

Examine Factors Affecting Employees' Attrition Using Logistic Regression

Benner Feng, Elizabeth Tong, Samita Prabhasavat, Zoie So

Apr 2, 2023

Abstract

This study examines the factors influencing employee attrition rates within an organization using a logistic regression model. Analyzing various employee-related variables, our findings highlight that working overtime is the lead factor that contributes to high attrition rates. The results underscore the importance of work-life balance, growth opportunities, and strong manager-employee relationships in reducing attrition rates and ensuring the organization's long-term success.

Introduction

Human resource management plays a crucial role in an organization's success. Consequently, comprehending the factors that contribute to employee attrition is imperative for effective talent management. This approach enables an organization to save costs on recruitment and training in the short run, while also identifying potential internal problems such as management structure and corporate culture that may need to be addressed in the long term. According to Chakraverty (2006), the costs of employee attrition, including the direct costs of recruitment, training, and lost productivity, as well as the indirect costs of decreased morale, reduced team cohesion, and knowledge loss. In light of these costs, organizations must take a strategic approach to address employee attrition and retain top talent. Therefore, proactive measures are necessary for organizations to maintain a competitive edge and ensure long-term success.

To achieve this goal, this project aims to identify the primary factors that drive employee attrition, which will enable us to predict whether a future employee is at risk of leaving the organization. Our current hypothesis suggests that an employee's age, marital status, years of experience in the organization, number of organizations one has worked for, and job satisfaction may influence their attrition. For instance, we believe that there may be a negative correlation between an employee's age, years of experience in the organization, job satisfaction, and likelihood of leaving. This is because older employees tend to settle into familiar work routines, and the cost of switching jobs is typically higher for them. The same principle applies to married individuals with children who require a stable income to

support their families. If an employee is content with their job and work environment, they are less likely to resign. Conversely, those who have worked for several organizations may have a higher likelihood of leaving the current organization since they have more experience switching between different work environments.

The dataset that we have selected can be found on Kaggle (Appendix A). The CSV file contains information on 2,940 employees. It includes 34 variables, including both categorical and numerical variables, such as demographic information (age, education, and marital status), job-related metrics (the total number of years worked in the current company, the number of companies worked for, and job satisfaction), and an attrition flag. The dataset provides comprehensive information that can be used to determine the primary factors that drive employee attrition.

Methodology

This project consists of five main steps. The first step is the general setup, which includes the installation and loading of packages, importing data, exploring data, and cleaning data. In the installation and loading of packages, 12 packages were used in this project. These packages were used for different purposes, such as manipulating, visualizing, and building models. After installing the necessary packages, the data was imported and explored. This allowed the data to be cleaned by removing unnecessary variables that would not add any value to the analysis.

The second step is exploratory data analysis (EDA). This involved the preparation of data, including separating numerical and categorical variables. The univariate analysis was then conducted, which investigated the distribution of each variable. This was done using two methods: summary statistics and histograms for numerical variables, and observing the percentages of each subcategory for categorical variables. Bivariate analysis was also conducted to investigate the relationships between each pair of variables. This was done using different methods, including scatter plot and correlation matrix.

The third step is model building. Before building a model using logistic regression, the data needed to be prepared to fit the model's requirements. This involved converting categorical variables to dummy variables, scaling data using z-score, separating the dependent variable from independent variables, and splitting the data into training and testing sets. Logistic regression was then used to build the model, and the performance of the model was evaluated using different metrics such as accuracy, sensitivity, and specificity.

The fourth step is model tuning. This step involved finding the optimal hyperparameters to improve the performance of the model. This was done using different methods such as grid search, random search, and Bayesian optimization.

The fifth and final step is the model interpretation. This involved interpreting the coefficients of the model and making predictions on new data. The coefficients were interpreted using odds ratios and p-values to determine the significance of each variable. The model was then used to make predictions on new data to determine if an employee is at risk of leaving the company or not.

Results

1. Exploratory Data Analysis

As mentioned in the methodology section, several methods were used in the exploratory data analysis. Statistics summary was used to observe the distribution of numerical variables included in the analysis. According to Table 2, the average employee age is around 37 years. The variable ‘Age’ has a wide range, from 18 years to 60, indicating that there is diversity in age within the organization. At least 50% of the employees live within a 7 km radius from the organization. However, there are some extreme values, given that the maximum value is 29 km. The average monthly income of an employee is 6500 USD. It has a wide range of values from 1,000 to 20,000 USD, which is to be expected for any organization's income distribution. There is a big difference between the 3rd quartile value (around 8,400 USD) and the maximum value (nearly 20,000 USD), showing that the company's highest earners have a disproportionately large income in comparison to the rest of the employees. Again, this is fairly common in most organizations. The average salary hike of an employee is around 15%. At least 50% of employees got a salary hike of 14% or less, with the maximum salary hike being 25%. The average number of years an employee is associated with the company is 7. On average, the number of years since an employee got a promotion is around 2.19. The majority of employees have been promoted since last year.

Table 2: Statistics summary of all numerical variables included in the analysis (Appendix B)

Variables	Minimum	1 st Quartile	Median	Mean	3 rd Quartile	Maximum
-----------	---------	--------------------------	--------	------	--------------------------	---------

DailyRate	102.0	465.0	802.0	802.5	1157.0	1499.0
Age	18.00	30.00	36.00	36.92	43.00	60.00
DistanceFromHome	1.000	2.000	7.000	9.193	14.000	29.000
MonthlyIncome	1009	2911	4919	6503	8380	19999
MonthlyRate	2094	8045	14236	14313	20462	26999
PercentSalaryHike	11.00	12.00	14.00	15.21	18.00	25.00
TotalWorkingYears	0.00	6.00	10.00	11.28	15.00	40.00
YearsAtCompany	0.000	3.000	5.000	7.008	9.000	40.000
NumCompaniesWorked	0.000	1.000	2.000	2.693	4.000	9.000
HourlyRate	30.00	48.00	66.00	65.89	84.00	100.00
YearsInCurrentRole	0.000	2.000	3.000	4.229	7.000	18.000

YearsSinceLastPromotion	0.000	0.000	1.000	2.188	3.000	15.000
YearsWithCurrManager	0.000	2.000	3.000	4.123	7.000	17.000
TrainingTimesLastYear	0.000	2.000	3.000	2.799	3.000	6.000

Figure 1 below is an example of a histogram that was plotted to examine the distribution of each numerical variable. According to the plotted histograms (Appendix C), the age distribution is close to a normal distribution with the majority of employees between the ages of 25 and 50. The percentage salary hike is skewed to the right, implying that employees are obtaining smaller increases. MonthlyIncome and TotalWorkingYears are skewed to the right, indicating that the majority of workers are in entry or mid-level positions in the organization. DistanceFromHome also has a right skewed distribution, meaning most employees live close to work but there are a few that live further away. On average, an employee has worked at 2 to 3 companies. Most employees have worked at only 1 company. The YearsAtCompany variable distribution shows a good proportion of workers with 10+ years, indicating a significant number of loyal employees at the organization. The YearsInCurrentRole distribution has three peaks at 0, 2, and 7. There are a few employees that have even stayed in the same role for 15 years and more. The YearsSinceLastPromotion variable distribution indicates that some employees have not received a promotion in 10 to 15 years and are still working in the organization. These employees are assumed to be high work-experience employees in upper-management roles, such as co-founders or C-suite employees. The distributions of DailyRate, HourlyRate and MonthlyRate appear to be uniform and do not provide much information. It could be that daily rate refers to the income earned per extra day worked while hourly rate could refer to the same concept applied for extra hours worked per day. Since these rates tend to be broadly similar for multiple employees in the same department, that explains the uniform distribution they show.

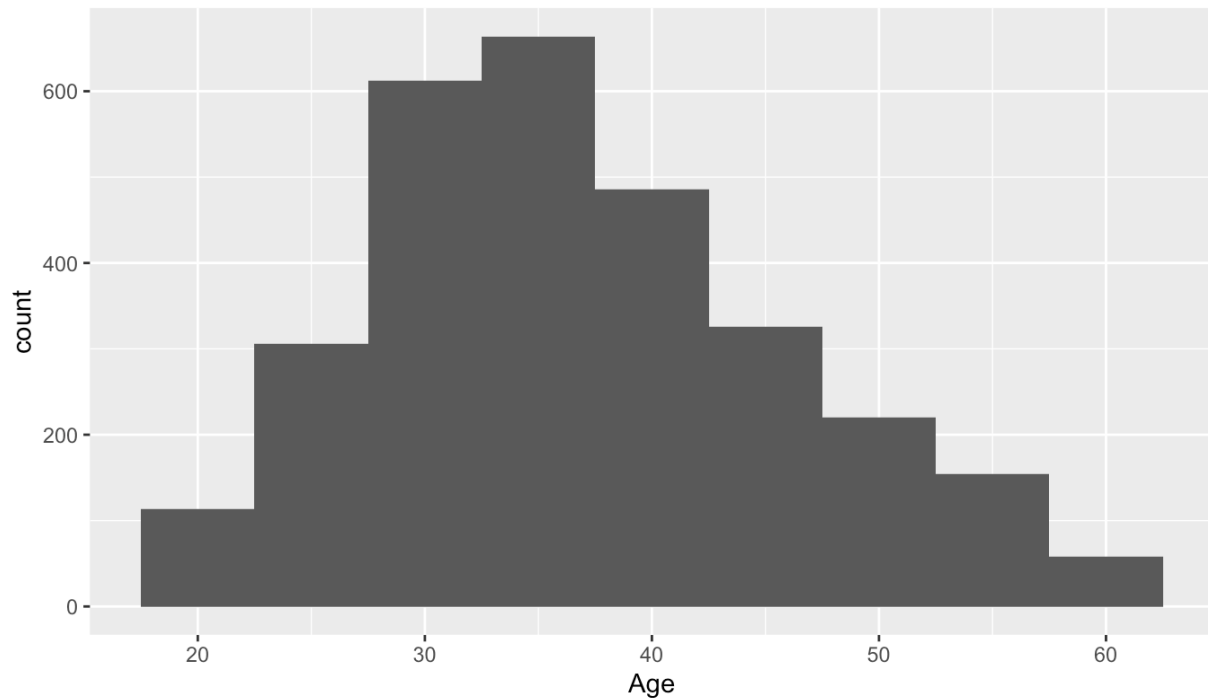


Figure 1: A histogram illustrating the distribution of a numerical independent variable 'Age'

Moving on to the univariate analysis for categorical variables (Appendix D), according to the computed percentage of subcategories in each categorical variable, the employee attrition rate is 16%. Around 28% of the employees are working overtime. This number appears to be on the higher side, and might indicate a stressed employee work-life. Around 71% of the employees have traveled rarely, while around 19% have to travel frequently. Around 73% of the employees come from an educational background in the Life Sciences and Medical fields. Over 65% of employees work in the Research & Development department of the organization. Nearly 40% of the employees have low (1) or medium-low (2) job satisfaction and environment satisfaction in the organization, indicating that the morale of the company appears to be somewhat low. Over 30% of the employees show low (1) to medium-low (2) job involvement. Over 80% of the employees either have none or very less stock options. In terms of performance ratings, none of the employees have rated lower than 3 (excellent). About 85% of employees have a performance rating equal to 3 (excellent), while the remaining have a rating of 4 (outstanding). This could either mean that the majority of employees are top performers, or the more likely scenario is that the organization could be highly lenient with its performance appraisal process.

