

Enginius

Segmentation

Puja Gundu, California state university- East bay

Copyright (c) 2024, DecisionPro Inc.

Warnings

The following warnings were triggered during execution. Although they did not interrupt the analyses, they might indicate that there is an issue with the data or with the options chosen. Please review them carefully before going any further.

These variables from discriminant data are removed because of high collinearity with variables present in the report - 7. What type of promotions do you prefer? = Loyalty rewards, 7. What type of promotions do you prefer? = Free shipping, 8. Where do you usually find out about Black Friday deals? = Store Flyers, 10. Do you typically prepare for Black Friday shopping? = Create a budget, 12. What influences your Black Friday purchase decisions the most? = Advertising and promotions, 13. How likely are you to buy items as gifts during Black Friday? = Somewhat unlikely, 5. What motivates you to shop during Black Friday? [Discounts] = 2-Slightly Agree, 5. What motivates you to shop during Black Friday? [Discounts] = 1 - Not Agree, 5. What motivates you to shop during Black Friday? [Convenience] = 5 - Highly Agree, 5. What motivates you to shop during Black Friday? [Gift Shopping] = 5 - Highly Agree, 5. What motivates you to shop during Black Friday? [New Product Releases] = 5 - Highly Agree

Segmentation options

Options selected

	Option	Selection
	Clustering method	Hierarchical
	Standardization method	standard
	Segments forced	2
	Run discriminant analysis	Yes
	Run classification analysis	Yes
	Date and time	2024-12-07 00:16:59 UTC

Options selected.

Data description

	Data	Number of Rows	Number of columns	Column names
1	Segmentation data	60	7	1. What is your age group?, 2. How do you plan your Black Friday shopping?, 3. How often do you shop during Black Friday?, 4. How do you prefer to shop on Black Friday?, 9. What is your preferred device for online Black Friday shopping?, ...
2	Discriminant data	60	10	6. How important are discounts when deciding what to buy?, 7. What type of promotions do you prefer?, 8. Where do you usually find out about Black Friday deals?, 10. Do you typically prepare for Black Friday shopping?, 12. What influences your Black Friday purchase decisions the most?, ...

Data description.

Data transformation

The segmentation data has been scaled column wise

	Mean	Standard deviation
2. How do you plan your Black Friday shopping? = Impulse buying without planning	0.3333	0.4754
2. How do you plan your Black Friday shopping? = Decide on the day	0.3833	0.4903
3. How often do you shop during Black Friday? = Every year	0.5333	0.5031
3. How often do you shop during Black Friday? = Occasionally	0.3167	0.4691
4. How do you prefer to shop on Black Friday? = In-store	0.2000	0.4034
4. How do you prefer to shop on Black Friday? = Online	0.1833	0.3902
9. What is your preferred device for online Black Friday shopping? = Smartphone	0.5500	0.5017
9. What is your preferred device for online Black Friday shopping? = Laptop/Desktop computer	0.4000	0.4940
14. Which categories do you shop for most during Black Friday? = Electronics	0.3500	0.4810
14. Which categories do you shop for most during Black Friday? = Other	0.1500	0.3601
14. Which categories do you shop for most during Black Friday? = Apparel	0.4167	0.4972
15. How much do you typically spend during Black Friday? = More than 500	0.3833	0.4903
15. How much do you typically spend during Black Friday? = Less than 100	0.1333	0.3428

Mean and standard deviation column wise.

Segment solution

2-segment solution

The ideal number of segments is a function of statistical fit (what the data say), managerial relevance (what makes the most sense from a managerial point of view), and targetability (can the segments be easily targeted).

When the three criteria do not perfectly converge, selecting the right number of segments becomes a judgment call.

