

8-2 Final Project Submission

Christian J. Kesler

Department of Computer Science, Southern New Hampshire University

DAT 220: Fundamentals of Data Mining

Professor Lee Edwards

August 22, 2021

Introduction

Business Problem

The business problem that Bubba Gump Shrimp Co is currently facing involves fighting the steady decline of business after a spike from external publicity. This decline is to be expected as trends come and go, but maintaining a steady business plan for the future is critical to keep Bubba Gump operational. The goal is to find an aspect of the business plan that has untapped potential to focus efforts towards. This will likely be found as a target demographic, specific type of sale platform, geographical location, or any combination therein. For example, the focusing in person sales at a location near a tourist hotspot that relates to the movie Bubba Gump was featured in would likely help appeal to casual tourists and fanatics alike. It is also possible that the number of fans who are not able to visit in person but would still order Bubba Gump products online is notable, making an effective online sales strategy highly important.

Analytic Method

The analytic method that will be employed throughout this report involves the observation of data clustering and predictive analysis. By examining the survey data provided by Bubba Gump customers, information about the most effective business strategy regarding sales platform, customer demographics, products sold, and geographical location. These clusters will offer insight into the best avenues to focus on for customer reception, and coupling this with past data will allow insight into future business sustainability. I highly suspect that the current level of expansion that Bubba Gump has attained will not be sustainable, requiring the down scaling of several business aspects. However, by increasing the business efforts in the most effective area(s), I am confident that Bubba Gump can continue operating at a notably

high level. The effectiveness of the focus area(s) may even prove enough to maintain other business aspects and reduce down scaling, or allow additional expansion into the focused business strategy.

Plan for Analysis

Analysis Tools

The data analysis tools that will prove most beneficial in this situation will likely be JMP so the data can be organized without having to read through the data manually. The data mining tools present for this scenario are minimal since the data has already been collected and presented to us. However, if that were not the case or Bubba Gump decides to evaluate data beyond the sample survey in the future, then I recommend using the rapid-miner platform. This software is entirely free and open source, making it a safe, reliable, and risk free option since no financial commitment is needed for its use. This software could also be used on the survey data in order to create predictions regarding future trends, but that is somewhat beyond the scope of this scenario. These are merely options for moving forward beyond evaluation of the survey data.

Data Visualization

The key data visualizations that I think will be most important for analyzing this data include histograms and boxplots. I will be examining these for every relevant category such as location, age, marital status, amount spent, and so on. The results of my findings will be easiest to illustrate with histograms and boxplots if the data type allows. The benefit that these types of data representations provide is in their ability to convey relationships and relational differences in a manner that is easy to comprehend. It will allow me to see trends that I might not notice otherwise, and

will allow me to more easily explain my findings to the client and whoever else may be present.

Research Question

My research question moving forward with this analysis scenario is as follows: “Which aspect(s) of the Bubba Gump business strategy has/have the most potential for sustainable success? Which aspect(s) of the Bubba Gump business strategy has/have the least potential for sustainable success? Which aspect(s) of the Bubba Gump business strategy is/are the holding back the company as a whole from sustainable success?” By answering these questions, we can begin to see which areas of Bubba Gump were only allowed to thrive due to the excessive media attention. This will provide insight into which business aspects will be the most critical to sustained success.

Research Measurement

The difficult aspect of this scenario is the highly fluctuating nature of the market. Measuring the certainty of my conclusions will be difficult if even possible considering that the interest in Bubba Gump is declining rapidly. Escalating that interest back to peak levels is likely impossible since the kind of exposure gained from Hollywood is truly not replicable. We can say with certainty that Bubba Gump is decreasing in popularity, and the unfortunate truth is that trend might be irreversible. The goal of my research is to find the business strategies that are most likely to succeed in this state of decreasing popularity. Even if I do uncover the best business strategy available, it might not be effective enough to sustain Bubba Gump at all. This makes the scenario somewhat impossible to fully quantify from beginning to end.

If my research does not yield results that can save the business, then I can never say for certain that my inquiries were fully explored. However, I can say with

certainty that my research questions have been fully explored if the business begins to succeed to the point of sustainability or greater. In summary, my research might not be able to save Bubba Gump, but the risk is negligible since Bubba Gump is declining sharply as is.

Follow-Up Questions

The best follow up questions that will need to be considered after I have identified which business aspects are the most beneficial and detrimental include assessing how much growth or reduction is needed in each department to achieve sustainability. For example, “If in-person restaurants are costing more than they are earning, how much must we reduce the in-person restaurants to eliminate that loss? Will all the restaurants need to be closed? Or only a select few since they could have potentially overlapping target demographics?” Another example might be, “If the merchandise sales online are generating more revenue than the other business ventures, how much can that be increased? Should the web portal be improved and advertising be explored? Or should a larger selection of items be offered?” By determining what business ventures are most and least effective, we can then begin to determine the level of adjustment needed in each department. When those questions have been answered, actionable plans can be implemented and begin getting Bubba Gump back on track.

Research and Support

The external resources that I would turn to are the observable strategies that other companies in similar markets have employed to great success. Red Lobster was performing quite poorly back in 2013, but began to generate much more revenue after being sold to Golden Gate Capital in 2014. Red Lobster remains a prominent high quality restaurant chain to this day, so attempting to replicate their success will give

us a fantastic starting point to build a business strategy around. If all goes well, then mimicking Red Lobster with realistic adjustments along the way might be the change that Bubba Gump needs. If this research yields less than ideal results, we will still have a good idea of what competitors in the market are doing to remain relevant. This might inspire ideas that will allow Bubba Gump to take a unique path that appeals to customers in an entirely different way.

