

GROUP 14 PRESENTS
EMPLOYEE ATTRITION
WHY DO WORKERS QUIT?

University of California, San Diego
Department of Electrical and Computer Engineering
Team Members : Nikhil Mohan, Xuezhu Hong, Qi Ma, Changhan Ge
03.13.2019

Contents

- Introduction
- Dataset
- Analysis
- Prediction
- Conclusion

Introduction

- Attrition is basically the turnover rate of employees within an organization. Our goal is to analyze the factors which determine attrition and suggest possible remedial measures.
- Variety of factors can determine attrition, and these are, but not limited to
 1. Hostile environment
 2. Long work hours with less pay
 3. Employees looking for better opportunities.
 4. Poor management

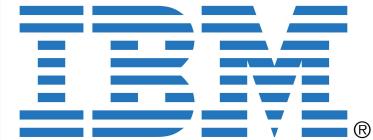


Contents

- Introduction
- Dataset
- Analysis
- Prediction
- Conclusion

Dataset

- **IBM Attrition Dataset**
 - ❖ A US Science and Technology Company
 - ❖ Source: Kaggle
<https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-attrition-dataset>
- **BRC Attrition Dataset**
 - ❖ A Chinese real estate & tourism investment group
 - ❖ Source: Provided by Ms. Jane Liu who served as a HR at BRC
- **Employee Review Dataset**
 - ❖ Company reviews written by employees of several silicon valley companies
 - ❖ Source: Kaggle
<https://www.kaggle.com/petersunga/google-amazon-facebook-employee-reviews>
- **Basics about dataset**



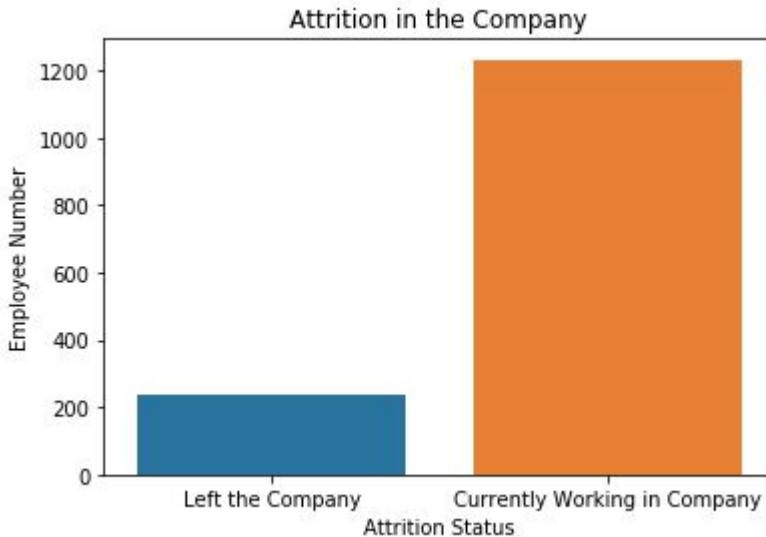
	Total Sample	Left the Company	Still in the Company	Attrition Rate
IBM	1470	237	1233	16%
BRC	838	752	86	90%
Review	67000		N/A	

Contents

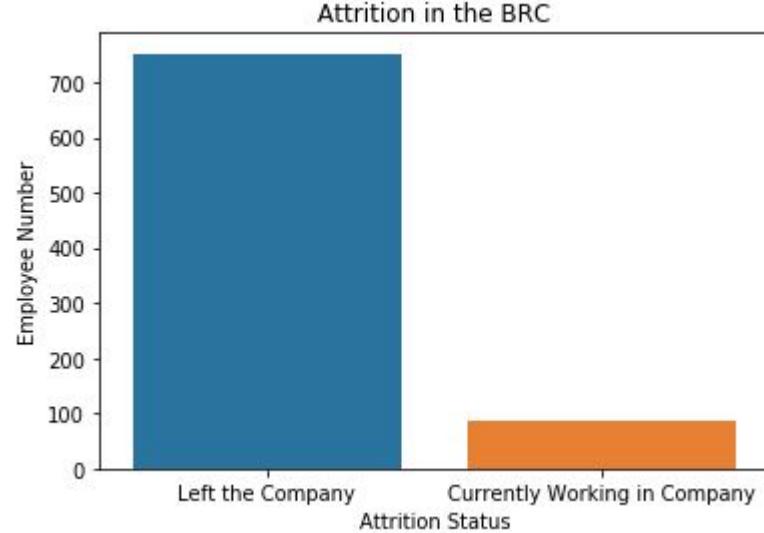
- Introduction
- Dataset
- Analysis
- Prediction
- Conclusion

