4-2 Exercise 4: Unsupervised Learning with Bubba Gump Data

Joel Meza

Professor Torre Roenne

March 27, 2021

DAT 220 -Fundamentals of Data Mining

Southern New Hampshire University

**Survey Information Source**

This is populated from five hundred clients that were used for the Bubba Gump’s database biased on the purchases from online webstores, in door restaurants, and its subcontracted third-party vendors. From each given client, there’s a precise history from the purchases they have made previously which the data given is used up from the Bubba Gump’s database. From this table chart, the data is stored into variables which are first name, last name, location, restaurants visits, webstore online spending, subcontracted third-party spending, age, income, and martial status.

**Visualization of Information Groups**

**Distribution – Restaurants Visits**

**Graphical user interface

Description automatically generated**

Given in this illustration above, it shows how the three times the restaurants have been visited from all five hundred clients that have visited this location at minimum once. However, this reflects only on the clients that were selected in the restaurant location and only have visited, not in other places like online webstore.

**Webstore Visits – Distribution**

**Graphical user interface

Description automatically generated**

This is similar to the restaurants visits distribution where the data given is irrelevant to the webstore visits distribution. This webstore is available to adults that have access to the internet to go in. This setup is helpful for clients that are not close to the restaurant or want to order to go while the website is available for anyone to have access.

**Subcontracted Third-Party Visits Distribution**

**A screenshot of a computer

Description automatically generated with medium confidence**

The relevant information as previous distribution charts and data this distribution has more than three client visits which could have sets of three visits along the way.

**Webstore Online Spending Distribution**

**A screenshot of a computer

Description automatically generated with medium confidence**

Distribution has not given much information to provide the clients actual spending rate in which money flow is set for clients spending less than fifty dollars. Now, if it’s made from another sector that doesn’t include any zeros then it’ll be more precise results.

**Subcontracted Third-Party Spending Distribution**

**A screenshot of a computer

Description automatically generated with medium confidence**

Similar to the previous distribution, clients that did not submit any cash flow are included into the zero through the three-hundred-dollar range. By including the zero, its better shown how the zero to hundred intervals counts for the eighty percent. Although, when the zero is apart from the webstore and the subcontracted third-party spending distributions then its clear to view the most clients that were not able to buy anything.

**Age Distribution**

Graphical user interface

Description automatically generated

**Income Distribution**

**Graphical user interface, application

Description automatically generated**

A picture containing text, electronics

Description automatically generated

For the five hundred entries from each client that already have taken the survey is shown above. The overview of the relations between the variables is that two are of the pairs are very alike while other pairs are not being demonstrated. As it spikes higher in the webstore visits joins along with the spending webstore. The increase of the subcontracted third-party vendor visits are joined with the spending subcontracted third-party vendor. Most of all the relationships have correlations that are low and feeble to be considered to use.

Graphical user interface, application

Description automatically generated

**Strengths and Weaknesses**

From the data given, lacking suitable removal, this is not as useful for analysis to improve the Bubba Gump Company with its economical issue. As I have seen the departure from its counts and calculations as it grows, it seems that it will be deceived. This is shown on all variables which clients are spending, where I think this issue is by not departing the client that spent with the other clients that only see and don’t spend to accumulate what services or products in general all clients have spent individually. Although, I was not able to identify other variables which were missing information. Like the restaurant expenses are in a large share of the revenue in Bubba Gump’s Company, which it’s not properly analyzed. So, the missing information is involved that would mean a lot and could be understood better on why these variables are also included.

Therefore, the variables that are very useful in this analysis are the Age and Income variables. Due to the precise and useful data given, which it can lead me to see the association from these two variables and the volume on expenses being spent. Also, with this data Bubba Gump’s Company will be wiser to make better decisioning whether they advertise, promote, and include offers for customers that do spend in products or services only which can control sales and future revenue.