Analysis

Analysis Organization

The analysis thus far has been guided by the research questions asked regarding Bubba Gump Shrimp Co. These questions were “Which aspect(s) of the Bubba Gump business strategy has/have the most potential for sustainable success? Which aspect(s) of the Bubba Gump business strategy has/have the least potential for sustainable success? Which aspect(s) of the Bubba Gump business strategy is/are the holding back the company as a whole from sustainable success?” By pursuing these questions in a manner that explores each topic and expands upon the discoveries made, a naturally organized research structure begins to appear. This is beneficial since reports regarding the findings will be made with a similar sequence, allowing the customers and other viewers to follow the logic through each step in the process.

That being said, there are several aspects of this project that are outside my control. The first and foremost being the data collection methods employed. I was tasked with analyzing data that had already been conjured and merged from multiple sources. While I was able to make meaningful discoveries with the data given, my insight would be greatly amplified had I been able to examine and manipulate the source data myself. As it stands, I have no way to verify that this data was merged properly from multiple databases, which database was used as the root for each merge,

or if there were conflicting fields that may have been overwritten in an erroneous way. I have no way of ensuring that the data I am examining was compiled correctly, and that makes the validity of this analysis impossible to confirm without additional information.

Sources of Error

As I previously mentioned, I have no way of knowing if this data set was compiled correctly. However, my complaints thus far seem like vague statements that carry more of a liability dismissal than constructive feedback. In order to make my concerns with the validity of this report clear, I will be highlighting specific errors I noticed while working with the data set. I hope that by bringing these errors to the forefront of the report, BGSC can account for them when making business decisions and adapt their data collection strategy for future analyses.

Several aspects of the data can be verified in a sense given that the values were nominal. We can confirm they are nominal since variables such as age group fit the expected one based on the movie that propelled Bubba Gump's popularity, but I think the certainty stops there. The trends that I observed regarding the relationship between web store spending and restaurant spending seem to indicate a strong correlation. However, there were no outliers of any kind, which indicates it is entirely possible that customers only began being tracked in the data warehouse after making their first restaurant purchase. This could mean that customers who purchased items from the web store before visiting a restaurant are not being listed as web store spenders, making that correlation entirely incorrect. While it is not impossible for this data to be accurate given the age of the target demographic, zero outliers of any kind rouses suspicion and warrants additional investigation.

While my prior concern is mostly conjecture and assumption, the zip code data can be condemned with a serious level of certainty. I found that the zip code variable has almost no correlation to the other variables, but the correlation between the zip code values strongly suggests that an error has occurred while creating or importing this data set. Most of the values move along a near perfect $x=y$ correlation, but a series of outliers exist that persist in the low zip code 1 values. This indicates that an error with entries beginning with zero has skewed the values, and potentially skewed the zip code 2 values if the additional digit was pulled from that column to fill the five digit capacity of the first column.

This zip code error is a serious one that could heavily skew any observations made regarding geographical location. Correcting it is incredibly important, but it might be causing even more errors than already suggested. Since I know very little about the way that this data was generated other than it was taken from the data warehouse, it is entirely possible that the city and state fields were automatically filled in using the skewed zip code value. This could list a customer as living in a completely different city and state than they actually do, making observations regarding geographical location impossible until that correction has been fully investigated.

There are also errors in the data set regarding the naming conventions of the categories and the vagueness of the income values. The inconsistencies therein make me suspect that the survey responses were merged with existing data from an automated database in the hopes of better understanding the context of the customer responses. This is a good strategy, but to be implemented to the fullest, appropriate standardized naming conventions along with clarification on the income rates need to be introduced. I was able to discern the meaning of each field based on the general

groups that were formed. While I did make an effort to be clear what the data meant in my report, the graphs are still poorly labeled.

Meaningful Patterns

During my cluster analysis, I was able to discern patterns and differences in the restaurant spending and web store spending as sources of income for Bubba Gump. The cluster for restaurant spending was very wide and shallow, meaning that all data points had highly similar values. The overall cluster pattern was rather uniform, suggesting that restaurants are a very stable and risk free source of income for Bubba Gump. On the other hand, the web store spending cluster had a huge gap in the data values. This is likely due to a significant portion of people spending little to no money on the web store whatsoever. The clear divide in this cluster suggests that the web store is a less reliable and more risk ridden source of income for Bubba Gump.

I generated a linear regression for web store spending in comparison to every other applicable variable in the data set so that correlations might be observed. Many of the resulting graphs contained few meaningful correlations of any kind, such as by city, county, zip code, restaurant visits, age, marital status, and income. However, there were a few graphs that stood out in a significant way. Web store spending by state showed that there are some states that have significantly less web store spending than others. This doesn't provide many actionable insights since web stores do not require physical establishments based on location, but it could guide targeted advertisement efforts in areas that already show great promise. Several states that seem like excellent candidates for such advertising include NJ, NY, CA, and FL.

I also discovered that there is a nearly perfect slope on the web store spending by restaurant spending graph, strongly suggesting that customers only visit the web store after visiting the restaurant. Because of this, web store advertising within the

restaurants themselves is a strategy that should be fully explored. There is a somewhat similar, if not as strong, correlation between web store spending and web store visits, but anything otherwise would be indicative of erroneous data mining practices. What is truly interesting to note is that the web store spending very consistently decreases as the customers spend money on third party merchandise. This might be due to third party merchandise being identical to web store products, but there could be other contributing factors. A similar slope can be seen on web store spending by third party visits, though it isn't as clearly defined. This leads me to wonder if web store sales could be increased by reducing or even eliminating third party sales, allowing Bubba Gump to gather data more directly and generate more profit per sale.

