Retail Customer Segmentation using SAS

April 2014
Calgary SAS Users Group meeting

Jenny Chen
Data Science, LoyaltyOne
Agenda

• Overview
• Applications
• Objectives
• Types of Segmentation
• A Real Example with SAS Code
• Further Reading
Overview

- **Customer Segmentation** is the practice of classifying your customers into distinct groups based on the similarities they share with respect to any characteristics you deem relevant to your business.

- **Key components in developing proper, actionable segmentation**
  - Understand business needs and objectives
    - Customer satisfaction
    - Share of wallet
    - Market share
  - Available customer information
    - Demographic
    - Geographic
    - Behavioral
    - Attitudinal
## Applications

<table>
<thead>
<tr>
<th>Applications</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Customer Investment Allocation</td>
<td>• Determine segment’s current and potential profitability</td>
</tr>
<tr>
<td></td>
<td>• Define differentiated value proposition by segment</td>
</tr>
<tr>
<td>Customer Relationship Management</td>
<td>• Better meet customer needs, improve customer satisfaction</td>
</tr>
<tr>
<td></td>
<td>• Develop customer loyalty to your brand</td>
</tr>
<tr>
<td></td>
<td>• Improve retention strategy</td>
</tr>
<tr>
<td>Tailor Marketing Program</td>
<td>• Build relevant campaign, communication</td>
</tr>
<tr>
<td></td>
<td>• Optimize marketing channel mix</td>
</tr>
<tr>
<td>Guide New Product Development</td>
<td>• Product preference</td>
</tr>
<tr>
<td></td>
<td>• Potential targeting audience</td>
</tr>
<tr>
<td>Guide Product, Service Pricing</td>
<td>• Price sensitivity</td>
</tr>
<tr>
<td></td>
<td>• Stimulate customer demand</td>
</tr>
</tbody>
</table>

…and many more
Objectives

Segmentation enables business to increase profitability while positively addressing customer needs

Create Relevancy
- Better defined target universe
- Right offer, communication, channel, time for right customers
- Personalized experience between clients and their customers

Change Behavior
- Higher response rates
- Combat attrition of high profit customers
- Grooming lower current value high potential customers
- Identify cross-sell & up-sell opportunities by customer

Improve Financials
- Measureable improvement in sales, traffic, revenue
- More cost effective campaign deployment and ROI
- Opportunity identification by category, department, store, etc.
Segmentation Type

Data source
• Demographic, geographic
• Behavior, attitudinal
• Life stage, life style

Approach
• Business rules
  - Profit ranking
  - RFM (Recency, Frequency, Monetary)
• Supervised clustering
  - Decision tree
• Unsupervised clustering
  - K-means clustering
Using unsupervised clustering segmentation for a grocery chain which would like better product assortment for its high profitable customers

**Potential Inputs**

- **Value**
  - Basket Size
  - Visit Frequency

- **Basket**
  - Spend by category
  - Type of category
  - Brand spend (i.e. private label)

- **Promotions**
  - % bought on targeted promotion
  - % bought from flyer

- **Time**
  - Time of day
  - Day of week

- **Location**
  - Store format
  - Area population density

**Deal Seeking Mom**

**Key Differentiators**

- Full store shop
- High avg. basket size / # trips
- High % purchased on promotion
- Rewards seeker
- High spend categories
  - Fresh produce
  - Organic food
  - Multipack juice, snack

**Clustering approach**
Segmentation Building Process I

Define business objective
- Identify high profitable customer groups
- Improve product assortment for them
- Design relevant, effective marketing campaigns for them
- Quantify campaign measurements

Gather information
- Customer transactional data
- Promotional data
- Client store, department, category merchandising hierarchy

Align with client on project scope
- Business Objectives
- Data Sources
- Methodology
- Measurement plan
Segmentation Building Process II

Data preparation

- Data merge purge, data cleaning, dealing with missing
- Convert all attributes to numeric metrics, cap outliers
- Standardize all metrics/variables
- Standardization enables each metric have similar contribution to the future cluster building

```sas
PROC STDIZE DATA=SAS_DATA.ALL_INFO
   OUT=SAS_DATA.STANDARD_STD
   OUTSTAT=SAS_DATA.SUMMARY
   (WHERE=( _TYPE_ IN ('LOCATION', 'SCALE')));
VAR &ALL_VAR;
RUN;
```
Data exploration, variable reduction

- Check descriptive statistics, correlation among all potential variables
- Use principle component, factor analysis (**PROC PRINCOMP, PROC FACTOR**)
- Use the Variable Clustering node in SAS Enterprise Miner to create variable cluster constellation plot and variable cluster tree diagram

Measure similarity among customers

- Using Euclidean distance - this measures the distance between 2 points in multi-dimensional space
- A very common method of unsupervised clustering with large data using Euclidean distance is K-means clustering
K-means clustering