According to Figure 2 below, The total work experience, monthly income, years at company and years with the current manager are highly correlated with each other and with employee age which is

easy to understand as these variables show an increase with age for most employees. Years at company and years in current role are correlated with years since last promotion which means that the company is not giving promotions at the right time.

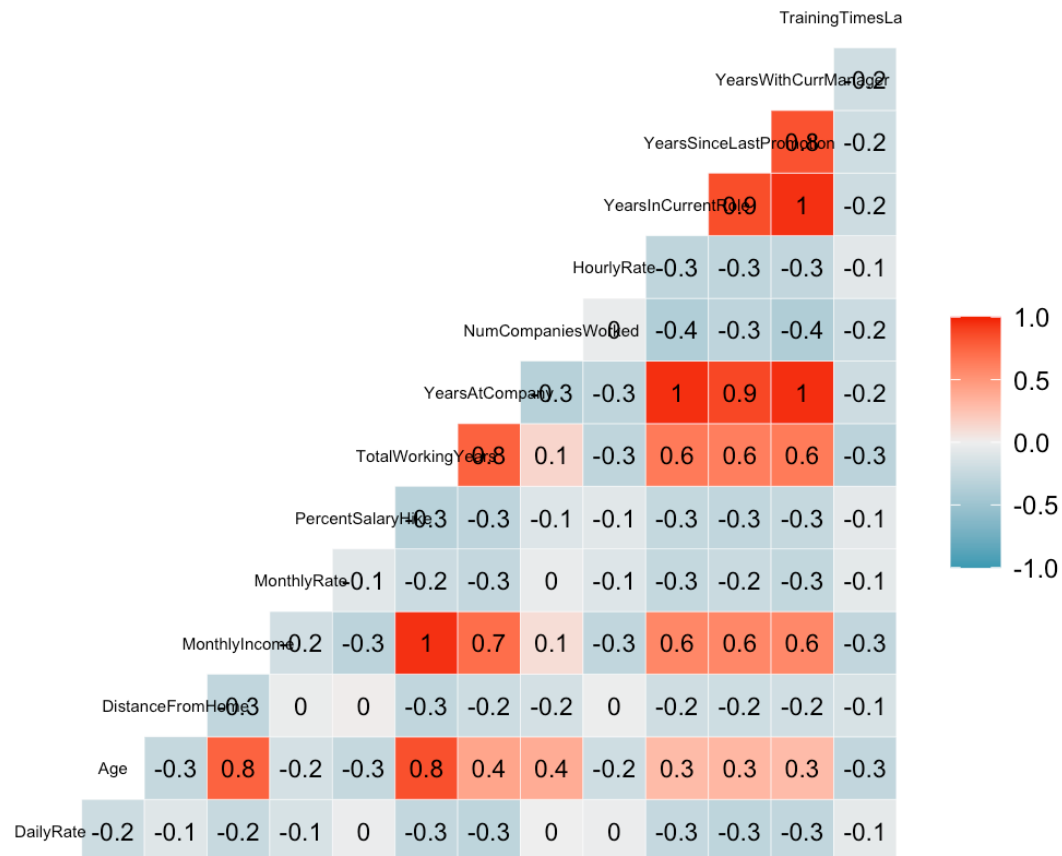


Figure 2: A correlation matrix illustrating correlation coefficients between each pair of numerical variables

Other methods such as bar graphs and scatter plots were also used in the exploratory data analysis. However, they do not provide much information about distributions of the variables included in the analysis.

2. Model Building

After the model was built, a part of the summary of the model was shown as in Figure 3 below (Appendix E). The exponentiated estimates can be seen in Figure 4 (Appendix F). Both will be discussed in the discussion section.

```
Call:
glm(formula = y_train ~ ., family = binomial(link = "logit"),
     data = X_train_scaled)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6831  -0.4860  -0.2516  -0.0843   3.3290

Coefficients: (1 not defined because of singularities)
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -3.368056    4.031600  -0.835  0.403485
Age            -0.432033    0.108010  -4.000  6.34e-05 ***
DailyRate     -0.122701    0.075200  -1.632  0.102750
DistanceFromHome  0.388033    0.075202   5.160  2.47e-07 ***
HourlyRate     0.023001    0.076758   0.300  0.764437
MonthlyIncome -0.016583    0.327029  -0.051  0.959557
MonthlyRate    0.036539    0.075359   0.485  0.627767
NumCompaniesWorked 0.477332    0.082971   5.753  8.77e-09 ***
OverTimeYes     1.931027    0.165343  11.679 < 2e-16 ***
PercentSalaryHike -0.038398    0.121411  -0.316  0.751804
TotalWorkingYears -0.512297    0.197597  -2.593  0.009524 **
TrainingTimesLastYear -0.222058    0.079169  -2.805  0.005034 **
YearsAtCompany   0.796277    0.199877   3.984  6.78e-05 ***
YearsInCurrentRole -0.631225    0.143446  -4.400  1.08e-05 ***
YearsSinceLastPromotion 0.566715    0.112301   5.046  4.50e-07 ***
YearsWithCurrManager -0.477905    0.135763  -3.520  0.000431 ***
BusinessTravelNon.Travel -0.367945    0.098642  -3.730  0.000191 ***
BusinessTravelTravel_Frequently 0.301594    0.071317   4.229  2.35e-05 ***
BusinessTravelTravel_Rarely      NA         NA      NA      NA
DepartmentResearch...Development 6.656993  232.961924   0.029  0.977203
DepartmentSales      6.492504  225.530106   0.029  0.977034
Education            0.004104    0.076077   0.054  0.956979
EducationFieldLife.Sciences -0.262694    0.343284  -0.765  0.444129
EducationFieldMarketing -0.003352    0.225883  -0.015  0.988161
EducationFieldMedical -0.313864    0.325019  -0.966  0.334204
EducationFieldOther -0.060677    0.169342  -0.358  0.720111
EducationFieldTechnical.Degree 0.134672    0.198733   0.678  0.497992
JobSatisfaction    -0.461321    0.076082  -6.063  1.33e-09 ***
EnvironmentSatisfaction -0.482249    0.077663  -6.210  5.31e-10 ***
WorkLifeBalance    -0.279051    0.072950  -3.825  0.000131 ***
StockOptionLevel   -0.037382    0.114785  -0.326  0.744676
GenderMale         0.137409    0.076778   1.790  0.073506 .
PerformanceRating   0.050329    0.119922   0.420  0.674715
JobInvolvement     -0.357067    0.072909  -4.897  9.71e-07 ***
```

Figure 3: A part of summary of the logistic regression model built for this analysis

(Intercept)	Age
0.03445656	0.64918790
DailyRate	DistanceFromHome
0.88452773	1.47407863
HourlyRate	MonthlyIncome
1.02326777	0.98355325
MonthlyRate	NumCompaniesWorked
1.03721525	1.61176919
OverTimeYes	PercentSalaryHike
6.89658767	0.96233031
TotalWorkingYears	TrainingTimesLastYear
0.59911782	0.80086894
YearsAtCompany	YearsInCurrentRole
2.21727149	0.53193966
YearsSinceLastPromotion	YearsWithCurrManager
1.76246778	0.62008131
BusinessTravelNon.Travel	BusinessTravelTravel_Frequently
0.69215515	1.35201153
BusinessTravelTravel_Rarely	DepartmentResearch...Development
NA	778.20696019
DepartmentSales	Education
660.17421691	1.00411236
EducationFieldLife.Sciences	EducationFieldMarketing
0.76897710	0.99665381
EducationFieldMedical	EducationFieldOther
0.73061837	0.94112713
EducationFieldTechnical.Degree	JobSatisfaction
1.14416163	0.63045052
EnvironmentSatisfaction	WorkLifeBalance
0.61739311	0.75650161
StockOptionLevel	GenderMale
0.96330825	1.14729687
PerformanceRating	JobInvolvement
1.05161740	0.69972575
JobLevel	JobRoleHuman.Resources
0.95740910	16.14688453
JobRoleLaboratory.Technician	JobRoleManager
1.71150393	1.02825799
JobRoleManufacturing.Director	JobRoleResearch.Director
1.02661242	0.68159430
JobRoleResearch.Scientist	JobRoleSales.Executive
1.10075026	1.33032893
JobRoleSales.Representative	MaritalStatusMarried
1.46365644	1.20797284

Figure 4: A part of the list of exponentiated estimates from logistic regression model

Discussion

According to the model summary, significant variables can be identified as in Table 3 below.

