EMPLOYEE ATTRITION PROJECT

Leveraging Data Analytics to Enhance Workforce Stability

Prepared by: Hailey Do

Date: 15 December 2024

Course/Organisation: University of Wollongong, Business Analytics

Hailey Do

ABSTRACT

Employee turnover presents significant challenges in the healthcare sector, particularly in maintaining organisational stability, service quality, and workforce performance. In the ever-growing age, the quantity and quality of a company's workforce are crucial for sustaining its competitive edge. This study investigates the key factors influencing employee attrition within the healthcare industry using the IBM employee attrition dataset. In the analysis part, a logistic regression model will be applied to identify significant variables as well as demographic variables in the company. Findings indicate that older employees with fewer business travel commitments are less likely to leave, while younger staff and those with extensive travel obligations show higher turnover rates. Therefore, these insights emphasise the importance of targeted retention strategies, such as improved work-life balance, development programs, and flexible work hours, to retain valuable healthcare individuals. This study offers recommendations for healthcare administrators to mitigate employee attrition rates and enhance retention initiatives, ultimately improving service quality and workforce stability.

TABLE OF CONTENTS

INTRODUCTION

The healthcare industry is one of the most essential sectors globally, providing critical services for human well beings and public health. However, it comes along with the intensive labour workforce with relying heavily on skilled professional or skilled individuals to maintain the prospects and effective quality care. Maintaining a core workforce is essential to an organisation's long-term viability and operational success (Das & Baruah 2013). According to Certified Ham Resource Management Professional (CHRMP), there are five main types of attrition: voluntary, demographic-specific, internal, involuntary and retirement-induced. Employee attrition leaves their company voluntarily which is the most common type one with the under control throughout any companies (CHRMP 2022). Some reasonings behind this problem are taking the better offer or to make a change for their careers, indicating significant and ongoing challenge in healthcare industry specifically and for other industries in generally. In fact, attrition rates far exceed the ideal benchmark of 10% for a healthy organisation (Lindquist 2023), and this problem has been exacerbated by the COVID-19 pandemic. In 2022, turnover rates increased dramatically, from 19.5% in hospitals to 94% in nursing homes, driven by factors such as burnout, high workloads, and inadequate work-life balance (Lindquist 2023). Additionally, the stress associated with workplace environment, data accessibility, and the emotional toll of caregiving can contribute to the increasing of the attrition rates. These trends in attrition rates pose both financial and operational challenges to healthcare providers, as organisations must constantly recruit, train, and replace new staffs.

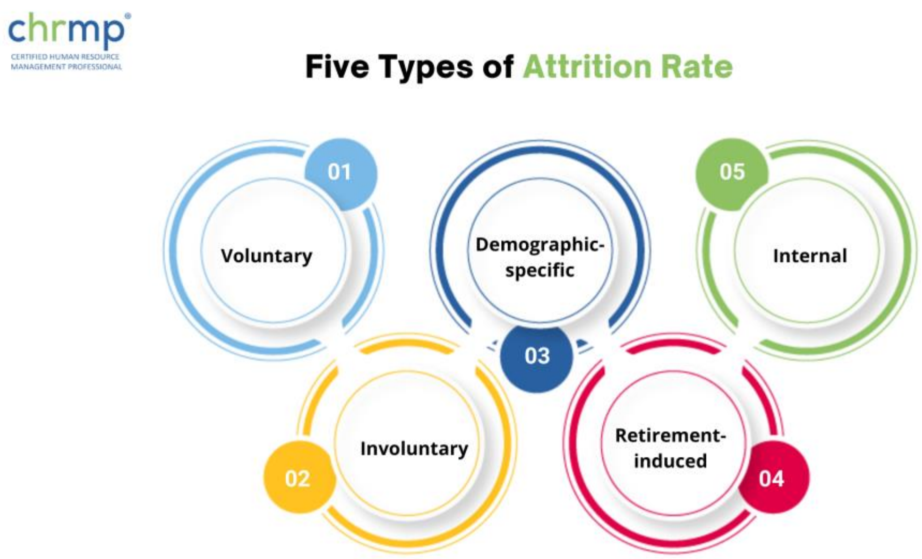


Figure 1. Five types of Attrition Rate (CHRMP)

Hailey Do

PROBLEM IDENTIFICATION

1. Problem statement

The rapidly expanding healthcare industry faces a critical challenge in managing employee turnover. High attrition rates, driven by competitive job markets and more attractive pay packages, cause newly hired healthcare employees to frequently leave within their first few years, indicating potential internal issues within workplace environments or workloads. This trend places the sector at risk, as highly skilled professionals are often drawn to opportunities in other places where they get better job with higher bases. The resulting staff shortages disrupt operations, compromise patient care, and lead to substantial financial losses such as planning, recruiting, training and contingent labour costs (Chervoni-Knapp 2022). Any disruption within healthcare structure can lead to unexpected consequences, including delays in patient care, team-based working condition, and an increased workload for remaining team members. Therefore, it will lead to the frequent change in the job description among staffs causing unexpected dilemmas. According to Oracle, the U.S. healthcare industry faced a costs equivalent to six to nine months of an employee's salary for new replacement, especially with cost up to 200% due to extensive onboarding and training programs. The level with approximately 500,000 workers leaving the industry every month in 2022 is an emergency for the hospital and healthcare providers' structure (Lindquist 2023). The United Kingdom have reported similar challenges, where the National Health Service (NHS) experienced record attrition during the pandemic, driven by increasing workloads and the "Great Resignation" (Poon et al. 2022). These number highlights the urgent stress for the healthcare sector, indirect impacts the customer satisfaction as understaffed facilities.

The COVID-19 pandemic introduced unprecedented challenges to the healthcare industry, causing significant disruptions and escalating attrition rates. The surge in patient numbers strained resources and led to severe staffing shortages, as many healthcare workers struggled to keep up with the overwhelming demand. Hospitals and healthcare facilities worldwide faced an unparalleled wave of resignations, as high infection rates and intense workloads took a toll on employees' mental and physical well-being (Poon et al. 2022). Combining with the inflexibility of healthcare schedules and the burden of excessive administrative tasks further push workers toward resignation in their first few years of working. The McKinsey hospital during the pandemic witnessed the increasing number of nursing in attrition rates by 4% to 5% just within a year, indicating the fragile in the workforce sustainability issue. This challenge forced many healthcare providers to alter their care models, reduce the patient inputs and reallocate to another divisions, significantly affecting operations. This situation intensified the need for robust support systems within healthcare, as the pandemic exposed vulnerabilities in employee retention strategies and workforce resilience. Recovery from this period has been slow, with many healthcare workers reconsidering their career paths due to the daunting conditions they faced and the impact of their mental health problems in the long-term.

Additionally, the growing demand for healthcare services, coupled with ineffective retention strategies, further exacerbates the turnover problem. Some potential factors such as inadequate work-life balance, short tenure, and insufficient career development contribute to employee dissatisfaction (Boatman 2021). Within the healthcare industry, there are notable disparities in attrition rates across different departments. Research shows that nurses working in areas such as medical/surgical services or Emergency Department (ED) experience some of the highest turnover rates, while those in children department and women's health report significantly lower rates. According to the 2023 NSI National Health Care Retention & RN Staffing Report, the registered nurses (RNs) in medical or surgical units have a rate as high as 22.5%, EDs consistently report at 27.1%, which are much higher compared to other departments where rates are lower at 18.4%. The disparities highlight the departments which is critical to be examined and deliver the better targeted retention initiatives such offering mental health support, flexible scheduling or well-trained sessions for new employees.

2. Scope of the study

This study aims to comprehensively explore the issue of employee attrition in the healthcare sector, focusing on identifying key factors such as job satisfaction, compensation, and tenure that influence the attrition rates (Appendix 1). The research will include the use of analysing the problem in terms of qualitative and quantitative methods, including logistic regression analysis, to identify key factors contributing to the likelihood of employee decisions to leave their companies. This analysis will consider the include demographic variables such as age, gender, and years of experience as the base to investigate the employee's background. Although the industry is healthcare focused, the findings are expected to have broader relevance across industries facing similar retention challenges, helping to deliver optimal solutions when it comes to the attrition rates.

The study will employ descriptive analytics to identify trends and typical patterns within the dataset, focusing on critical variables such as department, job satisfaction, tenure and compensation levels to highlight the areas and typical type of employees tend to leave the company. Following this, predictive analytics will be applied logistic regression model to examine and forecast which employees are most at risk of leaving based on factors such as job satisfaction, salary and demographics. Lastly, prescriptive analytics will be used for proposing optimised retention strategies, such as salary adjustments and leadership training to reduce attrition rate, ensuring the optimal allocation of resources across departments.

A key limitation of the study is the generalisability of the findings across different healthcare systems and regions. Moreover, the study focuses on a single dataset that may not fully capture the nuances of local healthcare environments or smaller organisations. Additionally, there is none model precisely capture all the variables affecting the attrition rate such as an organisational cultural or personal factor may be difficult to quantify. Despite these limitations, the study will offer valuable insights into the relationship between potential drivers and provide optimal retention strategies along with these insights.

3. Literature Review:

Research on employee turnover in healthcare has consistently identified critical factors driving high attrition rates, including job satisfaction, burnout, work-life balance, and organisational support as key elements affecting this rate. These elements are significant contributors to turnover, with job dissatisfaction often serving as the trigger for the likelihood of employees leaving their roles. As found in studies by Lartey, Cummings, & Profetto-McGrath (2014), some potential factors can significantly improve retention rates such as leadership quality, professional development opportunities, and recognition of employees' efforts. Those are recognised as effective strategies to mitigate attrition rate, especially for the healthcare sector where the patient quality care is a priority. This study shares similar findings, focusing on a comprehensive strategy that accounts for these potential factors to achieve lower attrition rates of the dataset (from over 16% to around 10%).

Additionally, studies show significant disparities across different healthcare departments, with high-stress environments such as Emergency Departments (EDs) experiencing higher attrition rates than others. According to the NSI National Health Care Retention & RN Staffing Report (2023), these departments struggle with increased staff turnover due to the physically and emotionally demanding nature of the work. The emotional and attitudinal aspects of job dissatisfaction significantly influence an individual's decision to leave. The stronger an employee's frustration or disappointed with their role, the more likely they are considering taking alternative employment opportunities. The mental health and the physical health need to be prioritised to address this disparity with new mentoring programs, flexible work arrangements, and job redesigns. Overall, this highlights an urgent need for improved support systems within the healthcare workforce to prevent future crises and support retention.

ANALYSIS AND ARGUMENTATION

1. DESCRIPTIVE ANALYSIS

1.1. DATASET OVERVIEW:

The IBM HR Employee Attrition was used in this analysis contains 1,470 observations and 35 features, capturing a range of demographic, job-related, and satisfaction-based attributes for employees in a company. Key variables include categorical features such as *Department, JobRole, BusinessTravel, and EducationField*, along with numeric features like *Age, DailyRate, DistanceFromHome*, and *MonthlyIncome*. This diverse set of attributes allows for a comprehensive analysis of potential factors influencing employee attrition. The dataset is clean, with no missing or duplicate values, providing a robust foundation for further analysis. The information about the features is examined in Table 1.

