Employee Attrition Analysis

The cost of training an employee for them to be situated in the company has a financial impact because the time it takes for the new hire takes at least 12 months to be fully productive. When an existing employee decides to leave the company. There are many factors that affect an employee's attrition and performance: family, kids, work-life balance, job satisfaction, complexity of the work, relationship status, etc.

Developing a classifier that more strongly weighs relevant factors using Logistic Regression, Decision Trees, or Random Forest could help solve this problem. This is an invaluable method because, if successful, it will be extremely important to improve employee satisfaction with the company and improve attrition and performance. The scope of this project is to find ways to improve employee satisfaction best and retain attrition. Other companies can benefit from similarly specified factors especially when it comes to improving the company's environment to retain employees.

Data

Having the IBM data as a reference point will allow us to use a representation of the workforce at a large-scale company. Given IBM's growth in products and individuals, IBM would serve as a good representation as a baseline for other companies whether to that size and magnitude or hopes to become of that size. Since the data has many variables as below:

- Work-life balance
- Relationship satisfaction
- Performance rating
- Job satisfaction
- Environment satisfaction
- Education level
- Manager
- Recent promotions

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

Exploratory Data Analysis

We graphed the different variables to determine the attrition of an employee:

- Overtime (figure 1)
- Number of companies worked (figure 2)
- Business Travel (figure 3)
- Stock option (figure 4)

