Segmenting Customers for Contract Plans

Business Analytics Using Data Mining

Submitted by: Mohali Mavericks (Team 4)

(Student name or group name)

Group Member Name	PG ID
Anu Kohli	61310639
Anupam Tripathi	61310471
Harshal Suthar	61310355
Namrata Raina	61310509
Naren Kolary	61310119

(Let us not waste paper, please continue writing your assignment from below)

Executive Summary:

This exercise is a continuation of the ideation project, wherein we had to identify a business goal from the mobile usage data available. Our business objective is to develop a model to profile the customers based on their usage patterns, the activities they indulge in and the preferred handsets in order to design a mobile handset based contract with service plans that will cater to their areas of interest. To realize our business objective, we need to apply unsupervised data mining techniques on given data to bring out clusters of people with similar data-usage attributes and mobile preferences. The clusters, so identified can then be targeted by service providers with customized data plans and contracts.

Business Objective:

The telecom industry is a highly competitive market, with frequent innovations required to sustain the business. The current business situation indicates a mismatch between user expectations and tariff plans available in the market. Furthermore, most service providers suffer from a high churn rate. The client (service provider) would therefore like to analyze the market and offer more customized plans that help increase the retention rate.

Service providers can collaborate with mobile handset manufacturers to sell contract based mobile handsets at a discounted price along with new and innovative product offerings to the customers. There is a plethora of service providers in the telecom industry today. In the industry, the customer switching costs are low and there are limited options to engage the client. For a service provider, having a well-defined idea about customer preferences can help them cement relationship with existing high value customers and attract new customers with tailored tariffs and promotions.

Such an exercise would allow service providers to target the right customers, offer them a bundled deal which has the highest value to them and also establish a long term relationship with the client. This would not only help them gain a loyal customer base but would give them an edge over competitors. By identifying different segments of customers on the basis of demographic data, usage patterns and handset preferences, the service provider can offer a series of contract based plans that matches exactly what the user's preferences are, while at the same time locking in the customer for a fixed period.

Data Preparation:

The churn dataset was used to figure out the usage pattern of customers and create segments that can be targeted byoffering different contract plans based on their preferences. The original dataset had 129 variables (columns) for 1859 customers (rows). Several of the 129 columns were redundant for the analysis. The columns were diligently analyzed to zero on a set of 28 columns that seemed most reasonable for the analysis. The selection of the columns was based on the critical factors that are most relevant to users. The columns that were used for the analysis are referenced below:

pref.Apple	Monthly.expenditure.on.mobile.service	Average.use.of.Value.Added.Services..VAS...Online.games
Pref.Blackberry	Average.minutes.per.day.voice.calls.	Average.use.of.Value.Added.Services..VAS...SMS.MMS
Pref.HTC	Average.SMSes.per.day	Average.use.of.Value.Added.Services..VAS...music.video.downloads
Pref.Karbonn	Average.use.of.Value.Added.Services..VAS...Caller.Tunes	Average.use.of.Value.Added.Services..VAS...Document.Reader..pdf..word.etc..
Pref.Lava	Average.use.of.Value.Added.Services..VAS...Ringtone.downloads	Current.data.plan
Pref.LG	Average.use.of.Value.Added.Services..VAS...E.mail.checking	Age
Pref.Micromax	Average.use.of.Value.Added.Services..VAS...ocial.networking	Gender
Pref.Motorola	Average.use.of.Value.Added.Services..VAS...Cricket..news.or.stock.alerts	Yearly.household.income
Pref.Nokia	Average.use.of.Value.Added.Services..VAS...Jokes..astrology.etc.	
Pref.Samsung	Average.use.of.Value.Added.Services..VAS...GPS.facility	

In addition, to increase the accuracy of the analysis, rows containing "blank" and "NA" valueswere deleted. The following transformations were performed on the dataset containing the aforementioned 28 columns:

1. Text values of the columns were segregated into numerical values to reduce the complexity of the analysis. The handset preference variables (column 1 to column 10) were categorized into 0 and 1 based on whether a person preferred

a particular handset. “Transform categorical data” feature of XLMiner was used to perform this transformation.

