7-2 Final Project Milestone Four

Joel Meza

Professor Torre Roenne

April 17, 2021

DAT 220 -Fundamentals of Data Mining

Southern New Hampshire University

**Display and Interpretation**

|  | **zip** | **Restaurant** | **RES\_VISITS** | **Webstore\_Spend** | **WEB\_VISITS** | **THIRD\_SPEND** | **THIRD\_VISITS** | **Age** | **Income** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| zip | 1.0000 | -0.0602 | -0.0906 | -0.1043 | -0.0928 | -0.0053 | 0.0465 | -0.0522 | 0.0085 |
| Restaurant | -0.0602 | 1.0000 | 0.5955 | 0.4534 | 0.2078 | -0.0427 | -0.0379 | -0.0033 | -0.0159 |
| RES\_VISITS | -0.0906 | 0.5955 | 1.0000 | 0.2943 | 0.1839 | -0.0801 | -0.0846 | 0.0045 | -0.0307 |
| Webstore\_Spend | -0.1043 | 0.4534 | 0.2943 | 1.0000 | 0.6119 | -0.0059 | -0.0034 | -0.0368 | -0.0299 |
| WEB\_VISITS | -0.0928 | 0.2078 | 0.1839 | 0.6119 | 1.0000 | -0.0409 | -0.0103 | -0.0037 | 0.0301 |
| THIRD\_SPEND | -0.0053 | -0.0427 | -0.0801 | -0.0059 | -0.0409 | 1.0000 | 0.7422 | -0.0827 | -0.0601 |
| THIRD\_VISITS | 0.0465 | -0.0379 | -0.0846 | -0.0034 | -0.0103 | 0.7422 | 1.0000 | -0.0768 | -0.0636 |
| Age | -0.0522 | -0.0033 | 0.0045 | -0.0368 | -0.0037 | -0.0827 | -0.0768 | 1.0000 | 0.1093 |
| Income | 0.0085 | -0.0159 | -0.0307 | -0.0299 | 0.0301 | -0.0601 | -0.0636 | 0.1093 | 1.0000 |

From the first step into my visualization from the given information was to allocate the table shown above while it precisely shows the highlights correlation along with the client’s information. Upon building this table, I found that this data is used for variables that are continuous which make a great fit into the multivariate analysis sort. Also, this table has helped discover which of the variables that are relevant to each other where it clearly leads me to the path to understand how the sales in Bubba Gump company have decrease over the years. Although, I was capable to find the relevant information from each category in restaurant visits, restaurant spend, webstore visits, and webstore spend where I could start to emphasis on the highlighted data given from the chart table above.

**Scatterplot Matrix**

Diagram

Description automatically generated

**Visual Evaluation**

On this table shown above, I was capable of discovering in the correlation table to focus down on the variables which are given from data within the scatterplot matrix table. This continues the process of information to gather along with the particular continuous variables listed as shown above.

****

**Resulting Decision Influence**

From this table above, I am capable to demonstrate a relevant relation from the continuous variables in a particular analysis model. The scatterplot correlation shows the best representation where the regression line specifies how clients are spending more in a Bubba Gump Company location that consumes more by spending online site rather spending inside a restaurant. Although, we know that if a client interest is in restaurant shopping, then it is more likely to spend more sales online shopping due to their use pattern of interest.



The above graph shows the regression logistic model which is supportive tool to calculate the future spending of Bubba Gump’s company existing clients’ information. From the scatter plot table specifies clients which have diverse visits online rather than the restaurant visit that may likely be attracted to purchase online.

**Validity, Reliability, Limitations**

Prior to working on the client information given, I will continue to categorize what the ideal points which I will strive to emphasis on. At first, I will use the correlation from the restaurant spend and the total spent in Bubba Gump web store. Another point would be the relevance in the number of online visits and the possibility of successful purchase transactions. From this correlation its observed from all the diverse methods that are applied on analyzing the client’s information given statistically.

Prior to validity and reliability given the information sets, there are a margin errors to think thoroughly before seeing actual outcome results. The information accumulated was only created upon client’s feedbacks that are given in assessment reviews from there point of interest views. However, there is some type of information that did not have quite enough data to emphasis in, but I continue looking for relevant information to continue my analysis and its validity from this project account.

**Next Steps**

For my next steps I will continue support my growth of sales from the Bubba Gump company to finally achieve the times or number of visits between the restaurant and the web online from its history. Prior to my project account here I most likely pursue the client future transactions in online purchases growth total where they have visited either online or in person at the restaurant. If I could find more relevant feedback, then I would start by extending my voice to our clients who made in the past huge amount cost in purchases to find what the interest was from the product of Bubba Gump company they were attracted. While we know as mentioned previously how clients could spend much in the restaurant, then they will proceed to make more purchases online.

Therefore, I find another possibility in reviews is to count on the geographic location which serves clients most often. In example, if there are big groups in online transactions by clients in a precise location, then Bubba Gump company will count on opening a restaurant close by that location. Within this data given, it’s very important to find potential sales with great sums of purchases to priorly open a new restaurant to receive a return of great investment. Overall, I do appreciate the work with Bubba Gump cliental information given for access to investigate and optimize revenue in sales which have been impacted in the past years while continuing to meticulously analysis, emphasis, and account in this project.