For my second project, I wanted to look into something on the business analytics side. I thought about my role doing membership sales at Boston Sports Clubs and tried to figure out what information could make me better at my job.

One of the most important metrics in sales is conversion rate. This can be a ratio of the dependent variable sales to many different independent variables: total leads, appointments, phone calls, hours worked, etc. In general, the conversion rate for each ratio remains about the same over time. This means that as you increase the independent variable, sales will increase at a constant rate. For example, I know that I make about 2 sales for every 3 appointments that I book. If I book 6 appointments per day, I can expect to make about 4 sales per day. If I increase the number of appointments that I book per day to 9, I can expect to make 6 sales per day.

Since I have been in the role for almost a year and a half, I have noticed one problem with these conversion rates. The problem is that every lead is treated the same. However, I know that they are not all equal. I tend to classify them at 3 different levels. There are Ones who are leads with the full intent of signing up. There are Twos who are on the fence and need some guidance or additional information before they decide. And there are Threes who have no intent of signing up.

In our business, leads originate in many ways. These can include walk-ins, trial requests, telephone inquiries, contact us submission, referrals, and prospecting, among others. It is important to track the sources of our leads in order to analyze the conversion rate of each source. From my experience, the lead source is not enough to conclude that one lead might be of higher quality than another. Ones, Twos, and Threes can come from each source of leads.

There is one thing, however, that all leads have in common. In order to be considered a lead, I need to have contact information. I have, at least, either the phone number or the email address of all of my leads. In this project I want to find out if I can draw any conclusions from that basic contact information. I want to know whether the email address or the phone number that I am given affects the likelihood that a lead will be converted into a member.

In this project, I used data from my 1st 12 months at BSC Newbury Street. There were 1380 new leads in the 12 month span. 477 of those leads bought memberships which means my overall conversion rate was about 34.5%.

I was able to break each lead down by the email address and phone number that they provided.

My data set can be found here:

In order to analyze whether any of this information could be used to determine if some of these leads were better than others, I decided to conduct 2 separate Chi-square tests. I chose the Chi-square test because I was working with 2 categorical variables in each test. The first test used the categorical variables email server and membership sales. The question I am trying to answer in this test is “Is the likelihood that a lead buys a membership dependent on the email address they provide?” The second test used the categorical variables area code and membership sales. The question I am trying to answer in this test is “Is the likelihood that a lead buys a membership dependent on the phone number they provide?”

The first step was to organize each of my data sets into two-way tables. One of the conditions of the chi-square test is that each expected cell count must be greater than or equal to 5. In order to comply with this condition, I first had to divide 5 by my overall success rate of .345. If my overall success rate was above .500, I would have to divide 5 by 1 minus my overall success rate. I found 5 divided by .345 was about 14.5. For this reason, I could only use the email servers and the area codes that appeared at least 15 times among the 1380 leads. I grouped the leads that did not qualify into a category of “Other”.

My two-way tables can be found below:

(Insert email two way)

(Insert phone two way)

With these two-way tables, I can analyze some probabilities right off the bat.

First I could look at marginal probabilities. These represent the portion of the entire group that belong in a single row or column. This means that it only considers a single variable. For example, the marginal probability that a lead used gmail.com for the email address is 833/1380 or about 60.3%. The marginal probability that a lead did not purchase a gym membership is 903/1380 or about 65.4%.

Secondly, I could look at joint probabilities. These represent the portion of the entire group that has two particular characteristics at the same time. For example, the joint probability that a lead used a hotmail.com email address and did not purchase a gym membership is 39/1380 or about 2.8%. The joint probability that a lead used a 617 area code and purchased a gym membership 71/1380 or about 5.1%.

Finally, I could look at conditional probabilities. Conditional probabilities look at subgroups in the sample. This is the element of the two-way table that is most important because it allows you to make comparisons between groups, which is why we are looking at these numbers in the first place. For example, the conditional probability that a lead purchased a gym membership given that they used an icloud.com email address is 8/17 or about 47.1%. On the other hand, the conditional probability that a lead purchased a gym membership given that they used an aol.com email address is 4/18 or about 22.2%. In these examples, it appears that it is much more likely that a lead will purchase a gym membership if they use an icloud.com email address than if they use an aol.com email address. However, without additional testing, we can not fully interpret these results.