I also generated logistical graphs depicting relationships between all variables and whether an online purchase was made. It became clear that at least most of the customers who visited the web store once made a purchase, and every customer who visited the web store twice made at least one purchase at some point. This means that the web store is highly effective at making sales once the customers actually visit the website, meaning that advertising the web store to increase web store traffic will almost certainly increase web store sales. This is by far the best indicator of whether a customer will make a web store purchase or not. Every other graph I generated had mostly level trendlines, with the strongest slopes showing that younger customers with low income are only slightly more likely to make a web store purchase.

Inaccurate Depictions of Data

An interesting trend to note is that the number of third party visits and spending are the opposite of what we might expect them to be based on prior observations. The likelihood of making at least one web store purchase actually

increases as the customer spends more on third party merchandise and makes more third party visits. This doesn't necessarily disprove our prior observations since this is simply a single web purchase binary, but it does show that interest in third party merchandise doesn't completely deter customers from web store products. It might even indicate that the web store makes customers more likely to visit and purchase from third party merchants. Unless of course, there is an error with the data collection or merging with third party merchant information. Considering the flaws within the data that came from within BGSC itself, I would find it difficult to believe that data from an external third party source would be merged in perfectly and fully accurate. I believe this counter-intuitive trend warrants that investigation.

The only other issues I have regarding inaccurate depictions of data have to do with the inaccurate data underneath. My concerns have been thoroughly expressed above, but in summary they surround the data collection of web store spending, zip code, and field naming conventions such as income and "Restaurant". All of these errors were kept in mind as I generated visual data depictions, and no erroneous data depictions arose to highlight new concerns. The erroneous data depictions that I witnessed only confirmed certain cases such as the zip code, or reinforced my suspicions such as with web store spending versus restaurant spending.

Alternative Analytic Methods

There are several methods of data analysis that can be used to gain actionable business insights, and they include Text Analysis, Statistical Analysis, Diagnostic Analysis, Predictive Analysis, and Prescriptive Analysis. Each of these methods focus on a unique aspect of the data even though the actual data being viewed might be identical. The purpose of these defined methods is to ensure that every relevant

and beneficial perspective be effectively employed in isolation, increasing the clarity of the observations made at each stage.

Text analysis involves looking at the raw data and converting it into what might be considered business attributes such as net worth or company size and so on. Text analysis for BGSC was already complete prior to the beginning of this research, albeit in a cursory fashion. Statistical analysis seeks to answer questions regarding what has happened within the business, connecting multiple attributes and tracking them across historical data. The data can range between a few days or hours to several years, even centuries if needed, depending on the event being observed. Diagnostic analysis usually follows statistical analysis, and attempts to answer questions surrounding why an event took place. This involves looking at correlations during key time periods both within the company and from external sources such as economic or environmental factors. Predictive analysis is usually built upon diagnostic analysis, and tries to answer questions surrounding what might happen in the future. This can be used to either replicate notable successes or avoid repeating failures. Prescriptive analysis is what might be considered the analysis capstone in that it brings together all the data derived from prior analyses to create a comprehensive perspective regarding the company data. Prescriptive analysis is often used to generate reports that advise influential individuals such as CEOs or other management positions within companies.

The types of analyses employed for this project were predictive and moderately statistical. Information regarding the text analysis was already provided in the form of BGSC size and scope. The company decline was the event that required attention, but the nature and origin of the decline was readily apparent, making statistical and diagnostic analysis surrounding that turning point redundant.

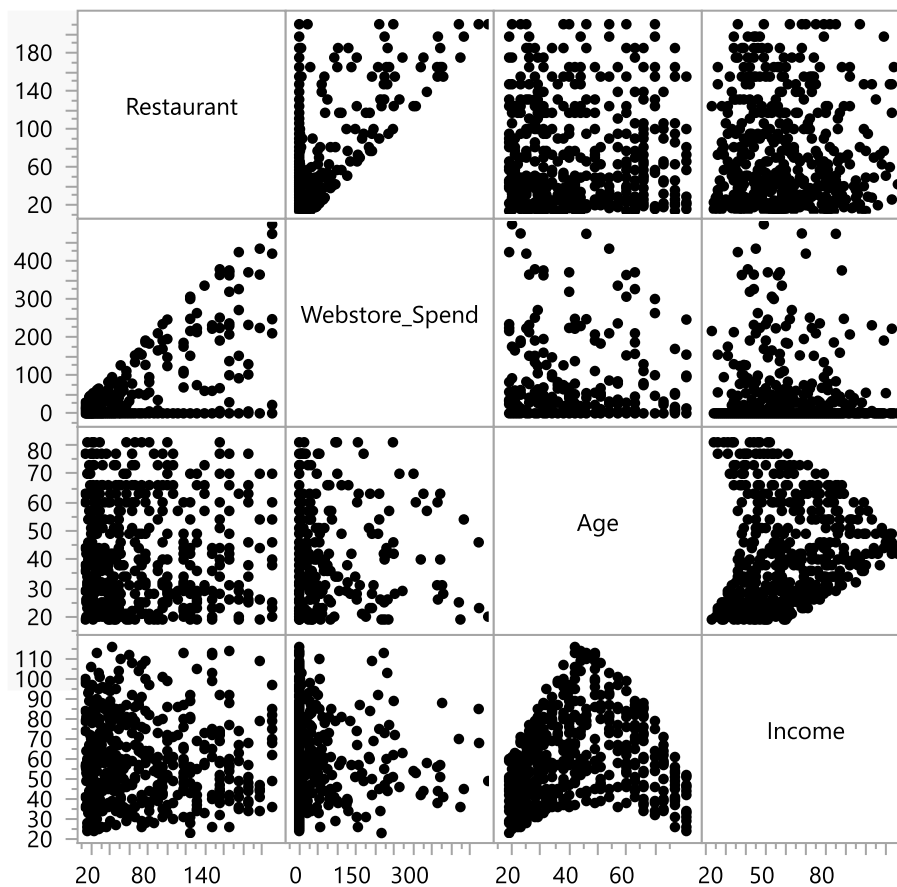
Statistical analysis was employed when reviewing what is currently happening within BBSC, but given the nature of the situation, it only counts as statistical analysis in the loosest sense. The majority of the analysis was predictive in that current trends were observed so that successful strategies could be amplified and inefficient strategies corrected or eliminated. While the data has been compiled in such a way that a prescriptive analysis seems like a valid description, I don't believe that my report can be considered as such in all honesty. While relevant text analysis, statistical analysis, and even diagnostic analysis were presented alongside the data set, the nature of the conclusions were shorthanded and far from academic or systematic. I was able to do my part to create a predictive analysis with the data provided and format it in a digestible report that offers actionable insights and guidance. However, for a proper prescriptive analysis to be compiled, all the other analyses need to be pursued thoroughly so that the prescriptive report can be deemed comprehensive with a confident level of certainty.

