

# IBM Employee Turnover Analysis

Abby Brown & Tao Wu

## Abstract

During Covid19, a lot of companies had difficulty hiring new employees, so we wanted to do some research and analysis to find out how to keep employees working for the company. The project uses the data of IBM resigned employees for analysis. On the basis of observing the influencing factors of turnover rate, we established a model to predict which employees are more likely to leave. Through the practice of data on IBM resigned employees, we discovered the factors that affect employee turnover, and reviewed the data analysis process using R language to deepen the understanding of the meaning of data analysis. Since the data contained no missing values and has various elements relevant to the employee, it is an excellent source to answer our questions.

## 1 Introduction

A company's turnover rate is the percentage of employees who left a company over a certain period of time. Many factors influence resignations including commute, overtime, environment satisfaction, job satisfaction, and traveling for work, to name a few. See below for the different variables relevant to employees that we explored.

Index	Variables
1	Age
2	BusinessTravel
3	Department
4	DistanceFromHome
5	Education
6	EducationField
7	EnvironmentSatisfaction
8	Gender
9	HourlyRate
10	JobInvolvement
11	JobLevel
12	JobRole
13	JobSatisfaction
14	MaritalStatus
15	MonthlyIncome
16	NumCompaniesWorked
17	OverTime
18	PercentSalaryHike
19	PerformanceRating
20	RelationshipSatisfaction
21	StandardHours
22	StockOptionLevel
23	TotalWorkingYears

Index	Variables
24	TrainingTimesLastYear
25	WorkLifeBalance
26	YearsAtCompany
27	YearsInCurrentRole
28	YearsSinceLastPromotion
29	YearsWithCurrManager

## 2 Obtaining the Data

The data for this project was obtained via Kaggle:

<https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-attrition-dataset>

We downloaded and renamed the file “IBM.csv” and read it in. After initial analysis, we noticed, among other things, that the average age of employees is about 36 years old, the oldest is 60 years old, and the youngest is 18 years old. Among all 1,470 employees, 237 resigned, with a turnover rate of 16%. The average employee income is \$6,500 a month, the median is \$4,919, the minimum is \$1,009, and the max is \$19,999 a month. See below for further descriptions.

## 3 Basic Identity Factors Influencing Turnover Rate

We visualized how different basic identity factors can influence turnover rate by looking at how often people resigned considering these factors. In Figure 1 we see that gender has little to do with turnover rate but people under the age of 33 are more likely to leave their jobs as shown in Figure 2. Figure 3 tells us that the higher level of education, the lower the turnover rate, but the difference is not particularly obvious. Figures 4 and 5 show us that unmarried singles are more likely to leave their jobs as are people who have worked for more than 5 companies.

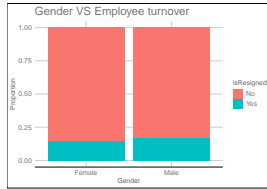


Figure 1: Employee turnover considering gender.

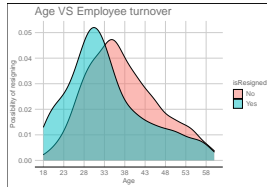


Figure 2: Employee turnover considering age.

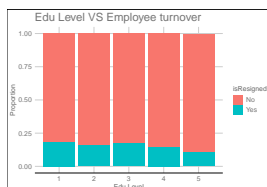


Figure 3: Employee turnover considering education.

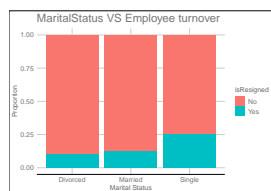


Figure 4: Employee turnover considering marital status.

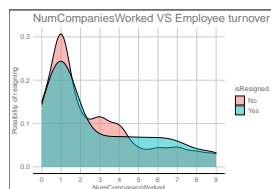


Figure 5: Employee turnover considering the number of companies the employee previously worked at.

#### 4 Work Identity Factors Influencing Turnover Rate

Regarding basic information about a person's professional life, seen below in figures 6 through 9, people with a total working experience of less than 8 years are more likely to leave as are people who have worked for IBM for less than 4 years. Low job level employees are more likely to leave the company. Think of job levels as power levels, so a 1 would be entry level and 4 would be a manager of some sort. Specifically, there is a high turnover rate for the sales department.

