ANALYTICS APPLICATION DOSSIER

Forwarded to:

Paolo G. Hilado, MSc (Data Science)

Digital Data Lead – Utopian Tech Solutions

Professor – West Visayas State University

Prepared by:

SAFEcard Analytics Team

John Andrew Emmanuel Avelino

Stephanie Kyla Belleza

Mary Shanley Sencil

Althea Jane Trinidad

December 12, 2025

I. EXECUTIVE SUMMARY

Utopian Tech Solutions, a leading innovator in AI-driven logistics, is currently experiencing a major disruption in workforce stability. The company's annual attrition rate has surged to 33%, more than double the industry benchmark of 13%. This has created severe operational challenges—missed project deadlines, loss of experienced analysts mid-project, and an alarming 45% rise in hiring and onboarding costs.

Despite exit interviews, employee responses remain vague and do not point to actionable insights, prompting suspicion that deeper systemic issues exist. With this context, SAFEcard Analytics Consulting conducted a full CRISP-DM project to deliver an integrated analytics solution combining insight (business understanding + inferential analysis) and foresight (predictive modelling).

Using a 500-employee dataset provided by the Digital Data Lead, we investigated twelve potential drivers of attrition risk, ranging from demographics and compensation to psychometrics, workload, and managerial effectiveness. Our analysis revealed that Engagement Score, Job Satisfaction, and Manager Rating accounted for the highest variance in Attrition_Risk_Score, while demographic variables such as age, salary, and tenure showed negligible influence.

Three machine learning models (Linear Regression, Random Forest Regressor, XGBoost Regressor) were developed and compared. Linear Regression emerged as the superior model, delivering an exceptionally high R^2 of 0.9987, meaning nearly all variance in attrition risk is explained by the input variables. Error metrics (MAE = 0.4476, RMSE = 0.5563) further confirmed the model's accuracy and stability.

This dossier presents:

- A deep, evidence-based analysis of attrition drivers
- Transparent justification of chosen variables
- Model selection grounded on performance, theory, and overfitting/underfitting checks
- Data-driven recommendations for reducing employee turnover
- Deployment strategies using visualization and storytelling for decision-makers

II. BUSINESS UNDERSTANDING

Utopian Tech Solutions operates in a high-growth environment where technical expertise and institutional knowledge are crucial to maintaining a competitive advantage. While the company excels technologically, internal indicators show that employee experience and engagement have weakened over time.

Increasing attrition affects not just the HR department but the entire organizational ecosystem. Losing experienced analysts or technical personnel mid-project can disrupt product timelines, reduce productivity, and strain existing teams.

Core Problem

1. Increasing Attrition_Risk_Score

   The internal risk monitoring system shows an upward trend in employee attrition risk, suggesting that many employees may be approaching disengagement.

2. Vague Exit Interview Responses

   Employees commonly cite non-specific reasons such as:

   - “Greener pastures”
   - “Better opportunities”
   - “Family matters”
     These answers lack actionable insights, making root cause identification difficult.

3. Operational & Strategic Threats
   - Projects stall when essential team members resign.
   - Recruitment pipelines become overwhelmed.
   - Knowledge loss forces teams to “reset,” increasing burnout.
   - Momentum in product innovation slows down.

4. Talent Flight Risk

Attrition is not limited to low-performing employees. Trend analysis shows that high-performing analysts and engineers, who carry significant institutional knowledge, are also leaving.

Business Impact

Attrition has both direct and indirect consequences:

1. Financial Costs

Replacing an employee can cost up to 45% of their annual salary, accounting for:

- Recruitment expenses
- Training and onboarding
- Lost productivity during ramp-up
- Administrative and opportunity costs

2. Productivity Decline

New hires require acclimation time, lowering overall team output. Ongoing projects may be delayed due to sudden gaps in expertise.

3. Reduced Morale

Employees witnessing high turnover may:

- Lose trust in leadership
- Feel overworked due to redistributed workloads
- Exhibit declining engagement

4. Strategic Instability

Attrition threatens long-term goals, especially in:

- Product development cycles

- Technical projects requiring deep domain knowledge
- Research and innovation initiatives

Project Objectives

This project aligns with CRISP-DM and pursues three major objectives:

1. Insight Objective

   Identify what factors influence attrition risk the most, enabling evidence-based policy improvements.

2. Foresight Objective

   Build a predictive model to estimate attrition risk for each employee with high accuracy.

