Reducing Employee Attrition

A data scientist approach

To understand *how* to stop a wave of departures from happening,

we first need to understand *why* waves happen in the first place

Introduction:

- Employees are the most valuable asset of a company.
- They are the most essential contributors toward profits and shareholder value.
- Employee churn is basically voluntary or involuntary departures of workforce.This is also called as employee turnover.
- It barely matters if a company's employee turnover is above or below average, its workforce becomes temporarily less efficient every time someone leaves.

Why we need to address this issue

1) **Reduction In The Cost of Employee Turnover** :

- Hiring a new employee is costly it includes advertising, interviewing, screening, hiring and on-boarding costs.
- The average replacement cost of a salaried employee to be six to nine months' salary or 16 percent of annual salary for high-turnover, low-paying jobs (earning under $30,000 a year).
- Retaining existing talent is much more cost efficient.

2) Overhead to existing employees:

- Before the new employee is hired, though, the open position will probably be covered by other employees, diverting them from their regular work or requiring overtime.
- Or activity is simply not done. If it's a sales position, that's a direct hit to the top line. Longer lead times on products could be costly as well.

3) **Tarnished company image:**

- It's the satisfaction level of your employees that matters the most. So, if an employee isn't happy, she might spread a negative word about the organization, even after leaving it.
- What's more, is that an unhappy employee will lack motivation and will not perform well, leading to unsatisfactory performance. This results in unachievable performance targets, low profits, and employee churn.

The questions we asked:

What factors are leading Employee to dissatisfaction and might make him leave in the near future?

Can these factors be analysed beforehand and proactive measures can be taken to help reduce employee turnover?

Methodology:

Dataset used: IBM Attrition dataset available on Kaggle

- The dataset consists of 1470 records and 35 variables,
- The target variable is “Attrition”
- Some of the features are: business travel, income, education. Job roles, overtime, stock-option-level, years with current manager , education etc:

Model Used: Random Forest

Evaluation Metrics: confusion matrics, roc_auc score

IDE : anaconda (jupyter notebooks)

1) Data Wrangling:

