AUGUST 2023

SPRINGBOARD CAPSTONE TWO REPORT

PREDICTING EMPLOYEE ATTRITION

FATIH DEMIROZ

EXECUTIVE SUMMARY

High employee turnover costs companies time, money, and institutional memory. This report analyzes the factors contributing employee attrition at CW Research and Analytics LLC.

CW has experienced 16% employee attrition over the last year, which is four time higher than the 4% industry standard. As a company that relies on high-skilled employees, CW executives worry that the company may lose its competitive position due to significant employee turnover.

The analysis of the data provided by the Chief Data Officer using Logistic Regression model identified a number of factors contributing to employee attrition.

- **The gap between promotions** and a perceived decline in prospects of career advancement is the primary contributor to attrition. **Recommendation:** Employees are given more opportunities for career advancement/development and promotion.

- **Overtime** is the second most significant contributor to attrition. **Recommendation:** minimize overtime.

- Employees working as **lab technicians**, **sales reps,** and **HR staff** are more likely to leave the company. **Recommendation**: Investigate working conditions and employee complaints in these job roles. Low pay can be a contributing factor.

- **Frequent job changes** in an employee's career also contributes to his/her attrition. **Recommendation:** The HR department shall give priority to job applicants who have not changed their jobs frequently in a relatively short period of time in their careers.

This page is intentionally left blank

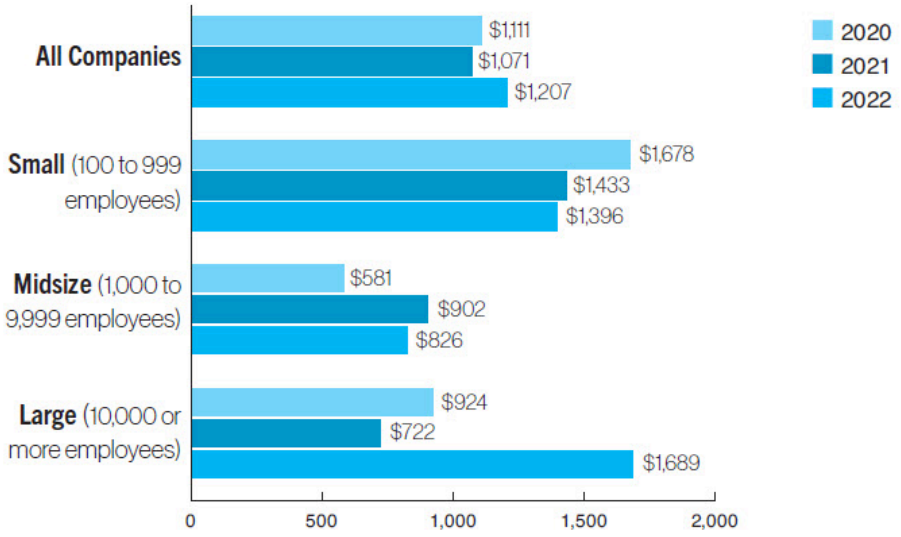
INTRODUCTION

PROBLEM

Employees are the greatest assets and liabilities of organizations. Depending on the industry, employee costs constitute between 50% to 80% of operating expenses of companies (Pynes, 2013, Deloitte, 2017).

Hiring, training, and retaining talent costs companies significant time and money. For example, according to a benchmarking data by Society for Human Resource Management, companies, on average, spend around $4,700 for hiring one employee (Navarra, 2022) while average executive cost per hire is over $28,000 (Miller, 2022). However, these numbers underestimate the true cost of hiring. Edie Goldberg, the founder of the talent management and development company *E.L. Goldberg & Associates* notes that "if you are hiring a job that pays $60,000, you may spend $180,000 or more to fill that role" (Navarro, 2022). Furthermore, according to the Training Magazine, in 2022, companies have spent over $1,200 (**see Figure 1 below**), on average, for employee training per learner (Training Magazine, 2022). Employee turnover means losing the investment made for training employees and covering extra costs of hiring new talent.

