Human Resources (HR) Strategy Analysis for Employee Retention

Arvind Ramkumar

Data Scientist | Data Analyst

Approach and Methodology

Employee Retention is a much necessary issue that a HR Department has to be worried about. There are various factors that affect the employee and eventually making them leave/ stay in the same office. But, investing so much on the employee for hiring them and training them, the companies need to make sure that they don't lose these assets. So various factors were studied and analyzed, and the approach taken is as follows.

1. Python (Jupyter Notebook) and Excel (Pivot tables) were used for the analysis.
2. Initially all the factors were individually studied against the Attrition. The major inference that was observed were that
   a. Average Age of Employees who leave the company is much lesser than the Average age of employees who do not leave. This is because, Employee at less age, tend to shift between jobs for a better relevant jobs and better paid jobs
   b. Average distance traveled by employees who left is higher than the Average distance traveled by employees who haven't left the job. This is because, if their home is located at a very longer distance, they might find it hard to commute and higher possibility for moving to another job which is closer to where they live.
   c. With the perspective of the job level, the Job Level seem to have a higher concentration of people leaving the job. At a low job level, probably the job might be a temporary with low pay. So, probably the employee jump between companies
   d. Employees who are much satisfied tend to stay longer in the job
   e. Monthly income has high correlation with the Job level and that seems pretty straight forward
3. Logistic Regression is employed to identify the parameters that are of interest, which is determined using the p-value. We use SMOTE Algorithm to balance the class imbalance in output class Attrition
4. All the important parameters are individually studied. All the continuous variables are converted into categorical, put together as bins and Pivoted using Excel, to see how Attrition is affected based on the categories.

Assumptions

1. All the variables were individually studied. No interaction was taken into study.
2. Only Near Normality was possible, during the t-test analysis of Monthly income to observe whether Average Monthly income of employee who do not leave was greater than the employees who leave
3. For better analysis, all the continuous variables were categorized into bins.

Major Findings

1. **Number of Years** in the company has no significant contribution to the model. Their **Department, Number of companies** worked before and Monthly Income has a very less effect on Attrition or no affect at all.
2. In general, the **Monthly Income** should be a greater contributor to the model to predict the employee retention. But here, we can see that they have a very low contribution to the model (0.0001) and the corresponding p-value is 0.001, which might be insignificant. But this very less significance is because, only one salary range has a higher impact than any other. (Less than $2500) The Employee who leave the company whose salary is < $2500 is 34%.
3. The Contribution of the **Age** to the model was -0.03 and the p value is 0. As expected, the employees with a less age have a higher percentage of leaving the company, than any other age group. This is quite evident because, the employees, at their young age will tend to leave the company real soon looking for promotions and salary hike. This is contrast to the higher age as well. Greater the age, they are more inclined towards resignation and they also have higher chances of leaving. But middle age people tend to stay longer because of job comfort and job security.
4. The **Business Travel** has a significant contribution to the model. For the frequent travelers, the percentage of employee leaving the company is higher than other categories. This might be because, the employee gets tired because of frequent travelling and with a family, they prefer to get settled in a place than traveling.
5. **Distance from home** is a significant factor but has a minimal effect on the model. If the company is located at a very farther distance, then the employees will find it hard to commute and apparently, they might look for a job closer to their work place. But with the given dataset, there is not much inference that could be made.
6. The **Job level** is significant. If the job level is low, the employee is highly likely to leave the job. This is because at the lower job level, they could be temporary employees who shift between jobs also, they could be underpaid for

7. **Job Role** plays a very minimal yet significant role. This is high importance is because of the variation caused by the Sales Representative role. The have a very high percentage of the employee leaving. Also, Human Resources and Lab Technician contribute to the next higher percentage.
8. The **Job Satisfaction** has a very high contribution to the model It is obvious if the job Satisfaction is low, the employee is likely to leave the job. This is because at the lower job satisfaction, they are less likely to continue, and they will try to move to another job where they are better satisfied.

