Micro-segmentation (clustering) and campaign management in telecoms:

A case study using two approaches via SPSS and Tanagra
Ultimately, the sole job of the CMO is to increase the level of Revenue for the company by the development of customer usage, while at the same time enhancing and improving the positioning and image of the brand. This usually means creating unique customer experiences that ensure retention and contribute to positive relationships between the customer and the company in a development perspective. The marketing campaigns involved by these objectives can not be effective without a segmented approach that only can make it possible to understand the needs and behaviour of all the customers. And then it will become possible to really adapt the offers to the individual needs identified in order to develop cross- and up-sell.
The case study developed here concerns a telecom company that addresses customers needs via 4 tariff plans.

While calculating the variance of some basic indicators (ASPU, MOU per user, tenure, etc) for each tariff plan, we observed strong variability. This strong dispersion means that each tariff plan involves different customer usage profile; hence the decision to perform a micro-segmentation of the base in order to identify more homogeneous clusters to address with specific offers.

The case study will be developed here with the help of two approaches:

• A classic clustering aiming at identifying the micro-segments. This approach will be completed with the deployment of the model via a dynamic scoring of the subscribers and a campaign management.

• Another approach will target a key variable (price sensitivity) in order to produce business rules that make it possible to affect a subscriber in a per-determined cluster.

For this case study, we selected a sample of 4,230 subscribers described via 14 generic descriptive variables (see listing below).
Segmentation in action

1. Classic data mining approach: clustering based

This approach makes use of clustering technique of data mining in order to build micro-segments that will be addressed with specific offers. The case study will primarily be treated with the help of IBM SPSS Modeler.

1.1. Process-flow of the project (Segm29-JayJay)

The process flow of micro-segmentation building as performed with IBM SPSS Modeler is depicted in the chart below. The 5 big steps are then:

• Importation of the data set and definition of the role of the variables. Note that three additional variables have been constituted by combining existing generic variables: ratio Peak/Off Peak, ratio SOI/SOR and Price_sensitivity.
• Preparation of the data with the help of filter node. For that purpose, we sorted the issues regarding the imputation of missing values and variable transformation in the same frame of mind as we did in the Churn prediction project (see page 300). Nevertheless concerning variable selection, we decided to involve all the variables in the construction of the clusters. But it is recommended to select the variables that have the biggest predictor importance when the dataset is huge.
• Application of various clustering techniques. Here we have selected K-Means. Note that IBM SPSS Modeller offers the possible to select Kohonen-SOMs in order to perform the same task (*). We performed various simulations in order to get the optimal number of clusters regarding rules like minimum size of the cluster, homogeneity, etc. At the end, despite the existence of 4 tariff plans, we identified 18 specific clusters.
• Description, visualization and summarization of the clusters. The segments identified were describe using the classic criteria: size, dispersion, proximities, business rules. In addition, we performed a Principal Components Analysis in order to summarize and visualize the 18 clusters.
• Test of the model. At this stage, we determine the accuracy of the model.
Segmentation in action

1. Classic data mining approach: clustering based
1.2. Outcomes of micro-segmentation

The fact that K-Mean algorithm has been able to identify 18 clusters from 4 tariff plans is an evidence of the existence of some usage profiles that need to be addressed specifically. The 18 segments are described in the sheet below produced under IBM SPSS Modeler. This description concerns the size of each segment as well as the distribution of each predictor (mean, standard deviation, skewness, etc) for the cluster selected.
Segmentation in action

1. Classic data mining approach: clustering based

Other outcomes may concern the proximity between the segments. As we can see in the first chart below:
• Customers of cluster 18 (sky blue circle) are mainly in tariff plan Business, their ASPU stands between $20 and $60, and they are low price sensitive.
• Customers of cluster 2 (violet square) are very far from those of cluster 18. They are mainly youngsters, very price sensitive, and their ASPU is below $20.

From the second chart, we can conclude for instance that: “Calls receivers (those who receive more calls than they make) are mostly low end users, they are in tariff plan Per Second”. 
1. Classic data mining approach: clustering based

1.3. Business rules and Accuracy of the model.
In the three outcomes superimposed below we can observe what follows.

• The outcome at the left hand side presents the business rules that serve in the affectation of the subscribers in the segments. For instance, we can point out that: Cluster 13 has 1 rule. If a customer is in tariff plan Per Second, and has an ASPU above $10.002, and has a ratio Peak/Off Peak above 200.017 (he/she makes twice as many calls in Peak compared with Off Peak); then he/she belongs to cluster 13. It is important to notice that a given cluster may contain several different business rules.

• The second outcome (in the middle) aims at ranking the descriptive variables according to the importance in defining the segments. As we can see, the top 6 variables are: ratio SOI/SOR, total value ASPU, price sensitivity, ratio Peak/Off Peak, the offer (tariff plan), the number of outgoing calls. If the analyst intents to perform further analyses, he/she might select these 6 variables only, and get accurate results while saving time and other resources.

• The third outcome (right hand side) presents the results of the test of the clustering model built via K-Means. IBM SPSS Modeler performed various iterations using in order to measure the accuracy of the model in terms of classification power. With an error rate of 4.68%, we can conclude that the model is satisfactory.
1. Classic data mining approach: clustering based

The approach adopted makes it possible to detect two kinds of outliers leading to two fundamental issues in telecoms:

- Customers whose usage pattern does not match with their tariff plan; for instance a subscriber in Classic Per Minute who make essentially very short calls (less than 30 seconds). This is a Marketing issue.
- Customers whose usage patterns lead to a loss of revenue for the company (fraudsters); for instance a subscriber with a very high “outgoing_calls” but for whom “total_value_aspu” is almost nil. This is a Revenue Assurance, Customer Relations and Finance issue.
1.4. Deployment of the model: Dynamic scoring and campaign management

In a dynamic environment like the one we observe in Africa and the Middle East regarding Prepaid customers, the scoring (affecting new customers to specific clusters) needs to be updated almost in real-time and to be used for regular and specific tactical campaigns. Dynamic scoring allows you to score an already-defined customer segment within your Campaign Management. Dynamic scoring both avoids mundane, repetitive manual chores and eliminates the need to score an entire database.

