Portfolio Project: Option 1

Jessica Ham

MIS543 – Enterprise Performance Management

Colorado State University – Global Campus

Steve Chung

May 10, 2020

Portfolio Project: Option 1

1. Introduction

In this assignment, my company is striving to identify customers who respond to direct marketing promotions in an effort to capitalize on those customers and others like them to increase sales. Our strategy is to maximize profits, which means selling the most product possible at the highest price possible. This does not mean that you do no promote or run specials when sales need a boost, it means recognizing what products sell well at full price, and what objects need promotions to improve sales numbers. Afterall, it is more profitable to sell 10 units at $25 and 10 units at $18 than just to sell the first ten units and be done. One thing many companies have recognized is that it is also more profitable to work with customers that have shopped with you before than it is to try and attract new customers. The marketing is less expensive and tends to work better, since you already have an idea of what this customer likes and some information about them. Put in “…more operational terms, the goal is to understand…how to implement relationship-marketing strategies with the most profitable ones in order to retain these customers and increase purchases made by them” (Estrella-Ramon et al, 2013).

By this logic, we agree to strive not for the strong reaction to a one time campaign, which will only drive sales in the short term and instead generate goals centered around drawing a reaction from existing loyal customers and marketing that will appeal to similar individuals that we can convert into lifetime customers. Creating customer lifetime value “…takes a powerful, integrated approach to turn occasional shoppers into loyal customers” (Sailthru, 2016). This includes actions after the marketing period, such as exceeding customer expectations, keeping your values at the front of view at every level, and maintaining a standard your loyal customers can rely on visit after visit.

With the goals of maximum profit and customer lifetime value in mind, we can take a look at our available sales data. Through Zip Codes we can get an idea of the geography of our customer and using the Microvision Lifestyle Cluster Type categories, we will be able to understand some basic characteristics of the customers we are most popular among. When you are looking at which customers respond to us, both of these are key characteristics. We also have some great data on our established customers, such as total number of lifetime purchases, and the value of those as well as average time between purchases. As we established at the start, one of our main questions is which customers respond to our direct marketing campaigns. This information can be understood by looking at the data included on the number of promotions on file for each customer. First it is important to review the description of the data. Looking at the histograms, we can see that none of the data is normally distributed, which will be important to bear in mind as we work through business questions. Additionally, we can review the means and distributions of each variable which will help determine the category splits on those variables we will later choose to categorize for clarity.

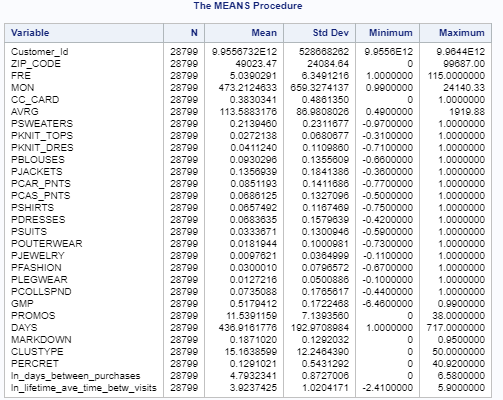


Figure Means Procedure on Variables

1. Business Questions and Analysis
2. Is there a correlation between shoppers that have promotions on file and lifestyle clusters?

H0: There is no significant relationship between customers that respond to promotions and their lifestyle clusters.

H1: There are significant relationships between lifestyle clusters and response to promotions.

This question is again more about groupings than anything else. If we treat both variables as categorical, we can understand the groupings through a chi-square test. To manage this question for meaningful interpretation, I created a new column detailing the top 10 lifestyle clusters and I converted the count of promotion responses to a frequent vs infrequent category, based around the mean value of 11 that I determined from the means procedure above.



Figure Chi Square Test of Independence Promo and Lifestyle Cluster

The results of this test show that while the overall odds of frequent responders vs non-frequent responders is roughly 50-50 across all clusters, however, if we are look at the third statistic, we can see that particularly the Upper Crust and Home Sweet Home clusters respond significantly more frequently than the others, with Mid-Life Success also responding well. Overall, the p-value indicates that these results are significant and that the null hypothesis should be rejected.