Table 1. Dataset Overview

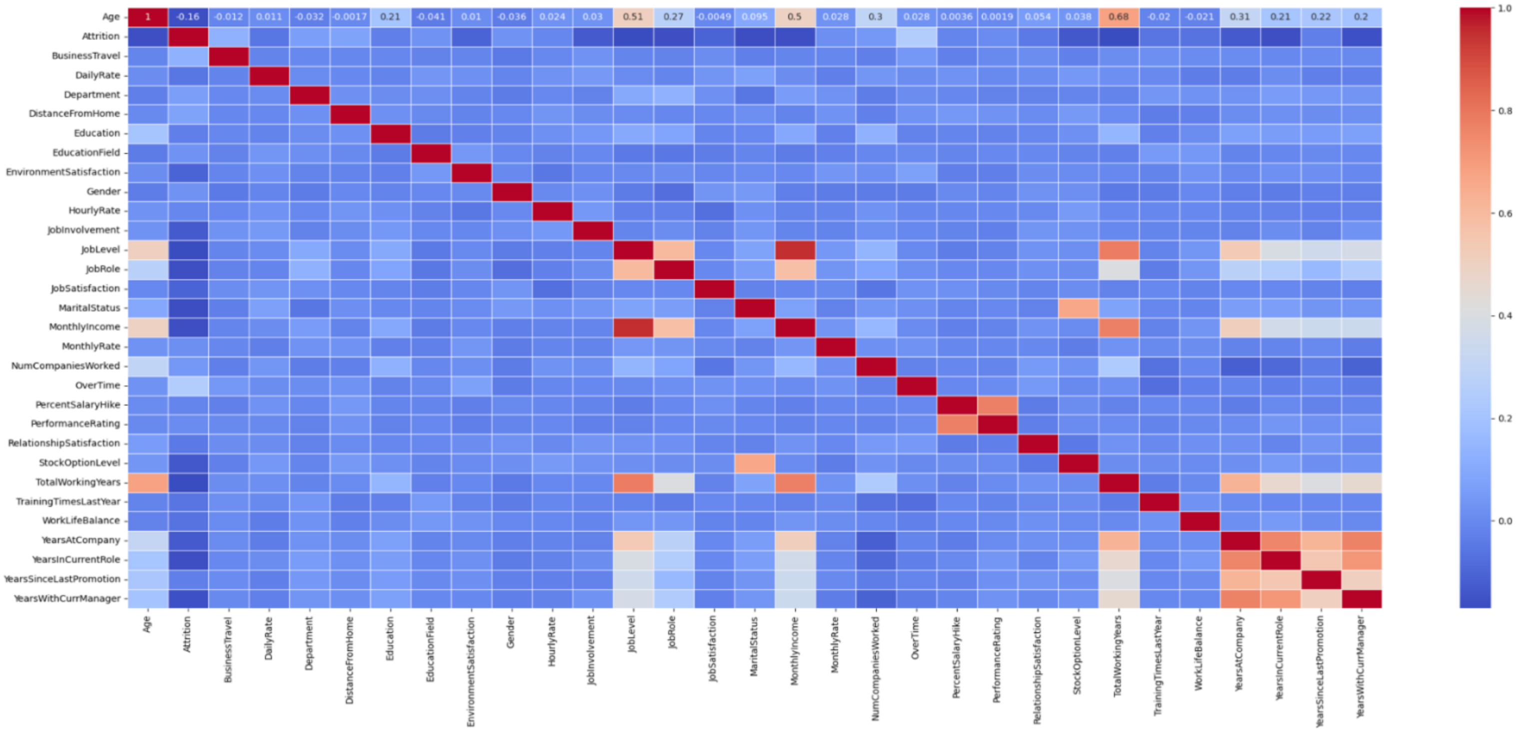
Column	Non-Null Count	Data Type	Column	Non-Null Count	Data Type
Age	1470	int64	MonthlyIncome	1470	int64
Attrition	1470	object	MonthlyRate	1470	int64
BusinessTravel	1470	object	NumCompaniesWorked	1470	int64
DailyRate	1470	int64	Over18	1470	object
Department	1470	object	OverTime	1470	object
DistanceFromHome	1470	int64	PercentSalaryHike	1470	int64
Education	1470	int64	PerformanceRating	1470	int64
EducationField	1470	object	RelationshipSatisfaction	1470	int64
EmployeeCount	1470	int64	StandardHours	1470	int64
EmployeeNumber	1470	int64	StockOptionLevel	1470	int64
EnvironmentSatisfaction	1470	int64	TotalWorkingYears	1470	int64
Gender	1470	object	TrainingTimesLastYear	1470	int64
HourlyRate	1470	int64	WorkLifeBalance	1470	int64
JobInvolvement	1470	int64	YearsAtCompany	1470	int64
JobLevel	1470	int64	YearsInCurrentRole	1470	int64
JobRole	1470	object	YearsSinceLastPromotion	1470	int64
JobSatisfaction	1470	int64	YearsWithCurrManager	1470	int64
MaritalStatus	1470	object			

Given the diversity of variables, this dataset is well-suited for examining correlations and group-wise trends to identify patterns essential for the analysis of an attrition rate, proving insights into the appropriate retention strategies.

1.2. CORRELATION MATRIX

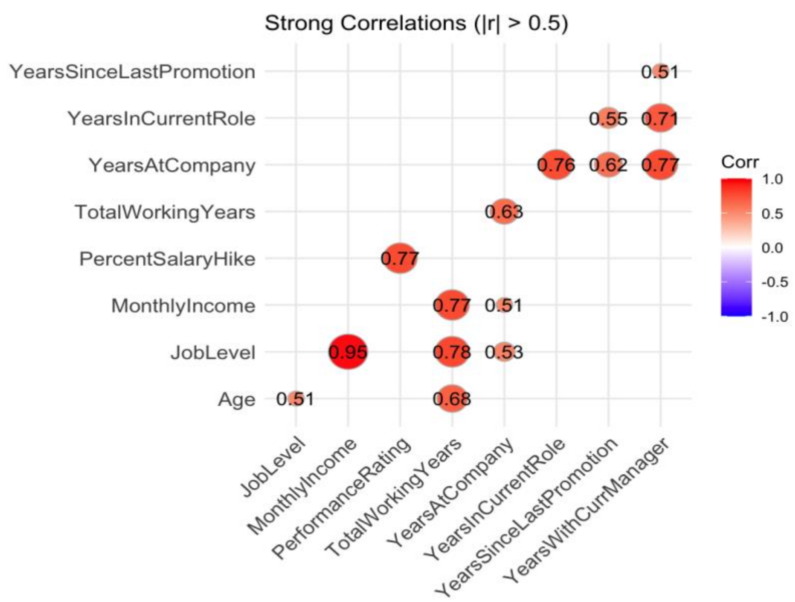
The implementation of this analysis is conducted using RStudio and Python for some visualisations, with an objective of demonstrating the patterns of who stay or left their jobs.

Figure 2. Correlation matrix



The overview of a correlation matrix above provides valuable insights into relationships between various explanatory factors in the dataset, focusing on attributes potentially influencing employee attrition. The strongest of correlations is spanning across the working years of the employees within the last 4 columns, indicating a potential multicollinearity problem between x variables.

Figure 3. Strong correlated variables

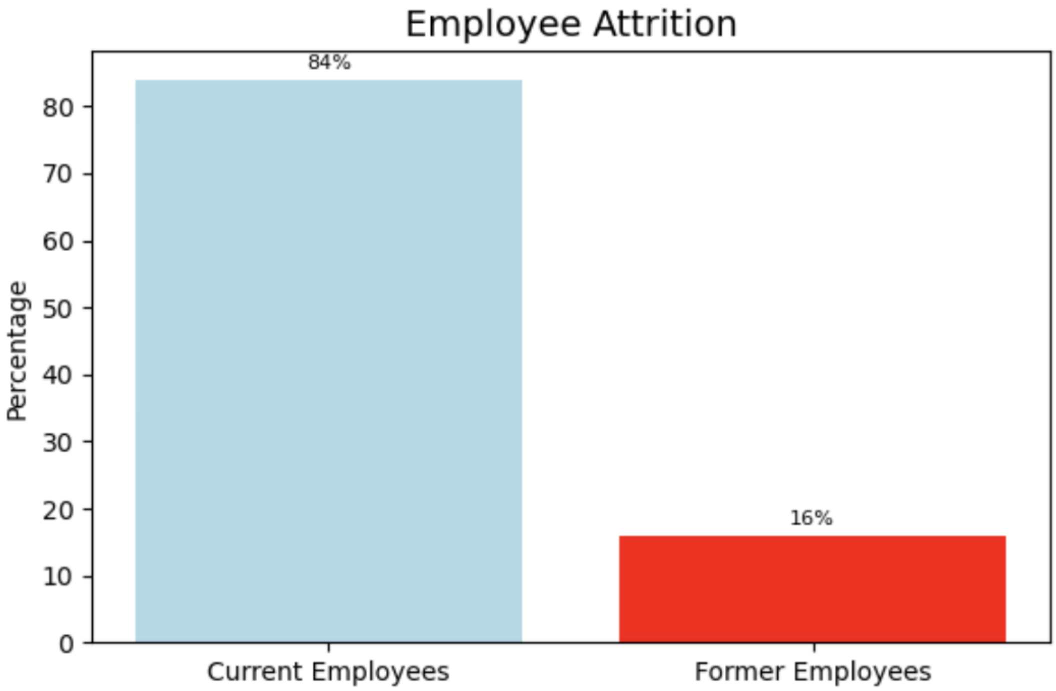


The significant correlations above highlight the strong linear relationship with absolute value of r > 0.5 among potential variables within the dataset. Notably, *JobLevel* and *MonthlyIncome* show the highest positive correlation (r = 0.95), suggesting that the common sense of employees with higher job levels tend to have higher monthly incomes. Other noteworthy correlations include *TotalWorkingYears* with *YearsAtCompany* (r = 0.76), and *YearsWithCurrManager* with *YearsSinceLastPromotion* (r = 0.77), indicating that longer tenures often coincide with more stable relationships with managers and fewer recent promotions. Overall, these correlations could imply a hidden multicollinearity problem within explanatory variables, it is worthy for dropping those variables to get more precise conclusions in the predictive analysis.

1.3. COMPREHENSIVE ANALYSIS:

The distribution of employee attrition within the organisation, revealing that majority (84%) are current employees, while 16% have left. This attrition rate reflects the organisation’s turnover, indicating potential areas for further investigation into the causes of employee separation.

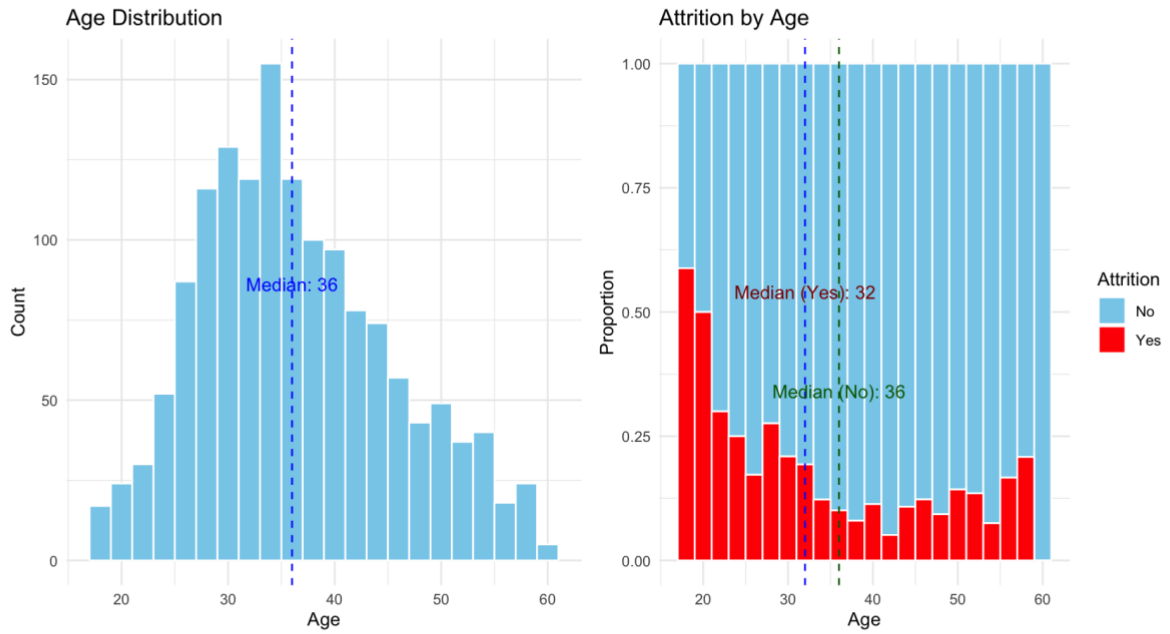
Figure 4. Overview of Attrition Rates



The attrition pattern does not take up much portion, but it is still an alertly high number suggesting for ultimately improvements in workforce management practices to retain employees, especially who work in the essential roles within a healthcare industry.