I am currently able to say that within this sample, if a lead used an icloud.com email address, they were more likely to purchase a gym membership than if they used an aol.com email address. I would, however, like to use these results to draw conclusions about the entire population. If I could find a statistically significant result about the entire population, I could tailor my sales strategy to make me a more efficient and higher performing salesman. In order to gauge these sample results in the scope of the entire population, I need to conduct a Chi-square test.

I collected this data from my first year at BSC Newbury Street in order to test whether there is a relationship between contact information and membership sales. More specifically, I want to determine whether membership sales is dependent on either email address or phone number. Since these variables are categorical, I will use a Chi-square test to look for a relationship.

To continue with the next step of my Chi-square hypothesis test, I need to set up my null and alternative hypotheses. I will go through the steps with my first test of email addresses and repeat the process afterwards with phone numbers. The null hypothesis in this case is Ho: email address and membership sales are independent. The alternative hypothesis is Ha: email address and membership sales are dependent. Another important step is to establish my predetermined cutoff of the p-value. I will use 0.05 to determine whether or not I can reject Ho and conclude dependence. My Chi-square test results can be found below:

(Insert email address chi-square)

This chart shows the observed and expected cell counts. These numbers are used to determine the Chi-square test statistic. In this case the test statistic came out to be (insert test statistic from picture). I am able to look up the p-value using the test statistic and the degrees of freedom. The p-value measures the likelihood that my results were just found by chance while Ho is still true. For example, in this case the p-value came out to be (insert p-value from picture). Because the p-value is greater than my predetermined cutoff of 0.05, I cannot reject Ho and I am therefore unable to conclude that email address and membership sales are dependent. There is no evidence that any lead is more likely to purchase a membership because of the email address they provided.

I will now conduct another Chi-square test with my data from the phone numbers of leads. I first went through the same steps of setting my null and alternative hypotheses as well as set my predetermined cutoff of the p-value at 0.05. My Chi-square test results can be found below:

(insert phone number chi-square)

This chart shows the observed and expected cell counts. These numbers are used to determine the Chi-square test statistic. In this case the test statistic came out to be (insert test statistic from picture). I am able to look up the p-value using the test statistic and the degrees of freedom. The p-value measures the likelihood that my results were just found by chance while Ho is still true. For example, in this case the p-value came out to be (insert p-value from picture). Because the p-value is greater than my predetermined cutoff of 0.05, I cannot reject Ho and I am therefore unable to conclude that area code and membership sales are dependent. Although it appears less likely that the results I found in the phone number data occurred by chance, I do not have sufficient evidence to conclude that any lead is more likely to purchase a membership because of the area code they provided in their phone number.

Although I did not find any of the statistically significant results that I was looking for, I am still able to use this information to become a better salesman.

One thing that is always preached as a basic principle of sales is not to cherry pick. This means that you need to treat every lead as equal. For example, if you are walking door to door to make sales, you should never pre-judge based on anything you see. Each door should be approached with the same expectations. The rule of thumb in sales is that the door you skip is the biggest sale.

This same principle can be applied to these results. These leads should not be pre-judged based on the basic contact information that was provided. Even the variable that resulted in the lowest percentage of membership sales in my sample, area code 917, resulted in 3 sales at a 13% rate. In sales, you cannot leave any stone unturned.

If I found a statistically significant result in my data and stopped working as hard on leads with the area code 917, I could miss out on 3 sales per year, or 1 every 4 months. That could make the difference in our key metrics of net member gain, year over year sales, etc. Also, 1 of those 3 could have bought personal training, and brought additional revenue to the club. Another 1 of those 3 could have referred his roommates and brought in 5 of our highest quality leads – referrals.

The bottom line of this project is that sales is a number game. As I stated in the intro, conversion rate is typically stagnant. The key to improving sales in the gym industry is generating more leads, making more touchpoints, and booking more appointments. It’s as simple as that. I hope you enjoyed this project! Cheers!