1. Is there a relationship between geography and customers that respond to promotions?

H0: There is no significant relationship between geographic location and response to marketing promotions.

H1: There is a significant relationship between certain geographic areas and how often they respond to marketing promotions.

One thought to consider is just how many Zip Codes are represented within our data. This could be visually prohibitive and create challenges in the distribution of data. One possible resolution would be to group zip codes into, for example, regional groups and add this new categorical variable to the data. In this way we can focus our testing to a few larger populations instead of many small groupings. Promotion response can again be used as categorical and reuse a chi-square test.

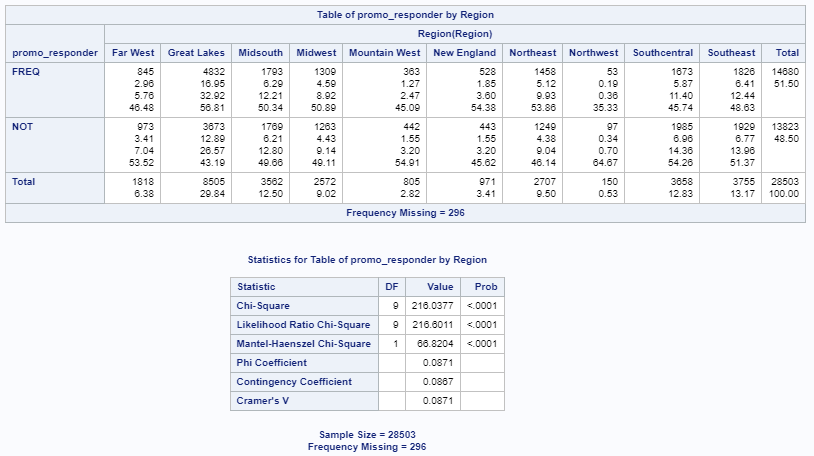


Figure Chi Square Test of Independence Promo and Region

Again, we can see a statistical significance in the p-value which points to a relationship existing between these variables. Notably standing out is the Great Lakes area, where we not only have more customers overall than other areas but a significant number of frequent promo responders.

1. Is there a relationship between frequent shoppers and shoppers that have more promotions on file?

H0: There is no correlation between visit frequency and frequency of responding to promotions.

H1: There is a correlation between frequency of visits and frequency of responding to promotions.

This business question can best be answered by treating both integers as continuous variables. By performing a simple linear regression we can determine if those people who frequently shop with are more likely to respond to our promotions. Linear regression is generally visualized using a scatter plot, where we can clearly see if the more often you shop, the more often you also respond to promotions.

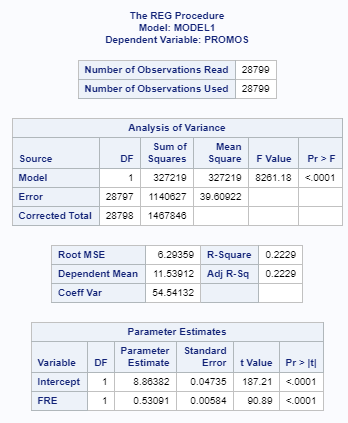


Figure Linear Regression of Promos from Frequency

This information again shows a statistical significance, and a fairly strong correlation between frequency of visits and frequency of promotional response. The equation for the relationship is expressed as: Promo = 8.86382 + 0.53091\*FRE, which would indicate that each visit to a store makes you nearly half as more likely to respond to a future promotion. The plot to represent this is below. While the fit of the model does tend to taper off beyond about the 40th visit and/or promotion, these are a relatively few number of outliers compared to the rest of the model. As a starting point, some fairly accurate predictions could be made with this model.

