Unpacking IBM Human Resource Analytics, Employee Attrition & Performance.

Arthur Richardson

30 October, 2022

Contents

1. Dedication:

This project is dedicated to my two daughters.

Do not become blinded by distractions and negativity. Let your goals and character push you towards positivity.

Love, Dad.

2. Introduction:

In this data set, we will explore, define and solve a series of factors that may or may not be attributed to employee attrition at IBM.

2.1 Objective:

IBM Data Scientists created a fictional data set exploring employee attrition. We will explore each category within the data set. We will also compare and contrast the former workers with the other employees. Eventually, we will categorize the former and current employees into three groups: Category Green: Retained, Category Yellow: At-Risk, and Category Red: Attrited. We will also create a trial data set to determine if we can retain "at-risk" employees based on the data provided.

2.2 Data Installation:

Upload following packages and libraries for data exploration.

```
library(tidyverse)
library(caret)
library(data.table)
library(RColorBrewer)
library(rmarkdown)
library(dslabs)
library(gtable)
library(hexbin)
library(gt)
library(dplyr)
library(ggpmisc)
library(gridExtra)
library(janitor)
library(lubridate)
library(highcharter)
library(viridisLite)
library(broom)
library(scales)
library(xfun)
library(htmltools)
library(mime)
library(ggfortify)
library(gtsummary)
library(tinytex)
library(vroom)
library(curl)
library(gtools)
library(hrbrthemes)
library(viridis)
library(latexpdf)
library(kableExtra)
library(knitr)
library(remotes)
library(extrafont)
library(plotrix)
library(readr)
library(ggforce)
```

2.3 Data Analysis:

Upload the data set. The file can be downloaded from https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset/

The data set has 1470 observations for 35 variables. 1233 employees are current employees and 237 are former workers.

3. Exploring IBM Employee Structure:

IBM data set can be divided into current and former employees by attrition. Attrition, in this case, is terminated employees. Listed below are the former and current workers' job roles. For large corporations, the industry attrition average is 9.9%, and IBM's average is 21%. That means an IBM employee is 2.3x as likely to resign than the average large corporation employee! Let us explore the data.

Note: The definitions and purpose of each category is located in Section 8.

Current Workers vs. Attrited Workers

Unpacking IBM HR Analytics Employee Attrition & Performance.

Current Workers	Current Total	Former Workers	Former Total	Attrition %
Healthcare Representative	131	Healthcare Representative	9	7%
Human Resources	52	Human Resources	12	23%
Laboratory Technician	259	Laboratory Technician	62	24%
Manager	102	Manager	5	5%
Manufacturing Director	145	Manufacturing Director	10	7%
Research Director	80	Research Director	2	3%
Research Scientist	292	Research Scientist	47	16%
Sales Executive	326	Sales Executive	57	17%
Sales Representative	83	Sales Representative	33	40%

Portions of this data is from the Reference Section.

Based on the table above, we observed that the Sales Representatives have the highest attrition rate, followed by Laboratory Technicians and Human Resources. Every position at this company is essential. IBM, as an enterprise, must retain a significant portion of its employees to achieve maximum performance. Unfortunately, we cannot keep the workers who resigned, but we can dive deeper into the data to prevent future attrition. Next, we will look at each column in the data set, create a series of visualizations to document the findings, and explore the differences between current and former workers.

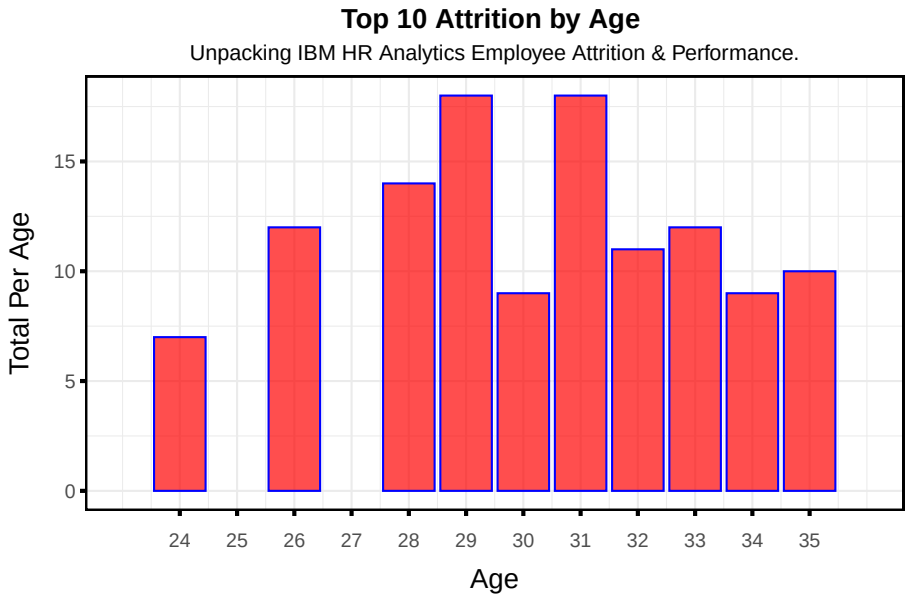
4. Exploration and Analysis: Comparing Current and Former Workers:

Now that we have all of the baseline data let us dive deeper into each appropriate category to grasp what may be the drivers for such a high attrition rate at IBM.

4.1 Age

	Attrited	Employed
Age Min	18	18
Age Mean	34	37
Age Max	58	60
Age Percentage <34	59%	39%
Age Percentage <37	70%	54%

Portions of this data is from the Reference Section.



Portions of this data is from the Reference Section. Ages 25 & 27 are not ranked in the top 10

As we can see, 59% of all former workers are below the age of 34, but 70% of all former workers are below the company's average age of 37. Of our current employees, 54% of the company is below the average age of 37, and 39% are below the age of 34.