Distribution of features - Attrition

IBM DATASET

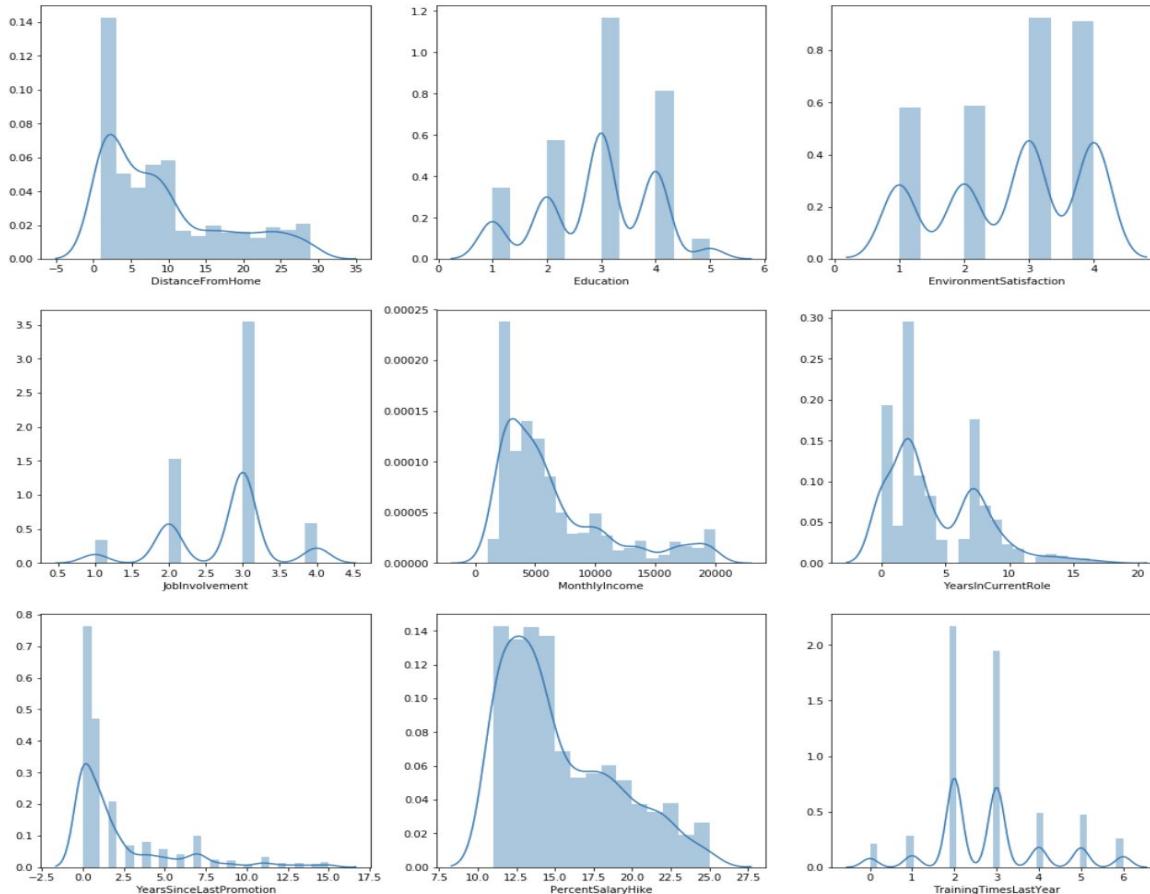


BRC DATASET



- An important aspect to consider is the fact that we are dealing with an imbalanced dataset in both the cases. This is crucial when we consider the accuracy of our prediction models.
- Due to the contrasting nature of the two datasets, we will try to draw suitable conclusions from both datasets wherever possible.

Distribution of auxiliary features - IBM Dataset



- A good practice is to look at the spread of features within a dataset.
- We use the IBM dataset as the data is more clean and thorough.
- The curves are more or less a mixture of gaussians and this is a good indicator that our dataset consists of mainly independent samples and is therefore unbiased.

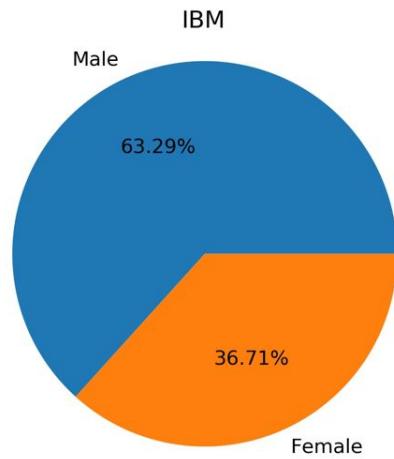
Gender Analysis

Points to consider

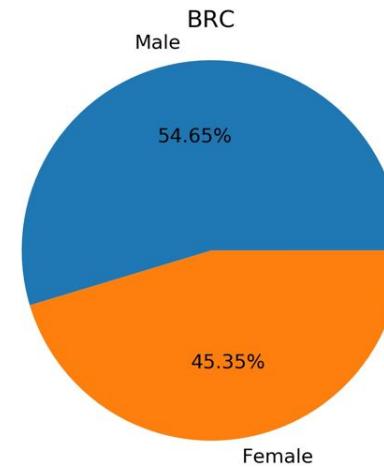
- What is the gender distribution within our dataset of those employees who have quit?
- Are there any disparities with respect to gender and income?
- On an average what is the correlation between gender and years served at the company?

Overall distribution of Gender among Employees who have quit

IBM DATASET

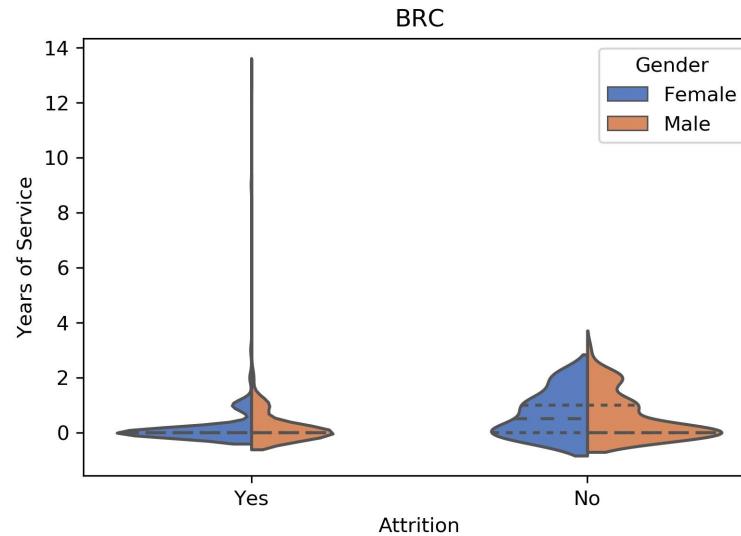
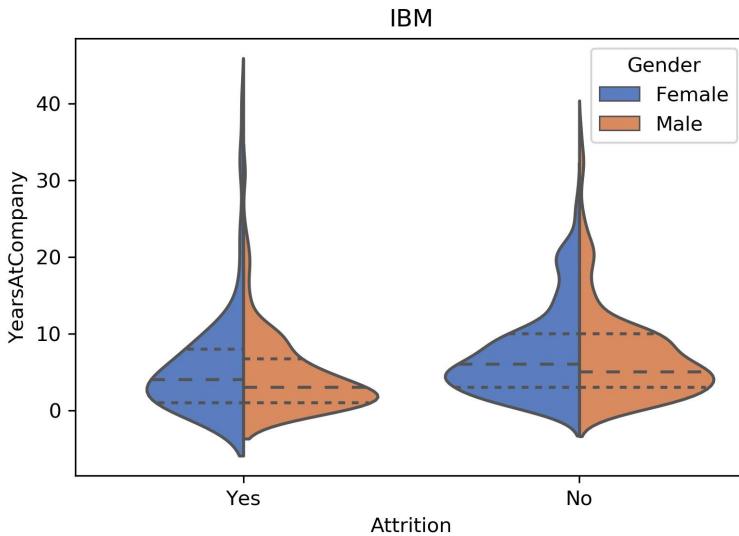