Table 3: Significant variables determined using p-values

Positive Relationship (Increased Attrition)	Negative Relationship (Decreased Attrition)
<ol style="list-style-type: none">1. DistanceFromHome2. NumCompaniesWorked3. OverTimeYes4. YearsAtCompany5. YearsSinceLastPromotion6. BusinessTravelTravel_Frequently7. JobRoleLaboratory.Technician8. MaritalStatusSingle	<ol style="list-style-type: none">1. Age2. TotalWorkingYears3. TrainingTimesLastYear4. YearsInCurrentRole5. YearsWithCurrManager6. BusinessTravelNon.Travel7. JobSatisfaction8. EnvironmentSatisfaction9. WorkLifeBalance10. JobInvolvement11. RelationshipSatisfaction

Based on the exponentiated estimates, it is clear that certain factors increase the odds of attrition, while others decrease it. Factors such as age, total working years, job satisfaction, and environment satisfaction have a negative relationship with attrition, implying that employees with more experience, higher job satisfaction, and a more favorable work environment are less likely to leave the company. This highlights the importance of creating a supportive work environment and offering opportunities for professional growth to retain experienced employees.

On the other hand, factors like distance from home and working overtime have a positive relationship with attrition, suggesting that employees living further away from their workplace and working overtime are more likely to leave the company. This highlights the need for organizations to consider flexible work arrangements, such as remote work or flexible hours, to accommodate employees' diverse needs and reduce attrition rates.

According to the right side of the table, more than half of the factors that decrease the odds of attrition are time-relevant, likely because they are related to an employee's growth, stability, and satisfaction within the organization. As employees gain more experience, skills, and knowledge over time, they often become more committed to their jobs and have a stronger connection to the company. These time-relevant factors that decrease attrition indicate that fostering a supportive work environment that focuses on employee growth and satisfaction can help retain experienced employees and reduce attrition rates.

It is important to note that while some factors related to time can reduce employee attrition, staying at the same company for many years seems to increase it. This unexpected result can be explained by a few reasons. When employees work at the same place for a long time, they might start feeling stuck or bored in their jobs. If they don't have chances for growth, learning new skills, or career advancement, they might become unhappy and start looking for other jobs. To help with this, employers should create ways for employees to openly talk about their challenges at work. Letting workers share their worries, frustrations, and goals with management can help bosses figure out how to make things better and offer new chances for growth. Additionally, giving employees access to training programs, mentors, or new projects can help them stay interested and motivated in their jobs, even if they've been with the company for a long time.

In the logistic regression model predicting employee attrition, several factors were found to have significant effects on the odds of an employee leaving the company. The exponentiated estimates represent the multiplicative effect of each factor on the odds of attrition, holding all other variables constant. In our discussion, we will focus on the significant factors and their relationships with attrition as outlined in Table 4 below:

Table 4: Relationship between each independent variable and the dependent variable according to the exponentiated estimates

Positive Relationship (Increased Attrition)	Negative Relationship (Decreased Attrition)
<ol style="list-style-type: none"> 1. OverTimeYes (6.90): Employees who work overtime are 6.90 times more likely to leave the company compared to those who do not work overtime, with all other independent variables held constant. 2. YearsAtCompany (2.22): A one-unit increase in years at the company is associated with a 2.22 times higher odds of employee attrition, with all other independent variables held constant. 3. MaritalStatusSingle (1.91): Single employees are 1.91 times more likely to leave the company compared to their non-single counterparts, with all other independent variables held constant. 4. YearsSinceLastPromotion (1.76): A one-unit increase in years since the last promotion is associated with a 1.76 times higher odds of attrition, with all other 	<ol style="list-style-type: none"> 1. YearsInCurrentRole (0.53): A one-unit increase in years in the current role is associated with a 47% decrease in the odds of attrition, with all other independent variables held constant. 2. TotalWorkingYears (0.60): A one-unit increase in total working years is associated with a 40% decrease in the odds of attrition, with all other independent variables held constant. 3. YearsWithCurrManager (0.62): A one-unit increase in years with the current manager is associated with a 38% decrease in the odds of attrition, with all other independent variables held constant. 4. EnvironmentSatisfaction (0.62): A one-unit increase in environment satisfaction is associated with a 38%

<p>independent variables held constant.</p> <p>5. JobRoleLaboratory.Technician (1.71): Laboratory technicians are 1.71 times more likely to leave the company compared to employees in the reference job role, with all other independent variables held constant.</p> <p>6. NumCompaniesWorked (1.61): A one-unit increase in the number of companies an employee has worked for is associated with a 1.61 times higher odds of attrition, with all other independent variables held constant.</p> <p>7. DistanceFromHome (1.47): A one-unit increase in the distance from home is associated with a 1.47 times higher odds of employee attrition, with all other independent variables held constant.</p> <p>8. BusinessTravelTravel_Frequently (1.35): Employees who travel frequently for business are 1.35 times more likely to leave the company compared to those who do not travel for business, with all other independent variables held constant.</p>	<p>decrease in the odds of attrition, with all other independent variables held constant.</p> <p>5. JobSatisfaction (0.63): A one-unit increase in job satisfaction is associated with a 37% decrease in the odds of attrition, with all other independent variables held constant.</p> <p>6. Age (0.65): A one-unit increase in age is associated with a 35% decrease in the odds of attrition, with all other independent variables held constant.</p> <p>7. BusinessTravelNon.Travel (0.69): Employees who do not travel for business are 31% less likely to leave the company compared to those who travel for business, with all other independent variables held constant.</p> <p>8. WorkLifeBalance (0.76): A one-unit increase in work-life balance is associated with a 24% decrease in the odds of attrition, with all other independent variables held constant.</p> <p>9. RelationshipSatisfaction (0.77): A one-unit increase in relationship satisfaction is associated with a 23% decrease in the odds of attrition, with all other independent variables held constant.</p> <p>10. TrainingTimesLastYear (0.80): A one-unit increase in training times last year is associated with a 20% decrease in the odds of attrition, with all other independent variables held constant.</p> <p>11. DailyRate (0.88): A one-unit increase in daily rate is associated with a 12% decrease in the odds of attrition, with all other independent variables held constant.</p>
--	--

According to Table 4, factors that increase the odds of attrition include working overtime, years being in the company, and being single are the top 3 factors that increase employees' attrition. Among these, working overtime has the most substantial influence, making employees 6.90 times more likely to leave the company. This could be due to increased job-related stress and work-life imbalance, which

might motivate employees to seek alternative employment opportunities. The company could address the overtime issue through strategies such as hiring additional staff, improving task delegation, or implementing flexible work arrangements. Reducing workload on current employees would also help increase their job satisfaction, which has a positive impact on reducing attrition. On the other hand, top 3 factors that decrease the odds of attrition include years spent in current role, total working years, and years spent with current manager, which suggests that stability, experience, and a strong relationship with one's manager contribute to lower attrition rates.

It is essential to recognize the limitations of this study, including potential confounding factors, interactions between variables, and non-linear relationships that might not be captured in this analysis. The dataset used for the analysis might not be representative of all industries or employee populations. Additionally, the cross-sectional nature of the data prevents us from drawing any causal conclusions. Longitudinal studies would be necessary to establish causal relationships between the identified factors and employee attrition rates.

In conclusion, the results of this study provide valuable insights into the factors affecting employee attrition rates and their positive or negative relationships with attrition. The research results support all the factors proposed in our hypothesis. Age, marital status, years of experience in the organization, the number of organizations an employee has worked for, and job satisfaction were found to significantly influence employee attrition. However, a negative relationship between the years spent in the company and the attrition rate is unexpected. These findings can inform HR practices and policies aimed at reducing attrition rates and improving employee satisfaction and engagement. However, future research should address the limitations of this study and explore the causal relationships between these factors and attrition rates.

Conclusion

In summary, this report has provided a comprehensive analysis of factors affecting employee attrition within the organization. The methodology of this analysis involves cleaning and manipulating data, conducting exploratory data analysis, and building a predictive model using logistic regression. Our logistic regression model identified several significant factors that influence an employee's decision to leave or stay at the company. Key findings reveal that working overtime is the strongest predictor of higher attrition rates, while working experience and the time spent in current roles and with managers could play significant roles in lower attrition rates.

The larger implications of our analysis emphasize the importance of organizations fostering a supportive and engaging work environment to ensure employee retention. Addressing work-life balance concerns, offering growth opportunities, and nurturing strong manager-employee relationships are essential in reducing attrition rates and contributing to the organization's long-term success.

This research opens doors for future studies to explore other potential factors influencing attrition, such as employee engagement levels, team dynamics, and workplace culture. Additionally, future research can delve deeper into the interaction of identified factors, which may provide a more nuanced understanding of the complex interplay between the variables affecting attrition. Moreover, it would be valuable to investigate industry-specific trends, as well as to compare these findings across different organizations to gain a broader perspective on employee attrition.

References

Chakraverty, S. (2006). *Employee attrition in the IT/BPO sector : cost & consequences*. IBS Case Development Centre.