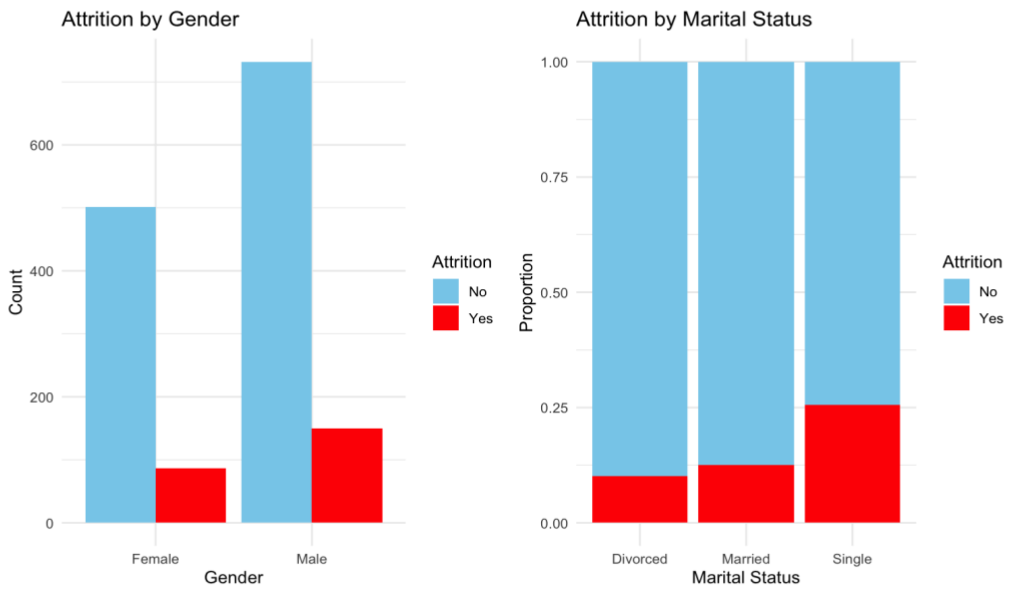
Demographic Analysis

Figure 5. Attrition by Age



The age distribution of employees shows a median age of 36 years, indicating a relatively young workforce. The histogram illustrates a higher concentration of employees between 30 and 40 years, with a noticeable decline when age increases beyond 50 years. The attrition analysis by age reveals that employees who left (indicated in red) tend to have a younger median age of 32 years compared to those who stayed (median age of 36). It implies that people will relatively have a higher likelihood of leaving at their 20s and more stable when reaching their 40s. Therefore, it is a need for launching retention program to address factors influencing early-career attrition.

Figure 6. Attrition by Gender and Marital Status



The attrition analysis by gender shows similar rates across male and female employees, suggesting that gender does not play a significant role in attrition. However, the gender distribution shows a notable male presence across various healthcare roles, challenging the traditional female dominance in the industry. This might reflect the evolving landscape where more men are entering nursing roles while also representing physician and administrative positions. Additionally, an attrition rate by marital status reveals a higher likelihood of single employees leaving the organisation compared to their married or divorced co-workers, likely due to the challenges of managing irregular shifts and high-stress environment without family support systems. This understanding is pivotal for healthcare organisation to develop targeted retention strategies and maintain stable patient care teams.

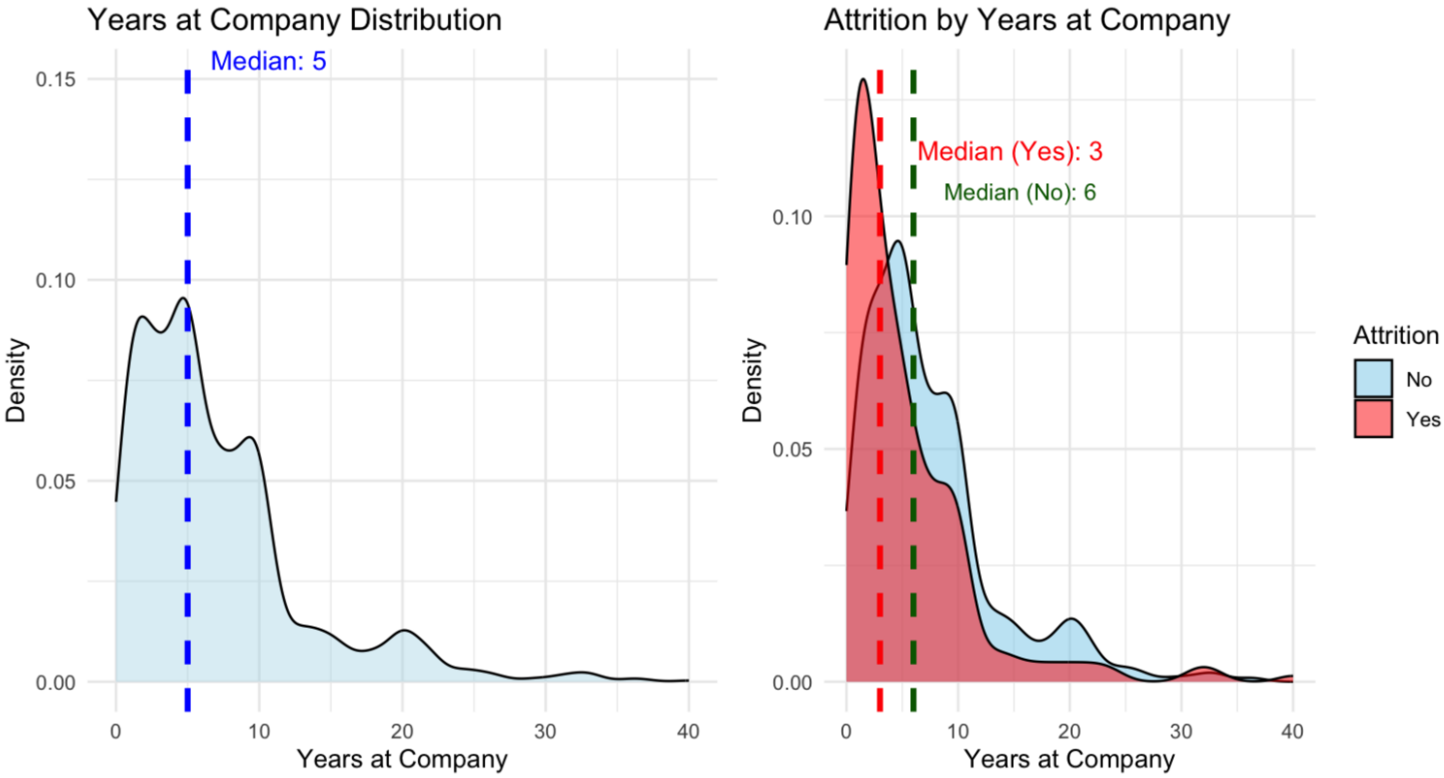
Compensation Analysis

Figure 7. Monthly Income and Attrition by Job Level

The scatter plot illustrates the relationship between *MonthlyIncome* and *JobLevel* with a focus on *Attrition* status. Employees at **Job Level 1** have low monthly incomes and represent a high density of attrition cases (red points). This pattern reflects a common sense of a low job level associated with low compensation tend to leave the company. In contrast, as the job levels increase, there is a noticeable rise in monthly income, and attrition rates become less frequent, although some cases exit in the higher income levels. Especially, in the healthcare providers, a lower attrition lower levels reflects healthcare's unique challenges: demanding patient care, strict protocols, 24/7 operations. Higher positions require more extensive medical knowledge and meticulousness throughout the work for patient outcomes, associated with better compensation and resulting in better retention.

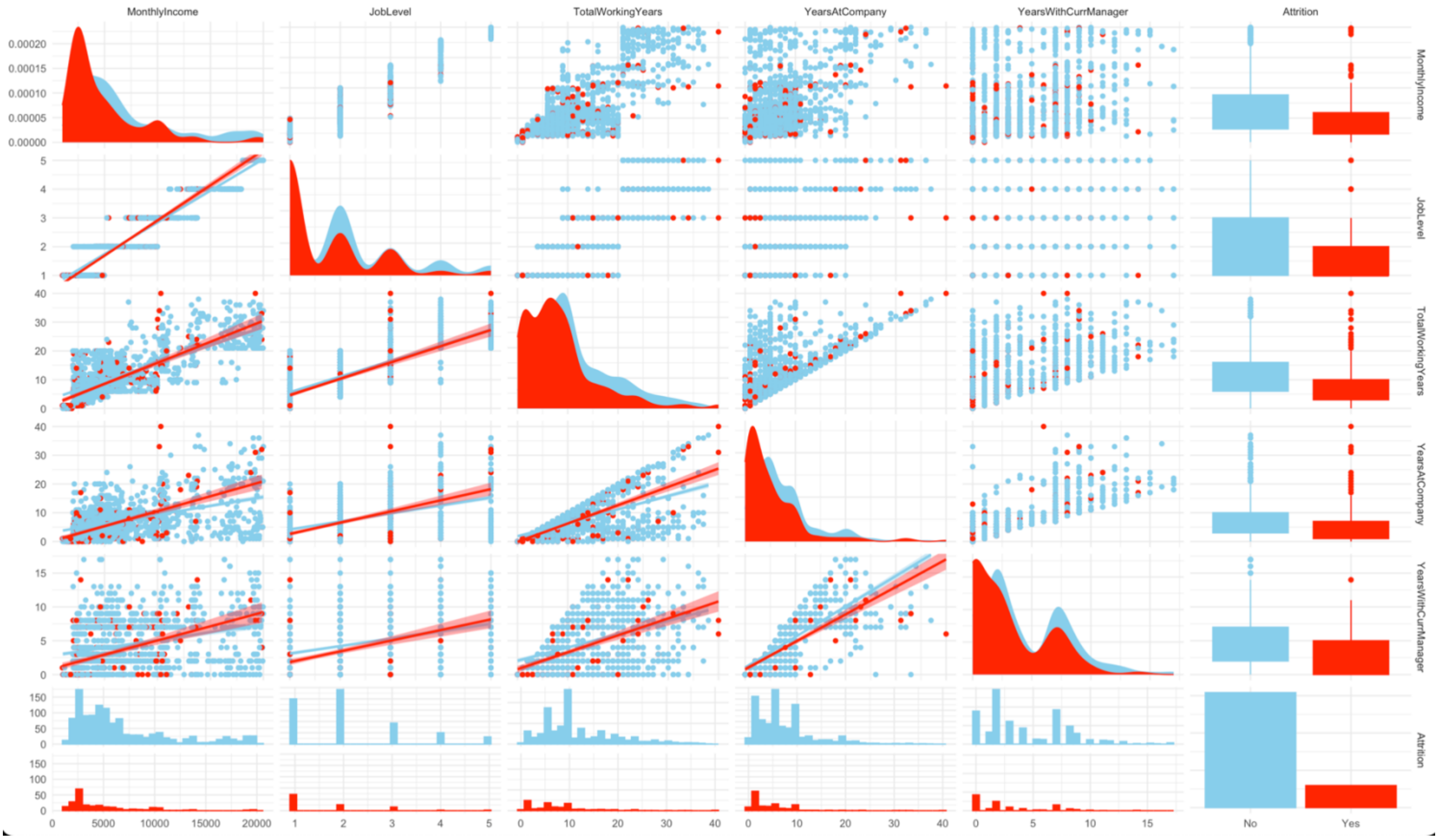
Tenure Analysis

Figure 8. Distribution of Years at Company and Attrition by Years at Company



The median tenure at the company is 5 years while the median for those who leave is only 3 years, indicating that there is a tendency for employees leaving within their early years at the organisation. This attrition pattern suggests underlying issues, possibly due to the unmet job expectations, the tedious and demanding nature of tasks, or lower growth opportunities. In the healthcare context, there is must to have a team-based cohesion and team dynamic of providing comprehensive quality. The high stress nature and unique demand for the specialised knowledge, especially who in the nursing roles or support roles to ensure the smooth of the operational procedures. An attrition rate is rising coming along with potential challenges for the labour force of a company structure, remaining staffs must handle with unaddressed tasks and take over more tasks. The domino effect will easily create with internal dilemma when the process of hiring new staff requires efforts, times and costs while those could be used for better operation decisions.