4.2 Business Travel

	Attrited	Employed
Non-Travel	5%	10.2%
Travel Frequently	29%	18.8%
Travel Rarely	66%	71.0%

Portions of this data is from the Reference Section.

As seen above, 66% of all former workers Travel Rarely while 71% of current employees Travel Rarely.

4.3 Daily Rate

	Attrited	Employed
Daily Rate Min	$103	$102
Daily Rate Mean	$751.81	$802.49
Daily Rate Max	$1,496	$1,499
Daily Rate Percentage <$802.48	56%	49%

Portions of this data is from the Reference Section.

Listed above, 56% of all former workers make less than the company's average daily rate. 49% of all current employees make less than the company's average daily rate.

4.4 Department

	Attrited	Employed
Human Resources	5%	4%
Research & Development	56%	65%
Sales	39%	30%

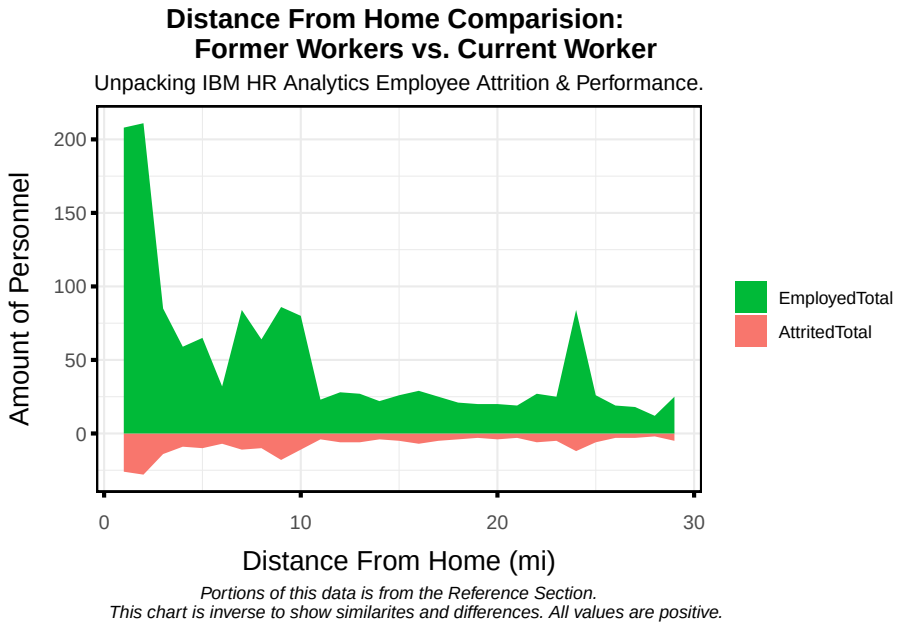
Portions of this data is from the Reference Section.

IBM's three major departments are Human Resources, Sales, Research and Development. As we can see, the Research & Development department has the most employees. 56% of all former employees worked in the Research & Development department. 65% of current employees work in the Research & Development department.

4.5 Distance From Home

	Attrited	Employed
Living <= 10miles of Company	61%	70%

Portions of this data is from the Reference Section.



61% of former workers lived within 10 miles of work while 70% of all current employees lives within 10 miles of work.

4.6 Education

	Attrited	Employed
Below College	13.1%	11.6%
College	18.6%	19.2%
Bachelor	41.8%	38.9%
Master	24.5%	27.1%
Doctor	2.1%	3.3%

Portions of this data is from the Reference Section.

89.6% of all former workers have some level of college education/degree and 88.5% of all current employees have some level of college education/degree.

4.7 Education Field

	Attrited	Employed
Human Resources	3.0%	1.8%
Life Sciences	37.6%	41.2%
Marketing	14.8%	10.8%
Medical	26.6%	31.6%
Other	4.6%	5.6%
Technical Degree	13.5%	9.0%

Portions of this data is from the Reference Section.

37% of former workers had their degree in Life Sciences while 41.2% of all current employees has their degree in Life Sciences.

4.8 Environment Satisfaction

Environment satisfaction uses a numerical value associated with corresponding levels. The groupings are as follows: 1. Low, 2. Medium, 3. High and 4. Very High. The attrition and company average for environment satisfaction was 2.5, which we will round to 3 (corresponding to High).

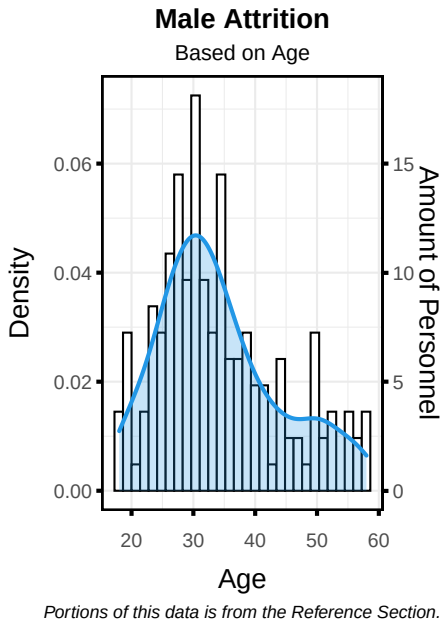
	Attrited	Employed
Low	30.38%	19.32%
Medium	18.14%	19.52%
High	26.16%	30.82%
Very High	25.32%	30.34%

Portions of this data is from the Reference Section.

