Data Driven Customer Segmentation

A whirlwind tour.

Jim Porzak
Data Science for Customer Intelligence (DS4Cl.org)
Global Predictive Analytics Conference
Santa Clara, CA
May 5, 2015
Why Segment Customers?

- For strategic and tactical business decisions
  - Strategic
    - Segments fairly stable over time
    - Conceptually high level
  - Tactical
    - Reflect customers current state
    - Actionable today – in the weeds
How to Segment?

“We’ve broken your list into eighty-four subgroups. Our work here is done.”

DS$_4$CI.org
Segmentation Examples

From simple to more complex:

- Customer Tenure
  - A really simple concept – everyone “gets it”
  - Can have huge impact on an organization

- RFM Based Segmentation
  - Not your grandfather’s RFM!
  - Intuitive & actionable segments

- Customer Classification
  - Based on ~ static properties (preference or properties)
  - Use unsupervised clustering

- Customer Life-stage Segments (a preview)
Customer Tenure Based Segments
Customer Tenure Segments

- Newbie
- Evolving
- Mature
- Core
Customer Tenure Segments

With Marketing Actions

- Newbie
- On-board
- Evolving
- Nurture
- Mature
- Retain
- Core
- Reward
Customer Tenure Segments

How to Determine Break Points?

Newbie → Evolving → Mature → Core

Tenure Days/Months/Years

0 ? ? ? ?
Customer Tenure Segments

*If subscription, look at hazard ratio curve.*

\[
HR_i = \frac{(# \text{ terminated during day } i)}{(# \text{ active at beginning of day } i)}
\]

Hazard Ratio

6/29/2015
Customer Tenure Segments

How to Determine Break Points?

KPI’s for the first three segments is the % of incoming customers passed on to next segment.

Newbie → On-board
Evolving → Nurture
Mature → Retain
Core → Reward

Tenure Days:
0
105
380
745
RFM Based Segments
What’s RFM?

It's all about your customer!

- **Recency** – How long since they did it?
- **Frequency** – How often have they done it?
- **Monetary** – What have they paid you to do it?

“it” depends on what you do & are measuring…

- Selling things => event is purchasing (classical RFM)
- Communicating => event is reading, clicking
- Delivering => event is using
Spoiler Alert: Final Segments

- **High Value**
- **Lost**
- **Lapsed**
- **Repeat**
- **New**
- **One-Time**

**Axes:**
- **Frequency (Highest to Lowest)**
- **Recency (Most Recent to Most Distant)**

---

6/29/2015

DS4CI.org
Recency

Breaks (weeks <=): 25, 51, 77, 103, <else>
levels = c("0-5", "6-11", "12-17", "18-23", "24-29")

Note levels labeled in months, not weeks
Frequency

Breaks (count <=): 1, 3, 7, <else> levels = c("8+", "7-4", "3-2", "1")

Note ordering for best is left.
Monetary

Breaks (value <=): 50, 100, 200, 400, <else>
levels = c("401+", "400-201", "200-101", "100-51", "50-0")

Again ordering is best is left.
Plot # Customers by R x F

Balloon Plot for Recency by Frequency.
Area is proportional to # Customers.

<table>
<thead>
<tr>
<th>Recency</th>
<th>0-5</th>
<th>6-11</th>
<th>12-17</th>
<th>18-23</th>
<th>24-29</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8+</td>
<td>1632</td>
<td>478</td>
<td>273</td>
<td>101</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>2487</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7-4</td>
<td>4478</td>
<td>2646</td>
<td>2040</td>
<td>1152</td>
<td>223</td>
</tr>
<tr>
<td></td>
<td>10539</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-2</td>
<td>959</td>
<td>7175</td>
<td>6337</td>
<td>6193</td>
<td>2916</td>
</tr>
<tr>
<td></td>
<td>39460</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>23195</td>
<td>21002</td>
<td>12212</td>
<td>21571</td>
<td>11463</td>
</tr>
<tr>
<td></td>
<td>86443</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

6/29/2015
Now Bake in Monetary

Balloon Plot for Recency by Frequency. Area is proportional to Annual Sales (000).

<table>
<thead>
<tr>
<th>Recency</th>
<th>0-5</th>
<th>6-11</th>
<th>12-17</th>
<th>18-23</th>
<th>24-29</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>8+</td>
<td>7-4</td>
<td>3-2</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

- Recency 0-5: 617.64
- Recency 6-11: 160.17
- Recency 12-17: 91.67
- Recency 18-23: 28.43
- Recency 24-29: 0.87

- Frequency 8+: 689.19
- Frequency 7-4: 302.45
- Frequency 3-2: 290.61
- Frequency 1: 153.71

- Area (000):
  - 800
  - 1656
  - 2056
  - 2256

6/29/2015

DS4CI.org
Too Complicated. Simplify!

