Research Proposal

Design and Creative Technologies

Torrens University, Australia

Student: Luis Guilherme de Barros Andrade Faria - A00187785

Subject Code: REM 502

Subject Name: Research Methodologies

Assessment No.: 3

Title of Assessment: Research Proposal

Lecturer: Dr. Bushra Naeem

Date: Dec 2025

Table of Contents

1. Introduction 3
2. Literature Review 4
3. Methodology and Methods 7
3.1. Design Paradigm 8
3.2. Data Collection 8
3.3. Data Processing Timeline 9
3.4. Quantitative Analysis 10
3.5. Triangulation Question 11
4. Conclusion 11
5. Proposed Timeline 12
6. Appendices 13
6.1. Appendix A – Dataset Overview 13
6.2. Appendix B – Data Preparation Pipeline 14
6.3. Appendix C – Descriptive Statistics 14
6.4. Appendix D – Correlation Preview 15
7. References 19

Exploring the Relationship between Net Promoter Score and Revenue Growth in Healthcare Clinics.

1. Introduction

This research aligns with Torrens University's "Here for Good" ethos by addressing healthcare accessibility and quality improvement through evidence-based decision-making.

By establishing empirical links between patient feedback and clinic performance, this study equips healthcare providers—particularly those serving diverse communities—with actionable intelligence to optimize resource allocation, improve patient outcomes, and ensure sustainable business models that can continue serving their communities long-term.

The research problem stems from a critical gap: while NPS is widely adopted in healthcare to measure patient satisfaction, its actual correlation with business viability (and thus clinic sustainability) remains empirically unvalidated. Without this validation, clinics risk investing in metrics that do not predict financial health, potentially jeopardizing their capacity to serve patients.

Research Questions:

- RQ1: To what extent is NPS correlated with monthly revenue growth in healthcare clinics?
- RQ2: Can NPS trends over time predict short-term revenue fluctuations?

Hypothesis:

- H0: There is no statistically significant correlation between NPS scores and monthly revenue.
- H1: There is a positive correlation between NPS scores and monthly revenue.

Objectives:

- Collect and prepare Net Promoter Score (NPS) and monthly revenue data from Pro-Corpo's clinics to ensure accuracy and comparability across time periods of the available data.
- Conduct quantitative analysis – including descriptive statistics, correlation, and regression – to examine the relationship between NPS scores and revenue performance.
- Perform correlation analysis (Pearson or Spearman) to quantify the strength and direction of the NPS to revenue relationship).
- Develop a practical ICT framework that demonstrates how NPS-based metrics can inform strategic decision-making and performance evaluation in healthcare.

2. Literature Review

2.1 Patient Experience as a Business Driver (200-250 words)

- Synthesize findings from Godovykh & Pizam (2023) about patient experience influencing loyalty

- Connect to Shankar & Yip (2024) showing NLP can extract operational insights

- **Gap you're addressing:** These studies show *operational* improvements but don't link to *financial* outcomes

2.2 The Net Promoter Score Debate (150-200 words)

- Summarize Dawes (2024) critique: NPS measures intention not behavior, has cultural bias

- Acknowledge NPS's widespread adoption despite limitations

- **Your research positioning:** Testing whether NPS actually predicts revenue in healthcare context

2.3 AI-Enabled Sentiment Analysis in Healthcare (200-250 words)

- Cover Alkhnbashi et al. (2024) and Xiao et al. (2022) on fine-grained sentiment analysis - Mention technical feasibility but note lack of business outcome validation

- **Your contribution:** This study provides the quantitative baseline that future AI-enhanced systems require

2.4 Emotions and Customer Engagement (150-200 words)

- Brief coverage of Angelis et al. (2024) on anger vs. fear in engagement - Healthcare relevance: different emotions drive different retention behaviors

- **Future extension:** Your NPS data has text comments that could enable emotion analysis (Assessment 3 future work)

