

Improving employee retention

A research proposal
By Cindy Barrientos

The problem

- Employee turnover is expensive for a company
- The cost can vary between 50-200% of an employee salary to replace and train someone new, according to a [Forbes article](#)

Data source

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

- Dataset was generated by IBM data scientists regarding employee attrition rates along with survey response data
- Scales are from 1-4
 - 1 referring to “low” or “bad,” while 4 referring to “very high” or “best”
- Some assumptions
 - Attrition is documented at one year post-survey release
 - Relationship satisfaction refers to amicable experiences with fellow employees

Cleaning the data

- Removed columns (boxed in red below) that returned variables that a company has no control over, such as marital status or distance from home
- Removed variables that yield the same or similar information, such as daily rate and monthly income, since all workers in this dataset are full-time

```
In [4]: df.columns
```

```
Out[4]: Index(['Age', 'Attrition', 'BusinessTravel', 'DailyRate', 'Department',  
              'DistanceFromHome', 'Education', 'EducationField', 'EmployeeCount',  
              'EmployeeNumber', 'EnvironmentSatisfaction', 'Gender', 'HourlyRate',  
              'JobInvolvement', 'JobLevel', 'JobRole', 'JobSatisfaction',  
              'MaritalStatus', 'MonthlyIncome', 'MonthlyRate', 'NumCompaniesWorked',  
              'Over18', 'OverTime', 'PercentSalaryHike', 'PerformanceRating',  
              'RelationshipSatisfaction', 'StandardHours', 'StockOptionLevel',  
              'TotalWorkingYears', 'TrainingTimesLastYear', 'WorkLifeBalance',  
              'YearsAtCompany', 'YearsInCurrentRole', 'YearsSinceLastPromotion',  
              'YearsWithCurrManager'],  
              dtype='object')
```

Cleaning the data

- Removed columns that returned variables that a company has no control over, such as marital status or distance from home
- Removed variables that yield the same or similar information, such as daily rate and monthly income, since all workers in this dataset are full-time

```
In [7]: df = df.loc[:, ['Age', 'Attrition', 'BusinessTravel', 'Department',  
    'EnvironmentSatisfaction', 'Gender', 'HourlyRate',  
    'JobInvolvement', 'JobLevel', 'JobRole', 'JobSatisfaction', 'RelationshipSatisfaction',  
    'OverTime', 'PercentSalaryHike', 'PerformanceRating',  
    'StockOptionLevel',  
    'TotalWorkingYears', 'TrainingTimesLastYear', 'WorkLifeBalance',  
    'YearsAtCompany', 'YearsInCurrentRole', 'YearsSinceLastPromotion',  
    'YearsWithCurrManager']]
```

Exploring attrition rates

- Company-wide, attrition is nearly 20%
- Could this be different by department?

```
In [15]: print("Attrition rates of IBM employees: ", len(attrition)/len(retention))
```

```
Attrition rates of IBM employees: 0.1922141119221411
```

Exploring attrition rates

- Company-wide, attrition is nearly 20%
- Could this be different by department?
 - R&D attrition is as low as 14% while Sales is as high as 21%

```
In [16]: depts = df.Department.unique()
print("Departments included: ", depts)
```

```
Departments included:  ['Sales' 'Research & Development' 'Human Resources']
```

```
In [17]: print("Attrition rates per department: ")
for dept in depts:
    print(dept, ": ", len(attrition[attrition.Department == dept]) / len(df[df.Department == dept]))
```

```
Attrition rates per department:
Sales :    0.2062780269058296
Research & Development :    0.1383975026014568
Human Resources :    0.19047619047619047
```

Exploring attrition rates

- HR department has too low an attrition rate sample size
 - Will not generate reliable information
- Focus on R&D and Sales

```
In [18]: attrition['Department'].value counts()
```

```
Out[18]: Research & Development    133
         Sales                      92
         Human Resources             12
         Name: Department, dtype: int64
```

```
In [19]: retention['Department'].value counts()
```

```
Out[19]: Research & Development    828
         Sales                      354
         Human Resources             51
         Name: Department, dtype: int64
```