Figure 9. Correlation Matrix of Employee's Tenure with *MonthlyIncome* and *JobLevel*



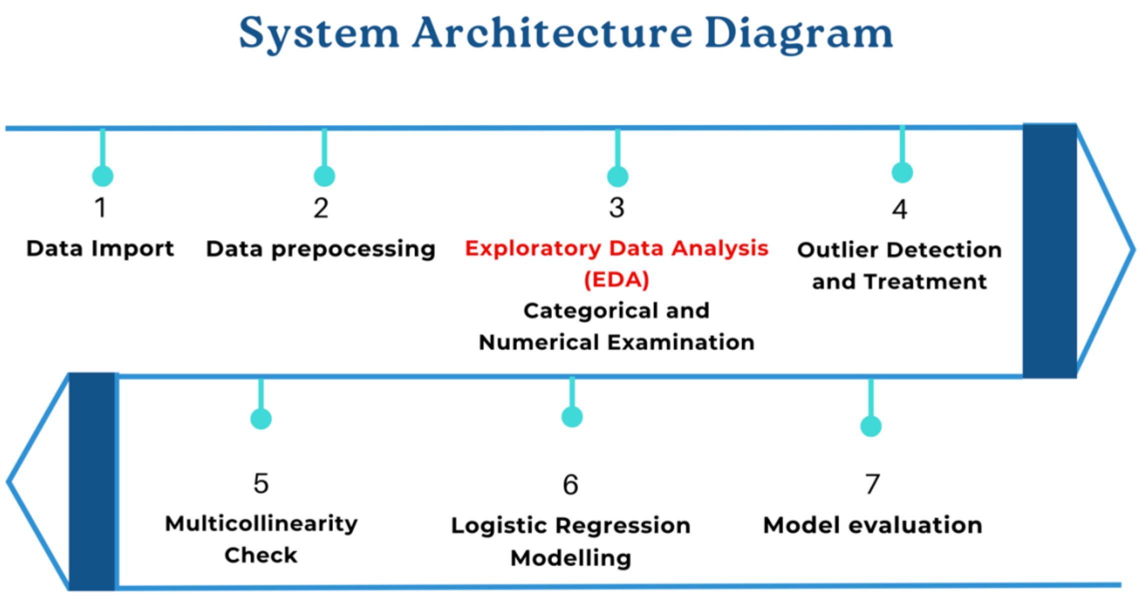
This pair plot is conducted based on the correlation matrix when the working year of employees having a strong relationship with each other. *MonthlyIncome* and *JobLevel* appear positively correlated, indicated that as employees progress in job level, their income increases, which is typical in structured career progression. However, the attrition pattern is shown clearly within the short tenures and younger employees with low compensation are likely to leave. In the context of healthcare, where continuity and expertise are critical for patient outcomes, such turnover can disrupt team stability and increase operational pressures on the remaining workforce. This finding empathises the need for launching the programs in the early employment which should be targeted for specific employees.

Overall, the analysis highlights that younger, early-career employees in lower-paying roles are more likely to leave, particularly in high-stress healthcare positions. Attrition is less frequent among higher-level, better-compensated roles, indicating a link between job satisfaction and compensation. Addressing this challenge may involve mentorship programs, early-career development initiatives, and compensation adjustments, to foster engagement and commitment among newer employees. This approach could mitigate early attrition, particularly for essential healthcare roles.

2. PREDICTIVE ANALYSIS

In this section, the research methodology and implementation approach are described comprehensively. It is noted that all programming tasks were carried out using Python along with descriptive statistics of variables. The main objective of this analysis is to use logistic regression to predict employee attrition based on various factors.

Figure 10. System Diagram



2.1. Data import

The data is ready for data cleaning and Exploratory Data Analysis (EDA) steps. The raw dataset containing 1,470 observations with varies categorical and numerical variables.

2.2 Data preprocessing

First, the data is expected as a dirty data and having missing values. The examination towards the data is necessary to enhance the data quality. Since the observations throughout the dataset have no missing values or duplicated values. Second, data preprocessing entailed separating the data into categorical and numerical subsets, preparing the dataset for model development.

The dataset was split into **categorical** and **numerical variables**. This allowed for targeted processing, such as encoding into binary variables for categorical variables and scaling for numerical data, essential for model consistency. Categorical variables are *Department, BusinessTravel*, *MaritalStatus, Education Field*, etc.., were converted into dummy variables (achieving 1 subset as a reference variable for others). This transformation was necessary to enable the logistic regression model to interpret the categorical variables.

The numerical variables are standardisation because of different ranges within all the numerical variables, for example, *MonthlyIncome* variable range is different from the range of the rating in terms of the company or workplace environment.

2.3. Exploratory Data Analysis (EDA)

Categorical variables

Table 2. The statistics of categorical variables

Column	Count	Unique	Top	Freq
Attrition	1470	2	No	1233
BusinessTravel	1470	3	Travel_Rarely	1043
Department	1470	3	Research & Development	961
EducationField	1470	6	Life Sciences	606
Gender	1470	2	Male	882
JobRole	1470	9	Sales Executive	326
MaritalStatus	1470	3	Married	673
Over18	1470	1	Y	1470
Overtime	1470	2	No	1054

Each categorical variables were examined to understand its distribution and potential association with the target variable which is Attrition (using the correlation matrix). For instance, Gender variable is defined as an insignificant variable with the less correlated with the Attrition status, considered to drop before implementing the logistic regression model. The contingency table for the *JobRole* variable is conducted to examine the attrition rate within each role, further highlighting that 62 laboratory technicians are likely to leave. This trend aligns with the broader issues with the medical laboratory, where there is an ongoing shortage of qualified professionals. According to a report by Forbes, there is a growing concern about the critical shortage of medical laboratory professionals in the workforce. This shortage is attributed primarily due to burnout, high workloads, and increased demand (Stone 2022). The problem is exacerbated in the COVID-19 period when the patient care and the research for the vaccination is advanced in a short time, causing the stress for all the laboratory employees and leading to the voluntarily attrition. The high turnover rate is not only disruptive to the hospital setting but also costly, due to the recruitment and training of new staff require considerable resources. Moreover, replacing laboratory analysts in the healthcare industry incurs significant costs, impacting not only finances but also operational efficiency (Lester-George 2023). The training process typical takes 3-6 months before new staff can sufficiently on board with the team and work independently and at full capacity (Lester-George 2023). With this costs and time, the hospital could enhance better patient services rather than spending too much time to train for new employee's overtime. Overall, to examine the analysis with the best prediction, recalling the data for the target variable is important to avoid the bias towards the majority class ('No' status).

Numerical variables

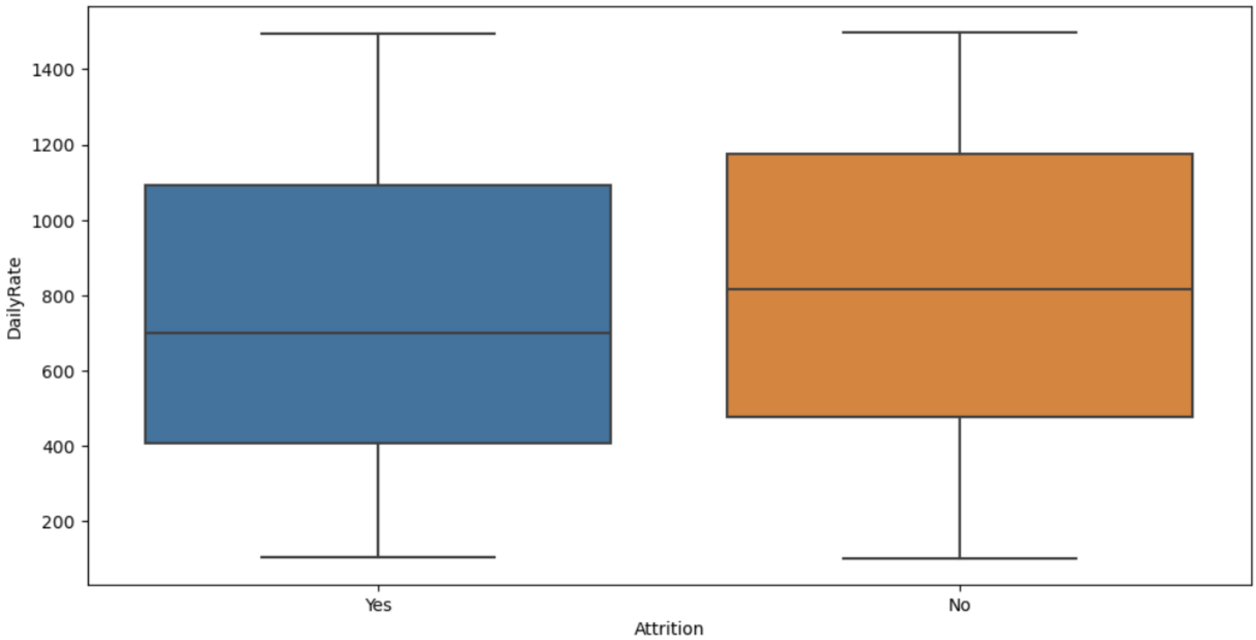
Table 3. The statistics of numerical variables

Column	Count	Mean	Std	Min	25%	50%	75%	Max
Age	1470	36.923810	9.135373	18.0	30.0	36.0	43.00	60.0
DailyRate	1470	802.485714	403.509100	102.0	465.00	802.0	1157.00	1499.00
DistanceFromHome	1470	9.192517	8.106864	1.0	2.00	7.0	14.00	29.0
Education	1470	2.912925	1.024165	1.0	2.00	3.0	4.00	5.0
EmployeeCount	1470	1.000000	0.000000	1.0	1.00	1.0	1.00	1.0
EmployeeNumber	1470	1024.865306	602.024335	1.0	491.25	1020.5	1555.75	2068.0
EnvironmentSatisfaction	1470	2.721769	1.093082	1.0	2.00	3.0	3.00	4.0
HourlyRate	1470	65.891156	20.329428	30.0	48.00	66.0	83.75	100.0
JobInvolvement	1470	2.729932	0.711561	1.0	2.00	3.0	3.00	4.0
JobLevel	1470	2.063946	1.106940	1.0	1.00	2.0	3.00	5.0
JobSatisfaction	1470	2.728571	1.102846	1.0	2.00	3.0	4.00	4.0
MonthlyIncome	1470	6502.931293	4707.956783	1009.0	2911.00	4919.0	8379.00	19999.0
MonthlyRate	1470	14313.103401	7117.786044	2094.0	8047.00	14235.5	20461.50	26999.0
NumCompaniesWorked	1470	2.693197	2.498009	0.0	1.00	2.0	4.00	9.0
PercentSalaryHike	1470	15.209524	3.659938	11.0	12.00	14.0	18.00	25.0
PerformanceRating	1470	3.153741	0.360824	3.0	3.00	3.0	3.00	4.0
RelationshipSatisfaction	1470	2.728571	1.102846	1.0	2.00	3.0	4.00	4.0
StandardHours	1470	80.000000	0.000000	80.0	80.00	80.00	80.00	80.0
StockOptionLevel	1470	0.793878	0.852077	0.0	0.00	1.0	1.00	3.0
TotalWorkingYears	1470	11.279592	7.780782	0.0	6.00	10.0	15.00	40.0

TrainingTimesLastYear	1470	2.799320	1.289271	0.0	2.00	3.0	3.00	6.0
WorkLifeBalance	1470	2.761224	0.706476	1.0	2.00	3.0	3.00	4.0
YearsAtCompany	1470	7.008163	6.126525	0.0	3.00	5.00	9.00	40.0
YearsInCurrentRole	1470	4.22925	3.623137	0.0	2.00	3.0	7.00	18.0
YearsSinceLastPromotion	1470	2.187755	3.222430	0.0	0.00	1.0	3.00	15.0
YearsWithCurrManager	1470	4.123129	3.568136	0.0	2.00	3.0	7.00	17.0

Numerical variables, such as *EmployeeCount, StandardHours*, and *EmployeeNume*r are unrelated to the target variables, are considered to drop due to its insignificant impact. This step aimed to identify all the numerical variables that might contribute to predicting attrition, such as whether higher *MonthlyIncome* and *DailyRate* correlate with reduced attrition rates (pay rate factors).

