

DATE: 5/17/2024

TO: IBM HR Manager

FROM: Collaboration

SUBJECT: Using Attrition to Increase Retention

Introduction

There are many reasons why employees will leave their current jobs and employers have the ability to use data analysis to understand why employees leave and how they can retain them. "...the majority of reasons why employees quit their job are under the control of the employer. In fact, any element of your current workplace, your culture, and environment, the employee's perception of his or her job, and opportunities are all factors that the employer affects". (Heathfield, 2021). These factors are key in understanding why people leave and how we can retain them. Utilizing employee analytics (also known as People or HR analytics) is the best option in figuring out what factors are an issue for that particular company. What is employee analytics? "...using statistical insights from employee data to make talent management decisions..." (Leonardi, 2018). It is basically using the data that has been collected in human resources to figure out better ways to retain employees, recruit new employees, etc. We found that while employee analytics has seen a rise, we also understand that companies have a hard time understanding what factors should be included for a much better picture. "A survey by Tata Consultancy Services found that just 5% of big-data investments go to HR, the group that typically manages people analytics. And a recent study by Deloitte showed that although people analytics has become mainstream, only 9% of companies believe they have a good understanding of which talent dimensions drive performance in their organizations" (Leonardi, 2018).

An employer's ability to retain employees no longer hinges on just money nor does it hinge on just being a "good" place to work. It is a combination of factors that can lead to employees to make the decision to stay or leave. "Employee turnover is costly. Turnover affects the performance of an organization, and it becomes increasingly difficult to manage as the competition for skilled employees continues to increase" (8 Essential Employee Retention Factors Modern Employers Ignore, 2019). This is especially true of IBM's employees, the company needs to keep their skilled employees and be more attractive to new

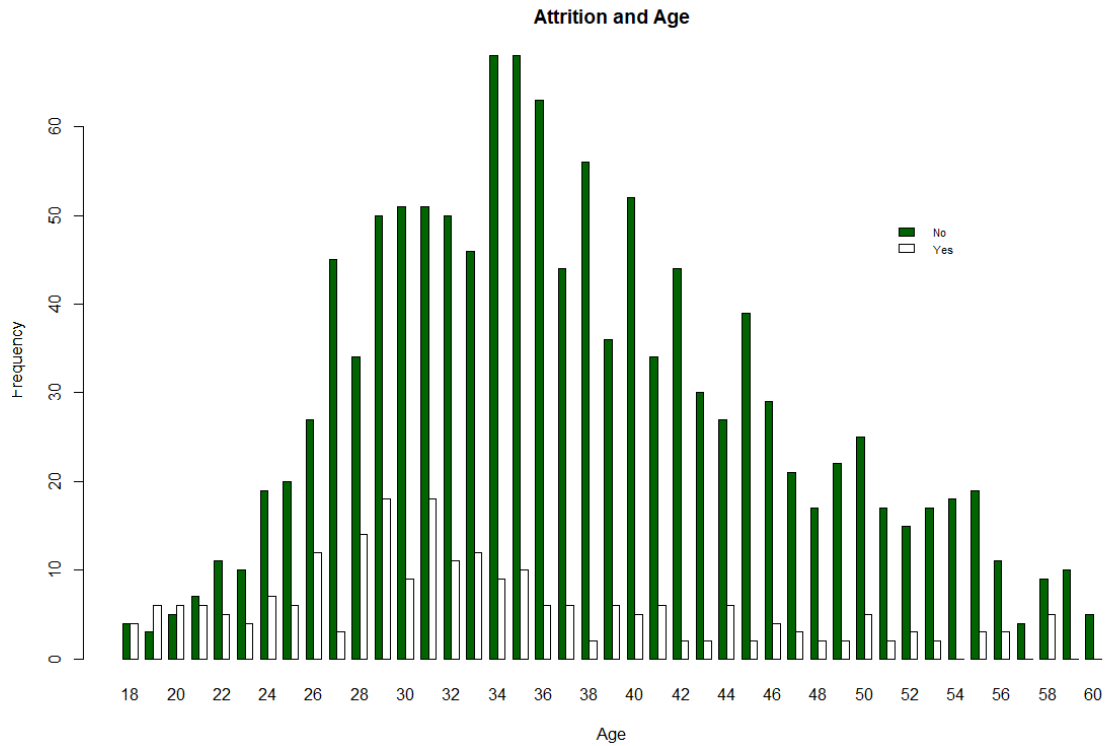
potential skilled employees. The article cites 8 ignored factors that can lead to retention, such as flexibility, benefits, personal development, etc. (8 Essential Employee Retention Factors Modern Employers Ignore, 2019). What an employee wants/needs for their jobs have changed drastically. We are finding that people value flexibility (working from home, hybrid scheduling, flexing in/out), work life balance, growth, etc. when it comes to their choice in jobs.

We were provided a small sample from IBM to figure out what factors at their company are causing attrition, so they can have a better chance at building a strategy to retain employees. For the remainder of the paper we will discuss the data we were provided, the analysis of that data, and conclude with results and recommendations.

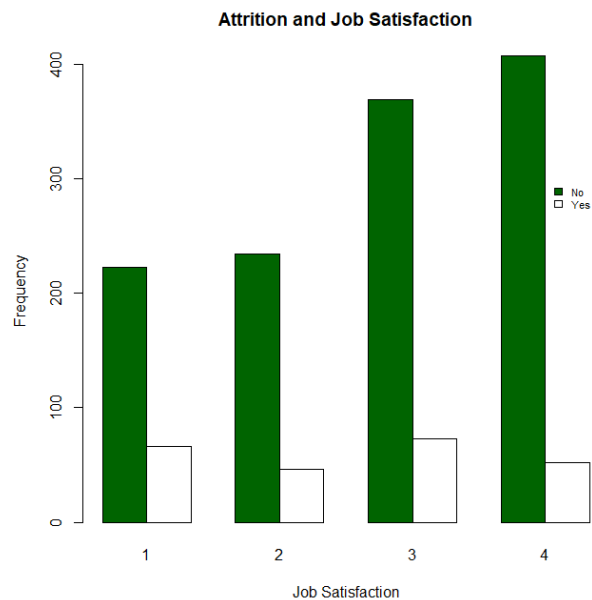
Data Description

As was mentioned in the introduction, IBM provided us with a small sample size of data. The data provided had observations from 1,470 employees, which as we know is very small in comparison to IBM as a whole. Included with those observations were 35 variables, but not all the variables were included in the analysis. We chose to exclude the variables Employee count, Over 18, and Standard Hours because the value for each variable was the same for all observations. Because of this, we can assume that these variables would have no impact on the analysis. Also, we excluded Yearly income because when we ran the regression the result returned “NA”, even though there were values listed. You will find the variable definition and the rating scales for some variables such as work life balance in the appendix. Below we are only going to touch on a couple of relationships with Attrition that we found interesting.

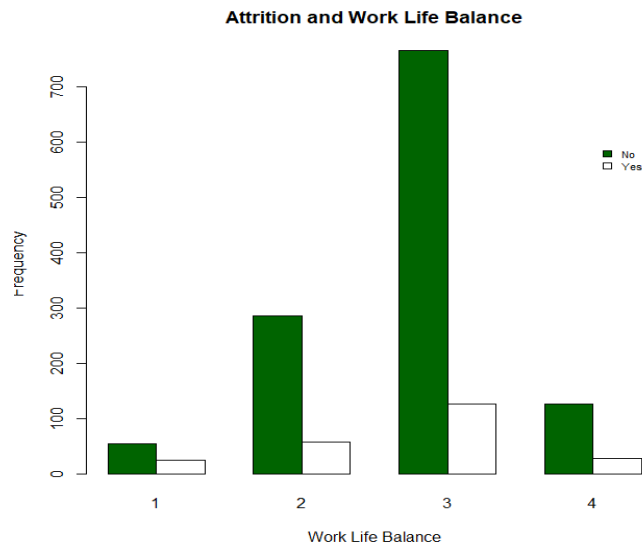
Above shows the summary breakdown of the measured variables from the data. It is interesting to note that the average age of the sample is closer to 37. That the average employee stays with their current manager is about 4 years. Also, it is interesting to note that Environmental, Relationship, and Job satisfaction have an average rank of 2.7, which can indicate that they affect each other. Work Life Balance is hovering around 2.8, which is only 0.1 point higher than the satisfaction ratings.



