David Bell

February 28, 2020

DAT 220

Final

**Business Problem**

In 1994, the movie Forrest Gump premiered. It garnered massive praise and popularity and won many awards. The movie was a household name in America and a favorite among all generations young and old. The Bubba Gump restaurant brand was able to monopolize on the movies popularity. The exposure catapulted it into new markets. After a few years of growth, the sales began to plateau and eventually slow. Now sales have declined at a steady rate and the business wants to use its collected data to find ways to increase its profits.

**Analytic Method**

I will use the JMP tool to assist me with clustering the data. With the use of clustering I will be able to identify groups that Bubba Gump’s restaurants attracts. The data I will use to populate my tables will come from the data warehouse that the Bubba Gump Brand has provided. This data will come from the restaurants point of sale, customer information via loyalty programs, the web store transactions and all data provided by third party sellers on customers and sales. I can review certain trends by comparing this data. With this data, I can find sales from different data marts to see things like customers restaurant purchases and their web purchases and how they compare. Are markets different for customers at a certain restaurant location in what they order? Do these customers in a certain demographic have buying habits that are different or similar to customers in another region? By comparing all the data together via the warehouse, I can find some helpful trends that could help the company increase revenue.

To mine the data from the Bubba Gump Company I will be using the JMP data analysis software. I can use the software to find relationships in the raw data that can display some trends on why the company has lost sales and show trends that can help increase these sales. The software gives me the ability to analyze the data by row or column and break it apart into sections for comparison.

For my visualizations, I will be using tables and graphs. These visualizations will help present a clear understanding of the comparisons and relationships in the data. I can also use these particular graphs and tables to build upon to find other relationships or trends that may exist. Using these visualizations simply makes the data more accessible and easier to understand for the company managers. The research question I would ask is how we can get customers to come back and what customer trends seem to have the best results for repeat business. Do we see trends from certain branches of the restaurants that have more repeat customers than other locations? We need to see why a customer is or is not returning.

If I find a proper hypothesis on why or how we can create more customer return or loyalty then it would need to be tested in the field. We would need the company to take this hypothesis and test it to see if there are results either negative or positive. If we see a trend where customers do return based on certain stipulations then we can move to a more concrete plan to initiate that change. What else can be done to increase customer retention and how can we transition to that plan effectively?

**Research and Support**

I have been looking at the Cracker Barrel business model to see what I can find about their growth. The Cracker Barrel model is very similar to Bubba Gump in that it is a food chain with an online store. The large difference here is that Bubba Gumps popularity was really helped by the movie. Cracker Barrels model makes use of the American Highway system. The advertising and word of mouth by using American vacationers and business travelers is a large boost to its sales. The cracker barrel has a fantastic system of placing its merchandise in the waiting area. The merchandise has become part of the restaurant experience, which was expanded by its online store. The online store has the same offerings as the physical restaurant with many additional items as well. Cracker Barrel has stated that their goal is customer loyalty so we should look into which Bubba Gump restaurant highlights that same loyalty.

**ANALYSIS ORGANIZATION**

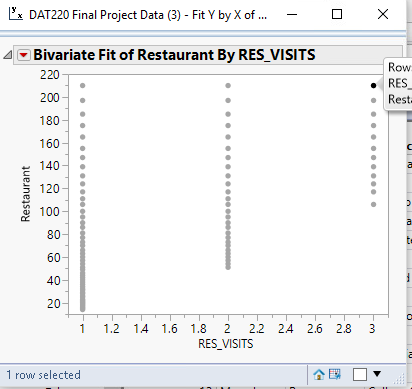
When I went to organize and compare the data from the Bubba Gump Company I used a few different analysis methods. The first method was Descriptive with the use of histograms to organize and correlate the data to attempt to find any trends. I set certain data points against another to see if any obvious trends could be found. While using the descriptive method I found obvious missing data, which could highly effect the outcome. I also used the unsupervised analysis method by creating hierarchical clusters and K-means clusters. My final analysis was supervised learning which was represented by linear regression, which would give me a continuous variable prediction, and a logistic regression, which gives me a binary answer in yes or no or true or false.