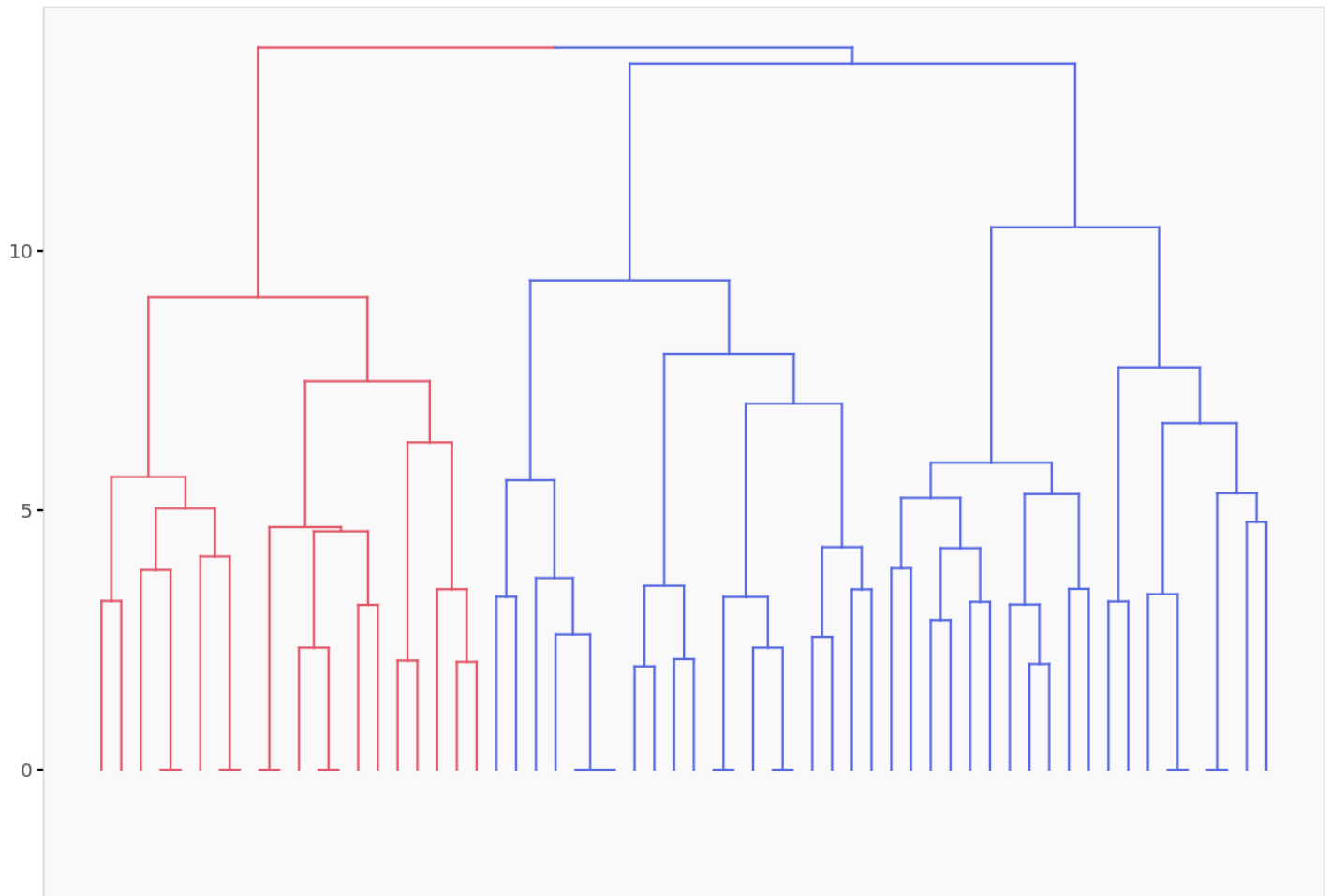
You have decided to perform the analysis with 2 segments.

The segmentation method relies on the hierarchical clustering approach. This approach generates a dendrogram that we display next.

Dendrogram

The dendrogram represents the grouping process of observations into clusters. The chart reads from bottom (all initial observations are separated) to top (all observations are clustered into one unique segment).

The height represents the distance between the two groups of observations being merged at each step. If two very distant groups are being merged, this will create a 'jump' in the dendrogram, indicating that it might be wise to stop the clustering process before.