- What cells can be combined?
  - Must make sense for marketing actions.
  - Way to many cells!
  - What is the story?

- How about two basic concepts?
  - Lifestage: New, Active, Lapsed, Lost
  - Value: Gold, Silver, Bronze

- Combine as:
  - New, Lapsed, & Lost
  - High Value, Repeat, & One-time
These Segments on Balloon Plot

Balloon Plot for Recency by Frequency. Area is proportional to Annual Sales (000).

<table>
<thead>
<tr>
<th>Recency</th>
<th>0-5</th>
<th>6-11</th>
<th>12-17</th>
<th>18-23</th>
<th>24-29</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td></td>
<td>8+</td>
<td>7-4</td>
<td>3-2</td>
<td>1</td>
</tr>
</tbody>
</table>

- **H**: High Recency, High Frequency
- **L**: Low Recency, Low Frequency
- **X**: Customer has made a purchase

Annual Sales (000):
- 2082
- 1591
- 1056
- 1541
- &gt;203.15097x280

DS4CI.org
Smoothing Out Adjacent Cells

- High Value
- Lost
- Lapsed
- Repeat
- New
- One-Time

Frequency vs. Recency

DS4CI.org
Clear Marketing Actions per Segment

- **High Value**
- **Lost**
- **Lapsed**
- **Repeat**
- **New**
- **One-Time**

Frequency vs. Recency grid:
- Highest Frequency: High Value
- Lowest Frequency: New
- Most Recent: Repeat
- Most Distant: One-Time
- Lost
- Lapsed
Cluster Based Segments
Cluster Based Segments

- Given a data set of non-required preferences and/or properties individual entities, find clusters to segment the entities.
- Totally unsupervised. We only specify the number of clusters we want.

- Two criteria for a “good” solution:
  - The cluster solution is stable
    - Repeatable with different random starts
  - The segments make sense
    - Our business partners can tell a story about each segment
Tool: flexclust by Fritz Leish

- Allows different distance measures
  - In particular, the Jaccard distance which is suited for optional response type survey or optional properties lists
    - 1 is a “yes” to the question - it is significant.
    - 0 is a didn't answer or does not apply. 0 is not the opposite of “yes.”
- Additionally flexclust had very good diagnostic and visualization tools. As an R package, it leverages the rest of the R ecosystem.
Using flexclust

- Generate cluster solutions for
  - A range of desired # of clusters (k), say 3 to 8
    - Over a set of random seeds (~ 6-10)
- Discard cluster solutions for k's which do not have stable solutions
- Examine variable plots by segment for remaining k's
  - Pick the one which makes most sense to your business partners.
Example 1 – Survey Responses

- 20k respondents to technical product use survey
- 35 check boxes or radio buttons
  - None are required
  - Coded as binary responses
- Goal: come up with “a few” segments which can be used to segment new respondents for follow up sales actions.
Example 1 - The 5-cluster solution

The 20k subjects plotted over the first two principal components:

The 5 clusters showing distribution of responses to each question:

Red dots are for the overall population
Example 2 – Business Attributes

- ~1k respondents to “nature of your business” survey
- 62 check boxes or radio buttons
  - In six topics
  - Some are required
  - Coded as binary responses
- Goal: come up with “a few” segments to characterize the fundamental nature of the on-line business
Example 2 – the 6-cluster solution
Customer Lifestage Segments

Preview
Sequence Analysis & Clustering

Example from TraMineR site – education / employment life stage clusters:
Can we find similar paths by our customers?

Stay tuned – we’re working on that now!
What We Covered

- Why & how to segment your customers
- Three examples & one preview:
  - Customer Tenure Segments
  - RFM Based Segments
  - Cluster Based Segments
  - Life Stage Cluster Segments

Questions? Comments? Now is the time!

DS4CI.org
Learning More

- Tenure based segmentation & subscription survival
  - Subscription Survival for Fun & Profit: [https://ds4ci.files.wordpress.com/2013/05/paw_sf2012_subscription_survival_for_fun_and_profit.pdf](https://ds4ci.files.wordpress.com/2013/05/paw_sf2012_subscription_survival_for_fun_and_profit.pdf)

- RFM based segmentation
  - Workshop at N Cal DMA lunch group [https://ds4ci.files.wordpress.com/2015/03/rfmb_dmanc_200905201.pdf](https://ds4ci.files.wordpress.com/2015/03/rfmb_dmanc_200905201.pdf)
    - Also has sample data set & flexclust example

- Customer Classification
  - See above useR! 2008 workshop for details on flexclust

- Lifestage with TraMinerR [http://traminer.unige.ch/](http://traminer.unige.ch/)


- Jim's Archives [www.ds4ci.org/archives](http://www.ds4ci.org/archives)

- Contact: Jim@DS4CI.org

6/29/2015