Marketing Analytics –MGMT 525

2015

Individual Assignment

HATCO Management

By Indhujha Natarajan

Executive Summary

HATCO is an industrial supplier of custom-built intermediate products. Its customer base consists of 2 segments with contrasting needs. Segment 1 wants Quality Products at Competitive prices whereas Segment 2 wants Delivery Speed and Flexible Pricing. It is difficult to satisfy both their needs. For example, if HATCO charges competitive prices there will no room to negotiate and thereby would not be able to accommodate Flexible Pricing. Also, since these are custom products, increasing quality may increase delivery times thereby faltering on the ‘Delivery Speed’ attribute. In this scenario, HATCO will have to either focus on one of the segments or create different value packages for each segment. The value package needed by each segment is as follows:

**For Segment 1:** Quality products need to be provided at Competitive Prices. However, the customer should be made aware that these products will have longer delivery times compared to other low quality products and that prices will be fixed as the best price has already been offered.

**For Segment 2:** Delivery Speed and Flexible Pricing need to be provided. To satisfy these needs, operational improvements are needed and the sales force also need to have good negotiation skills. However, the products may be of lower quality compared to products for Segment 1 and book prices may also be high for the quality provided in order to have room to negotiate.

From the cluster analysis we see that both the segments have given poor ratings for Overall Service Level. Also, from the final regression equation we see that a one point increase in the Overall Service rating will lead to an 8.260% increase in Usage Level. This increases Usage Level much more than other significant variables and hence HATCO should focus on improving overall service. Similarly, increase in Price Flex can increase Usage Level by 3.366%. However, an increase in Product Quality increases Usage Level only by 0.762%. This shows that HATCO’s current offerings are much more in line with Segment 1 which is also reflected in the actual higher Usage Level and Satisfaction Rating given by Segment 1 compared to those given by Segment 2. However, Segment 2 still has a Usage Level of 53.4%. So, HATCO will now have to decide whether it can adapt its skills to serve both segments’ needs or simply focus on Segment 1 and maintain status-quo with Segment 2.

**1. Correlations**

**Correlation: Prod Qlty, DelvySpeed, PriceLevel, PriceFlex, SalesForceIm, OverallServi, ...**

Prod Qlty DelvySpeed PriceLevel PriceFlex

DelvySpeed -0.642

PriceLevel 0.470 -0.367

PriceFlex -0.448 0.527 -0.487

SalesForceImg 0.177 0.018 0.186 -0.034

OverallService -0.055 0.496 0.513 0.067

MfrImage 0.199 0.014 0.266 -0.103

SalesForceImg OverallService

OverallService 0.241

MfrImage 0.801 0.297

Cell Contents: Pearson correlation

The highest Positive Correlation of 0.801is between MfrImage and SalesForceImg. This could be explained by the fact that SalesForce are the face of the company. Hence the image the Sales force leave upon the customers will be the Manufacturer’s Image as well.

The highest Negative Correlation of -0.642 is between Delivery Speed and Prod QLty. This could be explained by the fact that Product Quality and Delivery Speed are two different things in the value package that HATCO provides to customers. Product Quality may be good but may be delivered late and vice versa. Hence these two variables are negatively correlated with each other.

**2. Regression Analysis**

**Regression Analysis: UsageLvl versus Prod Qlty, DelvySpeed, PriceLevel, PriceFlex, ...**

Analysis of Variance

Source DF Adj SS Adj MS F-Value P-Value

Regression 7 5745.25 820.75 44.03 0.000

Prod Qlty 1 39.31 39.31 2.11 0.150

DelvySpeed 1 10.16 10.16 0.55 0.462

PriceLevel 1 46.36 46.36 2.49 0.118

PriceFlex 1 1305.35 1305.35 70.03 0.000

SalesForceImg 1 19.29 19.29 1.04 0.312

OverallService 1 804.19 804.19 43.14 0.000

MfrImage 1 7.36 7.36 0.39 0.531

Error 92 1714.94 18.64

Lack-of-Fit 91 1696.94 18.65 1.04 0.672

Pure Error 1 18.00 18.00

Total 99 7460.19

Model Summary

S R-sqR-sq(adj) R-sq(pred)