In addition to the monetary costs, high employee turnover impacts organizational productivity, damage organizational culture, and undermine institutional memory. Organizations relying on high-skilled employees (e.g., engineers, scientists) are particularly susceptible to attrition. Also, smaller companies and companies that strive to be competitive in the labor market feel the impacts the employee attrition more than large corporations.



Thus, understanding the factors contributing to employee attrition is crucial for minimizing it, and help companies reduce costs and stay competitive.

Figure 1: Training Expenditures per Learner 2020 - 2022
Source: Training Magazine

CONTEXT AND PURPOSE

The purpose of this report is to tackle employee attrition in the fictional company named CW Research & Analytics LLC. CW has been providing research and consulting services in the genetics/medical field for over 20 years. The company employs more than 1,400 high-skilled employees and has been experiencing significant employee turnover (16%) since last year. Considering the industry average of 4% attrition (BLS, 2023), the company executives worry that CW may lose its competitive position in the market due to attrition. The company CEO Brian Coon requested that the HR department find the root causes of the problem. This report is a response to the HR department's demand for analyzing the employee data and finding the factors contributing to employee attrition.

AUDIENCE

The primary audience of this report are the HR Department Director Rick Perry, the CEO Brian Coon, the Chief Data Officer Nicole Rosenbaum, and other C level executives in the company.

Secondary audience and beneficiaries of this report are managers dealing with recruitment, hiring, training, budgeting, strategic planning, and organizational governance in especially in high tech industries.

DATA AND METHOD

The data for this report was shared by Konapure and Uikey on Kaggle (Kaggle, 2023). This single CSV file contains 35 columns and 1,470 rows (i.e. observations). Columns include attrition (yes/no), age, business travel requirements, department, distance to work, education level, etc. Although the number of rows is relatively small for a machine learning project, a dataset of this size is not uncommon in organizational research.

FEATURE PROPERTIES

The dataset consists of 25 *Int64* and 12 *Object* features. However, 11 out of 25 *Int64* features are ordinal type. Also, 2 of the 12 *Object* columns are binary (see Table 1 for details).

Table 1: Feature Properties

Column Name	Data Type	Value Type	Column Name	Data Type	Value Type
Age	Int64	Discrete	MonthlyIncome	Int64	Discrete
Attrition	Object	Binary	MonthlyRate	Int64	Discrete
BusinessTravel	Object	Ordinal	NumCompaniesWorked	Int64	Discrete
DailyRate	Int64	Discrete	Over18	Int64	Discrete
Department	Object	Nominal	OverTime	Int64	Binary
DistanceFromHome	Int64	Ordinal	PercentSalaryHike	Int64	Discrete
Education	Int64	Ordinal	PerformanceRating	Int64	Ordinal
EducationField	Object	Nominal	RelationshipSatisfaction	Int64	Ordinal
EmployeeCount	Int64	Discrete	StandardHours	Int64	Discrete
EmployeeNumber	Int64	Discrete	StockOptionLevel	Int64	Ordinal
EnvironmentSatisfaction	Int64	Ordinal	TotalWorkingYears	Int64	Discrete
Gender	Object	Binary	TrainingTimesLastYear	Int64	Discrete
HourlyRate	Int64	Discrete	WorkLifeBalance	Int64	Ordinal
JobInvolvement	Int64	Ordinal	YearsAtCompany	Int64	Discrete
JobLevel	Int64	Ordinal	YearsInCurrentRole	Int64	Discrete
JobRole	Object	Nominal	YearsSinceLastPromotion	Int64	Discrete
JobSatisfaction	Int64	Ordinal	YearsWithCurrManager	Int64	Discrete
MaritalStatus	Object	Nominal			

DATA WRANGLING

The dataset was free of missing values and duplicate rows. *EmployeeCount*, *Over18*, and *StandardHours* columns have constant values. Also, *EmployeeNumber* is just an ID for each employee and has no value for analytical purposes. Thus, these four columns were dropped from the dataset.