2.5 Research Gap and Study Positioning (250-300 words)

``` Three converging gaps define this study's contribution:

1. **Metric Validation Gap:** NPS is widely used but empirically unvalidated as a predictor of clinic financial performance (Dawes, 2024).

2. **Healthcare-Business Intelligence Gap:** Studies demonstrate operational insights from patient feedback (Shankar & Yip, 2024) but fail to connect sentiment to revenue, retention, or growth metrics.

3. **ICT Foundation Gap:** AI-enabled sentiment analysis systems (Alkhnbashi et al., 2024) assume underlying correlations between patient sentiment and business outcomes without empirically establishing these relationships first.

This study addresses these gaps by conducting quantitative correlation analysis between NPS scores and clinic-level revenue across 27,000 patient feedback records spanning 36 months. Unlike prior research that remains within operational or technical domains, this study bridges patient experience measurement with financial performance—a prerequisite for evidence-based healthcare management and AI system design.

The availability of Pro-Corpo's longitudinal dataset (2022-2025) containing both NPS scores and monthly revenue data provides a unique opportunity to test these relationships empirically. This positions the research within ICT R&D practice: validating core assumptions before building complex computational systems. ```

**Visual Elements to Add:**
```

- **Figure 1:** Conceptual framework showing NPS → Patient Experience → Revenue pathway

- **Table 1:** Summary of literature gaps (adapted from your Assessment 1, Section 3)

3. Methodology and Methods

Healthcare business research operates at the intersection of human experience and organizational performance, requiring methodological approaches that balance measurement rigor with contextual understanding. The choice of methodology fundamentally shapes what can be known: quantitative methods enable hypothesis testing and generalization across populations, while qualitative methods reveal mechanisms and meanings that numbers alone can't capture (Creswell & Plano Clark, 2023). In ICT research and development (R&D), this tension is particularly salient when translating patient feedback, a qualitative phenomenon, into business intelligence metrics.

This study adopts a **pragmatic paradigm** (Morgan, 2014), prioritizing practical problem-solving over skepticism. Three methodological approaches were evaluated:

Methodology	Description	Strengths	Weakness
Qualitative	Explores human meaning through interviews or thematic coding	Rich context and interpretive depth.	Limited generalizability; prone to researcher bias.
Quantitative	Employs numerical measurement, hypothesis testing, and statistical inference.	Objectivity, replicability, scalability.	May overlook cultural tone.

Mixed Methods	Integrate both qualitative and quantitative strands.	Triangulation improves validity; merges AI outputs with human interpretation.	Requires time and data integration skills.

Given that this study seeks to establish a baseline relationship between existing business metrics (NPS and revenue), a quantitative approach has been chosen as the most appropriate. The availability of large-scale data (27,000 records across 36 months) makes statistical analysis both feasible and methodologically sound (Pallant, 2020). Qualitative methods, while valuable for understanding why correlations exist, require resources (patient interviews, thematic coding) beyond this study's scope. Mixed or qualitative methods may be relevant for future research that integrates NLP-derived sentiment with financial outcomes, but such extensions depend on the foundational quantitative correlation established in this study.

This methodological choice reflects ICT R&D principles: iterative development from simple to complex systems. By demonstrating statistical relationships first, subsequent research can build AI-enabled sentiment analysis with confidence that the underlying NPS-revenue connection merits computational investment.

3.1.Design Paradigm

Given that the research questions seek to quantify relationships and test predictions, a quantitative correlational design is most suitable to establish statistical validity and support evidence-based conclusions (Field, 2018).

3.2.Data Collection

- **Primary Source:** Anonymized operational dataset provided by Pro-Corpo Estetica (https://procorpoestetica.com.br/), comprising ≈ 27 000 records (2022–2025).