Figure 11. Box plot of *DailyRate* variable



Even if the box plot does not show a clear distinction, *DailyRate* could have a non-linear effect on attrition, possibly retained for validation process in a broader context. In a predictive model, removing it might reduce accuracy, so they have considered all pay-related variables for a comprehensive understanding behind the attrition rate among employees within any organisation.

2.4. Outlier Detection and Treatment

To ensure data integrity and enhance model accuracy, an outlier analysis was conducted on all numerical variables. The Interquartile Range (IQR) method was applied to identify and quantify outliers across the dataset. This method calculates the range between the first quartile (Q1) and the third quartile (Q3), with outliers defined as values lying below Q1−1.5×IQR or above Q3+1.5×IQR.

The outcomes of the analysis indicated that certain variables contained outliers:

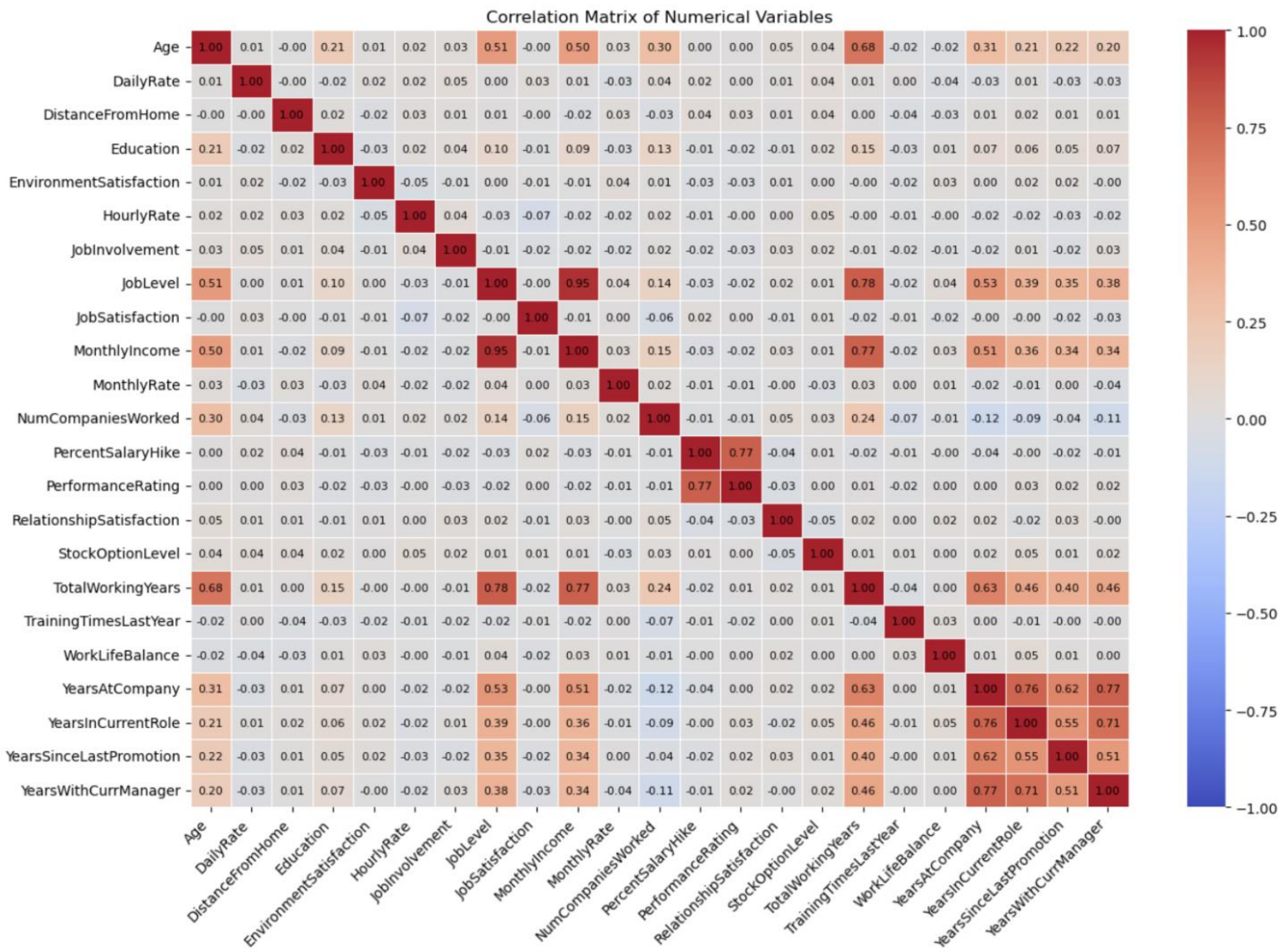
- *MonthlyIncome* has 114 outliers, taking up 7.76% of its values
- *NumCompanyWorked* has 52 outliers (3.54%)
- *PerformanceRating* has 225 outliers (15.37%)
- *StockOptionLevel* has 85 outliers (5.78%)
- *TotalWorkingYears* has 63 outliers (4.29%)
- *TrainingTimesLastYears* has 238 outliers (16.19%)
- *YearsAtCompany* has 104 outliers (7.07%)
- *YearsInCurrentRole* has 21 outliers (1.43%)
- *YearsSinceLastPromotion* has 107 outliers (7.28%)
- *YearsWithCurrManager* has 14 outliers (0.95%)

These outliers were reviewed for their potential impact on the model. Due to their relevance and the fact that they represent real variations in employee experience, these outliers were kept for further investigation, maintain the dataset's integrity and capturing authentic variability in employee attributes.

2.5. Multicollinearity check

Multicollinearity, which occurs when predictor variables in a model are highly correlated, can adversely affect model interpretation and the significance of individual predictors (Jim 2023). A correlation matrix heatmap was generated among predictors variables (x variables), focusing on high correlations (approximately above 0.7 or below -0.7) which could indicate problematic multicollinearity. A logistic regression model is sensitive with the large or imbalanced data as it will skew the data and generate an inaccurate result for the prediction. In the long-term, the model will not be in used for examining the attrition cases if they produce high relevance towards the majority class.

Figure 12. Correlation Matrix for Multicollinearity check

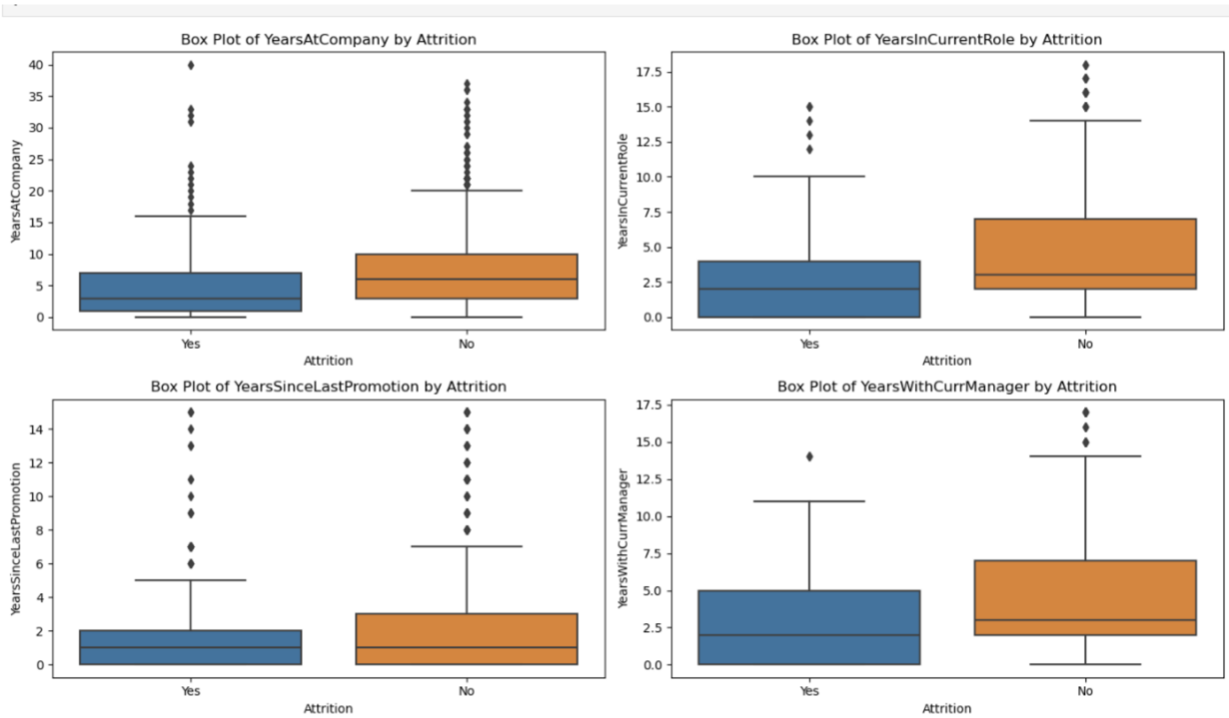


The correlation matrix revealed tenures related variables with high correlations:

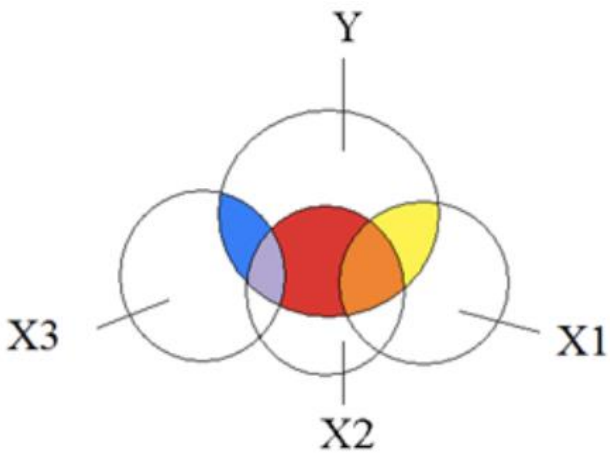
- *YearsAtCompany* had a high correlation with *YearsInCurrentRole* (0.76), *YearsSinceLastPromotion* (0.62), and *YearsWithCurrManager* (0.62), suggesting that an employee's overall tenure overlaps substantially with time spent in specific roles and intervals between promotions.
- *YearsInCurrentRole* correlated with *YearsWithCurrManager* at 0.71, which could reflect a strong relationship between their tenure spending in a role and the continuity with the same manger.
- *YearsSinceLastPromotion* shows moderate correlation with *YearsAtCompany* (0.62) and *YearsInCurrentRole* (0.46), indicating that promotion timing may be related to the overall tenure and role duration. As a result, *YearsInCurrentRole* is kept for further analysis due to its representative and distinct relationship with the *Attrition* variable. The box plot shown in the **Figure 12** demonstrating that there is a mean difference (central line in each box) of

YearsInCurrentRole between '**Yes**' and '**No**' categories. While *YearsAtCompany* and *YearsWithCurrManager* are related, they do not always indicate the same level of role management as *YearsInCurrentRole*. An employee might have the same manager in one hospital but have different roles overtime. Moreover, *YearsInCurrentRole* might be seen as a more direct measure of an employee's stability and satisfaction in their position, which could be closely tied to their decisions to leave.

Figure 13. Box plot of tenure variables



These tenure variables could introduce redundancy when it comes to the implementation of the regression model, when the independent variables are explaining each other, leading to the less precise prediction for the y variable (dependent variable).



Source: The Analysis Factor (n.d.).

If used together in the logistic regression, it will potentially establish the model with diluting interpretation because of wider confidence intervals (Karen n.d.).