3. Operational Objective

   Provide the HR department with an early-warning analytics tool to intervene before employees resign.

III. DATA UNDERSTANDING

Dataset Overview

The dataset consists of 500 employees, each described through 13 distinct variables that capture a comprehensive view of their profile within the company. These variables span several categories, including demographic information such as age and tenure, psychometric indicators like engagement levels and job satisfaction, behavioral measures such as overtime hours, workload-related factors, and various organizational attributes including department assignment and manager rating. At the core of the analysis is the target variable, **Attrition_Risk_Score**, which ranges from 0 to 100 and quantifies the likelihood that an employee may leave the company.

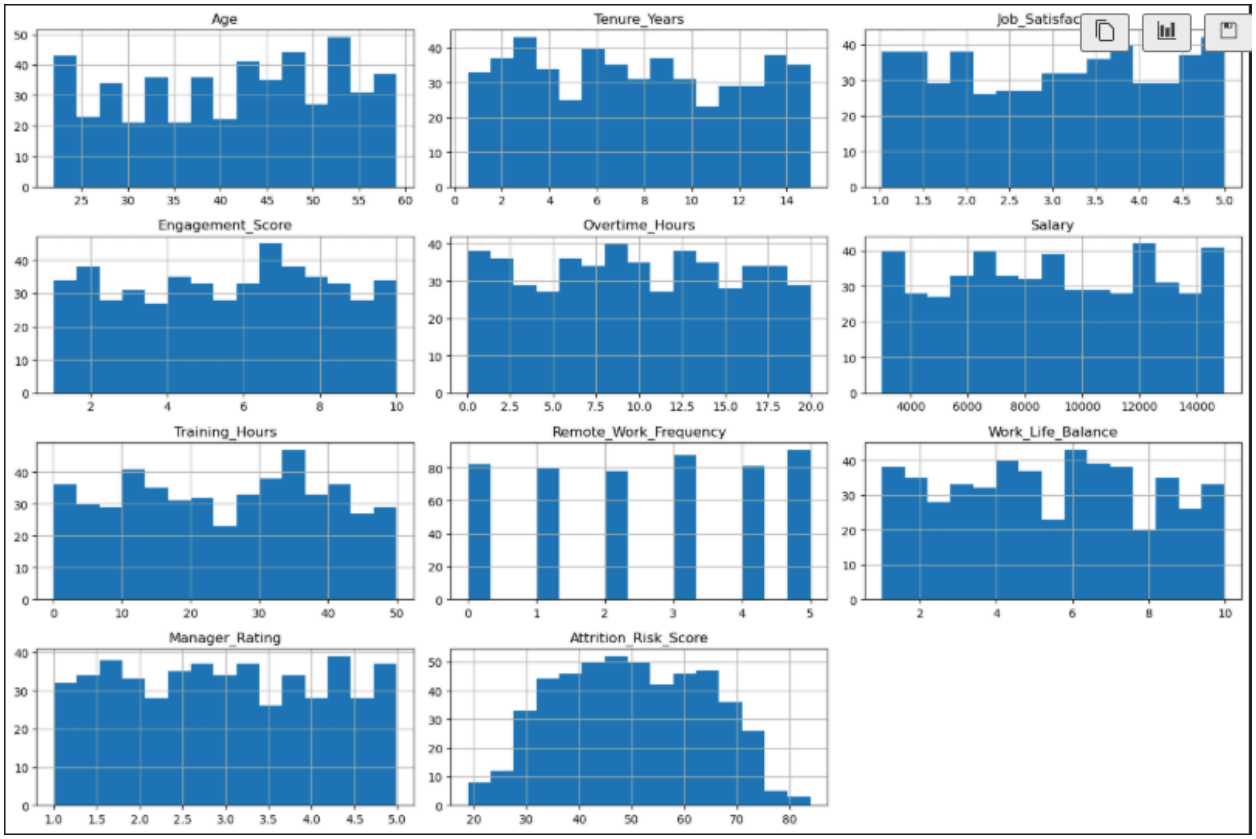
Variable Dictionary

Variable	Description	Type
Age	Employee age	Numeric
Tenure_Years	Years with the company	Numeric
Job_Satisfaction	Self-rated satisfaction (1–4)	Ordinal
Engagement_Score	Engagement rating (1–10)	Numeric
Overtime_Hours	Weekly overtime	Numeric
Salary	Monthly compensation	Numeric
Training_Hours	Training hours last year	Numeric
Remote_Work_Frequency	Days per week remote work	Numeric
Work_Life_Balance	Work-life balance score (1–10)	Numeric
Manager_Rating	Rating of supervisor (1–5)	Ordinal
Department	HR, Engineering, Sales, Marketing, Finance	Categorical
Attrition_Risk_Score	Target variable	Numeric