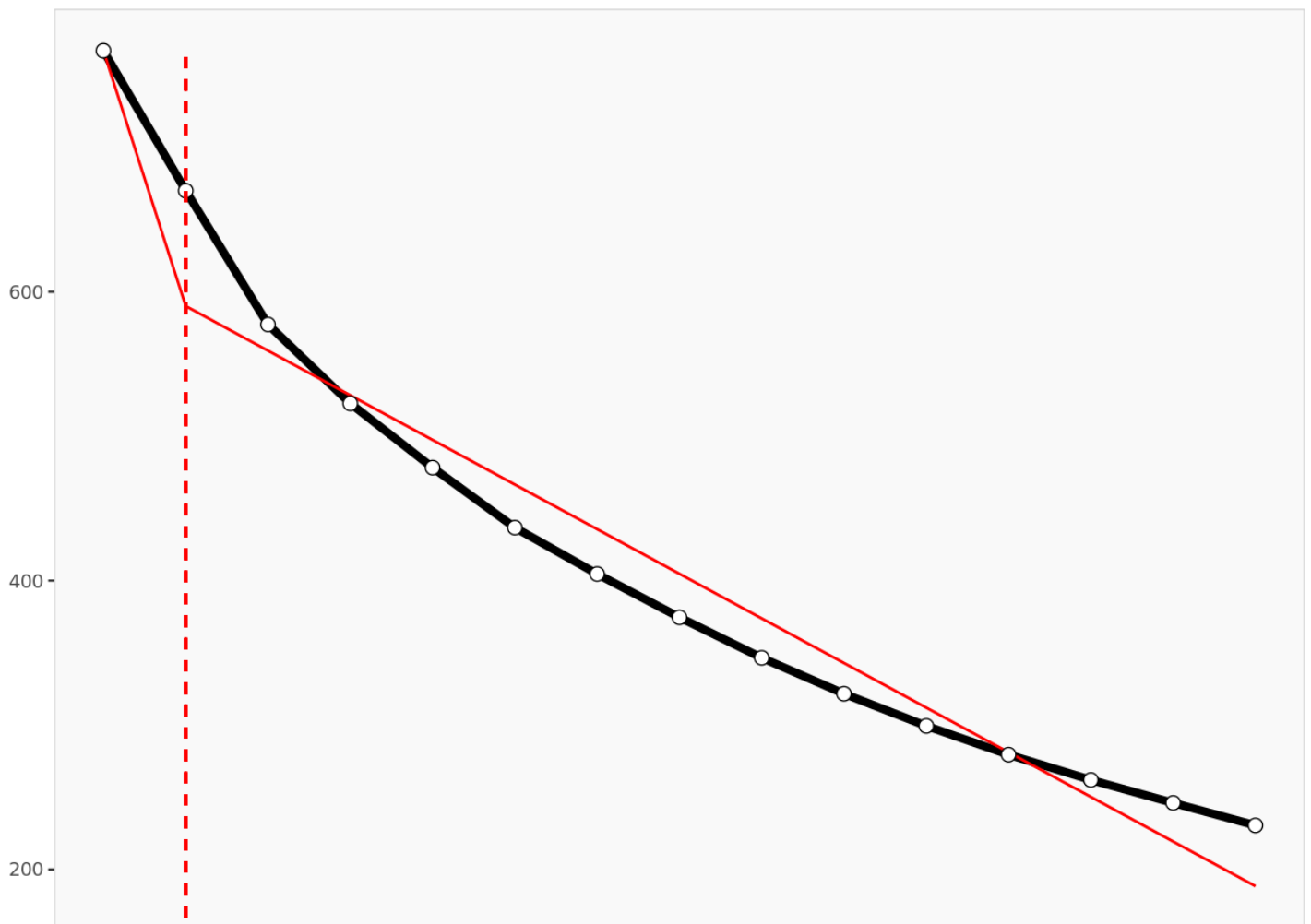
Dendrogram. The dendrogram is a tree diagram to illustrate the arrangement of clusters produced by hierarchical clustering, and how the observations are incrementally clustered together.

Scree plot

The screeplot displays, for each cluster solution, a measure of within-cluster heterogeneity. If clusters group observations that are widely different (which will happen if the number of clusters is too small to capture the variability in the data), the value will be high.

A good cluster solution might be where the screeplot displays an 'elbow', that is, where increasing the number of clusters beyond a certain point does not dramatically decreases within-cluster heterogeneity.

The measure displayed in the screeplot is related, but not equivalent, to the distance reported in the dendrogram.



Scree plot. The scree plot compares the sum of squared error (SSE) for each cluster solution. A good cluster solution might be when the SSE slows dramatically, creating an 'elbow'. Such elbow does not always exist. If number of segments is equal to maximum possible segments elbow cannot be created.

From a statistical point of view, the SSE reported in the screeplot is computed as the sum of squared error between each observation and its cluster centroid (or center), summed over all the observations.

Segment description

Segment size

	Population	Segment 1	Segment 2
Size	60	20	40
Relative size	100%	33%	67%

Segment size.

Segment description

	Population	Segment 1	Segment 2
2. How do you plan your Black Friday shopping? = Impulse buying without planning	0.333	0.400	0.300
2. How do you plan your Black Friday shopping? = Decide on the day	0.383	0.200	0.475
3. How often do you shop during Black Friday? = Every year	0.533	0.500	0.550
3. How often do you shop during Black Friday? = Occasionally	0.317	0.350	0.300
4. How do you prefer to shop on Black Friday? = In-store	0.200	0.100	0.250
4. How do you prefer to shop on Black Friday? = Online	0.183	0.300	0.125
9. What is your preferred device for online Black Friday shopping? = Smartphone	0.550	0.900	0.375
9. What is your preferred device for online Black Friday shopping? = Laptop/Desktop computer	0.400	0.100	0.550
14. Which categories do you shop for most during Black Friday? = Electronics	0.350	0.900	0.075
14. Which categories do you shop for most during Black Friday? = Other	0.150	0.000	0.225
14. Which categories do you shop for most during Black Friday? = Apparel	0.417	0.050	0.600
15. How much do you typically spend during Black Friday? = More than 500	0.383	0.300	0.425
15. How much do you typically spend during Black Friday? = Less than 100	0.133	0.150	0.125

Segment description. Average value of each segmentation variable, overall for each segment (centroid). Segmentation variables that

are statistically different from the rest of the population are highlighted in red (lower) or green (higher).



