People Analytics: Why IBM Workers Quit?

Bhawna Gupta, Hua Tu, Mingrui Zhou, Siyang Cheng

Index

- What is People Analytics
- Introduction - Attrition, Dataset
- Exploratory Data Analysis- Gender, Monthly Income, Department, Education Levels, Age, Job Title, Satisfaction
- Predictive Analysis
- Conclusions and Recommendations

What is People Analytics?

A data-driven approach to managing people at work

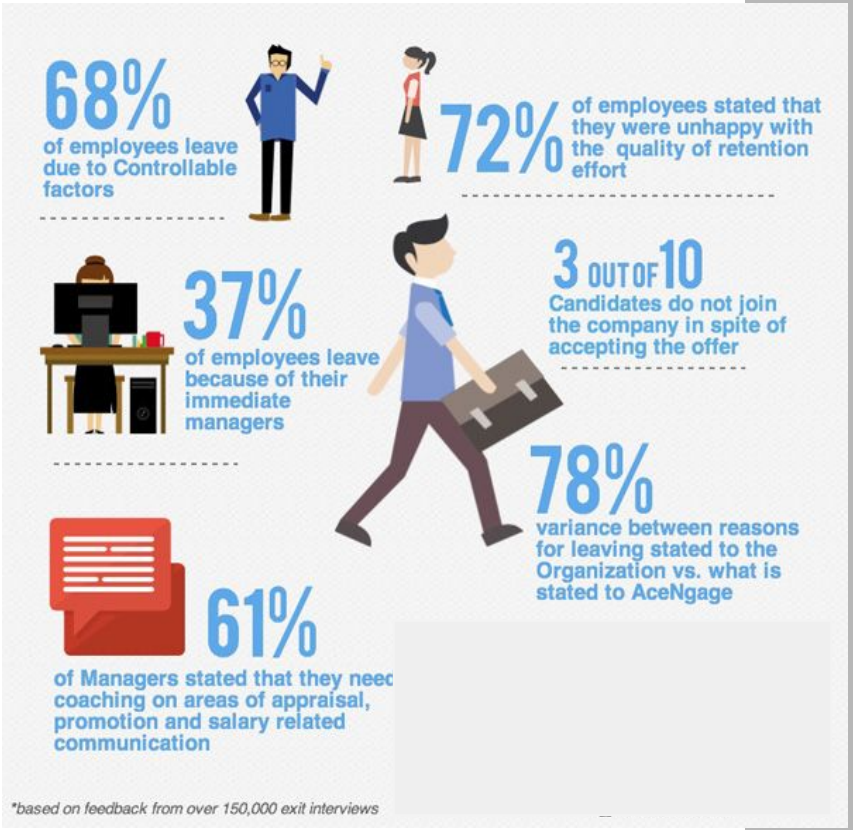
Analytics applied to people issues - hiring, retention, compensation etc.

Firm’s greatest assets are its people

Attrition

- Turnover rate of employees inside an organization
- Some possible Reasons:

1. Employees seeking for better opportunities.
2. A negative working environment.
3. Dissatisfaction with job
4. Bad management
5. Health condition of an employee (or even death)
6. Excessive working hours



About the Dataset

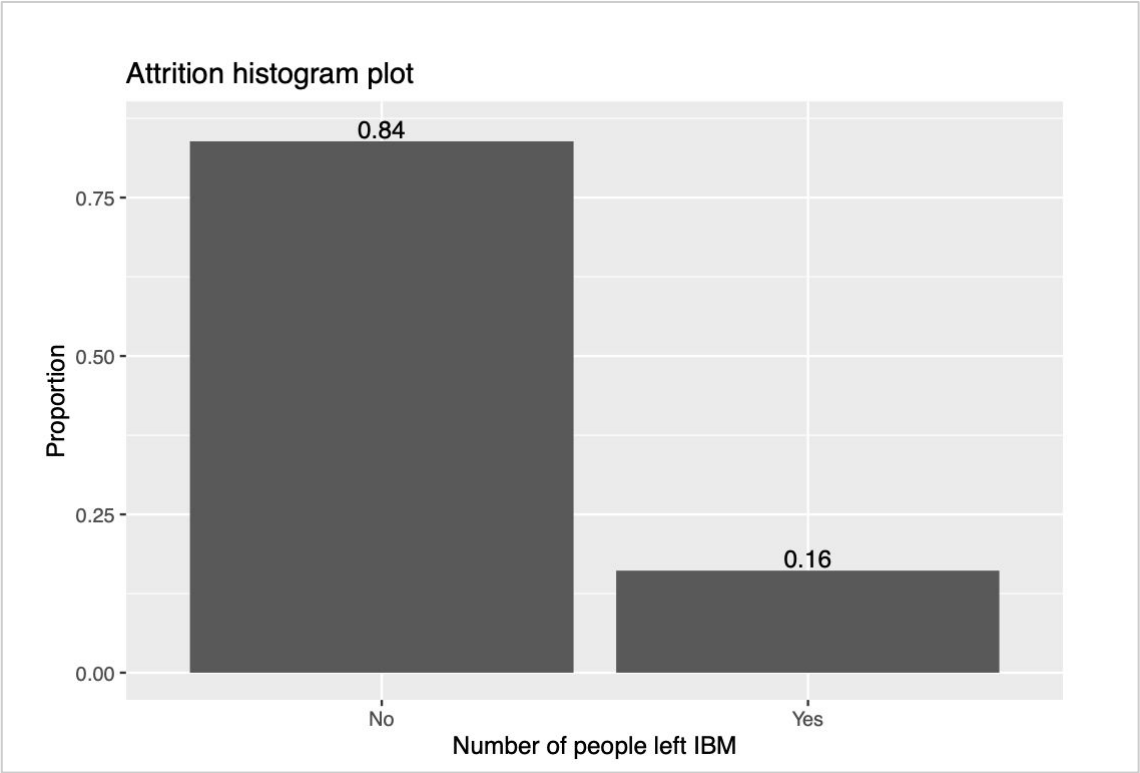
IBM HR Analytics Employee Attrition & Performance

