HEALTHCARE WORKER MENTAL HEALTH REPORT

Executive Summary

In early 2021, approximately 1,500 healthcare workers in Singapore resigned—a sharp rise from pre-pandemic levels, particularly among foreign staff. Overwork, burnout, and unused leave were major contributors (Channel News Asia, 2021). This report analyzes the factors influencing turnover intention among healthcare workers, with the goal of informing strategies to enhance staff retention and well-being.

Dataset

The analysis uses a synthetic dataset of 5,000 entries from Kaggle, representing healthcare workforce mental health data. Key variables include stress level, burnout frequency, job satisfaction, access to Employee Assistance Programs (EAPs), mental health absences, and turnover intention. The dataset was clean and well-structured, with no missing or duplicate values. However, as a synthetic dataset, it may not fully reflect real-world complexity, and some contextual factors are absent.

Exploratory Data Analysis (EDA)

Descriptive analysis used histograms, box plots, and bar charts to explore numeric and categorical variables. Key findings include:

- 66% of employees expressed turnover intention.
- Only 17% reported low or no burnout.
- 28% lacked access or awareness of EAPs.
- Heavy workload was the most commonly cited workplace stressor.

Inferential analysis was conducted to validate patterns observed in the data. A Chi-Square p-value matrix revealed statistically significant associations between several categorical variables—Employee Type, Department, Workplace Factor, and Burnout Frequency—and Turnover Intention. However, EAP access was not statistically significant ($p = 0.22$), though the fact that 28% of employees lack access suggests a potential gap in support worth further investigation.

For numeric variables, One-Way ANOVA tests showed highly significant differences ($p < 0.001$) in Stress Level, Job Satisfaction, and Mental Health Absences between those intending to leave and those who do not—confirming strong relationships with turnover intention.

Tableau visualizations were used to support exploratory analysis and communicate findings interactively. A dashboard was developed with department-level filters, allowing users to explore turnover intention, burnout frequency, and job satisfaction across various workplace factors. This enables stakeholders to drill down into department-specific trends and identify potential intervention areas.

<u>Predictive Modeling</u>

A logistic regression model was developed to predict turnover intention. Chosen for its interpretability, the model effectively identifies high-risk individuals and supports targeted intervention. It performed consistently across training and test sets, making it a viable tool for practical use.

<u>Recommendations</u>

- Support high-risk employees identified by the model with regular check-ins and resources.
- Address workload issues through staffing adjustments and process improvements.
- Expand and promote EAP access to improve awareness and usage across the organization.

<u>Next Steps</u>

- Monitor intervention outcomes using pre- and post-intervention data.
- Collaborate with HR and healthcare administrators to design tailored strategies.
- Enhance data collection to improve model performance and deepen future insights.

1. Problem, Goals, and Audiences

1.1. Problem Statement

Healthcare workers are experiencing high levels of stress, burnout, and job dissatisfaction, which impact their well-being contribute to rising attrition – posing risk to both staff retention and the quality of patient care.

The goal of this analysis is to identify patterns and predictors of burnout and turnover intention, enabling department heads to take early, data-driven action to support staff and reduce attrition risk.

1.2. Goals

- Retain healthcare staff by identifying and addressing the root causes of burnout.
- Support staff well-being through data-driven insights into mental health trends.
- Ensure continuity and quality of patient care by maintaining a stable and engaged workforce.
- Provide predictive tools and actionable recommendations to help department heads make informed decisions.

1.3. Criteria for Success

- Clear identification of key burnout and mental health risk factors.

- Accurate predictive model for identifying staff at risk of burnout.
- Practical and evidence-based recommendations for intervention and retention.
- Usability of insights through an interactive dashboard.

1.4. Target Audience

- Department head in the healthcare system who will be looking into reduce staff turnover and enhance workplace well-being as early intervention.

2. Data Sources, Cleaning, Dictionary and Limitation

2.1. Data Sources

- The dataset sourced from Kaggle is titled Healthcare Workforce Mental Health Dataset
- It has 5,000 synthetic employee records and contains key workforce-related mental health factors, including stress levels, burnout frequency, job satisfaction, mental health absences, and turnover intention.
- These would serve as a foundation for us to explore and visualizing workplace mental health trends in the healthcare industry

2.2. Data Cleaning

The dataset provided was already well-structured and free from common data quality issues. However, standard data validation steps were performed to ensure readiness for analysis:

- Duplicate Check: No duplicates were found.
- Missing Values: All variables were complete with no missing entries.
- Data Types: All columns were in appropriate formats (e.g., categorical, numeric).
    - All categorical variables had clean, standardized value sets.

2.3. Data Dictionary

S/N	Variable Name	Description	Type	Value/Scale
1	Employee Type	Role in the healthcare system	Categorical	Registered Nurse, Physician, Medical Assistant etc.
2	Workplace Factor	Primary workplace challenge faced by the employee	Categorical	Heavy Workload, Poor Work Environment, Career Stagnation etc.
3	Stress Level	Self-reported level of workplace-related stress	Ordinal (1-10)	1 = No stress, 10 = Severe stress
4	Burnout Frequency	Frequency of experiencing burnout symptoms	Categorical	Never, Occasionally, Often

5	Job Satisfaction	Self-reported job satisfaction	Ordinal (1-5)	1 = Very Dissatisfied, 5 = Very Satisfied
6	Access to EAPs	Whether the employee has access to Employee Assistance Programs (EAPs)	Binary	Yes, No
7	Mental Health Absences	Number of days taken off due to mental health-related issues	Numeric (Integer)	0, 1, 2, …
8	Turnover Intention	Whether the employee intends to leave their role	Binary	Yes, No

2.4. Limitation

- Synthetic Data: The dataset is artificially generated for analysis and training purposes. As such, it may not fully capture the complexities, nuances, or variability of real-world healthcare workforce behavior
- Lack of Personal Context: The dataset does not take into consideration of employee’s personal mental health histories, life stressors, interpersonal relationships at work, or coping mechanisms. These limit the depth of analysis around burnout and attrition risk.