Figure 1

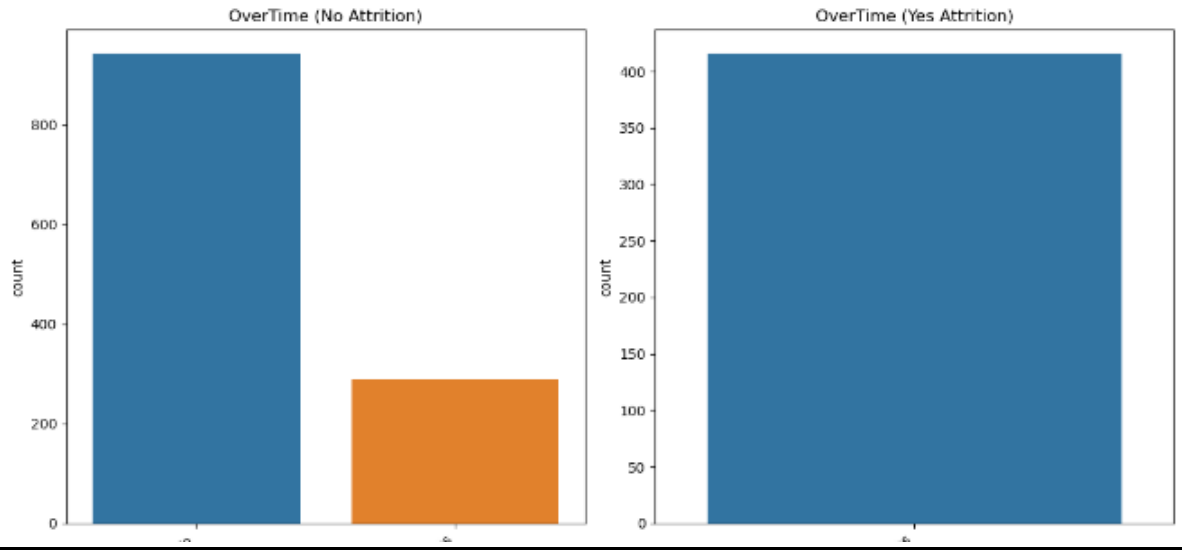


Figure 2

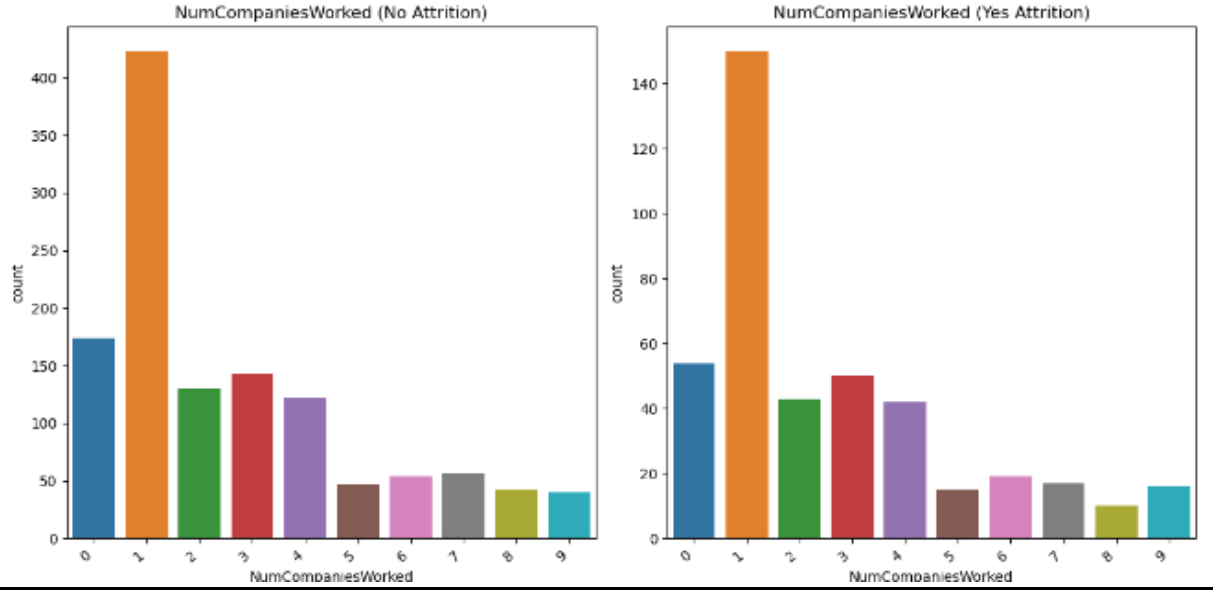


Figure 3

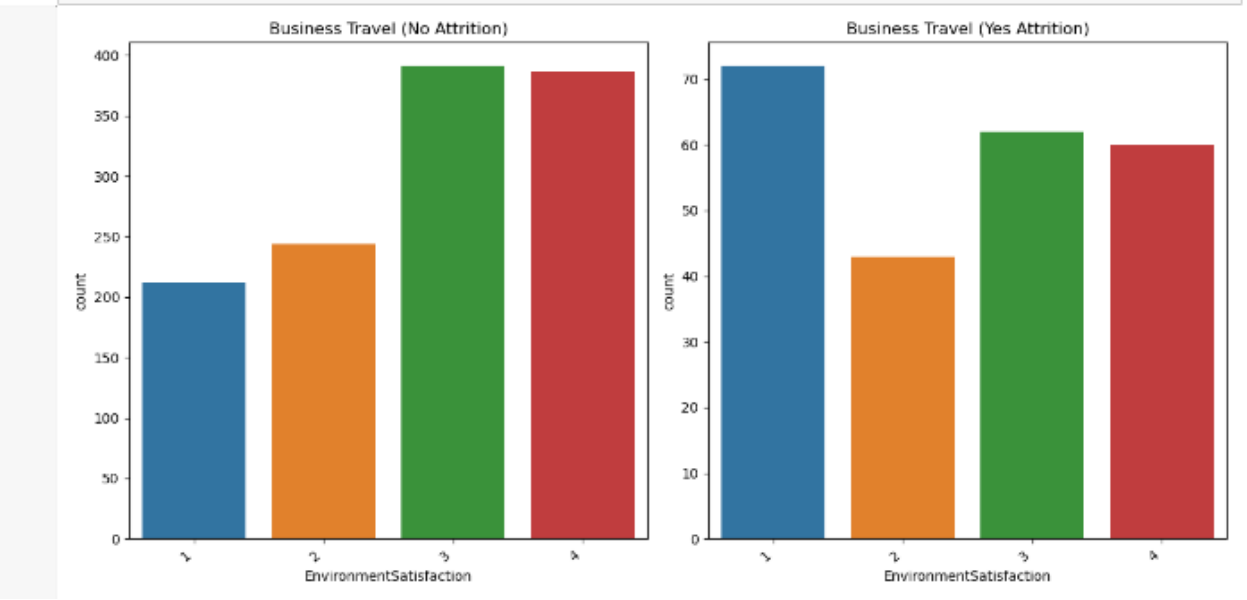
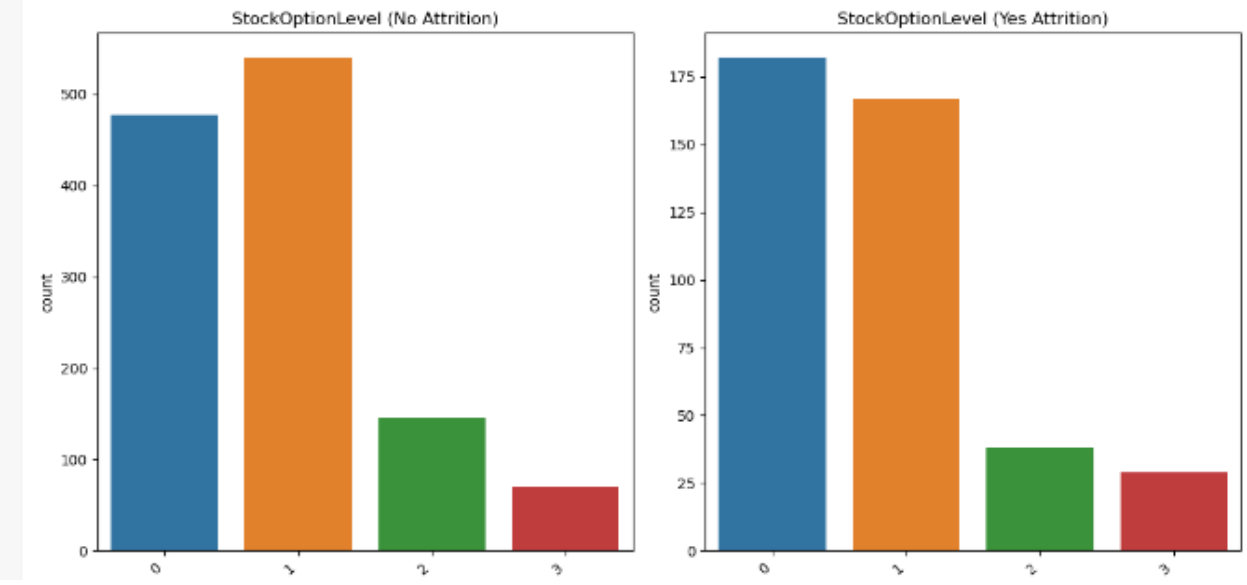