2. The average minutes per day of voice calls (item 12) were binned as follows –
   a. Bin 1 = less than 10 min per day
   b. Bin 2 = 11-30 min
   c. Bin 3 = 31 – 60 min
   d. Bin 4 = 1 to 2 hours
   e. Bin 5 = more than 2 hours
3. The average SMSes per day (item 13) were binned as follows –
   a. Bin 1 = None
   b. Bin 2 = 1 – 10 per day
   c. Bin 3 = 11 – 20 per day
   d. Bin 4 = More than 20 per day
4. The data for the average use of value added services (Items 14 through 24) was simplified and the data was simply binned as 0 and 1 based on whether the customer used the specific value added service or not.
5. The current data plan usage by the customers (item 25) was binned as follows –
   a. Bin 1 = 0 – 500MB
   b. Bin 2 = 500MB – 1GB
   c. Bin 3 = 1GB – 5GB
   d. Bin 4 = 5GB – 12GB
6. The gender data (item 27) was binned as follows –
   a. Bin 1 = Male
   b. Bin 0 = Female
7. The data on the monthly expenditure on mobile service, age as well as yearly household income were left as continuous variables for more insight into the data. All of the data was normalized before running the analysis.

Data mining method:

After transforming the data, a standard k-means clustering was performed on the data. The data was normalized prior to the clustering. Initially, 8 clusters were created. On analyzing the clusters, it was determined that there was too much variability amongst the clusters to segment customers on the basis of that. Next, 5 clusters were created. 5 clusters seemed adequate for the segmentation. Below this level there did not seem to be enough variability to create sensible clusters.

Customer Segmentation

Cluster	Apple	Blackberry	HTC	Karbonn	Lava	LG	Micromax	Motorola	Nokia	Samsung
Cluster-1	0.851485	0.55198	0.45297	0	0.007426	0.032178	0.019802	0.089109	0.403466	0.685644
Cluster-2	0.811502	0.792332	0.255591	0.003195	0.003195	0.019169	0.022364	0.047923	0.680511	0.41214
Cluster-3	0.99308	0.982699	0.982701	0.961938	0.979238	0.965397	0.948097	0.958479	0.99654	0.989619
Cluster-4	0.633621	0.482759	0.25	0.010776	0.006466	0.075431	0.056034	0.122845	0.732759	0.640086
Cluster-5	0.813559	0.559322	0.350282	0.016949	0.022599	0.084746	0.067797	0.107344	0.632768	0.610169