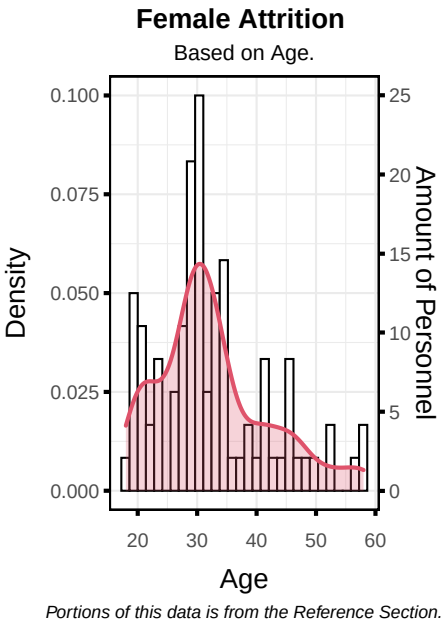
30.38% of former workers Environment Satisfaction was low but 51.48% had high or very high. 19.32% of current workers Environment Satisfaction was low but 61.16% had high or very high.

4.9 Gender

Let us explore the difference in the attrition values based on gender



Portions of this data is from the Reference Section.



Portions of this data is from the Reference Section.

Attrition Comparision: Male vs. Female Worker

Unpacking IBM HR Analytics Employee Attrition & Performance.

Female Age	Total Female Workers	Male Age	Total Male Workers
29	10	28	11
31	7	31	11
33	6	26	9
21	5	29	8
30	5	32	8

Portions of this data is from the Reference Section.

The average attrition age for both genders is 34. Attrited male average age is 37 while the female average age is 36. The former workers' breakdown for gender is 150 males to 87 females. The male to female ratio for attrited workers is almost 2:1 in favor of the males! The company current employees are split: 40% Female and 60% Male.

4.10 Hourly Rate

	Attrited	Employed
Hourly Rate Min	$31	$30
Hourly Rate Mean	$66	$66
Hourly Rate Max	$100	$100
Hourly Rate Percentage <$66	47%	37%

Portions of this data is from the Reference Section.

47% of former workers make less than the hourly rate but 37% of all current workers make less than the hourly rate.

4.11 Job Involvement

	Attrited	Employed
Low	11.8%	5.6%
Medium	30.0%	25.5%
High	52.7%	59.0%
Very High	5.5%	9.8%

Portions of this data is from the Reference Section.

52.7 % of former workers had a high job level involvement while 59% of all current workers have a high job level involvement.

4.12 Job Level

	Attrited	Employed
One	60.3%	36.94%
Two	21.9%	36.33%
Three	13.5%	14.83%
Four	2.1%	7.21%
Five	2.1%	4.69%

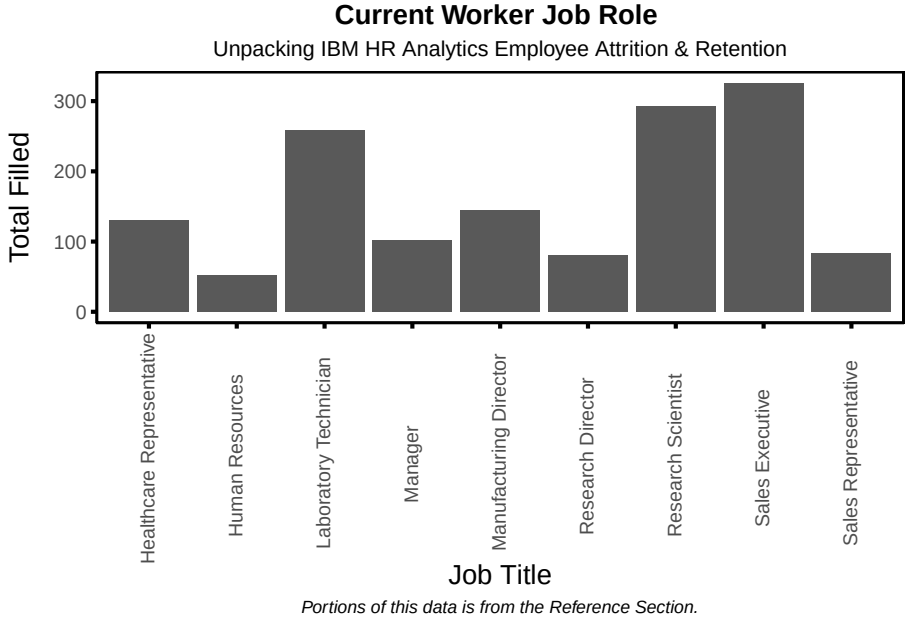
Portions of this data is from the Reference Section.

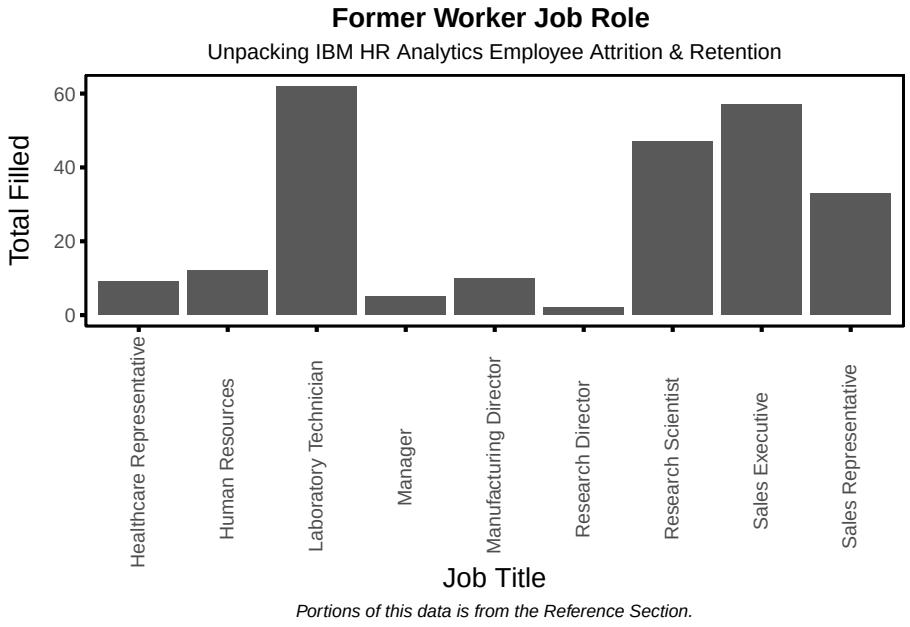
60% of former workers where level one employees while 36% of all current employees are level one employees. This is another glaring statistic that can be addressed by re-evaluating the talent at the company.