3. Patterns, Trends, and Insights

Exploratory data analysis (EDA) was conducted to understand trends, patterns, and relationships within the dataset that could inform department heads about factors contributing to staff burnout and turnover intention. Both descriptive and inferential statistical methods were used to examine insights from the dataset.

3.1. Descriptive Analysis

Numeric Variables includes Stress Level, Job Satisfaction, and Mental Health Absences were analyzed using histograms and boxplots to visualize distribution, central tendency, and detect outliers.

Findings:

- Stress Level: The majority of healthcare employees reported high stress levels, typically ranging between 8 and 9 on a 10-point scale.
- Job Satisfaction: Most staff rated their job satisfaction as 2 out of 5, indicating generally low morale across the workforce.
- Mental Health Absences: A significant portion of staff took between 5 to 7.5 days of mental health-related leave, suggesting widespread mental strain and there were also outliers observed.

Categorical Variables includes Burnout Frequency, Employee Type, Workplace Factor, Access to EAPs, and Turnover Intention were analyzed using frequency tables and bar charts to show category distributions.

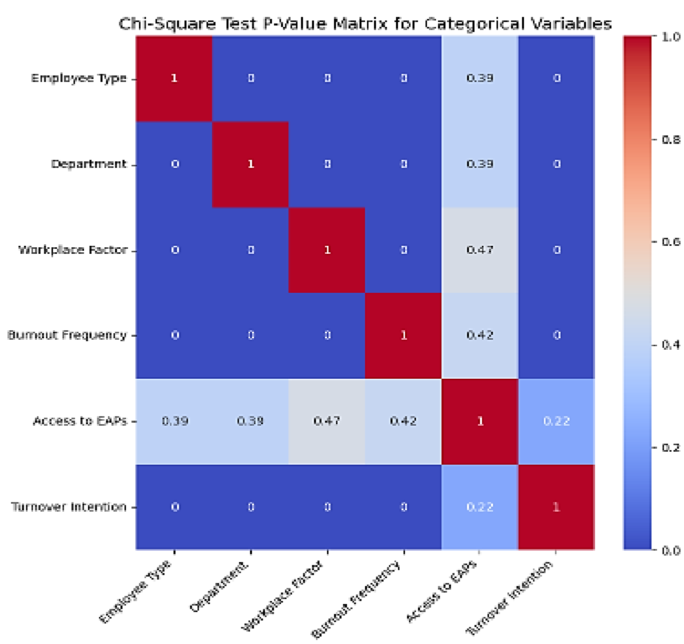
Findings:

- Employee Type & Department: These two fields appear to contain overlapping information, as each employee's type is uniquely tied to a specific department. Due to redundancy, only one of these columns was retained for analysis. Hence, department column was dropped.
- Workplace Factor: Among employees expressing a desire to leave, the most frequently cited contributing factor was "Heavy Workload", highlighting workload management as a priority area for intervention.
- Burnout Frequency: Burnout is a widespread issue across the organization as more than 83% experience occasional or frequent burnout episodes.
- Access to Employee Assistance Programs (EAPs): 72% of employees reported having access to EAPs, indicating that while support is available to a majority, 28% of employees remain without access. This presents an opportunity for expanding EAP availability to improve mental health outcomes.
- Turnover Intention: Approximately 66% of employees expressed an intention to leave their current roles. This reflects challenges in workforce retention and the needs to address burnout and job dissatisfaction.

3.2. Inferential Analysis

To statistically validate associations between variables and turnover intention, the following tests were applied:

3.2.1. Chi-Square P-Value Matrix



Chi-Square p-value matrix was generated to examine the associations between categorical variables: Employee Type, Department, Workplace Factor, Burnout Frequency, Access to EAPs and target variable: Turnover Intention and also to determine whether the distribution of turnover intention significantly differs across categories of other features.

A p-value of < 0.05 was used to identify statistically significant relationships.

It appears that categorical variables: Employee Type, Department, Workplace Factor, Burnout Frequency has statical significant relationship with target variables, Turnover Intention.

While the Chi-Square test ($p = 0.22$) did not find a statistically significant association between EAP access and turnover intention, there is a portion of employees lacking access (28%) highlight potential gap in staff support. Although this analysis does not confirm a direct relationship between EAP access and turnover, further exploration could examine whether EAPs indirectly support retention by mitigating stress or burnout.

3.2.2. One-Way ANOVA

A one-way ANOVA test was chosen to maintain consistency in method across multiple continuous variables and to support potential future comparisons involving more than two groups. It was used to determine whether there are statistically significant differences in numeric variables—Stress Level, Job Satisfaction, and Mental Health Absences—between employees who intend to leave and those who do not.

These results indicate that employees who intend to leave their roles report higher stress, lower job satisfaction, and more mental health-related absences, with all relationships being highly significant ($p < 0.001$).