- Using the distances to group customers into K clusters where each customer is with the nearest centroid
- The centroid is calculated as the multi-dimensional set of the means of the variables used for the particular cluster
- Pre-determine a range of number of clusters, use bottom-up approach
- Test the FASTCLUS procedure repeatedly using different starting points and different number of clusters until we obtain stabilized centroids and desired distance between clusters

```
PROC FASTCLUS DATA=SAS_DATA.STANDARD_STD MAXCLUSTERS=29
OUT=SAS_DATA.RESULT OUTSEED=SAS_DATA.CENTRES OUTSTAT=SAS_DATA.CLUSTER_STAT REPLACE=RANDOM RANDOM=187413849 CONVERGE=0.001 MAXITER=350;
VAR &VAR_LIST;
RUN;
```
Output of K-means Clustering

- Centroids info

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Var1</th>
<th>Var2</th>
<th>Var3</th>
<th>Var4</th>
<th>Var5</th>
<th>Var6</th>
<th>Var7</th>
<th>Var8</th>
<th>Var9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.68</td>
<td>0.28</td>
<td>0.37</td>
<td>0.20</td>
<td>0.77</td>
<td>0.46</td>
<td>0.29</td>
<td>1.11</td>
<td>1.70</td>
</tr>
<tr>
<td>2</td>
<td>0.81</td>
<td>0.91</td>
<td>2.44</td>
<td>0.19</td>
<td>0.93</td>
<td>0.95</td>
<td>0.41</td>
<td>0.62</td>
<td>0.56</td>
</tr>
<tr>
<td>3</td>
<td>0.68</td>
<td>0.62</td>
<td>0.73</td>
<td>0.44</td>
<td>0.77</td>
<td>0.67</td>
<td>0.53</td>
<td>1.11</td>
<td>1.66</td>
</tr>
<tr>
<td>4</td>
<td>0.49</td>
<td>0.54</td>
<td>0.45</td>
<td>0.45</td>
<td>0.51</td>
<td>0.33</td>
<td>0.48</td>
<td>0.46</td>
<td>0.56</td>
</tr>
<tr>
<td>5</td>
<td>0.17</td>
<td>0.04</td>
<td>0.10</td>
<td>0.02</td>
<td>0.13</td>
<td>0.73</td>
<td>0.70</td>
<td>0.43</td>
<td>0.55</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>0.68</td>
<td>0.68</td>
<td>0.84</td>
<td>0.47</td>
<td>0.77</td>
<td>1.64</td>
<td>0.58</td>
<td>2.27</td>
<td>2.40</td>
</tr>
</tbody>
</table>

- Statistics for variable comparison and further variable reduction
- RSQ/(1-RSQ) is the ratio of between cluster variance to within cluster variance
- Get client’s buy-in for the final list of variables for cluster building
Segmentation Building Process V

**Bottom-up approach**

- Flexible with how many clusters to choose initially
- Aggregate clusters up afterwards using hierarchy clustering `PROC CLUSTER`
- Create a tree diagram for clusters using `PROC TREE`
- Choose proper distance to fit into client’s request of final number of segments and distribution of segments

```plaintext
PROC CLUSTER DATA=SAS_DATA.CENTRES METHOD=AVE OUTTREE=TREE;
VAR &VAR_LIST;
RUN;

PROC TREE DATA=TREE;
RUN;
```
Bottom-up Approach

Tree diagram to illustrate the arrangement of the clusters

Seg1
Seg2
Seg3
Seg4
Segmentation Building Process VI

Segments profiling

- Using cluster building variables to profile the segments
- Using additional data source to profile them as well
- Paint a clear picture of your segments

Validate segmentation effectiveness

- Test campaigns with segments strategy
- Measure campaign results by segments
- Validation segments with market research survey

Review, iterate, deploy again

- Update segments regularly
- Monitor the migration of segments
- Gather feedback, campaign response results
- Maintain and improve the process
Pros and Cons of K-means Clustering

Pros

• Flexible with fewer assumptions about your customer population
• Can incorporate many relevant metrics
• Provide more comprehensive, diverse segments
• Simple algorithm fast to run with large dataset
• Prove to work well for behavioral targeting

Cons

• Purely data driven
• Pre-specify number of clusters or the minimum distance between clusters
• Time consuming to maintain: for example, certain metrics used to build the clusters were changing overtime dramatically
• May have difficulty to explain the concept to your marketing clients
Question or Comment
Reference

SAS website

http://support.sas.com/documentation/
http://support.sas.com/resources/papers/
http://support.sas.com/events/sasglobalforum/previous/online.html


A video about LoyaltyOne: Enriching Relationships
http://www.youtube.com/watch?v=NSed-m3Z8Ik