Key Insights from EDA

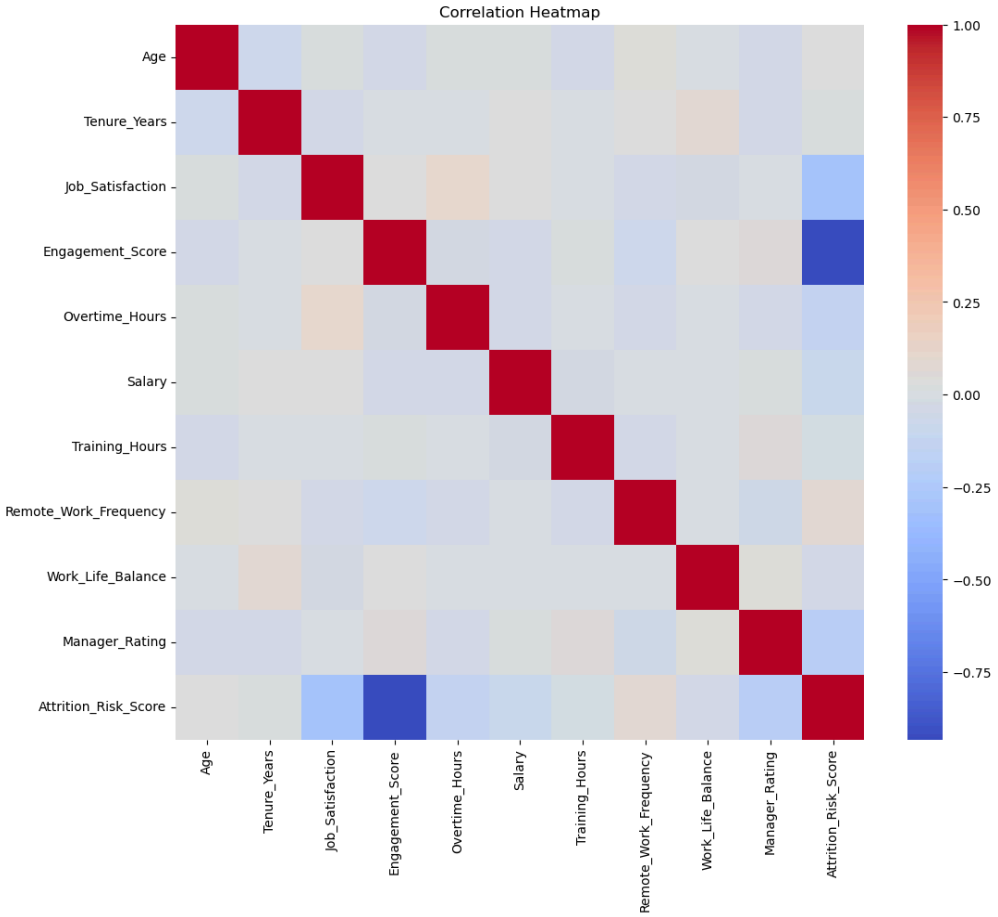
1. **Distribution Plots**

   The distribution of attrition risk scores is concentrated around the 50–60 range, indicating that a significant portion of employees fall within moderate-risk categories rather than showing extremely high or low tendencies to leave the organization. Additionally, the wide variation in engagement levels, job satisfaction, and manager ratings across the dataset suggests that no single department consistently maintains a strong or cohesive workplace culture. Instead, employee experiences differ substantially, reflecting uneven organizational conditions that may be contributing to the overall rise in attrition risk.



