MKT 6337 Predictive Analytics Using SAS

Group 2

| Jie Zheng | Jun Zhou | Liu Yang | Anukrati Tiwari | Jenny Shah | Roger Kim |

Assignment 5

Awesome! Awesome! 30/30

Problem:

How to cluster customers based on transaction data for promoting more suitable sales strategies and improving sales? In order to know some useful information regarding customers and profits, we do cluster analysis and ANOVA test using TI's Transaction Data.

Analysis:

Yes!

To find the segment, we use the numeric variables in TI transaction data. However, three variables (TRAN_NETTOTI_VALUE, TRAN_RESALE_VALUE, COBM) are highly correlated which means only one variable could be chosen. To get a more accurate model, we added some new variables, such as total cost, unit cost, unit profit and unit margin of distributor kept, and standardized them to avoid over-scaling. (Please see Appendix for detailed formulas.)

cool.

Absolutely.

At first, we cluster TI's customer into six groups by using these new variables and order quantities. However, percentage of both two groups (cluster 1 and cluster 4) are less than 10%. According to the nearest rule, we merge cluster 1 into cluster 2 and cluster 4 into cluster 3. After merging, TI'S customer can be divided into four different clusters: 21.94% of TI's customers are fallen into cluster 2, 43.03% in cluster 3 ,23.15% in cluster 5 and 11.88% in cluster 6, respectively.

Next, we analyze features of each cluster and name them. The table below shows the given name and description of each cluster by integrating and analyzing the mean of unit cost, the unit profit, the unit of the margin the distributor kept, order quantities and total margin.

I love this. I don't know if it makes sense but you showed insight and initiative.

Cluster No	Cluster Name	Descriptions
Cluster 2	Core Customers	low profit rate with high demand, most of them located in Asia and Europe
Cluster 3	Valuable Customers	medial-high profit rate, high value of cost but low demand
Cluster 5	Third Party Customers	very low demand from resales, price of these customers is not sensitive. All of these customers are from distribution channel.
Cluster 6	Entry Products Customers	Highest profit rate (around 586%) with very low unit cost and medial demand quantity. These customers are looking for the components.

Yes, it makes sense.

After that we use the ANOVA test by *ACCOUNT CLASSIFICATION*, treat 74 noisy data as missing values and reduce the degree of freedom from 6 to 2. The results show that the profit to TI, unit cost, order quantity are not equal between salesperson assigned and salesperson not assigned. Besides, the transaction amount is declined from Q1 2017 to Q2 2017 but the mean of total profits are ascending as the order quantities increased.

I love the idea!

Finally, In order to find variable changing patterns through time, we compared the transaction frequency, the total profit, average net to PI revenue per unit, and average unit cost by each cluster and

Wow.

season. In 2017 Q2, the transaction frequency of cluster 2 and cluster 6 are reduced 8% and 42% respectively. While it of cluster 3 and cluster 5 are increased 6% and 4% respectively. The total revenue from customers are all increased except customers in cluster 6. The performance of revenue per unit and cost per unit of each cluster are different. Both revenue per unit of cluster 2 and cluster 3 are reduced with the cost per unit reduced. This lead to the order quantity growth up. The cost per unit of cluster 5 is raised while the revenue per unit does not change too much. For cluster 6: In the second quarter of 2017, the unit cost is increased by 23%, meanwhile the corresponding net price is raised by 38%. This is one reason for the decreasing of their total profit. The customers of this cluster are little bit price sensitive.

Recommendations:

For Core Customers: Slightly reduce the unit price if it can further reduce the unit cost, which can stimulate the purchase quantities from customers and increase the total profit. Discount is a good idea.

For Valuable Customers: Slightly reduce the unit price meanwhile increase the unit profit margin.

For Third Party Customers: Don't change price strategies. But can do further research to find if some of them are new customers. If they are, it may have some opportunities to increase their purchase volume in further. This kind of customer could become direct high value customers.

For Entry Products Customers: Due to the medium quantity and pervious price marked up in the 2nd quarter of 2017, it is better to reduce the profit rate to ensure customer purchase product from TI, and grow their purchasing volume.

