LATENT DRIVERS OF EMPLOYEE RETENTION: A FACTOR ANALYSIS CASE STUDY USING R IN A U.S. TECH FIRM

1. Background

A mid-sized technology company headquartered in Austin, Texas, was facing higher-than-expected voluntary attrition among junior and mid-level employees. Although HR conducted regular engagement surveys, the leadership found the raw item-level analysis too granular to act on. They requested a factor analysis using R to extract major themes behind satisfaction and intent to stay.

2. Objective

- To reduce 35 survey items into core latent factors explaining retention-related sentiment
- To identify the strongest predictors of intent to stay using regression on factor scores
- To simplify future surveys by retaining only the most informative items per factor

3. Data Used

Source: Internal HR engagement survey (Q2 2024)

Sample Size: 540 employees (response rate 87%) **Variables**:

- 35 Likert-scale items (1 = Strongly Disagree, 5 = Strongly Agree)
- Outcome: Binary indicator for "Intent to Stay" (1 = Yes, 0 = No)

Themes covered:

- Team communication
- Career growth
- Work-life balance
- Trust in leadership
- Recognition
- Autonomy
- Compensation satisfaction

4. Methodology

4.1 Data Preparation

- Cleaned inconsistent scale formatting and missing values
- Normalized responses to a 1–5 scale
- Checked suitability using:

```
library(psych)
KMO(hr_data)  # Result: 0.84
cortest.bartlett(hr_data)  # p < 0.001
```

4.2 Factor Extraction

- Used principal axis factoring via psych::fa()
- Scree plot + parallel analysis supported a 5-factor solution
- Applied varimax rotation

```
fa_result <- fa(hr_data, nfactors = 5, rotate = "varimax", fm = "pa")
```

4.3 Factor Naming

Based on item loadings:

1. **Leadership Trust**
2. **Work Autonomy**
3. **Recognition and Feedback**
4. **Career Development**
5. **Work-Life Balance**

4.4 Regression on Retention

- Built logistic regression with factor scores as predictors of "Intent to Stay"

```
glm_model <- glm(Intent_Stay ~ F1 + F2 + F3 + F4 + F5, data = scored_df, family = "binomial")
```

5. Results Summary

Factor Name	Cronbach's α	Significant in Retention Model?
Leadership Trust	0.86	Yes ($p < 0.01$)
Work Autonomy	0.79	Yes ($p < 0.05$)
Recognition & Feedback	0.84	Yes ($p < 0.01$)
Career Development	0.74	No
Work-Life Balance	0.77	No

- Factors explained **68.2%** of total variance
- Model AUC = **0.81**

6. Interpretation and Recommendations

- Trust in leadership and frequent feedback drive retention far more than generic work-life programs
- 12 of the 35 items accounted for 93% of the factor structure's variance
- Suggested dropping repetitive or low-loading questions for future surveys
- Advised that "Recognition" and "Autonomy" be emphasized in performance reviews and manager training

7. Deliverables

- **PDF Report** (32 pages):
  - Factor loadings, rotation matrices, regression coefficients
  - Survey design recommendations
  - Visuals: Scree plot, factor score distribution, ROC curve
- **Reusable R Scripts**:
  - hr_factor_analysis.R
  - predict_retention_glm.R
  - survey_redesign_helper.R

- **Optional Module**:
    - R Shiny-based internal tool to visualize department-wise factor scores and retention risk

8. Business Outcome

- HR redesigned its engagement survey using only the top 3 factors
- Leadership launched a manager training series focused on Recognition and Autonomy
- Voluntary attrition fell by **17% in the next two quarters**
- The factor analysis framework is now applied annually to monitor workforce sentiment shifts