Final Report

Display and Interpretation

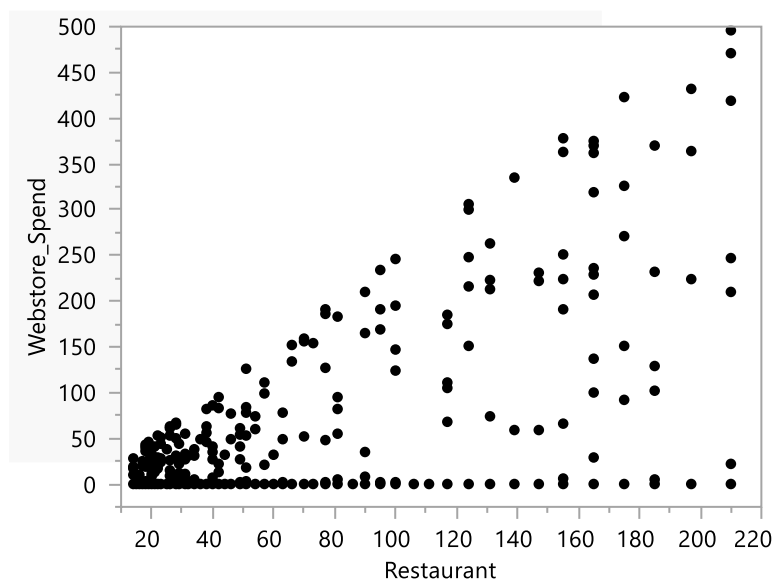
After reviewing the correlations between all variables in the entire data set, several relationships stood out as immediately meaningful. The relationships between Web Store Spending and Restaurant Spending, along with Age and Income present interesting patterns. Let's take a closer look at all four and their relationships below:

Scatterplot Matrix



We can see that those with the highest income tend towards the middle of the age ranges provided, which gives us information about our ideal target demographic for advertisement. The correlation between Web Store Spending and Restaurant Spending makes it clear that not a single customer from the survey spent money on the web store without also spending money on the restaurant, but there were a significant portion of customers who spent money on the restaurant without spending any money on the web store. A more detailed graph of the relationship between Restaurant spending and Web Store spending is as follows.

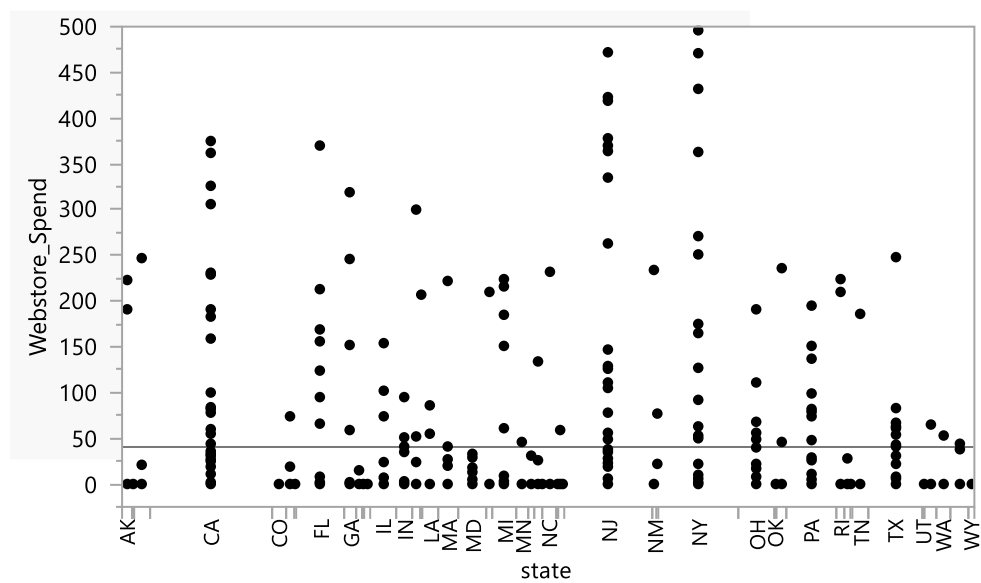
Bivariate Fit of Webstore_Spend By Restaurant



We can see that there is a nearly perfect slope on the web store spending by restaurant spending graph. This isn't conclusive by any means, but it does suggest that the restaurants are the first point of contact with Bubba Gump, further guiding our strategy for tailored advertisements to new and returning customers. Because of this, web store advertising within the restaurants themselves is a strategy that should be fully explored.