Interpretation:

These statistical tests confirm that both categorical excluding Department and numeric features have meaningful relationships with turnover intention. For example:

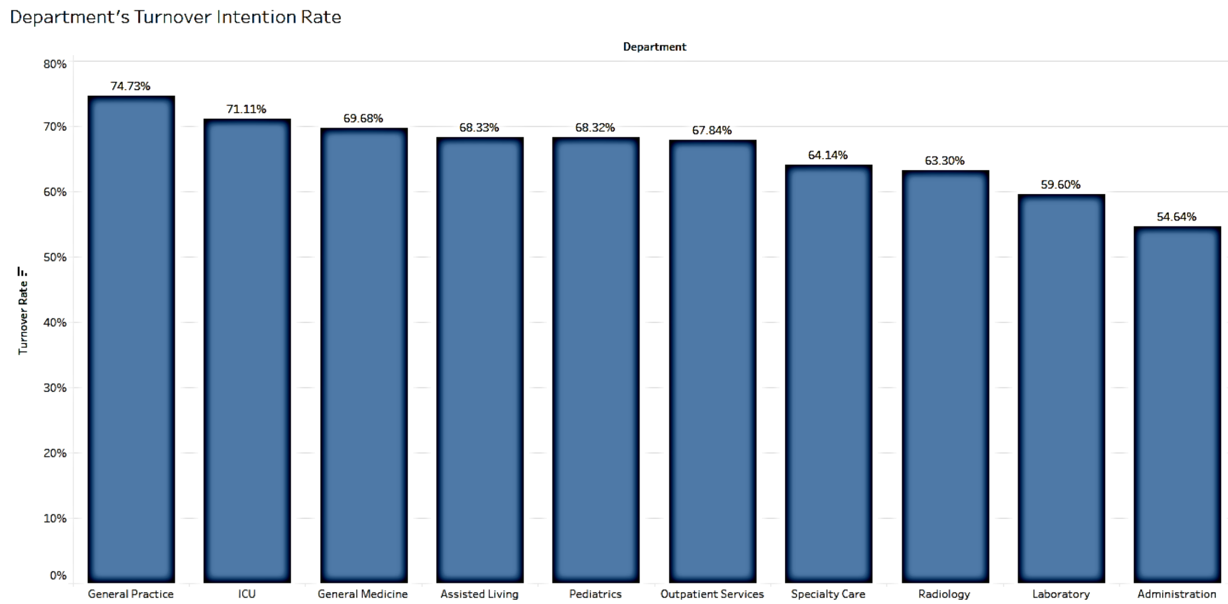
- Employees experiencing frequent burnout are significantly more likely to express intent to leave.
- Lower job satisfaction and higher stress levels are statistically linked to turnover.

Please refer to Appendix 7.1

3.3. Trend Analysis

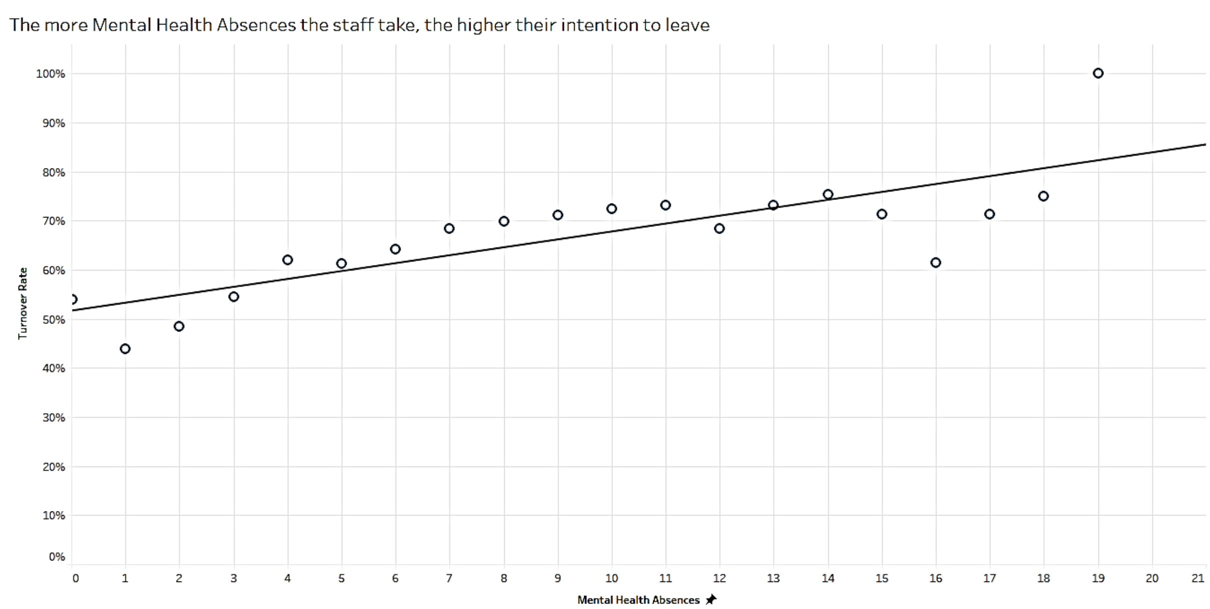
To gain deeper insights into the factors contributing to turnover intention among healthcare workers, several visualizations were developed using Tableau. These analyses explore patterns across departments, mental health-related absences, job satisfaction, stress levels, and access to support systems. Together, they provide a comprehensive view of how employees are coping within the workplace and how the organization is currently positioned in terms of supporting staff well-being and retention.