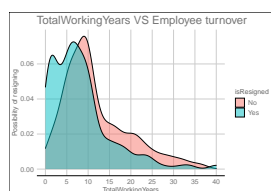


Figure 6: Employee turnover considering total working years for the employee.

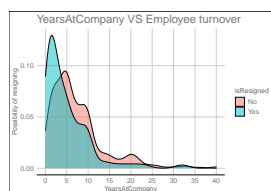


Figure 7: Employee turnover considering the number of years at the company.

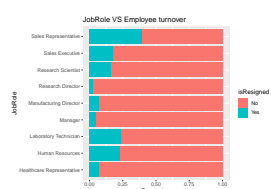


Figure 8: Employee turnover considering job role.

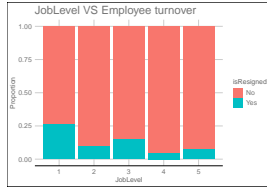


Figure 9: Employee turnover considering job level.

## 5 Analyzing how Salary, Involvement, & Performance Influence Turnover Rate

Analyzing how monthly income, job involvement, and performance rating affect turnover rates in figures 10-13, we discover some interesting results. The turnover rate of employees with a monthly income below \$4,000 and around \$10,000 is higher. And while performance rating has very little effect on turnover rate, we do see that the higher the work involvement, the lower the turnover rate. But with the same work involvement, the monthly income of resigned employees is lower than in-service employees.

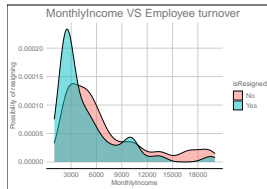


Figure 10: Employee turnover considering monthly income.

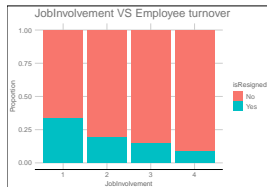


Figure 11: Employee turnover considering job involvement.



Figure 12: Employee turnover considering performance rating.

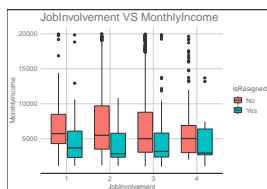


Figure 13: Employee turnover considering job involvement and monthly income.

## 6 Benefit Related Factors Influencing Turnover Rate

In this section, we analyzed figures 14-17 to see how stock option levels, salary hike percentage, training times in the previous year, and years since the last promotion can affect resignations. Results indicate that the highest and lowest stock option level are more likely to leave the company. Oddly, the salary hike percentage between 15%, 17%, and greater than 22% are more likely to leave. Generally speaking, the fewer the training times in the last year, the more likely employees are to leave. We also see that people whose salary increase interval is less than half a year or about 7 years are more likely to leave.

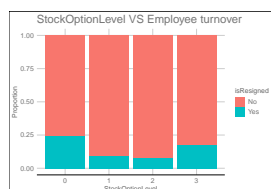


Figure 14: Employee turnover considering the level of stocks the employee participates in.

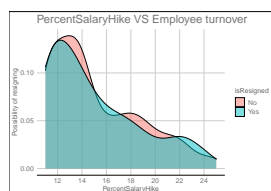


Figure 15: Employee turnover considering the percentage the employee's salary hiked in the past year.



Figure 16: Employee turnover considering training times last year for the individual employee.

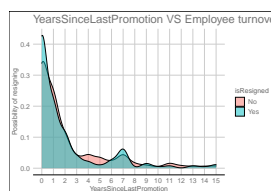


Figure 17: Employee turnover considering the years since the last promotion.

## 7 Quality of Life Influencing Turnover Rate

Factors analyzed in figures 18-20 below are environment satisfaction, job satisfaction, and relationship satisfaction. Unsurprisingly, people with lower levels of satisfaction in any or all of the three factors, are more likely to resign.

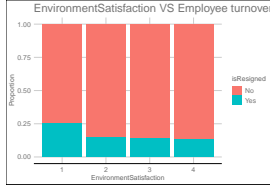


Figure 18: Employee turnover considering environmental satisfaction.

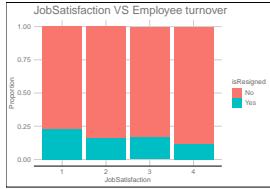


Figure 19: Employee turnover considering job satisfaction.

