Knowledge Discovery for HR Turnover Risk and Financial Exposure

Derek I. Medina Rosario
Master in Computer Science
Advisor: Nelliud Torres Batista, DBA
Polytechnic University of Puerto Rico
Graduate Project EXPO, May 2026

Abstract *— Employee turnover creates financial and operational challenges through recruiting, onboarding, training, productivity loss, and reduced workforce stability. This project applied a simplified Knowledge Discovery approach to estimate employee turnover risk and financial exposure using workforce indicators. A synthetic dataset of 15 employee records was created to simulate department, monthly salary, overtime status, job satisfaction, and years at company. Excel was used to transform the dataset by calculating turnover risk score, turnover risk class, retention priority segment, annual salary, and estimated financial exposure. Power BI was then used to create a basic data model, data analysis expressions measures, and an interactive dashboard. The results identified high-risk employee groups, department-level financial exposure, and workforce risk patterns related to overtime, job satisfaction, and tenure. The project demonstrates how simple workforce data can be transformed into decision-support insights for human resources planning and retention strategies.*

Keywords *— Data Modeling, Human Resources Analytics, Knowledge Discovery, Turnover Risk.*

INTRODUCTION

Employee turnover is an important workforce issue because it affects staffing stability, productivity, employee morale, and organizational costs. When employees leave, organizations may need to invest time and money in recruiting, selecting, onboarding, and training replacements. Turnover may also create indirect costs such as productivity loss, workload redistribution, and loss of organizational knowledge.

Traditional human resources reporting often focuses on turnover after it has already occurred. While this information is useful, it does not always help organizations identify risk early or estimate the financial impact of potential employee loss. Human resources analytics can help organizations move from reactive reporting to proactive workforce planning by using employee data to identify risk indicators and support decision-making [1], [2].

This project develops a simplified Knowledge Discovery model that estimates employee turnover risk and financial exposure. The model uses basic workforce indicators such as overtime, job satisfaction, and years at company to classify employees into turnover-risk categories. Excel was used for data preparation and calculated fields, while Power BI was used to model, visualize, and interpret workforce risk patterns.

PROBLEM STATEMENT

Organizations may understand that employee turnover is costly, but they may not have a simple method for estimating which workforce factors increase turnover risk or how much potential turnover could cost. Small organizations may not have advanced analytics systems, machine learning tools, or complete employee datasets. However, even a simplified Knowledge Discovery approach can help human resources professionals organize workforce data, identify risk patterns, classify employees into risk groups, and estimate financial exposure.

The main problem addressed in this project is: How can a simplified Knowledge Discovery model estimate employee turnover risk and financial exposure using basic workforce indicators? This project addresses the problem by creating a synthetic employee dataset, applying rule-based classification, modeling the data in Power BI, and visualizing risk patterns and estimated financial exposure through an interactive dashboard.

Project Objectives

- Create a simple synthetic human resources dataset.
- Develop a rule-based turnover risk scoring model.
- Classify employees into low, medium, and high turnover-risk categories.
- Segment employees into stable, monitoring, and priority retention groups.
- Estimate annual salary and financial exposure.
- Build a basic Power BI data model using fact and dimension tables.
- Create measures for employee count, high-risk employees, average risk score, and financial exposure.
- Visualize workforce risk patterns through an interactive dashboard.
- Demonstrate how Knowledge Discovery can support workforce planning and retention decisions.

Background

Human resources analytics, also known as people analytics or workforce analytics, uses employee data to support decisions related to hiring, retention, compensation, performance, and workforce planning. Data analytics can help organizations discover patterns, summarize trends, and support better decision-making [2]. In human resources, analytics may be used to evaluate turnover, absenteeism, engagement, staffing needs, and workforce costs.

Knowledge Discovery refers to the process of transforming data into meaningful patterns and actionable insights [3]. In this project, the Knowledge Discovery process included data selection, preprocessing, transformation, rule-based classification, pattern discovery, and interpretation. This structure allowed basic human resources data to be converted into turnover-risk classifications and financial exposure insights.