- **Variables**: textual feedback, NPS scores, month/year, clinic ID, and monthly revenue.
- **Data Security**: stored on encrypted drives compliant with the Australian Privacy Act (1988) and Brazilian Data Protection Law.
- **Authorization**: formal company consent letter ensuring confidentiality and academic use only.

The dataset originates from Pro-Corpo's post-service Net Promoter Score (NPS) program, which automatically invites clients to provide feedback within 24 hours of receiving treatment. Respondents can identify themselves or remain anonymous and answer four brief questions: (1) a 1-to-10 satisfaction rating, (2) optional comments, (3) confirmation of the store visited, and (4) optional mention of staff members for praise or concern. This process has generated approximately 27 000 records collected between 2022 and 2025, providing a rich source of structured (scores, store, month) and unstructured (text feedback) data. Monthly revenue data for each store are also available, enabling correlation between customer sentiment, NPS, and financial performance.

3.3. Data Processing Timeline

The plan is to develop the work in 4 phases detailed below:

Milestone	**Description**
Data extraction	• Extract NPS survey responses (score, date, clinic ID, optional text comments) • Extract monthly revenue records (clinic ID, month, total revenue) • Timeframe: January 2022 – December 2024 (36 months)
Data cleaning	• Remove duplicates (based on timestamp + clinic ID) • Handle missing values: ○ NPS missing: exclude record (cannot impute satisfaction scores)

	○ Revenue missing: interpolate if isolated gap (linear interpolation); exclude clinic-month if systematic missingness • Outlier detection: flag revenue values >3 standard deviations from clinic mean for review
Data transformation	• Aggregate NPS to clinic-month level (mean, median, standard deviation, response count) • Normalize revenue for clinic size: Revenue per appointment (if appointment data available) • Create derived variables such as **NPS category**: Detractors (0-6), Passives (7-8), Promoters (9-10), **Revenue growth**: Month-over-month percentage change and **Lagged NPS**: NPS from 1 and 2 months prior (for temporal analysis)
Integration	• Merge datasets on: "clinic_id + year + month" • Validate temporal alignment (ensure NPS survey dates precede revenue measurement) • Final dataset structure: Each row = one clinic-month observation

3.4. Quantitative Analysis

- **Step 1: Check Assumptions:** Before running statistical tests, the data must meet certain conditions (Field, 2018): Normality, Linearity, Outliers.
- **Step 2: Describe the Data: C**alculate basic statistics for each clinic and time period like Average NPS score and revenue, show distribution patterns.
- **Step 3: Test Correlation (RQ1):** Measure how strongly NPS and revenue move together.
- **Step 4: Test Prediction (RQ2):** Use regression to see if NPS can predict future revenue.

- **Step 5: Check Robustness:** Verify results are reliable by Re-running tests excluding, testing different time periods separately (2022 vs 2023 vs 2024).

3.5. Triangulation Question

- **Temporal triangulation:** Analyzing data across different time periods (from 2022 to 2025) to see if patterns hold
- **Clinic-**level triangulation: comparison between multiple Pro-Corpo stores at different locations
- **Methodological triangulation:** using both Pearson (parametric) and Spearman (non-parametric) correlation as a robustness check.

4. Conclusion

This research proposal establishes a systematic framework for empirically testing whether Net Promoter Score correlates with clinic-level revenue performance in healthcare settings. By employing correlation analysis and regression modeling on 27,000 patient feedback records across 36 months, the study will produce one of three possible outcomes:

- **Scenario 1: Strong positive correlation (r > 0.70):** NPS demonstrates robust predictive validity for revenue, validating its continued use and suggesting simpler tracking systems suffice for business intelligence.
- **Scenario 2: Moderate correlation (0.30 < r < 0.70):** NPS provides partial but incomplete business insight, indicating that supplementary metrics or AI-enhanced sentiment analysis systems would add value.
- **Scenario 3: Weak/no correlation (r < 0.30):** NPS fails as a revenue predictor, suggesting healthcare providers should invest in alternative patient experience

measurement systems—potentially NLP-based sentiment analysis that captures richer feedback dimensions.