4.31748 77.01% 75.26% 72.97%

Coefficients

Term Coef SECoef T-Value P-Value VIF

Constant 4.99 5.12 0.97 0.333

Prod Qlty 0.575 0.396 1.45 0.150 2.09

DelvySpeed -0.620 0.839 -0.74 0.462 5.86

PriceLevel -1.267 0.804 -1.58 0.118 4.90

PriceFlex 3.366 0.402 8.37 0.000 1.65

SalesForceImg 0.963 0.947 1.02 0.312 2.83

OverallService 9.00 1.37 6.57 0.000 5.63

MfrImage -0.411 0.654 -0.63 0.531 2.96

Regression Equation

UsageLvl = 4.99 + 0.575 Prod Qlty - 0.620 DelvySpeed - 1.267 PriceLevel + 3.366 PriceFlex

+ 0.963 SalesForceImg + 9.00 OverallService - 0.411 MfrImage

Fits and Diagnostics for Unusual Observations

ObsUsageLvl Fit ResidStdResid

5 58.00 66.85 -8.85 -2.21 R

15 73.00 64.41 8.59 2.10 R

16 69.00 59.76 9.24 2.24 R

21 63.00 53.36 9.64 2.28 R

31 62.00 52.48 9.52 2.25 R

57 72.00 71.33 0.67 0.18 X

R Large residual

X Unusual X

The proportion of variation in Usage level in the sample data explained by the seven perception variables is given by R-sq which is 77.01%.

The perception variables that have a significant relationship with Usage level are Price Flex and Overall Service. Both have p-values equal to 0.000 which is less than 0.05 (the significance level).Hence the null hypotheses that Price Flex and Overall Service are insignificant can be rejected.

Also since the VIF<10 for all variables, there is no severe multi-collinearity problem.

**3. Best Subsets Regression**

**Best Subsets Regression: UsageLvl versus Prod Qlty, DelvySpeed, ...**

Response is UsageLvl

                                                      O

                                                    S v

                                                    a e

                                                    l r

                                              D P   e a

                                            P e r P s l

                                            r l i r F l M

                                            o v c i o S f

                                            d y e c r e r

                                              S L e c r I

                                            Q p e F e v m

                                            l e v l I i a

             R-Sq    R-Sq  Mallows          t e e e m c g

Vars  R-Sq  (adj)  (pred)       Cp       S  y d l x g e e

   1  46.2   45.7    44.1    119.3  6.3994            X

   2  75.2   74.7    73.9      5.1  4.3633        X   X

   3  76.0   75.3    74.2      4.0  4.3172  X     X   X

   4  76.6   75.6    74.2      3.7  4.2879  X   X X   X

   5  76.8   75.5    73.7      5.0  4.2938  X   X X X X

   6  76.9   75.4    73.4      6.4  4.3034  X X X X X X

   7  77.0   75.3    73.0      8.0  4.3175  X X X X X X X

 As per the Best Subsets Regression, the set of variables that have the highest R-Sq (adj) of 75.6 are Prod Qlty, Pricelevel, PriceFlex and Overall Service. Hence these variables will be used to generate a nice and simple regression equation that does a good job of explaining the variation in Usage Level.

**4. Regression equation using variables chosen from Best Subsets Regression**

**Regression Analysis: UsageLvl versus Prod Qlty, PriceLevel, PriceFlex, OverallService**

 Analysis of Variance

Source            DF   Adj SS   Adj MS  F-Value  P-Value

Regression         4  5713.51  1428.38    77.69    0.000

  Prod Qlty        1    93.48    93.48     5.08    0.026

  PriceLevel       1    42.60    42.60     2.32    0.131

  PriceFlex        1  1323.74  1323.74    72.00    0.000

  OverallService   1  2084.17  2084.17   113.36    0.000

Error             95  1746.68    18.39

  Lack-of-Fit     93  1678.68    18.05     0.53    0.842

  Pure Error       2    68.00    34.00

Total             99  7460.19

 Model Summary

S    R-sq  R-sq(adj)  R-sq(pred)