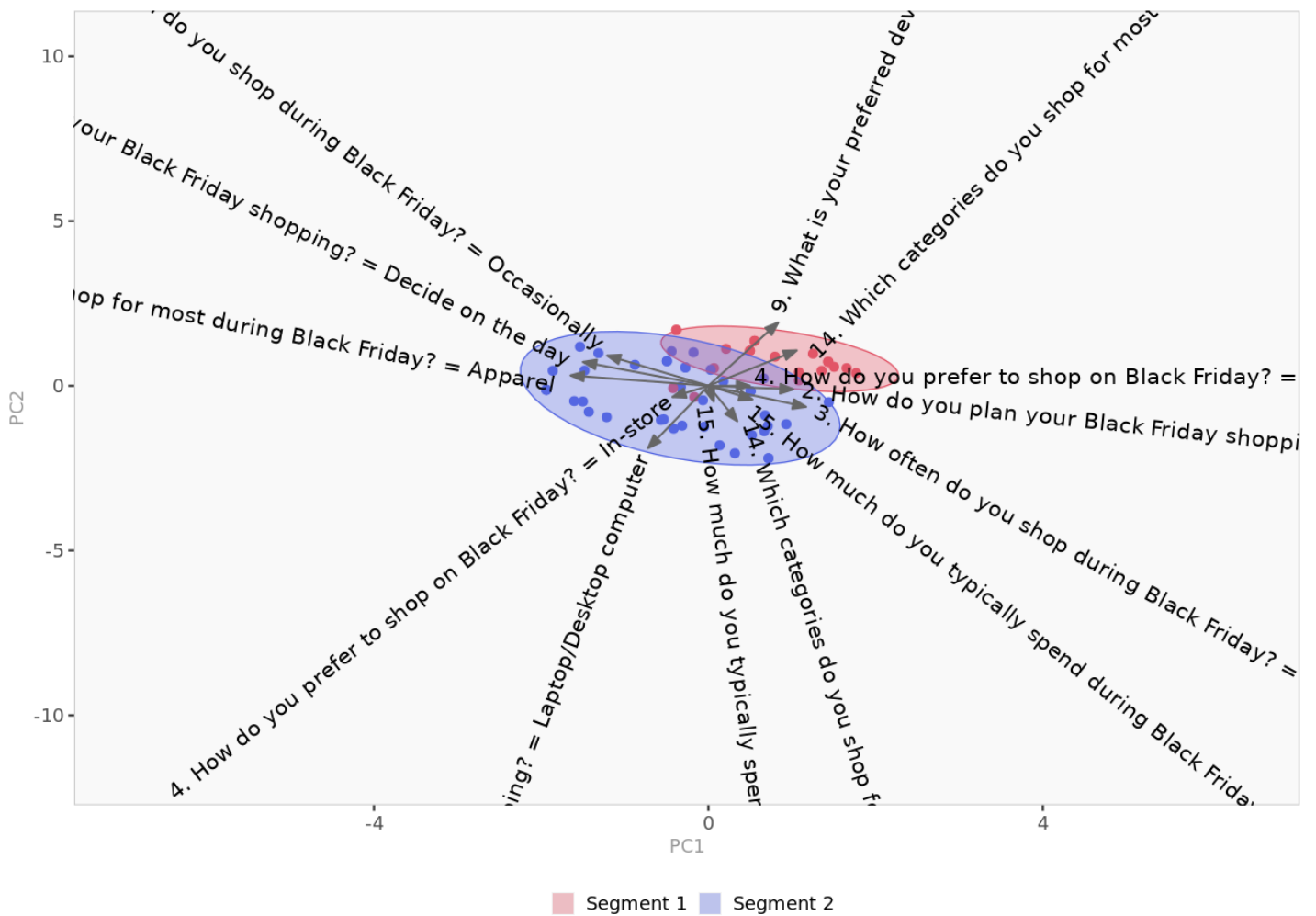
Segment differences per segment. Cell colors indicate to what extent a segment is statistically different from the rest of the population on each segmentation variable.

Segmentation space

The chart below is a graphical representation of the various segments, segment members, and segmentation variables. It is obtained by plotting the first two dimensions of a principal component analysis performed on the (standardized) segmentation data, on top of which segment information has been overlaid.