In addition to these variables, there are another pair of variables showing their strong relationship among all numerical variables:

- *MonthlyIncome* and *JobLevel* shows a strong correlation of 0.95, suggesting that as job level increases, corresponding to monthly income, creating redundancy between these variables.
- *TotalWorkingYears* and *JobLevel* displays a high correlation (0.78), further reinforcing the relationship between experience and job level in the dataset.
- *MonthlyIncome* and *TotalWorkingYears* shows a strong correlation of 0.77, suggesting the common sese of having longer tenure will corresponding to the higher salary.
- *PercentageSalaryHike* and *PerformanceRating* has the correlations of 0.77, suggesting their relationship when higher percentage in salary hike will correspond to the higher performance rating.

Overall, while those variables are insightful, by selectively dropping them will help to reduces noise, build an optimised logistic regression model for analysis because of its sensitive to explain the most dependent variable. With an imbalance in this dataset, focusing on the impactful variable solely with abilities to explain the attrition rate at most is crucial, suggesting a strong model without the confounding effects of interrelated predictors. This streamlined approach not only simplifies the model but also ensures more stable coefficient estimates, leading to a clearer interpretation of how each factor contributes to the likelihood of attrition.

2.6. Logistic Regression Modelling

The logistics regression is chosen as the classification model due to its suitability for the binary outcomes of Attrition status (‘Yes’ or ‘No’). Those steps involved in this modelling phase are detailed below:

Data Preparation for Modelling

The dataset was consolidated to ensure compatibility across separated data frames: numeric and categorical. The concating step is conducted for these data frames along columns using the inner join function, ensuring only rows where *Attrition* category variable exists in both df_numeric and df_cat. The *Attrition* column is isolated as the target variable y, while all reaming variables are considered as feature variables x.

To address class imbalance (the majority is No status, accounting for 84%), which can lead to biased predictions, the Synthetic Minority Oversampling Technique (SMOTE) is applied. SMOTE generated synthetic samples for the minority class, addressing an imbalance of Attrition rate in this dataset (balance 'Yes' and 'No' classes). The outcome has been balanced with the 50% for 'Yes' class and 50% for 'No' class.

Attrition	
1	1233
2	1233

With an aim of achieving the best model performance, this oversampling step helps the model avoid bias toward majority class, which improves generalisation. Balanced classes are especially important for logistic regression to generate fair and accurate in predicting attrition, as imbalanced class can lead to less precise for the outcomes and biased predictions.

Train-Test Split and Scaling

The dataset is split into training and testing subsets using an 80:20 ratio, ensuring the model can learn from the significant portion of the data while leaving enough data for an unbiased performance. Moreover, scaling is applied using the StandardScaler to standardise the range of numerical variables, such as 'Age' range is different from the 'MonthlyIncome' range, to ensure that larger numerical ranges do not dominate the model's process. For example, the 'Age' variable will range from 18 – 60 years old, while MonthlyIncome will range from $1,009 - $19,999USD. As the result, the model can capture the best portion of the data and ensure no single variables skews the results, enhancing a balanced analysis. In the healthcare industry, there are a broader range from the young interns to professionals with long tenure nearing their retirement period. In addition, the logistic regression model is trained on the scaled training data for x train and x test. By scaling the data to have mean of 0 and standard deviation of 1, the model's learning process becomes more stable and unbiased across predictor variables.

Table 4. Results summary

```
Accuracy: 0.93
Confusion Matrix:
[[240  10]
 [ 27 217]]
Performance Measure:
              precision    recall  f1-score   support

           0       0.90      0.96      0.93       250
           1       0.96      0.89      0.92       244

    accuracy                           0.93       494
   macro avg       0.93      0.92      0.92       494
weighted avg       0.93      0.93      0.92       494
```

In this result table, the model coefficients provide insights into the strength and direction of influence that various factors (tenure or work-life balance) have on attrition risk. The outcome achieved an accuracy of 93%, indicating a high overall success rate in classifying both ‘Yes’ (leave) and ‘No’ (stay) outcomes. It shows that out of 494 observations, 240 individuals are correctly predicted as ‘No’ attrition (stay) and 217 as ‘Attrition’ (likely to leave). The model has only 10 false positives (predicting attrition where there is none) and 27 false negative (predicting no attrition when attrition occurs). This balance is essential to be considered since the aim of the model is to achieve low false negative. A false negative means predicting an employee will stay, but in reality, they leave, leading to missed opportunities to intervention or launching any offers to retain, especially for talented employees. However, there are multiple reasonings to get the low number in false positives, due to specific business plans, and background of the finance prospects within the healthcare providers.

Confusion matrix

[240 10]

[27 217]

The confusion matrix highlights that the model distinguishing between two classes. In healthcare, where accurate predictions of attrition rate are pivotal for staffing and patient care performance, such metrics help determine how well the model split the employees who are likely to leave and stay. By achieving a positive number of the model, it helps the healthcare providers to measure its performance and internal workforce capacity, leading to effectively utilise costs and time for service quality.

Regarding the classification report, the performance measures (including precision, recall, and f1-score are balanced across ‘Yes’ and ‘No’ classes, with F1 score is 0.93 and 0.92, respectively. With the balanced score, it indicates that the model is effectively explaining at most the data with precise prediction for Attrition cases.

Performance Measure:

	precision	recall	f1-score	support
0	0.90	0.96	0.93	250
1	0.96	0.89	0.92	244
accuracy			0.93	494
macro avg	0.93	0.92	0.92	494
weighted avg	0.93	0.93	0.92	494

The outcomes of an optimised logistic regression model are an important feature for sensitive fields like healthcare, where misclassification can lead to waste in costs of hiring or retention mistakes. By taking the proactive steps for the retention programs, it is crucial to consider all the variables within the business context, suggesting meticulous steps to deal with the 'Yes' cases.

2.7. Model Evaluation

To confirm the validity and the application of the model, the logistic regression model is compared with the K-Nearest Neighbors (KNN) model with $k = 2$ to determine the most effective approach for predicting employee attrition rates. The logistic regression model shows a higher accuracy of 0.93, while the KNN model's accuracy of 0.91. In healthcare, where workforce stability is critical for maintaining patient care standards, the logistic regression model's recall of 0.89 for the attrition class, using the calculation $Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives}$, is particularly advantageous. The higher recall indicates a lower false negative rate, which reduce the risk of overlooking the 'Yes' attrition cases within the internal workforce. Such an approach supports proactive retention strategies, as failing to predict at-risk employees can result in unexpected staff shortages that may compromise patient care.

In addition, choice of scaler for each model highlights the different needs in data preprocessing based on model architecture. The logistic regression model benefits from standardisation via the StandardScaler, which assumes normally distributed data and scales it to a mean of 0 and standard deviation of 1. For the logistic regression model, it is less sensitive to data outliers. However, the KNN is a distance-based model, which is highly susceptible to outliers, necessitating the use of RobustScaler. The RobustScaler is more resilient to outliers by scaling the data based on the interquartile range, ensuring that extreme values do not disproportionately affect distance calculations between points in KNN. This differentiation highlights the importance of model-specific preprocessing to achieve optimal predictive accuracy and robustness in healthcare analytics.

The KNN model with $k = 2$ yields an accuracy of 0.90, which close to the outcomes achieving from the logistic regression model (0.93). The difference is slight lower, which is still critical in examining the current attrition status in high-stakes environment as healthcare, where even minor change in improvements can translate to significant cost-savings and operational benefits. The confusion matrix of KNN model shows that there are 216 true positives and 229 true negatives, with 34 false positives and 15 false negatives. The logistic regression model has lower false negative cases (who predicted to stay but likely to leave) than KNN. A missed capturing the attrition cases can disrupt patient care, staffing balance and the skilled roles are left vacant unexpectedly (especially the roles in ED or nursing). Thus, the slightly lower false negative in a logistic regression model can be preferable in terms of preventing attrition.

Table 5. Results summary for KNN model with $k = 2$

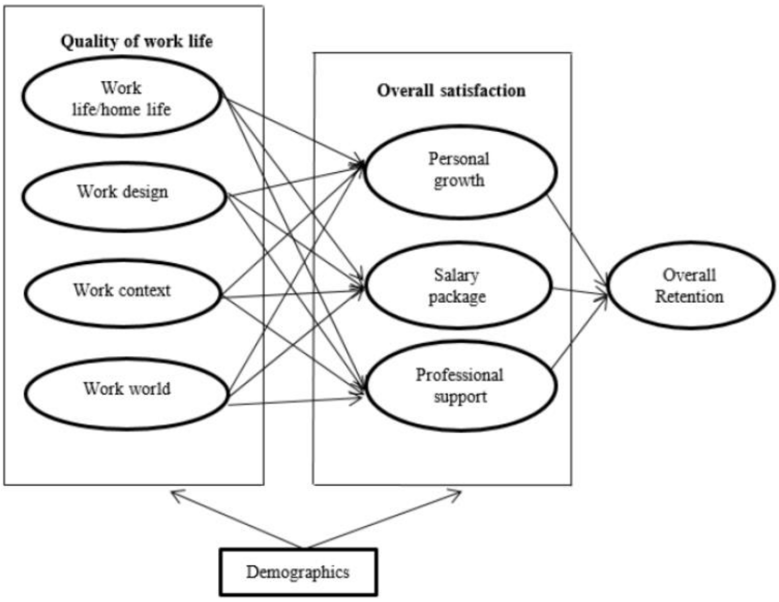
```
Accuracy: 0.90
Confusion Matrix:
[[216  34]
 [ 15 229]]
Performance Measure:
              precision    recall  f1-score   support

           0       0.94      0.86      0.90       250
           1       0.87      0.94      0.90       244

    accuracy                           0.90       494
   macro avg       0.90      0.90      0.90       494
weighted avg       0.90      0.90      0.90       494
```

Consequently, logistic regression provides more an interpretable model with coefficients that can be analysed using the insights from the predictor factors' impacts on attrition. This interpretability is valuable for healthcare administrators who need to understand the underlying factors driving attrition and devise new retention program suitably with the background and the context. Conversely, KNN, as a distance-based model, it will have less abilities to interpret, making it harder to derive actionable insights from individual predictors. In addition, KNN is sensitive with the outliers while the employee datasets can include variety of features with various limits (eg: age, experience, satisfaction scores…). Due to this disadvantages, logistic regression is generally more stable for capturing the nuances and performing more consistently across healthcare-focused applications.

3. PRESCRIPTIVE ANALYSIS

The analysis utilises an attrition framework developed specially for the healthcare sector, where there is a focus on the quality of work life, overall satisfaction impacting on the overall retention. According to Musrrat Parveen's model, key factors are work-life/home-life balance, work design, work context, and in broader workplace environment. All of which contribute significantly to employees' satisfaction and their decisions of straying or leaving. Parveen's model shows that components such as personal growth, salary package, and professional support driving the attrition cases, suggesting that the retention programs should be targeted towards those areas to ensure their significant impact on reducing the rates.

Figure 13: A model of attrition within a healthcare sector (Musrrat Parveen 2016)



Research has shown that healthcare employees often struggle with overloads, insufficient professional growth opportunities, and non-work schedules with repetitive tasks (Kossek, Rosokha & Leana 2019). These issues are exacerbated due to the growth of global pandemic COVID-19, when

Evening Supervisor, Site A: *'The constant changing in the scheduling make it extremely difficult to meet the workforce's needs consistently. This ongoing instability impacts our ability to support employees effectively'*

Nurse Practice Educator, Site F*: 'Our work involves caring people, not manufacturing objects, which requires us to have high responsibility within what we do'*

(Kossek, Rosokha & Leana 2019)

there is a flood of extremely ill COVID-19 cases, pushing pressure on carers and the whole healthcare system. The unstable internal workforce leads to the weak system of patient caring services, interrupting the operational efficiency. By enhancing these aspects, caring providers should launch the program to address these issues to improve satisfaction and engagement among members. The model and the employees' evaluations provide a foundation for designing targeted program to effectively align with special needs of healthcare workers, supporting well-being, and career aspirations.