Appendix

Appendix A

Link to dataset on Kaggle:

<https://www.kaggle.com/code/salehahmedrony/hr-analytics-employee-attrition-eda-prediction>

Appendix B

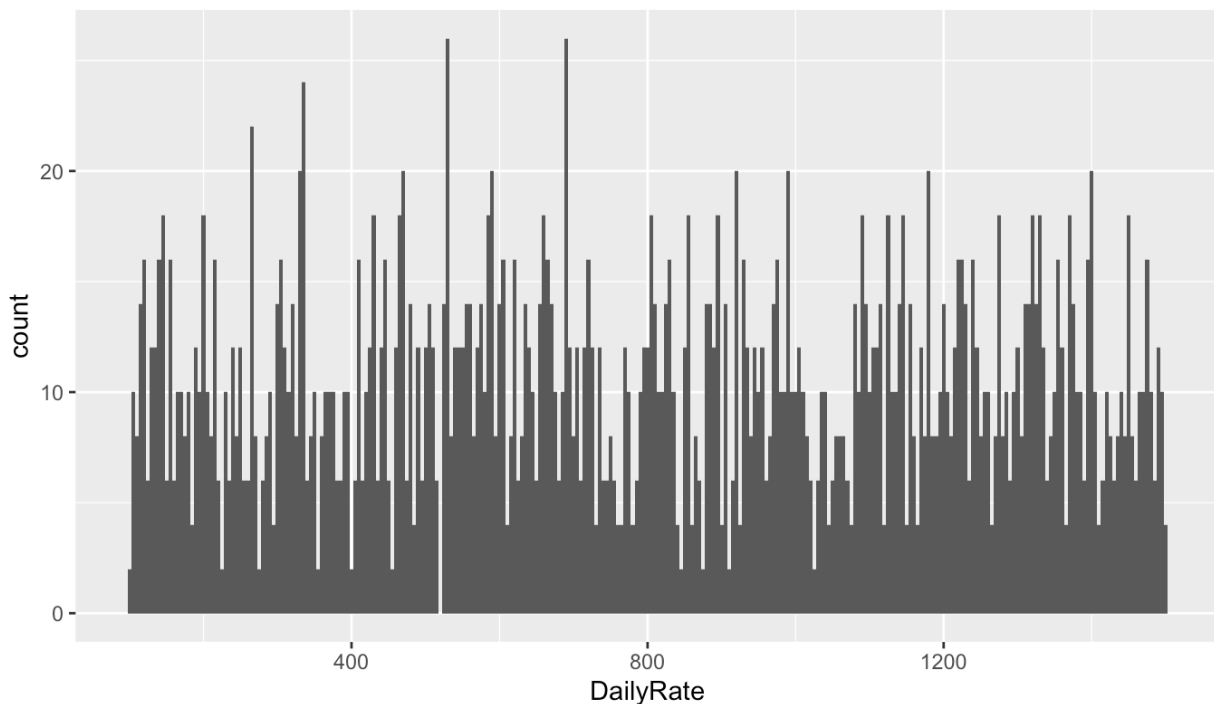
Statistics Summary

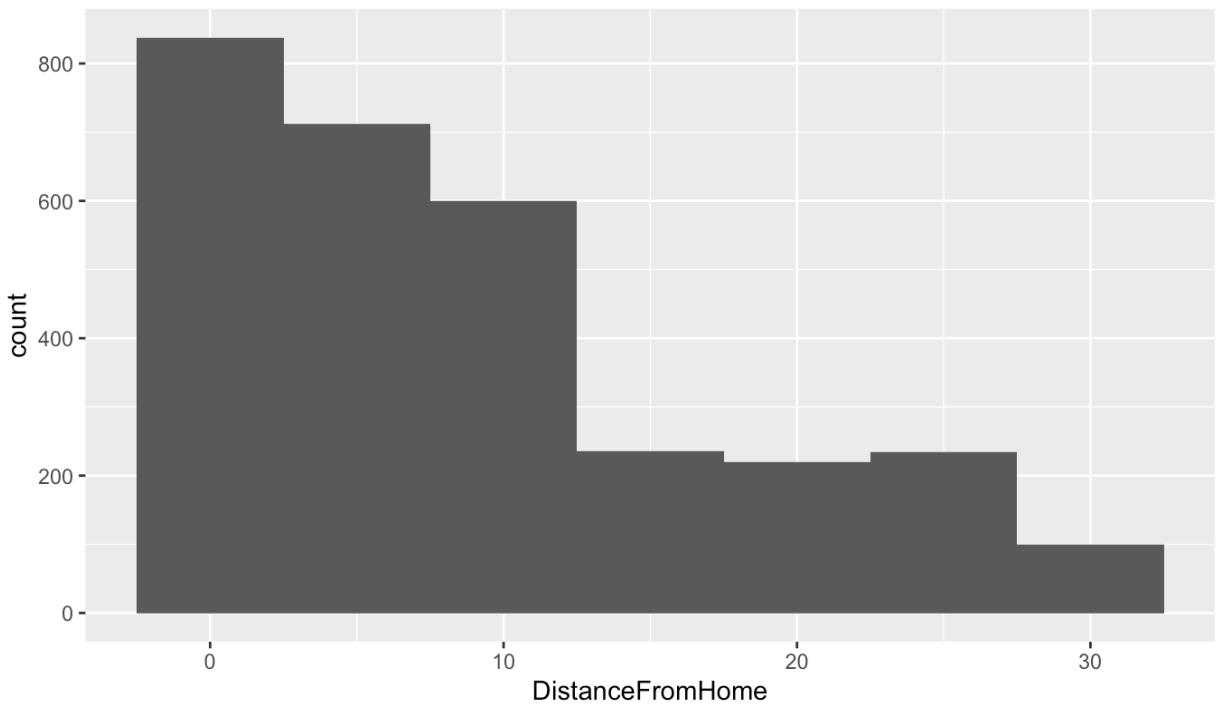
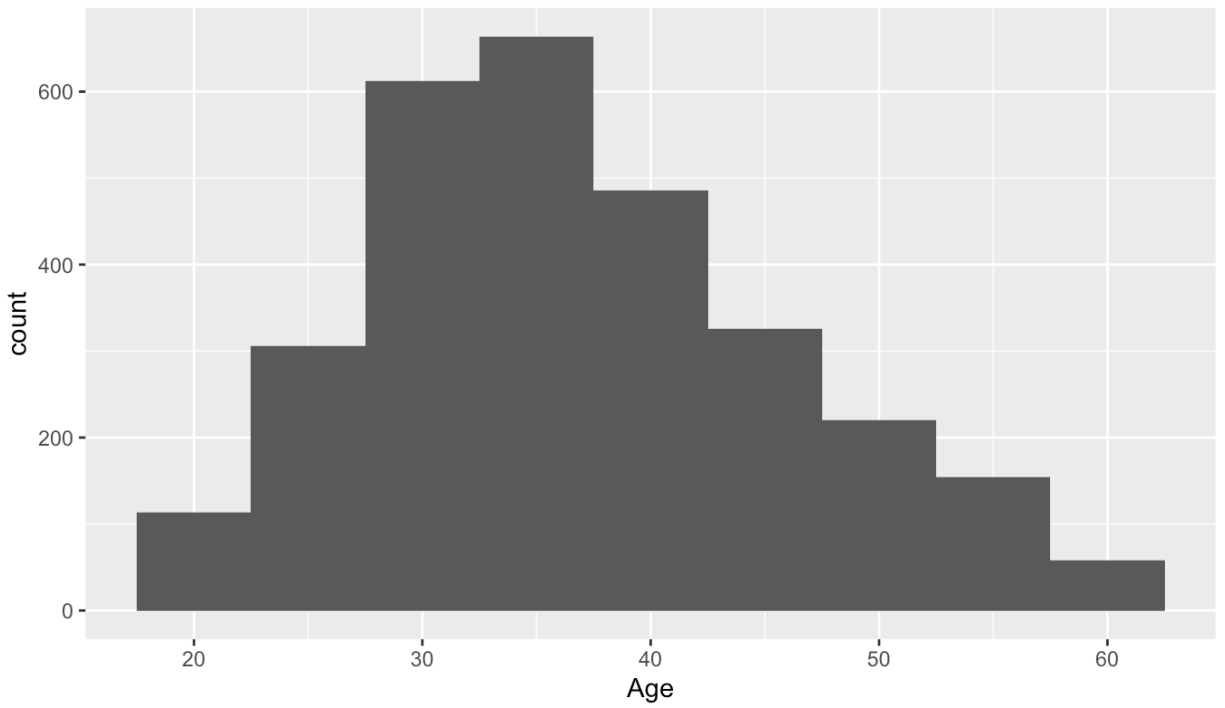


