Sushant Kumar* et al.
(IJITR) INTERNATIONAL JOURNAL OF INNOVATIVE TECHNOLOGY AND RESEARCH
Volume No.4, Issue No.6, October – November 2016, 4965-4970.

A Proposal For Customer Scoring Based On Collective Individual And Segment Profile

SUSHANT KUMAR
Associate Data Scientist, Blueocean Market Intelligence
Building 2A, East Tower, Lower Ground Floor, Embassy Tech Village,
Devarabeesana Halli, Bellandur, Bengaluru, Karnataka 560103

Abstract: **The challenge to target customers in a retail domain has been tremendous, given the uncertainty in the behaviour a customer possess. Data driven decisions from actionable insights have however reduced, if not vanished the underlying challenge. However, not to ignore, the arbitrary "buying pattern", constantly linked to customer behaviour suggests some amount of subjectivity that lies within the said "insights" the data derive. Having said that data driven decisions can reduce the uncertainty associated with consumer behaviour, below we present a methodology with an objective for the same. The proposed "customer score" in the suggested methodology ranks the customers in order of likelihood of a positive response in a campaign relative to other customers.**

Key words: **Customer Scoring; Data Driven Decision; Buying Pattern; Positive Response;**

I. INTRODUCTION

Though it is common to profile customer based on individual historical data, it is noteworthy that supporting it with "segment profiling" can lead to better selection of customers. The underlying behaviour of segment a customer fall into cannot and should not be ignored while making informed decisions. While quantifying the behaviour for customer scoring, instead of arbitrary selection of variables, itis suggested that metricsbe selected based on its significance, tested through supervised learning methods. For example, while scoring customers for an upcoming campaign, a supervised learning model based on a reference campaign, preferably launched in the near past, can be built in order to discover the variables which actually distinguished customers with distinct response to a campaign.Though not to forget, assigning equal weights to all the significant variables can lead to arbitrary scoring, since it is unlikely that they will have a proportional effect on the outcome. Supervised methods like logistic regression calculate the odds for a given significant variable, which can later be used to allocate weights to the said variables. Intrinsic allocation of weights and creation of latent/artificial variables can also be carried out using unsupervised methods like Principal Component Analysis. In order to brief the methodology, individual profiling has been done by allocating weights to variables through supervised method. On the other hand, segment profiling has been done by creating latent variables which intrinsically take cares of weight allocation. Though the experimentation on weights allocation methodology for individual and segment profiling could have been carried out, it has been left as future scope of the work.

II. PROBLEM DEFINITION

The marketing team of a retail giant plan to carry out a sale campaign for its existing customers. Due to budgetary constraints, only a given proportion of customers have to be targeted, who have high likelihood of positive response in the campaign.

Objective

Score the customers which would rank customers in order of likelihood of positive response in the upcoming campaign.

Methodology

Take a reference campaign in sale period.

The customers can be selected from all the segments, or from a subset of segments.

Using "Point-of-Sales" data, the targeted customers are classified into "respondents" and "non-respondents" customers, based on their positive/negative response in the reference campaign.

Create a dependent variable y and define as given below:

Y = 1, in case of positive response.
= 0, otherwise.

Calculate the following metrics for all the respondent customers (preferably for at least 6 months).

- Total Revenue

- Total Trips
- Average Transactional Value (ATV)
- Total Revenue in Sale Period
- Total Trips in Sale Period
- Sale Ratio
- Trips Ratio

Where,
ATV = Total Revenue / Total Trips
Sale Ratio = Total Revenue in Sale Period / Total Revenue
Trips Ratio = Total Trips in Sale Period / Total Trips

Test for multicollinearity, as the variables might be related. Exclude one at a time, variable with high Variation Inflation Factor (VIF). Here, VIF > 6 is considered to be high.

Variable	DF	Parameter estimate	Standard error	t Value	Pr > |t|	Variance inflation
Intercept	**1**	-0.02754	0.00065604	-41.98	<.0001	0
Total trips	**1**	0.01346	0.0000831	161.96	<.0001	1.32217
ATV	**1**	4.10E-07	2.51E-07	1.63	0.1022	1.85499
Total revenue in sale period	**1**	-0.00000194	7.43E-08	-26.11	<.0001	2.26327
Sale Ratio	**1**	0.01453	0.00176	8.25	<.0001	5.92295
Trips Ratio	**1**	0.0183	0.00182	10.06	<.0001	5.76711

Following variables are considered after removing multicollinearity.