- Fictional data set created by IBM data scientists
- Uncover the factors that lead to employee attrition
- Data structure: 1470 observations, 35 features
- Data Type: factors and integers
- Label: Attrition

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

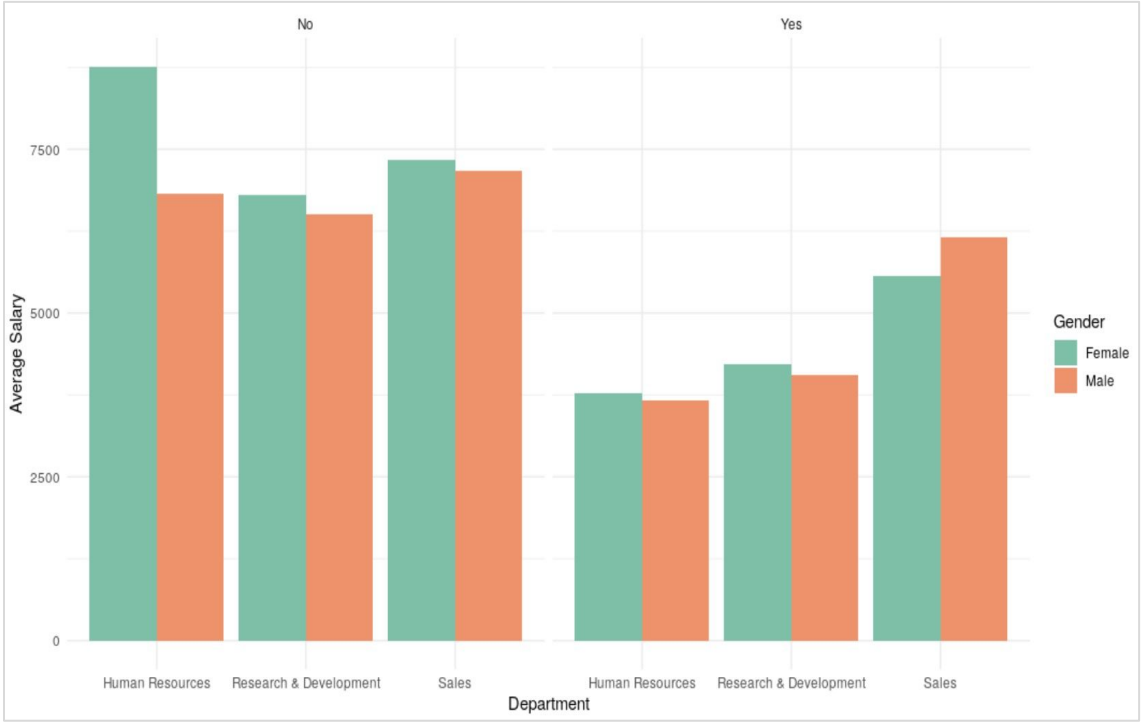
Exploratory Data Analysis on Attrition

Imbalanced Dataset



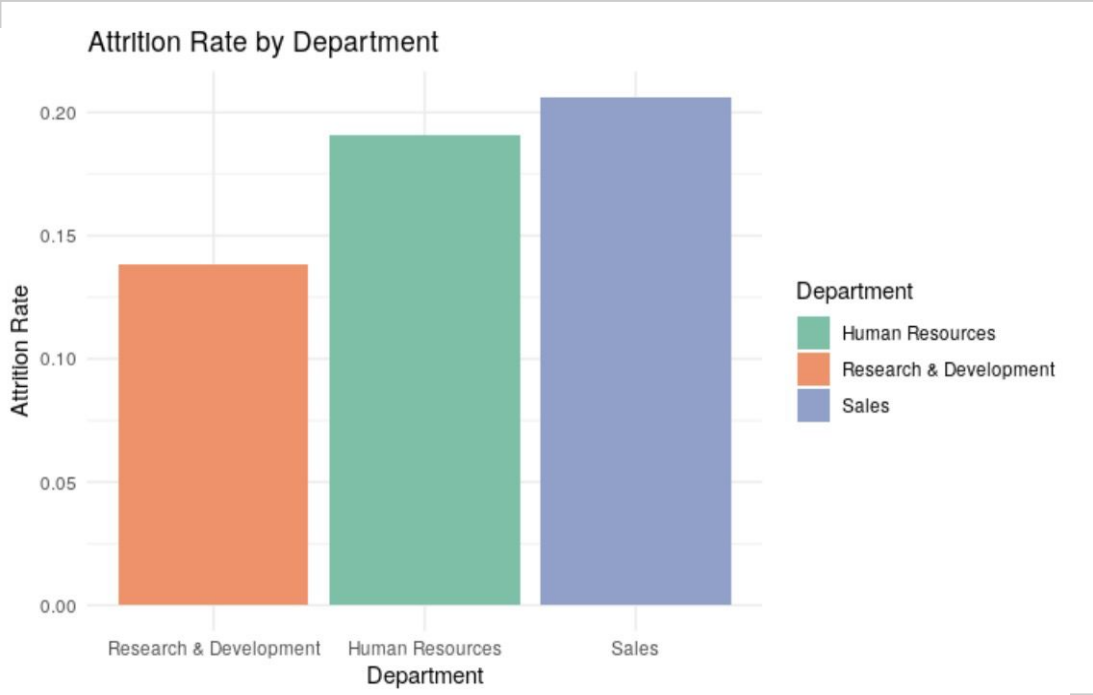
- 16% of the total employees left the organization while 84% of them stayed in the organization
- The number of data points available for different classes varies
- Dataset is **imbalanced**