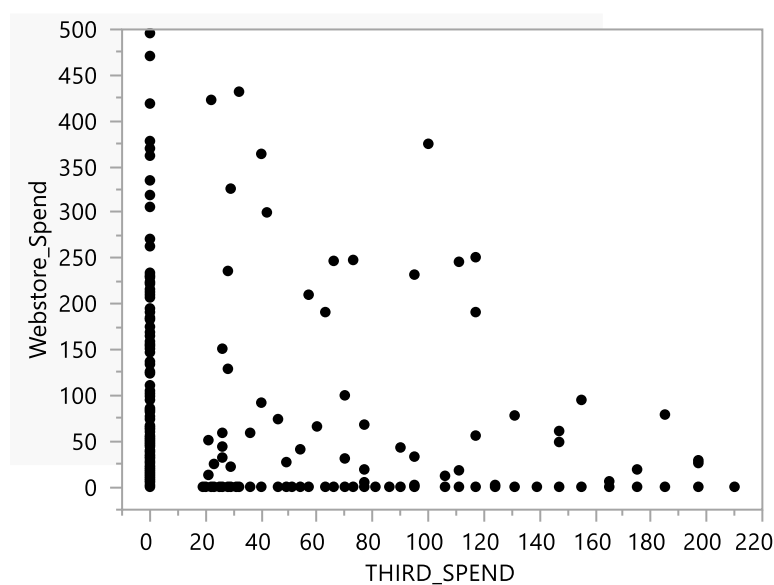
A linear regression for web store spending in comparison to every other applicable variable in the data set was also generated so that correlations might be observed. Many of the graphs contain few meaningful correlations of any kind, such as by city, county, zip code, restaurant visits, age, marital status, and income. However, there are a few correlations that stand out in a significant way. I have gathered the most useful graphs and relevant interpretations below.

Oneway Analysis of Webstore_Spend By state



Web store spending by state shows that there are some states that have significantly less web store spending than others. This doesn't provide many actionable insights since web stores do not require physical establishments based on location, but it can guide targeted advertisement efforts in areas that already show great promise. Several states that seem like excellent candidates for such advertising include NJ, NY, CA, and FL.

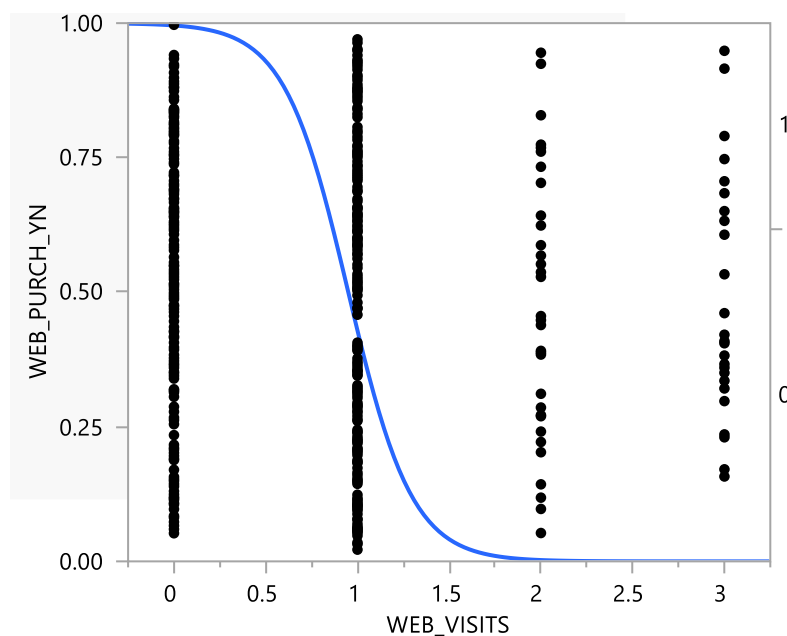
Bivariate Fit of Webstore_Spend By THIRD_SPEND



What is truly interesting to note is that the web store spending very consistently decreases as the customers spend money on third party merchandise. This might be due to third party merchandise being identical to web store products, but there could be other contributing factors. A similar slope can be seen on web store spending by third party visits, though it isn't as clearly defined. This leads me to wonder if web store sales could be increased by reducing or even eliminating third party sales, allowing Bubba Gump to gather data more directly and generate more profit per sale. The nature of third party data mining practices and sales competing with direct sales makes this a difficult question to tackle though.

After the linear graphs regarding web store spending had yielded all the information they could, I turned my attention towards logistical graphs regarding the web store purchased binary. The first logistical graph I want to point out is below.

