



Turas Segment

Transparent. Trustworthy. Tangible insights.

Discovers natural groups in your customer base—distinct segments that need different approaches. Uses statistical clustering algorithms to identify customers with similar needs, behaviors, or attitudes and creates actionable segments for targeting and strategy.

What This Module Does

Not all customers are the same. When you treat all customers identically, you waste resources on customers unlikely to respond, miss opportunities with high-value subgroups, and use generic messaging that resonates with nobody. Segmentation finds the natural groups.

Using K-means clustering and Latent Class Analysis, the module identifies how many distinct customer groups exist and what makes each unique—enabling precision targeting, tailored messaging, and strategic resource allocation.

What You Get

Exploration Mode (First Run)

- **K selection report:** Tests $k=2$ through $k=10$, recommends optimal number of segments
- **Quality metrics:** Silhouette scores, within-cluster SS, between-cluster SS
- **Elbow plot:** Visual guide to where adding segments stops helping

Final Mode (After Choosing K)

- **Segment assignments:** Each respondent assigned to their segment
- **Segment profiles:** Detailed characteristics of each segment
- **Discriminating variables:** Which variables best separate segments (ANOVA F-stats)
- **Targeting recommendations:** Business-ready segment cards with implications

How It Works

K-Means Clustering

Initialize k cluster centres randomly. Assign each customer to nearest centre. Recalculate centres as mean of assigned customers. Repeat until convergence. Result: groups where customers within each group are similar, groups are different from each other.

Choosing the Right K

Silhouette score measures how well each customer fits their segment vs. others (>0.5 is good). Elbow method shows where adding segments stops helping. Business actionability: 4-6 segments is usually optimal—manageable and distinct.

Latent Class Analysis (Alternative)

Model-based approach for categorical data. Estimates probability of segment membership. Provides BIC for model selection. Better for discrete, categorical clustering variables.

Packages Used

All packages are peer-reviewed, open-source R packages available on CRAN.

Package	Why We Use It
<code>stats::kmeans</code>	Core K-means (Hartigan-Wong algorithm)—fast, standard
<code>cluster</code>	Silhouette analysis and advanced clustering methods
<code>poLCA</code>	Latent Class Analysis for categorical data segmentation
<code>MASS</code>	Linear discriminant analysis for validation

Strengths

- ✓ **Exploration + Final workflow:** Guides you to optimal k systematically
- ✓ **Statistical validation:** Stability testing and discriminant analysis
- ✓ **Handles large datasets:** Mini-batch k-means for $n > 10,000$
- ✓ **Rich profiling:** Discriminating variables, indices, narratives
- ✓ **Business-ready outputs:** Segment cards, targeting recommendations
- ✓ **Reproducible:** Seed control ensures consistent results

Limitations

- ⚠ **Requires adequate sample:** Need 100+ per expected segment
- ⚠ **Variable selection is subjective:** Analyst judgment required
- ⚠ **K-means assumptions:** Assumes spherical clusters of similar size
- ⚠ **Not predictive:** Doesn't assign NEW customers to segments (need scoring model)
- ⚠ **Naming requires interpretation:** Statistical output needs business translation

Sample Size Requirements

Rule of thumb: $n \geq 100$ -150 per expected segment. For $k=5$ segments, need $n \geq 500$ -750. Small segments (< 100) are unstable and unreliable.

Best Use Cases

Ideal For

- ✓ Customer segmentation for targeting strategy
- ✓ Needs-based segmentation for product development
- ✓ Behavioural segmentation for personalization
- ✓ Attitudinal segmentation for messaging
- ✓ Market structure analysis

Not Ideal For

- ⚠ Homogeneous markets where no real groups exist
- ⚠ Very small samples (< 200 total)
- ⚠ When you just need demographic profiles (use cross-tabs)
- ⚠ Real-time segment assignment (this is batch-oriented)

Key Takeaways

- ✓ Silhouette score > 0.5 indicates good segment separation
- ✓ 4-6 segments is usually the actionable sweet spot
- ✓ Cluster on behaviours/attitudes, then DESCRIBE by demographics
- ✓ Validate stability by running multiple times with different seeds

The Bottom Line

Segment transforms undifferentiated customer data into actionable groups with distinct needs. Using proven clustering algorithms with rigorous validation, it creates segments that are both statistically sound and business-relevant. The two-mode workflow (exploration → final) ensures you choose the right number of segments before committing to detailed profiling. The alternative is one-size-fits-all strategy—which means mediocre results with everyone.

About The Research LampPost

Turas is developed and delivered by **The Research LampPost**, an organisational member of the Southern African Marketing Research Association (SAMRA). Duncan Brett is a SAMRA accredited researcher with 30+ years of market research experience. If you have questions about whether Turas is right for your project, if there is interest in the statistical documentation or module-specific information, or you would like a copy of our credentials, please get in touch at duncan@researchlamppost.co.za

Ready to discuss your project?

duncan@researchlamppost.co.za