Average Salary by Department, Gender and Attrition



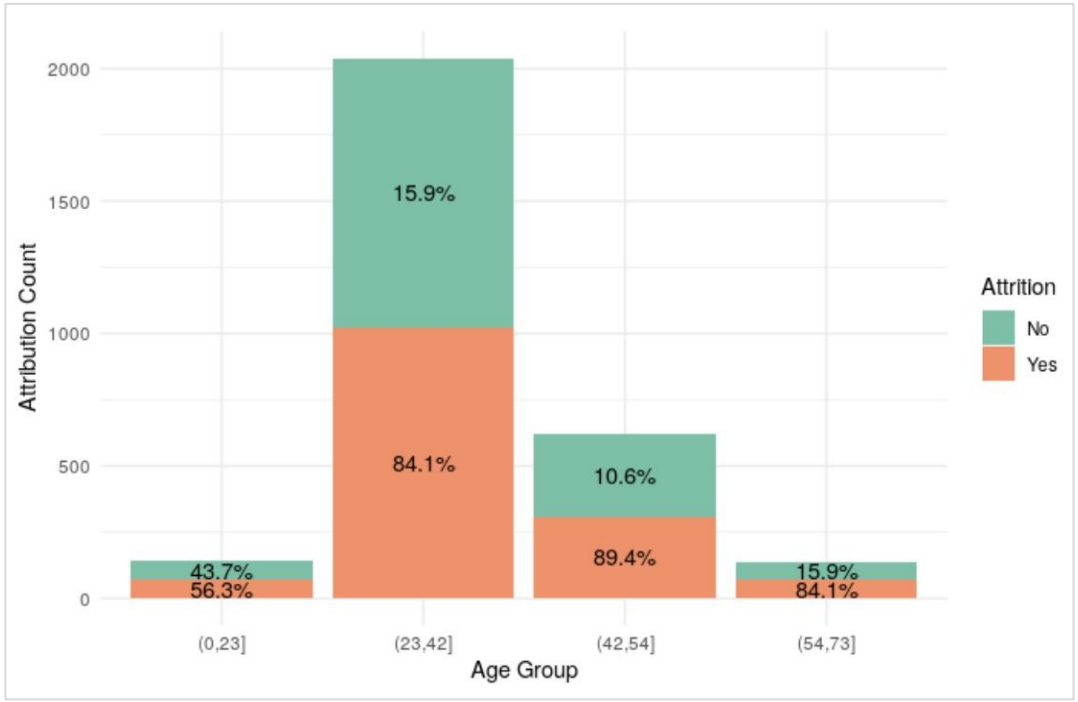
- In general, females earned higher salaries than males
- Employees who left the organization were earning much less than employee who stayed
- Among Attrition = No, average salary for human resources is the highest; When Attrition = Yes, average salary for sales is the highest
- Why did employees from sales leave when their salaries were much higher than other departments?