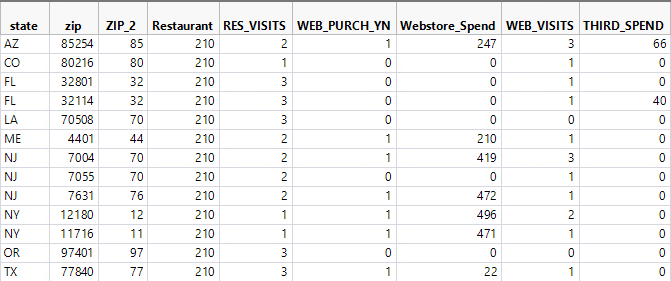
**SOURCES OF ERROR**

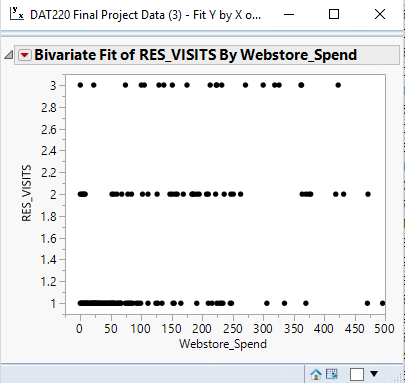
There were certain variables that I did not take advantage of because I did not find any value in them because they did not seem to matter. The second zip code was one that most likely did not need to be imported to the data warehouse. I also found that some of the data was missing that could have been useful such as what each customer spent on successive web store visits. The web store visits only listed the total spend at the store and how much they spent on the third visit. It is also possible that the income listed for those married individuals could have been the income of the entire family and not just of the individual customer. This could create large discrepancies with the outcome. I did find it easier to cluster the ages into ranges to find a clearer picture of how many of what age we had in the data. I also was more inclined to use the restaurant number attribute over the other location points because the customers are traveling to this restaurant and mattered more that they came to a certain restaurant then the county the restaurant was in. Now I could have probably dived a little deeper to see if certain customers had restaurants in their own county but were choosing to drive to another to visit the store. That could have been telling.

**MEANINGFUL PATTERNS**

I wanted to look at restaurants first that had a high amount of sales. I found that restaurant 210 not only had a large amount of sales it also had good web store purchase sales for visits under three. I also noticed that the restaurant visits were from people from all over the country, which was not the largest surprise since this restaurant chain is more or less a tourist attraction to some customers. I also noticed that well over half of the customers in general spent some money on the web store if they had a visit to the restaurant. The interesting fact I would like to find if true or not is that the customers have to give their information to the physical restaurant first before being included in the data? If the restaurant physical visit had to come first then maybe the attraction of the physical restaurant is driving sales to the web store. Is there advertising for the web store for people currently in the restaurant at the same time? Are people accessing the web store while dining? I would not mind having a time stamp on restaurant sales and web store sales. That could be a pattern worth diving into.







