PART A (40%) A segmentation based exploration of customers in the churn case study

Carry out an exploratory analysis to try and understand who these customers are and whether they have any behavioral patterns and tendencies which could be made use of. Although this analysis will not be directly linked to the earlier predictive analytics exercise, the results may provide useful incites when making decisions based on the predictive analysis results.

Add another copy of the churn_telecom data source to the churn case study diagram (you may use another new diagram). Go to the meta-data page and change variable roles to input (change the rejected ones to input), other than the ID roles.

Add a cluster node and a segment profile node to the diagram. Link the data source to the cluster node and the cluster node to the profile node.

Carry out the following clustering and profiling activities and report outcomes. (it is important to note that this is an exploration of the customer data set. Therefore there will be no correct or incorrect result. What is expected is a report on findings and where appropriate suggestions on possible value).

Open the variable information page of the cluster node (from the properties list). Since we are planning to conduct a cluster analysis using a limited number of variables, change the ‘use’ column to ‘no’ for all variables.

1. Carry out a demographics based profiling. Change the use of variables age, gender and customer value to ‘default’ (we will take customer value as a demographics variable although this may not be so – insufficient demographics in the data). Run the cluster node and see results. What can you say about the demography based segments. Run the segment profile node and comment on the results.

2. Include some customer status based information in the analysis – eg: tenure on network, no. of active services, total profitability of subscription, no of emails, internet/fix line revenue etc (use at least 3 variables). Run the cluster node and the segment profile node and discuss the outcomes. (Do you see any understandable groupings/segments?)
3. Remove the initial variables and carry out a cluster analysis of usage information with variables such as average number of outgoing calls, incoming calls, number of local/international calls etc (use at least 4 variables). **Note- we can further carry out cross cluster analysis to link these segments to segments from previous analysis in 1 and 2 above but will not be required for this assignment.**

Prepare a report (maximum 3 pages) based on the outcome of the first 3 steps. You may include screen shots of results and point out the variables of significance. The report must have a section discussing the potential value of these results when taking action based on a churn prediction and survival analysis.

**PART B – Discussion on the practical use of clustering, segmentation and profiling 20%**

Study the Roy Morgan value segments and Experian demographics segments and profiles in
http://www.roymorgan.com/products/values-segments

Using your knowledge on customer segmentation, clustering as well as module 2 lectures:

1. At what stages of a predictive analytics exercise will such information be useful?
2. How will you relate customer segments identified to Roy Morgan, Experian segments – how will this information be put to practical use?

**Search for suitable references – you are expected to provide at least 2 references. Max 1 page.**

You may use the following as reference material:

**PART C – Market Basket Analysis 40%**

In order to plan innovative promotions to move items that are often purchased together, a store is interested in market basket analysis of items purchased from the Health and Beauty Aids Department and the Stationary Department. The store chose to conduct a market basket analysis of specific items purchased from these two departments. The **TRANSACTIONS** data set contains information about more than 400,000 transactions made over the past three months. The following products are represented in the data set:

- bar soap
- bows
- candy bars
- deodorant
- greeting cards
- magazines
There are four variables in the data set:

<table>
<thead>
<tr>
<th>Name</th>
<th>Model Role</th>
<th>Measurement Level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>STORE</td>
<td>Rejected</td>
<td>Nominal</td>
<td>Identification number of the store</td>
</tr>
<tr>
<td>TRANSACTION</td>
<td>ID</td>
<td>Nominal</td>
<td>Transaction identification number</td>
</tr>
<tr>
<td>PRODUCT</td>
<td>Target</td>
<td>Nominal</td>
<td>Product purchased</td>
</tr>
<tr>
<td>QUANTITY</td>
<td>Rejected</td>
<td>Interval</td>
<td>Quantity of this product purchased</td>
</tr>
</tbody>
</table>

a. Create a new diagram. Name the diagram Transactions.

b. Create a new data source using the data set ABA1.TRANSACTIONS.

c. Assign the variables STORE and QUANTITY the model role Rejected. These variables are not used in this analysis. Assign the ID model role to the variable TRANSACTION and the Target model role to the variable PRODUCT. Change the data source role to Transaction.

d. Add the TRANSACTIONS data set and an Association node to the diagram.

e. Change the setting for the Export Rule by ID property to Yes.

f. Leave the remaining default settings for the Association node and run the analysis.

g. Examine the results of the association analysis.

1. What is the highest lift value for the resulting rules
2. Which rules have this value?
3. What is the significance of the lift value of a rule – explain using an example from the case study.
4. What is the relationship between lift, support and confidence values – describe using an example.
5. Based on the association rules, briefly describe 3 example product bundles and promotions that you might suggest?