4.28790  76.59%     75.60%      74.17%

 Coefficients

 Term              Coef  SE Coef  T-Value  P-Value   VIF

Constant          3.23     4.49     0.72    0.473

Prod Qlty        0.762    0.338     2.25    0.026  1.55

PriceLevel      -0.895    0.588    -1.52    0.131  2.66

PriceFlex        3.366    0.397     8.49    0.000  1.63

OverallService   8.260    0.776    10.65    0.000  1.83

Regression Equation

 UsageLvl = 3.23 + 0.762 Prod Qlty - 0.895 PriceLevel + 3.366 PriceFlex + 8.260 OverallService

 Fits and Diagnostics for Unusual Observations

                                Std

Obs  UsageLvl     Fit  Resid  Resid

16    69.000  60.056  8.944   2.14  R

21    63.000  53.525  9.475   2.24  R

31    62.000  52.270  9.730   2.30  R

 R  Large residual

As per the Best Subsets Regression, the set of variables that have the highest R-Sq (adj) of 75.6 are Prod Qlty, Pricelevel, PriceFlex and Overall Service. Hence these variables will be used to generate a nice and simple regression equation that does a good job of explaining the variation in Usage Level.

**4. Regression equation using variables chosen from Best Subsets Regression**

**Regression Analysis: UsageLvl versus Prod Qlty, PriceLevel, PriceFlex, OverallService**

Analysis of Variance

Source DF Adj SS Adj MS F-Value P-Value

Regression 4 5713.51 1428.38 77.69 0.000

Prod Qlty 1 93.48 93.48 5.08 0.026

PriceLevel 1 42.60 42.60 2.32 0.131

PriceFlex 1 1323.74 1323.74 72.00 0.000

OverallService 1 2084.17 2084.17 113.36 0.000

Error 95 1746.68 18.39

Lack-of-Fit 93 1678.68 18.05 0.53 0.842

Pure Error 2 68.00 34.00

Total 99 7460.19

Model Summary

S R-sq R-sq(adj) R-sq(pred)

4.28790 76.59% 75.60% 74.17%

Coefficients

Term Coef SECoef T-Value P-Value VIF

Constant 3.23 4.49 0.72 0.473

Prod Qlty 0.762 0.338 2.25 0.026 1.55

PriceLevel -0.895 0.588 -1.52 0.131 2.66

PriceFlex 3.366 0.397 8.49 0.000 1.63

OverallService 8.260 0.776 10.65 0.000 1.83

Regression Equation

UsageLvl = 3.23 + 0.762 Prod Qlty - 0.895 PriceLevel + 3.366 PriceFlex + 8.260 OverallService

Fits and Diagnostics for Unusual Observations

Std

ObsUsageLvl Fit ResidResid

16 69.000 60.056 8.944 2.14 R

21 63.000 53.525 9.475 2.24 R

31 62.000 52.270 9.730 2.30 R

R Large residual

The perception variables that have a significant relationship with Usage level are Price Flex (P value-0.000), Overall Service (P-value-0.000) and Prod Qlty (0.026). All three have p-values less than 0.05 (the significance level).Hence the null hypotheses that Price Flex, Overall Service and Prod Qlty are insignificant can be rejected.

Though the VIFs were less than 10 in Q.2 they still are up to 5. Hence a small multi-collinearity problem still exists. However, here we see that all VIFs are much lesser than 10. Therefore no multi-collinearity exists between the variables in this regression equation.

The small multi-collinearity problem in Q.2 combined with the observation about the small increase in the R-Sq (adj) value in Q.3 strongly suggests that the coefficient of **Prod Qlty** is likely to be truly significant. Hence we can say that Product Qlty was dropped in Q.2 due to a small multi collinearity problem.

Interpretation of Significant Slope Coefficients:

Coefficient of PriceFlex:

Controlling for the effects of all other independent variables, a one point increase for PriceFlex in the 0-10 scale is expected to result in a 3.366% increase in Usage Level.

Coefficient of Overall Service:

Controlling for the effects of all other independent variables, a one point increase for Overall Service in the 0-10 scale is expected to result in an 8.260% increase in Usage Level.

Coefficient of Product Quality:

Controlling for the effects of all other independent variables, a one point increase for Product Quality in the 0-10 scale is expected to result in a 0.762% increase in Usage Level.