Because only the first two dimensions of the PCA are displayed, and these two dimensions capture only 36.3% of the variance in the data, some differences between segments might not appear here. Note that segmentation variables with no variance, if any, have been excluded.

Two clusters that appear to overlap on the first two dimensions might be distinct on other dimensions. Consequently, this chart is a useful guide, for checking which variables are correlated, but may be misleading if used to select the optimal number of segments.



Segment space. Spatial representation of segments and segmentation variables, using principal component analysis.

Segment membership

	Segment
1	1
2	2
3	2
4	2
5	2
6	2
7	2
8	2
9	2
10	2

Segment membership (excerpt). Segment to which each member of the population belongs to. The complete membership list is only available in the Excel formatted output.

Descriptor analysis

Descriptors

This table reports the descriptor averages of each segment. The more differences can be found, the easier it will be to predict segment membership based on descriptors alone.

	Population	Segment 1	Segment 2
`7. What type of promotions do you prefer?` Buy-one-get-one deals	0.233	0.200	0.250
`7. What type of promotions do you prefer?` Flash sales	0.700	0.750	0.675
`7. What type of promotions do you prefer?` Free shipping	0.017	0.000	0.025
`7. What type of promotions do you prefer?` Loyalty rewards	0.050	0.050	0.050
`8. Where do you usually find out about Black Friday deals?` Emails	0.083	0.100	0.075
`8. Where do you usually find out about Black Friday deals?` Social Media	0.417	0.400	0.425
`8. Where do you usually find out about Black Friday deals?` Store Flyers	0.050	0.000	0.075
`8. Where do you usually find out about Black Friday deals?` Websites	0.333	0.350	0.325
`8. Where do you usually find out about Black Friday deals?` Word of Mouth	0.117	0.150	0.100
`10. Do you typically prepare for Black Friday shopping?` Create a budget	0.117	0.150	0.100
`10. Do you typically prepare for Black Friday shopping?` don t prepare in advance	0.467	0.450	0.475
`10. Do you typically prepare for Black Friday shopping?` Follow brands on social media	0.133	0.200	0.100
`10. Do you typically prepare for Black Friday shopping?` Make a wishlist	0.283	0.200	0.325
`12. What influences your Black Friday purchase decisions the most?` Advertising and promotions	0.100	0.050	0.125
`12. What influences your Black Friday purchase decisions the	0.333	0.250	0.375

most?` Brand reputation			
`12. What influences your Black Friday purchase decisions the most?` Product reviews and ratings	0.333	0.300	0.350
`12. What influences your Black Friday purchase decisions the most?` Recommendations from friends/family	0.233	0.400	0.150
`13. How likely are you to buy items as gifts during Black Friday?` Neutral	0.367	0.400	0.350
`13. How likely are you to buy items as gifts during Black Friday?` Somewhat likely	0.267	0.200	0.300
`13. How likely are you to buy items as gifts during Black Friday?` Somewhat unlikely	0.033	0.000	0.050
`13. How likely are you to buy items as gifts during Black Friday?` Very likely	0.250	0.350	0.200
`13. How likely are you to buy items as gifts during Black Friday?` Very unlikely	0.083	0.050	0.100
`5. What motivates you to shop during Black Friday? [Discounts]` 1 - Not Agree	0.067	0.050	0.075
`5. What motivates you to shop during Black Friday? [Discounts]` 2-Slightly Agree	0.067	0.050	0.075
`5. What motivates you to shop during Black Friday? [Discounts]` 3 - Moderately agree	0.267	0.100	0.350
`5. What motivates you to shop during Black Friday? [Discounts]` 4-Agree	0.167	0.250	0.125
`5. What motivates you to shop during Black Friday? [Discounts]` 5 - Highly Agree	0.433	0.550	0.375
`5. What motivates you to shop during Black Friday? [Convenience]`	0.083	0.050	0.100
`5. What motivates you to shop during Black Friday? [Convenience]` 1 - Not Agree	0.200	0.200	0.200
`5. What motivates you to shop during Black Friday? [Convenience]` 2-Slightly Agree	0.183	0.200	0.175
`5. What motivates you to shop during Black Friday? [Convenience]` 3 - Moderately agree	0.250	0.250	0.250
`5. What motivates you to shop during Black Friday? [Convenience]` 4-Agree	0.217	0.250	0.200