Recommendations

1. Job Satisfaction seems to be the most important. Efforts must be taken to form a committee that will oversee the satisfaction of the employee and helping them to improve their work environment and culture and with anything that's necessary
2. With respect to the job role, Sales representative, HR and Lab Technicians leave the firm at a higher ratio. More consideration should be done on why the employees are leaving. May be the job is not good, or the problem with the managers, or they have a dispute with the pay. Efforts should be taken to minimize this.
3. Monthly income can't be increased to everyone. Sometimes, perks, allowance and incentives need to be given to employee that will make them stay.
4. If the employee is a longer distance from the office, office might offer them reallocation bonus, or travel reimbursement or a bus to pick up and drop the employees as a trip.
5. Business Travel might be possibly reduced, or the employees can take turns to make business trips, allowing will allowances and salary bonus for every trip they make. This will avoid the employee to become fatigue.
6. Hire more employees in the middle age group, will increase their retention.
7. Increase the salary to a competitive salary, and the other companies can't provide salary above that. This will not allow the employee to move out to other firms for better salary.

# Supporting Graphs, Tables and Chart

Parameters of Interest

	coef	std err	z	P>|z|	[0.025	0.975]
const	3.5793	0.285	12.558	0.000	3.021	4.138
x1	-0.0300	0.006	-4.747	0.000	-0.042	-0.018
x2	-0.4261	0.070	-6.062	0.000	-0.564	-0.288
x3	-0.3906	0.119	-3.277	0.001	-0.624	-0.157
x4	0.0309	0.006	5.211	0.000	0.019	0.042
x5	-6.256e-05	7.6e-05	-0.823	0.411	-0.000	8.65e-05
x6	-0.8275	0.142	-5.838	0.000	-1.105	-0.550
x7	0.0995	0.026	3.792	0.000	0.048	0.151
x8	-0.5353	0.043	-12.361	0.000	-0.620	-0.450
x9	0.0001	3.33e-05	3.320	0.001	4.52e-05	0.000
x10	0.0620	0.021	2.977	0.003	0.021	0.103
x11	-0.0076	0.011	-0.712	0.477	-0.028	0.013

	Name of the Variable
x1	Age
x2	BusinessTravel
x3	Department
x4	DistanceFromHome
x5	EmployeeNumber
x6	JobLevel
x7	JobRole
x8	JobSatisfaction
x9	MonthlyIncome
x10	NumCompaniesWorked
x11	YearsAtCompany