**5. Discovering Segments through Cluster Analysis**



Dividing the market into 2 segments as follows:



**Cluster Analysis of Observations: Prod Qlty, DelvySpeed, PriceLevel, PriceFlex, ...**

Final Partition

Number of clusters: 2

Average Maximum

Within distance distance

Number of cluster sum from from

observations of squares centroid centroid

Cluster1 54 320.676 2.35212 4.31493

Cluster2 46 269.129 2.31979 4.28942

Cluster Centroids

Grand

Variable Cluster1 Cluster2 centroid

Prod Qlty 5.93519 8.18696 6.971

DelvySpeed 4.22963 2.49783 3.433

PriceLevel 1.63148 3.22391 2.364

PriceFlex 8.88333 6.73261 7.894

SalesForceImg 2.53333 2.81957 2.665

OverallService 2.95000 2.87609 2.916

MfrImage 4.98889 5.58261 5.262

Distances Between Cluster Centroids

Cluster1 Cluster2

Cluster1 0.00000 3.95866

Cluster2 3.95866 0.00000

Doing cluster analysis using K-means:

**K-means Cluster Analysis: Prod Qlty, DelvySpeed, PriceLevel, PriceFlex, SalesForceIm, Overall**

Final Partition

Number of clusters: 2

Average Maximum

Within distance distance

Number of cluster sum from from

observations of squares centroid centroid

Cluster1 52 306.933 2.342 4.294

Cluster2 48 279.232 2.317 4.281

Cluster Centroids

Grand

Variable Cluster1 Cluster2 centroid

Prod Qlty 5.9038 8.1271 6.9710

DelvySpeed 4.2962 2.4979 3.4330

PriceLevel 1.5808 3.2125 2.3640

PriceFlex 8.9000 6.8042 7.8940

SalesForceImg 2.5250 2.8167 2.6650

OverallService 2.9577 2.8708 2.9160

MfrImage 4.9519 5.5979 5.2620

Distances Between Cluster Centroids

Cluster1 Cluster2

Cluster1 0.0000 3.9675

Cluster2 3.9675 0.0000

On comparing cluster groupings of the K-means clusters with that obtained from hierarchical clustering for the two cluster solution as follows:

**Descriptive Statistics: Hits**

**Total**

**Variable Count Mean Sum**

**Hits 100 0.9800 98.0000**

We see that only 2% of the observations are reclassified (<5%) by K means. Since this is less than the threshold of 5%, I am satisfied with the robustness of hierarchical clustering.

**6. Profiling the two customer segments**

**K-means Cluster Analysis: Prod Qlty, DelvySpeed, PriceLevel, PriceFlex, SalesForceIm, Overall**

Final Partition

Number of clusters: 2

Average Maximum

Within distance distance

Number of cluster sum from from

observations of squares centroid centroid

Cluster1 52 306.933 2.342 4.294

Cluster2 48 279.232 2.317 4.281

Cluster Centroids

Grand

Variable Cluster1 Cluster2 centroid

Prod Qlty **5.9038** 8.1271 6.9710

DelvySpeed 4.2962 **2.4979** 3.4330

PriceLevel **1.5808** 3.2125 2.3640

PriceFlex 8.9000 **6.8042** 7.8940

SalesForceImg 2.5250 2.8167 2.6650

OverallService 2.9577 2.8708 2.9160

MfrImage 4.9519 5.5979 5.2620

Distances Between Cluster Centroids

Cluster1 Cluster2

Cluster1 0.0000 3.9675

Cluster2 3.9675 0.0000

**Descriptive Statistics: UsageLvl, SatisLvl**

Total

Variable KMeans Count Mean Sum

UsageLvl 1 52 60.50 3146.00

2 48 53.40 2563.00

SatisLvl 1 52 5.133 266.900

2 48 4.379 210.200

This is the same as the means obtained through K-means cluster analysis:

**Descriptive Statistics: Prod Qlty, DelvySpeed, PriceLevel, PriceFlex, SalesForceIm, ...**

Total

Variable KMeans Count Mean Sum

Prod Qlty 1 52 5.904 307.000

2 48 8.127 390.100

DelvySpeed 1 52 4.296 223.400

2 48 2.498 119.900

PriceLevel 1 52 1.5808 82.2000

2 48 3.212 154.200

PriceFlex 1 52 8.900 462.800

2 48 6.804 326.600

SalesForceImg 1 52 2.525 131.300

2 48 2.8167 135.2000

OverallService 1 52 2.9577 153.8000

2 48 2.871 137.800

MfrImage 1 52 4.952 257.500

2 48 5.598 268.700

The means for the seven perception variables can be summarized as follows:

|  |  |  |
| --- | --- | --- |
| Segment | S1 | S2 |
| Prod Qlty | 5.904 | 8.127 |
| DelvySpeed | 4.296 | 2.498 |
| PriceLevel | 1.5808 | 3.212 |
| PriceFlex | 8.9 | 6.804 |
| SalesForceImg | 2.525 | 2.8167 |
| OverallService | 2.9577 | 2.871 |
| MfrImage | 4.952 | 5.598 |

Profiling clusters based on the ratings given by each segment for the perception variables and the actual Purchase Outcomes:

|  |  |  |
| --- | --- | --- |
| Segment | S1 | S2 |
| Segment Size | 52% | 48% |
| Prod Qlty |  |  |
| DelvySpeed |  |  |
| PriceLevel |  |  |
| PriceFlex |  |  |
| Usage Level | 60.5 | 53.4 |
| Satisfaction level | 5.133 | 4.379 |
| Name | **Quality Product and Competitive Price Seekers** | **Delivery Speed and Flexible Pricing Seekers** |

It is seen that for the same HATCO products, S1 has given lower than average ratings for Product Quality and Price Level but higher than average ratings for Delivery Speed and PriceFlex whereas S2 has given lower than average ratings for Delivery Speed and PriceFlex but higher than average ratings for Product Quality and Price Level. The surprising point is that the two segments have given contrasting ratings for the perception variables with respect to the same HATCO products.

**S1 (Quality Product and Competitive Price Seekers):** Segment 1 wants better Product quality and Price Level and hence they have given low ratings for the same. This segment has higher Usage Level or “share of wallet” of 60.5% and has also given a higher satisfaction rating of 5.13 for Satisfaction Level compared to Segment 2. Hence Segment 1 uses and likes HATCO Products more than Segment 2.

**S2 (Delivery Speed and Flexible Pricing Seekers):** Segment 2 wants better Delivery Speed and PriceFlex and hence they have lower ratings for the same. This segment has lower Usage Level or “share of wallet” of 53.4% and has also given a lower satisfaction rating of 4.38 for Satisfaction Level compared to Segment 1. Hence Segment 2 uses and likes HATCO Products lesser than Segment 1.

**7. More Profiling for the two cluster solution**

**(A) Factor Analysis**

**Principal Component Analysis: Prod Qlty, DelvySpeed, PriceLevel, PriceFlex, SalesForceIm, Ove**

Eigenanalysis of the Correlation Matrix

Eigenvalue 2.6212 2.0586 1.1805 0.5499 0.3257 0.1934 0.0707

Proportion 0.374 0.294 0.169 0.079 0.047 0.028 0.010

Cumulative 0.374 0.669 0.837 0.916 0.962 0.990 1.000

Variable PC1 PC2 PC3 PC4 PC5 PC6 PC7

Prod Qlty 0.484 -0.189 0.045 -0.584 -0.609 0.013 -0.126

DelvySpeed -0.405 0.452 -0.181 0.134 -0.483 -0.005 -0.590

PriceLevel 0.475 0.103 -0.490 -0.081 0.495 -0.010 -0.521

PriceFlex -0.431 0.231 0.159 -0.777 0.353 0.071 -0.042

SalesForceImg 0.283 0.457 0.474 0.044 0.067 -0.692 -0.028

OverallService 0.092 0.520 -0.567 -0.116 -0.144 -0.052 0.602

MfrImage 0.318 0.465 0.392 0.122 0.033 0.716 0.012

83.7% of the information in the seven variables can be summarised by Factors 1, 2 and 3.

The Scree Plot is in **Exhibit 1**.