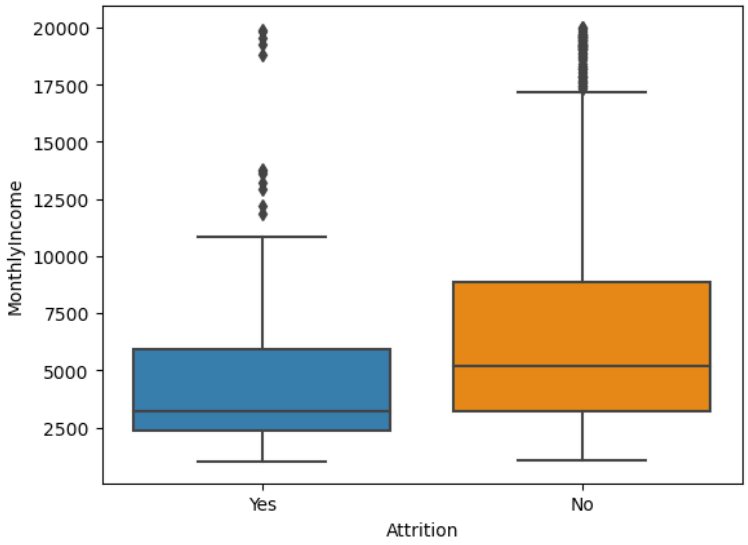
EXPLORATORY DATA ANALYSIS

MONTHLY INCOME AND ATTRITION

Compensation (i.e., salary), is one of the most important factors contributing to employee turnover. Exploratory data analysis results show that employees who leave the company tend to earn less than who don't (see **Figure 1**).

The median monthly income for *Attrition* cases is around $3,000 while the median monthly income for *No Attrition* cases is over $5,000. However, as shown in Figure 1, high earners also leave the company as well. This suggests that other factors contribute to turnover. Thus, further investigation of association between attrition and other is necessary.

Figure 1: Boxplot visualizing the MonthlyIncome distribution across Attrition and No Attrition cases