Attrition Rate by Department



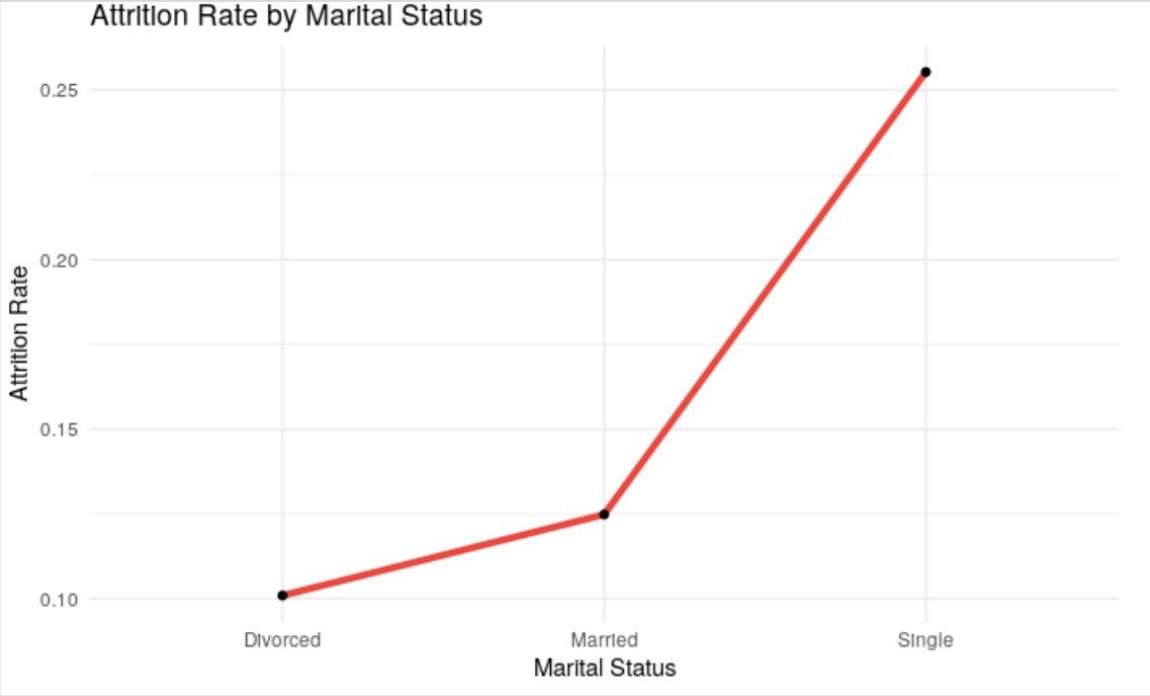
- Employees working at sales department left the organization the most.
- Employees working at Research and Development department are least likely to leave the company.

Attrition by Age Groups



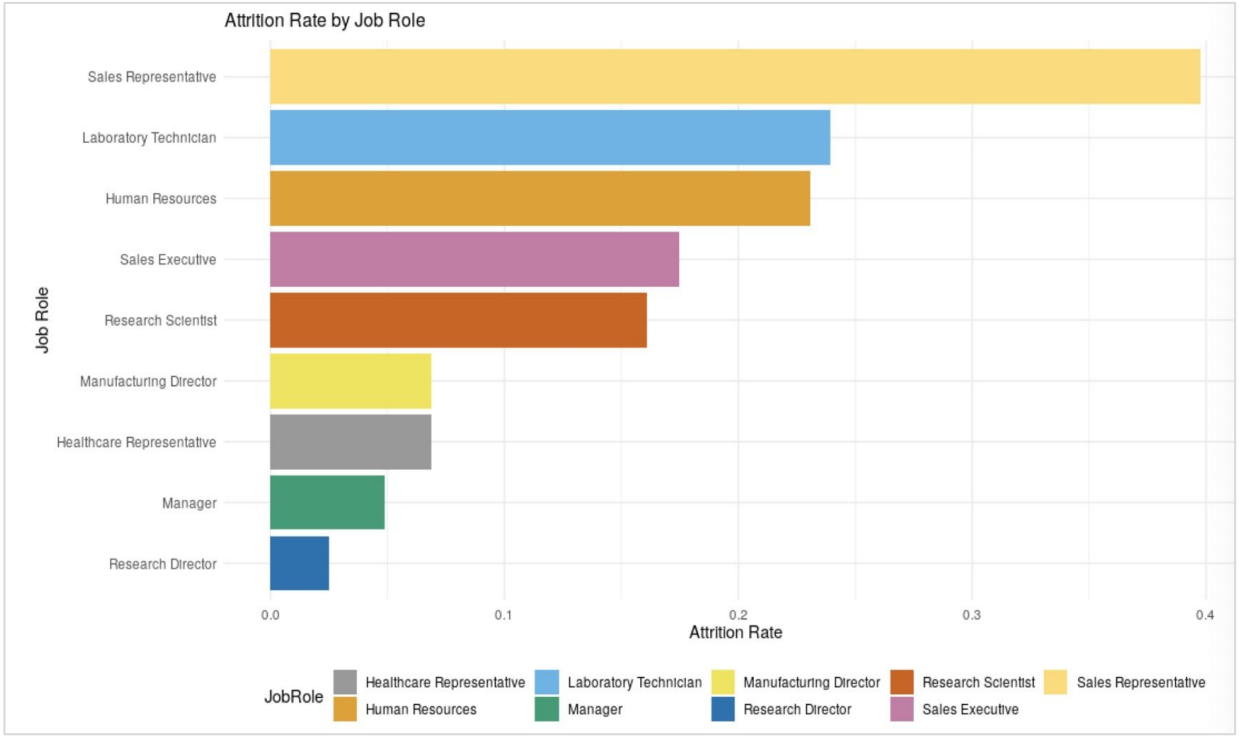
- Age groups :
  - 0-23 Gen Z
  - 24-42 Gen Y
  - 43-54 Gen X
  - 55-73 Baby Boomers
- For employees who are aged under 23, 43.7% of them left their job, which is the highest
- For the age group 42-54, only 10.6% of them left the organization