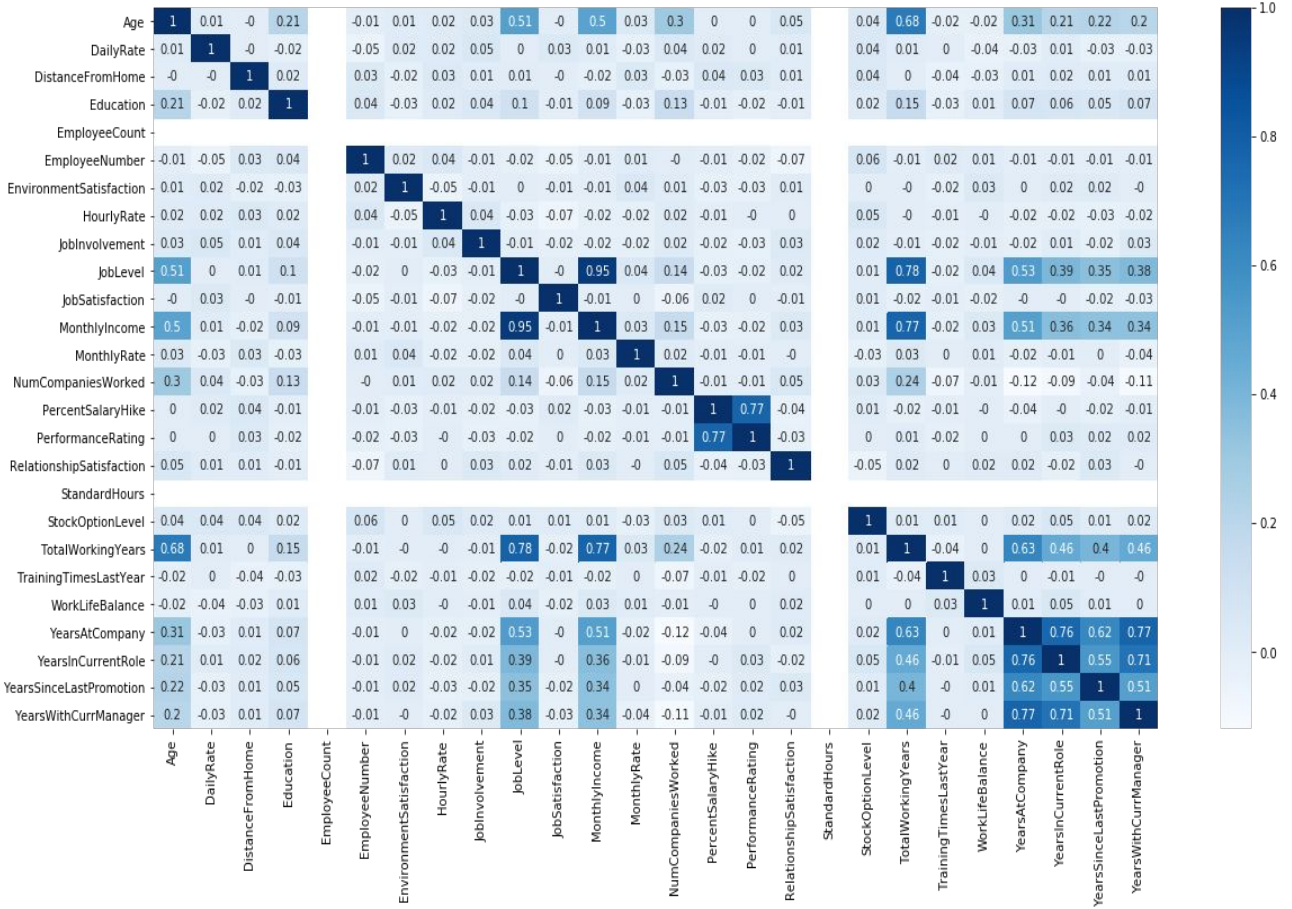
- First we checked for missing values(nan, 0 or any other value by checking unique value counts): None
- we checked outliers: none
- visualised correlation between variables:

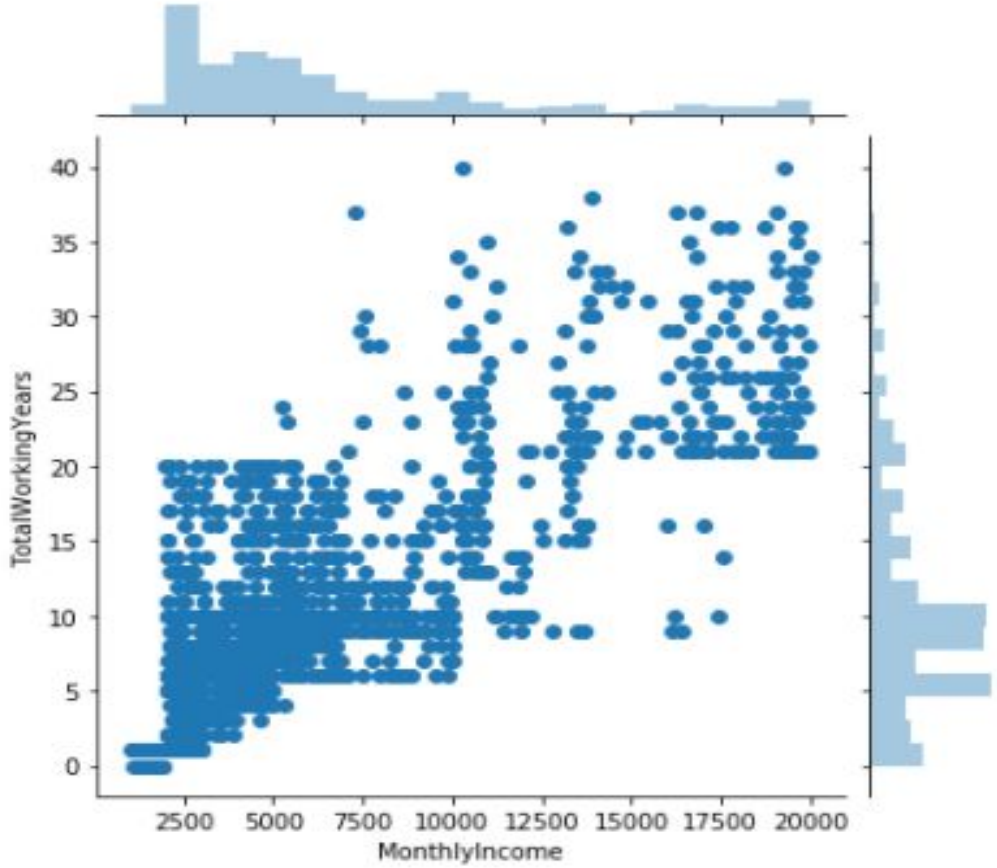
1. There is a high correlation between "joblevel" and "monthly income". which is plausible as higher your job level, higher income you get..another good correlation is between job level and total working years (more experience tends to better level and hence better monthly income)
2. columns such as Employeecount, Over18, StandardHours has same value for all employees with no new value so dropped those columns
3. just 2 values for performance rating....most people got 3 (3---4)
4. people with no stock option have higher attrition..
5. attrition is high in employees in their first year of joining, then it again increases around 5th year. and then goes down after 10 or 11 years of work experience
6. Age data for attrition is right skewed..people in their early career leave jobs more often as compared to later ones.. attrition is high around age 27-30.
7. Most of the people left job either the same year they got promoted or just an year after that.