3.1. Proposed Retention Programs

The proposed retention programs address key factors impacting attrition among healthcare professionals, aiming to reduce turnover rates to around 10% equivalent to two-thirds of the identified cases of potential attrition (217 cases in the predictive analysis). By targeting specific challenges faced by healthcare employees, three programs: **Employee Wellness Program**, **Professional Development Program**, and **Flexible Work Initiatives** seek to foster a more supportive and sustainable work environment that aligns with employees' unique needs.

3.1.1. Program 1: Employee Wellness Program

The Employee Wellness Program is specifically designed to address the mental and physical well-being of healthcare employees, targeting stress-induced attrition factors prevalent in high-stakes environments. Evidence from the healthcare sector suggests that healthcare employees, especially in private hospital settings, experience elevated levels of burnout due to prolonged exposure to high-stress situations, often leading to adverse health outcomes (Marković et al. 2024). The World Health Organization (WHO) advocates prioritising mental health in workplace settings, noting that enhancing psychological well-being can significantly reduce stress and improve employee retention, aligning with the objectives of this program (WHO 2024).

This program aims to alleviate symptoms of burnout by equipping employees with resources to manage mental health and promote physical fitness. Budgeted at approximately $400 per employee, the program will fund wellness coaches, mental health professionals, on-site wellness activities, and wellness resources, maximum up to 98 employees. Costs for wellness initiatives can vary based on program scope and delivery components; however, studies indicate that effective wellness programs typically cost between $200 and $500 per employee annually, depending on the intensity and comprehensiveness of the services offered (Wellsource 2024). According to research from the American Journal of Health Promotion, wellness programs incorporating counselling and fitness components tend to yield higher participation rates and deliver more substantial impacts on employee well-being (EAP Employee Wellness, 2024). It is strategically to launch from these programs in different periods of a year rather introducing all there concurrently, with the allocated budget at around $100 per employee in 3 months fixed duration. This approach allows adequate time for employee's evaluations to make any needed adjustments. This structure ensures a focused intervention period, allowing for tangible outcomes to be observed and evaluated effectively. Generally, some wellness activities could be introduced in a class with the highest intensive, well-suited with tight budget healthcare providers while saving money for the operational costs.

The Employee Wellness Program will strategically partner with local wellness centres to enhance accessibility and offer combined on-site and virtual wellness resources, ensuring program flexibility and effectiveness. Considering the predictive analysis that identifies 217 employees as likely to leave, it is essential to implement targeted programs like this to retain a significant portion of the workforce. It has the highest capacity (at 98 slots available), due to the typical factors driving rates are stress with high volume of workloads and vulnerability about their financial burdens, implying that wellness sessions are needed to cope with their mental and physical health issues. By building the supportive environment, this focus is expected to manage workplace stress and enhance resilience.

3.1.2. Program 2: Professional Development Program

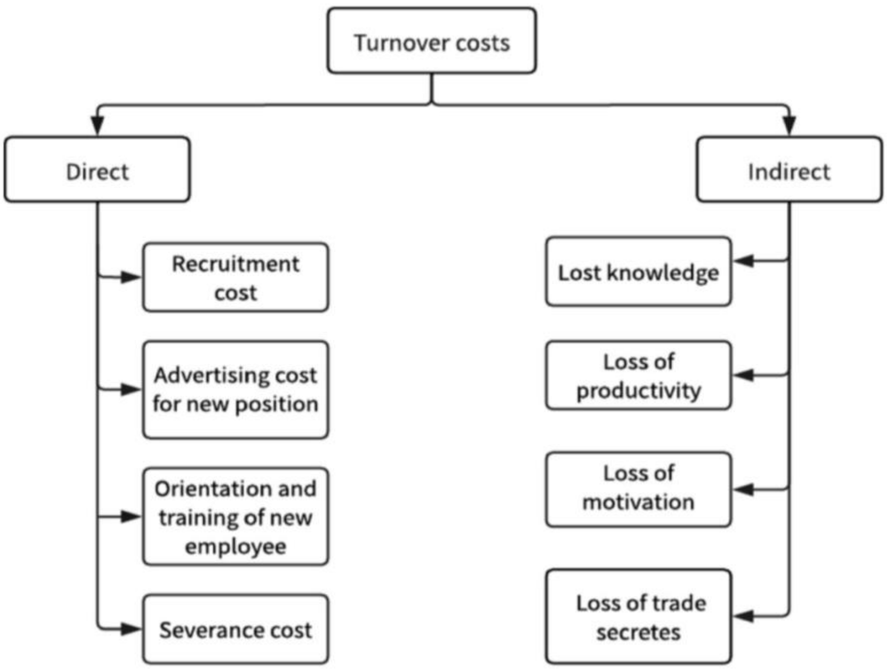
The Professional Development Program addresses the lack of career growth and skills development, which is defined as one of the most impactful factors driving the attrition rate. This program is structured to support employees' professional aspirations, providing them with skills-based training, mentoring opportunities, and access to career advancement resources. It is essential to offer them with a clear pathway for career progression and reducing dissatisfaction with perceived limited growth opportunities. These incentives could push and somewhat help employees' having a positive outlook and reducing perceived stagnation.

With a budget allocation per employee is around $350, it is designed for healthcare workers, included career counselling sessions. This program is estimated to sufficiently support up to 76 workers. While career growth and skill development are crucial for retention, the immediacy of these needs is secondary to direct wellness support. If the program is effective in the first two months, it might be justified in next period and scale up if received positive outcomes. Due to the typical trends from the dataset, it is a notable that there is a high number of employees leaving in the first few years (first 3 years), who could be the targeted employees for this program. Overall, this prioritisation is designed to retain employees by reinforcing a pathway for career advancement, thus helping healthcare workers feel more valued and fulfilled in their roles.

3.1.3. Program 3: Flexible Work Initiative

The implementation of flexible work policies often involves minimal overhead costs, especially if it primarily includes scheduling adjustments or remote work flexibility without significant infrastructure costs. For a flexible work initiative, the cost might include a small investment in scheduling tools or employee stipends for home office setups, roughly \$100–\$200 per employee. With the reasonable costs for this program, it should take between the range from \$8,000 to \$10,000, estimated to around \$8,500. The potential challenge of this program comes along with ethical, morale and responsible workers who need to ensure the job's quality in the remote areas. Some hospital systems are conducting the 'virtual workplace' or 'virtual nurses' which allow workers to control patient documents effectively without being on-premises floor workers (Preston 2023). This program suitable for employee who travel frequently, who could take time to check the patient's documentations and medications. Conversely, it is not preferable for nursing in ED because of the

Figure 13: A turnover costs model (Douaidi & Kheddouci 2022)



role characteristics and hectic paces.

In assessing the comparative benefits of retention programs versus new employee recruitment, studies show that turnover costs, including training and recruitment, are often higher than the investments needed to retain experienced staff (Douaidi & Kheddouci 2022). In addition, this program has the smallest capacity at 43 employees, reflecting the complexity and potential future risks associated with remote or flexible work in a healthcare setting, where responsibilities often require physical presence. Thus, this program is scaled to manage fewer participants initially, allowing for a controlled, measured approach to offering flexible work arrangements without compromising the essential service delivery standards. Overall, it is optimal to retain current employees with the retention programs well-aligned with their special needs, while still considering all the aspects related such as direct and indirect costs. It is a foundational step toward cultivating resilient workforce, reducing attrition, retaining competency employees.

3.2. Model Minimisation

In this analysis, the mathematical minimisation model will be utilised to evaluate the performance of the retention programs, based on the estimated budget and the attrition rates from the predictive analysis part after rescaling. Mathematical optimisation is powerful approach for solving complex decision-making problems by finding the most efficient allocation of resources under specified constraints. Especially, in the situation of having random number of employees with different slots for each program, defining an objective function and various constraints relevant to retention can provide optimal distribution of random employee across different programs to minimise attrition risk. Despite of allocating randomly employee to each program, it is essential to consider their background and their problems leading to the attrition rates. The mathematical model will be implemented using Gurobi Python, especially in the context of handling a binary variable ('Yes' or 'No') and constraints (for the retention programs). Gurobi's solver is particularly well-suited for applications requiring precision and adaptability, allowing this model to evaluate and compare the effectiveness of various retention programs based on capacity, budget, and specific goals for each program. This optimisation-based framework aims to reach the target of reducing the attrition rate to around 10%, by industry benchmarks, ultimately supporting a resilient and stable healthcare worker.

The decision variable will be defined as a binary variable, indicating the person who is likely to stay or leave. Variable x_{ip} illustrates whether an employee i is used so x_{ip} takes 1 if employee i is allocated to program p with $i \in \{1,2,..,217\}$, $p \in \{1,2,3\}$ which are Employee Wellness Program, Professional Development Program and Flexible Work Initiatives. The α_i is representing the benefit impact of assigning each employee to a program, is defined as:

$$\alpha_i = \frac{1}{MonthlyIncome_i}$$

This value indicates a converse relationship as lower-income employees are thus given higher α_i values, increasing the likelihood they are selected for retention support. The MonthlyIncome is chosen a primary variable to calculate alpha values because employees with lower income are often at the higher risk of leaving due to financial considerations. The model aims to effectively reduce the attrition rate in terms of economic pressures. These issues are happening within the role women taking and the people of colour, while it is witnessed that in healthcare industry, there are 86% are women, 55% people of colour and 23% immigrant workers (McGregor & Johnston 2022). With the majority of employees from the at-risk situation are likely to leave, it is essential to consider their incomes as a numerical factor to examine for further analysis. Besides, the other variables such as *YearsAtCompany, Age* or *JobSatisfaction* are excluded from this model because they introduce subjectivity and variability within their ranges, indicating that it is challenging to translate into a actionable retention strategy. Since the satisfaction scores are varies based on personal expectations or short-term factors, it is difficult to define consistent thresholds that would indicate the risk of attrition. Hence, the income factor has been chosen as a value to reduce proportionally the attrition rate aligning with the goal of supporting employees who may be financially motivated to leave.

The objective function (1) is defined with an aim of minimising the attrition rate with the impact of each employee within these programs.

$$Minimise\ z = \sum_{i=1}^{n}\sum_{p=1}^{m} \alpha_i\ .x_{ip} \qquad (1)$$

Based on the realistic situation, it is crucial to apply several constraints to this problem. Constraint (2) guarantee that the cost of enrolling employees in each program should not exceed the allocated program budget:

$$\sum_{i=1}^{n} c_p\ .x_{ip} \leq B_p \qquad (2)$$

where

- For the Wellness Program (c_1): c_1= 400 and $B_1 = 400\,.98 = 39200$
- For the Development Program (c_2): c_2= 350 and $B_2 = 350\,.76 = 26600$
- For the Flexible Work (c_3): c_3= 150 and $B_3 = 150\,.43 = 6450$

The justification table explaining each program's capacity in relation to its cost per employee and total employee allocation for each program:

Program	Cost per employee	Percentage of Total Employees	Final program capacity	Justification
Wellness	$400	45%	$0.45\,.217 \approx 98$	Higher cost per employee implies the most intensive support provided (eg: wellness classes, one-one consultations...), so this is allocated to a largest capacity
Development	$350	35%	$0.35\,.217 \approx 76$	Moderate cost of employee is the second-priorised program development workshops and counselling, reusing in moderate allocation
Flexible Work	$150	20%	$0.20\,.217 \approx 43$	Lowest cost per employee, supporting remote work and flexible scheduling, which allows broader access at a lower cost.

With this allocation method, it ensures that the total capacity aligns with 217 employees (utilised the number derived from the rescaling predictive analysis). Due to the different importance regarding the problem defined, the Wellness will be in the top priority for addressing the attrition rate, which cares about the physical and mental health. This approach allows the model to address the varied retention needs effectively with reasonable costs. As defined in **Figure 13**, the costs for the retention program contains two subparts such as direct costs and indirect costs, indicating that with the detailed and specific number of program capacity would help to achieve the goal of reducing the attrition rate.

