Christian Zink, David Kao, Peter Chen, Miloni Desai

MKTG 464

Group Project

6/10/2019

**Executive Summary**

Our findings tell us that there are two key segments of customers that HATCO has to target. The size of each of these segments is almost the same and thus to have an overall greater sale, HATCO must focus equally on both these segments. We found that the needs of both these segments are also significantly different. Thus, the marketing needs to have two campaigns to target all their customers.

The first segment is the Economy Customer Segment. This segment gives priority to delivery speed over product quality and is also highly price sensitive. Thus, from a managerial standpoint it is important to come up with offers, deals, and discounts to boost sales. Another strategy could be to come up with a two-day shipping scheme, for a members only level (Something drawing from Amazon Prime). As these customers are price sensitive and will be unwilling to pay for this kind of a membership, HATCO could offer this membership based on the number of purchases made or the frequency of purchase from them. Once they hit a certain number they could be included in this service and to maintain it, they have to continue their purchasing pattern. It is sort of like a loyalty program that promises better delivery times and discounts, as it appeals to this segment.

The second segment is the Premium Customer Segment. This segment is not price sensitive and puts its highest value in the quality of their product. They also seem to have a good brand image of HATCO and hence seem to be continuing their loyalty with them. To ensure that this segment remains a part of their prime market, HATCO must focus on two key areas. They need to make sure they focus on providing high quality products as well as a good overall customer service which will maintain their brand image. As price is not an issue with this segment, it is important to focus on the quality of products. Their marketing strategy must focus on marketing HATCO’s products as high quality and those of a trusted brand.

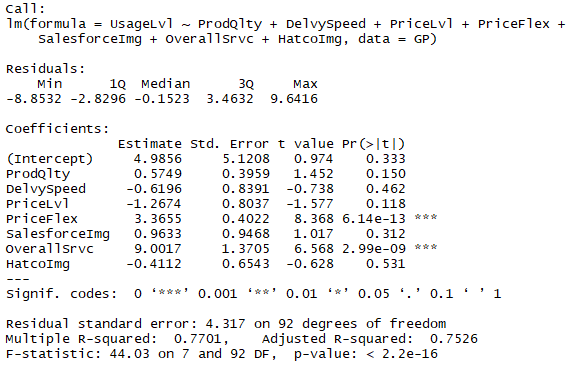
From our factor analysis, we associated key factors that HATCO must focus on for their products. (1) Product Perception (2) Company Image (3) Value (4) Price Flexibility.

For segment 1, factors 3 and 4 are of utmost importance and must be kept high to increase sales. While for segment 2 it is factors 1 and 2 that are valued most by the customers and thus business and marketing strategies must be made keeping these in mind.

1. 

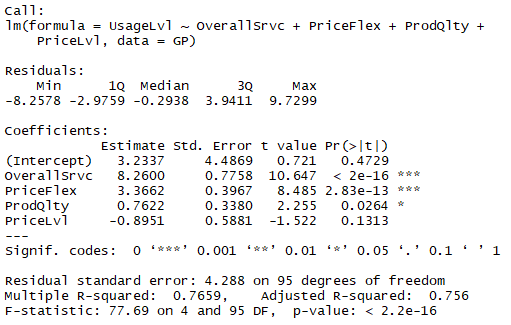
* The highest positive correlation is between Salesforce Image and HATCO Image. This means that these two variables are most closely associated with each other and move in the same direction. When salesforce image is higher, the image of HATCO is higher and when the salesforce’s image declines, so does HATCO’s.
* The highest negative correlation is between Product Quality and Delivery Speed. This means that these two variables have opposite associations and we can infer that as perceptions of product quality decreases, perceptions of delivery speed increase and vice versa as a higher product quality is associated with a lower delivery speed.

2.



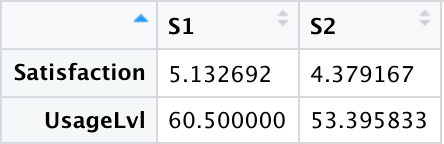
* Based on the regression, 77.01% of the variation in usage level can be explained by the perception variables.
* Perceptions on price flexibility and overall service have a significant relationship with usage level

3.