Most Interesting Basic Visualization

Box Plot of Age

Box Plot of Monthly Income

Box Plot of Distance from Home
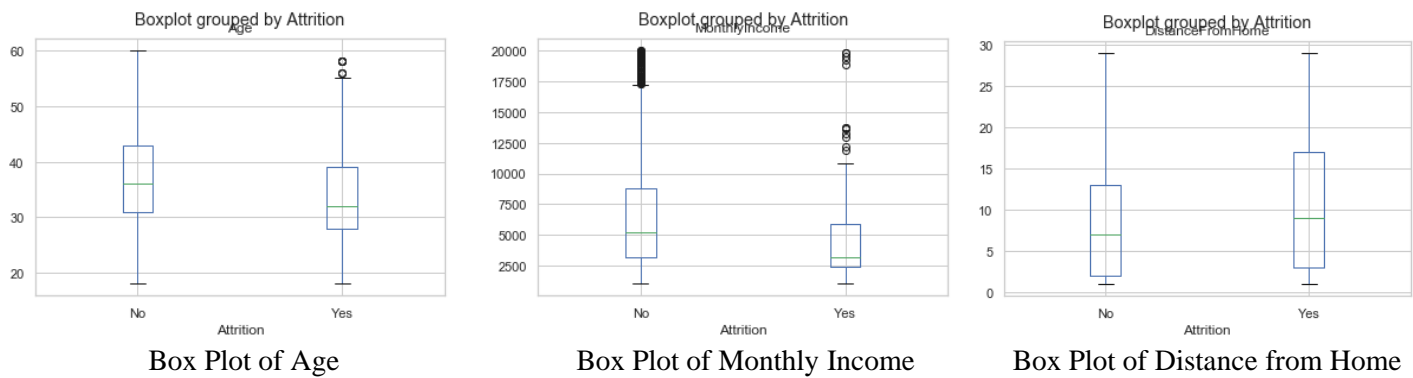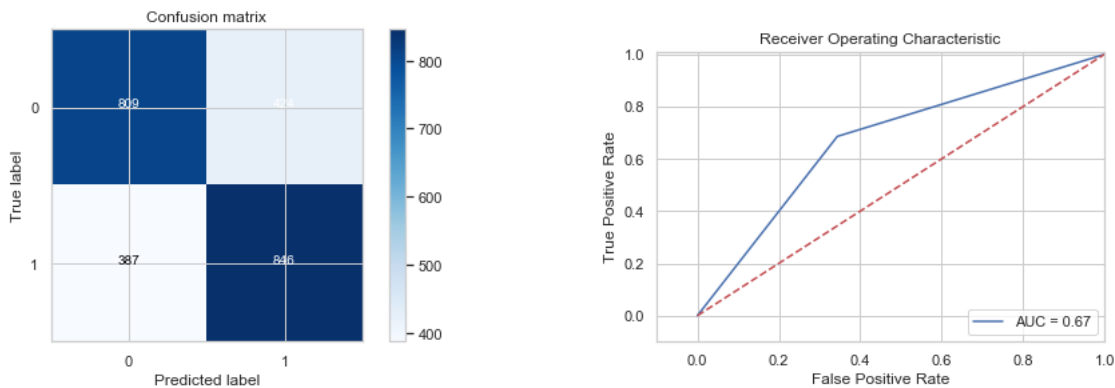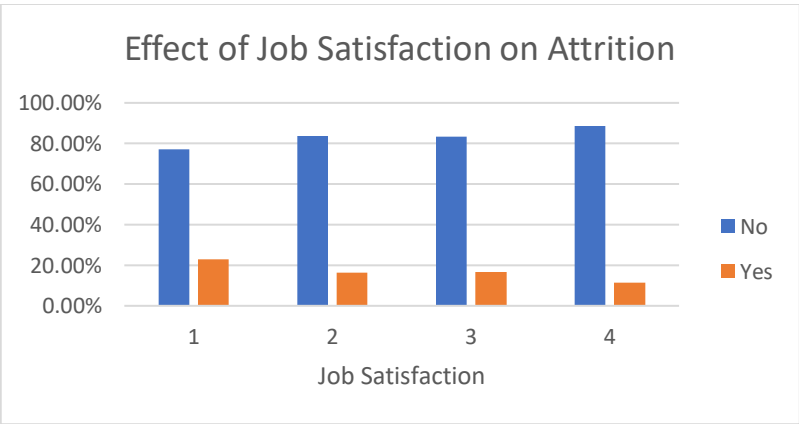
Class Imbalance of Attrition
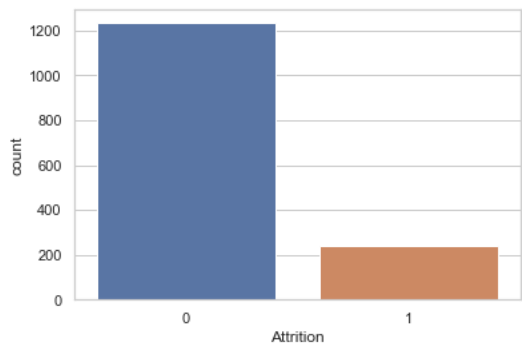
Confusion Matrix and ROC Curve of Logistic Regression Model

In depth Analysis of Most Important Factor


Effect of Job Satisfaction on Attrition

Count of JobSatisfaction	Column Labels	
Row Labels	No	Yes
1	77.16%	22.84%
2	83.57%	16.43%
3	83.48%	16.52%
4	88.67%	11.33%

