DAT 220 Final Project

June 26th, 2021

Bubba Gump Project

Christopher Haerle

**Business Problem:**

The Bubba Gump Shrimp company started off as a successful business. The company gained popularity from the popular movie “Forest Gump”. The company expanded to at least 45 different locations. As time went on the company started going through rough times and had to close some locations possibly due to this. Bubba gump took notice in this and started collecting their own data to try to figure out why some branches were not thriving as well as they once were. Some data points that they collected were point-of-sale, web channel sales performance, customer feed back through loyalty points, and sales transactions through their website and third-party vendors. Using the analytic method would help show customers segments and clusters, which will help show the patterns that are happening. Data warehouse would be helpful to show data patterns, trends and subgroups.

**Analytic Method:**

Bubba gump should try to identify the market. Identifying the market will help them start making better business decisions that will help them address their issues. They should understand the assets and manage their customers. The goal for Bubba gump shrimp is to entice new customers to want to come in and spend money, while at the same time showing their reoccurring customers that they are grateful for them to stick by them. One analytic tool that would be beneficial for Bubba Gump to consider using would be Azure. Azure is a business suite analytic tool that connects hundreds of data sources, data prep, and it drives ad-hoc analysis. The reason why I thought that this tool was a great choice for Bubba Gump is because, it is a low-cost BI solution tool. Azure is very affordable for a large company such as Bubba Gump, it is only $9.99 a month, they also have a pay as you go option to where you are only paying for the features that you are using. Companies can simply and affordably pay for tools that they use instead of a subscription and have a bunch of different tools that will never be touched.

**Research Question:**

What is the specific research question that needs to be addressed? What research question will you work from in order to analyze the given data for meaningful patterns?

The research questions that I feel that needs to be addressed is specifically how can we understand and identify natural clustering in customer population? If we can understand the high-density places where Bubba Gump shrimp see’s more customer’s then we can get a better understanding of the age group and income.

**Research Measurement:**

How will you determine if your research question was answered or if your hypothesis-generation was successful? How will you measure progress?

How I will determine whether the research question is answered is by identifying all the trends of customers that shop at Bubba Gump through data visualizations using metrics, such as Restaurant, age, and income.

**Follow-up Questions:**

What are cogent follow-up questions or explorations that should follow from your initial research?

A cogent follow-up question or explorations that should be followed from my initial research are, could this information really help increase sales and yes it can because it targets customers to analyze and monetize the gathered data.

**Research and Support:**

Are there any published sources or other resources that address your line of inquiry? Where do they fall short? How will they help guide your analysis?

Unfortunately, I have not found any sources that support my claims.

**Analysis Organization:**

To what extent does your analysis reflect an organized, stepwise approach? What aspects are beyond your control?

How I will determine if the research question was answered is by identifying all the shopping trends by specific age groups. The metrics that I will use are through data visualizations, using restaurants, age, and income.

**Source of Errors:**

When I initially started reviewing the data from the Bubba Gump data, I saw a few errors and some missing data. After selecting columns of data so that I could get a better idea of what the information was trying to tell me, I had to narrow down which columns to use. The reason why I had to keep selecting different columns was because some columns did not get us to much useful data that did not tell me much. When I used linear and logistic regression with the columns that I found that showed useful data it helped me understand what the data was showing. After running the logistic regression with restaurants, age, and income I was able to see which restaurants were performing better with a certain age group. People who are in their 20’s – 30’s showed that they spend more money. Those who are older they did not really come to the restaurants.

**Meaningful Patterns:**

One pattern that I noticed is that younger people visited the restaurants more and they were spending more money than older people. I saw this pattern from the data settings from Bubba Gump’s survey. When I added other data points to compared with, I noticed the same pattern. Younger people are the ones spending more money at the restaurants. They have an average income of 30K-60K. After reviewing the data more I noticed that the average return to the restaurants decreased. People are willing to try it once and then they hardly revisit the store. There should be a survey conducted to figure out why they do not want to come back. Is it the food? The atmosphere? The service?

**Inaccurate Depiction of Data:**

While using the data mining tools, I noticed some outliners. There were some points that were showing some big spending with younger people. I am not sure what caused the spike in the data, but it is something that should be reviewed further. It could be as simple as they were over served and spent their money gregariously, if not then why did they spend so much money?