3.3.1. Turnover Intention Across Departments



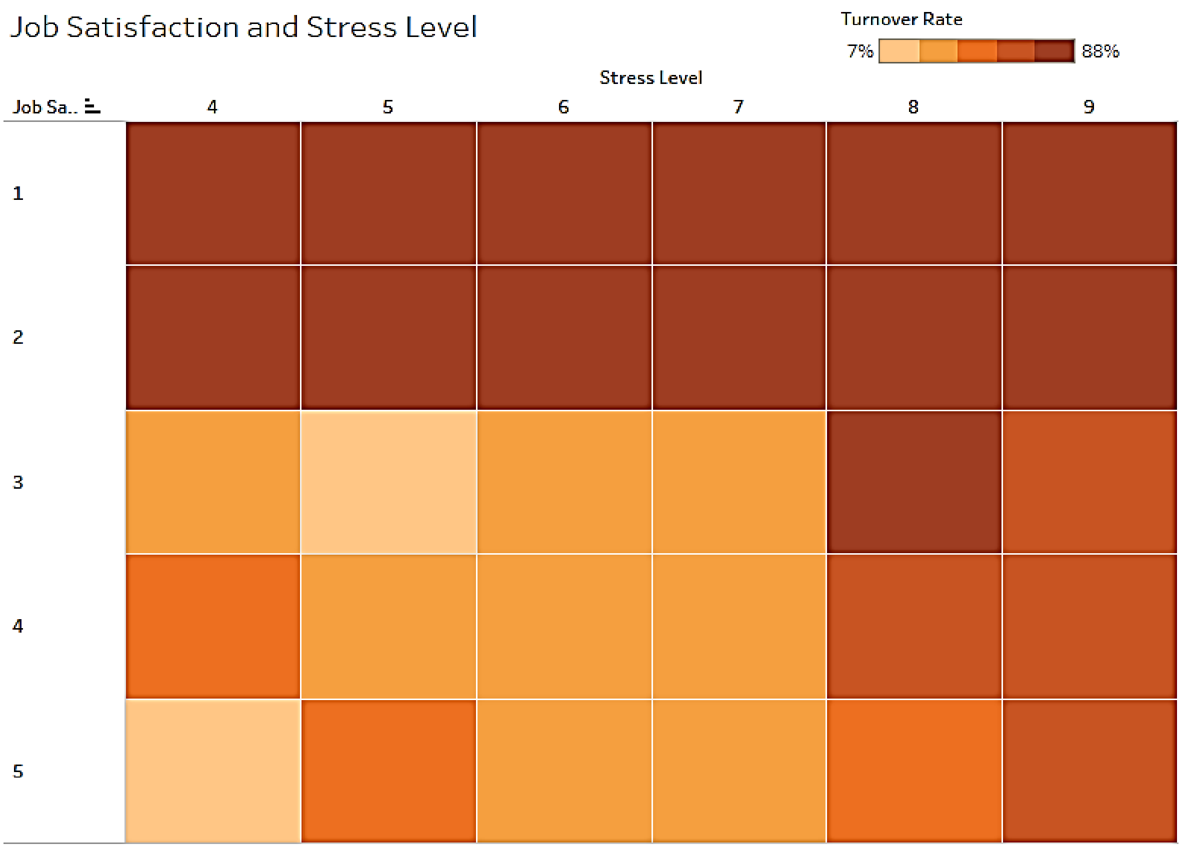
A column chart was used to display the distribution of turnover intention across different departments. This visualization highlights variability between the turnover intention across department with some departments showing apparent higher turnover intention than others. Further investigation is needed to understand the underlying causes—such as departmental work environments, exposure to workplace stressors, and differences in workload—which may be influencing staff decisions to leave.

3.3.2. Mental Health Absences vs Turnover Intention



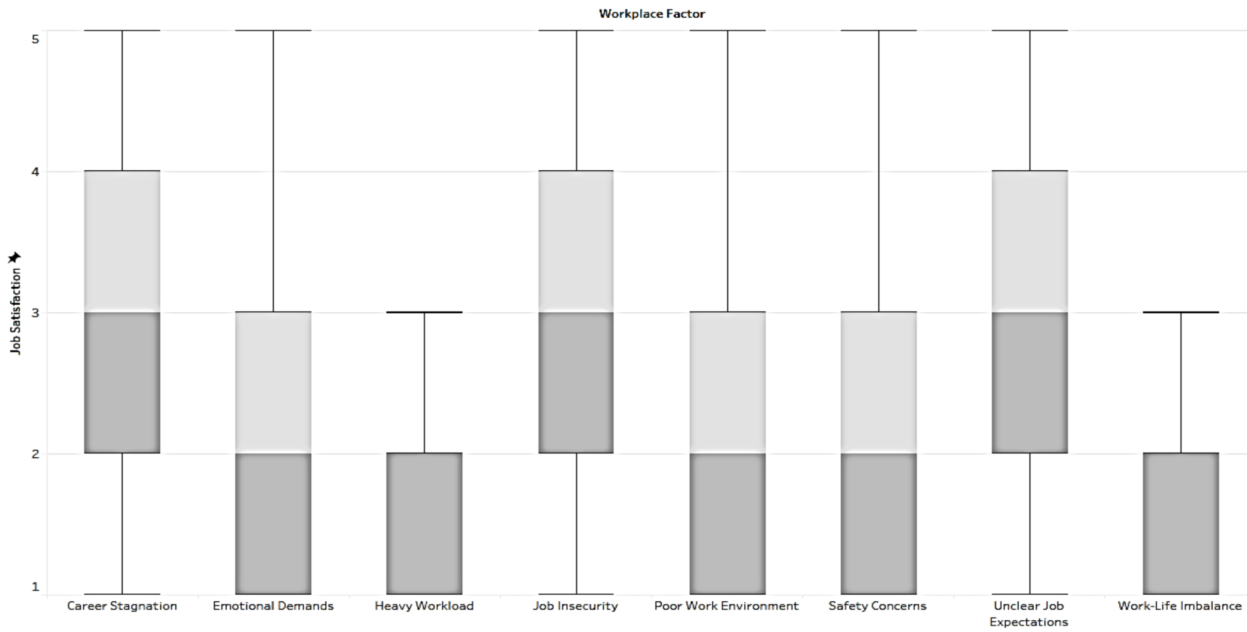
A scatter plot analyzing the relationship between mental health absences and turnover intention shows a positive correlation across the organization. Employees with more frequent mental health-related absences tend to report higher turnover intention. This finding underscores the potential impact of mental health challenges on staff attrition and highlights the need for early intervention and support systems to prevent burnout and disengagement.

