Technical Report: TechnoServe Customer Satisfaction Factor Analysis

1 Introduction

This report summarizes the methodology, data preparation, factor extraction, validation and results documented in the notebook `factor_analysis.ipynb`, which applies **Common Factor Analysis with Maximum Likelihood (FA–ML)** to TechnoServe Solutions' customer satisfaction survey.

The main objective is to identify the *latent dimensions* that drive the 23 satisfaction items and to quantify how these dimensions relate to key business outcomes:

- overall satisfaction,

- renewal likelihood,

- Net Promoter Score (NPS),

- and, as secondary indicators, revenue growth and referrals.

The factor model is designed to complement the business-oriented README and slide deck by providing a transparent, statistically grounded description of how the factors were obtained and how they support the final recommendations.

2 Data Overview

The dataset analysed in the notebook contains:

- **3,400** survey responses,

- corresponding to **850 enterprise customers**,

- **23 satisfaction items** scored on a 1–7 scale,

- and **5 outcome variables**: overall satisfaction, NPS score, renewal likelihood, revenue growth percentage and referrals generated.

The 23 items were originally grouped by the business team into five conceptual domains:

- **Technical Excellence:** technical expertise, problem solving, innovation, technical documentation, system integration.

- **Relationship Management:** account manager responsiveness, executive access, trust/reliability, long-term partnership, communication clarity.

- **Project Delivery:** project management, timeline adherence, budget control, quality of deliverables, change management.

- **Value & Cost:** cost transparency, value for money, ROI demonstration, competitive pricing, billing accuracy.

- **Support & Service:** support responsiveness, training quality, documentation help.

Initial checks in the notebook show that:

- Missingness per item is very low (below 1%), allowing for simple mean imputation.

- All items are numeric, on the same 1–7 Likert scale, and exhibit non-degenerate variance.

- The sample size (3,400 rows for 23 items) is more than adequate for FA–ML.

3 Exploratory Data Analysis (EDA)

The notebook performs a basic EDA before running factor analysis, including descriptive statistics, boxplots and a full correlation heatmap.

3.1 Univariate patterns

- Most satisfaction items are slightly left-skewed: customers tend to give positive evaluations, but all points of the 1–7 scale are used.

- No item is constant or near-constant; the standard deviations are similar and within a reasonable range.

- There are no extreme univariate outliers that would dominate the correlation structure.

3.2 Correlation structure

The correlation matrix of the 23 items shows:

- Very strong correlations among the Technical items (e.g., innovation vs. system integration $\rho \approx 0.67$).

- Strong internal coherence within Relationship and Project Delivery items.

- A visible block for Value & Cost items and a smaller block for Support & Service.

This visual structure strongly suggests that a relatively small number of latent factors is driving most of the covariation in the data, consistent with the original survey design.

3.3 Factorability diagnostics

Two standard diagnostics were computed in the notebook:

- **KMO index:** the global Kaiser–Meyer–Olkin measure is $KMO = 0.959$, which is typically described as "superb" and indicates that partial correlations are small compared to simple correlations.

- **Bartlett's test of sphericity:** the test rejects the null hypothesis that the correlation matrix is the identity (p-value < 0.0001).

Together, these results confirm that factor analysis is appropriate for this dataset.

4 Data Preprocessing

The following preprocessing steps are applied before factor extraction:

1. **Feature selection.** Only the 23 satisfaction items are used to estimate the factor model. Outcome variables are set aside for later correlation analysis.

2. **Imputation.** Missing responses are replaced by the mean of each item using a simple imputer. Given the low missingness, this has minimal impact and avoids dropping cases.

3. **Correlation matrix.** FA–ML is run on the correlation matrix of the items, which effectively standardizes variables and makes the solution scale-free.

After these steps, the correlation matrix is well-conditioned for FA–ML, with no Heywood issues in the initial diagnostics.

5 Factor Analysis Methodology

The notebook implements a structured FA workflow, summarised as follows.

5.1 Number of factors

The number of factors is determined using:

- **Eigenvalues from FA–ML:** five eigenvalues are greater than 1.

- **Scree plot:** a clear elbow occurs after the fifth component.

Both criteria converge on $k = 5$ **factors** as a good compromise between parsimony and explained variance.

5.2 Extraction method

The main extraction is performed using:

- **Common Factor Analysis**,

- **Maximum Likelihood (ML)** estimation.

FA–ML is preferred over PCA because the goal is to model latent constructs that explain the covariation of items, not simply to reduce dimensionality or maximise total variance.

5.3 Rotation strategy

Two rotation schemes are compared:

Varimax (orthogonal): assumes independent factors. This solution explains 50.4% of the common variance.

Promax (oblique): allows correlated factors. This solution explains 60.4% of the common variance.

The Promax solution shows substantial correlations between factors (maximum absolute correlation ≈ 0.617), which is realistic for customer experience—for example, technical performance and delivery quality tend to move together.

Given the higher variance explained and the conceptual plausibility of correlated dimensions, the final interpretation is based on the **five-factor Promax solution**.