4.13 Job Role

Attrition Jobtitle	Attrition %	Jobtitle	Current Workers %
Healthcare Representative	6.870%	Healthcare Representative	8.91%
Human Resources	23.077%	Human Resources	3.54%
Laboratory Technician	23.938%	Laboratory Technician	17.62%
Manager	4.902%	Manager	6.94%
Manufacturing Director	6.897%	Manufacturing Director	9.86%
Research Director	2.500%	Research Director	5.44%
Research Scientist	16.096%	Research Scientist	19.86%
Sales Executive	17.485%	Sales Executive	22.18%
Sales Representative	39.759%	Sales Representative	5.65%

Portions of this data is from the Reference Section.





Job role with the highest attrition is the Sales Representative at 39.759%. Keep in mind that the Sales Representatives only make up 5.65% of all current workers at the company. The job role with the most employees is the Sales Executive, followed by Research Scientist then Laboratory Technician.

4.14 Job Satisfaction

	Attrited	Employed
Low	27.8%	19.66%
Medium	19.4%	19.05%
High	30.8%	30.07%
Very High	21.9%	31.22%

Portions of this data is from the Reference Section.

21.9% of former workers job satisfaction was “Very High”. 31.22% of all current employees job satisfaction is “Very High”.

4.15 Marital Status

	Attrited	Employed
Divorced	14%	22.2%
Married	35%	45.8%
Single	51%	32.0%

Portions of this data is from the Reference Section.

51% of former workers were single. 32% of all current workers are single but 45% of current workers are married.

4.16 Monthly Income

	Attrited	Employed
Monthly Income Min	$1,009	$1,009
Monthly Income Mean	$4,787.09	$6,502.93
Monthly Income Max	$19,859	$19,999
Monthly Income % <$802.48	78%	66%

Portions of this data is from the Reference Section.

78% of former workers make less than the company average monthly income while 66% of all current employees make less than the company average monthly income. This is another eye popping statistic. To retain employees in such a competitive market, IBM must ensure that each employee is compensated appropriately for market conditions.

4.17 Monthly Rate

	Attrited	Employed
Monthly Rate Min	$2,326	$2,094
Monthly Rate Mean	$14,559.31	$14,313.10
Monthly Rate Max	$26,999	$26,999
Monthly Rate % <$802.48	78%	50%

Portions of this data is from the Reference Section.

78% former workers make less than the company average monthly rate while 50% of all current workers make less than the company average monthly rate

4.18 Number of Companies Worked

	Attrited	Employed
Zero	9.70%	13.40%
One	41.35%	35.44%
Two	6.75%	9.93%
Three	6.75%	10.82%
Four	7.17%	9.46%
Five	6.75%	4.29%
Six	6.75%	4.76%
Seven	7.17%	5.03%
Eight	2.53%	3.33%
Nine	5.06%	3.54%

Portions of this data is from the Reference Section.

41% of former employees had worked at only one company before quitting/retiring, while 35% of current employees had worked at only one company before being employed at IBM.

4.19 Workers Under the Age of 18

```
table(IBM_Data$Over18 <18)
```

```
##
## FALSE
##  1470
```

All workers are 18 and older.

4.20 Overtime

	Attrited	Employed
Worked Overtime	54%	28%

Portions of this data is from the Reference Section.

54% of former workers worker over time while 28% of all current employees worked over time.

4.21 Percent Salary Hike

	Attrited	Employed
Eleven %	17.30%	2.789%
Twelve %	13.92%	2.245%
Thirteen %	14.35%	2.313%
Fourteen %	10.13%	1.633%
Fifteen %	7.59%	1.224%
Sixteen %	5.91%	0.952%
Seventeen %	5.91%	0.952%
Eighteen %	5.49%	0.884%
Nineteen %	3.80%	0.612%
Twenty %	2.95%	0.476%
Twenty One %	2.11%	0.340%
Twenty Two %	5.06%	0.816%
Twenty Three %	2.53%	0.408%
Twenty Four %	2.53%	0.408%
Twenty Five %	0.42%	0.068%

Portions of this data is from the Reference Section.

63% of former workers and current workers had a Salary hike between 11-15%

4.22 Performance Rating

	Attrited	Employed
Excellent	84%	85%
Outstanding	16%	15%

Portions of this data is from the Reference Section.

84% of former worker had performance rating of Excellent while 85% of all current employees had a performance rating of Excellent.

4.23 Relationship Satisfaction

	Attrited	Employed
Low	24.1%	18.8%
Medium	19.0%	20.6%
High	30.0%	31.2%
Very High	27.0%	29.4%

Portions of this data is from the Reference Section.

57% of former employees relationship Satisfaction was High or Very High while 60.6% of former employees relationship Satisfaction was High or Very High.

4.24 Standard Hours

	Attrited	Employed
Worked an 80 Hour Week	100%	100%

Portions of this data is from the Reference Section.

100% of all current and former employees worked 80 hr weeks.

4.25 Stock Option Level

	Attrited	Employed
Stock Option Level Min	0	0
Stock Option Level Mean	1	1
Stock Option Level Max	3	3
Stock Option Level <1 %	65%	10%

Portions of this data is from the Reference Section.

65% of the former workers didn't have a stock option above level 1 while 10% of all current employees didn't have a stock option above level 1.