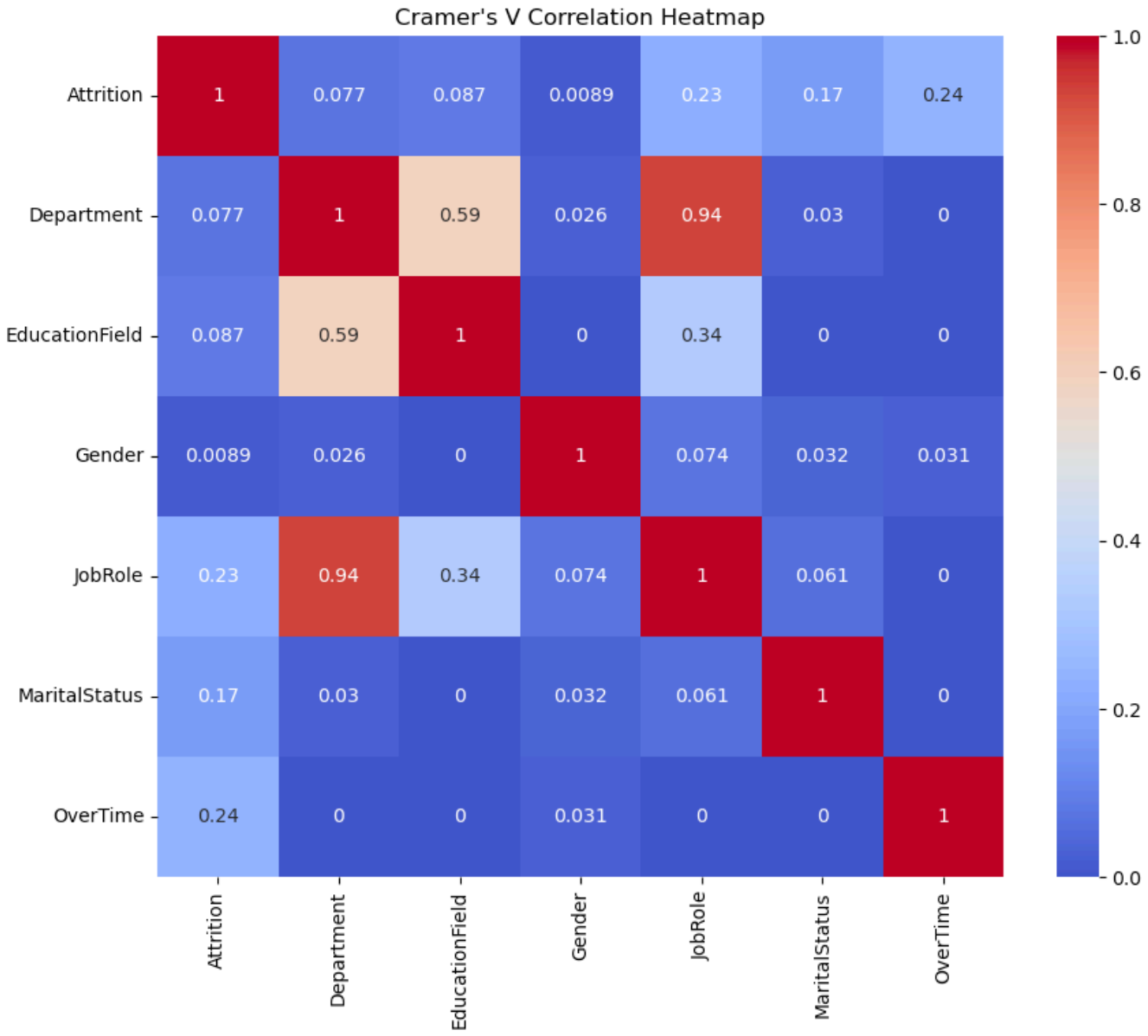
To further investigate the associations between features and their relationship to **Attrition**, separate correlation tests were conducted for nominal, ordinal, and numeric features.

CORRELATION HEATMAPS

Nominal Features

Figure 2 shows the Cramer's V correlation coefficients for nominal features. **Attrition** is weakly correlated with **JobRole** and **Overtime** features (0.23 and 0.24 respectively). The associations between **Attrition** and other features are not noticeable.