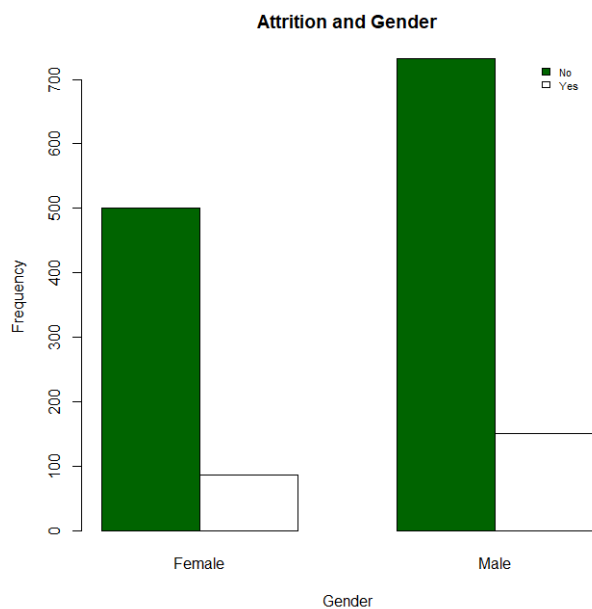
The image above shows that mainly individuals leave their jobs around their mid-30s, which is interesting because usually people are in the middle of their career path at that age. There could be some outside factors as well, but based on the graph that seems unlikely.



Based on the graph above, it looks like the people who rated their job satisfaction a 3 and have resigned is slightly more than the people that rated their job satisfaction a 1 and left.



The figure above shows the relationship between Attrition and Work Life Balance, and it is worth noting that the rank that shows the most people leaving are the ones that choose 3. Also, it should be noted that choosing 3 seems to be a popular answer for a lot of the variables that have a rank.



Gender		
Attrition	Female	Male
No	501	732
Yes	87	150

We can note from the graph and the table that there are substantially more males than there are females in the data set. About 15% of women vs 17% of men that have resigned. The percentage of men being

slightly higher can most likely be attributed to 294 more men than women in the data. Either way, this gives us an interesting look at how gender factors into the data set.

Analysis

Model Determination (we attached the R file alongside this report)

Many factors drive employees to quit their job for another. These factors based on this case study are, for example, work-life balance, job involvement, age, gender, environmental satisfaction, job satisfaction, relationship with managers, compensation (monthly income), etc. To further our analysis, we decided to create some models and compare them by applying some statistical techniques. The first technique in mind is linear regression. In this case scenario, the dependent variable is Attrition which happens to be a character with the form of “Yes” and “No” (they stayed or left). Before running a regression, we converted the Attrition variable into a dummy variable using the R programming function (see R script below):

```
Attrition.Y <-c(rep(0,1470))

for (i in 1:1470) {
  if(Attrition[i]=="Yes") Attrition.Y[i] <-1
}
Attrition.Y
```

We assigned 1 to the value “Yes”, which indicates an employee has resigned. To have reasonable parameter estimations in the regression model, all character variables in the dataset were converted into dummy variables. Then we combined the dummy variables with other variables using an R function called “cbind”. We ran three regression models to determine which one is the best model to use when discussing the results. The descriptive statistics showed a lot of skewed variables such as MonthlyIncome, DistanceFromHome, NumCompaniesWorked, PercentSalaryHike, TotalWorkingYears, etc. Due to this we determined that we would need to take the logarithm transformation of the highly skewed variables to normalize the distribution. However, the logarithm transformation is not always a good approach to improve a model because it can create some biases. We found that all the variables didn’t improve the significance of the variables, except for monthly income. Monthly has been shown as an important contributing factor. Therefore, taking the logarithm transformation of the variable MonthlyIncome would render the data less skewed or normalized. As we continue our analysis, we created a linear regression model using the Multi Linear Regression technique. In general, linear models are oftentimes easy to create and interpret, but can also be unreliable due using only binary variables.

Linear Regression: R function lm:

R output – Linear Model

```
Coefficients:
(Intercept)      1.868e+00  3.099e-01  6.026 2.13e-09 ***
Age             -3.401e-03  1.313e-03 -2.589 0.009710 **
TravelRare.dum   7.212e-02  2.848e-02  2.533 0.011431 *
TravelFreq.dum   1.603e-01  3.297e-02  4.861 1.29e-06 ***
Sales.dum        9.675e-02  6.328e-02  1.529 0.126519
Research.D.dum   -3.398e-03  6.073e-02 -0.056 0.955382
Other.dum        -3.388e-02  3.789e-02 -0.894 0.371277
Medical.dum      -2.782e-02  1.932e-02 -1.439 0.150229
Human.Resources.dum 9.505e-02  8.302e-02  1.145 0.252408
Marketing.dum     2.472e-02  3.289e-02  0.752 0.452320
DistanceFromHome 3.556e-03  1.045e-03  3.402 0.000688 ***
EnvironmentSatisfaction -4.145e-02  7.788e-03 -5.323 1.19e-07 ***
Gender.F         -3.613e-02  1.738e-02 -2.079 0.037825 *
JobInvolvement   -6.082e-02  1.196e-02 -5.086 4.14e-07 ***
JobLevel         4.514e-02  2.349e-02  1.921 0.054868 .
Lab.tech.dum     4.582e-02  3.811e-02  1.202 0.229447
Research.Scientist.dum -4.533e-02  3.770e-02 -1.203 0.229317
Sales.Executive.dum -5.243e-02  4.208e-02 -1.246 0.212919
Research.Director.dum 1.691e-02  5.230e-02  0.323 0.746532
Mgr.dum          9.566e-03  5.402e-02  0.177 0.859476
Mfg.Director.dum -1.477e-02  3.784e-02 -0.390 0.696427
JobSatisfaction  -3.849e-02  7.696e-03 -5.001 6.41e-07 ***
Single.dum       1.084e-01  3.130e-02  3.463 0.000550 ***
Married.dum      1.486e-02  2.292e-02  0.648 0.516947
log.MonthlyIncome -3.466e-01  8.347e-02 -4.152 3.49e-05 ***
HourlyRate       -1.894e-04  4.180e-04 -0.453 0.650528
DailyRate        -2.280e-05  2.114e-05 -1.079 0.280980
MonthlyRate      4.840e-07  1.191e-06  0.407 0.684415
NumCompaniesWorked 1.829e-02  3.802e-03  4.811 1.66e-06 ***
Overtime.Y       2.124e-01  1.894e-02  11.219 < 2e-16 ***
PercentSalaryHike -2.124e-03  3.668e-03 -0.579 0.562548
PerformanceRating 1.217e-02  3.714e-02  0.328 0.743110
RelationshipSatisfaction -2.401e-02  7.859e-03 -3.055 0.002289 **
StockOptionLevel -1.849e-02  1.359e-02 -1.360 0.173980
TotalWorkingYears -2.979e-03  2.396e-03 -1.243 0.213890
TrainingTimesLastYear -1.345e-02  6.624e-03 -2.031 0.042453 *
WorkLifeBalance  -2.960e-02  1.204e-02 -2.458 0.014071 *
YearsAtCompany    4.543e-03  2.981e-03  1.524 0.127781
YearsInCurrentRole -7.327e-03  3.886e-03 -1.885 0.059570 .
YearsSinceLastPromotion 1.154e-02  3.397e-03  3.398 0.000698 ***
YearswithCurrManager -9.024e-03  3.959e-03 -2.280 0.022782 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3216 on 1429 degrees of freedom
Multiple R-squared:  0.2565,    Adjusted R-squared:  0.2357 
F-statistic: 12.32 on 40 and 1429 DF,  p-value: < 2.2e-16
```

Estimations

We can see that the overall model has significant p-values < 0.05 but not all variables are significant. We can see that Age, business travel (TravelRarely and TravelFrequently) are significant which means that they are important factors that could have affected employee retention. DistanceFromHome significantly entails that employees prefer living close to their job rather than far away. Environment satisfaction is another factor that plays a role in employees' decision to quit their job or stay. Job involvement, Job satisfaction, Monthly Income, Number of Companies that Worked, and overtime are significant (probability values are less than 0.05). The adjusted R-square equals 0.2357 which is not great. To

develop a better linear model, we decided to do two stepwise regression to remove the insignificant variables and improve the model overall.