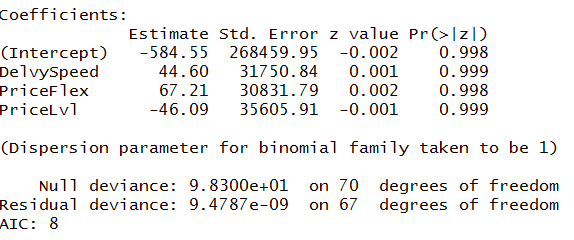
* Based on the stepwise regression, perceptions on overall service, price flexibility, and product quality have a significant relationship with the usage level
* When the rating of overall service increases by 1 point, usage level increases by 8.26%.
* When the rating of price flexibility increases by 1 point, usage level increases by 3.37%.
* When the rating of product quality increases by 1 point, usage level increases by 0.76%.

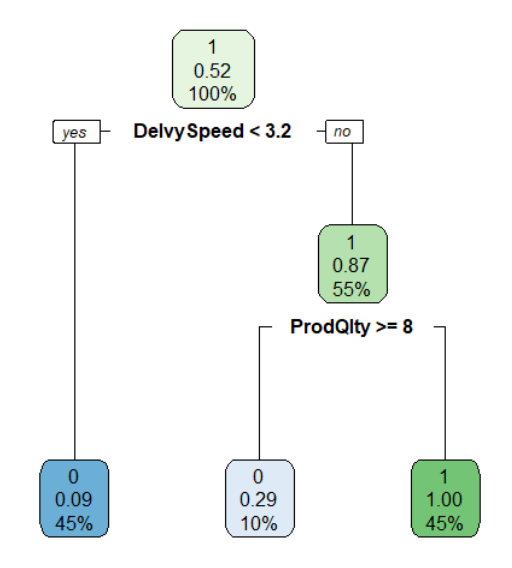
4.

* There are 52 customers in segment 1 and 48 in segment 2
* For segment 1 we can see that these customers give priority to the delivery speed over the quality of the product quality. They are price sensitive and feel like the products are overpriced. They believe that HATCO representatives will be flexible about prices and will be ready to negotiate. They have a fairly good image of the company and its workforce, and thus are happy with their services. Based on their preferences, we can classify segment one as the Economy consumers segment.
* For segment 2, we can see from the data that they give a very high priority to the quality of the product and are okay with the delivery speed being slow. Their rating for the price level tells us that they seem satisfied with the product quality and service, and don’t feel overcharged. They do expect HATCO to be flexible in terms of price up to a certain extent but not as much as the first segment does. They have a better image of the brand and its workforce and are almost as happy with their services as the first segment. Based on their priorities we can call this segment the Premium consumer segment.
* 
* From the purchase patterns we can see that the usage levels and satisfaction are both higher for the economic segment. This segment is marginally bigger than the second segment so it's a good thing that they have a higher satisfaction and usage level, but as segment two is almost as big as one but is not as price sensitive, it is advised that HATCO increase its customer satisfaction levels for the premium segment, which will also boost the usage levels.

5.

|  |  |
| --- | --- |
| **Logistic Regression Accuracy** | **Decision Tree Accuracy** |
| Test: 96.6% | Test: 90% |
| Train: 100% | Train: 93% |

* The misclassification rates of our models indicate that the logistic regression is a better model at using the perception variables to predict segment membership. 



6a.

* Factor RC1 - Product Perception - Positively correlated with Product Quality and negatively correlated with Delivery Speed
* Factor RC2 - Company Image - Positively correlated with Sales Force Image and Hatco Image.
* Factor RC3 - Value - Highly positively correlated with Price level and Overall Service.
* Factor RC4 - Price Flexibility - Highly positively correlated with Price flexibility.

6b.

* For Factor 1, p-value = 0.07 > 0.05 which means that the segments are not statistically different on this Factor with a 95% confidence interval.
* For Factor 2, p-value = 1.4e-13 < 0.05 which means that the segments are statistically different on this Factor with a 95% confidence interval.
* For Factor 3, p-value = 0.18 > 0.05 which means that the segments are not statistically different on this Factor with a 95% confidence interval.
* For Factor 4, p-value = 1.6e-11 < 0.05 which means that the segments are statistically different on this Factor with a 95% confidence interval.

6c.   
 The recent results tell us that there are four factors that define and affect Hatco’s key consumer segments. The economic segment is price sensitive, values delivery speed over product quality and has a favorable view of Hatco. The premium segment is not as price sensitive, values product quality, doesn’t mind longer delivery time, and has a very high perception of Hatco. The segments are fairly equal in size and of equal importance. The satisfaction and usage levels are greater for the economic segment so we should improve the purchase outcome variables for the premium segment.   
 The segments have different priorities and value factors differently, so its necessary to identify segment membership so Hatco can deliver the service experience suited for them. The regression model is more accurate so we should use it to predict segment membership instead of the classification tree.