3.3.3. Job Satisfaction and Stress Levels in Relation to Turnover Rate



The heatmap analysis shows that turnover rates are significantly higher when job satisfaction scores are low—particularly in the 1 to 2 range. In contrast, stress levels do not appear to have a consistent relationship with turnover rate; employees across various stress levels display similar turnover tendencies. This suggests that low job satisfaction is a more reliable predictor of turnover intention than stress alone, reinforcing the needs to foster positive and fulfilling work environment.

3.3.4. Workplace Factors and Job Satisfaction



A boxplot was used to analyze how different workplace factors relate to job satisfaction levels. The analysis covers a range of organizational stressors including:

- Career Stagnation
- Emotional Demands
- Heavy Workload
- Job Insecurity
- Poor Work Environment
- Safety Concerns
- Unclear Job Expectations
- Work-Life Imbalance

Across all factors, the median job satisfaction generally falls between 2 and 3, indicating moderate to low satisfaction. However, Heavy Workload and Work-Life Imbalance appear to be associated with lower satisfaction scores, with more concentrated distributions at the lower end of the scale - whereby more than more than 75% who indicated heavy workloads and work-life imbalance factor express job satisfaction of 2.

In contrast, factors like Career Stagnation and Unclear Job Expectations show a wider range in satisfaction levels, indicating more variability in how individuals experience these issues—possibly influenced by role, tenure, or departmental support.

Overall, the boxplot highlights key workplace stressors that negatively impact job satisfaction which can serve as actionable areas for improvement.

Interactive Dashboard

To complement the visual analyses, an interactive Tableau dashboard was developed to allow department heads to explore turnover-related trends in greater depth. The dashboard includes department-level filters and presents the following key components:

- Turnover Intention (Pie Chart): Provides a quick overview of the proportion of employees indicating intent to leave.
- Burnout Frequency (Pie Chart): Displays the distribution of burnout experiences among staff.
- Job Satisfaction vs Workplace Factors (Box Plot): Enables detailed comparisons across different workplace stressors by department, helping identify which factors correlate with lower satisfaction.

A full view of the dashboard is provided in Appendix 7.2 for reference.

4. Predictive Modeling

A predictive model was developed using key features such as stress level, job satisfaction, burnout frequency, access to EAPs, mental health absences, and workplace factors to predict whether an employee has the intention to leave. This model serves as a data-driven tool for healthcare administrators to pinpoint individuals with a high likelihood of turnover intention, enabling timely and targeted retention strategies.

4.1. Model Development

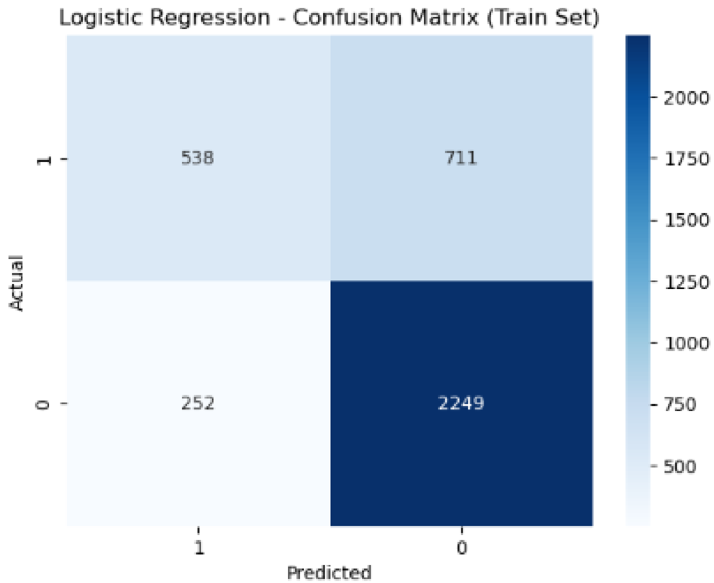
- **Algorithm Used**: Logistic Regression to predict the likelihood of burnout. It is selected for its theoretical fit with the binary classification task, its interpretability, stability across datasets, and reliability as a baseline model. While more complex models were explored, logistic regression offers a balanced trade-off between performance and explainability, which is especially important in the context of healthcare workforce analytics.

4.2. Model Evaluation

The model was trained on 70% of the data and evaluated on the remaining 30% test set. Performance metrics below are reported on the test set to reflect real-world predictive capability.

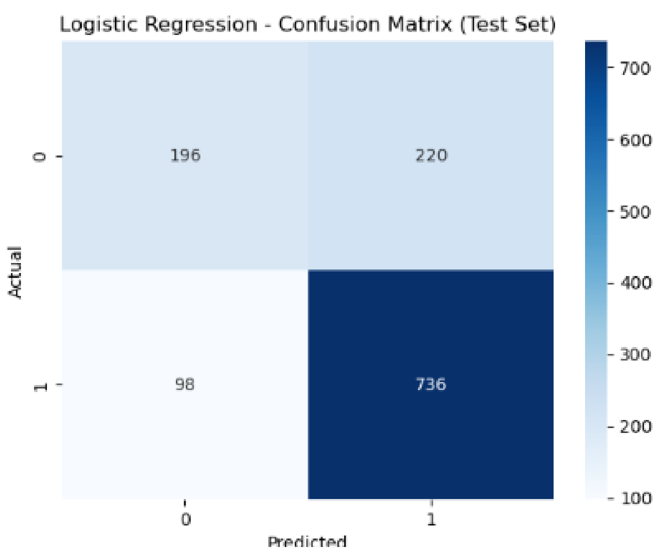
- Features Used
    - Employee Type (categorical)
    - Workplace Factor (categorical, encoded
    - Stress Level (numeric)
    - Burnout Frequency (categorical, encoded)
    - Job Satisfaction (numeric)
    - Access to EAPs (categorical)
    - Mental Health Absences (numeric)

- Confusion Metrics (Train Set):



  - Accuracy: 0.7432
  - Precision: 0.76
  - Recall (Sensitivity): 0.90

- Confusion Metrics (Test Set):



  - Accuracy: 0.7456
  - Precision: 0.77
  - Recall (Sensitivity): 0.88

The model generalizes well to unseen data, with consistent performance on the test set. It is particularly strong in identifying employees who intend to leave (as seen in the high recall), which supports the project's goal of enabling early intervention. Although the precision is moderate, in the context of employee retention, prioritizing recall is often more valuable, as identifying at-risk employees is more critical than occasionally flagging a false positive.