- converted categorical variables to numerical variables using one hot encoding.

Data exploration

There is a high correlation between "joblevel" and "monthly income". which is plausible as higher your job level, higher income you get..another good correlation is between job level and total working years (more experience tends to better level and hence better monthly income)





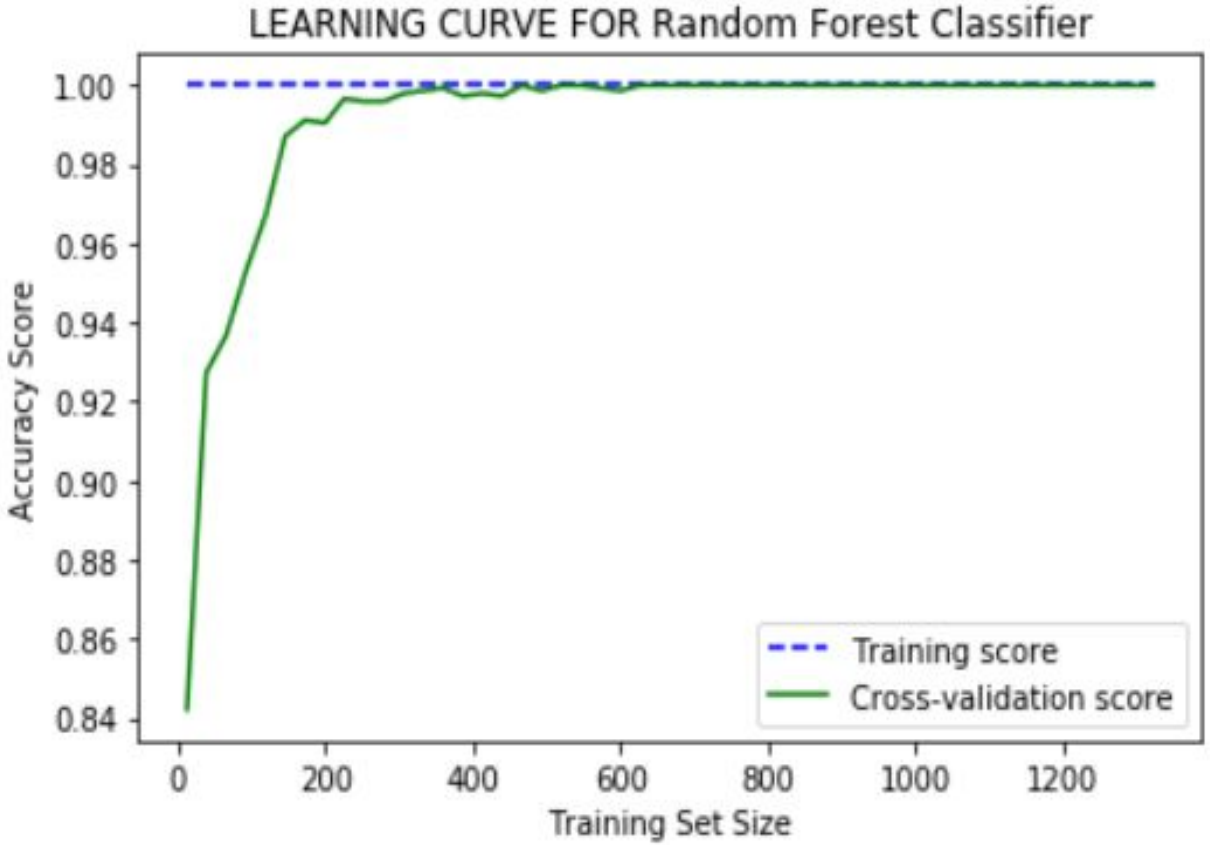
as the number of working years increases the income increases. positive linear but wont fit regression line..

