 30% of the visitors came from other means of referral. Regarding the frequency of submitting a quote, natural search has the highest frequency, and webmail has the lowest frequency. If a consumer performs a natural search, these customers have the highest frequency of purchasing a policy, while those responding to webmail will have the lowest frequency.
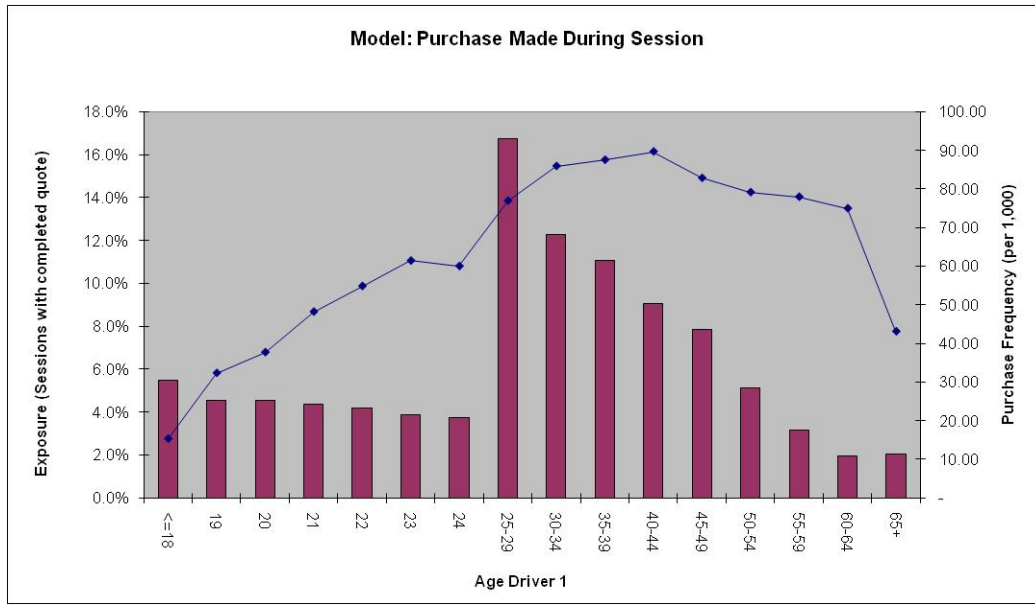
Model: Purchase Made During Session

Figure 4: Shopper Characteristics - Age

Figure 4 above shows the distribution of internet auto insurance shoppers by age. In the figure above, the purple horizontal bars represent the percentage of those customers receiving a quote in that particular age category. The blue line represents the purchase frequency of each category. As can be seen, about 30% of those customers that are receiving quotes on the internet are under 25 years of age. In addition, it can be seen that as the age increases, the proportion of the customers receiving quotes on the internet decreases. However, the purchase frequency tells a different story. The purchase frequency actually increases with age until age 40 – 44, when it begins decreasing again. For ages above 44, the decrease in purchase frequency is small until the 65 and older age group, where the purchase frequency decreases by about 40%.
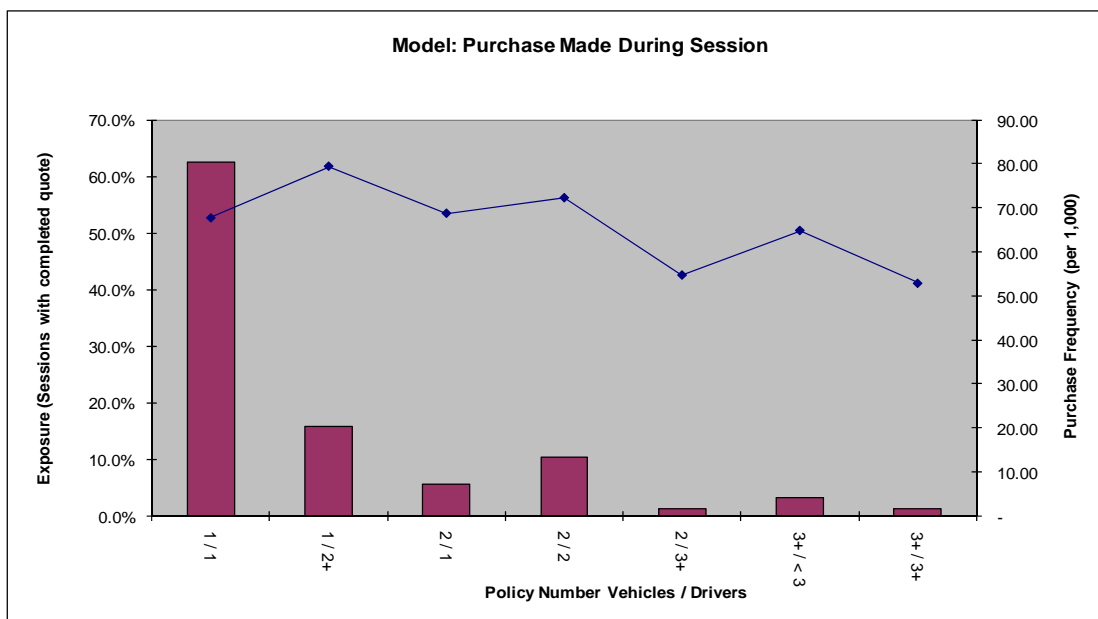
Model: Purchase Made During Session