Regardless of outcome, this quantitative foundation enables evidence-based decisions about customer feedback systems in healthcare. If correlations are strong, the study validates existing practices; if weak, it justifies investment in more sophisticated AI-enabled alternatives explored in Assessment 1's literature review.

The methodology's triangulation strategy (temporal, clinic-level, and methodological) ensures findings are robust across contexts, while the 12-week timeline (detailed in Section 5) positions completion as feasible within standard research project constraints. This progression from foundational correlation analysis (Assessment 2) to implementation- ready insights (Assessment 3) exemplifies iterative ICT R&D practice.

5. Proposed Timeline

The overlapping structure enables iterative refinement: preliminary findings from correlation analysis inform regression model specifications, while early visualization prototypes guide final dashboard design.

- Weeks 1-2: Data extraction and cleaning (Milestone: validated dataset ready for analysis)
- Weeks 3-4: Exploratory data analysis and assumption testing (Milestone: confirmed data meets parametric test requirements)
- Weeks 5-6: Correlation analysis for RQ1 (Milestone: Pearson/Spearman correlations computed)
- Weeks 7-8: Regression modeling for RQ2 (Milestone: predictive models validated)

- Weeks 9-10: Robustness checks and sensitivity analysis (Milestone: findings confirmed across subgroups)
- Weeks 11-12: Report writing and visualization development (Milestone: final deliverables complete).

Figure 4: Gantt Chart showing these phases with overlapping bars

6. Appendices

6.1. Appendix A – Dataset Overview

Column	Description	Example
store_name	Clinic identifier	"LAPA"
month	Reference month of transaction	2025-04
nps_score	Net Promoter Score from post-service	9
revenue	Monthly clinic revenue (BRL)	85.200
responses	Number of NPS responses in that period	112

Data are anonymized and aggregated by store and month, ensuring no personally identifiable information is retained.

6.2. Appendix B – Data Preparation Pipeline

```
# Merging by store and month
df = pd.merge(nps, revenue, on=["Store", "month"], how="inner")

# Computing monthly averages
df["avg_nps"] = df.groupby("Store")["NPS"].transform("mean")

# Dataframe preview
df.head()
```

✓ 0.0s

	month	Store	NPS	count_of_responses	revenue	count_of_orders	avg_nps
0	2024-01	BELO HORIZONTE	65.000000	40	111643.77	182	65.000000
1	2024-01	CAMPINAS	91.044776	67	114131.31	246	75.986733
2	2024-01	COPACABANA	80.281690	71	225878.13	250	74.070324
3	2024-01	IPIRANGA	69.230769	65	234789.84	305	69.230769
4	2024-01	ITAIM	30.769231	13	99281.93	128	55.494505

Figure 1: Code example of pandas' dataframe merging

The dataset was merged using store and month as common keys. Only complete entries were retained to maintain consistency in correlation analysis.

6.3. Appendix C – Descriptive Statistics

```
df.describe()[["NPS", "revenue"]]
```

[34] ✓ 0.0s

...

	NPS	revenue
count	67.000000	67.000000
mean	75.820018	179285.282836
std	28.704604	112676.584657
min	-100.000000	48331.610000
25%	66.666667	95715.870000
50%	81.818182	158653.140000
75%	92.397388	222530.025000
max	100.000000	601662.020000

Figure 2: Descriptive statistics of Net Promoter Score and revenue across all clinics.

6.4. Appendix D – Correlation Preview

The analysis has not been finished yet, but still it is possible to see the beginning of the studies of correlation between NPS score and Revenue income generated on the stores. More to come in the next analysis.

```
corr = df[["NPS", "revenue"]].corr(method="pearson")
corr
```

✓ 0.0s

	NPS	revenue
NPS	1.000000	0.010176
revenue	0.010176	1.000000

Figure 3: Preliminary Pearson correlation matrix..