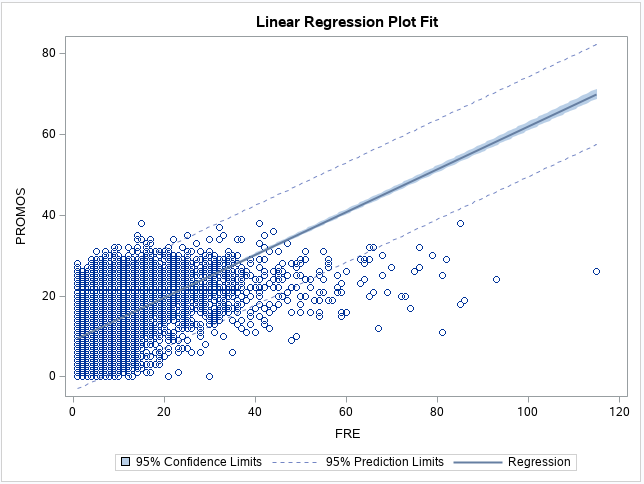


Figure Plot fit for Promo from Frequency

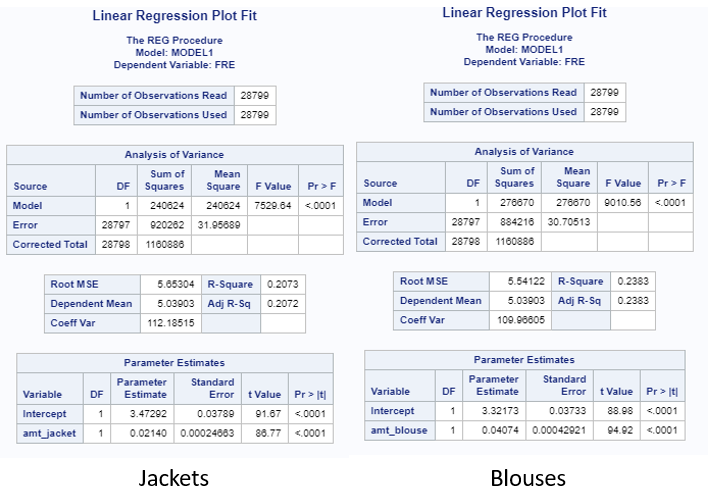
1. Is there a correlation between a particular product sales and frequency of visits?

H0: No relationship exists between product sales and frequency of visits.

H1: One or more products have a significant relationship with frequency of visits.

This is a question of whether a particular product is driving our business. Are there one or two products that are the most commonly bought by our frequent customers. The data provided on department sales is in a bit of an inconvenient format, given as the percent of total sales of each customer. This question could be answered either by creating a categorical variable of the best selling product for each customer and performing a chi-square test on the category of best seller and frequent visitor, or in an effort to keep the data continuous, we could review the statistics for the highest average by product category and run a linear regression to understand the relationship between the top two products and the frequency of visits. As this requires less data manipulation and is a more reliable test statistic, I will work with option two.

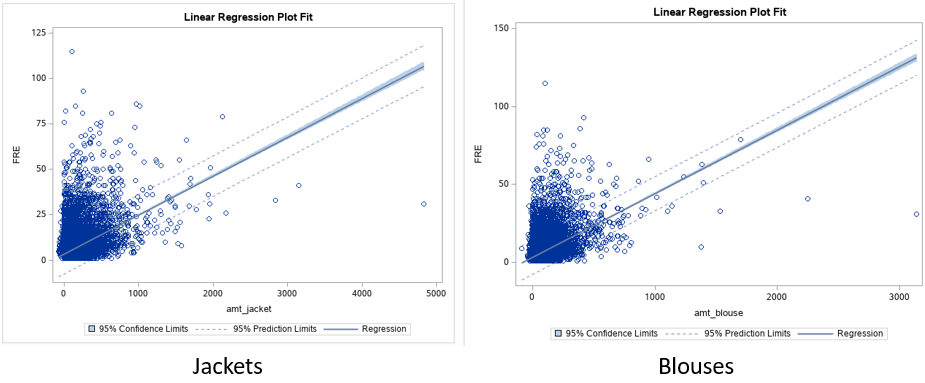
Going back to our descriptive statistics at the top, the highest means among the Percent Departments are in Jackets (13.57% mean) and Blouses (9.3% mean). Jackets do probably have a higher price point that accounts for some of the mean, but even still, these are the two best selling departments we have. My original thought was to run a linear regression model to see if these products were driving visits, however, as one would expect, even if a person goes in for that item on a visit, if they buy other items as well, then the total percent blouses or jackets make up of total sales will gradually go down. Instead I decided to change these percents into actual net sales amounts by department (multiplying the percent by total net sales).