Logistic Fit of WEB_PURCH_YN By WEB_VISITS

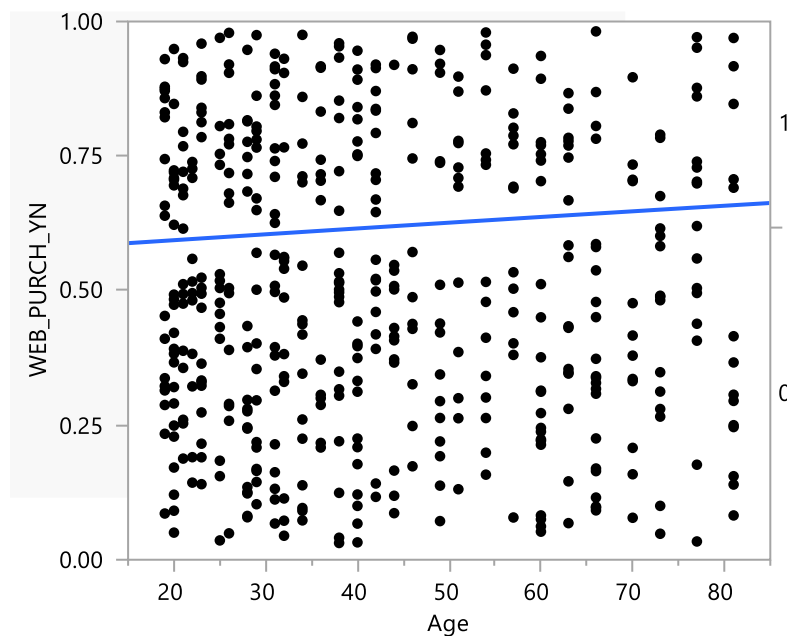


This graph best highlights the counter intuitive nature of the web purchased YN binary data. Those who visited the web store zero times clearly did not make a purchase, and the trendline for the web purchased at zero web store visits is all the

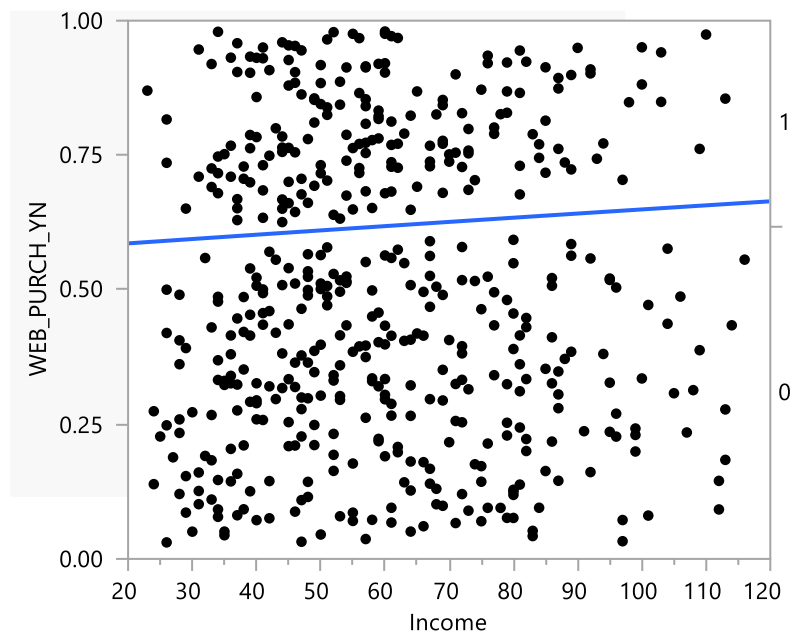
way up at the top by 1. This means that 1 means no purchases were made, and 0 means that at least one purchase was made. This is important to keep in mind for other graphs since it is not an ideal format to be working with and can easily cause confusion.

That being said, we can clearly see that at least most of the customers who visited the web store once made a purchase, and every customer who visited the web store twice made at least one purchase at some point. This means that the web store is highly effective at making sales once the customers actually visit the website, meaning that advertising the web store to increase web store traffic will most certainly increase web store sales. This is by far the best indicator of whether a customer will make a web store purchase or not.