Once the model is in the Campaign Management system, a user (usually Segment manager or any other person than the business intelligence analyst who created the model) can start to build marketing campaigns using the predictive models. Models are invoked by the Campaign Management System.

The diagram below depicts the process.

When a marketing campaign invokes a specific predictive model to perform dynamic scoring, the output is usually stored as a temporary score table. When the score table is available in the data warehouse, the Data Mining engine notifies the Campaign Management system and the marketing campaign execution continues.

The table next page (exportation from IBM SPSS Modeler to Microsoft Excel) presents the scoring of the subscriber base. We can read for instance that: subscriber #2 (MSISDN 20823300756 and in tariff plan Classic Talk) should belong to cluster 6 with a probability of 0.9365.
Segmentation in action

1. Classic data mining approach: clustering based
Here is how a dynamically scored customer segment might be defined for our example:

Where
Offer = Per Second
And
Total_value_ASPU > 10.002
And
Ratio_Peak/Offpeak > 200.017
In_Model(Segm29.JayJay).score > 0.80

For more clarification:
Offer = Per Second limits the application of the model to those customers in tariff plan Per Second.

Total_value_ASPU > 10.002 selects from the first split above only customers spending, on average, more than $10.002 each month.

Ratio_Peak/Offpeak > 200.017 selects from this second split only customers making, on average, at least twice as much calls on Peak period than on Off Peak. The marketer deemed that it would be unprofitable to send the offer customers.

Segm29.JayJay is the name of the logged predictive model that was created with a Data Mining application. This criterion includes a threshold score, 0.80, which a customer must surpass to be considered "in the model." This third criteria limits the campaign to just those customers in the model, i.e. those customers most likely to require a specific bonus and/or rewarding in order to boost the usage.
This approach is based on the identification of a target variable in order to identify the segments to address. Here, Price_Sensitivity is the target variable. In order to go ahead with our review of the various software which can serve for analytics, the case study will be sorted with the help of TANAGRA.

2.1. Process-flow of the project

The process appears in detail on the left hand side of the diagram on page 305. We will not insist on the phase of data pre-processing here as the reader is already comfortable with such tasks. The activities we will focus on are the following.

1. Summarisation of the variables in order to observe the proximities as well as the correlations between them. In this regard, we decided to choose Multiple Factor Correspondence Analysis, the aim of which is to represent in a 2-dimension map the categorical variables. Actually, what we are looking for is to detect which values of the predictors are correlated to a given value of the target variable: price_sensitivity. The four qualitative predictor used here are: offer, ratio SOI/SOR, value band and ratio Peak/Offpeak.

2. Clustering in order to identify the segments the price sensitive subscribers belong to. For that, we selected Kohonen- SOMs, and we set the number of clusters to build at fifteen. One of the advantages of Tanagra is that this software is able to produce business rules from most of the data mining models.
3.2. Outcomes of the project

The first outcome to present concerns the proximity between the values of the five categorical variables (one target variable and four predictors). As we can see in the map below:

- Subscribers who are strongly price sensitive belong mainly to the offer Classic_Talk and a little bit to Per-Second.
- Most of these subscribers are medium end users, and they tend to receive a few more calls than they make.
Segmentation in action

2. Other data mining approach: supervised learning based
2. Other data mining approach: supervised learning based

The second outcome provides more actionable results as we can get business rules. Here, these business rules take the shape of the so-called English Rules. But it is also possible to get a decision tree; we can see in the diagram that such a tree would have 29 nodes and 15 leaves.

As our goal is to identify the segments in which the price sensitive customers belong (so that we can get a significant increase of revenue from our marketing tactical initiatives), we have to look for these interesting rules. Let us set the price sensitivity threshold at 1.25 for instance. This means that if we implement a drop of price of 10%, the usage of these subscribers should increase by at least 25%.

As we can see in the diagram on page 305, with this threshold we can then identify three interesting segments:

- The first one is cluster 8, with a price sensitivity of 1.281.
- The second one is cluster 10, with a price sensitivity of 1.278.
- The third interesting group of subscribers is cluster 14, with a price sensitivity of 1.457.

These clusters are clearly described in the table below. And the final step regarding the data mining project will be the dynamic scoring of the current as well as the new subscribers in order to implement the campaign management programme.

<table>
<thead>
<tr>
<th></th>
<th>Cluster 1</th>
<th>Cluster 2</th>
<th>Cluster 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASPU</td>
<td>From $7.5 to $12.5</td>
<td>From $7.5 to $12.5</td>
<td>Above $7.5</td>
</tr>
<tr>
<td>Offer</td>
<td>Per Second</td>
<td>Business, Classic Talk, Per Second</td>
<td>Business</td>
</tr>
<tr>
<td>Call length</td>
<td>Below 45.5 seconds</td>
<td>Above 45.4 seconds</td>
<td>Above 61.5 seconds</td>
</tr>
<tr>
<td>Ratio SOI/SOR</td>
<td>Below 134</td>
<td>Below 96</td>
<td>Below 134</td>
</tr>
<tr>
<td>Price sensitivity</td>
<td>1.281</td>
<td>1.278</td>
<td>1.457</td>
</tr>
</tbody>
</table>