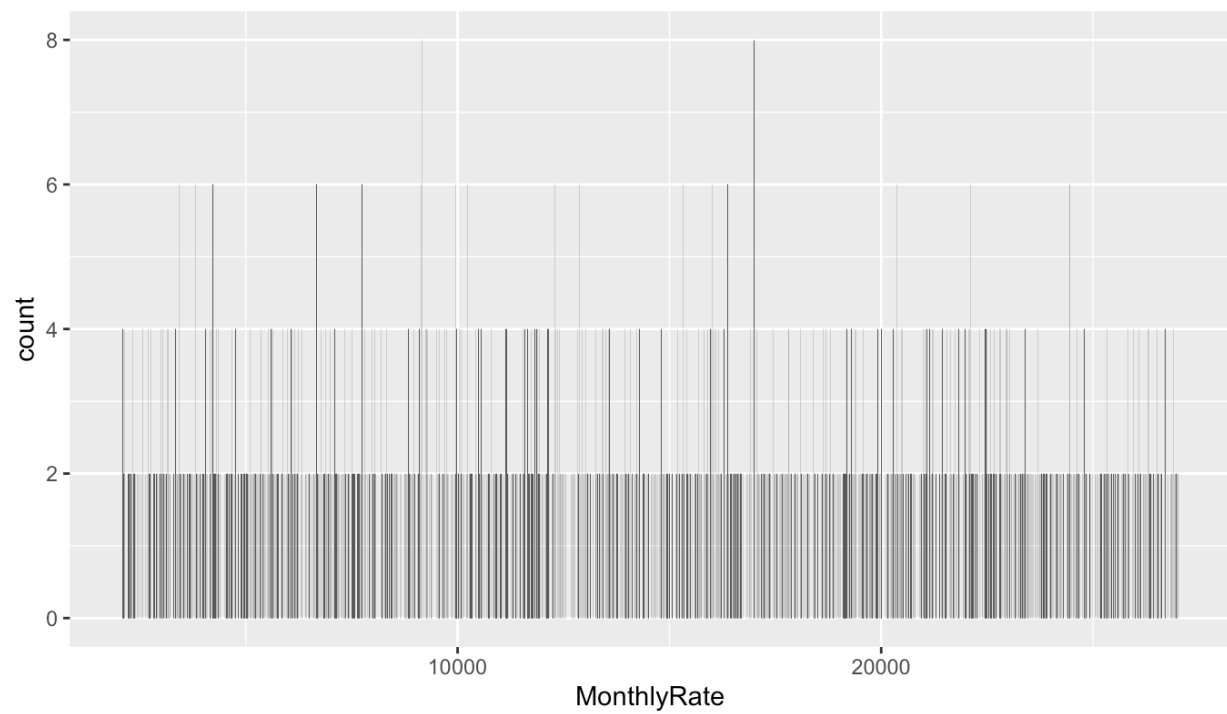
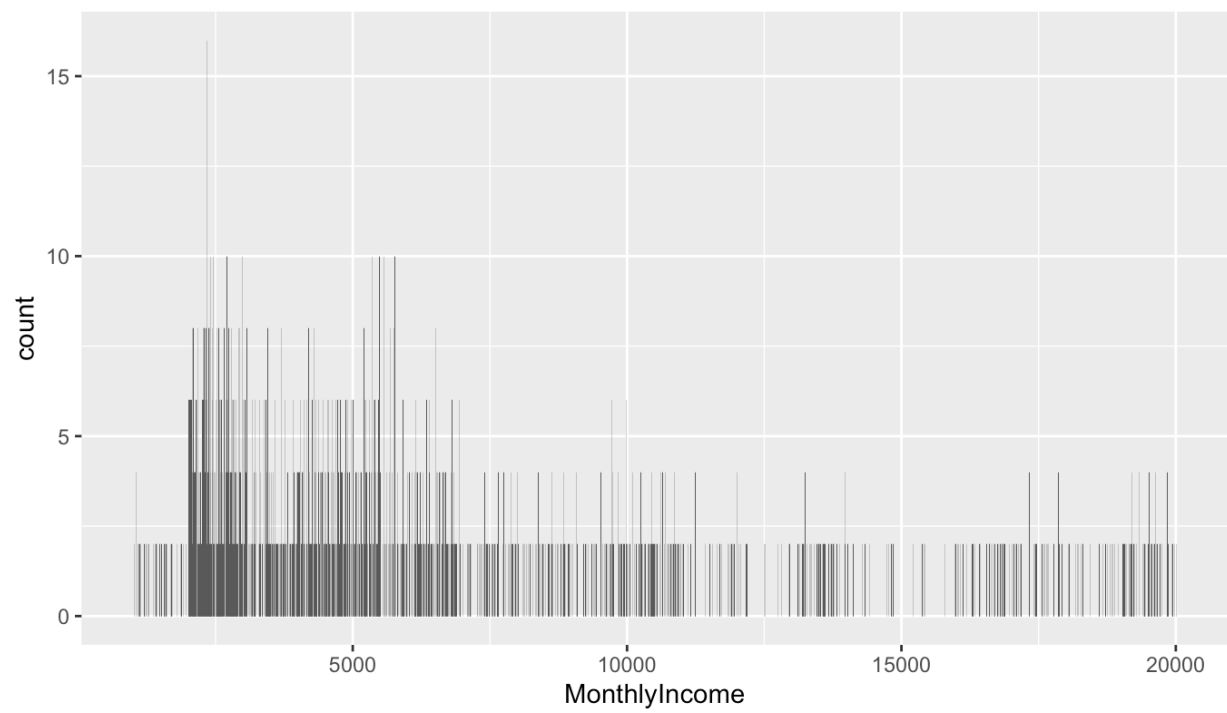
DailyRate	Age	DistanceFromHome	MonthlyIncome	MonthlyRate
Min. : 102.0	Min. : 18.00	Min. : 1.000	Min. : 1009	Min. : 2094
1st Qu.: 465.0	1st Qu.: 30.00	1st Qu.: 2.000	1st Qu.: 2911	1st Qu.: 8045
Median : 802.0	Median : 36.00	Median : 7.000	Median : 4919	Median : 14236
Mean : 802.5	Mean : 36.92	Mean : 9.193	Mean : 6503	Mean : 14313
3rd Qu.: 1157.0	3rd Qu.: 43.00	3rd Qu.: 14.000	3rd Qu.: 8380	3rd Qu.: 20462
Max. : 1499.0	Max. : 60.00	Max. : 29.000	Max. : 19999	Max. : 26999
PercentSalaryHike	TotalWorkingYears	YearsAtCompany	NumCompaniesWorked	HourlyRate
Min. : 11.00	Min. : 0.00	Min. : 0.000	Min. : 0.000	Min. : 30.00
1st Qu.: 12.00	1st Qu.: 6.00	1st Qu.: 3.000	1st Qu.: 1.000	1st Qu.: 48.00
Median : 14.00	Median : 10.00	Median : 5.000	Median : 2.000	Median : 66.00
Mean : 15.21	Mean : 11.28	Mean : 7.008	Mean : 2.693	Mean : 65.89
3rd Qu.: 18.00	3rd Qu.: 15.00	3rd Qu.: 9.000	3rd Qu.: 4.000	3rd Qu.: 84.00
Max. : 25.00	Max. : 40.00	Max. : 40.000	Max. : 9.000	Max. : 100.00
YearsInCurrentRole	YearsSinceLastPromotion	YearsWithCurrManager	TrainingTimesLastYear	
Min. : 0.000	Min. : 0.000	Min. : 0.000	Min. : 0.000	
1st Qu.: 2.000	1st Qu.: 0.000	1st Qu.: 2.000	1st Qu.: 2.000	
Median : 3.000	Median : 1.000	Median : 3.000	Median : 3.000	
Mean : 4.229	Mean : 2.188	Mean : 4.123	Mean : 2.799	
3rd Qu.: 7.000	3rd Qu.: 3.000	3rd Qu.: 7.000	3rd Qu.: 3.000	
Max. : 18.000	Max. : 15.000	Max. : 17.000	Max. : 6.000	

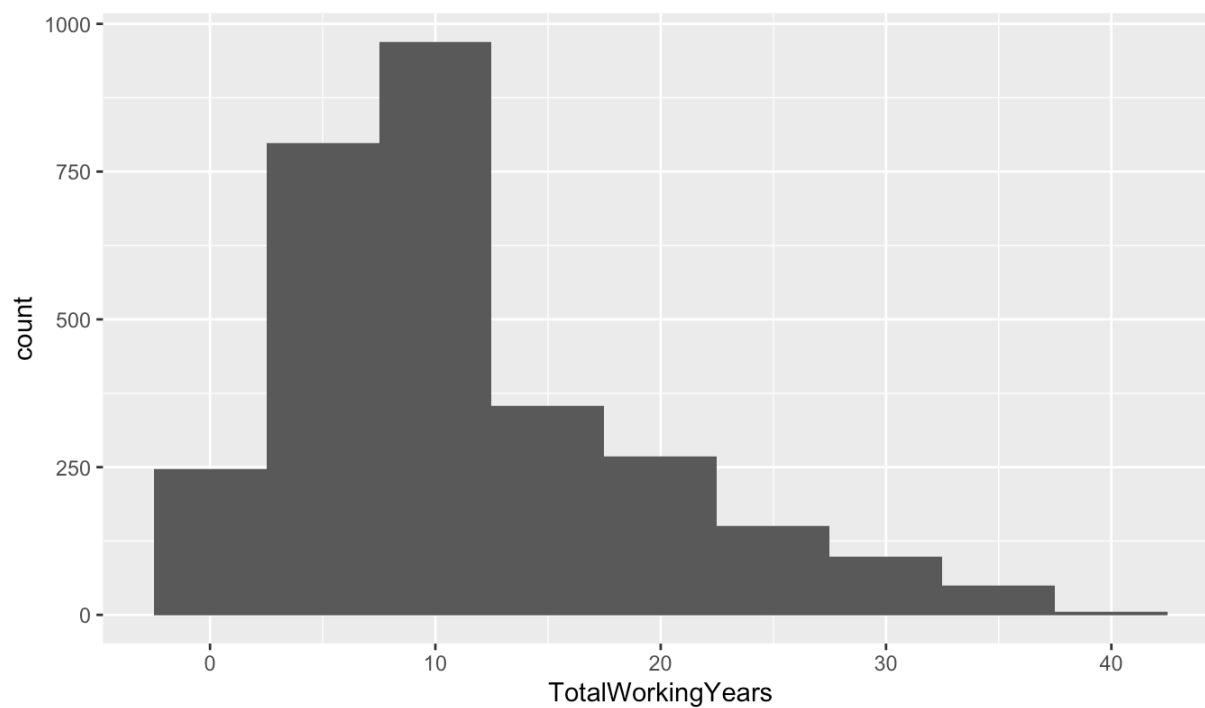
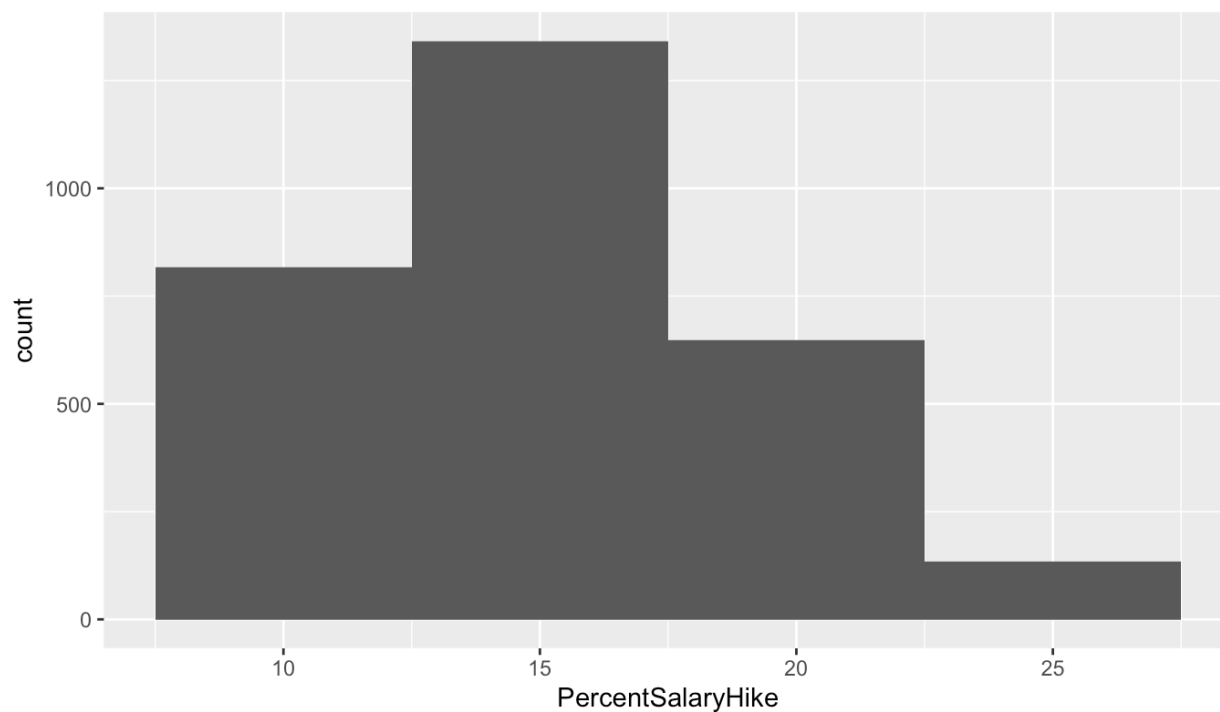
Appendix C