Logistic Fit of WEB_PURCH_YN By Age

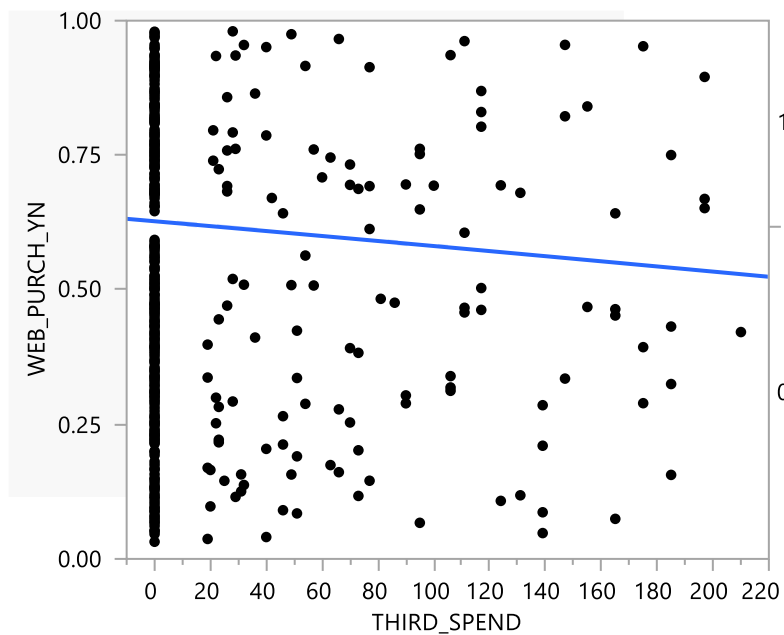


Logistic Fit of WEB_PURCH_YN By Income

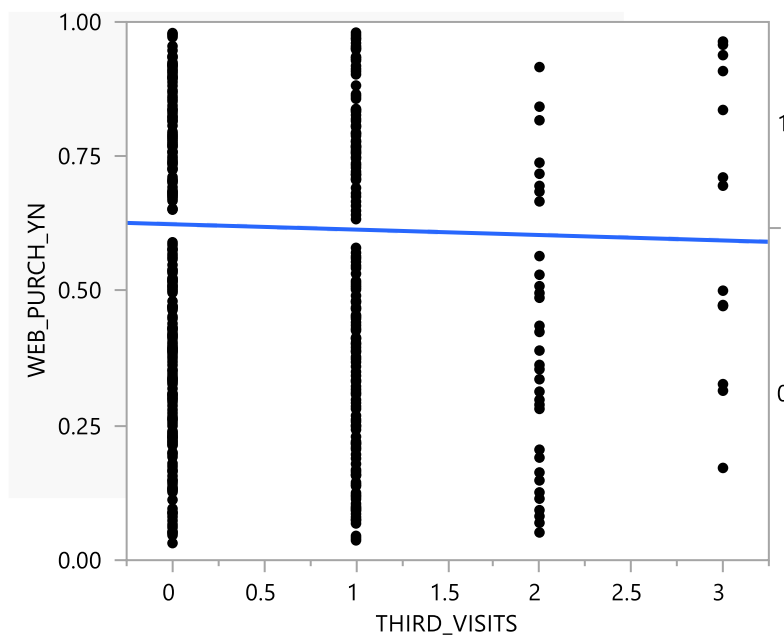


Every other graph had mostly level trendlines, with the strongest slopes showing that younger customers with low income are slightly more likely to make a web store purchase.

Logistic Fit of WEB_PURCH_YN By THIRD_SPEND



Logistic Fit of WEB_PURCH_YN By THIRD_VISITS



An interesting trend to note is that the number of third party visits and spending are the opposite of what we might expect them to be based on prior observations. The likelihood of having made at least one web store purchase actually increases as the customer spends more on third party merchandise and makes more third party visits. This doesn't necessarily disprove our prior observations since this

is simply a single web purchase binary, but it does show that interest in third party merchandise doesn't completely deter customers from web store products. It might even indicate that the web store makes customers more likely to visit and purchase from third party merchants.

Validity

Originally, Validity, Reliability, and Limitations were intended to be a single section. However, each topic requires such thorough elaboration that I decided to break them down into individual sections. As I mentioned earlier, there are several methods of data analysis that can be used to gain actionable business insights. Each method focuses on a unique aspect of the data even though the actual data being viewed might be identical. The purpose of these defined methods is to ensure that every relevant and beneficial perspective be effectively employed in isolation, increasing the clarity of the observations made at each stage. Ensuring that these methods are executed correctly and in order will ensure internal validity.

The types of analyses employed for this project were predictive and moderately statistical. Information regarding the text analysis was already provided in the form of BGSC size and scope. The company decline was the event that required attention, but the nature and origin of the decline was readily apparent, making statistical and diagnostic analysis surrounding that turning point redundant. Statistical analysis was employed when reviewing what is currently happening within BBSC, but given the nature of the situation, it only counts as statistical analysis in the loosest sense. The majority of the analysis was predictive in that current trends were observed so that successful strategies could be amplified and inefficient strategies corrected or eliminated. While the data has been compiled in such a way that a prescriptive analysis seems like a valid description, I don't believe that my report can

be considered as such in all honesty. While relevant text analysis, statistical analysis, and even diagnostic analysis were presented alongside the data set, the nature of the conclusions were shorthand and far from academic or systematic. I was able to do my part to create a predictive analysis with the data provided and format it in a digestible report that offers actionable insights and guidance. However, for a proper prescriptive analysis to be compiled that is properly internally validated, all the other analyses need to be pursued thoroughly so that the prescriptive report can be deemed comprehensive with a confident level of certainty.

External validity is much more difficult to measure, and usually involves taking the internally verified insights and converting them into effective strategies for the real world applications. External validation is less methodically approached since the process has to be tailored to each specific situation. The types of data that the insights can be checked against include environmental and economic factors, as well as cultural shifts and adaptive competitor strategies. I would recommend looking at the economy in general to see what data spikes were specific to BGSC, as well as examining other sea food based restaurants in local areas to examine the behavior of BGSC's target market. Red Lobster as mentioned earlier in this report is a fantastic example to look to.

Reliability

The reliability of this report after validations have been confirmed lies in the quality of the underlying data set. There are some major concerns I have before moving forward with formal business strategy suggestions. While it is safe to draw conclusions about customer age and income for targeted advertisement purposes given that the observed age group fits the expected one based on the movie that propelled Bubba Gump's popularity, I think the certainty stops there. The trends

mentioned above regarding the relationship between web store spending and restaurant spending seems to indicate a strong correlation. However, the lack of outliers indicates that it is entirely possible that customers only began being tracked in the data warehouse after making a restaurant purchase. This would mean that customers who purchased items from the web store before visiting a restaurant are not being listed as web store spenders, making that correlation entirely incorrect. If a customer named “Joe” were to buy a large amount of merchandise online before dining in at a restaurant, the data warehouse this survey was pulled from might not count his previous purchases. Let’s say that first online purchase was \$1,000, and the restaurant bill the week after was \$50. If Joe goes back to the online retailer and spends another \$100 after visiting the restaurant, his row in this data set might only have \$50 spent at the restaurant and \$100 spent on the website represented, even though his actual total for the website is \$1,100. The fact that online customers are excluded from this table at the moment could be indicative of a customer initialization error highly similar to what I am describing. The absence of online-only customers and the types of items being purchased are two issues that I would like to resolve before saying with any amount of certainty what this data means for Bubba Gump.

