Final Project

Joseph Costanzo

Southern New Hampshire University

Background and Introduction

The overall problem I am looking to solve for the Bubba Gump Shrimp Company is to increase the overall revenue growth of the company through various data mining techniques. This will be done using the data collected over the years that the Bubba Gump Shrimp Company have stored within a data warehouse. The strategy I plan to use is to go through the various information that they provide such as customer purchases and location sales to construct graphs and charts that will provide a clear view of what makes the Bubba Gump revenue grow and what products hurt the growth.

Analysis Tools

The tools I plan to use are JMP and Excel. JMP will be used to gather and organize the data while using different tools to help visualize patterns needed to demonstrate outlet and product satisfaction among customers. Excel will be used as an additional tool to help analyze the data, as well as provide additional support for various graph representations of the data.

Data Visualizations

Due to the task is to analyze the results of a survey, I believe the use of a social network analysis would help with the information gathered. With social network analysis, I can use the outlets that the customers previously purchased from with the survey result answers to see which outlets correlated with which survey answers. Another visualization I plan to use is the histogram. The histogram will show more relations based on the results of the survey and the customer’s purchase history.

Research

The biggest question that the survey results should answer is: Which of the Bubba Gump outlets receive the best results and which receive the worst? This question will help answer which of the Bubba Gump outlets promote customers to return and which will deter further making customers less likely to shop at Bubba Gump. To answer this question, I plan to use the various survey results categorized by Bubba Gump outlet to visualize customer satisfaction based on the products and their experience using the outlet.

Other questions that would help be answered are: Which products have the best customer response? Which outlets have the most traffic among the 500 customers? These questions will give more insight on what products draw customers to shop at Bubba Gump outlets and help us learn which locations draw the most customers. There are quite a few published sources that go over the topic of whether online shopping is better or not than in store shopping. These sources will help provide additional support to my analysis on which Bubba Gump outlets provide a bigger return.

Analysis Organization

Throughout my report, I made sure to keep the analysis of the data as organized as possible. My process was to take each technique one at a time to properly explain how they were used to show which data is both hurting or benefitting the Bubba Gump profits. For example, when I used cluster analysis, I first described what the graph showed. I then explained how customer age played a part in web store sales. Afterwards, I gave the suggestion that the company should target younger customers for the web store to increase sales.

Sources of Error

Some errors I found when running the tests occurred mostly when I used the wrong variables. When using these tests, I tried to use the location of stores in combination with web sales to find a correlation. This ended with a loss of time as I couldn’t find a successful chart that displayed anywhere that location of store would affect web sales. I decided to move to using age instead of location and found more success.

Meaningful Patterns

After switching out location to age in the graphs, I discovered that more younger customers were spending more money on the webstore. This indicates that the question of whether age matters when promoting certain avenues of revenue is a useful course to follow when looking to increase profits.

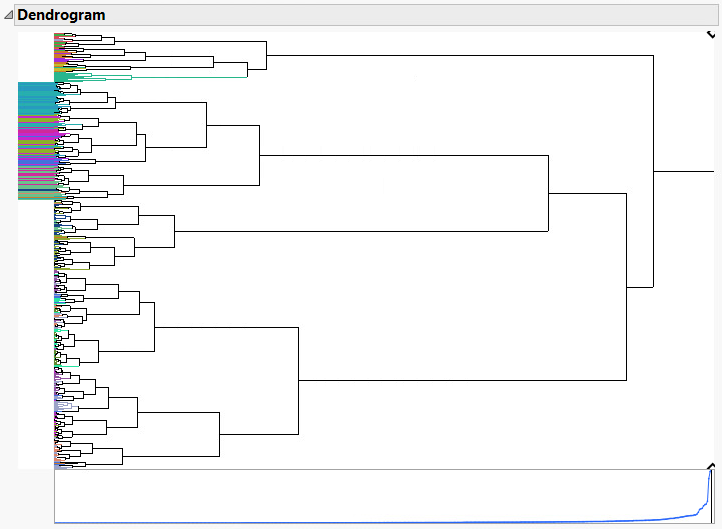
Inaccurate Depictions of Data

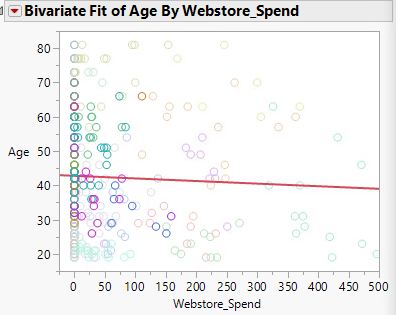
When I encountered results in the data that I believed to be inaccurate, I changed the way the data was organized within the tables. When something seemed off about how the data was represented, I would switch the variables around or even replace a variable with another to receive a better display.