`5. What motivates you to shop during Black Friday? [Convenience]` 5 - Highly Agree	0.067	0.050	0.075
`5. What motivates you to shop during Black Friday? [Gift Shopping]`	0.100	0.000	0.150
`5. What motivates you to shop during Black Friday? [Gift Shopping]` 1 - Not Agree	0.167	0.050	0.225
`5. What motivates you to shop during Black Friday? [Gift Shopping]` 2-Slightly Agree	0.150	0.200	0.125
`5. What motivates you to shop during Black Friday? [Gift Shopping]` 3 - Moderately agree	0.200	0.200	0.200
`5. What motivates you to shop during Black Friday? [Gift Shopping]` 4-Agree	0.300	0.550	0.175
`5. What motivates you to shop during Black Friday? [Gift Shopping]` 5 - Highly Agree	0.083	0.000	0.125
`5. What motivates you to shop during Black Friday? [New Product Releases]`	0.083	0.000	0.125
`5. What motivates you to shop during Black Friday? [New Product Releases]` 1 - Not Agree	0.283	0.250	0.300
`5. What motivates you to shop during Black Friday? [New Product Releases]` 2-Slightly Agree	0.200	0.150	0.225
`5. What motivates you to shop during Black Friday? [New Product Releases]` 3 - Moderately agree	0.200	0.250	0.175
`5. What motivates you to shop during Black Friday? [New Product Releases]` 4-Agree	0.150	0.300	0.075
`5. What motivates you to shop during Black Friday? [New Product Releases]` 5 - Highly Agree	0.083	0.050	0.100

Descriptor data per segment. Average value of each descriptor, overall and within each cluster. Descriptors that are statistically different from the rest of the population are highlighted in red (lower) or green (higher).



Descriptor differences per segment. Cell colors indicate to what extent the distribution of a descriptor in a segment is statistically different from the rest of the population.

Descriptor space

The chart below is a graphical representation of the various segments, segment members, and descriptors. It is obtained by outputting the first two dimensions of a principal component analysis performed on the (standardized) descriptors, on top of which segment information has been overlayed.

Because only the first two dimensions of the PCA are displayed, and these two dimensions capture only 20.4% of the variance in the data, some differences between segments might not appear here. Note that descriptors with no variance, if any, have been excluded.

If two or more segments fully overlap, it is unlikely that they could be clearly separated based on descriptors alone.

However, two segments that seem to overlap on two dimensions may be more clearly separated on other dimensions. Consequently, the confusion matrix is a better guide to assess the quality of segment classification based on descriptors.

Classification model

Introduction

Often, segmentation (needs) variables for each customer may not be available to managers, but descriptors variables for customers may be available.

In this section, we explore whether descriptors alone can predict segment membership with sufficient accuracy. The confusion matrix and hit rates (reported below) indicate whether the model is accurate enough.

For member classification based on descriptors, Enginius uses a multinomial logit model (similar to the one used to predict 'choices between multiple alternatives (A/B/C)' in the predictive modeling module.

The largest segment is selected as the default option (dummy), and the model identifies which descriptors are the most significant for predicting cluster memberships. If a descriptor is highly predictive, its p-values will be close to zero, and the cells will appear in green (or red).

Model coefficients

	Segment summary(model)\$coefficients
(Intercept)	-2 781
7. What type of promotions do you prefer? = Buy-one-get-one deals	198
7. What type of promotions do you prefer? = Flash sales	431
8. Where do you usually find out about Black Friday deals? = Social Media	1 037
8. Where do you usually find out about Black Friday deals? = Websites	1 314
8. Where do you usually find out about Black Friday deals? = Word of Mouth	2 012
8. Where do you usually find out about Black Friday deals? = Emails	1 191
10. Do you typically prepare for Black Friday shopping? = Follow brands on social media	1 754
10. Do you typically prepare for Black Friday shopping? = don t prepare in advance	718
10. Do you typically prepare for Black Friday shopping? = Make a wishlist	1 231
12. What influences your Black Friday purchase decisions the most? = Recommendations from friends/family	200
12. What influences your Black Friday purchase decisions the most? = Product reviews and ratings	24.9
12. What influences your Black Friday purchase decisions the most? = Brand reputation	-423
13. How likely are you to buy items as gifts during Black Friday? = Neutral	-335
13. How likely are you to buy items as gifts during Black Friday? = Very likely	-307