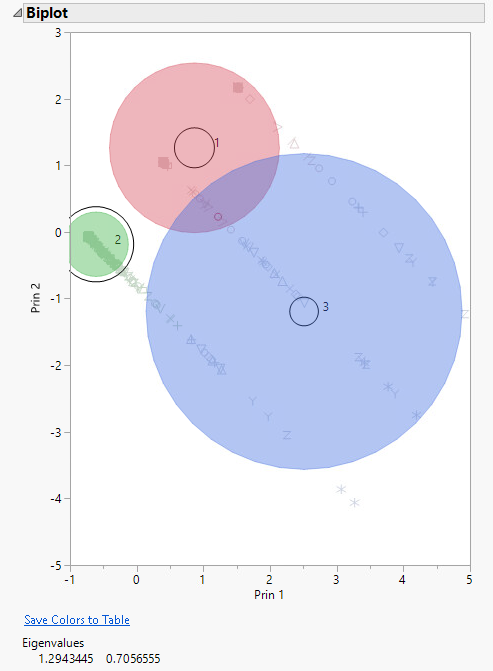
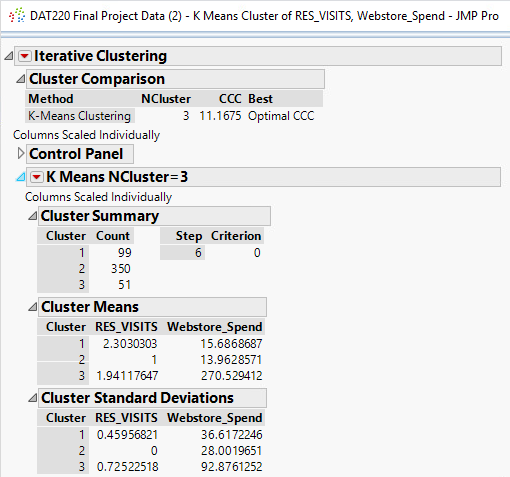
**INNACURATE DEPICTIONS OF DATA**

As I stated above the data that has been presented could be inaccurate because the income of the customer could represent the income of the family and not an individual. The money the customer is spending could be more significant if we have a better understanding of what value their income holds. If we have a low-income family that spends money at a restaurants website or we perceive someone to have higher income then they do because they represent a household then that will change how we approach the data.

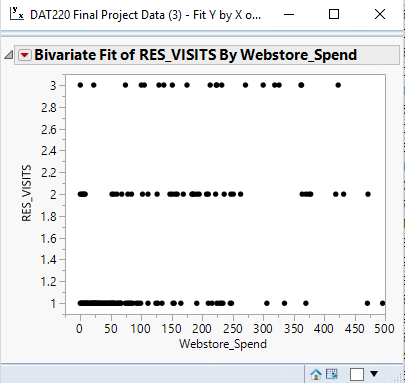
**ALTERNATIVE ANALYTIC METHODS**

A nice analytic method that I have used in the past is a decision tree. A visual predictor of a possible path that a customer could take once they enter the system depending on their variables. They outcome could become very predictable once the variables can be correlated and possibly logistic regression can be accessed to view with the tree. If a customer holds these qualities then the decision tree can predict an outcome that would benefit the company. Certain customers would hold more value than others would if we could find which variables are the most meaningful. The decision tree could set up several outcomes that we could analyze.

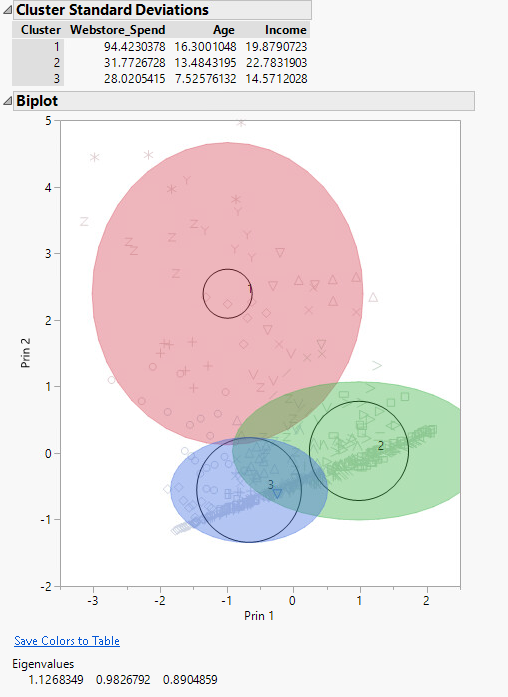
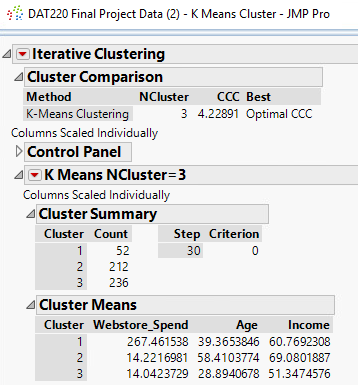
My first k-means cluster was looking at the restaurant visits compared to the web store spending. What I found at first was that the majority of restaurant visitors spend the most on the web store only after they had just over one visit.