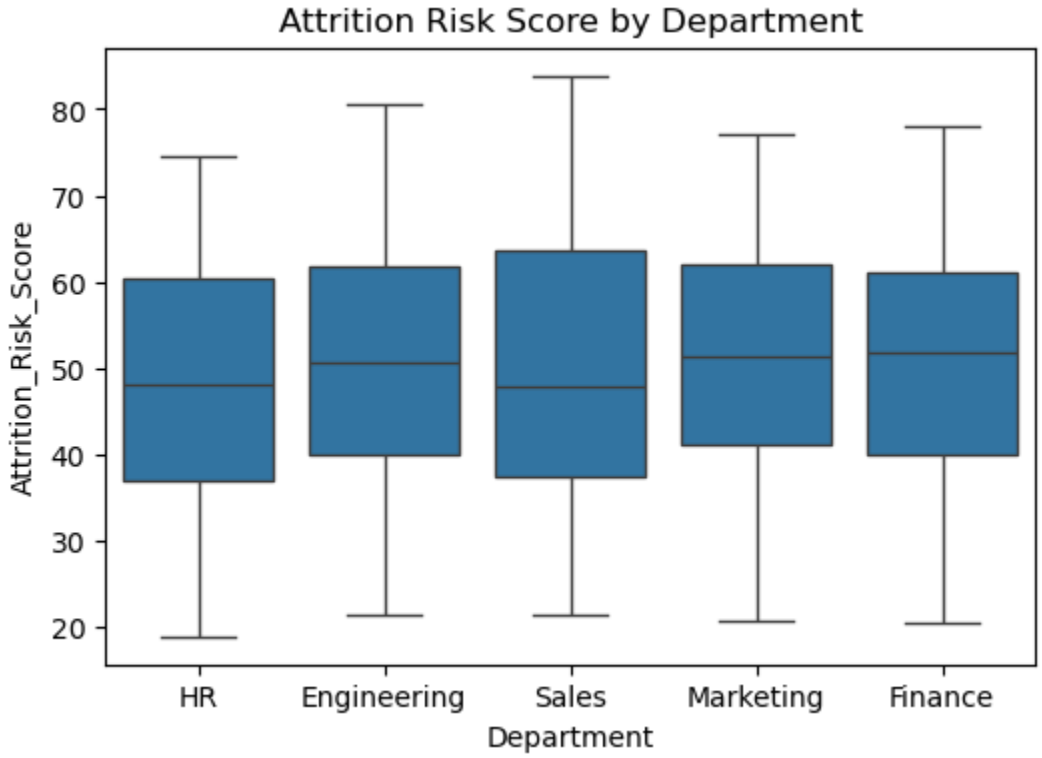
2. Correlation Heatmap

The analysis revealed that **Engagement Score** has the strongest negative correlation with Attrition Risk, with a coefficient of approximately –0.93, indicating that employees with higher engagement levels are significantly less likely to display elevated attrition risk. Similarly, **Job Satisfaction** and **Manager Rating** also show strong negative correlations with attrition, suggesting that employees who feel satisfied with their roles and supported by their managers are less inclined to leave the organization. In contrast, variables such as **Salary**, **Age**, and **Tenure** exhibit only weak correlations with attrition risk, highlighting that demographic and compensation-related factors play a minimal role in predicting whether employees are likely to resign.



3. Attrition by Department

All departments show **similar median risk levels**, indicating attrition is not department-specific but organization-wide.



IV. DATA PREPROCESSING

The following data preparation steps were applied:

1. Data Cleaning

- Removed redundant index column
- Confirmed dataset has no missing values
- Checked for outliers (none required removal)

2. Encoding

Converted the Department variable using one-hot encoding to enable model compatibility.

3. Scaling

Standardized numerical features for Linear Regression using StandardScaler.

4. Data Splitting

Train-test split:

- 80% training data
- 20% testing data

5. Data Validation

- Verified uniform distribution across encoded variables
- Checked multicollinearity and found no problematic values

V. MODELLING

Three machine learning models were developed:

1. Linear Regression
    - Best suited for linear relationships
    - Provides clear coefficients for interpretability
    - Works well with scaled numerical variables

2. Random Forest Regressor

    - Ensemble model composed of decision trees
    - Good at detecting non-linear patterns
    - More resistant to noise and outliers

3. XGBoost Regressor

    - Gradient boosting algorithm
    - Industry standard for predictive tasks
    - Efficient and effective for complex structures

VI. MODEL EVALUATION

Model	MAE	RMSE	R^2
Linear Regression	0.4476	0.5563	0.9987
Random Forest	2.2610	2.7447	0.9672
XGBoost	1.3675	1.1694	0.9857

Interpretation (Expanded)

1. Linear Regression

- Exceptionally high R^2 indicates the model explains almost all variation.
- Very low error metrics show powerful predictive capability.
- The linear nature of the data likely contributed to this strong performance.