Linear Model – Stepwise Model 1:

Below is the first stepwise regression model that we ran.

```
Call:
lm(formula = Attrition.Y ~ Age + TravelRare.dum + TravelFreq.dum +
    DistanceFromHome + EnvironmentSatisfaction + Gender.F + JobInvolvement +
    JobSatisfaction + Single.dum + NumCompaniesWorked + log.MonthlyIncome +
    Overtime.Y + RelationshipSatisfaction + WorkLifeBalance +
    YearsInCurrentRole + YearsSinceLastPromotion + YearsWithCurrManager)

Residuals:
    Min       1Q   Median       3Q      Max
-0.55203 -0.20539 -0.09281  0.06686  1.22606

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      1.301228   0.137367   9.473 < 2e-16 ***
Age             -0.003901   0.001118  -3.489 0.000499 ***
TravelRare.dum    0.071254   0.028558   2.495 0.012702 *
TravelFreq.dum    0.161005   0.033119   4.861 1.29e-06 ***
DistanceFromHome  0.003583   0.001050   3.413 0.000659 ***
EnvironmentSatisfaction -0.041684  0.007804  -5.341 1.07e-07 ***
Gender.F         -0.038226   0.017438  -2.192 0.028526 *
JobInvolvement   -0.064467   0.012006  -5.370 9.17e-08 ***
JobSatisfaction   -0.039034   0.007740  -5.043 5.15e-07 ***
Single.dum       0.121415   0.018462   6.577 6.70e-11 ***
NumCompaniesWorked 0.015351   0.003669   4.184 3.03e-05 ***
log.MonthlyIncome -0.182085   0.036971  -4.925 9.40e-07 ***
Overtime.Y       0.210705   0.019020  11.078 < 2e-16 ***
RelationshipSatisfaction -0.022906  0.007918  -2.893 0.003874 **
WorkLifeBalance  -0.026703   0.012089  -2.209 0.027338 *
YearsInCurrentRole -0.006162   0.003593  -1.715 0.086524 .
YearsSinceLastPromotion 0.013678   0.003258   4.198 2.86e-05 ***
YearsWithCurrManager -0.007286   0.003530  -2.064 0.039206 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3255 on 1452 degrees of freedom
Multiple R-squared:  0.2261,    Adjusted R-squared:  0.2171
F-statistic: 24.96 on 17 and 1452 DF,  p-value: < 2.2e-16
```

We can see that most of the independent variables such as, Age, business travel (Travel Rarely and Travel Frequently), distance from home, monthly income, job satisfaction, marital status, work-life balance, etc. are all significant except the years in the current role that has a higher p-value of 0.08 greater than 0.05. Years currently do not play a crucial role in employees' decision to leave. Moreover, overall the variables are more significant than the prior model, but we do see a slight dip in the adjusted R squared. This means that while the variables are more significant, the model as a whole has become less significant.

Linear Model – Stepwise Model 2:

We ran a second stepwise linear regression to refine the model, in hopes that it would approve the model overall by increasing the adjusted r squared. (See the model below)

```
Call:
lm(formula = Attrition.Y ~ Age + TravelRare.dum + TravelFreq.dum +
    DistanceFromHome + EnvironmentSatisfaction + Gender.F + JobInvolvement +
    JobSatisfaction + Single.dum + NumCompaniesWorked + log.MonthlyIncome +
    Overtime.Y + RelationshipSatisfaction + WorkLifeBalance +
    YearsSinceLastPromotion + YearsWithCurrManager)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-0.55346 -0.20717 -0.09056  0.06638  1.22993
```

```
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      1.332484   0.136243   9.780 < 2e-16 ***
Age              -0.003907   0.001119  -3.492 0.000494 ***
TravelRare.dum    0.071160   0.028577   2.490 0.012880 *
TravelFreq.dum    0.161392   0.033140   4.870 1.24e-06 ***
DistanceFromHome  0.003563   0.001050   3.391 0.000714 ***
EnvironmentSatisfaction -0.042083  0.007806  -5.391 8.15e-08 ***
Gender.F         -0.038982   0.017444  -2.235 0.025587 *
JobInvolvement   -0.064561   0.012013  -5.374 8.95e-08 ***
JobSatisfaction  -0.039393   0.007742  -5.088 4.09e-07 ***
Single.dum       0.123199   0.018445   6.679 3.41e-11 ***
NumCompaniesWorked 0.015741   0.003664   4.296 1.86e-05 ***
log.MonthlyIncome -0.191754   0.036563  -5.244 1.80e-07 ***
Overtime.Y       0.211054   0.019032  11.090 < 2e-16 ***
RelationshipSatisfaction -0.022507  0.007920  -2.842 0.004547 **
WorkLifeBalance  -0.028052   0.012071  -2.324 0.020268 *
YearsSinceLastPromotion 0.012129  0.003133   3.871 0.000113 ***
YearsWithCurrManager -0.010697  0.002919  -3.665 0.000256 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.3257 on 1453 degrees of freedom
Multiple R-squared:  0.2246,    Adjusted R-squared:  0.216
F-statistic: 26.3 on 16 and 1453 DF,  p-value: < 2.2e-16
```

After omitting the variable yearsInCurrentRole, the output shows that all the independent variables are significant, but the Adjusted R-squared decreased again. Although this gives an interesting insight into the data, it is not the best model to use because it uses binary variables to determine, which is unreliable.

Logistics Regression Model 1:

We converted some independent variables into categorical variables to generate adequate results.