Figure 5: Shopper Characteristics - Number of Vehicles and Drivers

6

Another interesting distribution of internet auto insurance purchasers was the distribution of vehicles and drivers. For most insurance companies, the majority of insurance policies insure two drivers and two vehicles. However, as can be seen from the above distribution, the majority of web site visitors receiving quotes are one driver, one vehicle policies. Also, comparing one vehicle policies versus two vehicle policies, potential customers with one vehicle have a purchase frequency of about 10% higher than potential customers with two or more vehicles.

QUOTE SUBMITTED

The first step in the insurance purchase process is the visitor to the website initiating the process to obtain a quote, and then submitting that quote to the insurance company. When a customer visits the insurance company website, insurers would benefit from being able to understand what characteristics are associated with customers that are more likely to submit a quote. Therefore, the first analysis undertaken was to create a model to predict the likelihood of a visitor to an insurance website actually submitting a quote. This could be helpful for an insurance company in understanding its effectiveness in presenting its products and prices, can assist an insurer in developing marketing plans that focus on customers that are most likely to purchase a policy from the insurer, and can help insurance companies identify segments of the market that they need to do a better job of presenting products and services to.

For this analysis, the target variable was a binary target, whether or not a quote was submitted. To analyze the target variable, we used SAS Enterprise Miner™ to develop neural network, decision tree, and linear regression predictive models for the target variable. Shown below is a simplified diagram of the models that were developed.
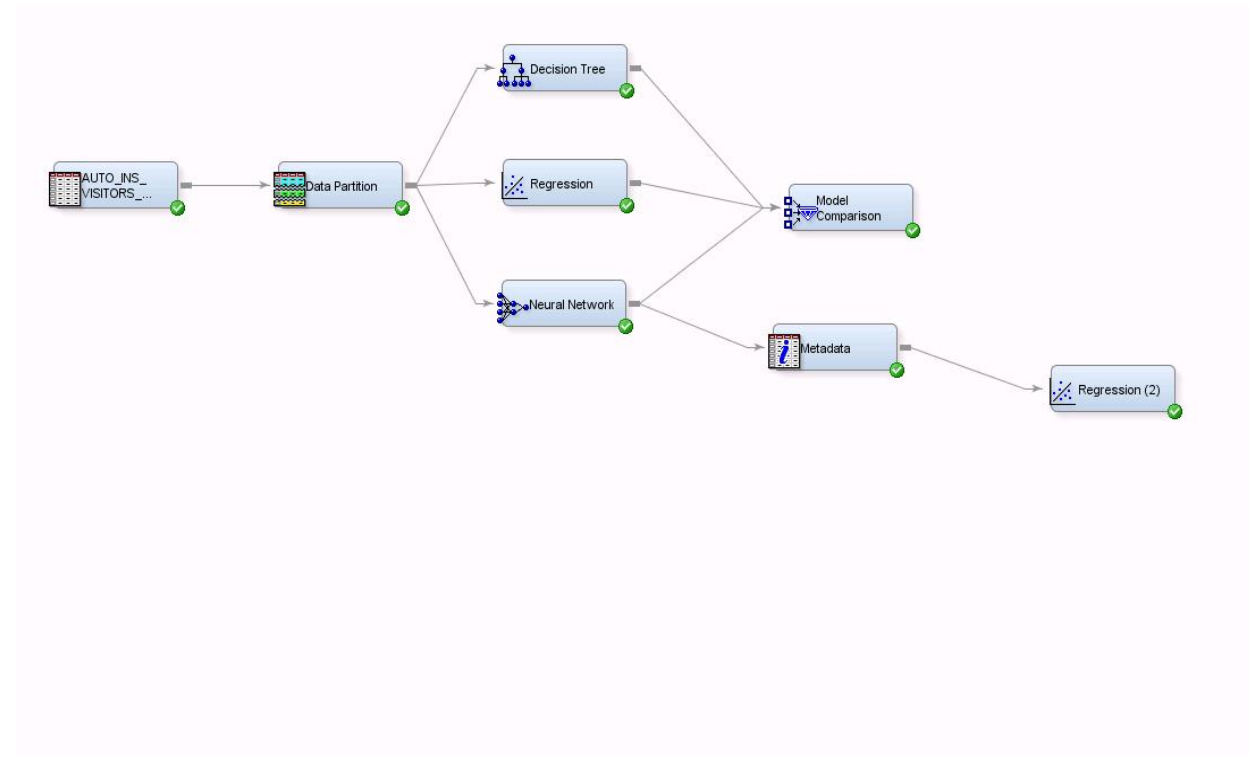