The two segments differ on the company image and price flexibility factors. The economic segment has a lower perception of Hatco and wants to negotiate price more. Hatco can use this understanding to give the key segments the optimal experience suited for them.

R Code:

#1.

round(cor(GP[,2:8]), digits = 3)

Perceptions <- GP[, c(2:8)]

Perceptions\_cor <- cor(Perceptions)

corrplot(cor(GP[,2:8]), order = "hclust")

#2.

usage.reg = lm(UsageLvl ~ ProdQlty + DelvySpeed + PriceLvl + PriceFlex + SalesforceImg + OverallSrvc + HatcoImg, data = GP)

summary(usage.reg)

#3.

Base = lm(UsageLvl ~ 1, data = GP)

Full = usage.reg

Step = step(Base, scope=list(lower = "Base", upper = "Full"), direction = "both")

summary(Step)

#4.

describe(GP[2:8])

GP.ca = data.frame(GP[2:8])

GP.dist = dist(GP.ca)

GP.hc = hclust(GP.dist, method = "ward.D2")

plot(GP.hc, col = "darkgreen")

rect.hclust(GP.hc, k=2, border = "red")

GP.ca[ncol(GP.ca)+1] = cutree(GP.hc, k = 2)

names(GP.ca)[ncol(GP.ca)] = "hc.seg"

GP.hc.mean = aggregate(.~hc.seg, data=GP.ca, mean)

GP.kmeans = kmeans(GP.ca[1:7], centers = GP.hc.mean[2:8])

GP[ncol(GP)+1] = GP.kmeans$cluster

names(GP)[ncol(GP)] = "kmeans.seg"

GP.seg.mean = aggregate(.~kmeans.seg, data=GP[10:11], length)

names(GP.seg.mean)[2] = "Seg.size"

GP.seg.mean[3:12] =  aggregate(.~kmeans.seg, data=GP[2:11], mean)

GP.seg.mean[3] = NULL

prctl = t(apply(GP[2:8], 2, quantile, probs = c(0.25,0.75)))

seg.df = data.frame(t(GP.seg.mean[3:9]), prctl)

names(seg.df) = c("S1", "S2", "25prctl", "75prctl")

seg.target = data.frame(t(GP.seg.mean[10:11]))

names(seg.target) = c("S1", "S2")

#5

full.model = kdummy ~ ProdQlty + DelvySpeed + PriceLvl + PriceFlex + SalesforceImg + OverallSrvc + HatcoImg

full.tree = rpart(formula = full.model, data = GP.train, control = rpart.control(cp = 0.01))

rpart.plot(full.tree)

#predicting click probability for classification tree

full.kdummy.prediction = predict(full.tree, GP.test, type="class")

mean(GP.test$kdummy == full.kdummy.prediction)

summary(full.kdummy.prediction)

full.kdummy.prediction2 = predict(full.tree, GP.train, type = "class")

mean(GP.train$kdummy == full.kdummy.prediction2)

summary(full.kdummy.prediction2)

#6a.

corrplot(cor(GP[,2:8]), order = "hclust")

GP.pc = prcomp(GP[,2:8], scale = TRUE)

summary(GP.pc)

plot(GP.pc, type = 'l')

GP.fa1 = principal(GP[2:8], nfactors = 4,

                  rotate = "none")

GP.fa1$loadings

GP.fa1$communality

GP.fa2 = principal(GP[2:8], nfactors = 4,

                  rotate = "varimax")

GP.fa2$loadings

GP.fa2$communality

#6b.

GP.fa2$weights

summary(GP.fa2$scores)

GP[12:15] = GP.fa2$scores

names(GP)[12:15] = c("Factor1", "Factor2", "Factor3", "Factor4")

pairwise.t.test(GP$Factor1, GP$kmeans.seg,

               p.adjust.method = "none", pool.sd = FALSE)

pairwise.t.test(GP$Factor2, GP$kmeans.seg,

               p.adjust.method = "none", pool.sd = FALSE)

pairwise.t.test(GP$Factor3, GP$kmeans.seg,

               p.adjust.method = "none", pool.sd = FALSE)

pairwise.t.test(GP$Factor4, GP$kmeans.seg,

               p.adjust.method = "none", pool.sd = FALSE)