At the end, TI should focus on core customer, Valuable Customer, and Entry products Customers, around 76.85% of total customers who place the order, in a sequence to improve their further performance.

ok.

Appendix

1. Formula of adding variables

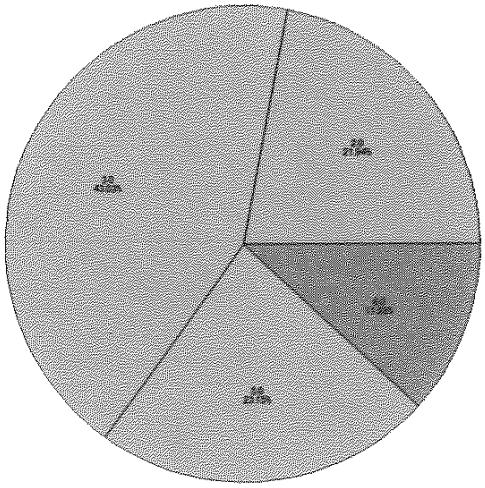
total_cost =TRAN_NETTOTI_VALUE -COBM;

unit_cost=total_cost/TRAN_QUANTITY;

Unit_Profit = COBM/TRAN_QUANTITY;

unit_DISTIMARGIN = DISTIMARGIN/TRAN_QUANTITY;