Heatmap:

```python
# Add correlation heatmap between revenue and NPS
import seaborn as sns
import matplotlib.pyplot as plt

# Create correlation heatmap
plt.figure(figsize=(5, 4))
correlation_matrix = df[['revenue', 'NPS']].corr()
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', center=0,
            square=True, linewidths=0.5)
plt.title('Correlation between Revenue and NPS')
plt.tight_layout()
plt.show()

# Print correlation coefficient
correlation = df['revenue'].corr(df['NPS'])
print(f"Correlation coefficient: {correlation:.3f}")
```

✓ 0.0s

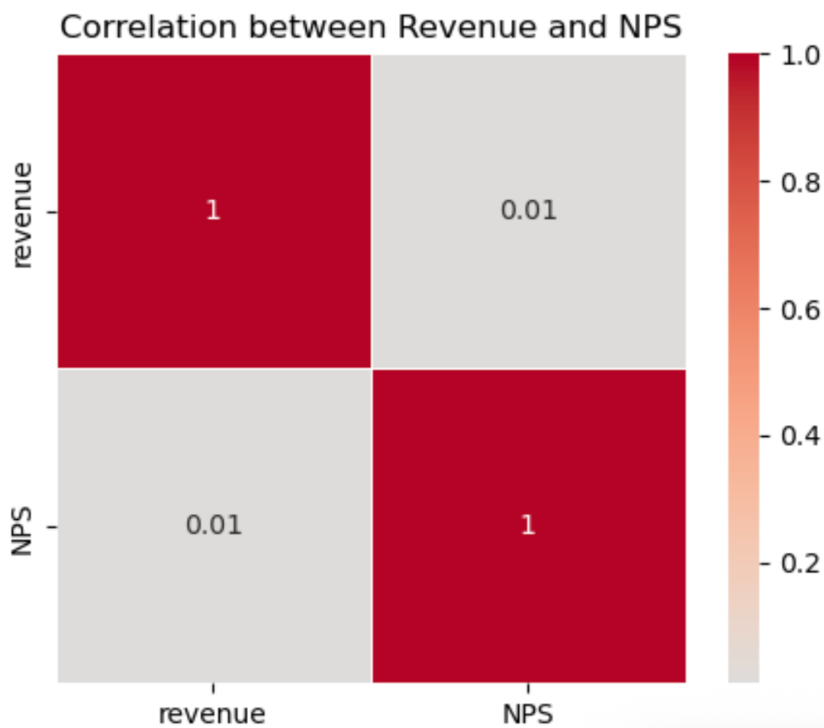

Figure 4: Preliminary Correlation Heatmap between Revenue and NPS score..

Scatter Plot

```python
# Add scatter_nps_vs_revenue between NPS and revenue
def scatter_nps_vs_revenue(df):
    """Create scatter plot of NPS vs revenue"""
    plt.figure(figsize=(5, 4))
    plt.scatter(df['NPS'], df['revenue'], alpha=0.6)
    plt.xlabel('NPS')
    plt.ylabel('revenue')
    plt.title('NPS vs revenue')
    plt.grid(True, alpha=0.3)
    plt.show()

# Call the function
scatter_nps_vs_revenue(df)
```