13. How likely are you to buy items as gifts during Black Friday? = Somewhat likely	-234
13. How likely are you to buy items as gifts during Black Friday? = Very unlikely	-517
5. What motivates you to shop during Black Friday? [Discounts] = 3 - Moderately agree	-882
5. What motivates you to shop during Black Friday? [Discounts] = 5 - Highly Agree	72.3
5. What motivates you to shop during Black Friday? [Discounts] = 4-Agree	600
5. What motivates you to shop during Black Friday? [Convenience] = 3 - Moderately agree	-382
5. What motivates you to shop during Black Friday? [Convenience] = 1 - Not Agree	331
5. What motivates you to shop during Black Friday? [Convenience] = 2-Slightly Agree	-165
5. What motivates you to shop during Black Friday? [Convenience] = 4-Agree	-133
5. What motivates you to shop during Black Friday? [Gift Shopping] = 2-Slightly Agree	897
5. What motivates you to shop during Black Friday? [Gift Shopping] = 3 - Moderately agree	511
5. What motivates you to shop during Black Friday? [Gift Shopping] = 4-Agree	1 043
5. What motivates you to shop during Black Friday? [Gift Shopping] = 1 - Not Agree	481
5. What motivates you to shop during Black Friday? [New Product Releases] = 3 - Moderately agree	325
5. What motivates you to shop during Black Friday? [New Product Releases] = 2-Slightly Agree	-195
5. What motivates you to shop during Black Friday? [New Product Releases] = 1 - Not Agree	-529
5. What motivates you to shop during Black Friday? [New Product Releases] = 4-Agree	-118

Model parameters. Segment 2 is the model baseline.

P-values

	Segment summary(model)\$coefficients
(Intercept)	0.000
7. What type of promotions do you prefer? = Buy-one-get-one deals	0.490
7. What type of promotions do you prefer? = Flash sales	0.036
8. Where do you usually find out about Black Friday deals? = Social Media	0.001
8. Where do you usually find out about Black Friday deals? = Websites	0.000
8. Where do you usually find out about Black Friday deals? = Word of Mouth	NaN
8. Where do you usually find out about Black Friday deals? = Emails	0.000
10. Do you typically prepare for Black Friday shopping? = Follow	0.000

brands on social media	
10. Do you typically prepare for Black Friday shopping? = don t prepare in advance	0.000
10. Do you typically prepare for Black Friday shopping? = Make a wishlist	0.000
12. What influences your Black Friday purchase decisions the most? = Recommendations from friends/family	0.196
12. What influences your Black Friday purchase decisions the most? = Product reviews and ratings	0.913
12. What influences your Black Friday purchase decisions the most? = Brand reputation	0.384
13. How likely are you to buy items as gifts during Black Friday? = Neutral	0.113
13. How likely are you to buy items as gifts during Black Friday? = Very likely	0.000
13. How likely are you to buy items as gifts during Black Friday? = Somewhat likely	0.291
13. How likely are you to buy items as gifts during Black Friday? = Very unlikely	0.000
5. What motivates you to shop during Black Friday? [Discounts] = 3 - Moderately agree	0.000
5. What motivates you to shop during Black Friday? [Discounts] = 5 - Highly Agree	0.950
5. What motivates you to shop during Black Friday? [Discounts] = 4-Agree	0.000
5. What motivates you to shop during Black Friday? [Convenience] = 3 - Moderately agree	0.232
5. What motivates you to shop during Black Friday? [Convenience] = 1 - Not Agree	0.000
5. What motivates you to shop during Black Friday? [Convenience] = 2-Slightly Agree	0.000
5. What motivates you to shop during Black Friday? [Convenience] = 4-Agree	0.382
5. What motivates you to shop during Black Friday? [Gift Shopping] = 2-Slightly Agree	0.139
5. What motivates you to shop during Black Friday? [Gift Shopping] = 3 - Moderately agree	0.366
5. What motivates you to shop during Black Friday? [Gift Shopping] = 4-Agree	0.001
5. What motivates you to shop during Black Friday? [Gift Shopping] = 1 - Not Agree	0.006
5. What motivates you to shop during Black Friday? [New Product Releases] = 3 - Moderately agree	0.609
5. What motivates you to shop during Black Friday? [New Product Releases] = 2-Slightly Agree	0.635
5. What motivates you to shop during Black Friday? [New Product Releases] = 1 - Not Agree	0.000
5. What motivates you to shop during Black Friday? [New Product Releases] = 4-Agree	0.512

p-values. Probabilities that parameter estimates are different from zero only by chance.