Factors affecting attrition rates per department

- Running selected variables across a Mann-Whitney U test to determine differences in attrition within R&D and Sales

```
In [22]: columns = ['Age',  
                  'EnvironmentSatisfaction', 'Gender', 'HourlyRate',  
                  'JobInvolvement', 'JobLevel', 'JobRole', 'JobSatisfaction', 'RelationshipSatisfaction',  
                  'OverTime', 'PercentSalaryHike', 'PerformanceRating',  
                  'StockOptionLevel',  
                  'TotalWorkingYears', 'TrainingTimesLastYear', 'WorkLifeBalance',  
                  'YearsAtCompany', 'YearsInCurrentRole', 'YearsSinceLastPromotion',  
                  'YearsWithCurrManager']  
  
for dept in depts:  
    print(dept, "\n")  
    for column in columns:  
        print(column)  
        print(mannwhitneyu(attrition[attrition.Department == dept][column], retention[retention.Department==dept][column]))  
        print("\n")  
    print("\n")
```

Research & Development

Age

MannwhitneyResult(statistic=39127.0, pvalue=4.021915896592603e-08)

EnvironmentSatisfaction

MannwhitneyResult(statistic=47004.0, pvalue=0.0024195682172469574)

Gender

MannwhitneyResult(statistic=50520.0, pvalue=0.035483500964093063)

HourlyRate

MannwhitneyResult(statistic=53829.0, pvalue=0.3391228846797934)

JobInvolvement

Factors affecting attrition rates per department

- Running selected variables across a Mann-Whitney U test to determine differences in attrition within R&D and Sales
- Key differences between attrition and retention:
- Both R&D and Sales
 - age
 - environment satisfaction
 - job involvement
 - job level
 - job satisfaction
 - overtime
 - stock option level
 - total working years
 - years at company
 - years in current role
 - years with current manager

Factors affecting attrition rates per department

- Running selected variables across a Mann-Whitney U test to determine differences in attrition within R&D and Sales
- Key differences between attrition and retention:
- R&D
 - gender
 - relationship satisfaction
 - training times last year
 - work life balance
 - years since last promotion
- Sales
 - Job role

Factors affecting attrition rates per department

- Though sales has higher attrition rates, I would like to explore R&D further as there seems to be more determinant factors.
- With R&D as a model, key differences between the departments can be studied as to implement changes that seem to have a positive impact.
- The company can later conduct another A/B test here before rolling out those changes company-wide.

Colleague relationship satisfaction relative to attrition

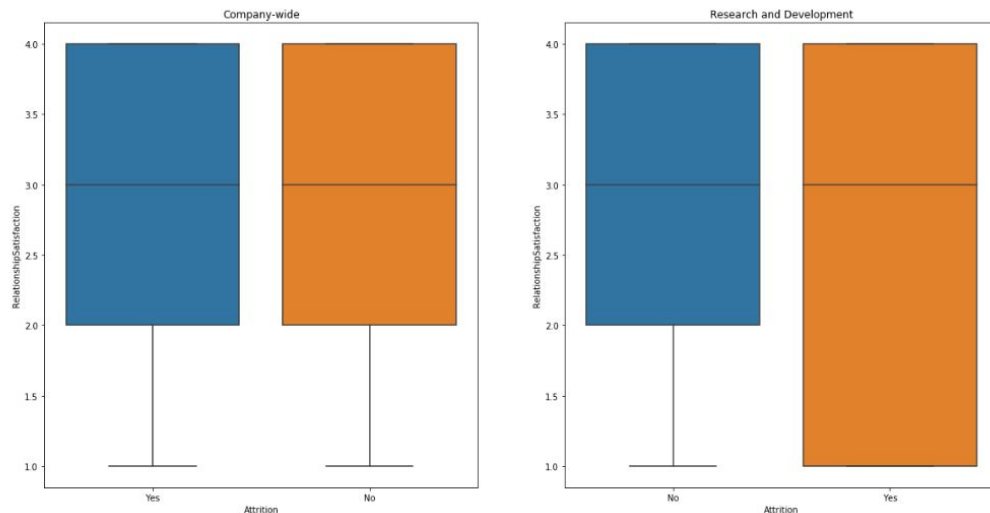
```
In [30]: fig, axs = plt.subplots(ncols=2,rows=1)

fig.suptitle("Colleague relationship satisfaction relative to attrition", fontsize=16)

ax = sns.boxplot(x='Attrition', y='RelationshipSatisfaction', data=df, ax=axs[0])
ax.set_title("Company-wide")

ax = sns.boxplot(x='Attrition', y='RelationshipSatisfaction', data=df_rnd, ax=axs[1])
ax.set_title("Research and Development")
plt.show()
```