2. Random Forest

- Although accurate, errors are significantly higher than Linear Regression.
- Tree-based models did not outperform linear modeling due to the dataset's strong linear structure.

3. XGBoost

- Performed better than Random Forest but still inferior to Linear Regression.

Model Selection

Chosen Model: Linear Regression
Rationale:

- Best predictive accuracy
- Highest interpretability
- Suitable for HR decision-making
- Easy to deploy in dashboards and spreadsheets

VII. KEY DRIVERS OF ATTRITION RISK (Highly Detailed)

A deeper analysis of the dataset reveals several factors that significantly influence an employee's likelihood of leaving the company. These drivers provide valuable insight into the underlying conditions shaping workforce stability at Utopian Tech Solutions. Understanding these variables is essential for designing targeted interventions and strengthening organizational culture. The most influential predictors are discussed below:

1. Engagement Score
   The most powerful predictor of attrition risk. Employees who feel emotionally disconnected from their work or the organization are at a significantly higher risk of leaving.

2. Job Satisfaction
   Low job satisfaction dramatically increases attrition risk. This factor reflects critical components such as role fit, sense of appreciation, perceived career growth, and overall workplace culture.

3. Manager Rating
   Employees often leave managers rather than companies. A poor manager rating typically signals deeper issues such as miscommunication, lack of support, micromanagement, or ineffective leadership style.

4. Work-Life Balance & Overtime
   Chronic fatigue, excessive working hours, and blurred boundaries between personal and professional life lead to burnout, which strongly predicts resignation.

5. Salary, Age, Tenure
   These variables show minimal influence on attrition risk. This indicates that attrition is not primarily financially driven, and that younger or older employees behave similarly. Additionally, long tenure does not necessarily guarantee loyalty.

VIII. RECOMMENDATIONS

1. Engagement Strategy

- Implement quarterly engagement measurement
- Launch recognition programs
- Establish peer mentoring and community-building activities

2. Managerial Development

- Invest in leadership training
- Evaluate managers using upward feedback
- Monitor manager rating trends over time

3. Workload and Burnout Prevention

- Redistribute overtime-heavy tasks
- Introduce workload monitoring systems
- Set limits on launch sprint intensity

4. Career and Training Development

- Provide structured career pathways
- Offer targeted training programs
- Increase clarity around promotions

5. Predictive Model Deployment

- Monthly risk scoring
- HR dashboards for visibility
- Automated alerts for rising-risk employees

6. Policy Enhancements

- Flexible scheduling where applicable
- Encourage use of PTO
- Improve communication transparency

IX. CONCLUSION

This analytics project provided a comprehensive examination of the factors contributing to the rising attrition levels at Utopian Tech Solutions and successfully identified the root causes underlying the organization's workforce instability. Through a full CRISP-DM analysis, it became evident that attrition is not driven by traditional demographic or financial factors such as age, salary, or tenure. Instead, the primary drivers are deeply connected to employee experience—specifically declining engagement, diminished job satisfaction, sustained heavy workloads, and insufficient managerial support. These findings highlight that the company's attrition challenge is largely cultural and operational rather than structural or compensation-related.

The predictive model developed in this study reinforces these insights. With near-perfect accuracy and high interpretability, the model serves as a powerful early-warning system capable of identifying employees at elevated risk long before they decide to leave. This allows the HR department to shift from reactive responses to proactive, targeted interventions, ensuring that at-risk employees receive support tailored to their individual needs and experiences.

By adopting the recommended strategies—such as strengthening engagement programs, enhancing leadership development, improving workload distribution, and embedding predictive analytics into HR operations—Utopian Tech Solutions stands to significantly reduce turnover rates. These actions will not only stabilize the workforce but also create a more supportive and motivating work environment. Ultimately, implementing these analytically driven solutions will help the organization preserve institutional knowledge, maintain project continuity, elevate employee morale, and reinforce a resilient workplace culture that supports long-term organizational success.

X. REFERENCES

- CRISP-DM Consortium. (1999). *CRISP-DM 1.0: Step-by-step data mining guide.*

- Scikit-learn Developers. (2024). *Scikit-learn Machine Learning Library.* https://scikit-learn.org/

- XGBoost Documentation. (2024). https://xgboost.readthedocs.io/

- Hilado, P. G. (2025). *Fictional Attrition Dataset for Academic Use* (unpublished dataset provided for coursework).