Tasha Mauer

The Charts bellow where done using a program called JMP and where done off of information off of a jmp file. The original project had to do with finding information to try and increase online sales for a restaurant. I used clusters, regressions, and correlations to show the relationship between different variables such as age and webstore purchases. These types of charts help to analyze the data in a way that can be read so you can see patterns that are harder to spot just by looking.

I did find some meaningful patterns in the data. I found that people who visited the restaurant multiple times where more likely to visit the online store, the more they visited the store the more they spent, people between 20 and 30 where more likely to purchase things off the webstore, and people with an income between 30 and 60 where more likely to buy off the store. These are just an example of how charts can be more useful than just having someone look at all the data and possibly miss something.

**Summary Plots**



**Parallel Coordinate Plot**

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

**Cluster Summary**

| **Cluster** | | **Count** |
| --- | --- | --- |
| 1 | | 10 |
| 2 | | 3 |
| 3 | | 3 |
| 4 | | 329 |
| 5 | | 4 |
| 6 | | 7 |
| 7 | | 8 |
| 8 | | 9 |
| 9 | | 3 |
| 10 | | 24 |
| 11 | | 4 |
| 12 | | 4 |
| 13 | | 10 |
| 14 | | 62 |
| 15 | | 20 |
| **Step** | **Criterion** | | |
| 8 | 0 | | |

**Cluster Means**

| **Cluster** | **RES\_VISITS** | **Webstore\_Spend** | **WEB\_VISITS** |
| --- | --- | --- | --- |
| 1 | 2 | 191.7 | 3 |
| 2 | 2 | 386 | 2.33333333 |
| 3 | 1 | 445.666667 | 1.66666667 |
| 4 | 1 | 8.76899696 | 0.50151976 |
| 5 | 2 | 413 | 1 |
| 6 | 3 | 174.714286 | 1.42857143 |
| 7 | 1 | 229.75 | 1.25 |
| 8 | 1 | 185.666667 | 3 |
| 9 | 3 | 382.666667 | 2.66666667 |
| 10 | 3 | 0.91666667 | 0.33333333 |
| 11 | 3 | 135 | 2.75 |
| 12 | 3 | 304 | 1.75 |
| 13 | 2 | 178 | 1.1 |
| 14 | 2 | 3.9516129 | 0.48387097 |
| 15 | 1.2 | 80.25 | 2 |

**Bivariate Fit of Webstore\_Spend By WEB\_VISITS**



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



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



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



**Whole Model Test**

| **Model** | **-LogLikelihood** | **DF** | **ChiSquare** | **Prob>ChiSq** |
| --- | --- | --- | --- | --- |
| Difference | 0.39863 | 1 | 0.797254 | 0.3719 |
| Full | 332.11874 |  |  |  |
| Reduced | 332.51737 |  |  |  |

|  |  |
| --- | --- |
| RSquare (U) | 0.0012 |
| AICc | 668.262 |
| BIC | 676.667 |
| Observations (or Sum Wgts) | 500 |

| **Measure** | **Training** | **Definition** |
| --- | --- | --- |
| Entropy RSquare | 0.0012 | 1-Loglike(model)/Loglike(0) |
| Generalized RSquare | 0.0022 | (1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n)) |
| Mean -Log p | 0.6642 | ∑ -Log(ρ[j])/n |
| RMSE | 0.4855 | √ ∑(y[j]-ρ[j])²/n |
| Mean Abs Dev | 0.4714 | ∑ |y[j]-ρ[j]|/n |
| Misclassification Rate | 0.3820 | ∑ (ρ[j]≠ρMax)/n |
| N | 500 | n |

**Parameter Estimates**

| **Term** |  | **Estimate** | **Std Error** | **ChiSquare** | **Prob>ChiSq** |
| --- | --- | --- | --- | --- | --- |
| Intercept[0] |  | 0.28796171 | 0.2347227 | 1.51 | 0.2199 |
| Age |  | 0.00456328 | 0.0051241 | 0.79 | 0.3732 |

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



**Whole Model Test**

| **Model** | **-LogLikelihood** | **DF** | **ChiSquare** | **Prob>ChiSq** |
| --- | --- | --- | --- | --- |
| Difference | 0.28311 | 1 | 0.566219 | 0.4518 |
| Full | 332.23426 |  |  |  |
| Reduced | 332.51737 |  |  |  |

|  |  |
| --- | --- |
| RSquare (U) | 0.0009 |
| AICc | 668.493 |
| BIC | 676.898 |
| Observations (or Sum Wgts) | 500 |

| **Measure** | **Training** | **Definition** |
| --- | --- | --- |
| Entropy RSquare | 0.0009 | 1-Loglike(model)/Loglike(0) |
| Generalized RSquare | 0.0015 | (1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n)) |
| Mean -Log p | 0.6645 | ∑ -Log(ρ[j])/n |
| RMSE | 0.4856 | √ ∑(y[j]-ρ[j])²/n |
| Mean Abs Dev | 0.4716 | ∑ |y[j]-ρ[j]|/n |
| Misclassification Rate | 0.3820 | ∑ (ρ[j]≠ρMax)/n |
| N | 500 | n |

**Parameter Estimates**

| **Term** |  | **Estimate** | **Std Error** | **ChiSquare** | **Prob>ChiSq** |
| --- | --- | --- | --- | --- | --- |
| Intercept[0] |  | 0.28145097 | 0.2806083 | 1.01 | 0.3159 |
| Income |  | 0.00334523 | 0.0044553 | 0.56 | 0.4527 |

The graphs above are the graphs that I created in analyzing the data from the survey that was given. The Parallel coordinate plot shows a cluster of fifteen, unfortunately the first and last clusters where mixed up in color, but this shows different groups of the same data which lets you see the information differently. The cluster information shows the count in the cluster so we can see that cluster 4 has the highest count with 329. The cluster means show the mean of all the data so that way we don’t have to do it, so for cluster 4 we can see that the means are RES\_VISITS 1, Webstore\_Spend 8.7, and WEB\_VISITES 0.5. We can also see that cluster 6, 9, 10, and 11 have the highest RES\_VISITS. Cluster 9 has the highest Webstore\_Spend at 382.7, and Cluster 8 and 1 have the highest WEB\_VISITS at 3. The next chart shows that on the first visit to the webstore people spend less with the majority of spending being 0 to 100. On the second the majority is 50 to 100 with more groupings in bigger spending areas such as 350 to 400. The third visit show spending at 100 to 250. The next chart shows that across the ages the most spending comes in at 0 to 50. It also shows that the older the customer the less they are likely to spend on the webstore. Ages 20 to 25 have the highest spending while 70 to 80 have the least. The next chart goes by zip and it shows that some zip codes spend more on the store than others. The whole model test shows that goes with the first logistic chart also shows some interesting data if you can understand it all, while I understand some I don’t understand it all. The model test breaks down the 500 observations that I feed to it to give me information that will help to understand the chart it goes with. The last chart and model test does the same but for separate variables.