A more specific concern is the correlation between the zip code values, which suggests that an error has occurred while creating or importing this data set. Most of the values move along a near perfect $x=y$ correlation, but a series of outliers exist that persist in the low zip code 1 values. This seems to indicate that an error with entries beginning with zero has skewed the values, and potentially skewed the zip code 2 values if the additional digit was pulled from that column to fill the five digit capacity of the first column. This zip code error is a serious one that could heavily skew any observations made regarding geographical location. Correcting it is incredibly

important, but it might be causing even more errors than already suggested. Since we know very little about the way that this data was generated other than it was taken from the data warehouse, it is entirely possible that the city and state fields were automatically filled in using the skewed zip code value. This could list a customer as living in a completely different city and state than they actually do, making observations regarding geographical location impossible until that correction has been made.

There are also errors in the data set regarding the naming conventions of the categories and the vagueness of the income values. The inconsistencies therein make me suspect that the survey responses were merged with existing data from an automated database in the hopes of better understanding the context of the customer responses. This is a good strategy, but to be implemented to the fullest, appropriate standardized naming conventions along with clarification on the income rates need to be introduced. After those changes have been made, actionable insights will be easier to glean and of a higher quality than they would be otherwise.

Limitations

The scope of this report is limited by the scope of the data set. This survey data is drawn from among the customers of Bubba Gump, and is taken from willing participants who have visited a restaurant at least once. All of Bubba Gump's data had recently been integrated in a data warehouse. That enterprise data warehouse was built specifically to support data mining initiatives like this one, by consolidating data from multiple operations and channels in one place and integrating the data across sources for a complete view of the customer experience. As a first step, a sample of 500 customers has been selected from the analytics data warehouse and given a survey in exchange for purchase credits at one of Bubba Gump's sales channels. The

survey sample was selected from the universe of customers who have made purchases from at least one Bubba Gump outlet (restaurant, web store, etc.). Responses to various customer satisfaction questions were recorded, and historical purchase information has been extracted from the data warehouse for each customer in the sample. This data set includes information regarding the customer zip code, the amount spent at a restaurant, the number of times the restaurant was visited, whether a purchase has been made on the web store, the amount spent on the web store, the number of times the web store was visited, the amount spent on third party merchandise, the number of visits to third party merchandise distributors, age, marital status, and income. While the findings within this report are made with sound analysis techniques, this portion of customers are not necessarily representative of the majority.

Resulting Decision Influence

I've made recommendations regarding specific correlations within the data set above, and I've also commented on what external data might be included to safeguard against poor decisions. While I have given plenty of information that can be helpful when making decisions, I have yet to weigh in on my personal perception of BGSC's situation and best strategy to move forward with. Rather than leaving you to craft your own business plans to check against external data, I am offering several proposals to consider, evaluate, and do with as you see fit. I'd like to note that these are merely proposals that I would conduct further research around before implementing, and are not strategies that I believe are ready for implementation at this time.

The first decision that I highly recommend is the reduction or elimination of third party merchandise. Third party sales have shown that they decrease web store

sales, making these two business facets in competition with each other. In the modern age, every company needs to have a website. Many restaurants have their menus online and require their customers to scan a QR code to view it. The web store clearly has promise and is highly effective at making sales once visited, which is enough to convince me it is a priority even before considering the critical importance of a strong internet presence in the digital age. If BGSC were to implement QR code menus in their restaurants that also advertised the web store on the online menu page and the QR code display, online sales would certainly see a sharp increase. By eliminating third party sales and increasing web store traffic, BGSC can keep more profits per sale, gather customer data directly, and remain relevant in the modern era.

Visual Evaluation

The data visualizations used in the above report are highly effective considering the limitations of the data set. The backwards nature of the web store purchased binary field is certainly less than ideal, but by thoroughly explaining the poor format before going over the meaning of the data allows the results to be properly interpreted. The scatterplot matrix does take some thought to navigate, but is still relatively intuitive even before my explanations. The bivariate fit graph is extremely easy to understand, as are the linear regressions. The logistical regressions are equally simple at a glance, but grow more complicated upon closer inspection. The general meaning behind each trendline and graph is readily apparent and intuitive to most, and my detailed explanations should have even the slowest reader keeping up in no time. Overall, I am confident that the visual representations are above adequate, and thoroughly conveyed the relationships I wished to highlight.

Next Steps

The best follow up questions to this report include questions regarding the nature of the web store itself. Is it outdated? Does it need to be redesigned? Can the servers handle increased web traffic should it become a massive success? If the web store is to be the new backbone of BGSC, very specific questions about how to update, improve, and maintain it will be necessary. I again recommend examining the approach of competitors in this regard, specifically Red Lobster.

It is also important to ask questions regarding external data and external verification of my observations, but that has already been expanded upon in detail. Questions regarding the merchandise products offered should be explored as well, using the most successful as models for new products and discarding the least successful variants. However, the most important next step to take before making any grand decisions is without a doubt related to the data warehouse and data mining methods. How was this data set merged? Where is the zip code error originating from? What can be done to prevent such errors and improve the data format for future analyses? Answers to these questions might not improve business success directly, but gaining a measure of reliability and certainty for both this and future reports will yield the greatest return on investment over time by a massive margin.