In addition to Logistic Regression, K-Nearest Neighbors (KNN) and Random Forest models were also explored to compare predictive performance. These models were selected for their ability to capture non-linear patterns and interactions among variables. Eventually, Logistic Regression was chosen for its interpretability and consistency between train and test dataset. Please refer to Appendix 7.3 for more details of KNN and Random Forest models.

5. Recommendations

Based on the insights gained from exploratory data analysis and predictive modeling, the following recommendations are proposed to address burnout and improve staff retention:

5.1. Prioritize High-Risk Groups Identified by the Model

- Use the predictive model to flag employees who exhibit high-risk characteristics, such as:
  - Individuals who indicate low job satisfaction in Staff Survey
  - Increased mental health absences
- They should be prioritized for check-ins, mental health resources, and managerial support.

5.2. Address Workplace Factors Contributing to Burnout

  - The most cited workplace factor among those intending to leave was 'heavy workload'.
  - Department head should consider workload redistribution, staffing adjustments, or workflow redesigns to address this challenge.

5.3. Strengthen Mental Health Support Initiatives

- Although the Chi-Square test did not show a statistically significant direct link between EAP access and turnover intention, 28% of employees lack access to support programs.
- Healthcare organizations should explore expanding access to Employee Assistance Programs (EAPs) and actively promote their use to better support staff and evaluate the impact of these interventions over time.

6. Next Steps

To build on the current analysis and strengthen workforce retention strategies, the following next steps are recommended:

6.1. Monitor and Measure Intervention Outcomes

- Use pre- and post-intervention data to evaluate the impact of Employee Assistance Programs and policy changes.
- Track changes in job satisfaction and turnover intention metrics following intervention efforts to evaluate effectiveness.
- Integrate the model into ongoing intervention and HR processes to regularly assess staff risk levels.

6.2 Collaborate with Other Stakeholders

- Work with HR and healthcare administrators to design tailored interventions for high-risk groups.
- Use data insights to advocate for policy changes aimed at improving workplace well-being.

6.3. Future Data Enhancements

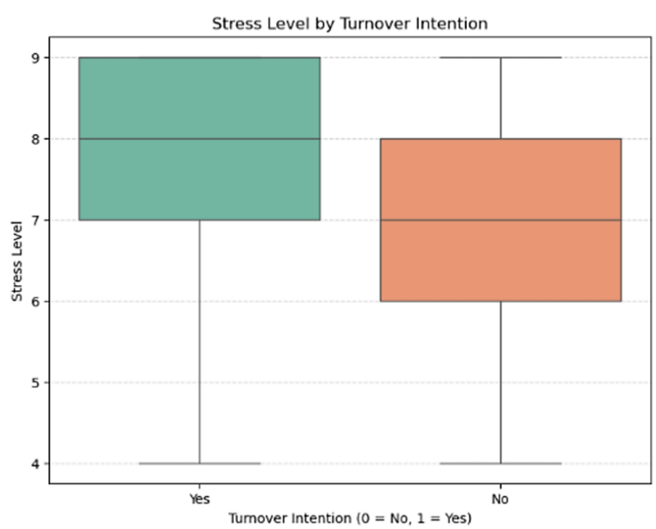
- Consider collecting additional data such as:
  - Work schedule variability - to examine its potential impact on job satisfaction, burnout frequency and stress levels
  - Managerial support ratings - to assess leadership's role in employee well-being
  - Engagement with wellness programs – to evaluate their effectiveness in reducing stress and improving retention
- These may improve model performance and allow for more refined intervention strategies.

7. Appendix

This appendix provides supplementary materials referenced in the report.

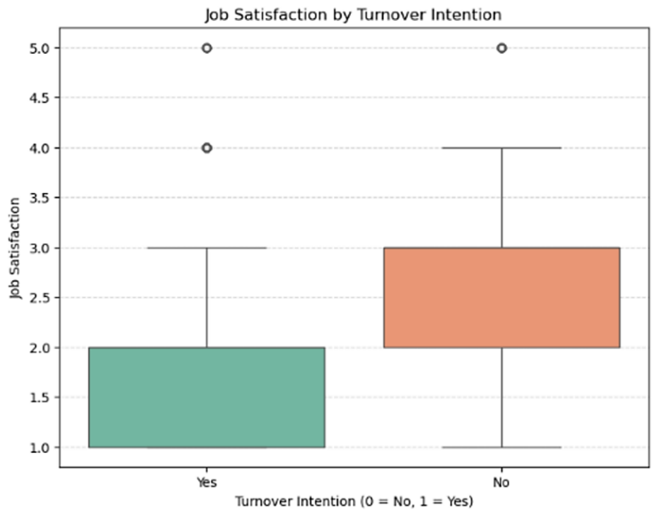
7.1 One-Way ANOVA

This section provides statistical outputs and visualizations used in Section 3.2 of the report to determine whether differences in stress level, job satisfaction, and mental health absences are statistically significant between employees with and without turnover intention.