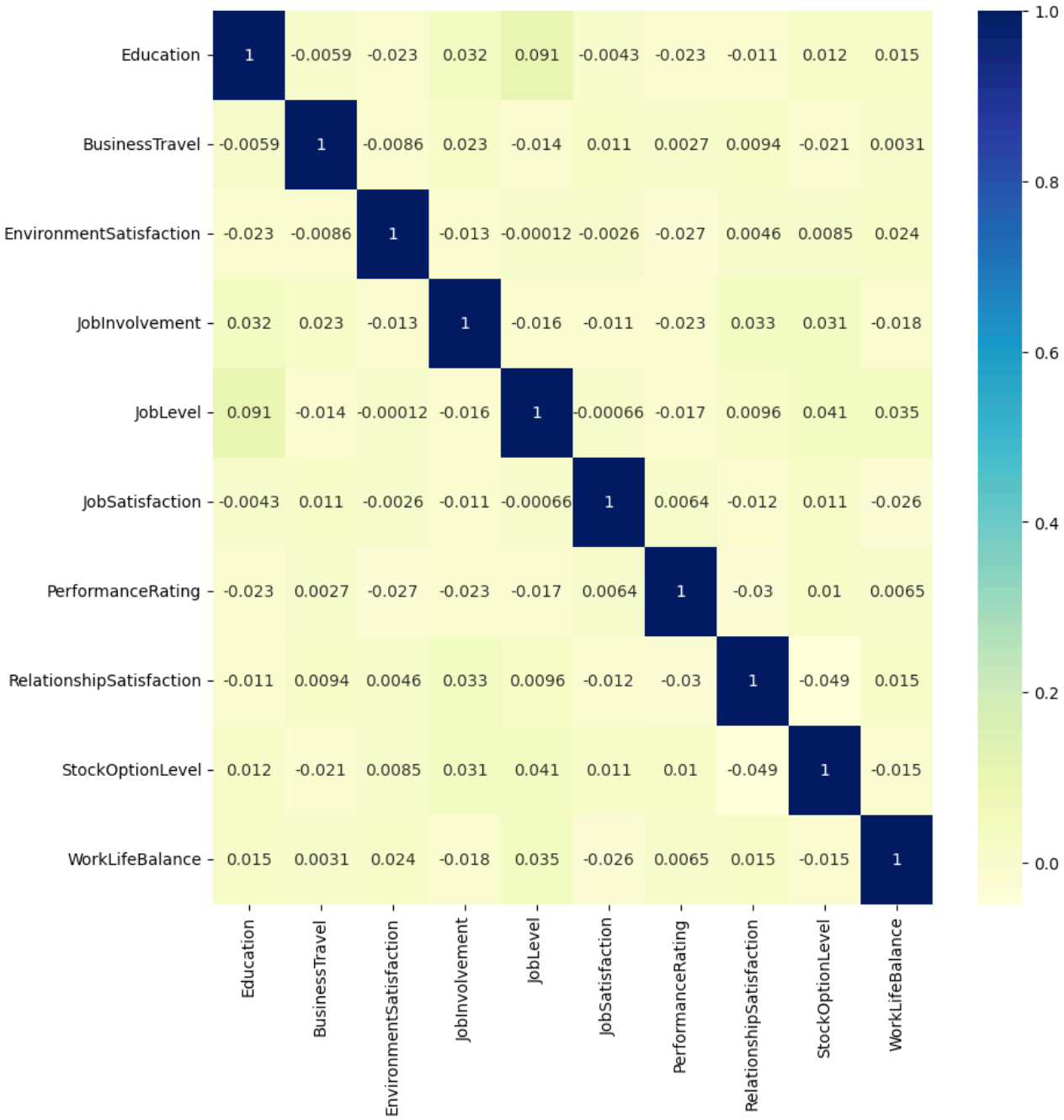
Figure 2: Correlation Heatmap for Categorical Features (Cramer's V)



Ordinal Features

Figure 3 visualizes the associations between ordinal features using Kendal's Tau. No obvious association was detected.

Figure 3: Correlation Heatmap for Ordinal Features (Kendall's Tau)