Figure 6: Diagram for Quote Submitted Predictive Model

The decision tree model was instructive because it allowed us to ascertain which variables in the dataset were predictive of the likelihood to submit a quote. Shown below is the output from the decision tree variable importance table.

Variable Name	Number of Splitting Rules	Importance	Validation Importance	Ratio of Validation to Training Importance
total_pages	6	1.000	1.000	1.000
means_of_entry	5	0.238	0.237	0.994
num_prior_visits	1	0.135	0.129	0.956
search_phrase_group	2	0.125	0.130	1.040
time_since_last_session	1	0.115	0.123	1.071
visit_time	3	0.112	0.122	1.087
insurance	1	0.066	0.059	0.888

Figure 7: Quote Submitted Analysis - Decision Tree Variable Importance

As you can see from the table, several variables show significance in predicting the likelihood of a customer submitting a quote. One of the key variables here has to do with the amount of time that customer spends on the website (total pages viewed, total time on the website). While this information is not known when the customer gets to the website, as the customer progresses through the visit it is something that can be monitored and tracked in real time. As the time spent on the website crosses some threshold, companies can implement something that targets the customer and provides encouragement or incentive for them to finish the quote process, such as the offer to chat with an advisor or to receive a call from a customer service representative. Also, another important element is understanding how the customer came to the website. This reinforces the chart we saw above, how the customer makes it to the insurance company site has an impact on whether they submit a quote.

The English rules for a node with a high probability of submitting a quote are shown below.

> *WHERE total_pages >= 15.5 AND total_pages >= 29.5 AND companyA 0 AND num_prior_visits < 20.5 AND total_pages < 58.5 AND total_ssl_page < 22.5*

It can be seen in these set of rules how the variables combined to identify a class of consumers that are more likely to submit a quote to the insurance company. Specifically, this node is centered around the number of pages visited within a particular range, with a limited number of prior visits excluding a specific company. These types of results can be used as an input to make actionable changes to the customer purchase process which would hopefully result in more successful submit rates.

Below is an example of one of the variables from the regression analysis. In the figure below, the regression results for the number of prior visits is displayed. As can be seen, the likelihood of submitting a quote goes down the more times a visitor has previously been to the company's website.
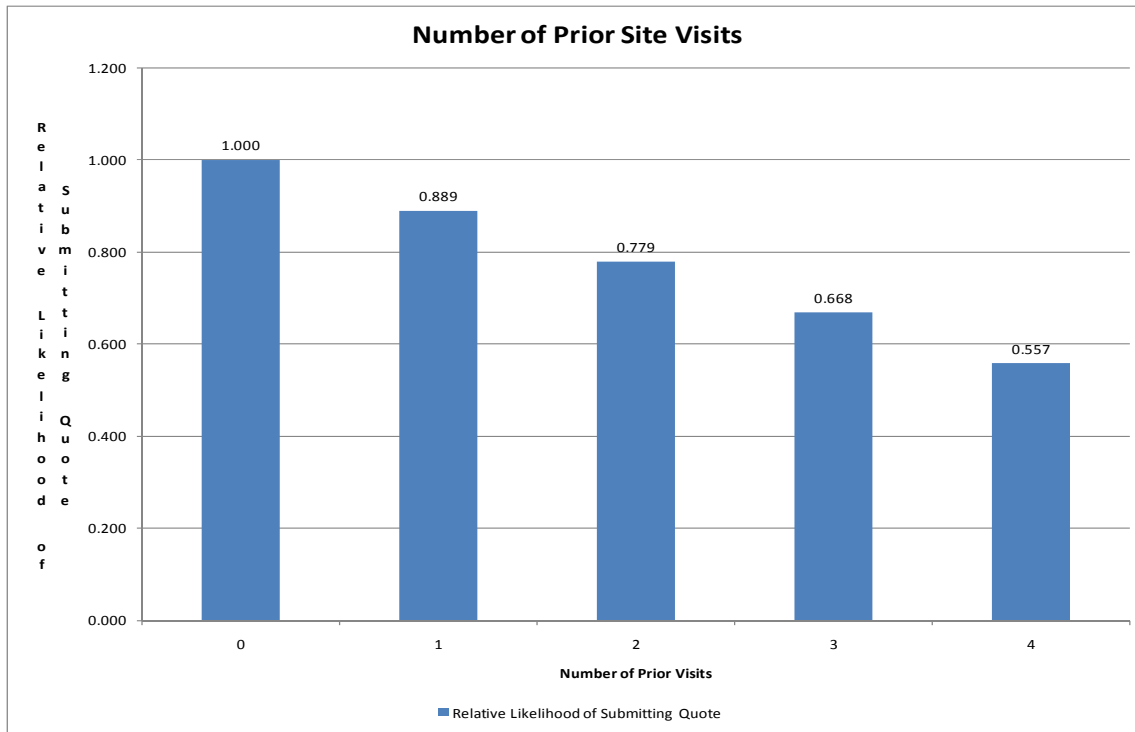
Number of Prior Site Visits