BRCA DATASET



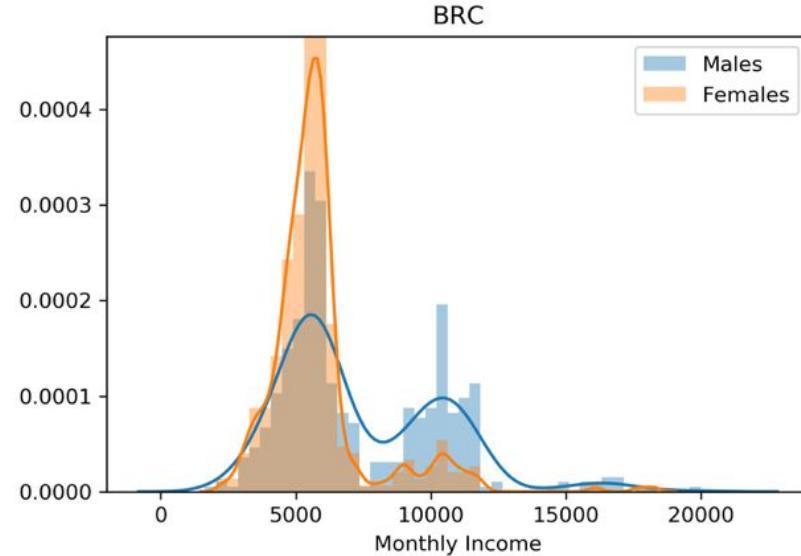
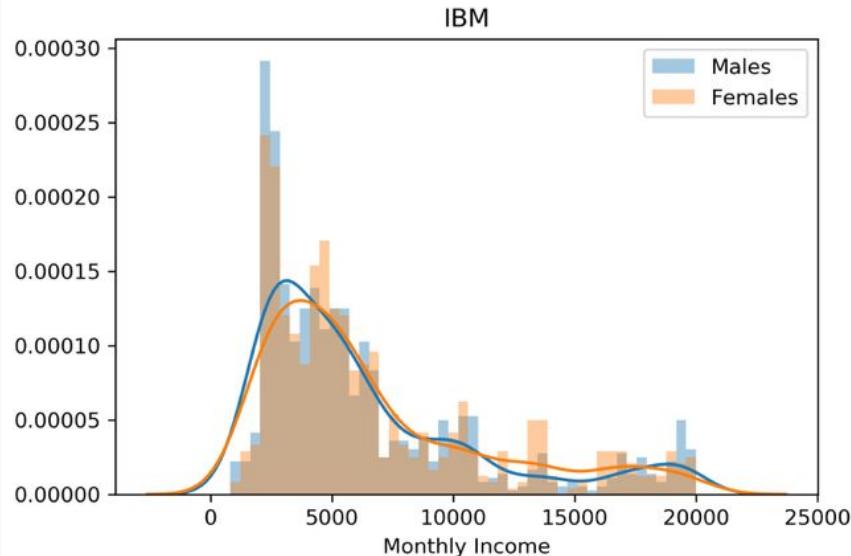
- In general the percentage of males in our dataset is more than females.
- The difference is not enough to be statistically significant though, so we can say with a fair amount of confidence that both our datasets are balanced with regard to gender

Correlation between Gender and Years of Service



- We see that on an average years of service doesn't depend much on gender.
- However we note that the longer an employee has served in a company, the less likelier he is to quit which makes a lot of intuitive sense!

Correlation between Gender and Income



- We observe that in the IBM dataset there's not much correlation between gender and income.
- However in the BRC dataset males are paid a little more than females.
- An interesting conclusion to note can be that the percentage of females who have left the company is higher when we consider the BRC data in contrast with the IBM dataset.
- This naturally suggests that income influences attrition!

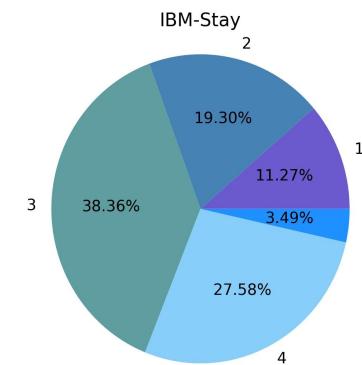
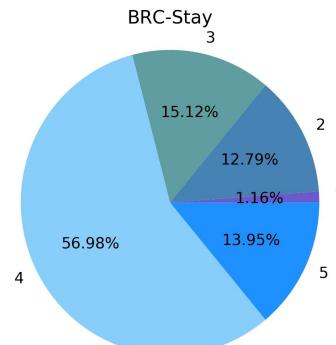
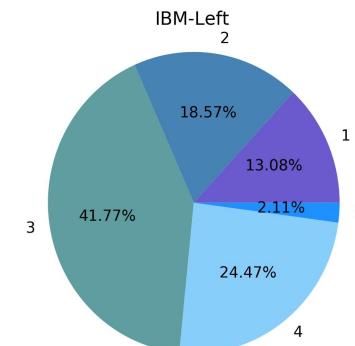
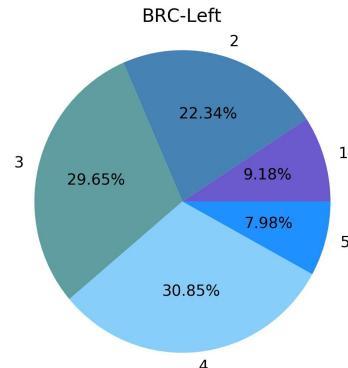
Analysis by Age and Educational Background