Call:

```
glm(formula = Attrition.Y ~ Age + BusinessTravel + Department +  
DistanceFromHome + EducationField + EnvironmentSatisfaction.cat +  
Gender + JobInvolvement.cat + JobLevel.cat + JobRole + JobSatisfaction.cat +  
MaritalStatus + log.MonthlyIncome + HourlyRate + DailyRate +  
MonthlyRate + NumCompaniesWorked + OverTime + PercentSalaryHike +  
PerformanceRating.cat + RelationshipSatisfaction.cat + StockOptionLevel +  
TotalWorkingYears + TrainingTimesLastYear + WorkLifeBalance.cat +  
YearsAtCompany + YearsInCurrentRole + YearsSinceLastPromotion +  
YearsWithCurrManager, family = binomial("logit"))
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.8175	-0.4424	-0.1914	-0.0534	3.5334

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.746e+00	5.958e+02	-0.003	0.997662
Age	-2.934e-02	1.425e-02	-2.059	0.039520 *
BusinessTravelTravel_Frequently	2.175e+00	4.526e-01	4.806	1.54e-06 ***
BusinessTravelTravel_Rarely	1.220e+00	4.174e-01	2.923	0.003469 **
DepartmentResearch & Development	1.456e+01	5.958e+02	0.024	0.980499
DepartmentSales	1.348e+01	5.958e+02	0.023	0.981945
DistanceFromHome	5.592e-02	1.158e-02	4.828	1.38e-06 ***
Education.cat2	1.455e-01	3.474e-01	0.419	0.675278
Education.cat3	1.339e-01	3.059e-01	0.438	0.661564
Education.cat4	1.847e-01	3.338e-01	0.553	0.580032
Education.cat5	2.439e-01	6.306e-01	0.387	0.698868
EducationFieldLife Sciences	-1.265e+00	8.808e-01	-1.436	0.151075
EducationFieldMarketing	-6.924e-01	9.240e-01	-0.749	0.453675
EducationFieldMedical	-1.208e+00	8.770e-01	-1.378	0.168342
EducationFieldOther	-1.128e+00	9.498e-01	-1.188	0.234950
EducationFieldTechnical Degree	-1.166e-01	8.922e-01	-0.131	0.896025
EnvironmentSatisfaction.cat2	-1.136e+00	2.919e-01	-3.892	9.93e-05 ***
EnvironmentSatisfaction.cat3	-1.290e+00	2.678e-01	-4.817	1.46e-06 ***
EnvironmentSatisfaction.cat4	-1.584e+00	2.747e-01	-5.766	8.13e-09 ***
GenderMale	4.617e-01	1.958e-01	2.358	0.018349 *
JobInvolvement.cat2	-1.324e+00	3.790e-01	-3.492	0.000479 ***
JobInvolvement.cat3	-1.655e+00	3.583e-01	-4.618	3.87e-06 ***
JobInvolvement.cat4	-2.189e+00	4.924e-01	-4.445	8.79e-06 ***
JobLevel.cat2	-1.332e+00	4.788e-01	-2.782	0.005410 **
JobLevel.cat3	5.051e-01	6.651e-01	0.759	0.447593
JobLevel.cat4	-5.716e-01	1.081e+00	-0.529	0.597068
JobLevel.cat5	2.397e+00	1.385e+00	1.731	0.083409 .
JobRoleHuman Resources	1.477e+01	5.958e+02	0.025	0.980215
JobRoleLaboratory Technician	3.366e-01	5.972e-01	0.564	0.573035
JobRoleManager	-2.384e-01	1.058e+00	-0.225	0.821675
JobRoleManufacturing Director	4.016e-01	5.574e-01	0.721	0.471179
JobRoleResearch Director	-1.960e+00	1.146e+00	-1.710	0.087270 .
JobRoleResearch Scientist	-7.557e-01	6.244e-01	-1.210	0.226198
JobRoleSales Executive	2.415e+00	1.288e+00	1.875	0.060815 .
JobRoleSales Representative	1.813e+00	1.364e+00	1.330	0.183669
JobSatisfaction.cat2	-6.424e-01	2.878e-01	-2.232	0.025629 *
JobSatisfaction.cat3	-6.367e-01	2.543e-01	-2.503	0.012306 *
JobSatisfaction.cat4	-1.333e+00	2.725e-01	-4.892	9.97e-07 ***
MaritalStatusMarried	4.743e-01	2.834e-01	1.674	0.094163 .
MaritalStatusSingle	1.425e+00	3.641e-01	3.912	9.14e-05 ***
log.MonthlyIncome	-2.866e+00	9.237e-01	-3.102	0.001921 **
HourlyRate	4.420e-03	4.718e-03	0.937	0.348822
DailyRate	-4.244e-04	2.339e-04	-1.815	0.069586 .
MonthlyRate	8.810e-06	1.318e-05	0.668	0.503899
NumCompaniesWorked	2.239e-01	4.151e-02	5.394	6.91e-08 ***
OverTimeYes	2.239e+00	2.124e-01	10.542	< 2e-16 ***
PercentSalaryHike	-1.674e-02	4.146e-02	-0.404	0.686440
PerformanceRating.cat4	4.634e-02	4.244e-01	0.109	0.913048
RelationshipSatisfaction.cat2	-9.690e-01	3.018e-01	-3.211	0.001321 **
RelationshipSatisfaction.cat3	-1.025e+00	2.681e-01	-3.824	0.000131 ***
RelationshipSatisfaction.cat4	-1.044e+00	2.684e-01	-3.888	0.000101 ***
StockOptionLevel	-1.481e-01	1.617e-01	-0.916	0.359733
TotalWorkingYears	-5.370e-02	3.040e-02	-1.766	0.077387 .
TrainingTimesLastYear	-1.908e-01	7.602e-02	-2.510	0.012079 *
WorkLifeBalance.cat2	-1.006e+00	3.879e-01	-2.595	0.009470 **
WorkLifeBalance.cat3	-1.447e+00	3.625e-01	-3.991	6.57e-05 ***
WorkLifeBalance.cat4	-1.011e+00	4.384e-01	-2.307	0.021079 *
YearsAtCompany	9.457e-02	4.220e-02	2.241	0.025026 *
YearsInCurrentRole	-1.253e-01	5.103e-02	-2.455	0.014080 *
YearsSinceLastPromotion	1.710e-01	4.523e-02	3.781	0.000156 ***
YearsWithCurrManager	-1.478e-01	4.981e-02	-2.967	0.003011 **

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

The R output in figure above shows the logistics regression model 1 results. We see that TrainingTimesLastYear, WorkingLifeBalance, YearsSinceLastPromotion, YearswithCurrentManager are significant. As well as Age, business travel, distance from home, job involvement, job satisfaction, and environmental satisfaction, and so on are also significant at a 95% confidence interval with low p-values less than 0.05. The logistic regression technique provides important results such as Residual deviance and AIC. Both statistical results play a vital role when comparing models. AIC of the overall logistics model 1 equals 902.96 with Residual deviance equal to 780.96. As we can see, there are variables that are insignificant and need to be removed.

Logistics Regression Model 2:

We ran a second logistic model 2 to improve the model

```
Call:
glm(formula = Attrition.Y ~ Age + BusinessTravel + DistanceFromHome +
    EnvironmentSatisfaction.cat + Gender + JobInvolvement.cat +
    log.MonthlyIncome + JobSatisfaction.cat + Single.dum + NumCompaniesWorked +
    OverTime + RelationshipSatisfaction.cat + WorkLifeBalance.cat +
    TrainingTimesLastYear + YearsInCurrentRole + YearsSinceLastPromotion +
    YearsWithCurrManager, family = binomial("logit"))
```

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.7993  -0.5042  -0.2713  -0.1111   3.9689
```

```
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    9.26707    1.53226   6.048 1.47e-09 ***
Age           -0.04058    0.01153  -3.519 0.000434 ***
BusinessTravelTravel_Frequently  1.93351    0.40765   4.743 2.11e-06 ***
BusinessTravelTravel_Rarely     1.11140    0.37963   2.928 0.003416 **
DistanceFromHome    0.04479    0.01040   4.306 1.67e-05 ***
EnvironmentSatisfaction.cat2    -1.00744    0.26643  -3.781 0.000156 ***
EnvironmentSatisfaction.cat3    -1.04329    0.24154  -4.319 1.56e-05 ***
EnvironmentSatisfaction.cat4    -1.34138    0.24715  -5.427 5.72e-08 ***
GenderMale         0.41330    0.17988   2.298 0.021583 *
JobInvolvement.cat2    -1.13051    0.34470  -3.280 0.001039 **
JobInvolvement.cat3    -1.47083    0.32288  -4.555 5.23e-06 ***
JobInvolvement.cat4    -2.16120    0.45537  -4.746 2.07e-06 ***
log.MonthlyIncome    -2.12213    0.41310  -5.137 2.79e-07 ***
JobSatisfaction.cat2    -0.55118    0.26299  -2.096 0.036099 *
JobSatisfaction.cat3    -0.64334    0.23107  -2.784 0.005366 **
JobSatisfaction.cat4    -1.36035    0.25333  -5.370 7.88e-08 ***
Single.dum         1.16995    0.18114   6.459 1.06e-10 ***
NumCompaniesWorked    0.15633    0.03585   4.361 1.29e-05 ***
OverTimeYes        1.88116    0.18663  10.080 < 2e-16 ***
RelationshipSatisfaction.cat2    -0.84606    0.27355  -3.093 0.001982 **
RelationshipSatisfaction.cat3    -0.71265    0.24165  -2.949 0.003187 **
RelationshipSatisfaction.cat4    -0.89983    0.24897  -3.614 0.000301 ***
WorkLifeBalance.cat2    -0.92488    0.34918  -2.649 0.008079 **
WorkLifeBalance.cat3    -1.21062    0.32573  -3.717 0.000202 ***
WorkLifeBalance.cat4    -0.72621    0.39899  -1.820 0.068744 .
TrainingTimesLastYear    -0.17000    0.07065  -2.406 0.016119 *
YearsInCurrentRole    -0.08090    0.04181  -1.935 0.052982 .
YearsSinceLastPromotion    0.20262    0.03828   5.293 1.20e-07 ***
YearsWithCurrManager    -0.11715    0.04170  -2.809 0.004962 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1298.58  on 1469  degrees of freedom
Residual deviance:  892.02  on 1441  degrees of freedom
AIC: 950.02
```

The figure above shows that most of the independent variables have significant probability values below 5%. YearsInCurrentRole does not play an urge in this case study and it shows no significance. We kept Work-Life Balance.cat4 because work life balance plays an important role overall and felt that removing it would not provide an accurate representation. The residual deviance for this model equals 892.02 which is greater than the previous residual deviance in model 1. The AIC of this model 2 is 950.02 which is also greater than 902.96 in model 1. It would be difficult to conclude that logistics regression model 1 is better based on its AIC and residual deviance indicate a better fit, especially when there were quite a few variables that were insignificant.