Figure 8: Quote Submitted Analysis - Number of Prior Visits

POLICY PURCHASED

Once a consumer has submitted a quote and received a price, the customer has the option of either purchasing the policy at the quoted price, or rejecting the quote. Insurance companies can analyze the likelihood of a policyholder that received a quote purchasing a policy based on the characteristics of the policyholder. This is known as a customer conversion analysis. These analyses can have a number of benefits:

1. For potential insureds with a higher likelihood of purchasing a policy, steps can be taken within the quote and purchase process to ensure that the customer does not drop out
2. For potential insureds with a higher than average likelihood of purchase, insurance companies would be able to market more aggressively to these potential insureds
3. For segments of potential insureds that have a lower than average likelihood of purchasing a policy, an insurance company can identify and investigate where these lower than average likelihoods are and try to increase the likelihood of purchase

Again, underlying Decision Tree, Regression, and Neural Network models were developed, and then were combined using an Ensemble model. Based on the results of the analysis, there were several categories of variables that were significant in predicting conversion.

One category of variables that was significant in predicting conversion was driver characteristics. Below is an example of one of the driver characteristics.
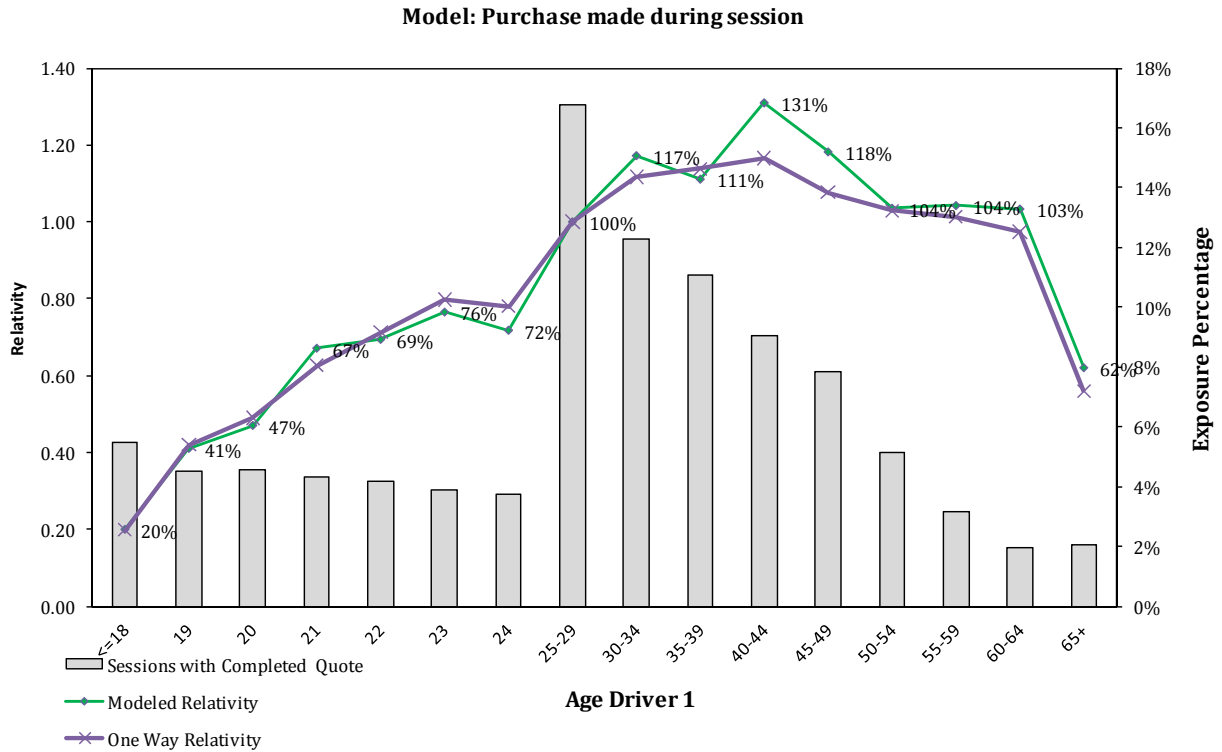
9