Points to consider

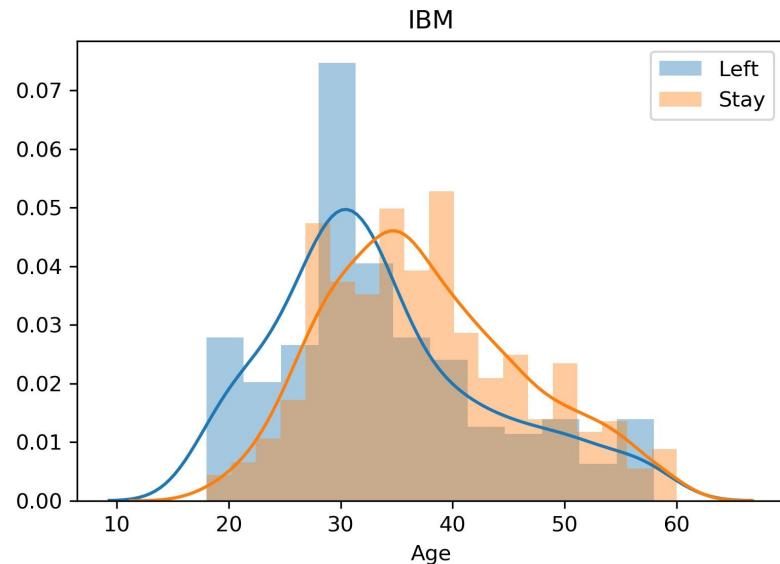
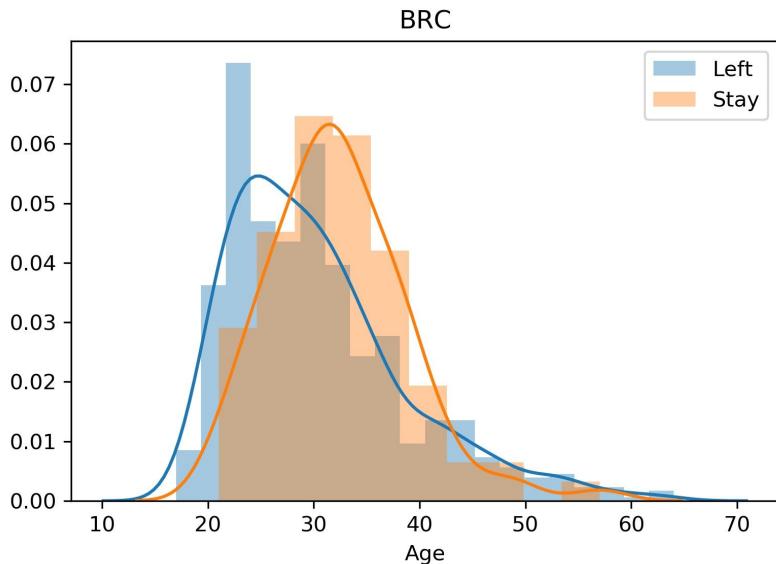
- Is there any link between age of an employee and their likelihood to quit?
- What is the correlation between an individual's education level and their attrition rate?

Correlation between Educational level and Attrition rates

- Education level is categorized from 1(High School) to 5 (Ph.D.)
- We observe that overall education level is not a firm determiner of attrition status.
- To summarize a highly educated employee is not any less likely to leave the company as compared to a less educated employee and vice versa!



Age vs Attrition - Are millennials more likely to quit?



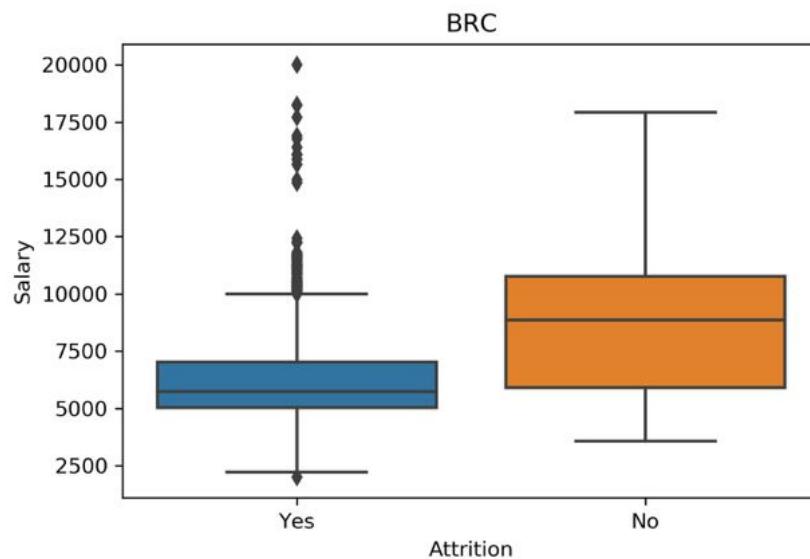
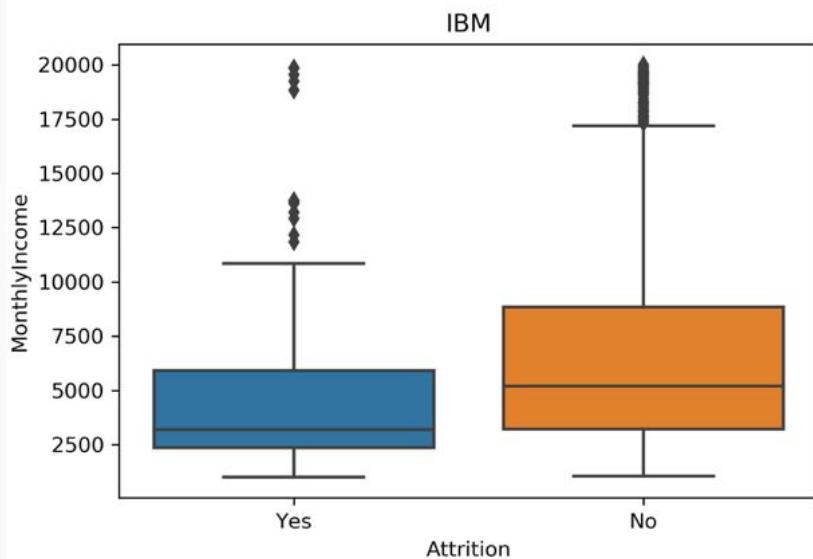
- We observe that there IS a correlation between age and attrition!
- The average age of employees who eventually leave the company is lesser than the employees that stay.
- One possible reason for this observation could be that older employees look to settle down which is in contrast with their younger counterparts who are still looking for a job that they are passionate about.
- Still begs the question whether there are other factors at play?