4.26 Total Working Years

Attrited Work Years	Total	%	Employed Work Years	Sum	Employed %
1	40	16.88%	10	202	13.741%
10	25	10.55%	6	125	8.503%
6	22	9.28%	8	103	7.007%
7	18	7.59%	9	96	6.531%
5	16	6.75%	5	88	5.986%
8	16	6.75%	1	81	5.510%
4	12	5.06%	7	81	5.510%
9	10	4.22%	4	63	4.286%
2	9	3.80%	12	48	3.265%
3	9	3.80%	3	42	2.857%

Portions of this data is from the Reference Section.

16% of former workers only had one working year. 13.741% of all current employees had ten working years but only 5.5% had only one working year.

4.27 Training Times Last Year

Attrited Worker

```
## # A tibble: 7 x 3
##   `Attrited Training Hours` Total `Training %`
##                       <dbl> <int> <chr>
## 1                         2    98 41.35%
## 2                         3    69 29.11%
## 3                         4    26 10.97%
## 4                         0    15 6.33%
## 5                         5    14 5.91%
## 6                         1     9 3.80%
## 7                         6     6 2.53%
```

Current Worker

```
## # A tibble: 7 x 3
##   `Current Employee Training Hours` Total `Training %`
##                                 <dbl> <int> <chr>
## 1                                   2   547 37.21%
## 2                                   3   491 33.40%
## 3                                   4   123 8.37%
## 4                                   5   119 8.10%
## 5                                   1    71 4.83%
## 6                                   6    65 4.42%
## 7                                   0    54 3.67%
```

41.35% of former workers trained two hours while 37% of current employees trained two hours. Training your employees creates an increased level of knowledge, can create a healthy learning environment, and drive innovation at your enterprise.

4.28 Work life Balance

	Attrited	Employed
Work life Balance Min	1	1
Work life Balance Mean	3	3
Work life Balance Max	4	4
Work life Balance %	35%	29%

Portions of this data is from the Reference Section.

Only 35% of former workers have a work life balance below the attrition average of "better" but 29% of all current employees have a work life balance below the company average of "better".

4.29 Years At Company

	Attrited	Employed
Years At Company Min	0	0
Years At Company Mean	5	7
Years At Company Max	40	40
Years At Company Percentage <5.14	68%	53%
Years At Company Percentage <7.01	77%	64%

Portions of this data is from the Reference Section.

68% of former workers work less than 5 years and 77% work less than the 7 year company average. ***This is a Red Flag for employees approaching 5 Years at company!*** 67% of current workers work less than 7 years and the company average is 7 years.

4.30 Years in Current Role

	Attrited	Employed
Years in Current Role Min	0	0
Years in Current Role Mean	3	4
Years in Current Role Max	18	40
Years in Current Role Percentage <2.90	64%	62%
Years in Current Role Percentage <4.23	77%	62%

Portions of this data is from the Reference Section.

Only 64% of former workers had at less than 3 Years in current role while 62% of all current workers have less than 3 Years in current role below the company average of 4.23.

4.31 Years With Current Manager

Attrited Worker	Total	Attrited %	Current Worker	Current Total	Employed %
0	85	35.86%	2	344	23.40%
2	50	21.10%	0	263	17.89%
7	31	13.08%	7	216	14.69%
3	19	8.02%	3	142	9.66%
1	11	4.64%	8	107	7.28%
4	11	4.64%	4	98	6.67%
8	10	4.22%	1	76	5.17%
9	6	2.53%	9	64	4.35%
5	4	1.69%	5	31	2.11%
6	4	1.69%	6	29	1.97%
10	3	1.27%	10	27	1.84%
14	2	0.84%	11	22	1.50%
11	1	0.42%	12	18	1.22%

Portions of this data is from the Reference Section.

35.86% of former workers spent 0 years with the current manager, 17.89% of all current workers spent 0 years with the current manager, and 23.50% spent two years with the manager.

4.32 Years Since Last Promotion

	Attrited	Employed
Years Since Last Promotion Min	0	0
Years Since Last Promotion Mean	2	2
Years Since Last Promotion Max	15	15
Years Since Last Promotion Percentage <2	22%	25%

Portions of this data is from the Reference Section.

22% of former workers have not received a promotion in more than 2 years and 25% of all current workers have not received a promotion in more than 2 years.

4.33 Review of IBM Categories

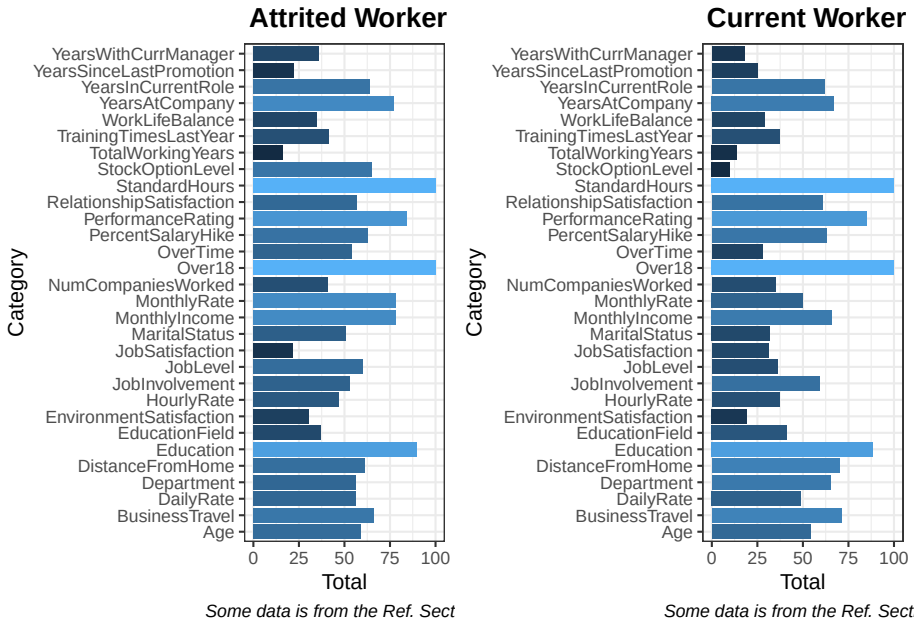
This company is structured well, but some weaknesses must be addressed. To ensure IBM is on par with the industry retention standard, we must attempt to retain all employees to prevent a high employee attrition rate. Below is a table with all the data we explored in section 4. Additionally, look at the Atrrited and Current Worker Category Chart to visualize the data.