Model: Purchase made during session

Figure 9: Policy Purchase Analysis - Driver Age

The figure above shows the likelihood of conversion based on the age of the first driver on the policy. The model results above are based on a multivariate model and are similar to the results shown on the data summaries. They have been rebased to set ages 25-29 as the base class and having a purchase relativity of 1.0. As can be seen, the likelihood of a customer purchasing a policy increases as the policyholder gets older until ages 40 – 44. At this point, the likelihood of purchase decreases moderately until age 64, at which point the likelihood to purchase drops dramatically.

Another category of characteristics which are significant in predicting conversion likelihood are demographic characteristics related to the drivers or the household. An example of the education of the driver is shown below.
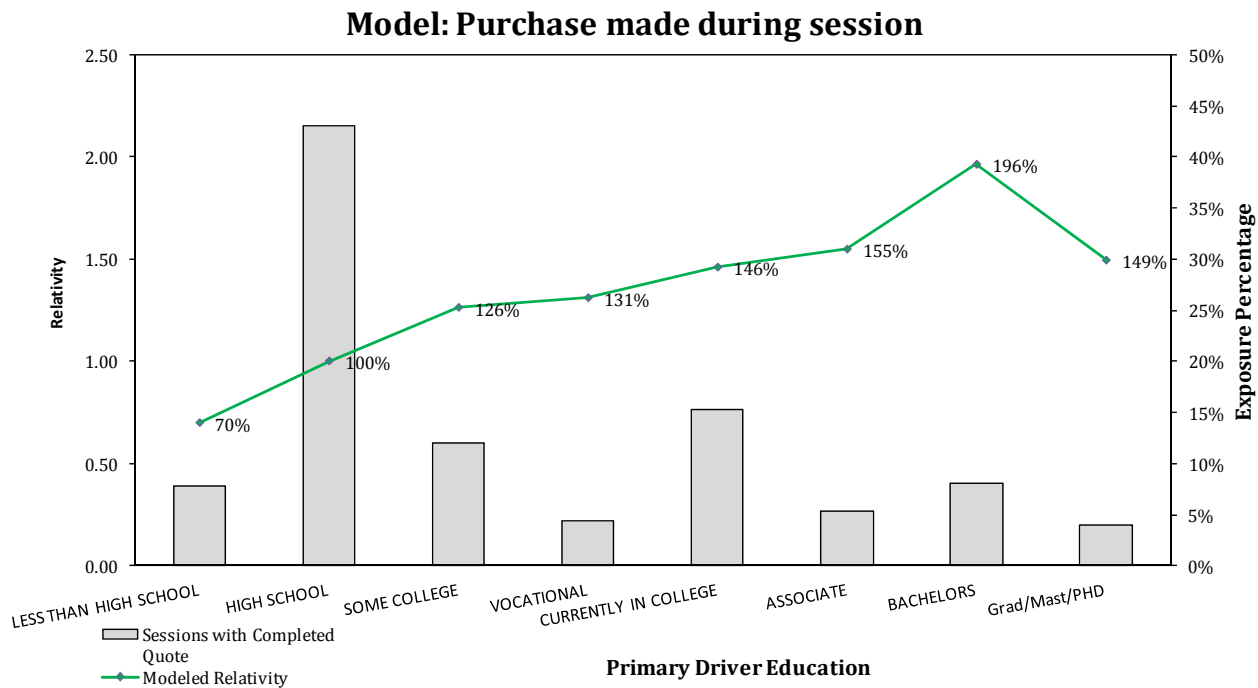
Model: Purchase made during session

Figure 10: Policy Purchased Analysis - Driver Education

There are a few interesting observations based on the chart above. First, it can be seen that about 45% of the web site visitors obtaining quotes had a high school education. In addition, about 45% of the quotes were obtained by policyholders with some college education or higher. The likelihood of a purchase being made increases as the education level of the customer increases. While the majority of the shoppers are at the lower end of the education range, visitors with a higher amount of education have a likelihood of purchase that is twice as high.

