Douglas Wiggins

DAT 220: Final Project

***Business Problem:***

Based on the scenario that was presented to me, there are a few problems that Bubba Gump Shrimp that can be attainable and can be solved. From the analysis, Bubba Gump Shrimp Co. gained its success with help of a blockbuster movie. Sales was growing at a high rate to the point that the company were able to expand into various locations. However, a few years later after the surges, the company’s sales declined within 2 years. The reason for this drop is because now that customers have experience of what Bubba Grump Shrimp has to offer, they have moved on to different things, making slower rates on its sales and ultimately leveled off.

***Analytic Method:***

Despite these setbacks however, Bubba Gump Shrimp Co. manage to collect large amounts of data through the years that contains the business point-of-sale (POS), web channel sales performance, its customer loyalty programs, and sales transaction data from its website and the retail partners. By having all of this data on-hand, we can plan and find the problem that causes the company to lose sales come up with a solution that can restructure its means of sales and customer growth over a long period of time. Since the data is integrated in the data warehouse, we can use that data from different areas and pinpoint the source for data that greatly affected the customer experience and can be heavily monitor and modify whatever the situation arises. By implementing consolidation of its data from the data warehouse, we will be able to link every sales transaction to very specific customers from the business restaurants. Moreover, this will give Bubba Gump a better understanding of what the customer wants and formulate new ideas and plans to bring past and presents back into its restaurants. Another good plan is to create data marts within the data warehouse. This will allow Bubba Gump to be more focus on the data that they were struggling the most and focus on how to better the customer experience as well as bringing more revenue within the business.

***Analysis Tools:***

The tool that would most likely be implemental and crucial for this problem is the software program JMP. This tool is essential to our success for a variety of reason pertain to data mining. JMP is the perfect tool for various algorithms that can be constructed from the data that is given and provides continuity within other software programs such as *Microsoft Excel* and for MacOS users’, *Numbers*. Moreover, what JMP has within its software is dynamic data visualization, which is effect when looking at its displays, it’s easy to import any sort of data of any business, displays any necessary analysis that is given then create those visualizations at the end result.

***Data Visualizations:***

The data visualization of this project is going to permit Bubba Gump management to make important decisions and draw conclusions that is based on the conclusions and analysis that is handed to them. Visualizations must be important for the business to understand easily and be able to spread the information given to them to various locations of the business in a clear and concise way. One way that can be help can depend on the customer clusters which can be segmented. Showing the customer clusters that are segmented is a good way to looks at the details or the basis of what the characteristics of a customer are such as its age, or whether the customer is married or single. Other useful visualizations can be scatterplots, which can be useful to show the correlation between variables, A geographical heat map that is useful to see the clusters that is segmented by its location. As long as management can understand their own data visually, any tools of visualizations can be useful.

***Research Question:***

The research question that has to be asked for this project is as follows: based on where Bubba Gump is located, what can we do to find customer clusters that spends the most money? The location of which sales the most is key so that way, the spending can divide itself into other important factors such as adding hotspot locations, updating the online store, and better options with 3rd party vendors.

***Research Management:***

There are a couple of steps when dealing with customer clusters and how it can be grouped and segmented. We have to find how certain segments of data that spends the most on its products, which in this case is Bubba Gump. We can measure this by finding that customer segment with increased spending. This can be a source as to how to create more revenue within all locations. Other research that can be measure is locating certain customer clusters that is worth segmenting which could be a more effective way to start.

***Follow-up Questions:***

* ***Can we assume that the geographical location can affect the customer spending?***
* ***Based on the restaurant location, which spends the least amount of money?***
* ***Is having subscriptions can increase customer awareness?***
* ***How does the location and age affect how much a customer is spending?***

***Research and Support:***

There are a couple of resources that utilize data mining in the restaurant industry and how data mining helps once struggling restaurant businesses. One of the sources I have read is an article in the New York Times “To survive rough times, Restaurants turns into Data-Mining” (Stabiner, 2017) summarizes how the implementation of data-mining helped restaurant businesses and the tools that help certain businesses get back on their feet. Another article “Mining Restaurant Data: Know your Customer” (Kasavana, PH. D, 2010) not only summarize how data mining is used in the restaurant industry, but also shows various tools and strategies that is implemented such as developing customer relationship management, customer profiling and forecaster. This is essentials as to how to develop a plan for Bubba Gump and its management. Moreover, there is also a JMP learning library that displays various techniques and tools for visualization that we can use in the JMP program.