Analysis by Income - Just how important IS money?

Points to consider

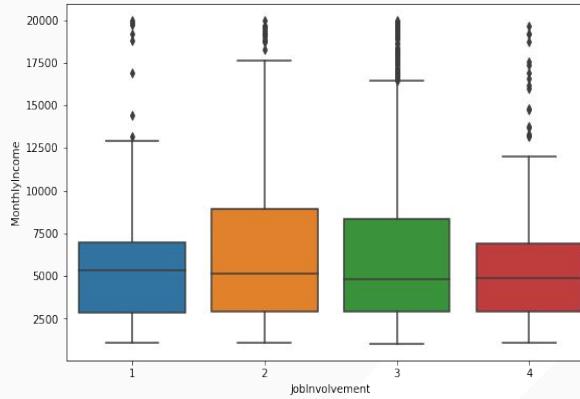
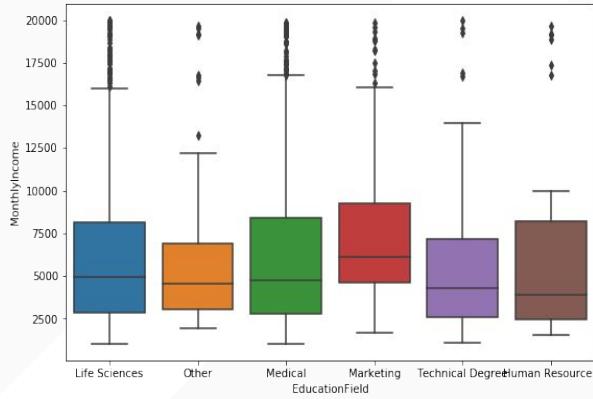
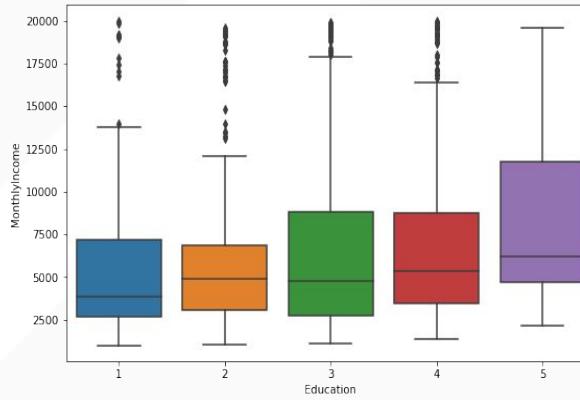
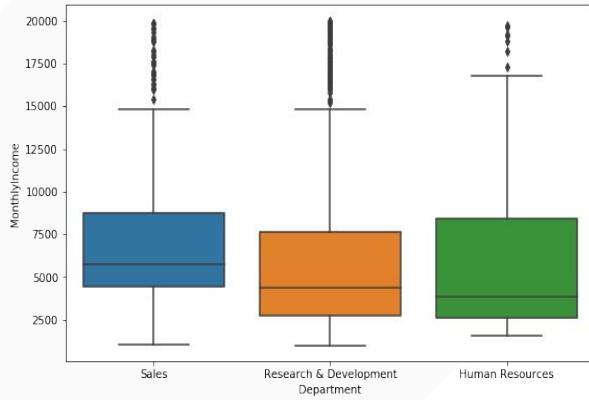
- How does monthly income affect attrition?
- What factors does monthly income depend on? Is there any correlation between age, job role and education level on income?
- In the coming slides we shall try to answer these questions.

Monthly Income vs Attrition



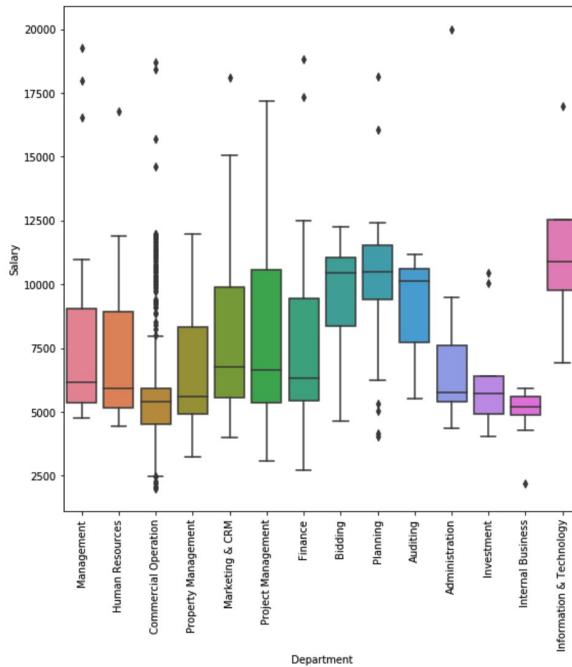
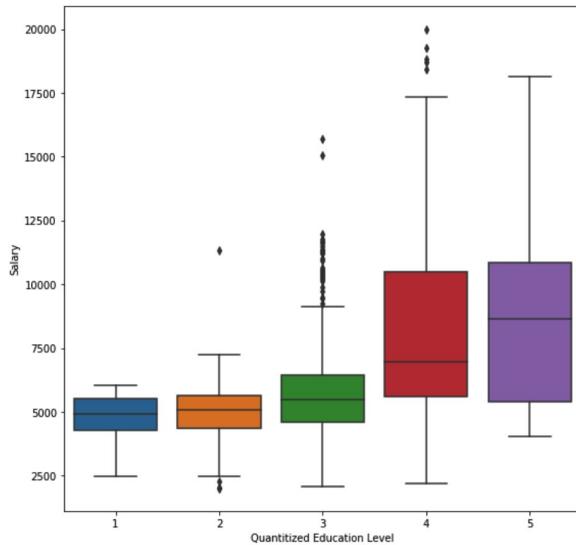
- The average income of the employees who quit is considerably lesser than those who stayed on.
- It is extremely evident from the above plots that income plays a very clear role in determining attrition!
- It now becomes crucial to consider the factors that income depends on.