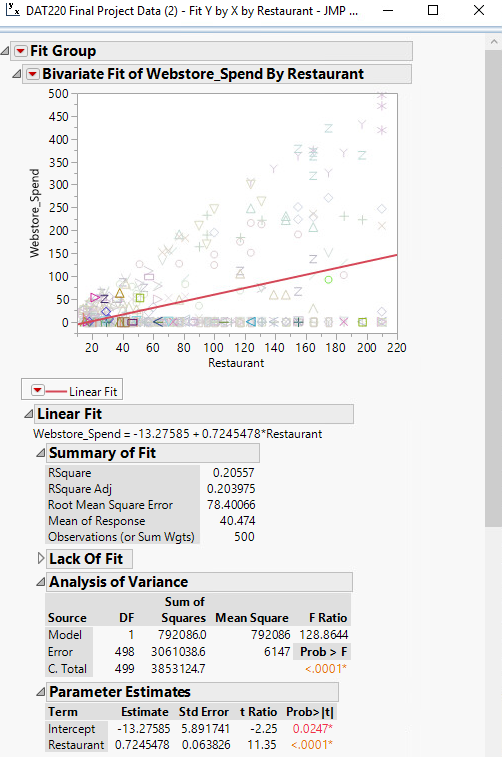


I also ran a bivariate analysis on restaurant visits in comparison to web store spending. A few different stores in certain states show a nice increase in web store spending when the customers visit the restaurant a few times.



I created a second cluster using the web store spend, age and income as the variables. The demographic of people in the middle of the income range as well as the age range spent more at the web store.





The strength of my RSquare result is low. This does not mean that the relationship is bad. It just means that my findings do not apply to every restaurant. This goes to show that the validity of my results are not validated but also are not destroyed by the RSquare result when I look at this model. It is a fact from the data that some restaurants successfully push online merchandise with a higher visitation rate in the restaurant. This of course does mean that I am relying on the data provided to push for this campaign. I would suggest as a data analyst that the data be double-checked and collected carefully. To validate my findings I need accurate data. The limitations of this data is that there is only 500 customers which most likely only counts for a fraction of the visitors every year to the Bubba Gump restaurant chain. I would push for a more thorough data collection system. Offer a chance to win a prize is the data is filled out and submitted.

From these results, I would like to collect more data on the restaurants in the states that have high web sales. I would like to see what is happening in the locations that are giving them high sales numbers with their return customers.

I would personally like to display the individual states that I chose and show a much smaller comparison between those states in a cluster or at least create a cluster that shows the states bunched with high sales and visits a bit more clearly. I believe this could help the company see some of their standouts.

A new hypothesis could be formed once I have more data from the restaurants that I found to be successful from my research. Is the income in these areas higher than others are? Is the standard of living effect the staff at these restaurants to where they perform better or is the management more professional? Are these locations more sought after for tourists?

**REFERENCES:**

Jones, D. (2019, January 21). Cracker Barrel: A Great Business, But It Could Be Better. Retrieved from <https://seekingalpha.com/article/4234477-cracker-barrel-great-business-be-better>

(n.d.). Retrieved from <https://shop.crackerbarrel.com/?_ga=2.248612039.1561139474.1582895757-1306788279.1582895757&_gac=1.95256046.1582895757.Cj0KCQiAkePyBRCEARIsAMy5SctjIwOnVOq4yD4IfYwWIZsuc9UWDQGw1r6Y4VXmBK6PAKEQL-zC69waAtkUEALw_wcB>