Attrition Rate by Marital Status



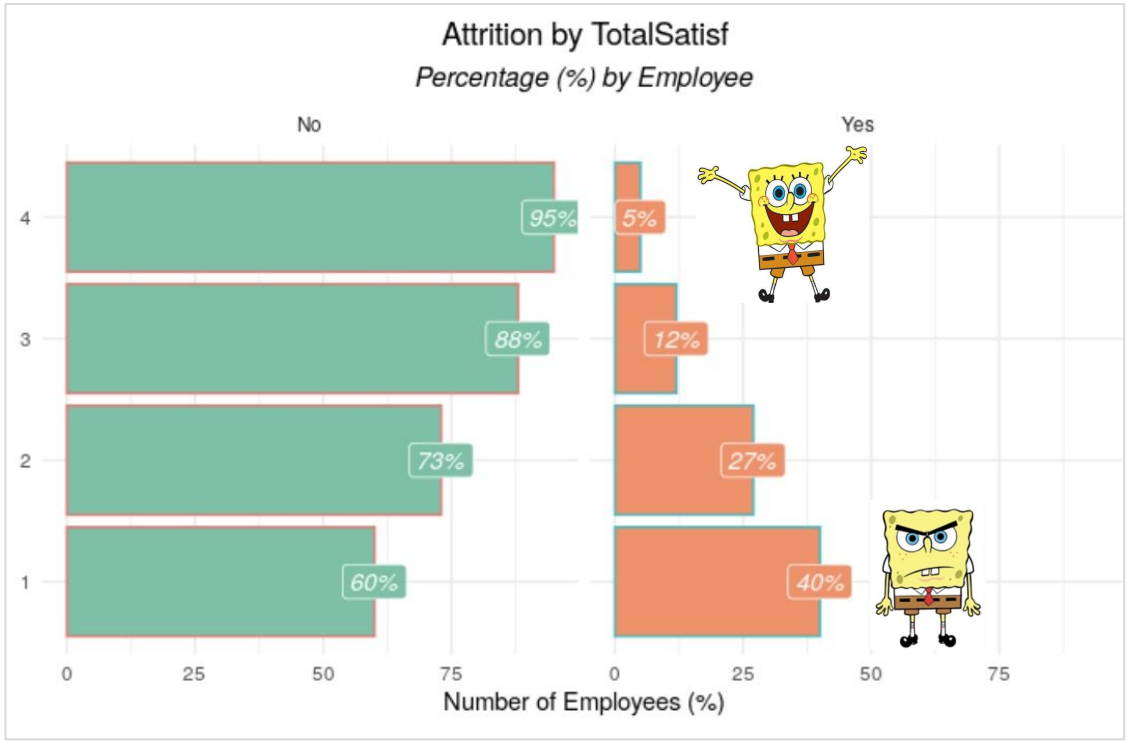
- Only 10% of employees who are divorced left the organization; they are least likely to leave the organization
- Employees who are single are most likely to leave the organization that more than 25% of them left.

Attrition Rate by Job Role



- Sales representatives have strong tendency to leave their company
- The attrition rate for laboratory technician and human resources are similar, with a percentage of around 0.23%
- Directors, representatives, managers tend to have low attrition rate