Colleague relationship satisfaction relative to attrition



- Relationship satisfaction does not seem to be a problem company-wide (left plot)
- In R&D (right plot), there is a greater amount of attrition in employees who rated relationships as a 2 or lower

Pre-rollout plan

Assuming these surveys were not recently taken, I suggest sending out an additional round of surveys as a pre-experiment baseline across all of R&D. With these new surveys, I would also change the scale in the perception-related questions from its original 1-4 to 1-10. It's much easier for the human brain to quantify their experience when there is a wider range of number to choose from (such as 5 if they feel entirely neutral - which 1-4 does not allow). In addition, it will allow the analysis to take note of more hidden patterns.

Rollout plan

IBM has many offices, select 2 that are most similar (i.e. 2 bustling cities or 2 suburb areas with similar department sizes and similar attrition rates). Select one office at random to be our control group for an A/B test and maintain status quo in this department. With the experimental group, resources will be put in to improve relationships across colleagues. Since the amount of resources utilized will depend on company budget, a few questions needs to be addressed first.

Rollout plan

Budgeting

1. What is the annual cost of training new employees to replace those that have left the company?
 - What is the minimum decrease in costs here that the company would consider this a worthwhile endeavor?
2. Considering the costs of turnover and loss of knowledge and skills that have been invested in each employee, what is a reasonable budget that the company is willing to invest in promoting retention?

Rollout plan

Budgeting

With these questions answered, the budget then needs to be divided into two categories

1. Training for management that will promote teamwork and camaraderie amongst their subordinates
2. Regular team-building events and picnics

Rollout plan

Events schedule

For the purposes of completing this proposal that can then be used as a backbone for any company looking to decrease turnover rates, I will assume a budget that can afford the following schedule:

1. Management training every 6 months that will facilitate improved coworker and manager-subordinate relations

2. Bimonthly team-building event or picnic

- Month 1: Softball tournament - Month 3: Picnic - Month 5: Escape room - Month 7: Picnic -
Month 9: Speed Go-Karts - Month 11: Potluck

Rollout plan

Data acquisition and length of study

Unfortunately, culture that leads to high turnover may take some time to improve. However, I would argue that it's better to take time to measure and experiment in ways that will lead to positive outcomes for the company, since the information generated would be beneficial for the rest of the company's lifetime.

A survey will be released again at the 6 month mark for both control and experimental group, and finally, at one year post-rollout. Employees who quit during this time will again be given the survey as part of their exit interview.

Rollout plan

Hypothesis

The experimental group will experience at least a 10% greater overall retention rate within the year, relative to the control group. A Mann Whitney-U test will determine this is correct with a 95% confidence or greater.

Success metric

Metric for success would be to retain 10% more employees in the experimental group. Once this is established, we can implement this program throughout every office's R&D department, and begin plans to establish such a program in other departments with much higher attrition rates, such as sales.