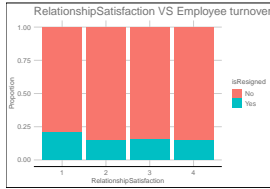


Figure 20: Employee turnover considering relationship satisfaction.

## 8 Work Life Balance Influences on Turnover Rate

Commute, overtime, and traveling for work are considered in this section as seen in figures 21, 22, and 23. We can see that who whose commute is further than 11 miles are more likely to resign, the more overtime an employee gets, the more likely they are to leave, and those who travel for work frequently are more likely to leave.

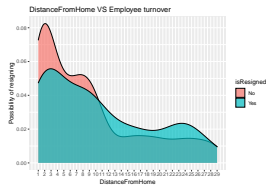


Figure 21: Employee turnover considering distance from home.

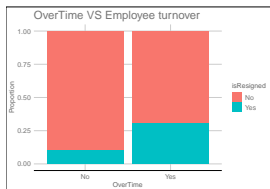


Figure 22: Employee turnover considering overtime.

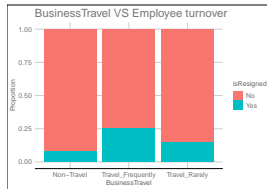


Figure 23: Employee turnover considering the amount of business travel.

## 9 Data preparation

There are a few steps that need to be taken to prepare the data. We cleaned it, checked for missing values, and shuffled to reduce variance and improve accuracy. Next, we created a train/test set and then built the models. Three models, glm, random forest, and gbm, were tested and compared for best accuracy. GLM had the best accuracy so we used that all the way through.

### 9.1 Clean the data:

First, the dataset has no missing values. Then, the columns: EmployeeCount, EmployeeNumber, Over18, StandardHours are irrelevant for the prediction, so we need to remove them.

- 1) Check if there are any missing values.

```
[1] 0
```

- 2) Shuffle the data randomly to reduce variance and improve model accuracy.

### 9.2 Create train/test set

- 1) Split the data: 80% for train set, 20% for test set.

Train set:

```
[1] 1176 32
```

Test set:

```
[1] 294 32
```

- 2) Verify if the randomization process(shuffle) is correct - Results for male are about 60% means all good.

```
Female    Male
0.3988095 0.6011905
```

```
Female    Male
0.4047619 0.5952381
```

## 10 Build the models

### 1. glm model

	No	Yes
No	231	8
Yes	32	23

[1] 86.39%

### 2. Random forest model

	No	Yes
No	238	1
Yes	49	6

[1] 82.99%

### 3. GBM model

```
pred_gbm
  No Yes
No 239  0
Yes 43 12
```

	No	Yes
No	239	0
Yes	43	12

[1] 85.37%

We see that GLM has the best accuracy so below we used it to display the most important factors for resignation using a scale of 1-10 with 0 being not important and 10 being 100% important.



