



Sample Data Guide

Typing Tool (Segment Assignment)

File	Respondents	Columns
typing_tool_sample.sav	1,000	29

This file contains the same attitudinal data as the segmentation sample (1,000 U.S. online shoppers, 19 shopping-behavior statements on a 7-point scale), plus a pre-assigned segment membership column (Seg_kmeans_5) from a prior k-means segmentation run with 5 clusters. The typing tool uses these segmentation variables and the known segment assignments to train a classification model (e.g., discriminant analysis or random forest) that can type new respondents into the established segments based solely on their attitudinal responses.

Business Questions

The typing tool answers these operational questions:

1. Given a new respondent's attitudinal answers, which of the 5 established segments do they belong to?
2. Which survey questions are the strongest discriminators between segments (i.e., which minimal set of questions could a short screener use)?
3. How accurately can the model classify held-out respondents, and which segments are most often confused?

Segment Assignment Column (required for typing tool)

This column contains the segment each respondent was assigned to in the original segmentation.

Column	Description	Values
Seg_kmeans_5	Segment membership (k-means, 5 clusters)	1 = Segment 2, 2 = Segment 3, 3 = Segment 4, 4 = Segment 5, 5 = Segment 6

Segmentation Variables (19 columns)

These are the input variables used to build the typing model. All use a 7-point Likert scale: 1 = Completely disagree ... 7 = Completely agree

Column	Statement
Q5r2	I like to explore new products within brands or stores I already shop at
Q5r6	I consider my go-to online store to be part of my personal routine

Q5r9	I shop from brands I aspire to own
Q5r11	I prefer to discover products that are not yet mainstream
Q5r12	I am not loyal to any particular online retailer
Q5r71	I generally prefer to shop in physical stores rather than online
Q5r25	I check online shopping apps multiple times throughout the day
Q5r29	I keep shopping apps on my phone so I can browse whenever I have free time
Q5r34	I like trying the newest products as soon as they launch
Q5r35	I prefer to buy products that most people do not own yet
Q5r36	I generally stick to well-known, established brands that have been around for a while
Q5r37	I find all the product choices available online to be overwhelming
Q5r41	If I like a product when browsing online, I will buy it immediately
Q5r47	I sometimes purchase products immediately after seeing them for the first time
Q5r51	I tend to only buy products when they are on sale or discounted
Q5r53	When visiting a website for one item, I like to browse other categories the store carries
Q5r54	I like to browse the new arrivals section of online stores to see what is new
Q5r91	I only buy a new product after I have used up what I already have
Q5r92	I actively compare prices across multiple retailers before making a purchase

Profiling Variables (4 columns)

Demographic variables available for segment profiling.

Column	Description	Values
hAge	Age group	1 = 18-24, 2 = 25-34, 3 = 35-44, 4 = 45-50
hRegion	Geographic region	1 = North East, 2 = Midwest, 3 = South, 4 = West
DM1	Household income	1 = Under \$20K, 2 = \$20-35K, 3 = \$35-50K, 4 = \$50-75K, 5 = \$75-100K, 6 = \$100-150K, 7 = \$150-200K, 8 = \$200K+, 9 = Refused
DM4	Education level	1 = High school, 2 = Some college, 3 = Bachelor's, 4 = Post-graduate

Additional Variables for Reporting (3 columns)

Extra context for human-in-the-loop review and enriched reporting.

Column	Description	Values
Q1a	Online spending for self (past 12 months)	1 = Under \$50, 2 = \$50-149, 3 = \$150-249, 4 = \$250-499, 5 = \$500-999, 6 = \$1000+, 7 = Don't know
Q1b	Number of online retailers shopped regularly	1 = 1, 2 = 2, 3 = 3, 4 = 4, 5 = 5, 6 = 6, 7 = 7 or more

Q2b	Online gift spending (past 12 months)	1 = \$0, 2 = \$1-49, 3 = \$50-149, 4 = \$150-249, 5 = \$250-499, 6 = \$500-999, 7 = \$1000+, 8 = Don't know
-----	---------------------------------------	---

Weight Column

Apply this column as a survey weight to adjust for sampling design.

Column	Description	Details
FinalWeightV2	Survey weight	Continuous numeric, mean ~0.88, range 0.10 - 7.66

Note: This data file is in SPSS (.sav) format with full metadata including variable labels, value labels, and measure types. Column names and value labels are embedded in the file and will display automatically when opened in SPSS or uploaded to CrowdmindAI.