***Analysis Organization:***

When it comes to cluster analysis, this method is used to divide a population which in this case Bubba Gump’s customers, into various groups that have similar attributes. How this was done is by displaying certain characteristics of the customer such as the age, sex, income, marital status and his or her spend rate. Moreover, there are very distant customer clusters within all of the characteristics of data that was available. How there are segmented customer groups are formed is by having the population of that amount being spent. The results show that customers spend more online whenever they make a purchase on their website or in this case its web store. The results from the data can be implemented as a tool to target certain customer groups that could potentially have a higher return when it comes to its marketing dollars. There is a con to this analysis unfortunately for this only as a single variable

Another form of data model was used is linear regression. Linear regression method was used for bubba gump to predict and determine the web store spend data. This model was useful in parts of evaluating different characteristics of the customer that affects how the amount is being spent.

***Sources of Error:***

There were errors that pertaining to obtaining to data was putting the columns on the data set in the JMP data set. This error was fixed by remove this data from the data set. Doing this method will enable all of the rows to be checked thoroughly before it given a full analysis. Another error is from where the zip code is incorrected specifically the first 2 digits of zip code was missing. This is a crucial error because this erects the entire data set where the leading zeros are removed. However, I corrected this issue by added leading zeroes to all zip codes that were only four digits, changed the datatype to character, and recalculated the first two digits. Other errors can be from the customer standpoint whereas the customer can have errors will filling out the surveys which in case affect the dataset.

***Meaningful Patterns:***

The most meaningful yet common pattern that is very noticeable is that the customers that spend at the restaurant also spend more on the website when they make a purchase. This results in additional questions: can we also identify this customer group and target its sales? What causes the customer to spend a large amount of money at its web stores and restaurants?

***Inaccurate Depictions of Data:***

One data that has an inaccurate depiction of data is the plots that was generated from the correlation matrix and regression analysis. Because there were many zero-value data when it comes to spending amounts, the plots were difficult to understand. I realize however, that once I exclude the zero values form the analysis, the plot becomes more easier to read.

***Alternative Analytic Methods:***

Other alternative analytic method can be the use of hierarchical and k-means clustering, decision trees, and regression methods for instance stepwise and partial least squares. These methods can be useful, but it only applies to the necessary analysis. For instance, k-means clustering is essential when dealing with various variables as well as producing tighter cluster, but it can be difficult to predict the amount of numbers of clusters and the initial seeds will have a strong impact on its final results. Another method that can be implemented is to segment the customers that is based on its demographic data that isn’t included in the data set but instead in its location.

***Display and Interpretation:***

The following charts below shows different comparisons of customers that actually completed the essay for Bubba Gump. By comparing the age and income of the customers and how the customer spends an item at its webstore, we can see that in the first chart, customers with an income between 30k-65k didn’t spend any money from the webstore. Moreover, from the first chart, the customers who spend money at the webstore are customer that are within the average income. As far as the age in the second chart, specifically ages 20-35 however, there were some customers that bought a good number of items from the webstore. From this analysis, Bubba Gump’s plan of success has to start by focusing and targeting a younger audience as well as finding ways of keeping this group from returning to the restaurants and its webstore.

A screenshot of a social media post

Description automatically generated

A screenshot of a cell phone

Description automatically generated

***Validity, Reliability and Limitations***

By proving the validity, the regression chart below shows that based on the survey, the demographic that responded the most from the survey is the younger generation that is within the average income. On the other hand, there is a decline from the chart that shows that people who are older and make more money didn’t respond much to the survey. Moreover, based from the logistic regression chart, it shows that most customers didn’t visit the webstore after they completed the survey and the decline drop dramatically. Therefore, with these graphs we can determine the main problem that Bubba Gump has is the loss of profits.

A screenshot of a social media post

Description automatically generated

***Resulting Decision influence***

From the resulting data that was gather, I believe the source of Bubba Gump’s issue is the lack of steady demographics within the business. I believe that in order for Bubba Gump to have a long-term success, they will have to draw in a younger demographic specifically between from ages 18-35 and it’s within the average income. One way that will draw this group is develop an incentive program such as loyalty and rewards. Doing this will always enable the group to return from the restraint which will stabilize Bubba Gump’s profits.

***Visual Evaluation***

Throughout this project, I have seen various forms of data and charts that should be visually comprehensible. For instance, based on all of the charts and data that was available from the survey, is considered to be the color red and blue. The red plotted dots represent how many people didn’t purchased an item from the webstore. The blue plotted dots indicate customers that visits and purchase anything from the webstore. Given that I believe that this the visual from the charts is clear and concise for Bubba Gump’s upper management to interpret.

***Next Steps***

Based on all of the data and research that is given for this project, for me to move forward with a plan for Bubba Gump to gain income, I would first recommend the company to increase their revenue. Moreover, I believe that Bubba Gump should mine and gather more data from the restaurant end, since that’s where they make the more revenue. Also, find more way to promote the company to gain more customers such as merchandising and a more enhance way to brand and add merchandising in the idea.