As discussed earlier, one of the challenges with analyzing customer behavior is the availability of price change or competitive price information, and how sensitive customers are to these price differences. There are a number of different approaches to understand customer sensitivity to price. One approach is to compare different quotes that a customer receives from multiple insurance sites and determine how the differences in price affect purchasing behavior. Another approach is to review multiple quotes received during the same session, and to determine how these differences in price drive consumer behavior. An example of this is shown below.
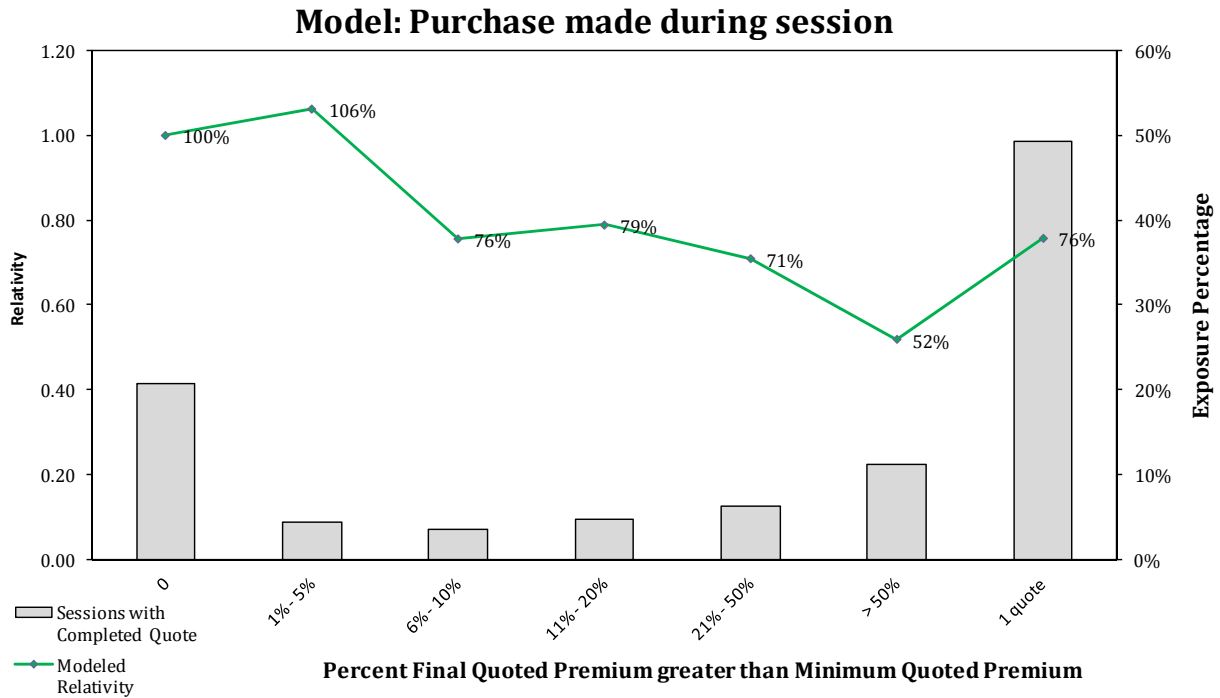
Model: Purchase made during session

Figure 11: Policy Purchase Analysis - Multiple Quotes

The chart above shows the ratio of the final quoted premium to the minimum quoted premium during that session. There are a number of reasons why the final premium could be greater than the minimum premium. There could be changes to limits or deductibles which could cause the price to change. There also could be additional information that the company finds out after the quote is submitted that cause the price to change, such as uncovering prior accidents or moving violations. In the chart above, it can be seen that about 50% of the customers that receive insurance quotes during a web session receive only one quote. For those customers that receive multiple quotes, as the final quoted premium increases relative to the minimum premium quoted, the likelihood of purchase decreases. A premium increase of 5% did not cause a decrease in purchase likelihood, but as the premium difference increased above 5%, the likelihood of purchase decreased by 20 – 50%. While there are a number of different measures that can be used to measure customer price sensitivity, in this case it appears that a premium increase of over 5% causes a significant impact on customer purchase likelihood.

Search Phrase

In addition to the structured data that is available related to auto insurance internet purchasing, there are also several unstructured data element fields that were incorporated into the analysis. One of these unstructured fields is the search phrase. If a customer arrived at the insurance company website using a search engine, the search engine used is captured as well as the search phrase typed in by the user. Given that the search phrase is free form, there are a large number of different phrases that website visitors use to get to the insurance company's site. These search phrases fell into several categories:

1. Searching for specific insurance companies
2. Quotes: auto, homeowners, life insurance
3. Value: cheap, affordable
4. Specific insurance agencies
5. Life circumstances: teen, new car
6. Other