2. The pie chart for the percentage of each cluster in the TI transaction data



3. Anova test Table.

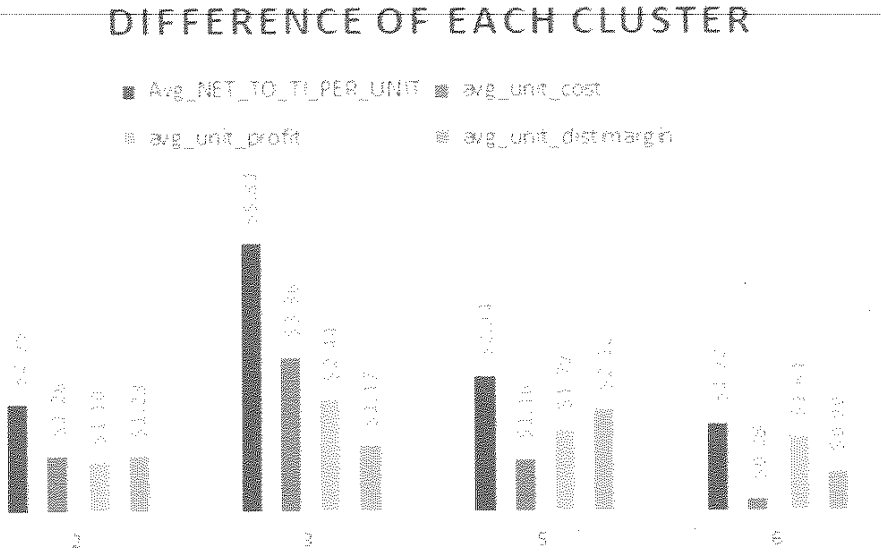
ANOVA TEST CLASS:ACCOUNTCLASSIFICATION

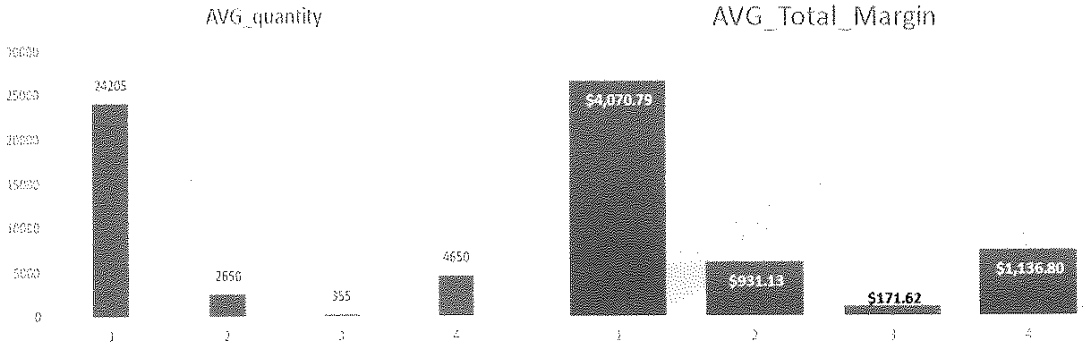
dependent variable	Sum of Squares	Mean Square	F value	Pr>F	R-square	Coeff Var	Roost MSE	Mean
COBM (profit)	6.08346E+12	3.04173E+12	10865.4	<.0001	0.010871	1139.362	16731.61	1468.508
Unit_Profit	323348	161674	107.11	<.0001	0.000108	2043.676	38.85189	1.901079
TRAN_QUANTITY	1.74796E+14	8.73981E+13	5929.32	<.0001	0.005962	1713.586	121408.3	7085.047
unit_cost	34035	17017	18.27	<.0001	0.000018	1507.392	30.51673	2.024473
unit_DISTIMARGIN	76619.2	38309.6	150.79	<.0001	0.000153	1046.802	15.93935	1.522671
Avg NET TO TI PER UNIT	147363	73681	78.36	<.0001	0.000079	1799.235	50.97132	3.92318

MKT 6337 Predictive Analytics Using SAS
Group 2
| Jie Zheng | Jun Zhou | Liu Yang | Anukrati Tiwari | Jenny Shah | Roger Kim |
Assignment 5

4. Performance comparison among clusters

Note: the label of bar chart shows 1,2,3,4 means cluster 2, 3, 5,6 respectively





TOTAL PROFIT

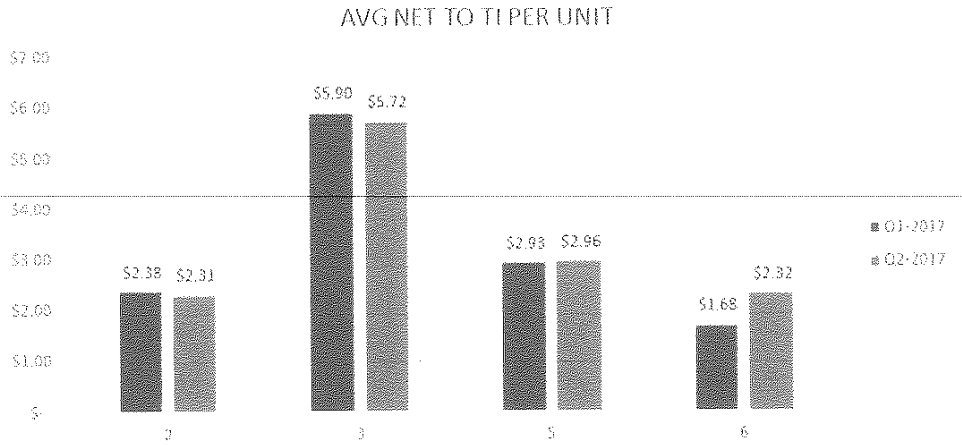
cluster	quarter	sum	mean	median	variance
2	Q1-2017	$ 842,218,110.00	$ 3,729.80	$ 241.00	848861452
2	Q2-2017	$ 923,450,184.00	$ 4,441.09	$ 264.40	995844889
3	Q1-2017	$ 391,056,573.00	$ 945.43	$ 64.55	159390235
3	Q2-2017	$ 401,084,605.00	$ 917.61	$ 56.85	108964676
5	Q1-2017	$ 37,788,355.49	$ 168.55	$ 3.37	7350288.88
5	Q2-2017	$ 40,767,630.93	$ 174.57	$ 3.22	38281368.44
6	Q1-2017	$ 147,839,947.00	$ 995.74	$ 25.55	96272952.57
6	Q2-2017	$ 119,253,110.00	$ 1,378.98	$ 42.25	212226167