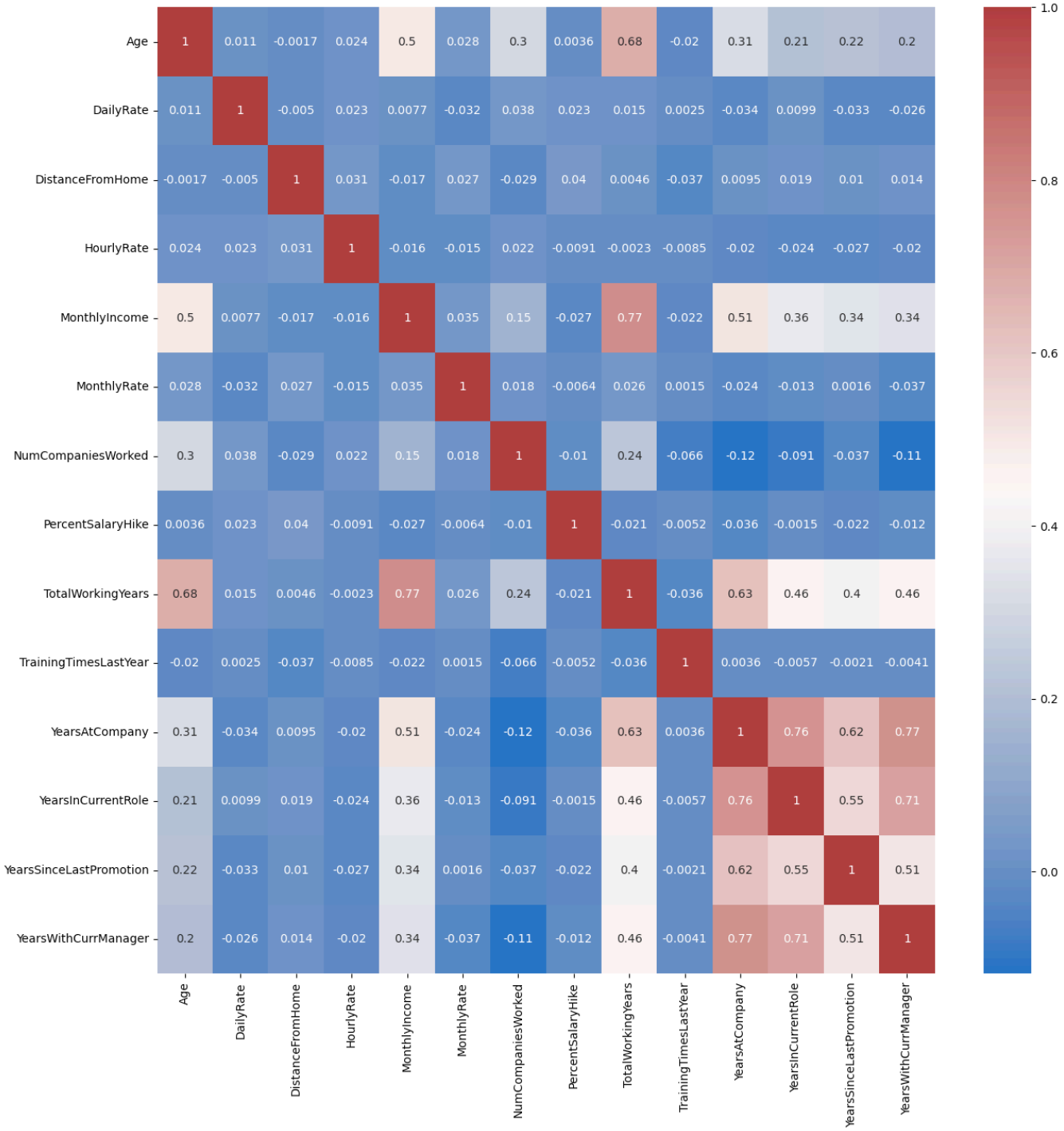


Numeric Features

Figure 4 visualizes the associations between discrete numeric features using Pearson's r correlation coefficient. The results show moderate association between **Age - TotalWorkingYears,** and **MonyhlyIncome-TotalWorkingYears**. Also, features about years (i.e., **YearsAtCompany, YearsInCurrentRole, YearsSinceLastPromotion, YearsWithCurrentManager, TotalWorkingYears)** were also moderately associated with each other.

Exploratory data analysis did not identify any major patters in the data, except for the difference in MonthlyIncome of *Attrition* and *No Attrition* cases.

Figure 4: Correlation Heatmap for Discrete Numeric Features (Pearson's r)



PREPROCESSING

In this step, the data was split into three separate dataframes according to the feature value types. These were **nominal features** (e.g., gender, department), **ordinal features** (e.g., job satisfaction), and **numeric features** (e.g., monthly income).

Next, each data frame were prepared for the modeling step:

- nominal features were converted into dummy variables using Pandas,
- ordinal feature values were mapped and each observation was replaced with an appropriate numeric value
- Numeric features were scaled and standardized using MinMaxScaler.