In order to process the search phrase, first we removed punctuation and non-characters. Then, each search phrase was parsed into individual words. Once the phrase was parsed, the frequency of each word was determined. Words were also corrected for misspellings, and then indicators were added to the database for the presence of word in the search phrase for each website visit. Based on the frequency of words in the search phrase, the top 80 words accounted for about 80% of the words present in the search phrases. Once these indicators were built, a cluster analysis was developed to analyze the likelihood of words showing up together in search phrases. Based on this analysis, a series of phrases were developed which were then analyzed a part of the distinction in purchase likelihood.
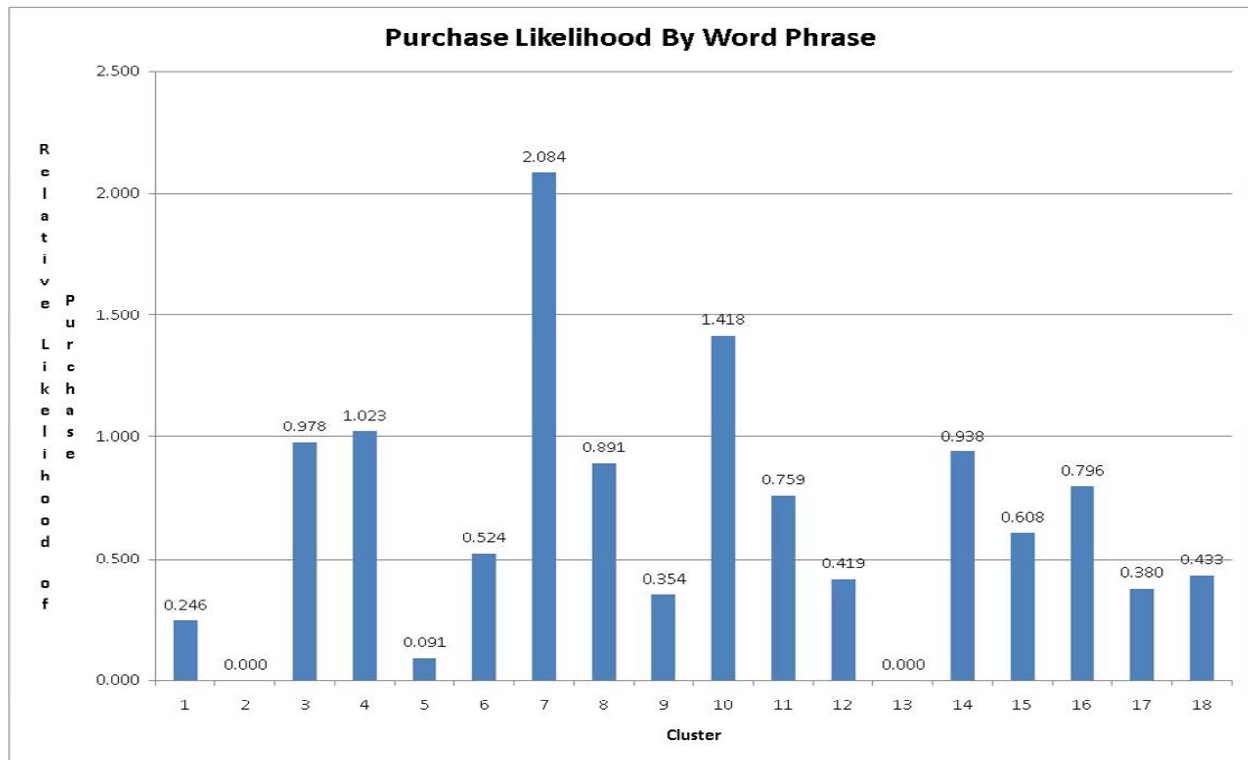

Figure 12: Purchase Likelihood by Word Phrase

In this example, there were 18 search phrases which were identified, and the associated purchase likelihoods are shown below. As can be seen, the likelihood of purchase varies significantly by the search phrase. Understanding how the difference in search phrases can impact purchase likelihood can help an insurer begin to react in real time to help close sales the instant a potential customer arrives at their website.