Effect of Age on Attrition
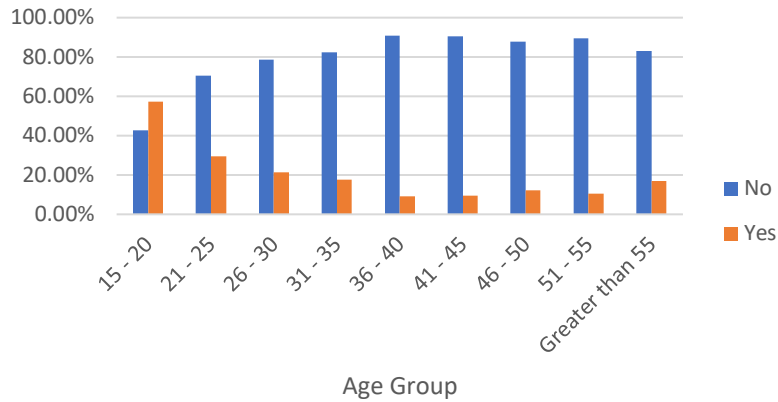
Count of Age	Column Labels	
Row Labels	No	Yes
15 - 20	42.86%	57.14%
21 - 25	70.53%	29.47%
26 - 30	78.71%	21.29%
31 - 35	82.51%	17.49%
36 - 40	90.94%	9.06%
41 - 45	90.63%	9.38%
46 - 50	87.69%	12.31%
51 - 55	89.58%	10.42%
Greater than 55	82.98%	17.02%

Effect of Job Level on Attrition
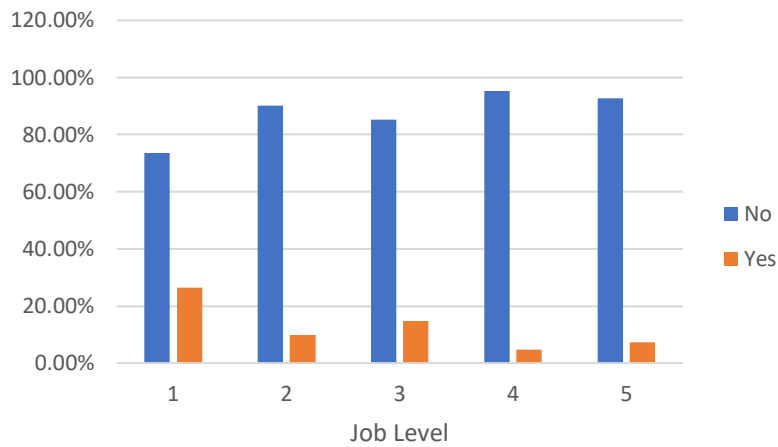
Count of JobLevel	Column Labels	
Row Labels	No	Yes
1	73.66%	26.34%
2	90.26%	9.74%
3	85.32%	14.68%
4	95.28%	4.72%
5	92.75%	7.25%

Effect of Business Travel on Attrition
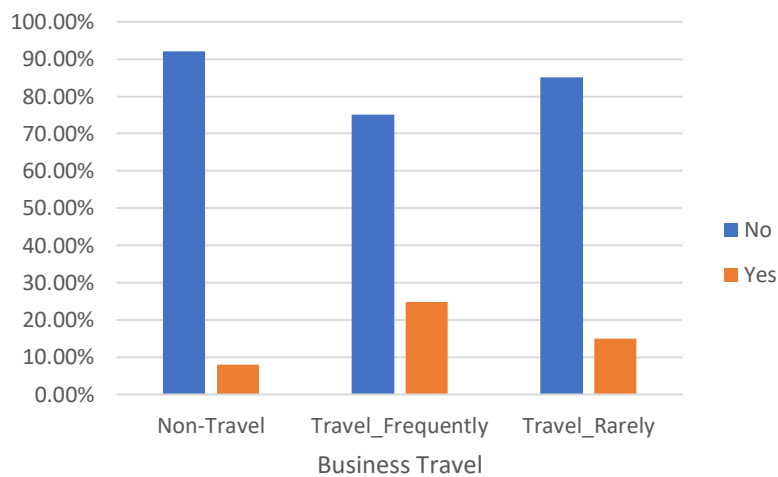
Count of BusinessTravel	Column Labels	
Row Labels	No	Yes
Non-Travel	92.00%	8.00%
Travel_Frequently	75.09%	24.91%
Travel_Rarely	85.04%	14.96%