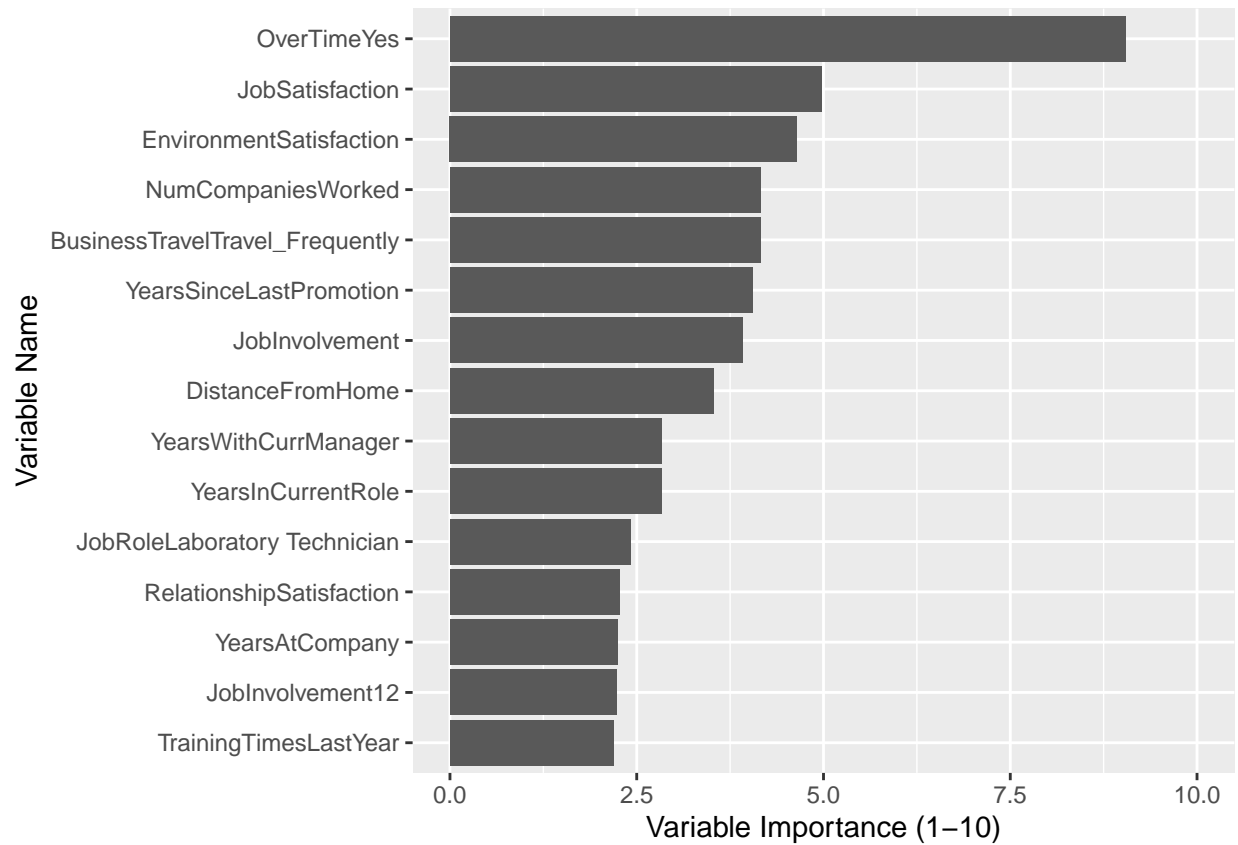


Figure 24: This contains a scale of the top 15 variables most importance for resignation.

—What we can know based on the model— Use this model to predict the possibility of employees resigning.