Overall Model Summary: Policy Purchased Analysis

Based on the overall model development of purchase likelihood, there is a significant difference in purchase likelihood based on the characteristics of the potential insured. The lift chart below shows the comparison between the predicted and actual likelihood of purchase.
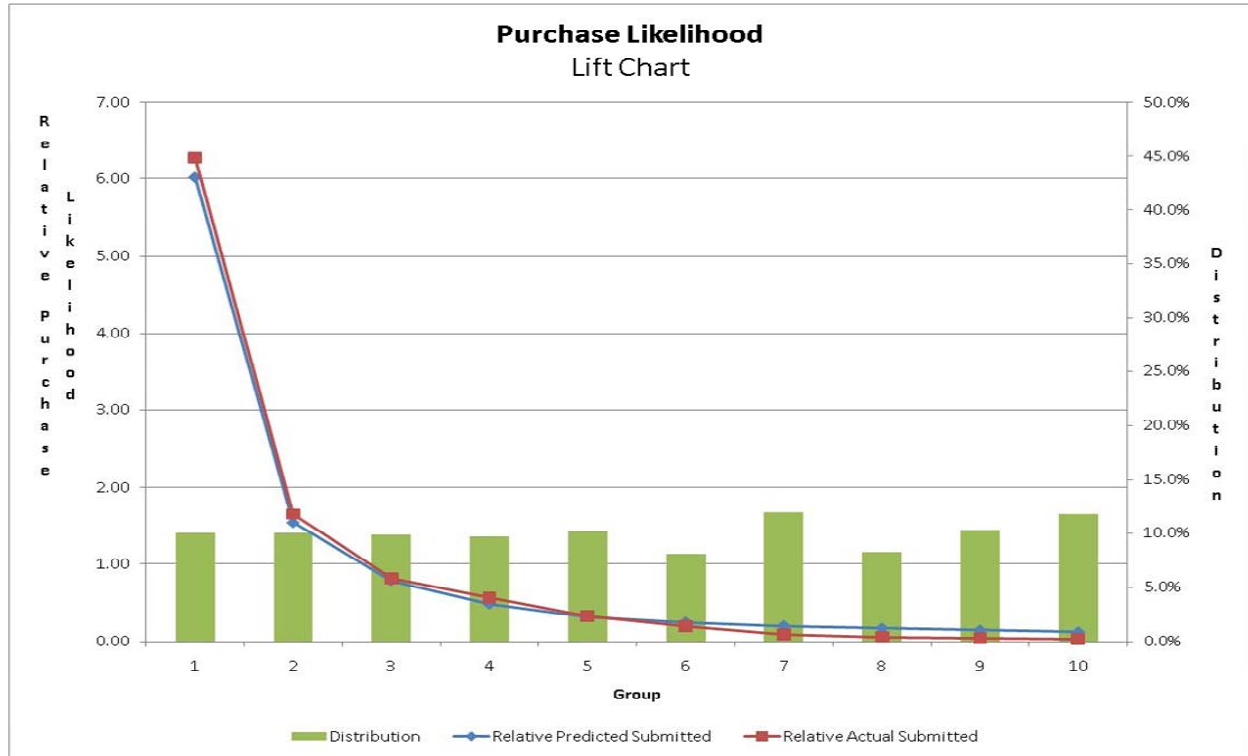
Figure 13: Policy Purchase Likelihood Lift Chart

As can be seen above, the predicted purchase likelihood for the group most likely to purchase is more than six times greater than average, and is consistent with the actual purchase likelihood. The difference between the purchase likelihood for the lowest and highest groups is more than ten to one. If the insurance company was able to identify the likelihood of purchase based on the risk characteristics, different steps could be taken improve the likelihood of closing those risks with the highest probability of purchase. In addition, insurers could work to market to customers with a higher likelihood of purchase. This would improve the close rate for insurance companies, allow companies to make smarter use of their marketing dollars, and ultimately lead companies to improve financial performance due to these insights.

DATA CONSIDERATIONS

As discussed earlier, there are a number of data challenges that a company must face when undertaking customer response analyses. While the analysis of the comScore data helps to address some of those data challenges, care must be taken in interpreting the results of these analyses. First, the comScore data is based on information from a panel. In order to apply this data accurately, one must understand how the characteristics of this panel compares with their understanding of the market. Secondly, the auto insurance purchasing information is related to internet purchasing only. It does not capture information from insurance purchases through traditional channels. As a result, care should be taken in generalizing results. Also, no offline purchase information is captured. If an insured received a quote on line and then purchased the policy through a traditional channel, it will not be recorded here. Lastly, if the policy was purchased from a different computer that is not tracked, this would not be captured as a purchase either.

CONCLUSION

Auto insurance companies have become very competitive. Insurers are using a number of methods to attract potential customers, and once they attract these customers, they are working hard to close the sale. In order to improve their likelihood of attracting and converting these potential customers, insurance companies can undertake

14

customer response analyses to understand more about their potential customers, more effectively focus their resources to improve their overall attraction and conversion rate, and to address potential competitive issues. These analyses can incorporate traditional insurance data elements, price change information, competitive information, and available unstructured elements. Ultimately, understanding the customers can help in establishing marketing strategies, optimizing marketing dollars spent, and establishing pricing policies that drive overall volume and profitability.