Employee turnover is a common area of human resources analytics because it directly affects organizational performance and cost. Public employee attrition datasets, such as the IBM HR Analytics Employee Attrition dataset, show that workforce variables such as job role, overtime, income, job satisfaction, tenure, and work-life balance are commonly used in attrition analysis [4]. Although this project does not use the IBM dataset directly, it supports the idea that attrition can be studied using workforce indicators.

This project uses a synthetic dataset instead of real employee data. Synthetic data was selected for simplicity and privacy. Real employee data may contain sensitive information and should be handled carefully. A synthetic dataset allows the project to demonstrate a Knowledge Discovery workflow without exposing private employee records.

Methodology

This project followed a simplified Knowledge Discovery process: data selection, preprocessing, transformation, rule-based classification, pattern discovery, and interpretation. A synthetic human resources dataset was created in Microsoft Excel. The dataset included 15 employee records and the following variables: employee ID, department, monthly salary, overtime status, job satisfaction, years at company turnover risk score, turnover risk score, turnover risk class, retention priority segment, annual salary, and estimated financial exposure.

A rule-based scoring model was developed to classify employees into low, medium, and high turnover-risk categories. The model assigned risk points based on three workforce indicators: overtime, job satisfaction, and tenure. Employees received two points if they worked overtime, two points if their job satisfaction score was two or lower, and one point if they had one year or less at the company.

The risk score formula was: Turnover Risk Score = Overtime Points + Job Satisfaction Points + Tenure Points. Employees were then classified according to the rule shown in Table 1.

Table 1
Turnover Risk Class Classification

Risk Score	Turnover Risk Class
0-1	Low
2-3	Medium
4-5	High

Financial exposure was estimated by calculating annual salary and estimated financial exposure. Annual salary was calculated as monthly salary multiplied by twelve. Estimated financial exposure was calculated using 50% of annual salary as a conservative replacement cost assumption. This assumption was selected because the Society for Human Resource Management reports that replacement cost can range from 50% to 200% of annual salary, depending on the role level [1].

The results were then modeled in Power BI. A basic star-schema structure was created using one fact table and four-dimension tables. The fact table, FactEmployeeRisk, contained employee-level records, calculated risk scores, annual salary, and estimated financial exposure (see Table 2). Dimension tables were created for department, turnover risk class, overtime status, and retention priority segment. Star-schema modeling is recommended for Power BI semantic models because it improves filtering and analysis structure [4]. Measures were developed in Power BI using data analysis expressions to calculate total employees, high-risk employees, high-risk percentage, total financial exposure, high-risk financial exposure, and average risk score. These measures were used to create dashboard visuals and support interactive filtering by department, risk class, overtime status, and retention segment (see Table 2).

Figure 1
Star Schema Structure

Dataset Description

The dataset was synthetic and included 15 employees from four departments: Human Resources, Sales, Information Technology, and Operations. Each employee record included monthly salary, overtime status, job satisfaction score, and years at company. The use of synthetic data allowed the project to focus on the design of the model, the logic of the calculations, and the interpretation of the results without requiring access to confidential employee records.

The dataset was expanded through calculated fields. Turnover risk score was calculated from overtime, job satisfaction, and tenure. Turnover risk class categorized employees as low, medium, or high risk. Retention priority segment grouped employees into stable, monitoring, and priority retention segments. Annual salary and estimated financial exposure connected workforce risk to business cost.