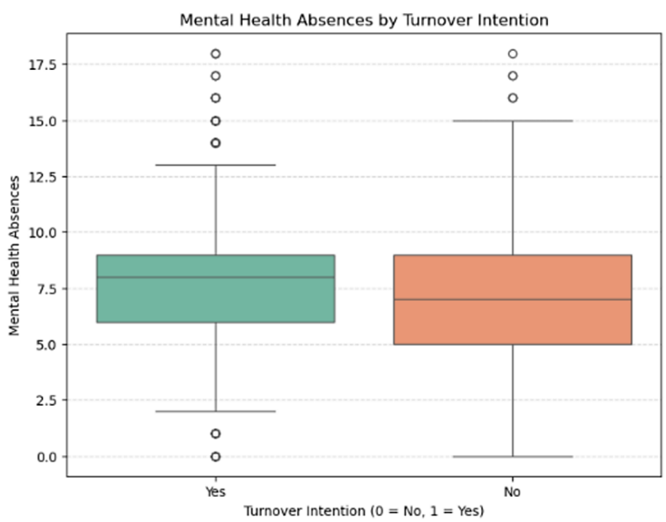

```
ANOVA F-statistic: 256.0254
P-value: 0.0
✅ The difference in stress level between the groups is statistically significant.
```

Figure A1. Box Plot & One-way ANOVA result: Job Satisfaction vs Turnover Intention



ANOVA F-statistic: 572.1792
P-value: 0.0
✅ The difference in Job Satisfaction between the groups is statistically significant.

Figure A2. Box Plot & One-way ANOVA result: Stress Level vs Turnover Intention



ANOVA F-statistic: 40.0570
P-value: 0.0
✅ The difference in Mental Health Absences between the groups is statistically significant.

Figure A3. Box Plot & One-way ANOVA result: Mental Health Absences vs Turnover Intention

7.2 Tableau Interactive Dashboard

This section presents snapshots of the interactive Tableau dashboard developed to support data exploration and stakeholder engagement.

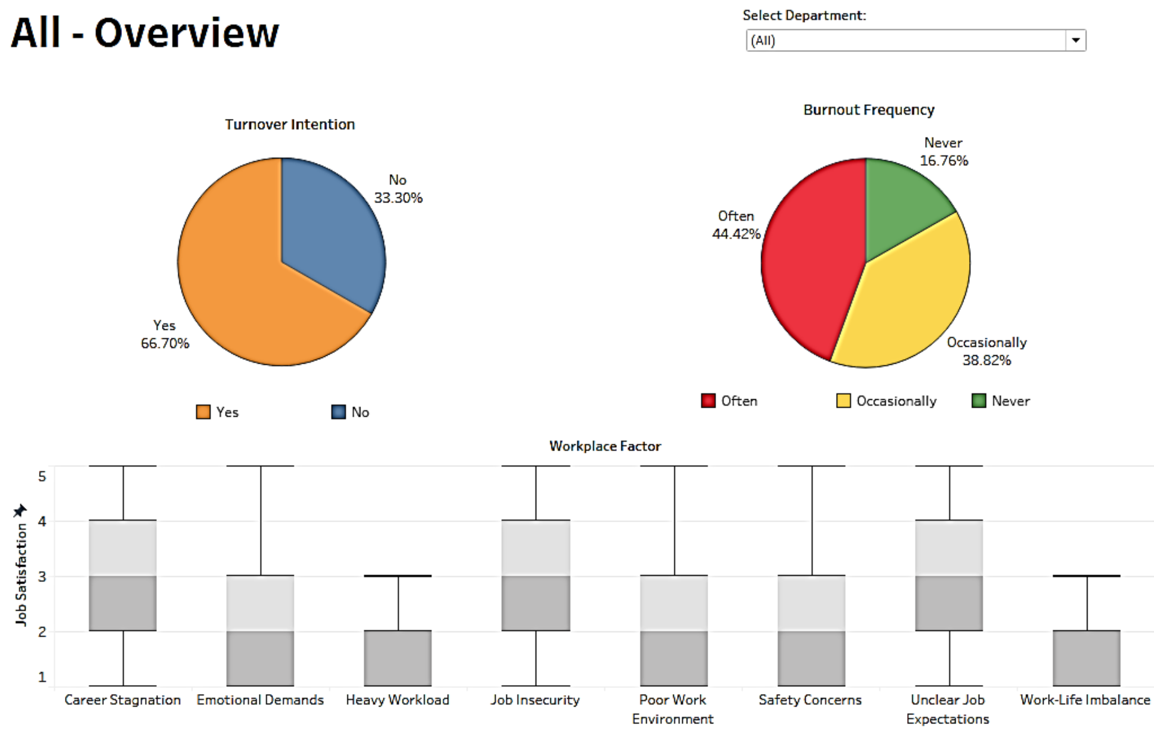
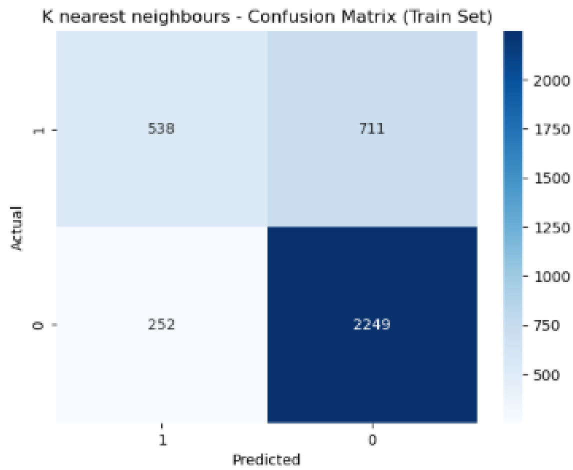


Figure A4. Tableau Interactive dashboard

7.3 Results from Other Models Explored (KNN and Random Forest)

- KNN performed moderately well, but it is sensitive to the number of neighbors (*k*) and requires proper feature scaling. Its performance varied depending on these settings, making it less stable and harder to interpret.
- Random Forest showed high accuracy on the training set but dropped significantly on the test set, indicating overfitting.