Equations would appear as follows:

FRE = 3.47292 + 0.0214\*Jackets or FRE = 3.32173 + 0.04074\*Blouses

Fit Plots are shown below:



The plots are similar to each other and the fits are statistically significant, suggesting that both are driving sales in some way, but again, the plot appears to diverge as the visit frequency grows above 25.

1. Analysis of Findings and Further Recommendations

Each of these questions recognizes either the goal of understanding our lifetime customer or an understanding which customers respond to marketing (regardless of if they are lifetime or not). As we begin to look deeper into the customers who are most profitable, knowing the similarities in these groups can help us anticipate which other customers will be likely to respond positively to other marketing efforts, with a goal of attracting the gold standard of customers- those that are both loyal and likely to respond to promotions.

There are a few themes that are coming out in the initial business questions that we could explore further to reach a point where predicting these customers is possible. There is a definite geographic component to our frequent visitors and promotion responders. There is also a definite cluster component to these. Other questions that would be worth asking is whether these visits are resulting in sales? Are there characteristics of the geography and the clusters that overlap and can help us to identify our target customer? Other information that could help with insights at this level is not just how many promotions these customers have bought from, but which ones specifically? If we had the responses of customers for each specific campaign, this could again lead us to the most profitable promotions for attracting new customers versus getting our core clients back through the door. These questions lead us to an understanding of the profitable, responsive customer. To help with the prediction of whether a customer is or is not likely to become a lifetime customer, this data would be fed into a logistic regression model, which is visualized as a curve “…constructed using the natural logarithm of the “odds” of the target variable, rather than the probability” (Brid, 2018).

One other theme that came of all the linear regressions was the change in behavior of lifetime customers around the 25 visit mark. Each of the models began to have significant outliers at this level. With the mean visits at 5, I don’t know that I would put my frequent customer split as high as 25, but perhaps it would be useful to split this into 3 groups: Not Frequent (<=5), Frequent (>5), and Lifetime (>=25). In this way you could focus on customers that are visiting more than the mean but are not yet at that point where they stray from the model as separate from those who are lifetime customers and seem to have a different shopping pattern. It would also be important to verify the use of linear regression for each model as you go. As mentioned at the first of the paper, none of our variables held a normal distribution, which could lead to skewed results in a simple linear equation. It is worth noting that in my research I did run question three through a generalized linear model (GLM), which does not require normalized distribution. The results were identical, so I chose to exclude it from the report, however that is not guaranteed to be the case in all further testing. If I were going to suggest next steps for modeling our customers, it would be to isolate these levels of shoppers and then take a deeper dive into the cluster and geographic details for each and how they impact their sales and promotion responses.

References

Brid, Rajesh (2018). *Logistic Regression.* Retrieved from:

<https://medium.com/greyatom/logistic-regression-89e496433063>.

Estrella-Ramón, A. M.; Sánchez-Pérez, M.; Swinnen, G.; VanHoof, K. (2013). A Marketing

View of the Customer Value: Customer lifetime value and customer equity. *South African Journal of Business Management*44:4, p47-64. Retrieved from: [http://eds.a.ebscohost.com.csuglobal.idm.oclc.org/eds/detail/detail?vid=0&sid=b03c98e9-f4db-4902-aac9-880f3eca9010%40sdc-v-sessmgr01&bdata=JnNpdGU9ZWRzLWxpdmU%3d#AN=96352228&db=f5h](http://eds.a.ebscohost.com.csuglobal.idm.oclc.org/eds/detail/detail?vid=0&sid=b03c98e9-f4db-4902-aac9-880f3eca9010%40sdc-v-sessmgr01&bdata=JnNpdGU9ZWRzLWxpdmU%3d)

Sailthru (2016). 7 Strategies for Turning Holiday Shoppers into Lifetime Customers. Retrieved

from: <https://www.businessinsider.com/sc/convert-holiday-shoppers-to-lifetime-buyers-2016-11>.

SimpleMaps (2010-2020) US Zip Codes Database. Retrieved on 05/02/20 from:

<https://simplemaps.com/data/us-zips>.