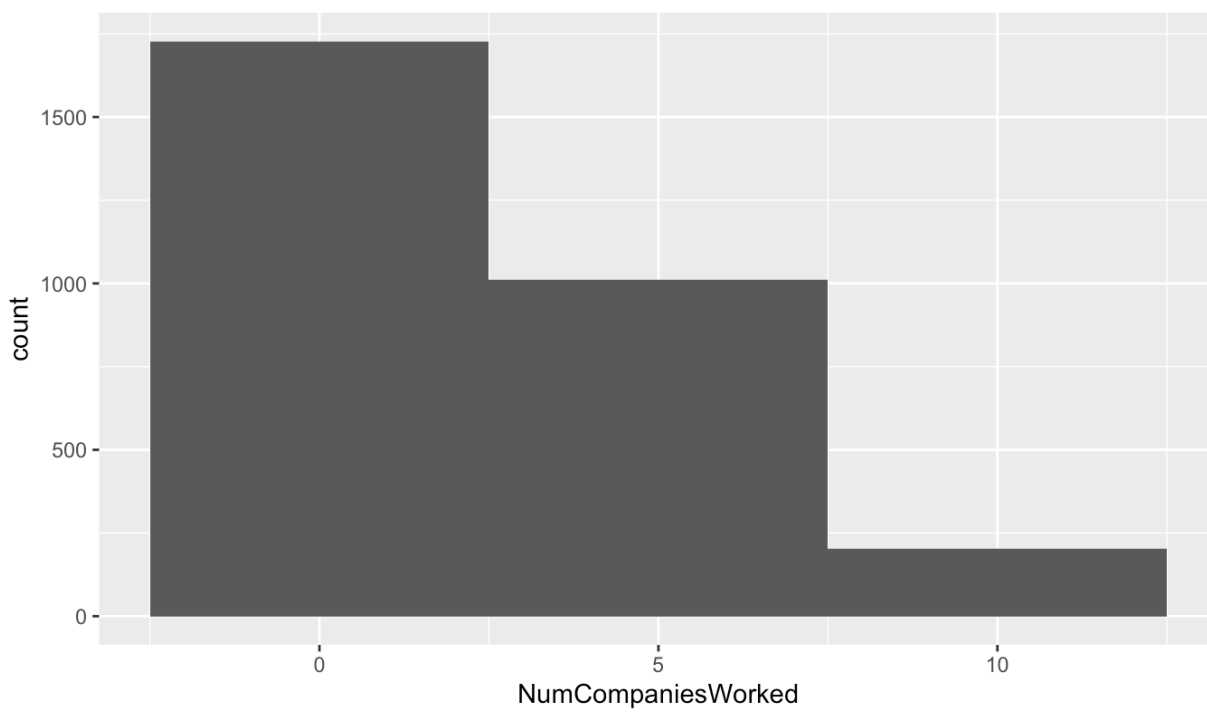
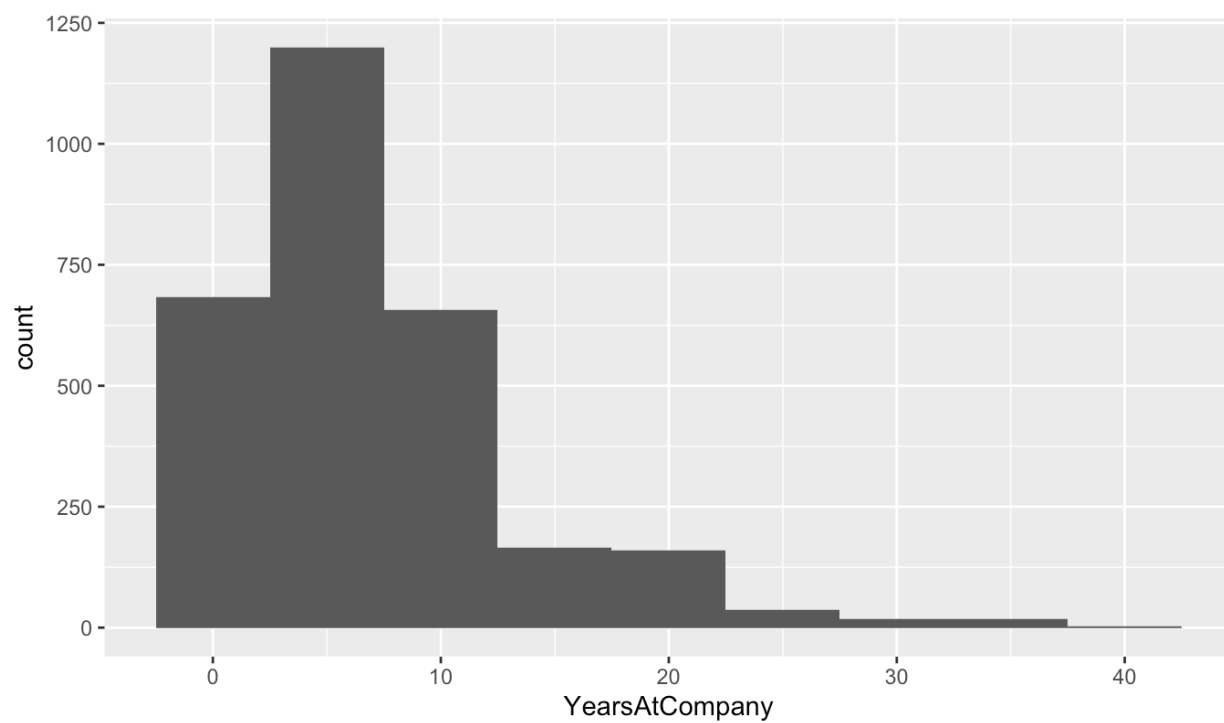
Histograms plotted in univariate analysis for numerical variables