6 Factor Solution: Structure and Interpretation

6.1 Extracted factors

Inspection of the Promax loadings leads to the following interpretation:

- **Factor 1 – Technical Excellence.** High loadings on technical expertise, problem solving, innovation, technical documentation and system integration. This factor captures how strong, reliable and advanced the technology side of the service feels to clients.

- **Factor 2 – Value & Cost.** High loadings on billing accuracy, ROI demonstration, competitive pricing, cost transparency and value for money. This dimension represents the perceived fairness of pricing and the clarity of the economic value delivered.

- **Factor 3 – Relationship Quality.** High loadings on account manager responsiveness, executive access, trust/reliability, long-term partnership and communication clarity. It measures how responsive, trustworthy and strategically aligned the relationship is.

- **Factor 4 – Delivery & Project Management.** High loadings on project management, timeline adherence, quality of deliverables, change management and budget control. It reflects whether projects are delivered on time, on scope and with robust quality.

- **Factor 5 – Support & Enablement.** High loadings on support responsiveness, training quality and documentation help. This factor summarises post-sale support and user enablement.

Overall, the empirical factor structure aligns closely with the theoretical design of the survey, which is a strong validity check for the instrument.

6.2 Variance explained

The Promax solution explains approximately:

- 60.4% of the common variance across the 23 items,

- with Factors 1–4 accounting for the majority of this share and Factor 5 adding a smaller but meaningful contribution.

This level of variance explained is typical for multi-item satisfaction instruments in B2B contexts.

7 Link to Business Outcomes

To connect the latent factors to business metrics, the notebook computes Pearson correlations between factor scores and the available outcomes. Correlations are interpreted as association, not as causal effects.

7.1 Correlation with satisfaction, renewal and NPS

The main patterns are:

- **Overall satisfaction.** The strongest associations are Delivery (Factor 4, $\rho \approx 0.651$), Technical Excellence (Factor 1, $\rho \approx 0.644$) and Relationship Quality (Factor 3, $\rho \approx 0.615$). Value & Cost (Factor 2, $\rho \approx 0.526$) and Support (Factor 5, $\rho \approx 0.444$) also contribute but play a secondary role.

- **Renewal likelihood.** The same pattern emerges: Delivery ($\rho \approx 0.515$), Technical ($\rho \approx 0.510$) and Relationship ($\rho \approx 0.493$) are the main levers, followed by Value & Cost ($\rho \approx 0.435$) and Support ($\rho \approx 0.362$).

- **NPS.** Delivery ($\rho \approx 0.445$), Relationship ($\rho \approx 0.428$) and Technical ($\rho \approx 0.425$) again form the top trio, with Value & Cost and Support slightly lower but still positively associated.

In other words, *what is delivered, how it is delivered and the quality of the relationship* are the strongest drivers of whether customers are satisfied, renew and recommend TechnoServe Solutions.

7.2 Illustrative predictive check

As an illustrative check, the notebook fits a linear regression model where renewal likelihood is predicted by the five factor scores. The model yields:

- $R^2_{\text{train}} \approx 0.391$,

- $R^2_{\text{test}} \approx 0.379$.

This confirms that the latent factors capture a substantial portion of the variance in renewal intentions, without evident overfitting. Similar patterns are observed when inspecting simple correlations with overall satisfaction and NPS.

8 Discussion

From a methodological standpoint, the FA–ML workflow implemented in the notebook is coherent and well-justified:

- Data quality and factorability were checked with KMO and Bartlett.

- The number of factors was decided using both eigenvalues and the scree plot.

- Extraction used Maximum Likelihood, and rotation compared orthogonal vs. oblique options.

- The chosen Promax solution respects the empirical correlations between underlying dimensions.

From a business standpoint, the analysis provides a compact view of the customer experience:

- **Delivery & Project Management** and **Technical Excellence** are the primary levers for satisfaction and renewal.

- **Relationship Quality** acts as a strong amplifier: even good delivery feels fragile if the relationship is weak.

- **Value & Cost** and **Support & Enablement** are important hygiene factors and boosters, particularly for recommendations.

These insights align with intuition in enterprise technology consulting: customers are most loyal when solutions work, are delivered predictably and there is a reliable team behind them.

9 Conclusion

The `factor_analysis.ipynb` notebook successfully applies multivariate factor analysis to TechnoServe Solutions' customer satisfaction data and produces a stable, interpretable five-factor model.
In summary:

- The 23 items are well modelled by five correlated factors matching the intended survey dimensions.

- The Promax solution explains around 60% of common variance, with sensible loadings and communalities.

- The latent factors are strongly associated with overall satisfaction, renewal likelihood and NPS, especially Delivery, Technical and Relationship.

- The model offers a statistically grounded backbone for the qualitative recommendations and the 30–60–90 day action plan described in the accompanying documents.

This Technical Report consolidates the notebook's methodology, diagnostics and results into a concise narrative suitable for academic evaluation and business discussion.