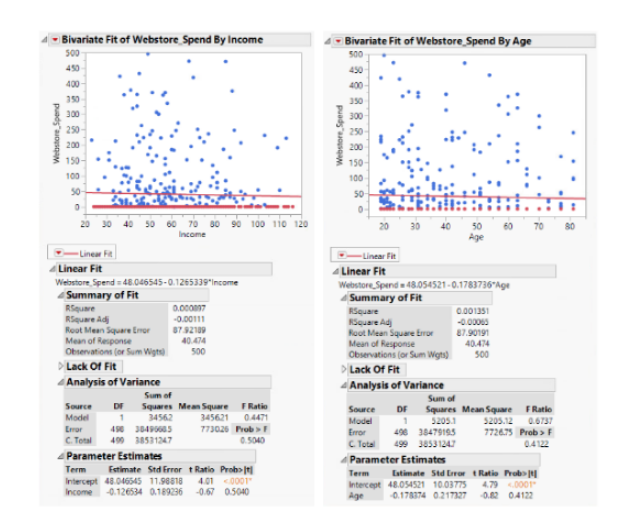
Alternative Analytic Methods

Another analytic method that I would have liked to use for this data is another grouping method such as discriminant analysis. There were times when I felt that the data was too diverse with the amount of survey results to properly show definite results. With discriminant analysis, I could separate two or more groups from the results to better find how a certain demography trend differently than another. Path analysis is another I’d like to use to show how many customers followed the same route when purchasing from Bubba Gump.

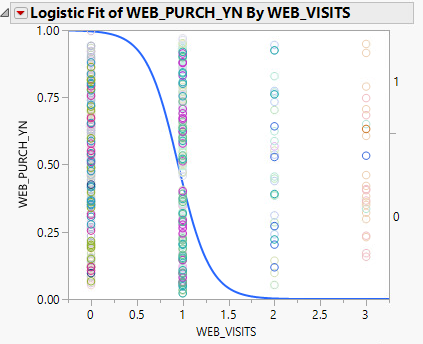
Display and Interpretation



 The dendrogram above shows the clustering relations between the web spending, age of the customer, and income. A natural cluster that appears from this graph is the fact that older customers are less likely to spend as much as younger customers. Going further down that trail of thought, the linear regression graph on the left further displays the relation between age and the amount a customer spends in the webstore. As seen by the line, we can say that on average customers under the age of 45 are more likely to spend more money using the webstore.

**** Adding to the age, I’ve also incorporated these two charts above. The graphs compare age and income of customers against how much money the customer spent in the store. The first graph shows that many visitors with an average income of 30k-65k, spend a lot more than those with a higher income. This means the webstore should appeal more to the average income customer by having more reasonably priced items in the store rather than items with big price tags that would only appeal to people with more disposable income.

The second chart above further proves that the Bubba Gump Shrimp company should focus more on the younger customers when advertising the webstore. As you can see in the graph, there are far more customers who spent money in age range of 20-35 than those who were older. The line also represents that as the age of customer raises, the number of purchases lowers. This can be seen by the red line declining as it goes to the right.

The logistic regression graph shows whether a customer has made a purchase on the web store based on the number of times they have visited. The line shows that over half the customers that have visited the web store at least once have purchased something and that multiple visits to the store is a definite yes to buying a product. This means that if the web store, when promoted, can draw purchases from customers over 50% of the time and any return trip to the web store equates to additional revenue for the Bubba Gump company.

Validity, Reliability, and Limitations

These reports are valid and reliable due to focus of looking mostly at the age of the customers and how different statistics that are still within the general area of the web store are affected by just the age alone. The amount of survey responses in combination with the variance of locations and backgrounds give more credibility to the reports as results seem to correlate across the board.

Resulting Decision Influence

The resulting decision that comes from these results is that the Bubba Gump company should focus on promoting their web store towards their younger customer base. Another suggestion should be to try to promote customers to return to the web store. This is because of the data stating that customers who visited the store more than once at least made a purchase. For example, providing a coupon code for previous customers for a web store purchase should help push customers to return.

Visual Evaluation

The style I chose to display the data was based on how well it showed the information I wanted to explain. For example, I started with the clustering dendrogram to introduce the amount of different natural clusters that are within the data set. I next chose a linear graph to display a clear and easy view of how age affects web store visits. Finally, the logistic regression graph provides additional detail to give an extra bonus of how to effectively use the web store.

Next Steps

The next steps would be to look at the other data that was provided by survey results. Younger customers are drawn more to the web store so would that mean older customers might prefer visiting a physical gift shop? I’d also like to look at new survey results that provides more details on what products are more profitable. Do different locations have different tastes among the customers? Or is age a factor to which products are sold? Could older customers be more interested in the web store if it featured different items?