Retained vs. Attrited Employees Comparision

Unpacking IBM HR Analytics Employee Attrition & Retention

Attrited	Employed	Attrition_Stats	Employed_Stats
Age	Age	59.00%	54.00%
BusinessTravel	BusinessTravel	66.00%	71.00%
DailyRate	DailyRate	56.00%	49.00%
Department	Department	56.00%	65.00%
DistanceFromHome	DistanceFromHome	61.00%	70.00%
Education	Education	89.60%	88.50%
EducationField	EducationField	37.00%	41.20%
EnvironmentSatisfaction	EnvironmentSatisfaction	30.38%	19.32%
HourlyRate	HourlyRate	47.00%	37.00%
JobInvolvement	JobInvolvement	52.70%	59.00%
JobLevel	JobLevel	60.00%	36.00%
JobSatisfaction	JobSatisfaction	21.90%	31.22%
MaritalStatus	MaritalStatus	51.00%	32.00%
MonthlyIncome	MonthlyIncome	78.00%	66.00%
MonthlyRate	MonthlyRate	78.00%	50.00%
NumCompaniesWorked	NumCompaniesWorked	41.00%	35.00%
Over18	Over18	100.00%	100.00%
OverTime	OverTime	54.00%	28.00%
PercentSalaryHike	PercentSalaryHike	63.00%	63.00%
PerformanceRating	PerformanceRating	84.00%	85.00%
RelationshipSatisfaction	RelationshipSatisfaction	57.00%	60.60%
StandardHours	StandardHours	100.00%	100.00%
StockOptionLevel	StockOptionLevel	65.00%	10.00%
TotalWorkingYears	TotalWorkingYears	16.00%	13.74%
TrainingTimesLastYear	TrainingTimesLastYear	41.35%	37.00%
WorkLifeBalance	WorkLifeBalance	35.00%	29.00%
YearsAtCompany	YearsAtCompany	77.00%	67.00%
YearsInCurrentRole	YearsInCurrentRole	64.00%	62.00%
YearsSinceLastPromotion	YearsSinceLastPromotion	22.00%	25.00%
YearsWithCurrManager	YearsWithCurrManager	35.86%	17.89%

Portions of this data is from the Reference Section. NAs = No data. Just place holders



5. Tackling Attrition:

After our data exploration and analysis, we found exciting statistics that can assist us in retaining employees. One way we can retain our current employees would be to identify who is "at risk." At this point, we will not be able to recall our former workers but what we can do is utilize the data that may have caused them to depart the company. First, let us identify statistics that were below the company average. Next, we will take the best categories and statistics and create three categories featuring the attrited, at-risk, and retained employees. Lastly, in this section, we will remove categories from the IBM data that may not assist us in defining our new category of workers.

5.1 Defining Each Category of Worker

After reviewing all the data in section 4, we will not utilize specific data columns due to redundancies in data, not essential to retention, or insufficient information to expound upon the data to realize its actual effects. We will remove the following: hourly, monthly, and daily rates.

Now lets define the three categories. Category Green will be comprised of employees that are not at-risk of attrition and currently are safe for retention based on the following factors:

1. Category Green:

- Years With Current Manager >= 0,
- Years In Current Role >=0,
- Years At Company >= 0,
- Stock Option Level >= 0,
- Monthly Income <= 13500,
- Job Level >= 1)

Employees in Category Green will be labeled as “Retained”.

Category Yellow will be comprised of employees who are at-risk of attrition and currently are employed. This category is defined by the following factors:

2. Category Yellow:

- Age <= 37,
- Attrition == “No”,
- Years With Current Manager <= 2,
- Years In Current Role <= 4,
- Years At Company <= 7,
- Stock Option Level <= 1,
- Monthly Income < 6502.931,
- Job Level <= 2,
- Business Travel == “Travel_Rarely”)

Employees in Category Yellow will be labeled as “At-Risk”.

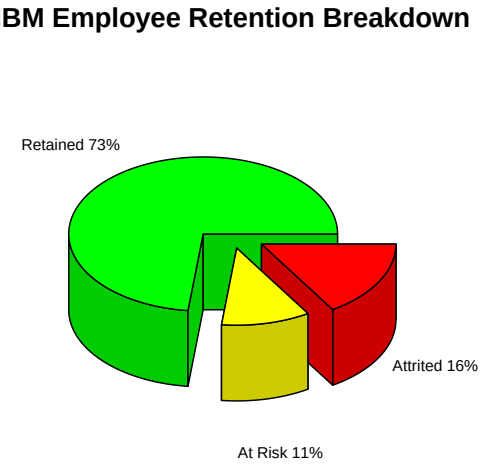
Category Red will be comprised of employees who have resigned or retired. This category is defined by the following factors:

3. Category Red:

- Attrition == “Yes”

Former workers in Category Red will be labeled as “Attrited”.

The following visualizations below will give us a better understanding of the company attrition woes we must correct. One is a pie chart of each category. Another is two tables: one of retained vs. at-risk employees comparison, and the other is the top at-risk jobs.



Retained vs. At-Risk Employees Comparison.