Table 2
Fact Risk Table, Power BI

Employee ID	Department	Monthly Salary	Overtime	Job Satisfaction	Years at Company	Turnover Risk Score	Turnover Risk Class	Retention Priority Segment	Annual Salary	Estimated Financial Exposure
DMR001	HR	$ 3,500.00	No	4	4	0	Low	Stable Segment	$ 42,000.00	$ 21,000.00
DMR002	Sales	$ 4,200.00	Yes	2	1	5	High	Priority Retention Segment	$ 50,400.00	$ 25,200.00
DMR003	IT	$ 5,200.00	Yes	3	2	2	Medium	Monitoring Segment	$ 62,400.00	$ 31,200.00
DMR004	Operations	$ 3,800.00	No	5	5	0	Low	Stable Segment	$ 45,600.00	$ 22,800.00
DMR005	Sales	$ 3,900.00	Yes	2	1	5	High	Priority Retention Segment	$ 46,800.00	$ 23,400.00
DMR006	HR	$ 3,600.00	No	3	2	0	Low	Stable Segment	$ 43,200.00	$ 21,600.00
DMR007	IT	$ 6,100.00	No	4	6	0	Low	Stable Segment	$ 73,200.00	$ 36,600.00
DMR008	Operations	$ 4,100.00	Yes	2	1	5	High	Priority Retention Segment	$ 49,200.00	$ 24,600.00
DMR009	Sales	$ 4,500.00	Yes	3	2	2	Medium	Monitoring Segment	$ 54,000.00	$ 27,000.00
DMR010	IT	$ 5,800.00	No	5	4	0	Low	Stable Segment	$ 69,600.00	$ 34,800.00
DMR011	HR	$ 3,400.00	Yes	2	1	5	High	Priority Retention Segment	$ 40,800.00	$ 20,400.00
DMR012	Operations	$ 4,000.00	No	4	3	0	Low	Stable Segment	$ 48,000.00	$ 24,000.00
DMR013	Sales	$ 4,700.00	Yes	1	1	5	High	Priority Retention Segment	$ 56,400.00	$ 28,200.00
DMR014	IT	$ 5,300.00	Yes	3	3	2	Medium	Monitoring Segment	$ 63,600.00	$ 31,800.00
DMR015	Operations	$ 3,700.00	No	3	2	0	Low	Stable Segment	$ 44,400.00	$ 22,200.00

Table 3
Dataset Variables

Variable	Description
Employee ID	Unique employee identifier
Department	Employee department
Monthly Salary	Employee's monthly salary
Overtime	Whether the employee works overtime
Job Satisfaction	Satisfaction score from 1 to 5
Years at Company	Employee tenure
Turnover Risk Score	Calculated score based on overtime, satisfaction, and tenure
Turnover Risk Class	Low, medium, or high turnover-risk classification
Retention Priority Segment	Stable, monitoring, or priority retention group
Annual Salary	Monthly salary multiplied by twelve
Estimated Financial Exposure	50% of annual salary

Power BI Data Model

Power BI was used to move the project beyond a flat spreadsheet and into a basic analytical data model. The model contained one fact table and four-dimension tables. The fact table stored the employee-level measures, while the dimension tables stored the categories used for slicing and filtering. This structure allowed the dashboard to summarize risk by department, risk class, overtime status, and retention segment.

Table 4
Power BI Model Structure

Table	Type	Purpose
FactEmployeeRisk	Fact table	Stores employee records, salaries, scores, and exposure
DimDepartment	Dimension table	Filters results by department
DimRiskClass	Dimension table	Filters results by low, medium, and high risk
DimOvertime	Dimension table	Filters results by overtime status
DimSegment	Dimension table	Filters results by retention priority segment

The use of a fact table and dimension tables supports the Knowledge Discovery objective because it organizes raw and calculated fields into an analytical structure. Instead of viewing the dataset only as rows of employee records, the model allows the user to explore patterns by risk class, department, overtime status, and estimated financial exposure.

Figure 2
Power BI HR Turnover Risk and Financial Exposure Dashboard

Results and Discussion

The model classified employees into low, medium, and high turnover-risk categories based on overtime, job satisfaction, and years at company. Employees with overtime, low job satisfaction, and low tenure received higher risk scores. These indicators were selected because they are practical workforce variables that can be monitored by organizations.

The first analysis summarized the number of employees in each risk category. The model classified seven employees as low risk, three employees as medium risk, and five employees as high risk. This type of summary can help human resources professionals understand whether turnover risk is concentrated among a small group of employees or spread across the workforce.