Attrition by Total Satisfaction

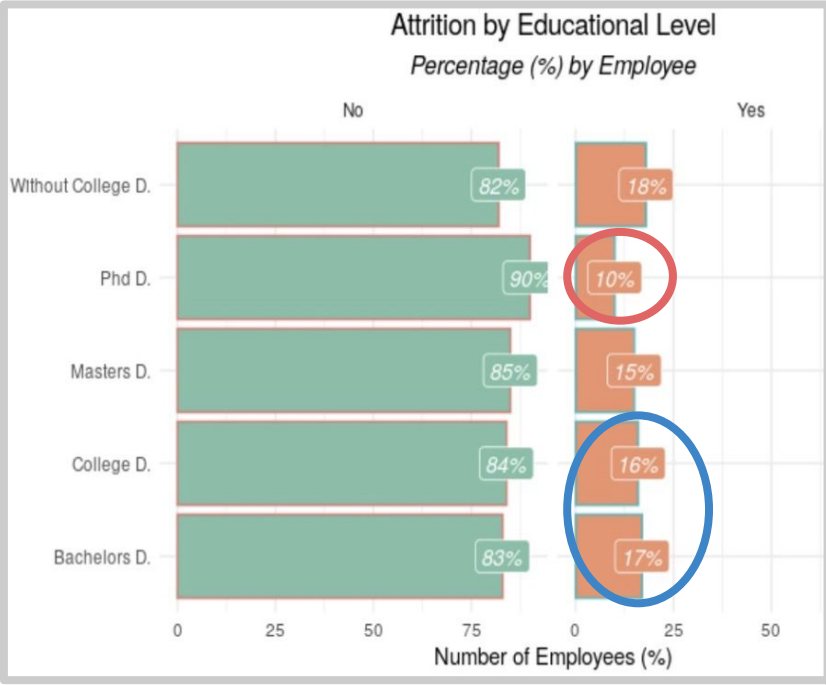


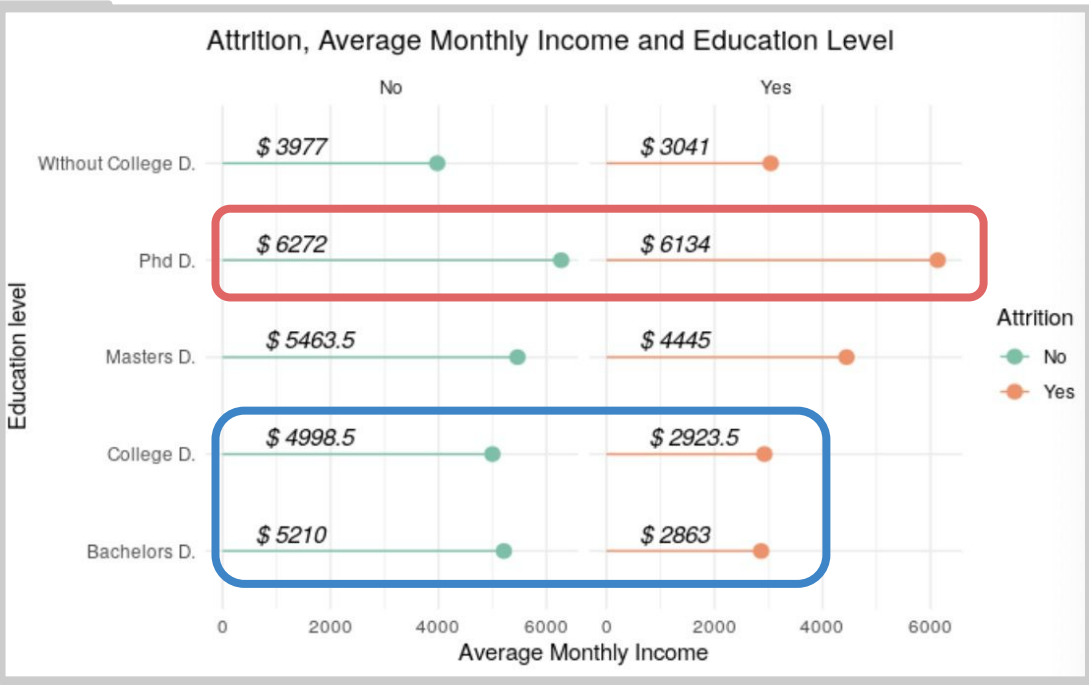
Total Satisfaction ≈

1/5*
(Relationship Satisfaction +
Environment Satisfaction +
Job Satisfaction +
Job Involvement +
Work-Life Balance)

Total Satisfaction Level:
1 'Low' 2 'Medium' 3 'High' 4 'Very High'

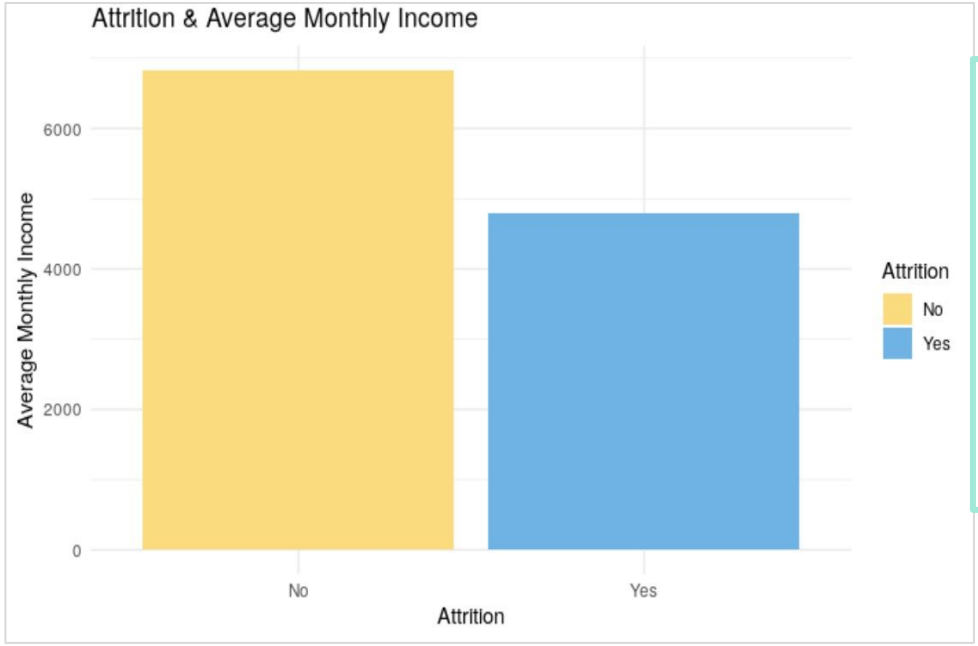
Attrition by Average Monthly Income& Education Level





PhD - Small income gap & low attrition rate (red)
College, Bachelors - Huge income gap & high attrition rate(blue)