✓ 0.0s

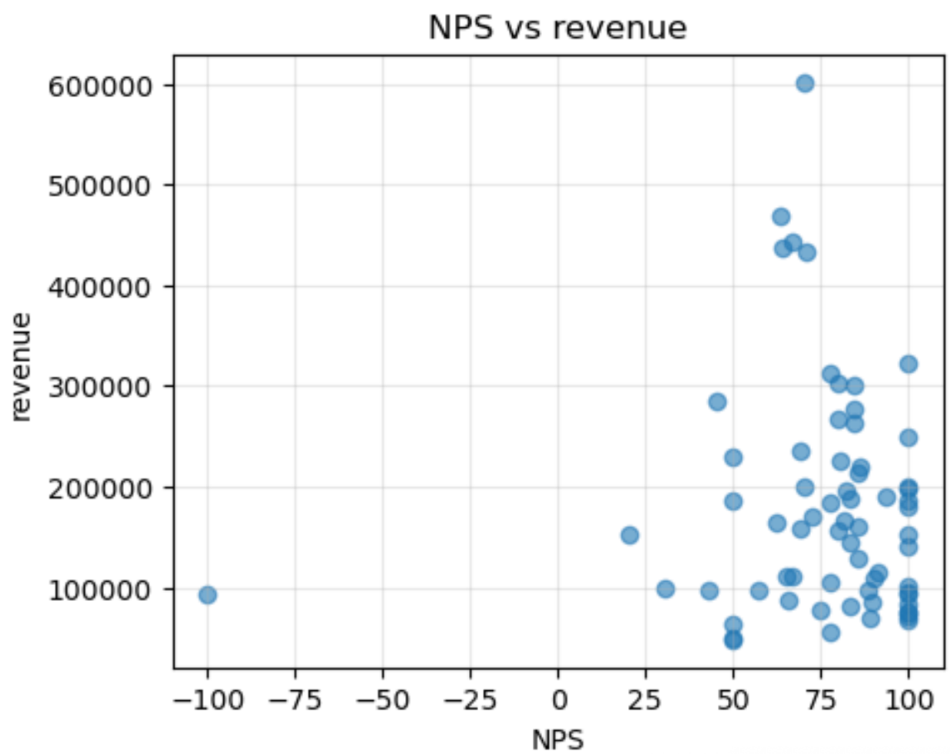

Figure 5: Preliminary Scatter Plot Graphic between Revenue and NPS score.

End of Appendix Section

Statement of Acknowledgment

I acknowledge that I have used the following AI tool(s) in the creation of this report:

- OpenAI ChatGPT (GPT-5): Used to assist with outlining, refining structure, improving clarity of academic language, and supporting APA 7th referencing conventions.

I confirm that the use of the AI tool has been in accordance with the Torrens University Australia Academic Integrity Policy and TUA, Think and MDS's Position Paper on the Use of AI. I confirm that the final output is authored by me and represents my own critical thinking, analysis, and synthesis of sources. I take full responsibility for the final content of this report.

7. References

Bryman, A. (2016). *Social research methods* (5th ed.). Oxford University Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge.

Creswell, J. W., & Plano Clark, V. L. (2023). *Designing and conducting mixed methods research* (4th ed.). SAGE Publications.

Dawes, J. G. (2024). *The net promoter score: What should managers know? International Journal of Market Research*, 66(2–3), 182–198. https://doi.org/10.1177/14707853231195003

Field, A. (2018). *Discovering statistics using IBM SPSS Statistics* (5th ed.). SAGE Publications.

Godovykh, M., & Pizam, A. (2023). *Measuring patient experience in healthcare.* International Journal of Hospitality Management, 112, 103405. https://doi.org/10.1016/j.ijhm.2022.103405

Morgan, D. L. (2014). *Pragmatism as a paradigm for social research*. Qualitative Inquiry, 20(8), 1045–1053. https://doi.org/10.1177/1077800413513733

Pallant, J. (2020). *SPSS survival manual* (7th ed.). Routledge.

Polonsky, M. J., & Waller, D. S. (2019). Quantitative data analysis. *In Designing and managing a research project* (4th ed., pp. 222-254). SAGE Publications. https://doi.org/10.4135/9781544316499

Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., & Wesslén, A. (2012). *Experimentation in software engineering*. Springer.

Wohlin, C., & Runeson, P. (2021). Guiding the selection of research methodology in industry–academia collaboration in software engineering. *Information and Software Technology*, 140, 106678. https://doi.org/10.1016/j.infsof.2021.106678