Finally, these three dataframes were combined into a final dataframe for modeling step.

MODELING

In this step, five different classification models were used for predicting **Attrition**. **Table 2** displays the *Accuracy*, *Precision*, *Recall*, and *F1-Score* for each model used. Because the goal is to minimize attrition, we are more tolerant to False Positive results than False Negatives.

Logistic regression performed the best among the five models. Decision Tree and Random Forest models were the worst performers. SVC and XG Boost did better than either tree models but they were behind the logistic regression.

The logistic regression model predicted 22 attrition cases correctly and missed 39 of them, which is significantly better than other models. It also identified 13 false positive cases.

Table 2: Accuracy, Precision, Recall, and F1- Score Results for SVC, Decision Tree, Random Forest, XG Boost, and Logistic Regression Models

		Models				
METRIC		**SVC**	**Decision Tree**	**Random Forest**	**XG Boost**	**Logistic Regression**
Accuracy		0.89	0.83	0.87	0.88	0.88
Precision	No Attrition	0.89	0.89	0.87	0.88	0.9
	Attrition	0.77	0.33	0.64	0.71	0.63
Recall	No Attrition	0.99	0.92	0.99	0.99	0.97
	Attrition	0.28	0.26	0.11	0.2	0.36
F1-Score	No Attrition	0.94	0.9	0.93	0.93	0.93
	Attrition	0.41	0.29	0.19	0.31	0.46
False Positive		5	32	4	8	13
False Negative		44	45	54	45	39
True Negative		375	348	376	372	367
True Positive		17	16	7	16	22

Table 3 shows the features that contribute to Attrition. The top three contributors to attrition in CW are **YearsSinceLastPromotion**, **NumCompaniesWorked**, and **OverTime_Yes**. These features increase the odds of attrition by 380%, 368%, and 266% respectively. Also, **JobRole**, **MaritalStats**, BusinessTravel, and **EducationField** contribute to attrition at different levels.

The results show that attrition is higher among Laboratory Technicians and Sales Representatives. Single employees are more likely to leave the company. Frequent business travel increases the odds of attrition. Finally, employees with technical degrees are more likely to leave company.

Table 3: Features in the Logistic Regression Models that Increase the Odds of Attrition

Feature	Coefficient	% Increase in Odds of Attrition
YearsSinceLastPromotion	1.569122	380.24
NumCompaniesWorked	1.543734	368.20
OverTime_Yes	1.299739	266.83
MaritalStatus_Single	1.064923	190.06
JobRole_Laboratory Technician	0.947741	157.99
BusinessTravel	0.855795	135.32
JobRole_Sales Representative	0.80235	123.08
EducationField_Technical Degree	0.758158	113.43
JobRole_Human Resources	0.566221	76.16
Gender_Male	0.560651	75.18
EducationField_Human Resources	0.501213	65.07
Department_Sales	0.480375	61.67
YearsAtCompany	0.388956	47.54

Table 4 shows the features that decrease attrition in CW. **YearsInCurrentRole**, **TotalWorkingYears**, **Age**, **JobRole**, **YearsWithCurrManager**, and **Overtime** decrease the odds of attrition.

The results show that the longer the employees stay in a role and work with the same manager, it is less likely that they will leave the company.

Age is an important factor in attrition. As the age of an employee increases by one year his/her odds of leaving the company decreases by 68%.

The odds of attrition for employees at Healthcare Representative and Research Director roles are significantly lower than other job roles (56% and 54% respectively).

Finally, eliminating Overtime, providing training to employees, increasing job involvement, and increasing environment satisfaction decrease attrition.