Attrition & Average Monthly Income

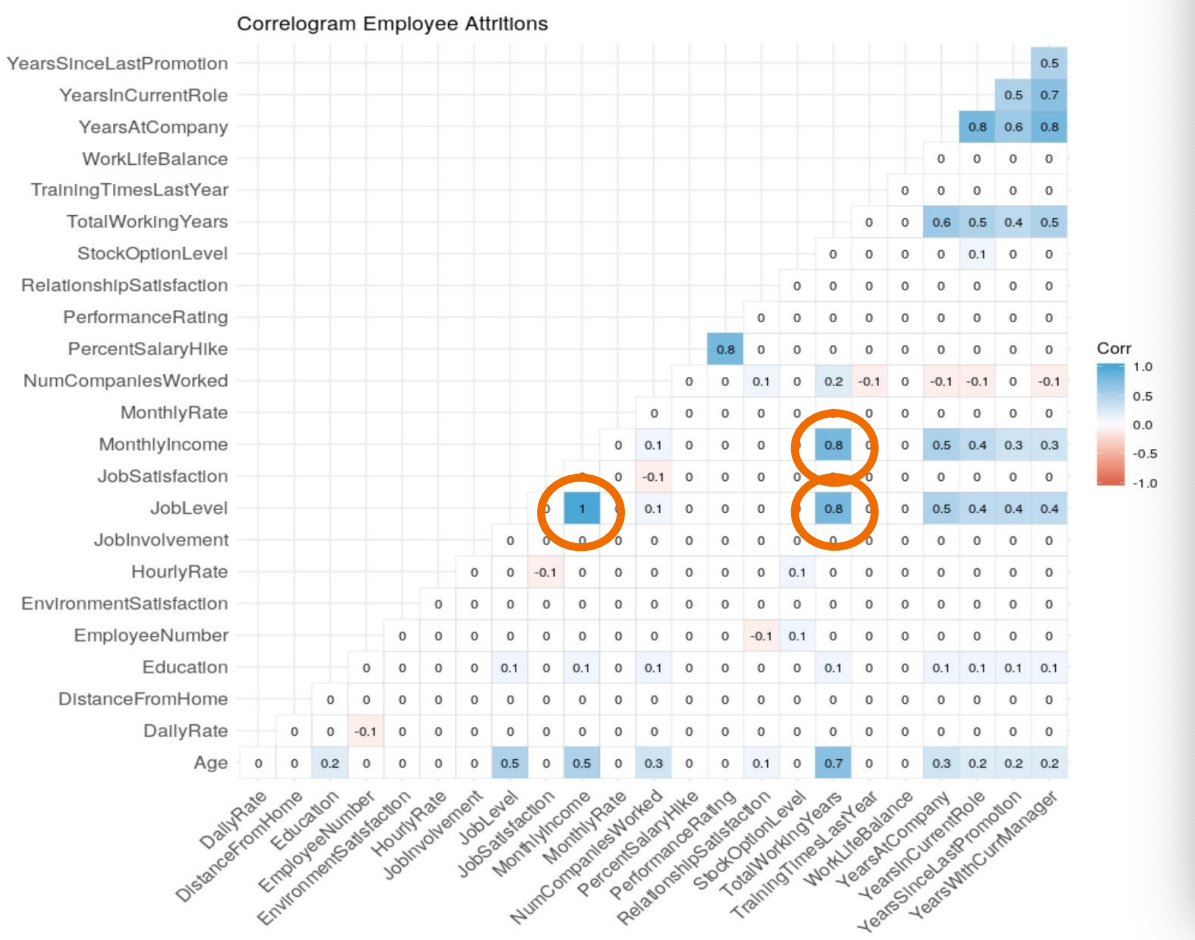



```
        Welch Two Sample t-test

data:  income_attr$MonthlyIncome and income_not_attr$MonthlyIncome
t = -7.4826, df = 412.74, p-value = 4.434e-13
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2583.050 -1508.244
sample estimates:
mean of x mean of y
 4787.093  6832.740
```

Significant difference!

Feature Correlations

- Job level and monthly income are highly correlated
- Job level and total working years are highly correlated
- Monthly income and total working years are highly correlated



Predictive Analysis

Feature Engineering and Selection

- Based on our previous exploratory data analysis and correlation table, we selected features which we think contribute most to the predictive variable: Attrition
- Fidelity is a feature generated by ourselves using *Number of Companies worked / Total working years*
- A total of 22 features: *'Business Travel', 'Daily Rate', 'Department', 'Distance From Home', 'Education', 'Education Field', 'Employee Count', 'EmployeeNumber', 'Gender', 'Hourly Rate', 'Job Role', 'Marital Status', 'Monthly Income', 'Monthly Rate', 'OverTime', 'Percent Salary Hike', 'Performance Rating', 'Stock Option Level', 'Training Times Last Year', 'YearsAtCompany', 'Years In Current Role', 'Years Since Last Promotion', 'Years With Current Manager', 'overall_satisf', 'Age_bin', 'fidelity'*

Model Design

- 4 Models: Logistic Regression, Random Forest, Gradient Boosting Classifier, Ada Boost Classifier
- Split data with 70% of training data and 30% of test data
- Dealing with unbalanced data:

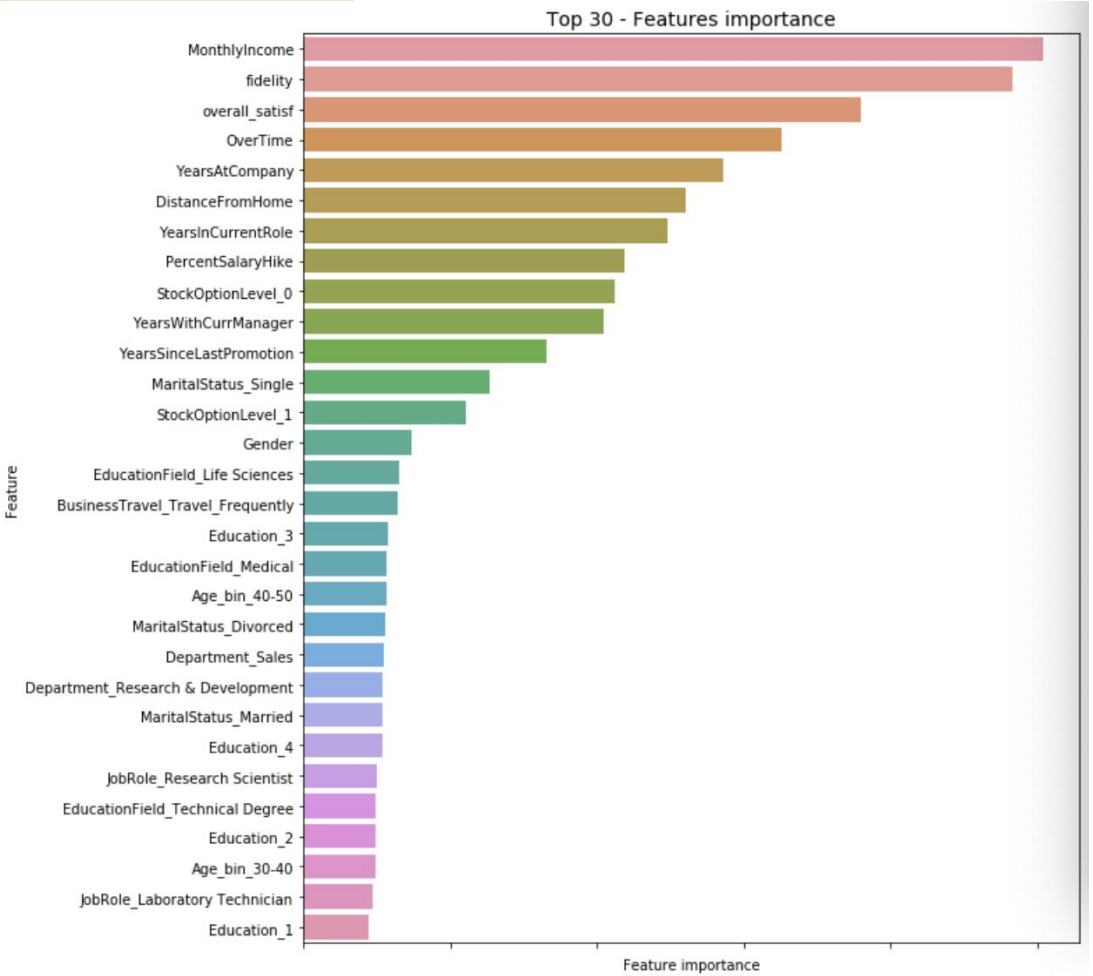
```
sample_weight = compute_sample_weight ('balanced', y_train)

gb_clf = GradientBoostingClassifier(n_estimators=20)

gb_clf.fit(X_train, y_train, sample_weight)
```

- Test the trained model using the test set and cross validation

Comparison between Models

Models	Accuracy rate on test set	Accuracy rate of cross validation
Random Forest	87.30%	87.53%
Gradient Boosting Classifier	86.62%	86.6%
Ada Boost Classifier	86.62%	86.74%
Logistic Regression	87.12%	87.06%

- Random Forest has the highest accuracy rate on both test set and cross validation



Top 30 important features

- **Monthly Income:** People who earned a higher income have low tendency to leave
- **Fidelity:** If people switch jobs less often, it means they are also less likely to quit their jobs
- **Overall Satisfaction:** People enjoy working when they feel satisfied with their jobs, environment and relationships, therefore have low possibility for attrition
- **Over-time:** People don't prefer jobs which require long working hours
- **Years at Company:** If people work for a company for too long, it is easier to get a raise at a new job

Conclusions and Recommendations

- The main general reason for attrition is the **effort-reward imbalance.**
- People who work **overtime** or have **low income** have high tendency of leaving the job
  - ❏ IBM should check with their overtime policy that whether the overtime payoff is effective and limit on overtime hours
- For IBM's current work force, sales representatives have high possiblity for attrition
  - ❏ HR can check with recruiters' fedelity when they hire sales representatives

Conclusions and Recommendations

- With our logistic regression model, IBM can apply it to calculate the possibilities of employees prone to leaving the company
- With known information of an employee, we can calculate his probability of attrition

Gender	Male
MonthlyIncome	8463
OverTime	No
PercentSalaryHike	18
PerformanceRating	3
YearsAtCompany	5
YearsInCurrentRole	4
YearsSinceLastPromotion	1
YearsWithCurrManager	3
overall_satisf	3
BusinessTravel	Rarely Travel
Department	Sales Deparment
Education	Bachelor
Age_bin	Under 30 Years Old
EducationField	Medical
JobRole	Sales Executive
MaritalStatus	Single
StockOptionLevel	Level 0
Prediction of Attrition Probability	**13%**

GROUP 2