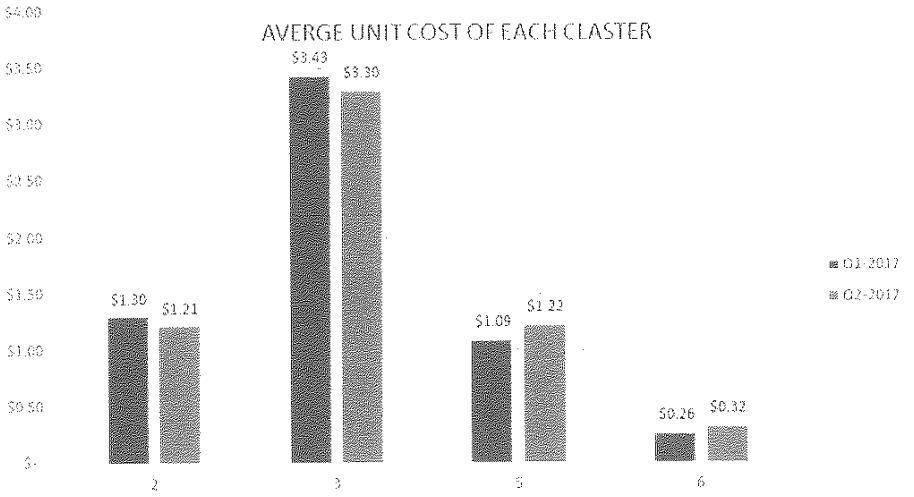
MKT 6337 Predictive Analytics Using SAS

Group 2

| Jie Zheng | Jun Zhou | Liu Yang | Anukrati Tiwari | Jenny Shah | Roger Kim |

Assignment 5





	TOP 2 frequency in MTL_MG1			Top 2 Frequency in Level 1 name			Sales Channel			Top 2 Account Category		
Cluster 1	Commodity Material	148602	34.3%	Europe	198990	45.9%	d	22774	5.3%	5	257648	59.4%
433741	Multi Source Material	101846	23.5%	Asia	165485	38.2%	resale	410967	94.7%	3	86560	20.0%
cluster2	Commodity Material	278851	32.8%	Americas	334420	39.3%	d	19181	2.3%	5	661231	77.7%
850728	Sole Source Material	225794	26.5%	Europe	319416	37.5%	resale	831547	97.7%	3	108037	12.7%
cluster 3	Commodity Material	152874	33.4%	Americas	215425	47.1%	d	0	0.0%	5	411682	89.9%
457728				europe	171695	37.5%	resale	457722	100.0%	3	24428	5.3%
cluster 4	Commodity Material	88949	37.9%	europe	88570	37.7%	d	2997	1.3%	5	179272	76.3%
234952				americas	78877	33.6%	resasle	231955	98.7%	3	28406	12.1%

MKT 6337 Predictive Analytics Using SAS

Group 2

| Jie Zheng | Jun Zhou | Liu Yang | Anukrati Tiwari | Jenny Shah | Roger Kim |

Assignment 5

5. Details in four clusters

a) 1 st Cluster

The FREQ Procedure

Cluster=2

MTL_MG1

MTL_MG1	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Commodity Material	148602	34.27	148602	34.27
Functional Equiv.	84036	19.38	232638	53.65
Multi Source Materia	101846	23.49	334484	77.14
Sole Source Material	99118	22.86	433602	100.00

Frequency Missing = 139

LEVEL1_NAME

LEVEL1_NAME	Frequency	Percent	Cumulative Frequency	Cumulative Percent
AMERICAS	46507	10.72	46507	10.72
ASIA	165485	38.15	211992	48.88
EUROPE	198990	45.88	410982	94.75
JAPAN	22759	5.25	433741	100.00

SALESCHANNEL

SALESCHANNEL	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Direct	22774	5.25	22774	5.25
Resale	410967	94.75	433741	100.00

ACCOUNTCATEGORY

ACCOUNTCATEGORY	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	11842	2.73	11842	2.73
2	52150	12.02	63992	14.75
3	86560	19.96	150552	34.71
4	25514	5.88	176066	40.59
5	257648	59.41	433714	100.00

Frequency Missing = 27

b) 2 nd Cluster

Cluster=3

MTL_MG1

MTL_MG1	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Commodity Material	278851	32.79	278851	32.79
Functional Equiv.	154967	18.22	433818	51.01
Multi Source Materia	190838	22.44	624656	73.45
Sole Source Material	225794	26.55	850450	100.00