Logistics Regression Model 3:

We ran the third model to remove the variable YearsInCurrentRole. See R output in the figure below.

```

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    9.64166    1.52543   6.321 2.61e-10 ***
Age           -0.04142    0.01154  -3.591 0.000330 ***
BusinessTravelTravel_Frequently  1.91101    0.40515   4.717 2.40e-06 ***
BusinessTravelTravel_Rarely     1.07443    0.37693   2.850 0.004365 **
DistanceFromHome    0.04445    0.01034   4.297 1.73e-05 ***
EnvironmentSatisfaction.cat2    -1.03633    0.26587  -3.898 9.70e-05 ***
EnvironmentSatisfaction.cat3    -1.06429    0.24082  -4.419 9.90e-06 ***
EnvironmentSatisfaction.cat4    -1.35816    0.24664  -5.507 3.66e-08 ***
GenderMale         0.42056    0.17957   2.342 0.019175 *
JobInvolvement.cat2    -1.14021    0.34354  -3.319 0.000903 ***
JobInvolvement.cat3    -1.47465    0.32156  -4.586 4.52e-06 ***
JobInvolvement.cat4    -2.15192    0.45381  -4.742 2.12e-06 ***
log.MonthlyIncome    -2.23274    0.41016  -5.444 5.22e-08 ***
JobSatisfaction.cat2    -0.53674    0.26271  -2.043 0.041045 *
JobSatisfaction.cat3    -0.62628    0.23031  -2.719 0.006542 **
JobSatisfaction.cat4    -1.36076    0.25310  -5.376 7.60e-08 ***
Single.dum         1.17796    0.18063   6.522 6.96e-11 ***
NumCompaniesWorked    0.16102    0.03560   4.523 6.10e-06 ***
OverTimeYes        1.87777    0.18601  10.095 < 2e-16 ***
RelationshipSatisfaction.cat2    -0.84053    0.27302  -3.079 0.002079 **
RelationshipSatisfaction.cat3    -0.70601    0.24112  -2.928 0.003410 **
RelationshipSatisfaction.cat4    -0.87627    0.24806  -3.533 0.000412 ***
WorkLifeBalance.cat2    -0.95372    0.34829  -2.738 0.006176 **
WorkLifeBalance.cat3    -1.24547    0.32444  -3.839 0.000124 ***
WorkLifeBalance.cat4    -0.75506    0.39752  -1.899 0.057506 .
TrainingTimesLastYear    -0.16847    0.07049  -2.390 0.016851 *
YearsSinceLastPromotion    0.18073    0.03587   5.038 4.70e-07 ***
YearsWithCurrManager    -0.15764    0.03596  -4.384 1.16e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1298.58  on 1469  degrees of freedom
Residual deviance:  895.81  on 1442  degrees of freedom
AIC: 951.81

```

Most of the variables are found to be significant and it's worth noting that WorkLifeBalance.cat4 became more significant. The AIC = 951.81 is greater than AIC in model 2 as well as the Residual deviance = 895.81 > 892.02.

Probit model 1: see R output below

We ran three logistic regression using the probit function.

Coefficients:					
	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	1.894e+00	8.691e+01	0.022	0.982613	
Age	-1.407e-02	7.730e-03	-1.820	0.068727	.
BusinessTravelTravel_Frequently	1.153e+00	2.368e-01	4.867	1.13e-06	***
BusinessTravelTravel_Rarely	6.620e-01	2.181e-01	3.035	0.002404	**
DepartmentResearch & Development	4.660e+00	8.689e+01	0.054	0.957231	
DepartmentSales	4.241e+00	8.689e+01	0.049	0.961071	
DistanceFromHome	2.858e-02	6.218e-03	4.597	4.28e-06	***
Education.cat2	6.766e-02	1.875e-01	0.361	0.718273	
Education.cat3	6.812e-02	1.655e-01	0.412	0.680586	
Education.cat4	8.932e-02	1.799e-01	0.497	0.619444	
Education.cat5	1.693e-01	3.329e-01	0.509	0.611017	
EducationFieldLife Sciences	-7.117e-01	4.767e-01	-1.493	0.135425	
EducationFieldMarketing	-4.504e-01	5.016e-01	-0.898	0.369285	
EducationFieldMedical	-6.648e-01	4.753e-01	-1.399	0.161893	
EducationFieldOther	-6.157e-01	5.140e-01	-1.198	0.231003	
EducationFieldTechnical Degree	-5.702e-02	4.863e-01	-0.117	0.906668	
EnvironmentSatisfaction.cat2	-6.231e-01	1.571e-01	-3.967	7.28e-05	***
EnvironmentSatisfaction.cat3	-6.832e-01	1.433e-01	-4.769	1.85e-06	***
EnvironmentSatisfaction.cat4	-8.289e-01	1.467e-01	-5.652	1.59e-08	***
GenderMale	2.353e-01	1.053e-01	2.234	0.025493	*
JobInvolvement.cat2	-7.444e-01	2.096e-01	-3.552	0.000382	***
JobInvolvement.cat3	-9.010e-01	1.976e-01	-4.561	5.10e-06	***
JobInvolvement.cat4	-1.198e+00	2.666e-01	-4.493	7.02e-06	***
JobLevel.cat2	-6.000e-01	2.467e-01	-2.432	0.014999	*
JobLevel.cat3	3.968e-01	3.520e-01	1.127	0.259677	
JobLevel.cat4	-2.122e-01	5.683e-01	-0.373	0.708915	
JobLevel.cat5	1.465e+00	7.275e-01	2.014	0.043974	*
JobRoleHuman Resources	4.734e+00	8.689e+01	0.054	0.956549	
JobRoleLaboratory Technician	2.319e-01	3.031e-01	0.765	0.444154	
JobRoleManager	-3.668e-01	5.641e-01	-0.650	0.515581	
JobRoleManufacturing Director	2.070e-01	2.826e-01	0.733	0.463787	
JobRoleResearch Director	-1.136e+00	5.896e-01	-1.927	0.053968	.
JobRoleResearch Scientist	-3.441e-01	3.154e-01	-1.091	0.275179	
JobRoleSales Executive	1.117e+00	6.730e-01	1.660	0.096895	.
JobRoleSales Representative	9.388e-01	7.134e-01	1.316	0.188198	
JobSatisfaction.cat2	-3.276e-01	1.560e-01	-2.100	0.035700	*
JobSatisfaction.cat3	-3.692e-01	1.386e-01	-2.663	0.007739	**
JobSatisfaction.cat4	-7.107e-01	1.458e-01	-4.875	1.09e-06	***
MaritalStatusMarried	3.171e-01	1.526e-01	2.077	0.037787	*
MaritalStatusSingle	7.947e-01	1.970e-01	4.034	5.48e-05	***
log.MonthlyIncome	-1.476e+00	5.030e-01	-2.935	0.003337	**
HourlyRate	2.257e-03	2.543e-03	0.888	0.374766	
DailyRate	-2.327e-04	1.263e-04	-1.843	0.065326	.

MonthlyRate	3.870e-06	7.095e-06	0.545	0.585424	
NumCompaniesWorked	1.212e-01	2.239e-02	5.414	6.15e-08	***
OverTimeYes	1.162e+00	1.111e-01	10.467	< 2e-16	***
PercentSalaryHike	-1.198e-02	2.221e-02	-0.539	0.589766	
PerformanceRating.cat4	7.038e-02	2.261e-01	0.311	0.755629	
RelationshipSatisfaction.cat2	-5.051e-01	1.626e-01	-3.106	0.001896	**
RelationshipSatisfaction.cat3	-5.389e-01	1.452e-01	-3.711	0.000206	***
RelationshipSatisfaction.cat4	-5.139e-01	1.441e-01	-3.566	0.000362	***
StockOptionLevel	-6.649e-02	8.682e-02	-0.766	0.443796	
TotalWorkingYears	-3.224e-02	1.614e-02	-1.998	0.045766	*
TrainingTimesLastYear	-1.038e-01	4.094e-02	-2.537	0.011188	*
WorkLifeBalance.cat2	-5.247e-01	2.126e-01	-2.468	0.013569	*
WorkLifeBalance.cat3	-7.572e-01	1.989e-01	-3.806	0.000141	***
WorkLifeBalance.cat4	-4.810e-01	2.406e-01	-1.999	0.045601	*
YearsAtCompany	6.534e-02	2.161e-02	3.023	0.002499	**
YearsInCurrentRole	-7.160e-02	2.691e-02	-2.660	0.007808	**
YearsSinceLastPromotion	8.340e-02	2.345e-02	3.556	0.000376	***
YearsWithCurrManager	-8.373e-02	2.598e-02	-3.224	0.001266	**

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1298.58 on 1469 degrees of freedom
 Residual deviance: 796.74 on 1409 degrees of freedom
 AIC: 918.74

The probit model 1 above shows different results from the logistic regression model explained previously. The AIC = 918.74 and Residual deviance = 796.74. In logistics models 1, 2, and 3, Age was significant. With the probit, model Age is not significant with a p-value = 0.98. Does it mean that Age is not an important factor in retention? No conclusion can be made at this point until we choose the best model.