The number of employees will be assigned to each program cannot exceed the program's maximum capacity, defined as constraint (3):

$$\sum_{i=1}^{n} x_{ip} \le C_p \qquad (3)$$

Based on the table justification, the $C_1 = 98,\ C_2 = 76,\ C_3 = 43$ for Wellness Program, Development Program and Flexible Work Program, respectively. Additionally, the program assignment is defined in constraint (4), ensuring that one employee must be enrolled in only one program:

$$\sum_{p=1}^{m} x_{ip} = 1 \qquad (4)$$

The base of attrition rate is around 16%, after the optimal allocation, the estimated attrition rate is calculated by subtracting the total impact reduction achieved by the model from the base attrition rate:

$$Estimated\ attrition\ rate = Base\ attrition\ rate - \sum_{i=1}^{n}\sum_{p=1}^{m} \alpha_i \ . x_{ip}$$

The model checks if the estimated rate is close to the target rate of 10% by assessing whether it falls within the range of 10% (acceptable tolerance $\pm 1\%$). This minimisation model provides the robust framework of examining the attrition while introduce the impactful programs for future planning and decision making.

With the integration of the retention programs, the attrition rate is estimated reducing from over 16% to around 10%, suggesting the healthy attrition rate in any certain organisation. This model is using α_i as the monetary value to implying the factor driving the attrition rate while it depends on business goals, budgets and contexts. The α_i could be considered to add another variable with the significant impacts into the model with an objective of covering more aspects in the analysis. By doing so, there is a need to consider the variable's types and specific ranges, due to some limitations in the differences between numerical variables and categorical variables.

LIMITATION

Notably, the findings of this study are limited only to its current form, especially regarding the variety of new variables could be added. Thus, it is not generalised across all prostitutions due to different environment factors and business goals. Considering of including more variables to uncover the pattern could lead to the risk of overfitting, make the model more complex and reduce its precision. Therefore, it is a challenge for stakeholders to deliver the optimal method to address the internal workforce problem. For example, some categorical variables, such as job roles or departments, introducing handling challenges that may not align with the current numeric-focused structure, and as the different scales could interrupt the model's balance, reflecting the delays in decision making in the long-term. The solution for this limitation is including a scaling factor, dividing the impact portion for each new variable, while ensuring the consistent throughout the iterations. Overall, the model would benefit from ongoing evaluation and potential adjustments to address these limitations while maintain the effectiveness as organisational needs changes overtime.

RECOMMENDATION

As a recommendation, the above analysis implies that effective retention strategies must balance the diverse needs of all stakeholders with the delivery of healthcare – employees, managers, and the patients. Employees seek stability, career development, and work-life balance; manager aim to maintain the performance with the stable administrative, and employee engagement, while the patients are looking for the quality of care. Achieving the balance requires the robust framework of workforce within any healthcare providers, suggesting introducing effective retention programs to support employees' working lives. Hence, the flexibility in program design will indirectly help to achieve the operational efficiency. In the landscape of employee needs continuous to evolve, understanding and sympathising the employee recognition program or promotion opportunities are essential to cope with these dynamics and help to retain competency employee, contributing to a healthy and productive workplace. Also, the integration of new programs will yield both pros and cons towards the internal workforce, leaders will have to confront unexpected factors due to the constantly opinion's changes between employees. As a result, it is crucial to consider the compatible of new policies or program regarding the healthcare providers' background and long-term objectives.

CONCLUSION

Effective employee retention strategies not only benefit individual employees but also contribute to organisational success by enhancing well-being, service quality and overall satisfaction among clients or patients. This analysis identifies key retention strategies such as promoting healthy work environment, fostering the relationship between manager-employee and employee-patients. All of which are projected to the reduce the defined attrition rate from over 16% to around 10%, suggesting a healthy rate within the labour force. Additionally, seniors and leaders play a crucial role in these efforts, as leaders who prioritise employee satisfaction, professional growth and supportive relations are more likely to retain talented and cultivate a positive workplace environment. Hence, it effectively reduces the attrition rate with an objective of encouraging team-based structure within healthcare sector and strengthen morales, contributing to the long-term retention. Strong relational between mangers, employees and patients will help employee feel valued and motivated to stay while the patients are benefit from having the quality services. Overall, these strategies are necessary to build the resilient, committed workforce and supports the organisation's mission in the long term.

REFERENCES

Boatman, A 2021, 8 Causes of Job Dissatisfaction & How to Combat It, *Academy to Innovative HR*, viewed 22 October 2024, <https://www.aihr.com/blog/job-dissatisfaction/>

CHRMP 2022, What is Attrition Rate? 5 Types, Importance, Formula and More 2022, *CHRMP*, viewed 26 October 2024, < https://www.chrmp.com/what-is-attrition-rate/>

Chervoni-Knapp, T 2022, 'The Staffing Shortage Pandemic', *Journal of radiology nursing,* vol. 41, no.2, pp. 74-75, doi:10.1016/j.jradnu.2022.02.007.

Douaidi, L & Kheddouci, H 2022, 'A new approach for employee attrition prediction', in T Braun, D Görz & R Molitor (eds), *Graph-based representation and reasoning: ICCS 2022*, Lecture Notes in Computer Science, vol. 13403, Springer, Cham, Switzerland, pp. 115–128, viewed 6 November 2024, <https://doi.org/10.1007/978-3-031-16684-3_10>

Das, BL & Baruah, M 2013, 'Employee retention: A review of literature', *Journal of business and management*, 14(2), 8-16, doi:10.9790/487X-1420816

EAP Employee Wellness 2024, 'Employee wellness program cost', *EAP Employee Wellness*, viewed 6 November 2024, <https://eapemployeewellness.com/corporate-wellness/employee-wellness-program-cost/>

Grace-Martin K, n.d., 'What is multicollinearity? A visual description', *The Analysis Factor*, viewed 4 November 2024, <https://www.theanalysisfactor.com/multicollinearity-explained-visually/>

Jim F, 2023, 'Multicollinearity in regression analysis', *Statistics By Jim*, viewed 3 November 2024, <https://statisticsbyjim.com/regression/multicollinearity-in-regression-analysis/>

Hayes, LJ, O'Brien-Pallas, L, Duffield, C, Shamian, J, Buchan, J, Hughes, F & Stone, PW 2006, 'Nurse turnover: A literature review', *International Journal of Nursing Studies*, vol. 43, no. 2, pp. 237–26, doi:10.1016/j.ijnurstu.2011.10.001

Kossek, EE, Rosokha, LM & Leana, C 2019, ‘Work schedule patching in health care: exploring implementation approaches’, *Work and Occupations*, vol. 47, no. 2, pp. 228–261, doi:10.1177/0730888419841101.

Lindquist M, 2023, The Real Costs of Healthcare Staff Turnover, *Oracle*, viewed 21 October 2024, < https://www.oracle.com/au/human-capital-management/cost-employee-turnover-healthcare/>

Lartey, S, Cummings, G, & Profetto-McGrath, J 2014, Interventions that promote retention of experienced registered nurses in health care settings: A systematic review, *Journal of Nursing Management*, 22(8), 1027-1041, doi:10.1111/jonm.12105.

Lester-George A, 2023, The hidden costs of replacing lab talent, *Binocs Technology*, viewed 3 November 2024, <https://www.bluecrux.com/blog/the-hidden-costs-of-replacing-lab-talent/>

McKinsey 2022, ‘Assessing the lingering impact of COVID-19 on the nursing workforce’, *McKinsey & Company*, viewed 26 October 2024, <https://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/assessing-the-lingering-impact-of-covid-19-on-the-nursing-workforce>.

McGregor, K & Johnston, H 2022, 'Why improving low-wage health care jobs is critical for health equity', *AMA Journal of Ethics*, vol. 24, no. 9, viewed 7 November 2024, <https://journalofethics.ama-assn.org/article/why-improving-low-wage-health-care-jobs-critical-health-equity/2022-09>

Marković, S, Kostić, O, Terzić-Supić, Z, Mihajlović, S, Milovanović, J, Radovanović, S, Zdravković, N, Stojić, V, Jović, L, Jocić-Pivač, B, Tomić Lučić, A, Kostić, M & Šorak, M 2024, ‘Exposure to Stress and Burnout Syndrome in Healthcare Workers, Expert Workers, Professional Associates, and Associates in Social Service Institutions’, *Medicina (Kaunas)*, vol. 60, no. 3, p. 499, doi:10.3390/medicina60030499.

NSI Nursing Solutions 2023, NSI National Health Care Retention & RN Staffing Report, *NSI Nursing Solutions*, viewed 22 October 2024, <https://www.nsinursingsolutions.com/Documents/Library/NSI_National_Health_Care_Retention_Report.pdf>

Poon, YS, Lin, YP, Griffiths, P, Yong, KK, Seah, B & Liaw, SY 2022, 'A global overview of healthcare workers' turnover intention amid COVID-19 pandemic: a systematic review with future directions', *Human Resources for Health*, vol. 20, no. 70, doi:10.1186/s12960-022-00764-7.
Preston, R 2023, '7 strategies for improving healthcare workforce retention', *Oracle Human Capital Management*, viewed 6 November 2024, <https://oracle.com/au/human-capital-management/>

Stone J, 2022, 'We're facing a critical shortage of medical laboratory professionals', *Forbes,* viewed 3 Novermber 2024, <https://www.forbes.com/sites/judystone/2022/04/28/were-facing-a-critical-shortage-of-medical-laboratory-professionals/>

Winstanely, G 2023, Employee Disengagement: Navigating the Disconnect, *Mentorloop Mentoring Software*, viewed 28 October 2024, <https://mentorloop.com/blog/employee-disengagement/>

World Health Organisation 2024, 'Mental health at work', *WHO,* viewed 6 November 2024, <https://www.who.int/news-room/fact-sheets/detail/mental-health-at-work>

Wellsource 2024, 'How much should a good wellness program cost?', *Wellsource Blog*, viewed 6 November 2024, <https://blog.wellsource.com/much-good-wellness-program-cost>

APPENDIX

Appendix 1

Column	Details
Age	Represents the employee's age.
Attrition	Indicates whether the employee has left the company (Yes/No).
BusinessTravel	Shows how often the employee travels for business purposes (eg: rarely, frequently).
DailyRate	The employee's daily earnings.
Department	Specifies the department where the employee works, such as Research & Development.
DistanceFromHome	Measures the distance between the employee's home and workplace.
Education	Represents the educational attainment level of the employee.
EducationField	Field of study of the employee, for example, Life Sciences or Medical.
EmployeeCount	Total count of employees (likely set as 1 for each record).
EmployeeNumber	Unique identifier assigned to each employee.
EnvironmentSatisfaction	Indicates satisfaction level with the work environment, on a scale (1-4).
Gender	Employee's gender
HourlyRate	Hourly wage rate of the employee.
JobInvolvement	Level of the employee's involvement or engagement in their role, on a scale (1-4).
JobLevel	Indicates the position level or rank within the company.
JobRole	The specific role or title of the employee, such as Laboratory Technician or Research Scientist.
JobSatisfaction	Satisfaction with the job, on a scale (1-4).
MaritalStatus	Marital status of the employee (eg: Single, Married).
MonthlyIncome	The employee's monthly income.
MonthlyRate	Monthly payment rate for the employee.
NumCompaniesWorked	Number of different companies the employee has previously worked for.
Over18	Specifies if the employee is over 18 years old
OverTime	Indicates if the employee works overtime (Yes/No).
PercentSalaryHike	Percentage increase in the employee's salary.
PerformanceRating	Performance assessment rating, on a scale (3-4).
RelationshipSatisfaction	Satisfaction with relationships at work, on a scale (1-4).
StandardHours	Standard work hours, possibly a constant (80 hours).

StockOptionLevel	Level of stock options granted to the employee.
TotalWorkingYears	Total years of work experience the employee has.
TrainingTimesLastYear	Number of training sessions attended in the last year.
WorkLifeBalance	Satisfaction level with work-life balance, on a scale (1-4).
YearsAtCompany	Total number of years the employee has been with the company.
YearsInCurrentRole	Number of years spent in the current job role.
YearsSinceLastPromotion	Number of years since the employee's last promotion.
YearsWithCurrManager	Number of years the employee has worked under their current manager.