- Total Trips
- Average Transactional Value (ATV)
- Total Revenue in Sale Period
- Sale Ratio
- Trips Ratio

Run a logistic regression to identify the significant variables. Record the odds ratio for the significant variables.

Analysis of Maximum Likelihood Estimates					
Parameter	**DF**	**Estimate**	**Standard Error**	**Wald Chi Square**	**Pr > ChiSq**
Intercept	1	-5.0578	0.0289	30540.5958	<.0001
Total trips	1	0.1692	0.00152	12336.6381	<.0001
ATV	1	-0.00014	0.000013	116.5122	<.0001
Total revenue in sale period	1	-0.00002	1.91E-06	140.6887	<.0001
Sale Ratio	1	0.8495	0.0535	252.0888	<.0001
Trips Ratio	1	0.6996	0.0561	155.7592	<.0001

Odds Ratio Estimates			
Effect	**Point Estimate**	**95% Wald**	
Total trips	1.184	1.181	1.188
ATV	1	1	1
Total revenue in sale period	1	1	1
Sale Ratio	2.338	2.106	2.597
Trips Ratio	2.013	1.803	2.247

Significant Variables with odds ratio not equal to 1:

- Total Trips
- Sale Ratio
- Trips Ratio

Alternatively, Variable Clustering can also be used to identify the variables which would explain a major proportion of variance in the data. It would however be an unsupervised way of determining the variables as opposed to supervised method of logistic regression.

Cluster Summary for 3 Clusters				
Cluster	**Members**	**Cluster Variation**	**Variation Explained**	**Proportion Explained**
1	3	3	2.497641	0.8325
2	2	2	1.907219	0.9536
3	2	2	1.91554	0.9578
Total variation explained = 6.320401 Proportion = 0.9029				

3 Clusters		**R-squared with**		
Cluster	**Variable**	**Own Cluster**	**Next Closest**	**1-R**2 Ratio**
Cluster 1	**Total Revenue**	0.921	0.1816	**0.0965**
	ATV	0.658	0.007	0.3444
	Total revenue in sale period	0.9187	0.1691	0.0979
Cluster 2	**Sale Ratio**	0.9536	0.0618	0.0494
	Trips Ratio	0.9536	0.0498	**0.0488**
Cluster 3	**Total Trips**	0.9578	0.1067	**0.0473**
	Total Trips in Sale Period	0.9578	0.1316	0.0486

Variable Clustering inferred that Total Revenue, Trips Ratio and Total Trips explained major proportion of variance in the data.

We proceed with results of logistic regression.
Calculate for different segments, the following metrics.

- Average Transactional Value (ATV)
- Sale Ratio
- Trips Ratio
- Revenue per customer
- Trips per customer
- Sale revenue per customer
- Sale trips per customer

Here a segment is defined as any combination of Status, Group, Region and Ethnicity of a Customer.

- Status: Active / Inactive / Potential as defined by the business.
- Group: Target / Reference as defined by the business.
- Region: The Region in which the customer shops the most.
- Ethnicity: Self defined.

Create bins of all the above variables (here 4 bins are created)
Rank the bins from 1 – 4, 1 being the lowest.

For example:

ATV	Rank
92 – 250	1
250 – 600	2
600 – 1000	3
> 1000	4

Similarly for others.
For a given customer, allocate the rank based on the segment for all the variables.
Create optimal number of Factors of the ranked variables and record the factor scores.

Volume No.4, Issue No.6, October – November 2016, 4965-4970.

Scree Plot of Eigenvalues

```
  3.5 ˆ
      ,
      ,
      ,
      ,
      ,          1
  3.0 ˆ
      ,
      ,
      ,
      ,
      ,
  2.5 ˆ
      ,
      ,
      ,
E     ,                2
i     ,
g 2.0 ˆ
e     ,
n     ,
v     ,
a     ,
l     ,
u 1.5 ˆ
e     ,
s     ,
      ,
      ,
      ,
  1.0 ˆ
      ,
      ,
      ,
      ,
      ,                    3
  0.5 ˆ
      ,
      ,
      ,
      ,                         4
      ,                              5
  0.0 ˆ

        0     1     2     3     4     5
```

Number

Rotated Factor Pattern				
	Factor1		**Factor2**	
ATV Rank	92	*	-25	
Sale Ratio Rank	-9		-74	*
Revenue per customer rank	94	*	30	
Trips per customer rank	11		95	*
Sale Revenue per customer rank	96	*	22	
Sale Trips per customer rank	1		95	*
Printed values are multiplied by 100 and rounded to the nearest integer. Values greater than 0.4 are flagged by an '*'.				