Unpacking IBM HR Analytics Employee Attrition & Retention

Retained Title	Retained	At Risk Title	At Risk
Sales Executive	197	Research Scientist	61
Research Scientist	189	Laboratory Technician	39
Laboratory Technician	144	Sales Representative	22
Manufacturing Director	85	Sales Executive	14
Healthcare Representative	75	Manufacturing Director	7
Sales Executive	62	Human Resources	4
Research Scientist	55	Healthcare Representative	4
Laboratory Technician	51	Laboratory Technician	3
Sales Representative	44	Research Scientist	2
Manufacturing Director	40	Human Resources	1
Healthcare Representative	39	NA	NA
Human Resources	23	NA	NA
Research Director	21	NA	NA
Human Resources	13	NA	NA
Manager	9	NA	NA
Sales Executive	7	NA	NA
Sales Representative	6	NA	NA
Manufacturing Director	6	NA	NA
Human Resources	4	NA	NA
Healthcare Representative	3	NA	NA
Laboratory Technician	2	NA	NA
Research Scientist	1	NA	NA

Portions of this data is from the Reference Section.

Top 3 At-Risk Jobs by %
Unpacking IBM HR Analytics Employee Attrition & Retention

Research Scientist Level 1	Lab. Tech. Level 1	Sales Rep. Level 1
32%	27%	50%

Portions of this data is from the Reference Section. Manu. Dir. is a Manufacturing Director

5.2 Testing a Model to Retain At-Risk Employees

Now that we have an understanding of the data, let us create a model employee structure to drive down the attrition rate at IBM.

5.2.1 Adjust Stock Option for At-Risk Employees

In section 4, we noticed the disparity with the attrited worker stock option level. Let us take a look at the at-risk employees stock option level.

```
table(Cat_Yellow$StockOptionLevel)
```

```
##
##  0  1
## 85 72
```

Let us increase every at-risk employee stock option level. If the stock option level is zero we will upgrade it to one. If the stock option level is one we will upgrade it to level two.

```
RetainatRisk <- Cat_Yellow
RetainatRisk$StockOptionLevel <-replace(RetainatRisk$StockOptionLevel,
                                        RetainatRisk$StockOptionLevel == 1, 2)
RetainatRisk$StockOptionLevel <-replace(RetainatRisk$StockOptionLevel,
                                        RetainatRisk$StockOptionLevel == 0, 1)
table(RetainatRisk$StockOptionLevel)
```

```
##
##  1  2
## 85 72
```

5.2.2 Adjusting Income At-Risk Employees

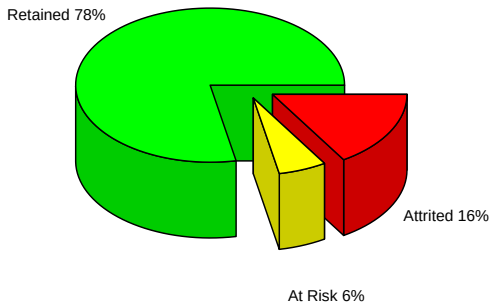
Another issue we highlighted in section 4 is the monthly income pay gap. Since Sales Reps had one of the highest attrition levels, we will give them a pay raise of 30% and a one-time monthly bonus equal to 2.04 times the difference between the company average and the attrition average.

```
RetainatRisk$MonthlyIncome <- ifelse(RetainatRisk$JobRole ==
                                       "Sales Representative",
                                     RetainatRisk$MonthlyIncome*.30+510,
                                     RetainatRisk$MonthlyIncome)

a <- RetainatRisk %>% filter(Age <= 37,
                             Attrition == "No",
                             YearsWithCurrManager <= 2,
                             YearsInCurrentRole <= 4,
                             YearsAtCompany <= 7,
                             StockOptionLevel <= 1,
                             MonthlyIncome < 6502.931,
                             JobLevel <= 2,
                             BusinessTravel == "Travel_Rarely")
nrow(a)
```

```
## [1] 85
```

Update IBM Employee Retention Model



We now cut our At-Risk employees by almost half by making the updates! Let us work on our Level One Research Scientists. First, let us increase their pay by 10% of the company average for that particular job and give them a one-time bonus equal to the difference in pay for the company employee and attrited worker's average monthly pay.

```
RetainatRisk$MonthlyIncome <- ifelse(RetainatRisk$JobRole == "Research Scientist",
                                     RetainatRisk$MonthlyIncome*.10+332.8122,
                                     RetainatRisk$MonthlyIncome)
```

```
a <- RetainatRisk %>% filter(Age <= 37,
                             Attrition == "No",
                             YearsWithCurrManager <= 2,
                             YearsInCurrentRole <= 4,
                             YearsAtCompany <= 7,
                             StockOptionLevel <= 1,
                             MonthlyIncome < 6502.931,
                             JobLevel <= 2,
                             BusinessTravel == "Travel_Rarely")

nrow(a)
```

```
## [1] 85
```

5.2.3 Adjusting Stock Option Level for Research Scientist

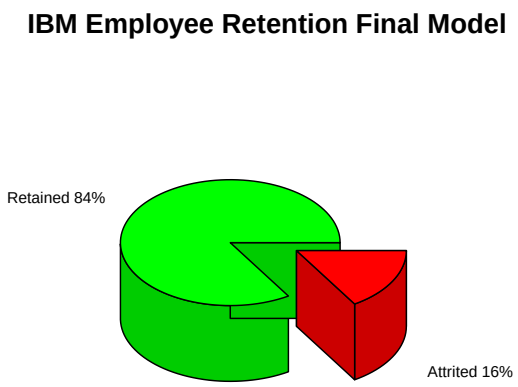
Our Research Scientist's stock options level is one after we made our adjustments, so let us upgrade them to stock options level two to incentivize our at-risk Research Scientist to stay at IBM. Then we will see if any more at-risk employees remain.

```
RetainatRisk$StockOptionLevel <- ifelse(RetainatRisk$JobRole ==
                                          "Research Scientist",
                                        RetainatRisk$StockOptionLevel <- 2,
                                        RetainatRisk$StockOptionLevel)

a <- RetainatRisk %>% filter(Age <= 37,
                             Attrition == "No",
                             YearsWithCurrManager <= 2,
                             YearsInCurrentRole <= 4,
                             YearsAtCompany <= 7,
                             StockOptionLevel <= 1,
                             MonthlyIncome < 6502.931,
                             JobLevel <= 2,
                             BusinessTravel == "Travel_Rarely")
nrow(a)
```

```
## [1] 0
```

6. Results:



Amazing! We retained all at-risk workers by increasing their monthly salaries and stock options.

