Marketing Analysis

The goal of the analysis is to use data to determine if improving the brand's image among a young demographic will positively impact consideration. The data is survey data from a third party. We will analyze this data to determine if this hypothesis is true for our company. We are given a .txt file with survey responses, and a legend with the variable names matched to the survey question (see data folder).

To start, the .txt file will be loaded into *R*. Next we can look at the legend given to see what variables we have in the data and the survey questions they correspond to. Because we are targeting a young demographic, we create indicator age variables for ages 18-24, 25-44, 45-64 and 65 or older, and add it into the data frame. Because we are measuring consideration, we can go to the legend and find the variable that measures this. This variable is called *consideration* based on how likely they are to consider the company on a scale from 1 (not very likely) to 10 (extremely likely). We will need our data to be binary to produce a logistic regression model. We can take all non-demographic variables and classify them as 1 if they fall into the top three category (after confirming the scale stayed the same between all variables, response was between 8-10) and 0 if they do not.

Something that could be interesting to look at is if there are significant differences in the means of the metrics (top 3 transforms of Attributes, Familiarity, Favorability, and Consideration) between males and females. We can perform a two sample, two sided t-test at a 99% confidence level to determine this. We find that we cannot reject the null hypothesis for 4 of the variables, T3Bimageattr2 (fixes problems quickly), T3Bimageattr4 (has consistent pricing), T3Bimageattr19 (offers the most reliable service, even in bad weather), T3Bfamiliarity (familiarity with the company), however we can reject it for 17 other variables.

In order to determine if improving the brand's image among a young demographic will positively impact consideration, I would first subset the data to just the young, before I fit a logistic model. I picked the ages from 18-44 as the young audience. After I created the model, I changed the ages a bit however still concluded that 18-44 should be used as the young audience. Using the top 3 category format of the data and looking at the young, I would then use R to test the significance of different predictor variables with and without interaction, keeping the consideration variable as the response variable. During this process, I would decide which variables to drop and finally pick the model with the lowest AIC value. This would allow me to determine what would happen to consideration if brand attitudes improved among the young.

After looking at some different models I determined the following was the best model:

```
Call:
glm(formula = T3Bconsideration ~ . - T3Bimageattr4 - T3Bimageattr7 -
    T3Bimageattr10 - T3Bimageattr20 - T3Bimageattr15 - T3Bimageattr22 -
    T3Bimageattr23 - T3Bimageattr24 - T3Bimageattr2 - T3Bimageattr5 -
    AG - T3Bimageattr21, family = binomial, data = T3B)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.2856  -0.5853   0.3904   0.5673   2.2740

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)     -2.50719    0.05674 -44.189  < 2e-16 ***
T3Bimageattr1    0.22741    0.07048   3.226 0.001254 **
T3Bimageattr6    0.22225    0.07224   3.077 0.002094 **
T3Bimageattr8    0.20437    0.06610   3.092 0.001989 **
T3Bimageattr17   0.34082    0.07036   4.844 1.27e-06 ***
T3Bimageattr9    0.33333    0.07735   4.309 1.64e-05 ***
T3Bimageattr11   0.29262    0.07764   3.769 0.000164 ***
T3Bimageattr19   0.41005    0.06863   5.975 2.30e-09 ***
T3Bfamiliarity   0.82961    0.05927  13.996  < 2e-16 ***
T3Bfavorability  2.18259    0.05500  39.681  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 15622.0  on 11268  degrees of freedom
Residual deviance:  9653.1  on 11259  degrees of freedom
AIC: 9673.1

Number of Fisher Scoring iterations: 4
```

After obtaining the model with the lowest AIC value and eliminating the image attributes with insignificant P-values, the only significant image attributes remaining were image attribute 1, 6, 8, 17, 9, 11, 19, familiarity and favorability. These attributes should be the focus of marketing efforts going forward. Based on this model increasing these image attributes among the young, to survey scores of 8, 9, or 10, will improve brand consideration.

Let's look at a different model that includes media exposure. The model being fit was done via Logistic Regression. The goal is to determine what drives likelihood to purchase a product in a given category.

Dependent Variable: Likelihood to Purchase Product X

- 0: not likely to purchase
- 1: likely to purchase
- Mean of variable is 0.25.

Potential Independent Variables:

- Media
    - 0: not exposed
    - 1: exposed
- Demographics
    - 0: not in that demo group
    - 1: in that demo group
- Brand
    - 0: not in that group
    - 1: in that group
- Interactions

Table of Coefficients

Variable	Value	P-Value	T-Value
Brand: Attribute A	1.22	0.00	5.90
Brand: Attribute B	1.48	0.00	7.74
Brand: Intent to make Purchase in Category in the next 3 months	0.43	0.04	2.09
Brand: Intent to make Purchase in Category in the next 6 months	0.39	0.03	2.22
Brand: Price perception	0.36	0.00	3.12
Brand: Prior Experience with the Brand	0.38	0.00	2.90
Demo: Age 55-64 (older age group)	-0.33	0.01	-2.61
Demo: Hispanic	0.48	0.03	2.23
Demo: Income < $25,000	-0.27	0.11	-1.60
Media: Digital	1.27	0.00	3.57
Media: Magazine	0.43	0.09	1.68
Media: Prime TV	0.17	0.02	2.36
Media: Sports TV	0.15	0.07	1.82
Interaction: (Demo: Income < $25,000) & (Brand: Attribute A)	0.77	0.00	4.93
Interaction: (Demo: Male) & (Media: Sports TV)	-0.70	0.00	-3.54

Key findings: Looking at this model we have insufficient evidence to assume the baseline probability of the likelihood of purchase is 0%. The data also demonstrated that younger people are more likely to make a purchase than older individuals. The probability of the older age group making a purchase is exp(ß0-.33)/(1+exp(ß0-.33)). Income is usually insignificant as to the purchasing decision. While it is better to improve both attributes (A and B), if the budget does not allow this, then it would be better to choose to improve B over A. The information is inconclusive that marketing dollars should be moved away from males. Additionally, the data does not demonstrate that we should move marketing dollars away from low income groups as their income does not affect their buying decision. The data shows that sports TV is more effective for females than males.