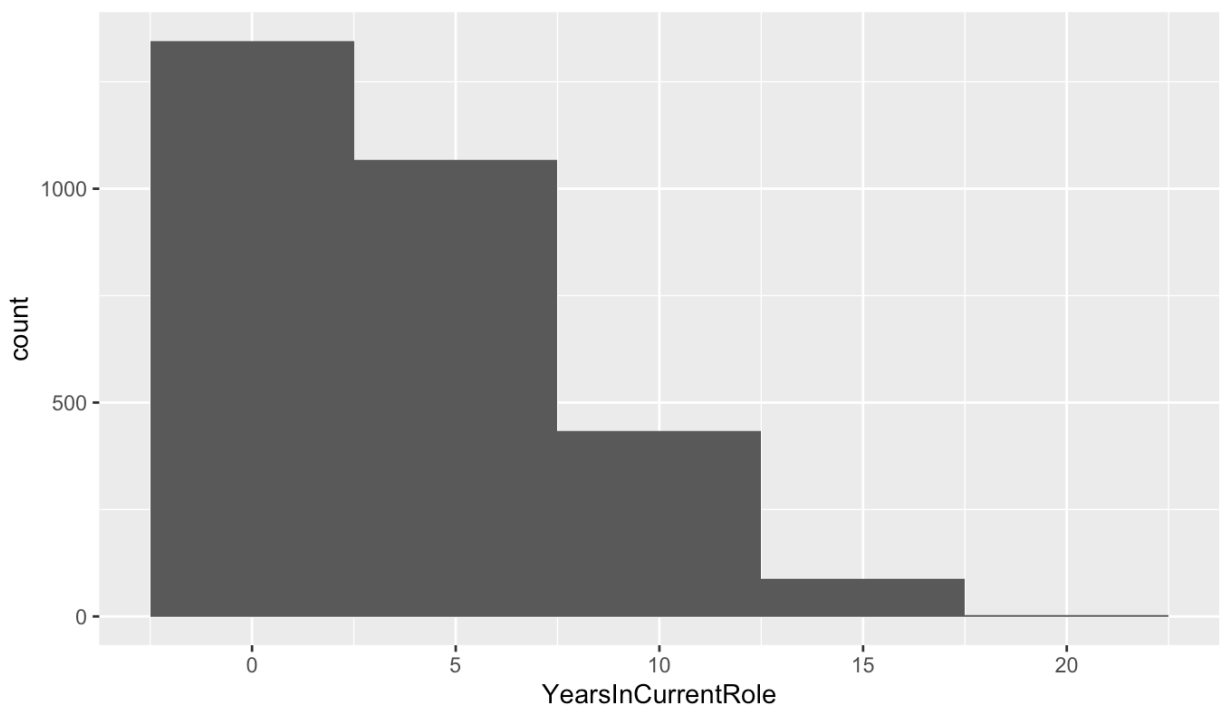
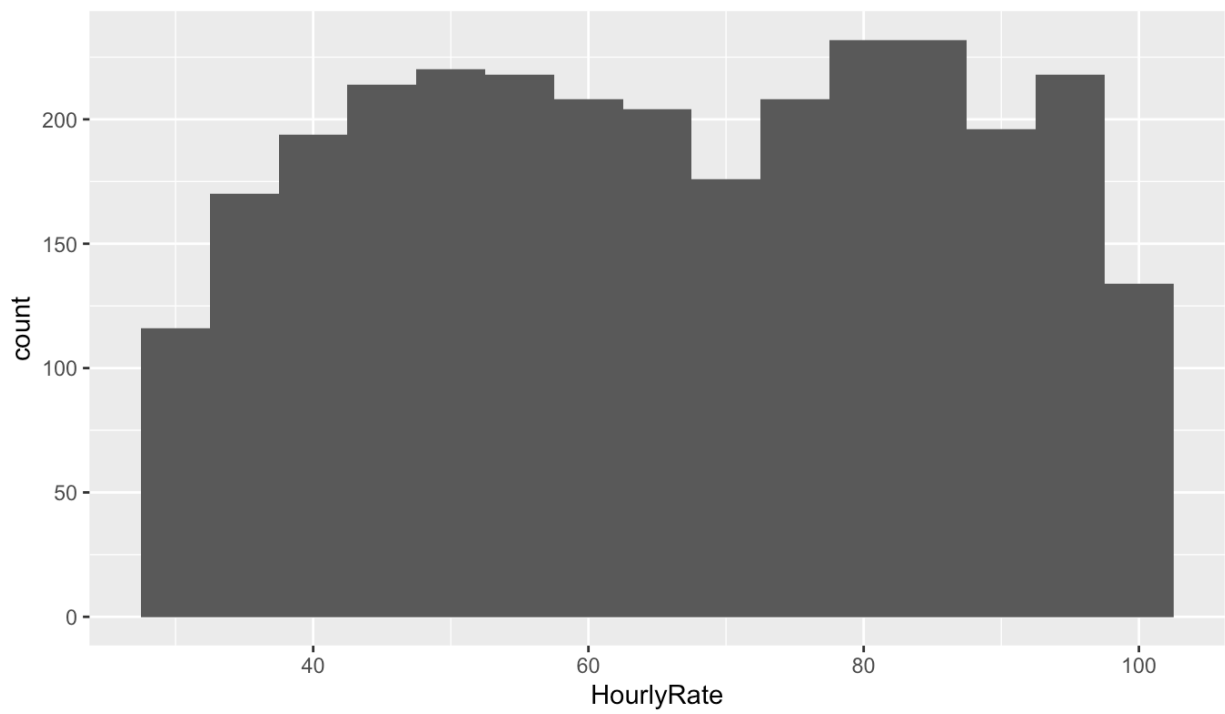


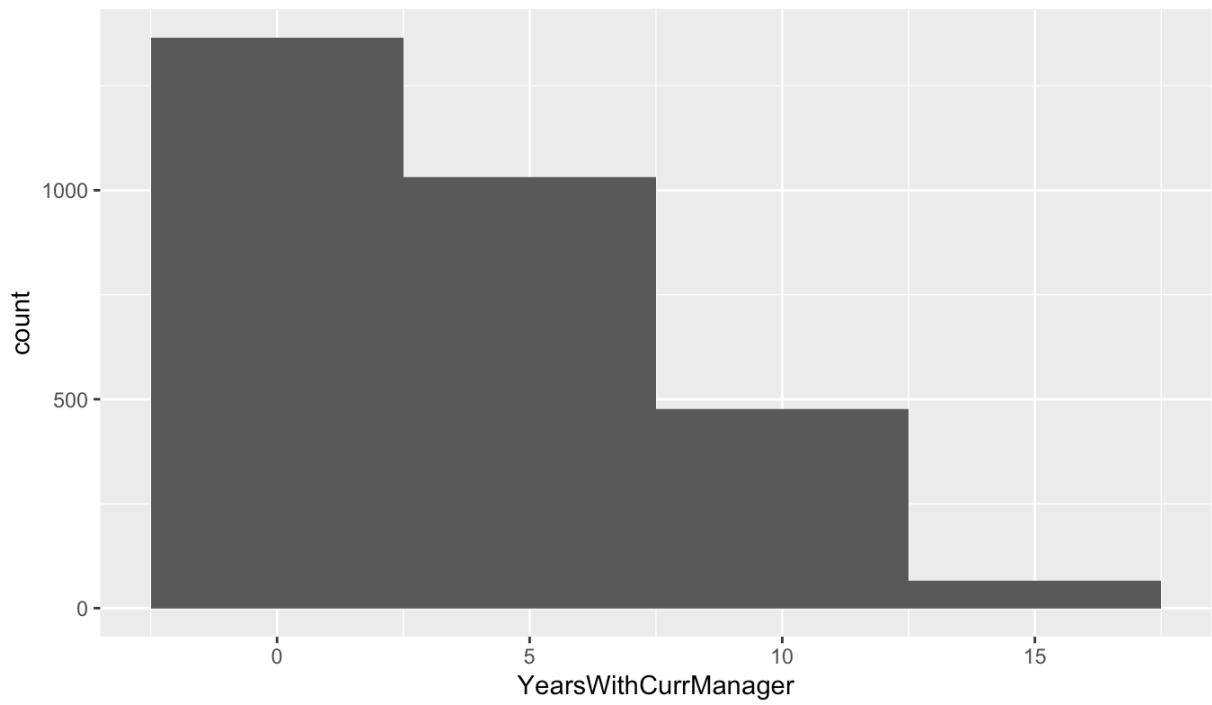
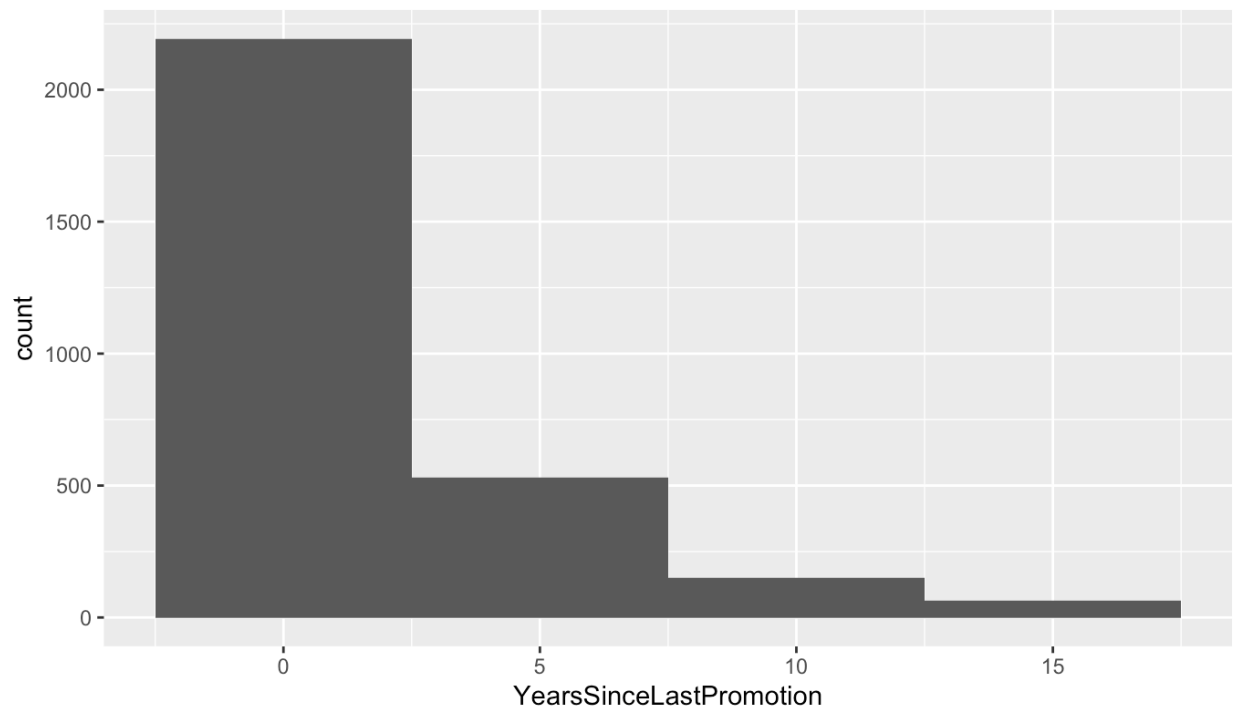