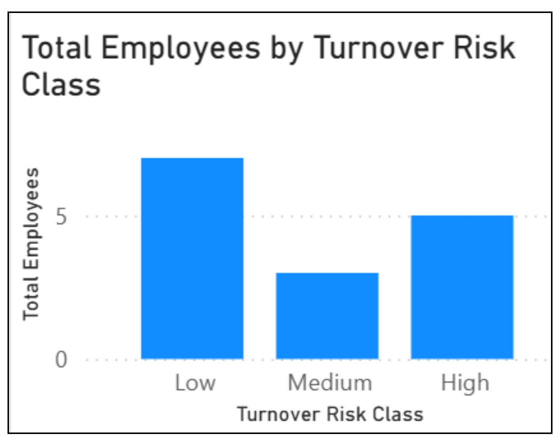

Figure 3
Employees by Turnover Risk Class

The second analysis summarized high-risk employees by department. In this synthetic data set, Sales had the highest number of high-risk employees, with three employees classified as high risk. This finding suggests that turnover risk may not be evenly distributed across departments.

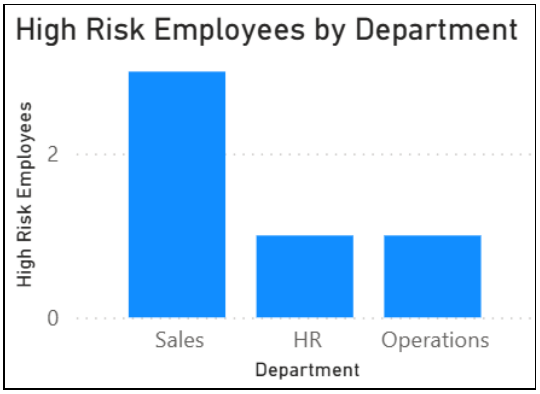

Figure 4
High-Risk Employees by Department

The third analysis estimated financial exposure by department. The total estimated financial exposure across all 15 employees was $394,800. The estimated exposure among high-risk employees was $121,800. Although the Information Technology department had no high-risk employees in the synthetic dataset, it had the highest total financial exposure by department because its employees had higher monthly salaries. This finding shows why risk class and financial exposure should be interpreted together.

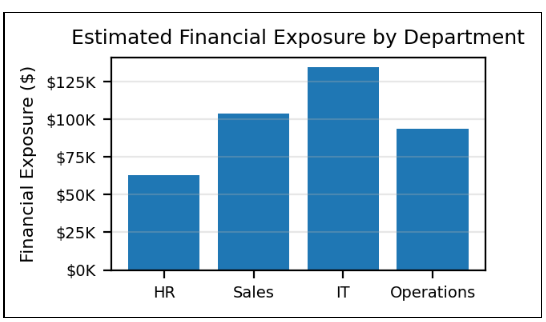

Figure 5
Estimated Financial Exposure by Department

Power BI improved the interpretation of the results by presenting risk and cost information through interactive visuals. The dashboard allowed employee turnover risk to be reviewed by department, risk class, overtime status, and retention segment. The Knowledge Discovery analysis supported pattern discovery by comparing turnover risk class across department, overtime status, job satisfaction, and average risk score.

Table 5
Financial Exposure by Turnover Risk Class

Turnover Risk Class	Number of Employees	Total Financial Exposure
Low	7	$183,000
Medium	3	$90,000
High	5	$121,800

The results suggest that a simple Knowledge Discovery model can support workforce planning by identifying risk groups, estimating financial exposure, and helping organizations prioritize retention strategies. The model does not predict turnover with certainty, but it provides a structured method for identifying potential areas of concern.

Financial Impact Analysis

The financial impact analysis estimated the potential cost of employee turnover using monthly salary and a conservative replacement cost assumption. Annual salary was calculated by multiplying monthly salary by twelve. Estimated financial exposure was calculated as 50% of annual salary.

This approach provides a simple way to connect human resources data to business costs. For example, if an employee earns $4,000 per month, the annual salary is $48,000. Using a 50% replacement cost assumption, the estimated financial exposure would be $24,000. The formula used was: Estimated Financial Exposure = Monthly Salary x 12 x 0.50.

The total estimated financial exposure was summarized by department and by turnover risk class. This helped identify where financial exposure was concentrated. A department with several high-risk employees and higher salaries may represent greater financial concern than a department with fewer high-risk employees or lower estimated replacement costs.

