Technical Report — Customer Satisfaction Factor Analysis

Dataset: data/customer_satisfaction_data.csv
Notebook source: notebooks/costomer_satisfaction.ipynb
Author(s): Team 5 — Sophia Gabriela Martínez Albarrán Sibyla Vera Ávila, Regina Pérez Vázquez.
Date: 02/11/2025 (analysis completion)

Executive Summary:

- **Objective**: Identify latent dimensions (factors) underlying customer satisfaction survey data and prioritize business actions that increase satisfaction, NPS, retention, referrals and revenue growth.

- **Key finding:** A five-factor solution explains ~61.85% of the variance. The dominant driver of customer outcomes is the "Technical Excellence & Innovation" factor (Factor 1), which consistently shows the largest predictive effect on satisfaction and related outcomes ($R^2 \approx 0.60–0.64$ for regression models).

Data & Preprocessing:

- **Observations / variables**: 3,235 observations and 23 survey items used for factor analysis (observation-to-variable ratio ≈ 140.7:1).

- **Missing values:** Rows with missing values were dropped prior to analysis.

- **Scaling**: Variables standardized with `StandardScaler` before factor analysis.

Suitability for Factor Analysis:

- **KMO (overall):** 0.959 (excellent sampling adequacy).

- **KMO (per variable):** All items > 0.60.

- **Bartlett's Test of Sphericity:** significant ($p < 0.05$), implying correlations differ from identity.

- **Intercorrelations:** mean absolute correlation ≈ 0.34; ~48.2% of correlations have |r| ≥ 0.3; ~11.6% of variable pairs had |r| ≥ 0.5 (supporting factorability).

Factor Extraction & Selection:

- **Method:** Principal factor extraction with orthogonal rotation (Varimax) for final interpretation; Promax (oblique) tested and reported similar structure with modest inter-factor correlations.

- **Number of factors retained**: 5 (Kaiser criterion: 5 eigenvalues > 1; scree plot elbow after factor 5).

- **Explained variance:** Five factors explain ~61.85% of total variance (Factor1 ≈ 38.0%; Factors 2–5 provide the remaining variance in decreasing order).

Factor Interpretation (business labels):

- **Factor 1 — Technical Excellence & Innovation:** technical_expertise, problem_solving, innovation_solutions, technical_documentation, system_integration. Main driver of outcomes.

- **Factor 2 — Relationship Management & Trust**: trust_reliability, long_term_partnership, communication_clarity, account_manager_responsive, executive_access.

- **Factor 3 — Financial Transparency & Perceived Value:** value_for_money, cost_transparency, roi_demonstration, competitive_pricing, billing_accuracy.

- **Factor 4 — Project Execution & Delivery:** project_management, timeline_adherence, budget_control, quality_deliverables, change_management.

- **Factor 5 — Customer Support & Service Excellence**: support_responsiveness, training_quality, documentation_help.

Factor Solution Quality:

- **Communalities:** Most variables have communalities in the 0.60–0.74 range; highest communalities observed for technical items (~0.73–0.74). Some items (e.g., change_management, value_for_money) show lower communalities (~0.47–0.51), indicating more unique variance.

- **Simple structure**: ~87% of variables load strongly (|loading| ≥ 0.4) on a single factor; cross-loading items are limited (~13%), primarily linking technical and project-delivery constructs.

Predictive Models & Outcomes:

- **Outcomes modeled:** `overall_satisfaction`, `nps_score`, `renewal_likelihood`, `revenue_growth_pct`, `referrals_generated`.

- **Approach**: Linear regression using factor scores as predictors for each outcome.

- **Performance:** R^2 values ranged from moderate to strong (examples reported between ~0.42 and 0.64 depending on method and outcome). For overall satisfaction the factor-based regression achieved $R^2 \approx 0.60–0.64$ and RMSE generally < 0.55.

- **Factor importance (cross-outcome mean absolute impact):** Factor 1 (Technical Excellence & Innovation) ranks highest; Factor 2 (Value/Transparency) and Factor 4 (Project Execution) follow; Factors 3 and 5 show smaller average effects.

Business Insights:

- Customers primarily value high technical competence and innovation (Factor 1). Improvements here yield the strongest gains in satisfaction, NPS, renewal, and referrals.

- Financial transparency and clearly communicated ROI (Factor 3) influence renewal and revenue growth—important for retention and monetization.

- Project delivery reliability (Factor 4) reinforces trust and satisfaction, and is a tactical area for process improvement.

- Relationship management and support (Factors 2 & 5) matter for referrals and NPS, and provide reinforcing effects after technical performance is established.

Strategic Recommendations (prioritized):

- **Priority 1 — Invest in Technical Excellence & Innovation:**

  - Expand technical training, R&D initiatives, and cross-functional innovation teams.

  - Improve quality control and technical documentation to scale capabilities.

- **Priority 2 — Strengthen Financial Transparency & ROI Communication:**

  - Provide clear cost breakdowns and ROI visualizations in client reports.

  - Audit and improve billing accuracy; present cost/value narratives aligned to client outcomes.

- Priority 3 — Optimize Project Delivery:

  - Adopt standardized project management frameworks (Agile / PMI), KPI dashboards for timeline/adherence tracking, and quality gates.

- **Priority 4 — Reinforce Relationship & Support Channels:**

  - Improve account manager responsiveness, executive accessibility, and support workflows (training, documentation, response SLAs).

Action Plan (timeline & KPIs):

- **Short-term (0–6 months):** Rapid response task force for technical reliability; documentation updates; baseline KPIs for support SLAs.

- **Mid-term (6–12 months):** Project KPI dashboards; billing and ROI templates; technical training rollouts.

- **Long-term (1–2 years):** Innovation labs; ongoing ROI measurement framework; continuous improvement cycles.

Suggested KPIs to track progress:

- % change in `overall_satisfaction` and `nps_score` after interventions.

- `renewal_likelihood` improvement and retention rate.

- `revenue_growth_pct` attributable to upsells/new client retention.

- Support response time and ticket resolution rates.

Limitations & Next Steps:

- The analysis used dropped rows (complete-case) which may bias results if missingness is not MCAR. Consider multiple imputation if missingness is material.

- Regression models are linear and assume constant effects; consider generalized models or regularization for robustness.

- Visuals (heatmaps, scree plot, loadings, communalities, factor-score distributions) were generated in the notebook; embedding high-resolution figures into this report (or exporting to PDF) is recommended for stakeholder presentations.

Next analytical extensions:

- Validate factor structure via confirmatory factor analysis (CFA) on a holdout sample.

- Perform cross-validation for predictive models and consider non-linear or ensemble models where appropriate.

- Quantify projected ROI for the top recommended initiatives and use a prioritization matrix combining impact (cross-outcome) and implementation cost/effort.

Appendix — Key code pointers:

- Full reproducible analysis is in `notebooks/costomer_satisfaction.ipynb` (includes data loading, KMO/Bartlett tests, scree plot, factor extraction, Varimax/Promax rotations, communalities, factor scores, regression models and visualizations).

- Recreate environment: ensure `factor_analyzer`, `scikit-learn`, `statsmodels`, `plotly`, `seaborn`, and `pandas` are installed.

Files added:

- `reports/technical_report.md` — this document.