1. To find out the possibility of employees resigning based on the correlation between JobInvolvement and

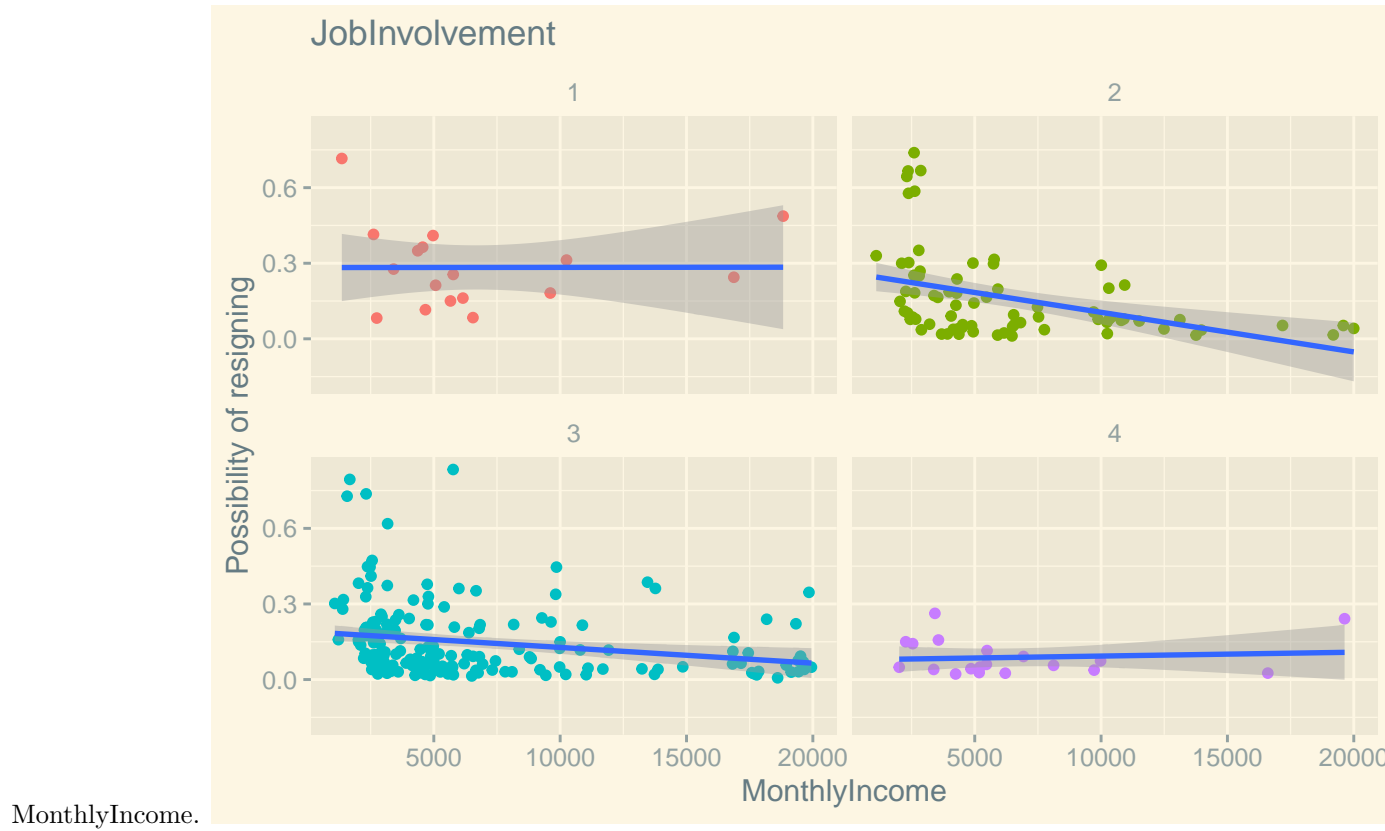
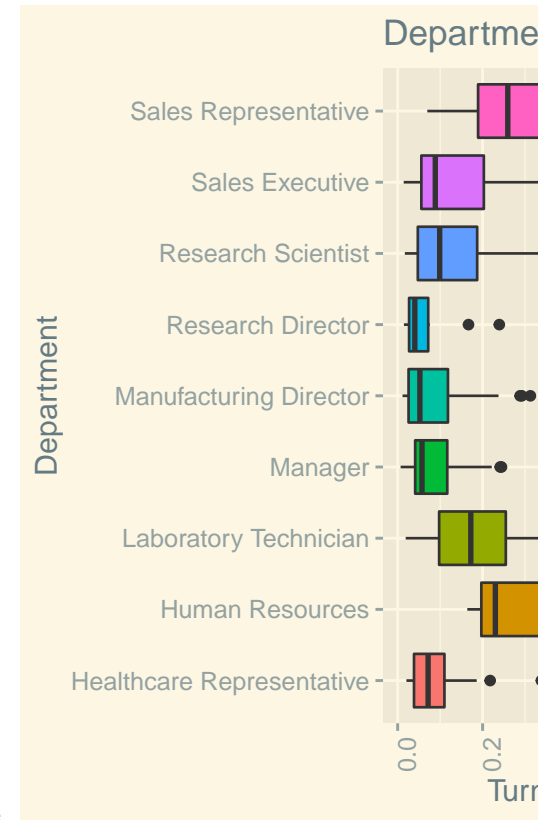


Figure 25: This displays how monthly income influences resignation on different levels of job involvement.

Above is a different view of figure 13. It is easier to see now that for those with higher job involvement, as their income increases, the turnover curve will be relatively stable unlike the lower job involvement. Therefore, in IBM, the most concerning factor for the most involved employees may not be about money, and further exploration is needed.



2. Here we used the model to predict which Job Role has a high turnover rate.

Figure 26: This figure explores resignation rates in different departments.

The boxplot above shows sales persons are most likely to resign while Managers and Healthcare Representatives are most likely to stay.

## 11 Conclusion

With many variables to consider in employee retention, the GLM model has helped us find the most important ones. The 5 most important factors for resignation are Overtime, Environmental satisfaction, Business Travel, Job Satisfaction, and Number of Companies Worked. Considering this information with the goal of retaining employees, employers should work hard to ensure minimal overtime and to create a good work environment.