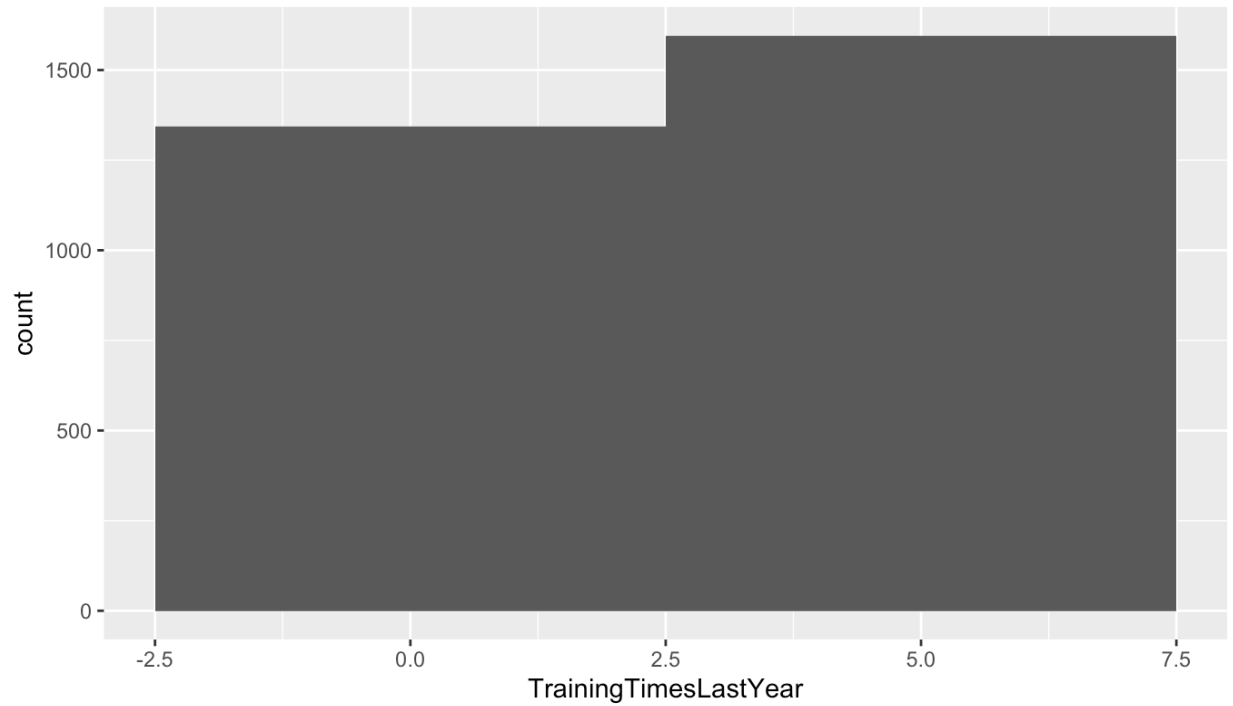












Appendix D

Percentages of subcategories in each categorical variable

No	Yes
0.8387755	0.1612245
* * * * *	

No	Yes
0.7170068	0.2829932
* * * * *	

Non-Travel	Travel_Frequently	Travel_Rarely
0.1020408	0.1884354	0.7095238
* * * * *		

Human Resources	Research & Development	Sales
0.04285714	0.65374150	0.30340136
* * * * *		

1	2	3	4	5
0.11564626	0.19183673	0.38911565	0.27074830	0.03265306
* * * * *				

Human Resources	Life Sciences	Marketing	Medical	Other
0.01836735	0.41224490	0.10816327	0.31564626	0.05578231

Technical Degree

0.08979592

* * * * *

1	2	3	4
0.1965986	0.1904762	0.3006803	0.3122449

* * * * *

1	2	3	4
0.1931973	0.1952381	0.3081633	0.3034014

* * * * *

1	2	3	4
0.05442177	0.23401361	0.60748299	0.10408163

* * * * *

0	1	2	3
0.42925170	0.40544218	0.10748299	0.05782313

* * * * *

Female Male

0.4 0.6

* * * * *

3	4
0.8462585	0.1537415

* * * * *

1	2	3	4
0.05646259	0.25510204	0.59047619	0.09795918

* * * * *

1	2	3	4	5
0.36938776	0.36326531	0.14829932	0.07210884	0.04693878

* * * * *

Healthcare Representative

0.08911565

Human Resources

0.03537415

Laboratory Technician

0.17619048

Manager

Manufacturing Director

Research Director

0.06938776

0.09863946

0.05442177

Research Scientist	Sales Executive	Sales Representative
0.19863946	0.22176871	0.05646259
* * * * *		
Divorced	Married	Single
0.2224490	0.4578231	0.3197279
* * * * *		
1	2	3
0.1877551	0.2061224	0.3122449
4	0.2938776	

Appendix E

Call:

```
glm(formula = y_train ~ ., family = binomial(link = "logit"),
     data = X_train_scaled)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.6831	-0.4860	-0.2516	-0.0843	3.3290

Coefficients: (1 not defined because of singularities)

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-3.368056	4.031600	-0.835	0.403485
Age	-0.432033	0.108010	-4.000	6.34e-05 ***
DailyRate	-0.122701	0.075200	-1.632	0.102750
DistanceFromHome	0.388033	0.075202	5.160	2.47e-07 ***
HourlyRate	0.023001	0.076758	0.300	0.764437
MonthlyIncome	-0.016583	0.327029	-0.051	0.959557
MonthlyRate	0.036539	0.075359	0.485	0.627767
NumCompaniesWorked	0.477332	0.082971	5.753	8.77e-09 ***
OverTimeYes	1.931027	0.165343	11.679	< 2e-16 ***
PercentSalaryHike	-0.038398	0.121411	-0.316	0.751804
TotalWorkingYears	-0.512297	0.197597	-2.593	0.009524 **
TrainingTimesLastYear	-0.222058	0.079169	-2.805	0.005034 **
YearsAtCompany	0.796277	0.199877	3.984	6.78e-05 ***
YearsInCurrentRole	-0.631225	0.143446	-4.400	1.08e-05 ***
YearsSinceLastPromotion	0.566715	0.112301	5.046	4.50e-07 ***
YearsWithCurrManager	-0.477905	0.135763	-3.520	0.000431 ***
BusinessTravelNon.Travel	-0.367945	0.098642	-3.730	0.000191 ***
BusinessTravelTravel_Frequently	0.301594	0.071317	4.229	2.35e-05 ***
BusinessTravelTravel_Rarely	NA	NA	NA	NA
DepartmentResearch...Development	6.656993	232.961924	0.029	0.977203

DepartmentSales	6.492504	225.530106	0.029	0.977034	
Education	0.004104	0.076077	0.054	0.956979	
EducationFieldLife.Sciences	-0.262694	0.343284	-0.765	0.444129	
EducationFieldMarketing	-0.003352	0.225883	-0.015	0.988161	
EducationFieldMedical	-0.313864	0.325019	-0.966	0.334204	
EducationFieldOther	-0.060677	0.169342	-0.358	0.720111	
EducationFieldTechnical.Degree	0.134672	0.198733	0.678	0.497992	
JobSatisfaction	-0.461321	0.076082	-6.063	1.33e-09	***
EnvironmentSatisfaction	-0.482249	0.077663	-6.210	5.31e-10	***
WorkLifeBalance	-0.279051	0.072950	-3.825	0.000131	***
StockOptionLevel	-0.037382	0.114785	-0.326	0.744676	
GenderMale	0.137409	0.076778	1.790	0.073506	.
PerformanceRating	0.050329	0.119922	0.420	0.674715	
JobInvolvement	-0.357067	0.072909	-4.897	9.71e-07	***
JobLevel	-0.043524	0.303217	-0.144	0.885862	
JobRoleHuman.Resources	2.781727	88.429085	0.031	0.974905	
JobRoleLaboratory.Technician	0.537372	0.157179	3.419	0.000629	***
JobRoleManager	0.027866	0.192609	0.145	0.884966	
JobRoleManufacturing.Director	0.026264	0.132973	0.198	0.843422	
JobRoleResearch.Director	-0.383321	0.218259	-1.756	0.079043	.
JobRoleResearch.Scientist	0.095992	0.167970	0.571	0.567673	
JobRoleSales.Executive	0.285426	0.378980	0.753	0.451364	
JobRoleSales.Representative	0.380938	0.232908	1.636	0.101930	
MaritalStatusMarried	0.188944	0.112856	1.674	0.094091	.
MaritalStatusSingle	0.649026	0.139193	4.663	3.12e-06	***
RelationshipSatisfaction	-0.257317	0.075835	-3.393	0.000691	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1812.1 on 2057 degrees of freedom
Residual deviance: 1199.9 on 2013 degrees of freedom
AIC: 1289.9

Number of Fisher Scoring iterations: 15

Appendix F

<u>(Intercept)</u>	<u>Age</u>
0.03445656	0.64918790
<u>DailyRate</u>	<u>DistanceFromHome</u>
0.88452773	1.47407863

<u>HourlyRate</u>	<u>MonthlyIncome</u>
1.02326777	0.98355325
<u>MonthlyRate</u>	<u>NumCompaniesWorked</u>
1.03721525	1.61176919
<u>OverTimeYes</u>	<u>PercentSalaryHike</u>
6.89658767	0.96233031
<u>TotalWorkingYears</u>	<u>TrainingTimesLastYear</u>
0.59911782	0.80086894
<u>YearsAtCompany</u>	<u>YearsInCurrentRole</u>
2.21727149	0.53193966
<u>YearsSinceLastPromotion</u>	<u>YearsWithCurrManager</u>
1.76246778	0.62008131
<u>BusinessTravelNon.Travel</u>	<u>BusinessTravelTravel_Frequently</u>
0.69215515	1.35201153
<u>BusinessTravelTravel_Rarely</u>	<u>DepartmentResearch...Development</u>
NA	778.20696019
<u>DepartmentSales</u>	<u>Education</u>
660.17421691	1.00411236
<u>EducationFieldLife.Sciences</u>	<u>EducationFieldMarketing</u>
0.76897710	0.99665381
<u>EducationFieldMedical</u>	<u>EducationFieldOther</u>
0.73061837	0.94112713
<u>EducationFieldTechnical.Degree</u>	<u>JobSatisfaction</u>
1.14416163	0.63045052
<u>EnvironmentSatisfaction</u>	<u>WorkLifeBalance</u>
0.61739311	0.75650161
<u>StockOptionLevel</u>	<u>GenderMale</u>
0.96330825	1.14729687
<u>PerformanceRating</u>	<u>JobInvolvement</u>
1.05161740	0.69972575
<u>JobLevel</u>	<u>JobRoleHuman.Resources</u>
0.95740910	16.14688453
<u>JobRoleLaboratory.Technician</u>	<u>JobRoleManager</u>
1.71150393	1.02825799
<u>JobRoleManufacturing.Director</u>	<u>JobRoleResearch.Director</u>
1.02661242	0.68159430
<u>JobRoleResearch.Scientist</u>	<u>JobRoleSales.Executive</u>
1.10075026	1.33032893
<u>JobRoleSales.Representative</u>	<u>MaritalStatusMarried</u>
1.46365644	1.20797284
<u>MaritalStatusSingle</u>	<u>RelationshipSatisfaction</u>
1.91367661	0.77312333