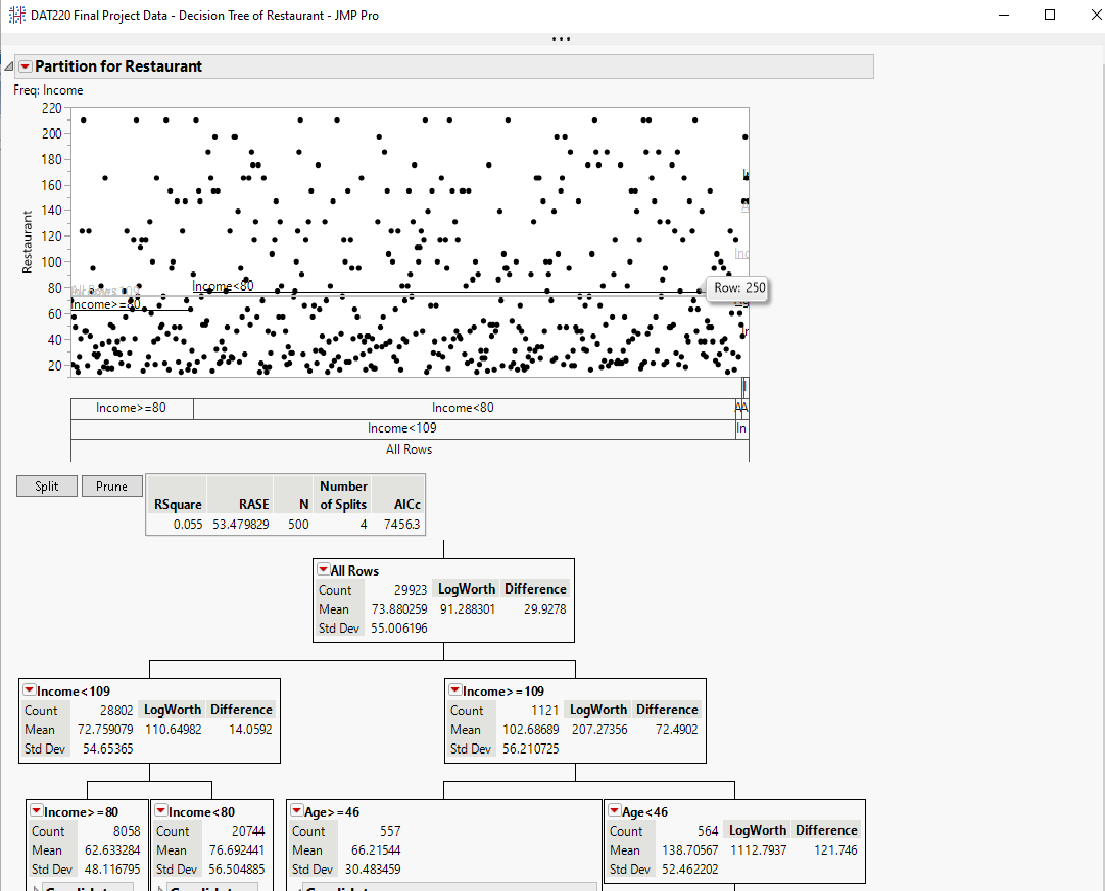
**Alternative Analytic Methods:**

One method that I feel that would be a good alternative analytical method is machine learning method. It seems that the data can be inputted into an algorithm and the machine can see the trends and what would be ideal fixes for Bubba Gump to move forward with.