As a result, Logistic Regression was chosen for its consistent performance across both train and test sets, as well as its clarity in explaining the relationship between predictors and turnover intention. Full confusion matrices and classification reports for KNN and Random Forest are shown below.



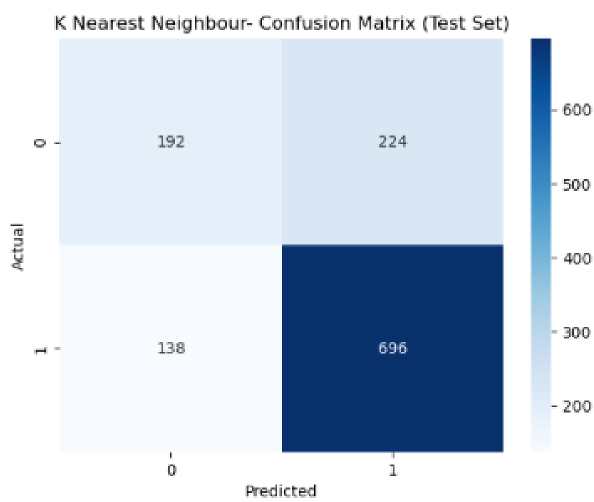

```
🔍 Accuracy on training set: 0.7824

📋 Classification Report:
              precision    recall  f1-score   support

           0       0.73      0.56      0.63      1249
           1       0.80      0.90      0.85      2501

    accuracy                           0.78      3750
   macro avg       0.76      0.73      0.74      3750
weighted avg       0.78      0.78      0.77      3750
```

Figure A5. Confusion Matrix and Classification Report – KNN (Train Set)



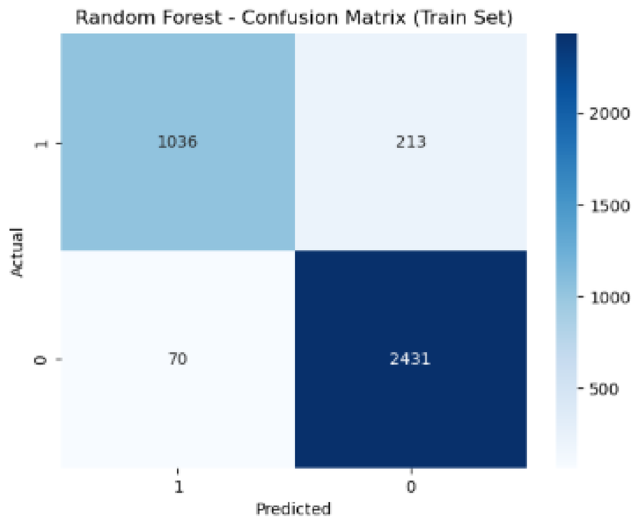

```
🔍 Accuracy on training set: 0.7104

📋 Classification Report:
              precision    recall  f1-score   support

           0       0.58      0.46      0.51       416
           1       0.76      0.83      0.79       834

    accuracy                           0.71      1250
   macro avg       0.67      0.65      0.65      1250
weighted avg       0.70      0.71      0.70      1250
```

Figure A6. Confusion Matrix and Classification Report – KNN (Test Set)

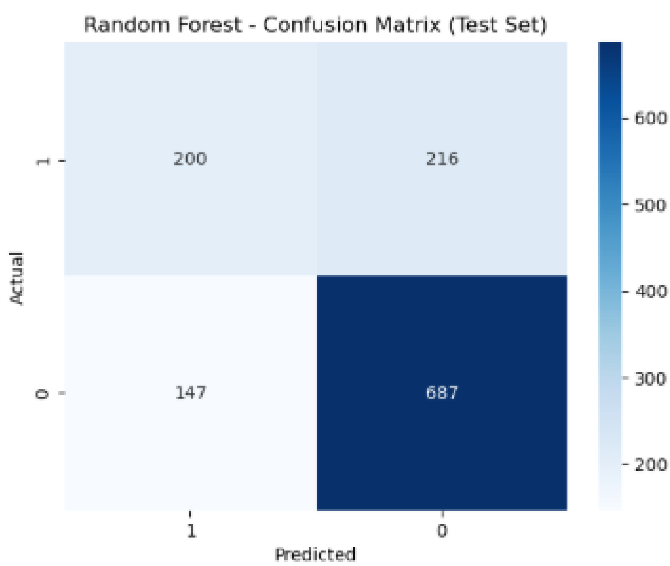



```
Classification Report:
              precision    recall  f1-score   support

           0       0.94      0.83      0.88      1249
           1       0.92      0.97      0.94      2501

    accuracy                           0.92      3750
   macro avg       0.93      0.90      0.91      3750
weighted avg       0.93      0.92      0.92      3750
```

Figure A7: Confusion Matrix and Classification Report – Random Forest (Train Set)




```
Classification Report:
              precision    recall  f1-score   support

           0       0.58      0.48      0.52       416
           1       0.76      0.82      0.79       834

    accuracy                           0.71      1250
   macro avg       0.67      0.65      0.66      1250
weighted avg       0.70      0.71      0.70      1250
```

Figure A8: Confusion Matrix and Classification Report – Random Forest (Train Set)

8. Reference

1. Baker, J. A. (2021, November 1). Resignation rates among healthcare workers in Singapore up this year; MOH to increase ICU capacity. *CNA*. https://www.channelnewsasia.com/singapore/resignation-rates-among-healthcare-workers-singapore-year-moh-increase-icu-capacity-2282766

2. Rivalytics. (2023). *Healthcare workforce mental health dataset* [Data set]. Kaggle. *https://www.kaggle.com/datasets/rivalytics/healthcare-workforce-mental-health-dataset*