Table 4: Features in the Logistic Regression Models that Decrease the Odds of Attrition

Feature	Coefficient	% Decrease in Odds of Attrition
YearsInCurrentRole	-1.49803	-77.64
TotalWorkingYears	-1.34368	-73.91
Age	-1.156059	-68.53
JobRole_Healthcare Representative	-0.84335	-56.97
JobRole_Research Director	-0.793108	-54.76
YearsWithCurrManager	-0.752833	-52.9
OverTime_No	-0.609669	-45.65
TrainingTimesLastYear	-0.59165	-44.66
JobInvolvement	-0.464995	-37.19
MaritalStatus_Divorced	-0.423062	-34.5
EnvironmentSatisfaction	-0.3512	-29.62
EducationField_Life Sciences	-0.349265	-29.48

CONCLUSION AND RECOMMENDATIONS

This report is prepare in response to significant employee turnover at CW Research and Analytics LLC. Based on the data provided, the best results were obtained from Logistic Regression model.

The findings from the Logistic Regression model identified a number of factors contributing to attrition.

1. **Year gap between promotions**: The findings suggest that as the time since the last promotion increases the odds of employee attrition increases significantly. Employees might perceive the gap between their previous promotion and the next promotion as an impediment for their advancement in the company and may want to increase their career development in other companies.

   Recommendation: Thus, it is recommended CW invests in employee training and development programs and find ways to promote employees or show other paths for career advancement.

2. **Overtime:** Overtime is a significant contributor to attrition. When an employee is asked to overtime work, his/her odds of leaving the company increases by 266%.

   Recommendation: Avoid requiring employees overtime.

3. **Job Roles**: The odds of Lab Technicians, Sales Representatives, and Human Resources staff leaving the company are significantly higher than other employees.

   Recommendation: Further investigations are needed to understand why Lab Technicians and Sales Reps, and HR staff leave the company at a faster rate. Burnout might be a factor to explore. Another cause of attrition in these three roles might be the low pay. The gap between median monthly income of *Attrition* and *No Attrition* cases is noticeable and this gap may be caused by the low salaries of these three job roles.

4. **Number of Companies Worked:** Employees who worked in multiple companies throughout their careers are more likely to leave the company. Thus, it is recommended that the HR department pay attention to career trajectories of job applicants during hiring process. Employees who have changed multiple companies in a relatively short period of time in their careers are more likely to leave CW. However, there are differences between millennials, Gen X, and Gen Z employees with respect to their company loyalty. Younger generations are tend to stay in a company a lot shorter period than older generations. Millennials, for example, change their jobs approximately every 2 years and this rate might be higher for Gen Z who has started joining the labor force recently. The HR department and the CW executives shall consider structural changes to the company to adapt to the characteristics of younger generation of employees.

REFERENCES

Deloitte. 2017. LaborWise. Accessed on 08/12/2023 via https://www2.deloitte.com/content/dam/Deloitte/us/Documents/human-capital/us-cons-laborwise-core-data-sheet.pdf

Navarra, Katie. 2022. The Real Costs of Recruitment. Accessed on 8/12/2023 via https://www.shrm.org/resourcesandtools/hr-topics/talent-acquisition/pages/the-real-costs-of-recruitment.aspx

Kaggle. 2023. HR Analytics. Accessed on 8/12/2023 via https://www.kaggle.com/datasets/rishikeshkonapure/hr-analytics-prediction

Miller, Stephen. 2022. SHRM HR Benchmarking Report Launch as a Free Member-Exclusive Benefit. Accessed on 08/12/2023 via https://www.shrm.org/resourcesandtools/hr-topics/benefits/pages/shrm-hr-benchmarking-reports-launch-as-a-member-exclusive-benefit.aspx

Pynes, Joan E. 2013. *Human Resources Management for Public and Nonprofit Organizations: A Strategic Approach (4th Edition)*. Jossey-Bass. San Francisco: CA.

Training Magazine. 2022. 2022 Training Industry Report. Accessed on 8/12/2023 via https://trainingmag.com/2022-training-industry-report/