Income vs Background/ Technical Areas-IBM Dataset



- There is positive correlation between education level and income.
- The income spread about different educational fields are more or less the same
- Counterintuitively higher job involvement does not always correspond to higher pay!
- Employees in Sales are often paid higher than other sectors.

Income vs Background/ Technical Areas-BRC Dataset



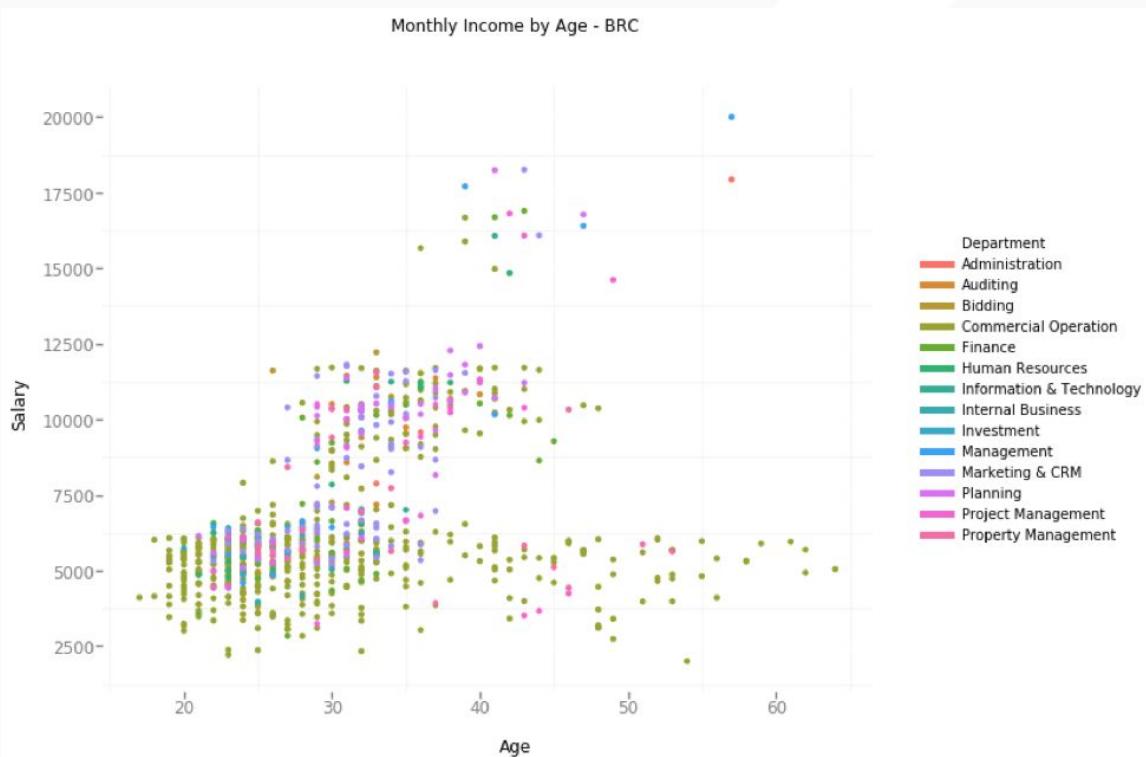
- Education level is positively correlated with monthly income.
- On average the employees in IT sector receive the highest pay in BRC.

Income vs Age - IBM dataset



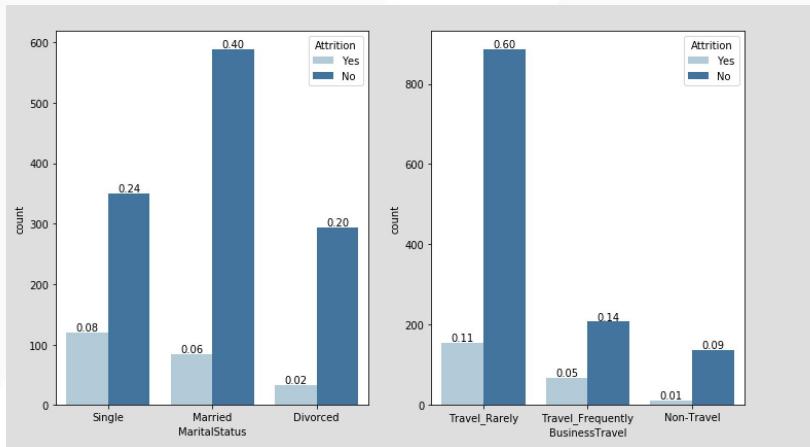
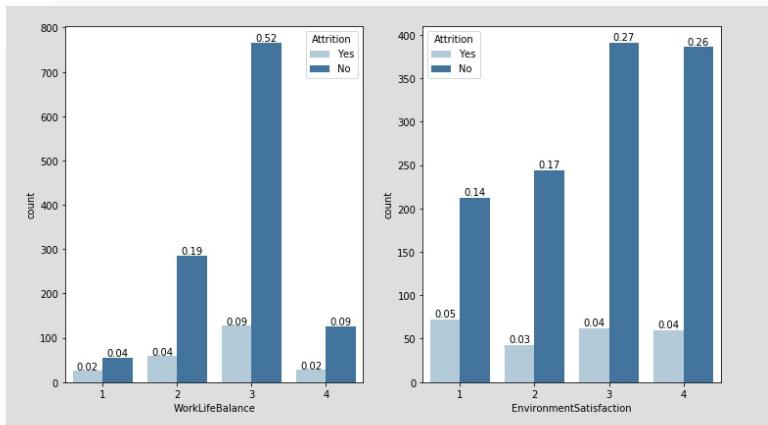
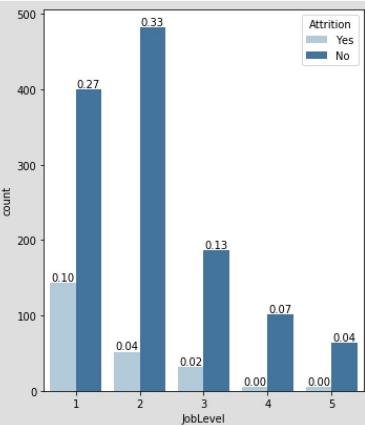
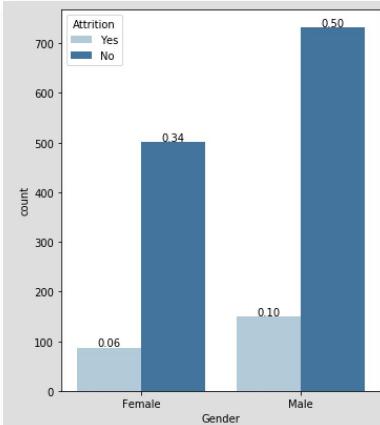
- Younger employees are typically paid lesser than the more experienced employees.
- This can be accounted for, by observing that more experienced employees tend to hold more prestigious job roles.