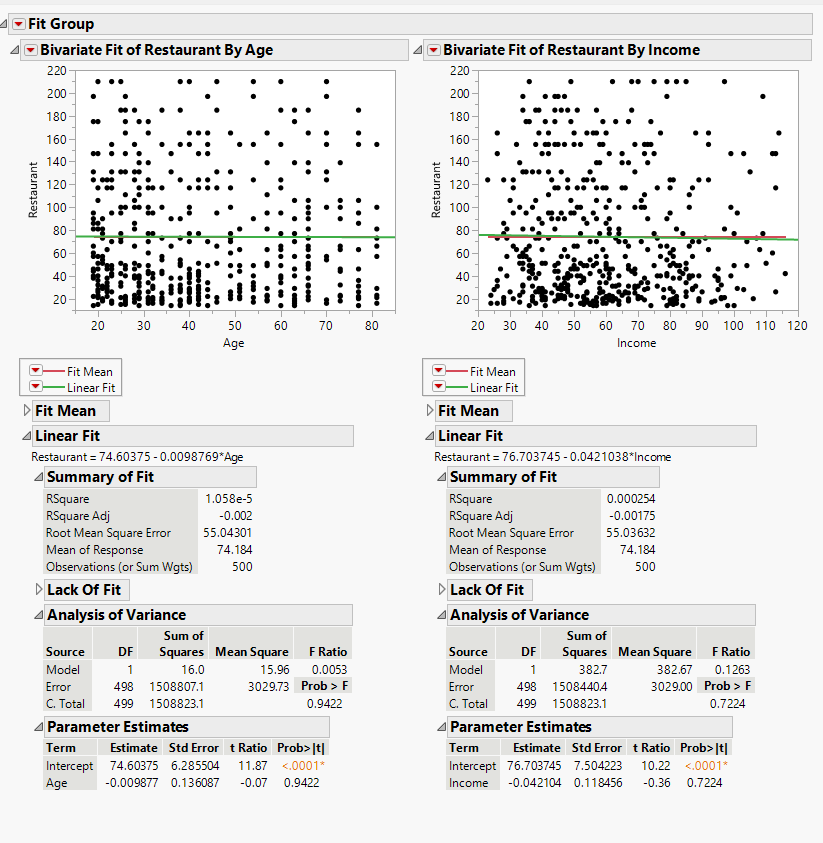
**Background and Introduction:**

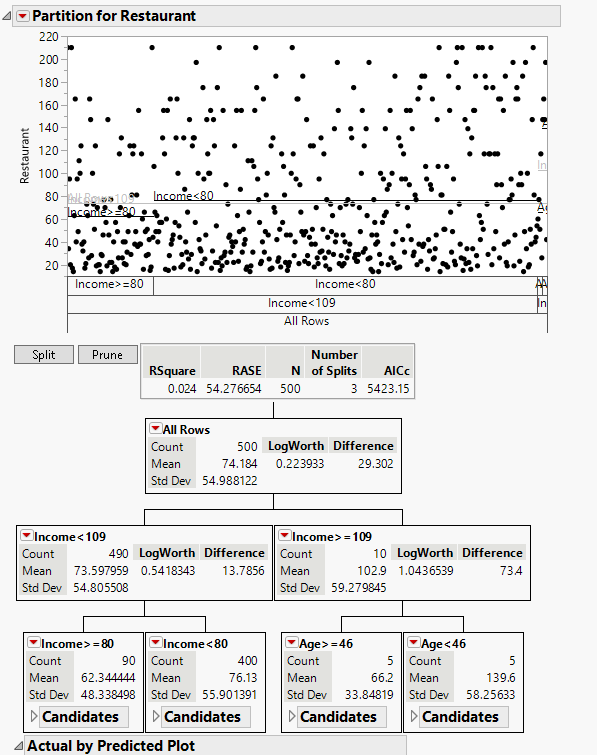
The problem that I am trying to solve is what age group and their income that visit which stores regularly so that we can start making it more appealing to them and try to get them to come to the restaurants more frequently. The results that I was trying to display was the average income, age by what restaurant is that they visited the most.

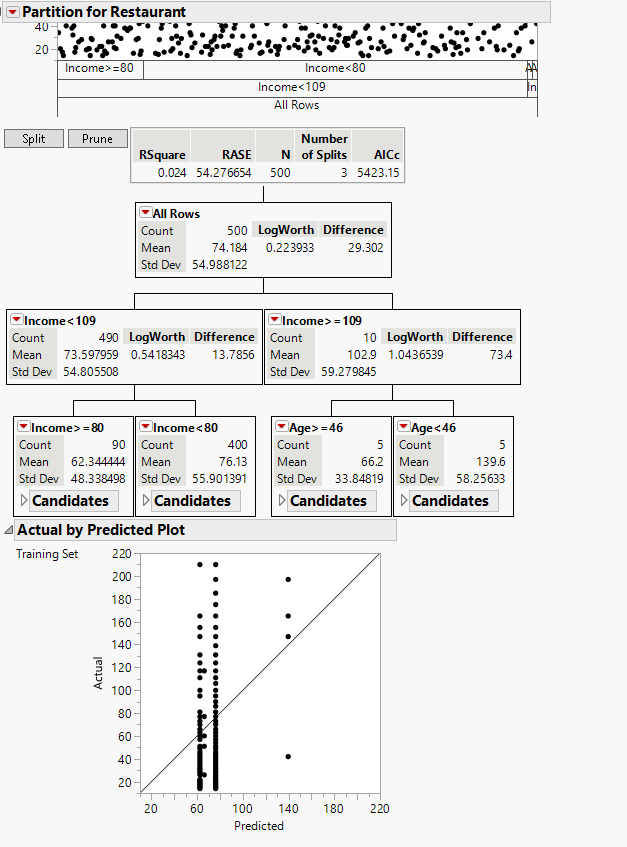
Restaurnts\_Age\_by income



Restaurant by Age: Restaurant by Income:







**Validity, Reliability, Limitations:**

The validity of the reports show that the average income and age for people who visit these restaurants have an average income of lower than 80k and they are younger than 46 years old. This is validated between the three different reports that were ran. These results are reliable because I ran them 3 different ways and they show that the group of people that Bubba Gump should focus on are those younger than 46 years old and that they make less than 80k. These results do have some limitations, the limitations are that I do not show if any of these people are married. I think this is another piece of information that should be added to the results so that we can see if the people coming to the restaurants are single or if they are more family oriented. That way Bubba Gump can make their locations more inviting.

**Resulting Decision Influence:**

The way I would show these results to a potential client or a superior, I would show them all three graphs to show that the results all show similar results. I would explain to them that if we were able to start having the restaurants that have this foot traffic then we should start modifying the menu and some of the seating to make the stores more inviting to them.

**Next Step:**

The next step that I would like to include in my searches is those that are married. I would like to start adding that because it will really help identify if the people visiting the restaurant are either planning on starting a family soon or have already started a family. I feel this information would really help Bubba Gump start making their business more inviting to those people so that they will start coming back more and more.