Variance Explained by Each Factor	
Factor1	**Factor2**
2.6631964	2.5604535

Standardized Scoring Coefficients		
	Factor1	**Factor2**
ATV Rank	0.37312	-0.16386
Sale Ratio Rank	0.01375	-0.29243
Revenue per customer rank	0.34189	0.05847
Trips per customer rank	-0.01979	0.37564
Sale Revenue per customer rank	0.35543	0.02396
Sale Trips per customer rank	-0.0598	0.38032

Create bins of Factor scores and assign rank to bins from 1 to 5, 1 being the lowest.

Factor 1	Rank
< 0.95	1
0.96 – 0.10	2
0.11 – 0.57	3
0.58 – 1.22	4
>1.22	5

Factor 2	Rank
< -1.81	1
-1.82 – 0.16	2
0.17 – 0.37	3
0.38 – 0.97	4
>0.97	5

Calculate at Customer level, the value of all the significant variables.

Create bins and rank them from 1 – 5, 1 being the lowest.

Trips	Rank
1	1
2 – 3	2
4 – 5	3
6 – 8	4
>8	5

Sale Ratio	Rank
< = 0.34	1
0.35 – 0.91	2
0.92 – 0.94	3
0.95 – 0.97	4
> 0.97	5

Trips Ratio	Rank
< = 0.50	1
0.51 – 0.86	2
0.87 – 0.90	3
0.90 – 0.95	4
> 0.95	5

Assign weights to significant variables as follow:

- **$W_i = (\text{odds ratio})_i / \sum (\text{odds ratio})$**

Total Score = $\sum$ Factor Rank + $\sum$ (W_i* Variable Rank)

Total Score = factor1 rank + factor2 rank + (0.21 * trips rank) + (0.42 * sale ratio rank) + (0.36 * trips ratio rank)

III. CONCLUSION

Thus we find the following as conclusions based on our proposal :-1.The response rate of target group was twice as compared to reference group. 2.Average member spend of target group was 2.5 times more than reference group.

IV. ACKNOWLEDGEMENTS

The author thanks Dr.S.Sridhar, Professor and Director, RV Cognitive Technologies, R.V.College of Engineering, Bangalore, India for guiding to present my ideas in the form of a note by giving a good shape and communicating to this journal for publication.

V. REFERENCES

[1]. Thompson, Ed and Kolsky, Estaban (2004-12-27). "How to Approach Customer Experience Management". Gartner.com. Retrieved 2016-07-15. Summary[:] Managing the detail of customer experiences is an important part of any customer relationship management strategy.

[2]. Verhoef, Peter C.; Lemon, Katherine N.; Parasuraman, A.; Roggeveen, Anne; Tsiros, Michael; Schlesinger, Leonard A. (2009-03-01). "Customer Experience Creation: Determinants, Dynamics and Management Strategies". Journal of Retailing. Enhancing the Retail Customer Experience. **85** (1): 31–41. doi:10.1016/j.jretai.2008.11.001.

[3]. Janakiraman, Narayan; Meyer, Robert J.; Morales, Andrea C. (2006-12-01). "Spillover Effects: How Consumers Respond to Unexpected Changes in Price and Quality". Journal of Consumer Research. **33** (3): 361–369. doi:10.1086/508440. ISSN 0093-5301.

[4]. C, Meyer; A, Schwager (2007). Understanding Customer Experience. Harvard Business Review. pp. 1–12.

[5]. Andajani, Erna (2015-11-25). "Understanding Customer Experience Management in Retailing". Procedia - Social and Behavioral Sciences. 2nd Global Conference on Business and Social Sciences (GCBSS-2015) on "Multidisciplinary Perspectives on Management and Society", 17–18 September 2015, Bali, Indonesia. **211**: 629–633. doi:10.1016/j.sbspro.2015.11.082.

[6]. Gentile, Chiara; Spiller, Nicola; Noci, Giuliano (2007-10-01). "How to Sustain the Customer Experience:: An Overview of Experience Components that Co-create Value With the Customer". European Management Journal. **25** (5): 395–410. doi:10.1016/j.emj.2007.08.005.

[7]. Zwilling, Mark (10 March 2014). "'Customer Experience' Is Today's Business Benchmark". Forbes. Retrieved 7 May 2015.

[8]. Manning, Harley; Bodine, Kerry (28 August 2012). "The 6 Disciplines Behind Consistently Great Customer Experiences". Fast Company. Retrieved 7 May 2015.