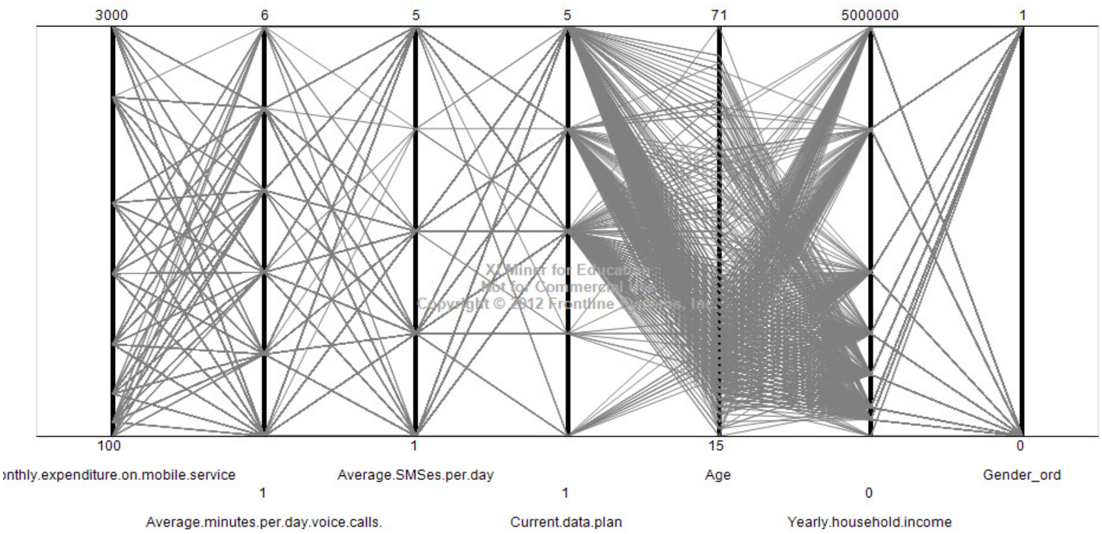
Cluster	Monthly.exp	Avg.mins.per.day	Avg.SMSes.per.day	Caller.Tunes	Ringtone.downloads	E.mail.checking	Social.networking	Cricket.news.or.stock.alerts	Jokes.astrology.etc.
Cluster-1	1341.46022	2.641089	2.571782	0.096535	0.002475	0.972772	0.935644	0.653465	0.009901
Cluster-2	1015.654986	3.00639	2.546325	0.169329	0.035144	0.923322	0.859425	0.182109	0.025559
Cluster-3	933.218391	2.66782	2.778547	0.249135	0.117647	0.695502	0.653979	0.422145	0.148789
Cluster-4	699.461251	2.689655	2.974138	0.118535	0.034483	0.090516	0.051723	0.079741	0.021552
Cluster-5	1275.988649	2.790961	2.384181	0.497175	0.446326	0.920904	0.915254	0.898307	0.790962

Cluster	GPS.facility	Online.games	SMS.MMS	music.video.downloads	Document.Reader.pdf.word.etc.	Current.data.plan	Age	Yearly.household.income	Gender
Cluster-1	0.886138	0.289604	0.955446	0.470297	0.913366	3.415841	28.908415	1917574.3	0.85396
Cluster-2	0.290735	0.070288	0.859425	0.111821	0.402556	4.121406	27.271567	1301597.4	0.507987
Cluster-3	0.49827	0.231834	0.885813	0.380623	0.557093	4.034602	26.989622	1191868.5	0.647059
Cluster-4	0.06681	0.032328	0.711207	0.068965	0.056034	4.743535	31.010779	1139655.2	0.62931
Cluster-5	0.870057	0.638418	0.966102	0.841808	0.830509	3.683616	28.468926	1591807.4	0.751413

Cluster Statistics

Cluster	#Obs	Average distance in cluster
Cluster-1	405	4.163
Cluster-2	312	4.178
Cluster-3	289	4.297
Cluster-4	464	4.161
Cluster-5	177	4.816
Overall	1647	4.259

Parallel Coordinates Plot



Conclusions and Recommendations:

The centroid values were then used to deduce measurable patterns and assign managerially relevant names. The five segments into which the cluster values were segregated are:

1. **High Income Business Users**: This group comprises high income people who have an affinity towards iPhones and incur a high monthly expenditure on mobile services. Such customers can be targeted by offering contract based iPhones with data plans having high speed connectivity as a part of the contract.
2. **Blackberry Office Users**: This group of customers prefers Blackberrys and uses it for email and voice calls. The best way to target this class of customers is by offering email packages on Blackberry.
3. **Handset Agnostic Users**: This group is agnostic to the brand of the phone. Contract based phones (phones with data plans from network providers) would probably not be a good option for them.
4. **Value Maximizers:** This segment has the most prolific users of data plans but spend the least on phone bills. An effective way of targeting these customers would be selling low rate GPRS plans bundled with free SMS service. Since they are heavy users of GPRS services they can be offered plans in which the rates are low for high usage.
5. **Online Generation:** This group consists of new generation users who are highly socially active on their phones and as such spend a lot on mobile services. Such customers can be targeted by selling high speed data plans in the form of social networking packages. The packages can have unlimited access to social networking sites on payment of a fixed rental fees every month.