Income vs Age - BRC dataset



- The similar trend is also observed in the BRC dataset.
- BRC's salary distribution is more hierarchical

Attrition vs Auxiliary features - IBM dataset



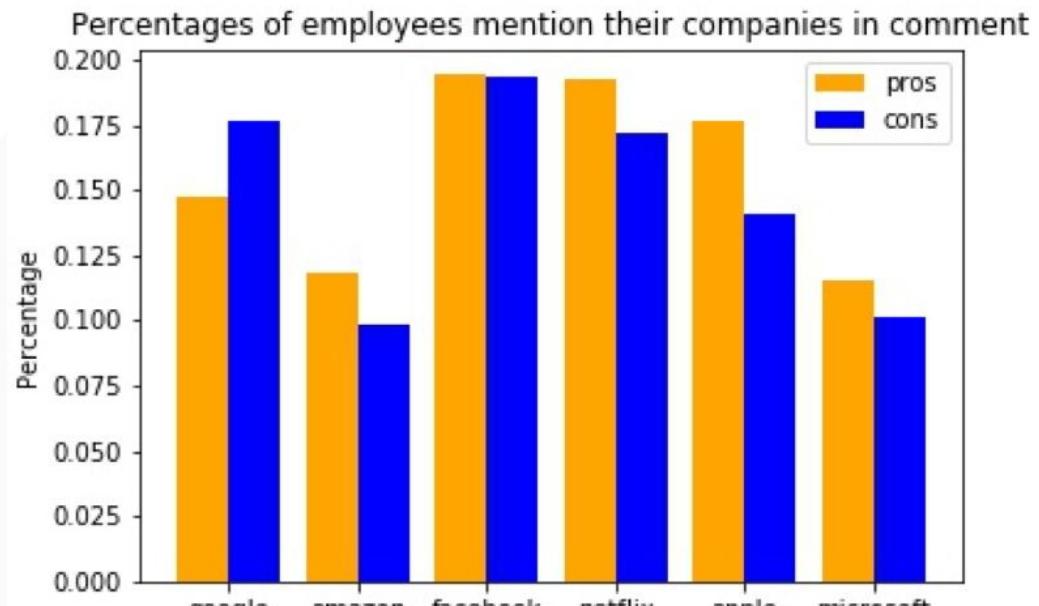
- Those with more demanding work have a much lower attrition rate.
- Work life balance and Environment satisfaction are independently not very discerning features but contribute to the holistic experience of an employee.
- Those who travel less and are settled down in life tend to show a lower attrition rate as well.

Analysis of Employee reviews

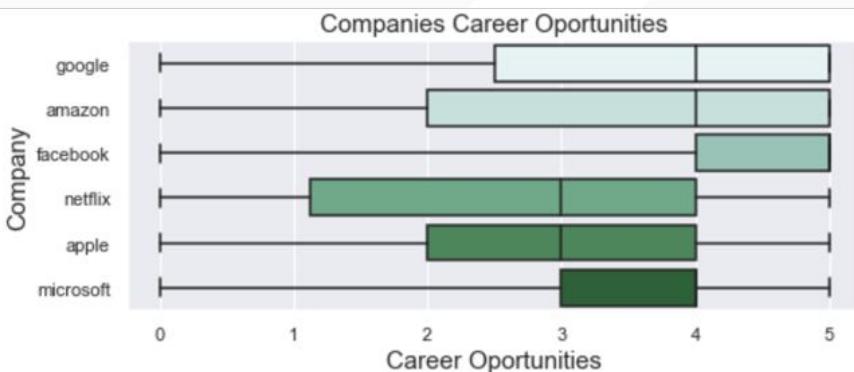
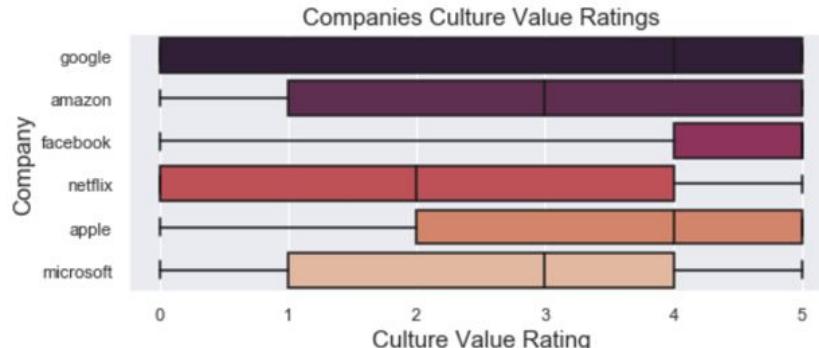
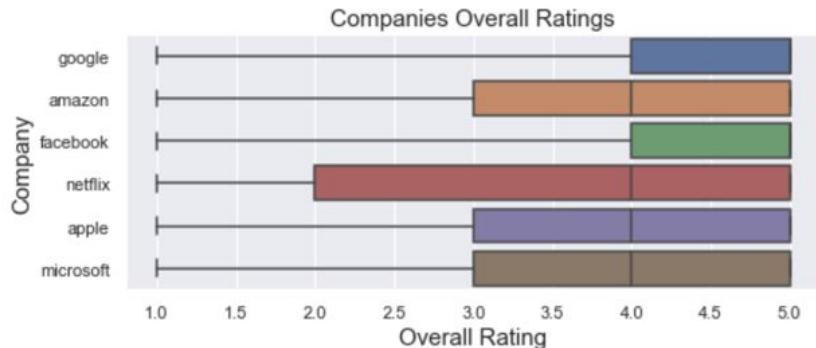
- This presentation would not be complete without taking into account of what the opinions of current employees are, and determining their likes and dislikes about their company.
- From a dataset sourced from Glassdoor of 67000 employee reviews we have analyzed and determined what the employees feel about their company and our aim is to suggest preemptive measures to prevent employee attrition.
- Using the concept of threading and distributed systems we were able to analyze this huge voluminous data.