Unrotated factor loadings. (The Unrotated Loading Plot is in **Exhibit 2**)

From the Unrotated Factors loadings, it is clear that it is difficult to interpret the above factors.

Hence considering the Varimax Rotated Factors:

Varimax rotated factor loadings. (The rotated loading plot is in **Exhibit 3**)

Rotated Factor Loadings and Communalities

Varimax Rotation

Variable Factor1 Factor2 Factor3 Communality

Prod Qlty 0.810 0.187 -0.008 0.691

DelvySpeed -0.875 0.029 -0.351 0.890

PriceLevel 0.632 0.107 -0.698 0.897

PriceFlex -0.780 0.024 0.128 0.626

SalesForceImg 0.035 0.948 -0.075 0.905

OverallService -0.195 0.187 -0.941 0.960

MfrImage 0.084 0.926 -0.166 0.892

Variance 2.4758 1.8386 1.5460 5.8603

% Var 0.354 0.263 0.221 0.837

Factor 1: Product Qlty, Price level, no delivery speed, no price flexibility

A high score for Factor 1 corresponds to the attitudes of need for Prod Qlty (+) , need for Price Level (Competitive Pricing) (+), no need for delivery speed (-) and no need for PriceFlex (-). This factor denotes a no frills simple attitude - quality product at a competitive price.

**-Quality product at Competitive Pricing**

Factor 2: need SalesForce Image, Mfr Image

A high score for Factor 2 corresponds to the attitudes of need for SalesForceImage (+) and need for Mfr Image (+). This factor denotes a Brand Image conscious attitude.

-**Image Consciousness**

Factor 3: No Price level, no overall service

A high score for Factor 3 corresponds to the attitudes of no need for price level (-) and no need for overall service (+). This factor is confusing as it does not have any need. Hence it can be called as a factor denoting ‘no requirements’.

**-No Requirements**

**(B) Statistically different clusters**

**Descriptive Statistics: F1\_Avg, F2\_Avg, F3\_Avg**

Total

Variable KMeans Count Mean StDev Sum

F1 1 52 -0.8560 0.4894 -44.5131

2 48 0.9274 0.3968 44.5131

F2 1 52 -0.165 1.140 -8.572

2 48 0.179 0.796 8.572

F3 1 52 0.115 0.793 5.974

2 48 -0.124 1.180 -5.974

**Hypothesis test For F1:**

**Two-Sample T-Test and CI**

Sample N Mean StDev SE Mean

1 52 -0.856 0.489 0.068

2 48 0.927 0.397 0.057

Difference = μ (1) - μ (2)

Estimate for difference: -1.7834

95% CI for difference: (-1.9597, -1.6071)

T-Test of difference = 0 (vs ≠): T-Value = -20.08 P-Value = 0.000 DF = 96

The two clusters are significantly different based on Factor 1.

**Hypothesis test For F2:**

**Two-Sample T-Test and CI**

Sample N Mean StDev SE Mean

1 52 -0.17 1.14 0.16

2 48 0.179 0.796 0.11

Difference = μ (1) - μ (2)

Estimate for difference: -0.344

95% CI for difference: (-0.732, 0.044)

T-Test of difference = 0 (vs ≠): T-Value = -1.76 **P-Value = 0.082** DF = 91

The two clusters are NOT significantly different based on Factor 2.

**Hypothesis test For F3:**

**Two-Sample T-Test and CI**

Sample N Mean StDev SE Mean

1 52 0.115 0.793 0.11

2 48 -0.12 1.18 0.17

Difference = μ (1) - μ (2)

Estimate for difference: 0.239

95% CI for difference: (-0.164, 0.642)

T-Test of difference = 0 (vs ≠): T-Value = 1.18 **P-Value = 0.242** DF = 81

The two clusters are NOT significantly different based on Factor 3.

Hence the two clusters are statistically different only on Factor 1 (**Quality product at Competitive Pricing).**

**Exhibit 1:**

****

**Exhibit 2: Unrotated Factor Loadings**

**Loading Plot of Prod Qlty, ...,MfrImage**



**Exhibit 3: Rotated Factor Loadings**

**Loading Plot of Prod Qlty, ...,MfrImage**