Probit model 2: see R output below.


```

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    3.653545   0.922312   3.961 7.45e-05 ***
BusinessTravelTravel_Frequently  1.016082   0.212426   4.783 1.73e-06 ***
BusinessTravelTravel_Rarely      0.601828   0.196777   3.058 0.002225 **
DistanceFromHome    0.022648   0.005574   4.064 4.83e-05 ***
EnvironmentSatisfaction.cat2    -0.570079   0.144234  -3.952 7.74e-05 ***
EnvironmentSatisfaction.cat3    -0.559158   0.129194  -4.328 1.50e-05 ***
EnvironmentSatisfaction.cat4    -0.709114   0.131986  -5.373 7.76e-08 ***
GenderMale         0.187889   0.096189   1.953 0.050780 .
JobInvolvement.cat2    -0.665306   0.189801  -3.505 0.000456 ***
JobInvolvement.cat3    -0.877874   0.178836  -4.909 9.16e-07 ***
JobInvolvement.cat4    -1.218112   0.243231  -5.008 5.50e-07 ***
JobSatisfaction.cat2    -0.313411   0.142550  -2.199 0.027906 *
JobSatisfaction.cat3    -0.372498   0.126570  -2.943 0.003250 **
JobSatisfaction.cat4    -0.671235   0.133403  -5.032 4.86e-07 ***
MaritalStatusMarried    0.254433   0.134377   1.893 0.058301 .
MaritalStatusSingle    0.786255   0.136465   5.762 8.33e-09 ***
log.MonthlyIncome    -0.968666   0.249280  -3.886 0.000102 ***
NumCompaniesWorked    0.073357   0.019498   3.762 0.000168 ***
OverTimeYes         0.967771   0.098649   9.810 < 2e-16 ***
RelationshipSatisfaction.cat2    -0.421034   0.146690  -2.870 0.004102 **
RelationshipSatisfaction.cat3    -0.384494   0.131189  -2.931 0.003380 **
RelationshipSatisfaction.cat4    -0.433764   0.132756  -3.267 0.001085 **
WorkLifeBalance.cat2    -0.472773   0.192846  -2.452 0.014224 *
WorkLifeBalance.cat3    -0.620492   0.179490  -3.457 0.000546 ***
WorkLifeBalance.cat4    -0.353867   0.218495  -1.620 0.105326
TotalWorkingYears    -0.023376   0.010911  -2.142 0.032153 *
TrainingTimesLastYear    -0.088566   0.037718  -2.348 0.018867 *
YearsInCurrentRole    -0.032625   0.022111  -1.476 0.140073
YearsSinceLastPromotion    0.102747   0.020216   5.083 3.72e-07 ***
YearsWithCurrManager    -0.053560   0.022134  -2.420 0.015531 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1298.58  on 1469  degrees of freedom
Residual deviance:  915.86  on 1440  degrees of freedom
AIC: 975.86

```

We notice that gender male is barely insignificant and we chose to keep it because it might play a role. We felt the same way regarding marital status pertaining to the value “married”. Work-life balance cat 4 and year in the current role are not significant. The AIC and Residual deviance on the other hand increase. To continue improving the model, we removed the non-significant variables in probit model 3.

Probit model 3: see R output below.

We used the GLM in R to create a logistic regression with a probit function. Attrition is the dependent variable but age was removed from the model.

```

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)      3.817030    0.912708   4.182 2.89e-05 ***
BusinessTravelTravel_Frequently  1.012161    0.209875   4.823 1.42e-06 ***
BusinessTravelTravel_Rarely      0.601750    0.194063   3.101 0.001930 **
DistanceFromHome    0.022724    0.005555   4.091 4.30e-05 ***
EnvironmentSatisfaction.cat2    -0.564382    0.143311  -3.938 8.21e-05 ***
EnvironmentSatisfaction.cat3    -0.562777    0.128695  -4.373 1.23e-05 ***
EnvironmentSatisfaction.cat4    -0.705678    0.131401  -5.370 7.86e-08 ***
GenderMale          0.189676    0.095896   1.978 0.047936 *
JobInvolvement.cat2    -0.649591    0.188452  -3.447 0.000567 ***
JobInvolvement.cat3    -0.862754    0.177354  -4.865 1.15e-06 ***
JobInvolvement.cat4    -1.191022    0.241429  -4.933 8.09e-07 ***
JobSatisfaction.cat2    -0.307704    0.142137  -2.165 0.030400 *
JobSatisfaction.cat3    -0.365751    0.125928  -2.904 0.003679 **
JobSatisfaction.cat4    -0.668886    0.132927  -5.032 4.85e-07 ***
Single.dum          0.611002    0.096218   6.350 2.15e-10 ***
log.MonthlyIncome    -0.972728    0.246937  -3.939 8.18e-05 ***
NumCompaniesWorked    0.073243    0.019302   3.795 0.000148 ***
OverTimeYes          0.962870    0.098150   9.810 < 2e-16 ***
RelationshipSatisfaction.cat2    -0.432324    0.146050  -2.960 0.003075 **
RelationshipSatisfaction.cat3    -0.386210    0.130571  -2.958 0.003098 **
RelationshipSatisfaction.cat4    -0.428124    0.132279  -3.237 0.001210 **
WorkLifeBalance.cat2    -0.492527    0.192261  -2.562 0.010414 *
WorkLifeBalance.cat3    -0.644105    0.178857  -3.601 0.000317 ***
WorkLifeBalance.cat4    -0.366962    0.217376  -1.688 0.091383 .
TotalWorkingYears    -0.025597    0.010861  -2.357 0.018433 *
TrainingTimesLastYear    -0.086041    0.037534  -2.292 0.021884 *
YearsSinceLastPromotion    0.093880    0.019090   4.918 8.75e-07 ***
YearsWithCurrManager    -0.068654    0.019087  -3.597 0.000322 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1298.58  on 1469  degrees of freedom
Residual deviance:  921.51  on 1442  degrees of freedom
AIC: 977.51

```

After removing some insignificant variables from the probit model 2, we notice that the AIC and Residual deviance did not decrease. The model is better and Gender has become more significant.

Results

R output – Model comparison

```

> glance(hrdata.SW2) # MLR model 2
# A tibble: 1 x 12
  r.squared adj.r.squared sigma statistic p.value    df logLik   AIC    BIC deviance df.residual  nobs
  <dbl>      <dbl>      <dbl>      <dbl>    <dbl>   <dbl> <dbl> <dbl> <dbl>      <int>    <int>
1    0.225      0.216  0.326      26.3 3.73e-69    16 -428.  893.  988.    154.    1453  1470
> glance(hrdata.glm3) # Logistics reg with logit model 3
# A tibble: 1 x 8
  null.deviance df.null logLik   AIC   BIC deviance df.residual  nobs
  <dbl>      <int> <dbl> <dbl> <dbl> <dbl>      <int>    <int>
1    1299.    1469 -448.  952. 1100.   896.    1442  1470
> glance(hrdata.probit3) # Logistics reg with probit model 3
# A tibble: 1 x 8
  null.deviance df.null logLik   AIC   BIC deviance df.residual  nobs
  <dbl>      <int> <dbl> <dbl> <dbl> <dbl>      <int>    <int>
1    1299.    1469 -461.  978. 1126.   922.    1442  1470
> |

```

In each regression model approach we ran, we picked a good model that we could use to compare to other ones. We chose the stepwise linear regression model 2, logistics regression model 3 with logit function, and the probit regression model 3. We used the R function called “glance” for model comparison. The “glance” function provides various statistical results such as R-square, Log-likelihood, AIC, BIC, and Residual deviance. As mentioned at the beginning of this analysis, the linear regression model was not a reliable technique for this data. We focused on using Lower AIC and Residual deviance to determine which model would be the best model to use in our analysis. Using “glance” we can determine that the logistic regression model 3 is the best model to use because it has the lower AIC = 952 and Residual deviance=896 compared to the probit regression model 3, which had an AIC = 978 and Residual deviance=922. In conclusion, our best model for this study is the logistic regression model 3.

Conclusion

The introduction touched on why people leave in general and the analysis was about why people are leaving IBM. Based on the results we can conclude that some significant variables could be attributed to people resigning, and those are the variables we need to focus on when we strategize on how to retain the employees. I am going to briefly discuss some of the variables that are significant. Here are some of the significant variables that might play a role in people leaving that can help us figure out how to retain employees.

Age on its own without any context, it could be inferred that people in their 20s will leave for the betterment of their career or people retiring. Due to the graph that we shared earlier, we know that isn't the case. Employees are leaving in the middle of their careers.

Business travel was significant in all 3 regression models we ran. It would suggest that most people tend to look for stability. It is understandable that traveling no matter how much can take its toll on any individual.

Gender was shown as significant which may infer that the environment may not be as inclusive as was thought and we should look into this to see why this is a factor

Environmental, Job, and Relationship Satisfaction can play a huge role in how people feel about their job. The better the culture/environment the more productive the company, the opposite can also hold as well.