Review Dataset - How do People Mention Their own Companies

- Employees of Amazon and Microsoft tends not mentioning their companies in reviews
- Different from others, Google' employees are more willing to mention their company when talking about disadvantages



Review Dataset-Ratings on different features



❖ From most rating range, Facebook might be the best one!

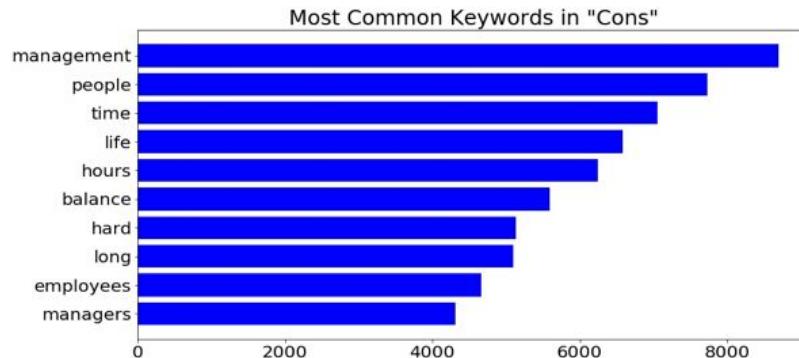
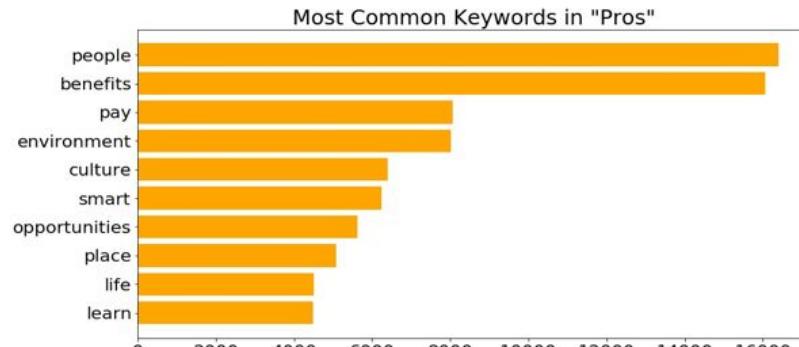
Summary of People's comments

- Common 'good' features:

Colleague, Benefits, Salary, Environment,
Culture, Opportunities, Place, Learn, Life

- Common 'bad' features:

Management, Colleague, WorkingTime,
Life, Balance, Hard



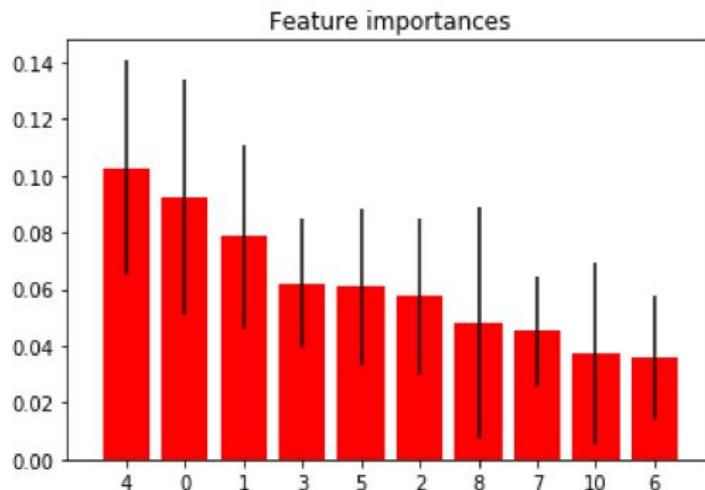
Contents

- Introduction
- Dataset
- Analysis
- Prediction
- Conclusion

Feature Importance via Extra Decision Tree Classifier

Feature ranking:

1. feature 4 - MonthlyIncome (0.102904)
2. feature 0 - Age (0.092525)
3. feature 1 - DailyRate (0.078458)
4. feature 3 - HourlyRate (0.062046)
5. feature 5 - MonthlyRate (0.060752)
6. feature 2 - DistanceFromHome (0.057355)
7. feature 8 - TotalWorkingYears (0.048040)
8. feature 7 - PercentSalaryHike (0.045227)
9. feature 10 - YearsAtCompany (0.037490)
10. feature 6 - NumCompaniesWorked (0.035893)



- Performance analysis on the IBM dataset using extremely randomized trees.
- Features ranked on the basis of Gini importance which counts the number of times a feature is used to split a node weighted by the number of samples it splits

Contents

- Introduction
- Dataset
- Analysis
- Prediction
- Conclusion

Conclusion

Contribution 1

Major reasons to quit

- Age of the employee : The employee might be relatively young and still exploring his/her passions
- Salary : Employees should be paid competitive wages as this is a very crucial factor that determines attrition rate

Contribution 2

Results given by the Machine Learning Model

- We see that the features deemed important by decision trees are very similar to what we concluded by our data visualization.
- Using these key features we were able to achieve a testing accuracy rate of 86.34% in determining attrition.

Contribution 3

Measures to be taken to prevent attrition

- Employees must be paid competitive wages!
- The younger employees must be allowed to explore a variety of departments within the company to ensure that they have the requisite freedom to explore their passions.
- Travel time for employees needs to be cut down as this contributes to a considerable amount of stress on the employees.
- Younger employees should be given higher job roles if their work is good, despite the fact that they are young.

THANK YOU!

Q & A