Turas KeyDriver

Transparent. Trustworthy. Tangible insights.

Identifies which factors most influence your outcome metric (such NPS, satisfaction, likelihood to recommend). Uses machine learning and SHAP values to reveal not just which drivers matter, but how they interact and drive outcomes in complex ways.

What This Module Does

You measure many attributes (quality, price, service, etc.) and one key outcome (satisfaction, loyalty, NPS). But which attributes actually matter? KeyDriver finds the strongest relationships and ranks drivers by importance using gradient boosting and SHAP value-based importance.

Unlike simple correlations, KeyDriver captures non-linear patterns, detects when drivers work together (interactions), handles correlated predictors correctly, and provides individual-level driver importance—not just population averages.

What You Get

Analysis Outputs

- **SHAP importance scores:** How much each driver contributes to predictions
- **Feature importance rankings:** Definitive order from most to least impactful
- **Impact categorization:** High/Medium/Low impact zones for prioritization
- **Individual-level attribution:** What drives each person's response, not just averages

Model Diagnostics

- **R-squared:** How much variance the model explains (>0.5 is typical for survey data)
- **RMSE/MAE:** Average prediction error in original scale units
- **Cross-validation score:** Performance on held-out data (checks overfitting)

How It Works

Gradient Boosting (XGBoost)

Builds many simple decision trees, each learning from previous trees' mistakes. Combined, they capture complex patterns while avoiding overfitting. Learning rate of 0.1, max depth of 6, with early stopping based on validation performance.

SHAP Values

Lundberg's TreeSHAP decomposes each prediction into contributions from each driver. For every respondent, you can see exactly which drivers pushed their score up or down. Aggregate SHAP values show overall importance.

Advantages Over Correlation

Captures non-linear effects (diminishing returns, thresholds). Handles multicollinearity correctly. Detects interactions automatically. Provides individual-level explanations, not just population averages.

Packages Used

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

Package	Why We Use It
xgboost	Industry-leading gradient boosting—used by Kaggle winners, 10M+ downloads
shapviz	Lundberg's TreeSHAP implementation for driver importance interpretation
ggplot2	Professional visualization of driver rankings and importance-performance matrices

Strengths

- ✓ **Captures complex patterns:** Machine learning detects non-linear relationships correlations miss
- ✓ **SHAP-based interpretation:** Advanced explainability shows how each driver contributes
- ✓ **Handles interactions:** Gradient boosting automatically captures when drivers work together
- ✓ **Mixed data types:** Works with continuous, ordinal, and categorical drivers
- ✓ **Fast computation:** Analyzes hundreds of drivers in seconds
- ✓ **Individual insights:** Shows what drives each person, not just population averages

Limitations

- ⚠ **Correlation not causation:** Shows relationships, not necessarily cause-and-effect
- ⚠ **Model complexity:** Gradient boosting less transparent than simple correlations (though SHAP helps)
- ⚠ **Sample size requirements:** Recommend n≥500 for reliable SHAP values
- ⚠ **Continuous outcomes:** Designed for rating scales, NPS, satisfaction scores—use CatDriver for categorical outcomes

Sample Size Requirements

Minimum n≥200 for basic importance rankings. Recommend n≥500 for stable SHAP values and reliable non-linear effect detection. For segment-level driver analysis, need n≥200 per segment.

Best Use Cases

Ideal For

- ✓ Customer satisfaction drivers
- ✓ NPS driver analysis
- ✓ Brand health tracking (what drives consideration/preference)
- ✓ Product optimization (which features drive purchase intent)
- ✓ Service improvement (which touchpoints drive loyalty)

Not Ideal For

- ⚠ Categorical outcomes (won't buy/might buy/will buy)—use CatDriver instead
- ⚠ Small samples (<200 respondents)—unreliable importance estimates
- ⚠ When you need formal coefficient interpretation with confidence intervals

Key Takeaways

- ✓ SHAP values show driver contribution to predictions—higher = more impactful
- ✓ Machine learning captures patterns that simple correlations miss
- ✓ Individual-level analysis reveals that different things matter to different people
- ✓ R^2 >0.5 indicates a good model fit for survey data
- ✓ Focus resources on High impact drivers; maintain Medium; deprioritize Low

The Bottom Line

KeyDriver is your advanced "what matters most" analysis tool. Using machine learning and SHAP values, it reveals not just which factors matter, but how they interact and drive outcomes in complex ways. This goes beyond simple correlation to capture real-world driver dynamics—giving you actionable priorities backed by rigorous statistical modeling.

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