7. Conclusion:

This data set provided different categories of information that aided us in making a trial set to categorize IBM's current and former employees. Each category adjustment aided us in our journey to retain as many employees as possible by reducing the risk of attrition based on the former workers' attributes. Even with the great strides we made, there are lingering adjustments IBM can make as a corporation that does not involve financial compensation or incentives. Here is a list of recommendations:

1. Encourage employees to travel more.

- Managers within each department should create a tracker to see if their employees travel more often, especially those who travel rarely.

2. Re-evaluate Overtime per employee.

- All workers are averaging 80 hrs per week. Overtime could place a considerable toll on employees. Request managers and engineers create a company-wide overtime tool to monitor Overtime effectively.

3. Create a training plan for all employees at all levels.

- A portion of the attrited workers did not receive the company average of 2 years of training. Request the company immediately evaluate the company training program and require each employee to have at least 3 hours of training per year

4. Re-evaluate Career Goals and Milestone program.

- 68% of former workers work less than five years, and 77% work less than the 7-year company average. Request Managers create a career ladder for employees detailing milestones centered on years at the company. Request management immediately implements a Year 1, 2, 3, and 4 milestone evaluation to prevent at-risk employees from quitting.

5. Create a Mentorship Program.

- 35.86% of former workers spent 0 years with the current manager. Request Managers create a mentorship program, requiring employees to spend at least two years with the current manager unless an HR issue or promotion opportunity is available.

8. IBM Data Set Glossary and Terminology:

1. age numerical value
2. attrition employee leaving the company (0=no, 1=yes)
3. business travel (1=no travel, 2=travel frequently, 3=tavel rarely)
4. daily rate numerical value - salary level
5. department (1=hr, 2=r&d, 3=sales)
6. distance from home numerical value - the distance from work to home
7. education numerical value
8. education field (1=hr, 2=life sciences, 3=marketing, 4=medical sciences, 5=others, 6= tehcnical)
9. employee count numerical value
10. employee number numerical value - employee id
11. enviroment satisfaction numerical value - satisfaction with the enviroment
12. gender (1=female, 2=male)
13. hourly rate numerical value - hourly salary
14. job involvement numerical value - job involvement
15. job level numerical value - level of job
16. job role (1=hc rep, 2=hr, 3=lab technician, 4=manager, 5= managing director, 6= reasearch director, 7= research scientist, 8=sales executieve, 9= sales representative)
17. job satisfaction numerical value - satisfaction with the job
18. marital status (1=divorced, 2=married, 3=single)
19. monthly income numerical value - monthly salary
20. monthy rate numerical value - monthly rate
21. numcompanies worked numerical value - no. of companies worked at
22. over 18 (1=yes, 2=no)
23. overtime (1=no, 2=yes)
24. percent salary hike numerical value - percentage increase in salary
25. performance rating numerical value - erformance rating
26. relations satisfaction numerical value - relations satisfaction
27. standard hours numerical value - standard hours

28. stock options level numerical value - stock options
29. total working years numerical value - total years worked
30. training times last year numerical value - hours spent training
31. work life balance numerical value - time spent bewtween work and outside
32. years at company numerical value - total number of years at the compnay
33. years in current role numerical value -years in current role
34. years since last promotion numerical value - last promotion 1, years with current manager numerical value - years spent with current manager
35. Education

    a. Below College
    b. College
    c. Bachelor
    d. Master
    e. Doctor

36. Environment Satisfaction

    a. Low
    b. Medium
    c. High
    d. Very High

37. Job Involvement

    a. Low
    b. Medium
    c. High
    d. Very High

38. Job Satisfaction

    a. Low
    b. Medium
    c. High
    d. Very High

39. Performance Rating

    a. Low
    b. Good
    c. Excellent
    d. Outstanding

40. Relationship Satisfaction

    a. Low
    b. Medium
    c. High
    d. Very High

41. Work Life Balance

    a. Bad
    b. Good
    c. Better
    d. Best

9. Reference Section:

1. Irizarry, R. A. (2022, July 7). Introduction to Data Science. HARVARD Data Science. Retrieved August 8, 2022, from Https://rafalab.github.io/dsbook/ This project utilized "Introduction to Data Science Data Analysis and Prediction Algorithms with R" by our course instructor Rafael A. Irizarry published 2022-07-07.

2. Kaggle: Your Home for Data Science. (n.d.). Retrieved October 30, 2022, from https://www.kaggle.com/data+sets/pavansubhasht/ibm-hr-analytics-attrition-data+set

3. Jain, R. A. F.-. R. S. (n.d.). IBM HR Analytics Employee Attrition & Performance. Retrieved October 30, 2022, from https://inseaddataanalytics.github.io/INSEADAnalytics/groupprojects/January2018FBL/IBM_Attrition_VSS.html

4. Industries with the Highest (and Lowest) Turnover Rates. (n.d.). Retrieved October 30, 2022, from https://www.linkedin.com/business/talent/blog/talent-strategy/industries-with-the-highest-turnover-rates