Frequency Missing = 278

LEVEL1_NAME

LEVEL1_NAME	Frequency	Percent	Cumulative Frequency	Cumulative Percent
AMERICAS	334420	39.31	334420	39.31
ASIA	157771	18.55	492191	57.86
EUROPE	319416	37.55	811607	95.40
JAPAN	39121	4.60	850728	100.00

MTL_MG1

MTL_MG1	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Commodity Material	278851	32.79	278851	32.79
Functional Equiv.	154967	18.22	433818	51.01
Multi Source Materia	190838	22.44	624656	73.45
Sole Source Material	225794	26.55	850450	100.00

Frequency Missing = 278

SALESCHANNEL

SALESCHANNEL	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Direct	19181	2.25	19181	2.25
Resale	831547	97.75	850728	100.00

ACCOUNTCATEGORY

ACCOUNTCATEGORY	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	5625	0.66	5625	0.66
2	45303	5.33	50928	5.99
3	108037	12.70	158965	18.69
4	30509	3.59	189474	22.27
5	661231	77.73	850705	100.00

Frequency Missing = 23

MKT 6337 Predictive Analytics Using SAS

Group 2

| Jie Zheng | Jun Zhou | Liu Yang | Anukrati Tiwari | Jenny Shah | Roger Kim |

Assignment 5

c) 3 rd Cluster

Cluster=5

MTL_MG1

MTL_MG1	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Commodity Material	152874	33.40	152874	33.40
Functional Equiv.	92287	20.17	245161	53.57
Multi Source Materia	105678	23.09	350839	76.66
Sole Source Material	106809	23.34	457648	100.00

Frequency Missing = 80

LEVEL1_NAME

LEVEL1_NAME	Frequency	Percent	Cumulative Frequency	Cumulative Percent
AMERICAS	215425	47.06	215425	47.06
ASIA	53555	11.70	268980	58.76
EUROPE	171695	37.51	440675	96.27
JAPAN	17053	3.73	457728	100.00

MTL_MG1

MTL_MG1	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Commodity Material	152874	33.40	152874	33.40
Functional Equiv.	92287	20.17	245161	53.57
Multi Source Materia	105678	23.09	350839	76.66
Sole Source Material	106809	23.34	457648	100.00

Frequency Missing = 80

SALESCHANNEL

SALESCHANNEL	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Direct	6	0.00	6	0.00
Resale	457722	100.00	457728	100.00

ACCOUNTCATEGORY

ACCOUNTCATEGORY	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	875	0.19	875	0.19
2	14675	3.21	15550	3.40
3	24428	5.34	39978	8.73
4	6068	1.33	46046	10.06
5	411682	89.94	457728	100.00

d) 4 th cluster

The FREQ Procedure

Cluster=6

MTL_MG1

MTL_MG1	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Commodity Material	88949	37.87	88949	37.87
Functional Equiv.	41407	17.63	130356	55.50
Multi Source Materia	54673	23.28	185029	78.77
Sole Source Material	49867	21.23	234896	100.00

Frequency Missing = 56

LEVEL1_NAME

LEVEL1_NAME	Frequency	Percent	Cumulative Frequency	Cumulative Percent
AMERICAS	78877	33.57	78877	33.57
ASIA	39091	16.64	117968	50.21
EUROPE	88570	37.70	206538	87.91
JAPAN	28414	12.09	234952	100.00

MTL_MG1

MTL_MG1	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Commodity Material	88949	37.87	88949	37.87
Functional Equiv.	41407	17.63	130356	55.50
Multi Source Materia	54673	23.28	185029	78.77
Sole Source Material	49867	21.23	234896	100.00

Frequency Missing = 56

SALESCHANNEL

SALESCHANNEL	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Direct	2997	1.28	2997	1.28
Resale	231955	98.72	234952	100.00

ACCOUNTCATEGORY

ACCOUNTCATEGORY	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	1839	0.78	1839	0.78
2	16124	6.86	17963	7.65
3	28406	12.09	46369	19.74
4	9287	3.95	55656	23.69
5	179272	76.31	234928	100.00

Frequency Missing = 24