Confusion matrix

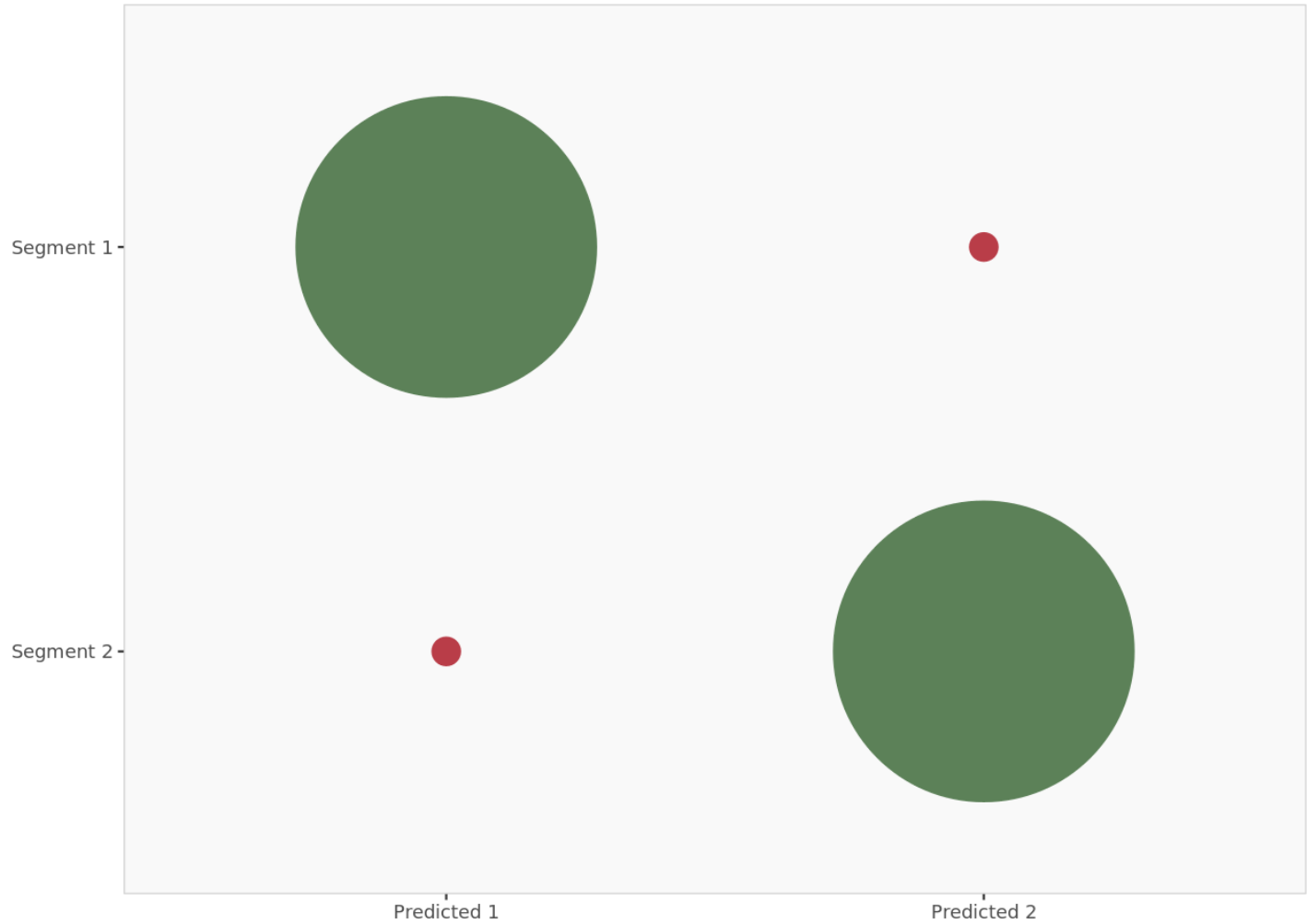
The confusion matrix compares actual segment membership (obtained from the segmentation analysis and the original segmentation variables) and predicted segment membership (obtained from the in-sample classification analysis and the descriptors alone). When actual and predicted segment memberships coincide, the diagonal elements will be comparatively large, indicating that the classification model based on available descriptors is accurate.

	Predicted 1	Predicted 2	Total
Segment 1	20	0	20
Segment 2	0	40	40
Total	20	40	60

Confusion matrix (count). The model has correctly classified 60 of the 60 observations. The off-diagonal elements are classification errors.

	Predicted 1	Predicted 2	Total
Segment 1	100%	0%	100%
Segment 2	0%	100%	100%

Confusion matrix (%). The global hit rate of the model is 100%. The diagonal elements represent segment-specific hit rates.



Confusion matrix (plot). Graphic representation of the confusion matrix: actual segment membership versus predicted segment membership. Bubbles in the diagonale represent correct classification.

Model predictions

	Prob(cluster 1)	Prob(cluster 2)	Predicted	Actual	Correct
1	100%	0%	1	1	1
2	0%	100%	2	2	1
3	0%	100%	2	2	1
4	0%	100%	2	2	1
5	0%	100%	2	2	1
6	0%	100%	2	2	1
7	0%	100%	2	2	1
8	0%	100%	2	2	1
9	0%	100%	2	2	1
10	0%	100%	2	2	1

Model predictions (in-sample) (excerpt). This table details the probabilities of each member of the segmentation dataset to belong to each cluster (as predicted by the in-sample classification model and the descriptors alone). The segment with the highest probability is retained, and is compared to the actual segment membership to measure model accuracy and classification errors.