5-2 Exercise 6: Supervised Learning with Bubba Gump Data

Joel Meza

Professor Torre Roenne

April 2, 2021

DAT 220 -Fundamentals of Data Mining

Southern New Hampshire University

By using JMP, I used two cluster analyzes on the Bubba Gump Shrimp Company information. The cluster was through the web store by spending and the persuasion is the twenty natural clusters where there broken up depending on web spending sum. The other cluster analyze was on web store was spent and the visits recommended thirty natural clusters. This clustered information by who has visited the web store the furthermost and spending more which visited the least but spend the maximum from the visited the most but spending the minimum. Last cluster was the clients that have not visited the online store along with not spending at the online store.

Now, I made the linear regression models for online webstore has spending against all the clients features the age, status, and income gross. Where nothing is these models were necessary while the looking a webstore spent by the restaurant spending there a pure connection. Both have parameters estimating indication in the precise model as together possibilities are very low. From the graph scatter plot with the linear regression line are showing the total of clients spending at a Bubba Gump restaurant are very relevant to spend more on the online web store than clients who did not spent as great within a restaurant.

I made a logistic regression model to expect whether a client would build an online purchase. Although, when I attempt to diverse variables together with the dual online webstore purchase. The logistic regression model of online website purchase against online visits that while the clients are familiarizing from the online visits specifies the first online purchase on their first visit. Clients that return to the web store with their second appearance are more likely to purchase. Otherwise, clients are returning for their third visit can decide better to execute the purchases.

**Dendrogram**

A picture containing diagram

Description automatically generated

**Bivariate Fit of Webstore\_Spend By Restaurant**

Chart, scatter chart

Description automatically generated

A picture containing icon

Description automatically generated

**Linear Fit**

Webstore\_Spend = -13.27585 + 0.7245478\*Restaurant

**Summary of Fit**

|  |  |
| --- | --- |
| RSquare | 0.20557 |
| RSquare Adj | 0.203975 |
| Root Mean Square Error | 78.40066 |
| Mean of Response | 40.474 |
| Observations (or Sum Wgts) | 500 |

**Analysis of Variance**

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Ratio** |
| --- | --- | --- | --- | --- |
| Model | 1 | 792086.0 | 792086 | 128.8644 |
| Error | 498 | 3061038.6 | 6147 | **Prob > F** |
| C. Total | 499 | 3853124.7 |  | <.0001\* |

**Parameter Estimates**

| **Term** | **Estimate** | **Std Error** | **t Ratio** | **Prob>|t|** |
| --- | --- | --- | --- | --- |
| Intercept | -13.27585 | 5.891741 | -2.25 | 0.0247\* |
| Restaurant | 0.7245478 | 0.063826 | 11.35 | <.0001\* |

**Logistic Fit of WEB\_PURCH\_YN By WEB\_VISITS**

Histogram, rectangle

Description automatically generated

**Whole Model Test**

| **Model** | **-LogLikelihood** | **DF** | **ChiSquare** | **Prob>ChiSq** |
| --- | --- | --- | --- | --- |
| Difference | 170.28874 | 1 | 340.5775 | <.0001\* |
| Full | 162.22863 |  |  |  |
| Reduced | 332.51737 |  |  |  |

|  |  |
| --- | --- |
| RSquare (U) | 0.5121 |
| AICc | 328.481 |
| BIC | 336.886 |
| Observations (or Sum Wgts) | 500 |

**Parameter Estimates**

| **Term** | **Estimate** | **Std Error** | **ChiSquare** | **Prob>ChiSq** |
| --- | --- | --- | --- | --- |
| Intercept[1] | -5.4419748 | 1.002145 | 29.49 | <.0001\* |
| WEB\_VISITS | 5.72760999 | 1.0079514 | 32.29 | <.0001\* |