Evaluation of model:

Accuracy Score= 0.8724279835390947

Confusion matrix:

	precision	recall	f1-score	support
1	0.83	0.14	0.24	70
2	0.00	0.00	0.00	0
3	0.00	0.00	0.00	0
avg / total	0.83	0.14	0.24	70


LEARNING CURVE FOR Random Forest Classifier
Accuracy Score
1.00
0.98
0.96
0.94
0.92
0.90
0.88
0.86
0.84
0
200
400
600
800
1000
1200
Training Set Size
Training score
Cross-validation score

Most Important Features:

```
'OverTime_Yes', 0.05265691806627671),
'MaritalStatus_Single', 0.022133027953885248),
'MaritalStatus_Married', 0.007741567972110167),
'JobRole_Sales Representative', 0.009442143756342702),
'JobRole_Sales Executive', 0.006982511928536768),
'JobRole_Research Scientist', 0.0065205178594668475),
'JobRole_Research Director', 0.0009879775515702271),
'JobRole_Manufacturing Director', 0.0027888066249858734),
'JobRole_Manager', 0.0012279263691265983),
'JobRole_Laboratory Technician', 0.00889099269370487),
'JobRole_Human Resources', 0.0033989847847188353),
'Gender_Male', 0.0094197911218145982),
'EducationField_Technical Degree', 0.00956722396286529),
'EducationField_Other', 0.00363118516657585),
'EducationField_Medical', 0.0082993978577814O5),
'EducationField_Marketing', 0.007233379421285697),
'EducationField_Life Sciences', 0.0079516809997849),
'Department_Sales', 0.00921384564428427),
'Department_Research & Development', 0.00920311239992901),
'BusinessTravel_Travel_Rarely', 0.0077923512736848876),
'BusinessTravel_Travel_Frequently', 0.0132017797O315635),
'YearsWithCurrManager', 0.033540606566684124),
'YearsSinceLastPromotion', 0.024392371459546747),
'YearsInCurrentRole', 0.031143866707609814),
'YearsAtCompany', 0.04180079410033298)]
```

Conclusion

we mounted a study to investigate the motivations of employees to stay or leave and reasons behind It.
This data was then fed into a Random Forest Classifier for training.

“ Our main aim is to develop an automatic system that takes survey forms from employees and figure out if employee is dissatisfied and has intent to leave and the reason that is making him/her do so, then organization can act to reinforce the right reasons to make them stay and stop reinforcing the wrong reasons. “

In other words, they can take a positive approach to managing retention in present , which will be more effective over the long run than the ordinary approach of simply conducting the exit interview in order to hope for reducing future turnover so that others don't leave.