If an employee doesn't feel fulfilled in their job due to unappreciation, lack of training, and not understanding the expectations that are set from the beginning.

Years with the current manager affect job satisfaction, relationship satisfaction, and environmental satisfaction. The manager can play a huge role in whether an employee is dissatisfied and leaves a job.

Work-life balance was shown to be significant, which makes sense because overtime was also shown as a factor. We can infer that most employees do not feel like they can maintain a work-life balance.

Recommendations

“According to the U.S. Bureau of Statistics, the average turnover rate in the U.S. is about 12% to 15% annually...According to LinkedIn, an average annual worldwide employee turnover rate is 10.9%” (Zojceska, 2020). Based on this information, we can conclude that 19.1% Attrition rate in this data is quite high. For a global company, IBM has about double the Attrition rate of the worldwide average. These are some recommendations on how we can increase our retention and lower our attrition rate

Improving onboarding and training, so that newly hired employees feel appreciated from day 1. People feel invested when there is a seamless onboarding, and they feel like the company is invested in their success when a training schedule is established. It is very important to set up check-ins every 30 days, 60 days, 90 days, and a year. On those check-ins, it is important to reinforce the expectations that were set at hire and discuss what training is needed to help meet those expectations.

Employee satisfaction in areas of relationships with colleagues and managers, the company environment, and ultimately their job can be very important. We can start by changing the company environment by looking into whether the environment is inclusive and equitable based on gender being a significant variable of people leaving. Also, it is important to train managers to better handle questions/concerns by employees. Emotional Intelligence is great training, especially for managers who have issues with handling questions/concerns. Training is a great way to improve employees' knowledge and lead to growth. Also, providing training is a good way for employees to stay on top of changes in the world of technology.

Improving work-life balance is a great way to make employees feel important. While it is understandable that there might be some times that there will be overtime that should not be the norm. Bolster the

employee assistance program (EAP) by adding resources for employees to access for their mental and physical wellness. Making flexibility an option by either offering 100% remote and/or hybrid scheduling. It allows employees to be able to work on a schedule that works best for them and be more productive.

Overall, we think it would be best to regather the data from a much bigger sample size, so we can rerun the models to see if we get the same results.

Discussion

The implications to IBM will depend on the recommendations that will be implemented. If they are implemented, we foresee a positive impact on the employees' satisfaction level and in turn the company's bottom line. For some recommendations, there may be some money that will need to be invested. This can affect the company's profit short term, but improve it long term because employees productivity is expected to increase.

We did recommend regathering the data because we wanted to see if we get the same results with a bigger sample size, which would help us see if there was implicit bias when the data was gathered initially. We noticed that not all roles, such as software engineers, were included in the data. This could affect the significant variables. IBM is a huge global company and we are not sure where this sample was taken and if that affected how the company is viewed.

We would need to take into account a few things such as the impact of covid and diversity, equity, and inclusion.

Let's first discuss how covid has impacted the job market and what that means for IBM. "'One result of the pandemic has been an accelerated rate of technological adoption to remain competitive, and therefore a need for learning skills,' Falzon says" (Liu, 2021). We have seen a push for working remotely during the pandemic, which introduced all of us to new ways to connect with our colleagues and teammates. We needed to learn new skills to stay connected and maintain similar productivity. "Workers who want to quit overwhelmingly say they're looking for a new job with more flexibility. Indeed, even among those who aren't considering changing jobs, half of people currently working remotely say if their current company doesn't continue to offer remote-work options long-term, they'll look for a job at a company that does" (Liu, 2021). We found that people have become accustomed to working remotely and proved

not only to their employers but themselves that they could be productive working from home. Overall, it has made employees think about what is important in a job.

We, also, saw that something was lacking in terms of factoring in variables such as ethnicity. To understand if the company environment is inclusive, equitable, and diverse, we need to do a better job of collecting data that includes those variables and to give other options other than “Female” and “Male” to include employees that identify as non-binary.

References

8 Essential Employee Retention Factors Modern Employers Ignore. (2019, June 13). 8 Essential Employee Retention Factors Modern Employers Ignore.

<https://risepeople.com/blog/employee-retention-factors/>

Heathfield, S. M. H. (2021, February 24). *10 Critical Reasons Why People Quit Their Jobs.* The Balance Careers. <https://www.thebalancecareers.com/top-reasons-why-employees-quit-their-job-1918985>

Leonardi, P. L. (2018, October 23). *Better People Analytics: Measure Who They Know, Not Just Who They Are.* Harvard Business Review. <https://hbr.org/2018/11/better-people-analytics>

Liu, J. (2021, April 20). *1 in 4 workers is considering quitting their job after the pandemic—here's why.* CNBC. <https://www.cnbc.com/2021/04/19/1-in-4-workers-is-considering-quitting-their-job-after-the-pandemic.html>

Zojceska, A. (2020, April 3). *HR Metrics: How and Why to Calculate Employee Turnover Rate?* Blog. <https://www.talentlyft.com/en/blog/article/242/hr-metrics-how-and-why-to-calculate-employee-turnover-rate#:~:text=Average%20employee%20turnover%20rate&text=According%20to%20the%20U.S.%20Bureau,employee%20turnover%20rate%20is%2010.9%25>.

Appendix

Variable	Variable Definition
EmployeeNumber	Using a number instead of names to keep the data anonymous
Age	Age of Employee
Attrition	Whether they stayed or left
BusinessTravel	How often they travelled for work
DailyRate	What the daily rate is per employee
Department	What the department is per employee
DistanceFromHome	What the distance from home is per employee
Education	What the education is per employee
EducationField	What the education field is per employee
EnvironmentSatisfaction	What the environment satisfaction is per employee
Gender	What the gender is per employee
HourlyRate	What the hourly rate is per employee
JobInvolvement	What the job involvement is per employee
JobLevel	What the job level is per employee
JobRole	What the job role is per employee
JobSatisfaction	What the job satisfaction is per employee
MaritalStatus	What the marital status is per employee
MonthlyIncome	What the monthly income is per employee
YearlyIncome	What the yearly income is per employee
MonthlyRate	What the monthly rate is per employee
NumCompaniesWorked	The number of companies worked per employee
OverTime	What the over time is per employee
PercentSalaryHike	What the percent salary hike is per employee
PerformanceRating	What the performance rating is per employee
RelationshipSatisfaction	What the relationship satisfaction is per employee
StockOptionLevel	What the stock option level is per employee
TotalWorkingYears	What the total working years is per employee
TrainingTimesLastYear	Training times last year per employee
WorkLifeBalance	What the work life balance is per employee
YearsAtCompany	Years at company is per employee
YearsInCurrentRole	Years in current role is per employee
YearsSinceLastPromotion	Years since last promotion is per employee
YearsWithCurrManager	Years with current manager is per employee

Rating Definition

Education

- 1 'Below College'
- 2 'College'
- 3 'Bachelor'
- 4 'Master'
- 5 'Doctor'

PerformanceRating

- 1 'Low'
- 2 'Good'
- 3 'Excellent'
- 4 'Outstanding'

EnvironmentSatisfaction

- 1 'Low'
- 2 'Medium'
- 3 'High'
- 4 'Very High'

RelationshipSatisfaction

- 1 'Low'
- 2 'Medium'
- 3 'High'
- 4 'Very High'

JobInvolvement

- 1 'Low'
- 2 'Medium'
- 3 'High'
- 4 'Very High'

WorkLifeBalance

- 1 'Bad'
- 2 'Good'
- 3 'Better'
- 4 'Best'