Table 5
Financial Exposure by Department

Department	High-Risk Employees	Total Financial Exposure
HR	1	$63,000
Sales	3	$103,800
IT	0	$134,400
Operations	1	$93,600

Power BI measures were used to calculate total financial exposure, high-risk financial exposure, and average risk score. These measures allowed the dashboard to respond dynamically to filters such as department, risk class, overtime status, and retention segment [5]. This section supports the practical value of the project because it shows how human resources analytics can help organizations understand not only who may be at risk, but also how much turnover may cost.

RECOMMENDATIONS

- Monitor employees with overtime because overtime may indicate workload pressure or burnout risk.
- Track job satisfaction scores regularly to identify early signs of disengagement.
- Pay attention to employees with low tenure because newer employees may require stronger onboarding and retention support.
- Review department-level financial exposure to prioritize human resources interventions.
- Use Power BI dashboards to support recurring workforce risk reviews.
- Interpret risk classifications as decision-support signals, not as final judgments about employees.

The model should be used as a decision-support tool, not as a final judgment about individual employees. Human review is still necessary before making any employee-related decision.

ETHICAL CONSIDERATIONS

Because this project relates to employee data, ethical considerations are important. Real human resources analytics projects should protect employee privacy, avoid unfair treatment, and ensure that employee data is used responsibly. Predictive or rule-based risk models should not be used to punish employees. Instead, they should support positive interventions such as workload review, retention conversations, engagement improvement, and career development.

Synthetic data was used in this project to avoid privacy concerns. If the model were applied to real organizational data in the future, the organization should limit access to sensitive information, review the model for fairness and bias, and ensure that employees are not negatively affected by automated or rule-based risk classifications.

CONCLUSIONS

This project developed a simplified Knowledge Discovery model to estimate employee turnover risk and financial exposure using workforce indicators. A synthetic dataset was created with employee department, monthly salary, overtime status, job satisfaction, and years at company. A rule-based scoring model classified employees into low, medium, and high turnover-risk groups.

The project demonstrated that basic workforce indicators can be transformed into a practical decision-support model. Excel was used to prepare the dataset and create calculated fields, while Power BI was used to build a data model, create measures, and visualize workforce risk patterns. Overtime, low job satisfaction, and low tenure were used as risk signals. Annual salary and estimated financial exposure connected workforce risk to financial impact.

The main contribution of the project is its practical application. The model can be recreated with accessible business tools and understood by human resources professionals without advanced programming. It supports proactive workforce planning by helping organizations identify risk patterns, estimate financial exposure, and prioritize retention strategies.

Future Work

Future work may expand the model by using real organizational human resources data and adding additional variables such as absenteeism, performance ratings, promotion history, training participation, engagement survey results, and manager feedback. The Power BI dashboard could also be expanded with additional pages for trend analysis, department-level drilldowns, and retention intervention tracking.

A future version of this project could compare the rule-based approach with machine learning models such as logistic regression, decision trees, or random forests. This would allow the organization to evaluate whether advanced analytics improves prediction accuracy. Fairness and bias analysis should also be included before applying the model to real employee decisions.

References

[1] Society for Human Resource Management. (2025, Jan. 21). *The Myth of Replaceability: Preparing for the Loss of Key Employees* [Online]. Available: https://www.shrm.org/mena/executive-network/insights/myth-replaceability-preparing-loss-key-employees.

[2] IBM. (n. d.). *What is big data analytics?* [Online]. Available: https://www.ibm.com/think/topics/big-data-analytics.

[3] P. Subhash. (n. d.). *IBM HR Analytics Employee Attrition & Performance* [Online]. Available: https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset.

[4] Microsoft. (2024, Dec. 30). *Understand star schema and the importance for Power BI* [Online]. Available: https://learn.microsoft.com/en-us/power-bi/guidance/star-schema.

[5] Microsoft. (2025, Oct. 22). *Create measures for data analysis in Power BI Desktop* [Online]. Available: https://learn.microsoft.com/en-us/power-bi/transform-model/desktop-measures