Figure 4



Method/Model

We determined the algorithm and model that would best determine whether an employee would stay at the current company. When finding that, we found that the Logistic Regression was the most accurate model to determine that.

	Algorithm	Model accuracy score
0	Logistic Regression	0.873016
1	KNN	0.863946
2	SVM	0.875283
3	Random Forest	0.863946
4	Gradient Boost	0.873016
5	Naive Bayes	0.773243

Out[89]:

	Algorithm	ROC-AUC train score	ROC-AUC test score
0	Logistic Regression	0.819373	0.750928
1	KNN	0.772569	0.702564
2	SVM	0.819323	0.770563
3	Random Forest	0.789897	0.701586
4	Gradient Boost	0.800767	0.681495
5	Naive Bayes	0.775175	0.695749

Feature Importance

In determining the best attributes/features that would determine whether the employee would stay at the company. We found that the most important features were years with the current manage, years in current r ole, work life balance, years at company, and years since last promotion.

	Features	Importance scores
0	BusinessTravel	0.000000
1	DistanceFromHome	0.000000
2	Education	0.000000
3	HourlyRate	0.000090
4	JobInvolvement	0.003412
5	JobLevel	0.005203
6	JobSatisfaction	0.006807
7	MonthlyIncome	0.010308
8	MonthlyRate	0.011241
9	NumCompaniesWorked	0.016310
10	Over18	0.019039
11	OverTime	0.022374
12	PercentSalaryHike	0.025136
13	PerformanceRating	0.025248
14	RelationshipSatisfaction	0.026968
15	StandardHours	0.027394
16	StockOptionLevel	0.029004
17	TotalWorkingYears	0.031550
18	TrainingTimesLastYear	0.031713
19	WorkLifeBalance	0.032638
20	YearsAtCompany	0.035097
21	YearsInCurrentRole	0.035745
22	YearsSinceLastPromotion	0.036281
23	YearsWithCurrManager	0.043648

Future Improvements

While IBM is a good source of data, having one organization's data prevents the users of the data from predicting for small or medium size companies. As companies continue to grow, there are other variables and factors that could impact an employee's decision to stay at the company such as mission statement, growth, stock price, etc. Incorporating these variables will allow for a holistic view and a better prediction whether an employee stays at a company.