JobSatisfaction

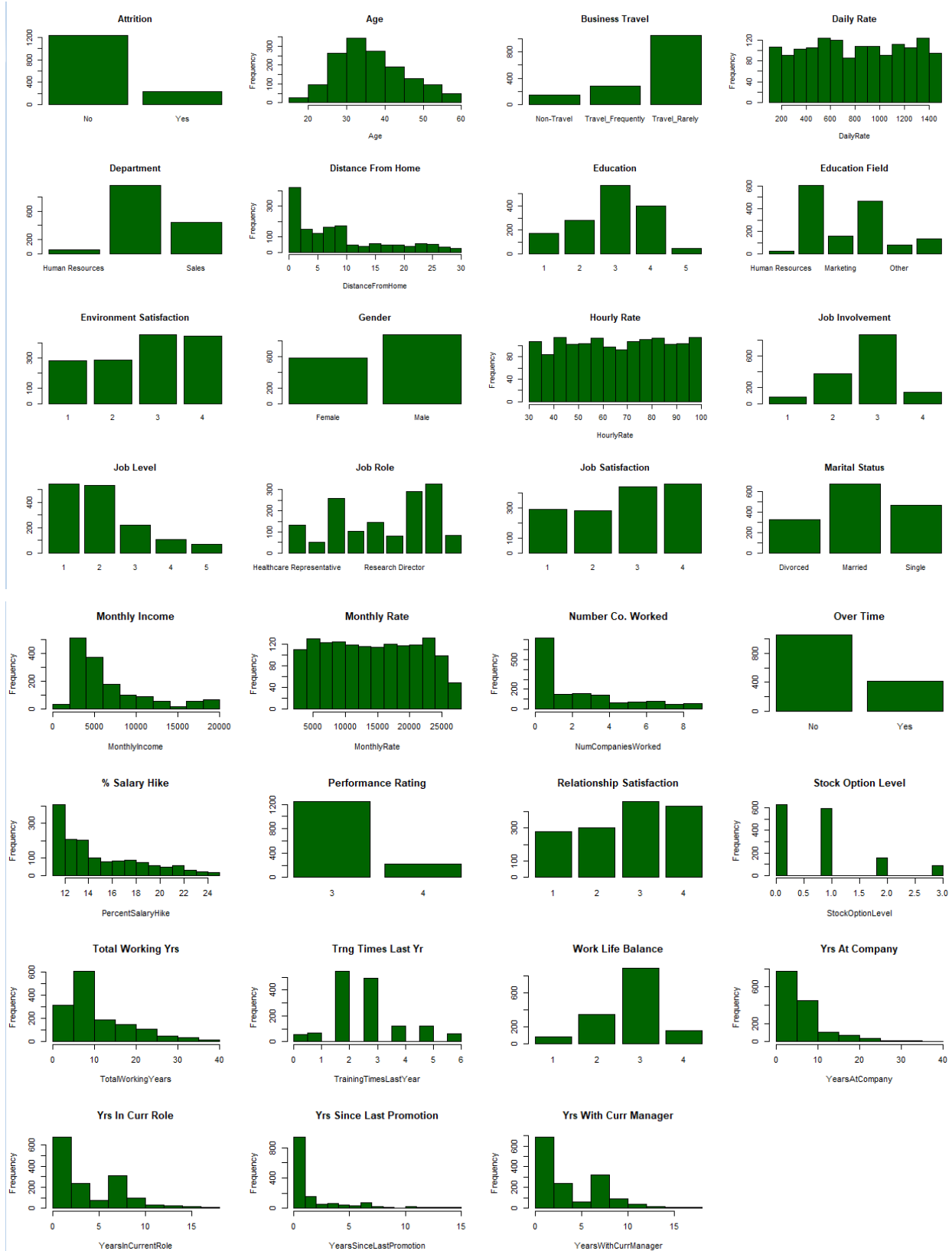
- 1 'Low'
- 2 'Medium'
- 3 'High'
- 4 'Very High'

□

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis
Age	1	1470	36.92	9.14	36.0	36.47	8.90	18	60	42	0.41	-0.41
Attrition*	2	1470	1.16	0.37	1.0	1.08	0.00	1	2	1	1.84	1.39
BusinessTravel*	3	1470	2.61	0.67	3.0	2.76	0.00	1	3	2	-1.44	0.69
DailyRate	4	1470	802.49	403.51	802.0	803.83	510.01	102	1499	1397	0.00	-1.21
Department*	5	1470	2.26	0.53	2.0	2.25	0.00	1	3	2	0.17	-0.40
DistanceFromHome	6	1470	9.19	8.11	7.0	8.08	7.41	1	29	28	0.96	-0.23
Education	7	1470	2.91	1.02	3.0	2.98	1.48	1	5	4	-0.29	-0.56
EducationField*	8	1470	3.25	1.33	3.0	3.10	1.48	1	6	5	0.55	-0.69
EnvironmentSatisfaction	9	1470	2.72	1.09	3.0	2.78	1.48	1	4	3	-0.32	-1.20
Gender*	10	1470	1.60	0.49	2.0	1.62	0.00	1	2	1	-0.41	-1.83
HourlyRate	11	1470	65.89	20.33	66.0	66.02	26.69	30	100	70	-0.03	-1.20
JobInvolvement	12	1470	2.73	0.71	3.0	2.74	0.00	1	4	3	-0.50	0.26
JobLevel	13	1470	2.06	1.11	2.0	1.90	1.48	1	5	4	1.02	0.39
JobRole*	14	1470	5.46	2.46	6.0	5.61	2.97	1	9	8	-0.36	-1.20
JobSatisfaction	15	1470	2.73	1.10	3.0	2.79	1.48	1	4	3	-0.33	-1.22
MaritalStatus*	16	1470	2.10	0.73	2.0	2.12	1.48	1	3	2	-0.15	-1.12
MonthlyIncome	17	1470	6502.93	4707.96	4919.0	5667.24	3260.24	1009	19999	18990	1.37	0.99
MonthlyRate	18	1470	14313.10	7117.79	14235.5	14286.48	9201.76	2094	26999	24905	0.02	-1.22
NumCompaniesWorked	19	1470	2.69	2.50	2.0	2.36	1.48	0	9	9	1.02	0.00
OverTime*	20	1470	1.28	0.45	1.0	1.23	0.00	1	2	1	0.96	-1.07
PercentSalaryHike	21	1470	15.21	3.66	14.0	14.80	2.97	11	25	14	0.82	-0.31
PerformanceRating	22	1470	3.15	0.36	3.0	3.07	0.00	3	4	1	1.92	1.68
RelationshipSatisfaction	23	1470	2.71	1.08	3.0	2.77	1.48	1	4	3	-0.30	-1.19
StockOptionLevel	24	1470	0.79	0.85	1.0	0.67	1.48	0	3	3	0.97	0.35
TotalWorkingYears	25	1470	11.28	7.78	10.0	10.37	5.93	0	40	40	1.11	0.91
TrainingTimesLastYear	26	1470	2.80	1.29	3.0	2.72	1.48	0	6	6	0.55	0.48
WorkLifeBalance	27	1470	2.76	0.71	3.0	2.77	0.00	1	4	3	-0.55	0.41
YearsAtCompany	28	1470	7.01	6.13	5.0	5.99	4.45	0	40	40	1.76	3.91
YearsInCurrentRole	29	1470	4.23	3.62	3.0	3.85	4.45	0	18	18	0.92	0.47
YearsSinceLastPromotion	30	1470	2.19	3.22	1.0	1.48	1.48	0	15	15	1.98	3.59
YearsWithCurrManager	31	1470	4.12	3.57	3.0	3.77	4.45	0	17	17	0.83	0.16

--

	se
Age	0.24
Attrition*	0.01
BusinessTravel*	0.02
DailyRate	10.52
Department*	0.01
DistanceFromHome	0.21
Education	0.03
EducationField*	0.03
EnvironmentSatisfaction	0.03
Gender*	0.01
HourlyRate	0.53
JobInvolvement	0.02
JobLevel	0.03
JobRole*	0.06
JobSatisfaction	0.03
MaritalStatus*	0.02
MonthlyIncome	122.79
MonthlyRate	185.65
NumCompaniesWorked	0.07
OverTime*	0.01
PercentSalaryHike	0.10
PerformanceRating	0.01
RelationshipSatisfaction	0.03
StockOptionLevel	0.02
TotalWorkingYears	0.20
TrainingTimesLastYear	0.03
WorkLifeBalance	0.02
YearsAtCompany	0.16
YearsInCurrentRole	0.09
YearsSinceLastPromotion	0.08
YearsWithCurrManager	0.09



Conversion into categorical variables

```
## Convert to categorical variables

Education.cat<-as.factor(Education)
EnvironmentSatisfaction.cat<-factor(EnvironmentSatisfaction)
JobInvolvement.cat<-factor(JobInvolvement)
JobLevel.cat<-factor(JobLevel)
JobSatisfaction.cat<-factor(JobSatisfaction)
PerformanceRating.cat<-factor(PerformanceRating)
RelationshipSatisfaction.cat<-factor(RelationshipSatisfaction)
WorkLifeBalance.cat<-factor(WorkLifeBalance)
|
```

(Dispersion parameter for binomial family taken to be 1)

```
Null deviance: 1298.58  on 1469  degrees of freedom
Residual deviance:  780.96  on 1409  degrees of freedom
AIC: 902.96
```

```
Number of Fisher Scoring iterations: 15
```