Attrition – A common problem that all the Organization Faces.

Introduction

The key to success in any organization is attracting and retaining top talent. Major task of HR analyst at any organization is to determine which factors keep employees at the organization and which prompt others to leave. They need to know what factors can be changed to prevent the loss of good people.

This problem is there in all the organizations, but that involve lots of analysis to pinpoint the issues, I have used the data from Kaggle to find some of the reasons for the attrition.

Attrition is a big problem of all the organization, sometime people leave their current job, when they are not satisfied with their work or salary or promotion or could be due to some other reason. Human Resource Team collect different data from employees to identify the reasons for attrition.

This data is created by IBM data scientist, and it have several attributes that can help to identify the attrition reasons. Some of the important fields are Business Travel, Department, Distance from Home, Employment satisfaction, Hourly Rate, Job Involvement, Monthly Income, Overtime, Performance Rating, Salary hike, Years Since Last promotion etc.

I am planning to do exploratory data analysis to do the analysis on the attributes, I will also try to find the relationship between attributes and attrition field.

Will plan to fit few models, that can predict an employee if he/she will be having attrition.

Attribute List in the Dataset

Attribute	Datatypes
Age	int64
Attrition	object
BusinessTravel	object
DailyRate	int64
Department	object
DistanceFromHome	int64
Education	int64
EducationField	object
EmployeeCount	int64
EmployeeNumber	int64
EnvironmentSatisfaction	int64
Gender	object
HourlyRate	int64

JobInvolvement	int64
JobLevel	int64
JobRole	object
JobSatisfaction	int64
MaritalStatus	object
MonthlyIncome	int64
MonthlyRate	int64
NumCompaniesWorked	int64
Over18	object
OverTime	object
PercentSalaryHike	int64
PerformanceRating	int64
RelationshipSatisfaction	int64
StandardHours	int64
StockOptionLevel	int64
TotalWorkingYears	int64
TrainingTimesLastYear	int64
WorkLifeBalance	int64
YearsAtCompany	int64
YearsInCurrentRole	int64
YearsSinceLastPromotion	int64
YearsWithCurrManager	int64

Some of the categorical data description is as below.

Education

1 'Below College' 2 'College' 3 'Bachelor' 4 'Master' 5 'Doctor'

EnvironmentSatisfaction

1 'Low' 2 'Medium' 3 'High' 4 'Very High'

JobInvolvement

1 'Low' 2 'Medium' 3 'High' 4 'Very High'

JobSatisfaction

1 'Low' 2 'Medium' 3 'High' 4 'Very High'

PerformanceRating

1 'Low' 2 'Good' 3 'Excellent' 4 'Outstanding'

RelationshipSatisfaction

1 'Low' 2 'Medium' 3 'High' 4 'Very High'

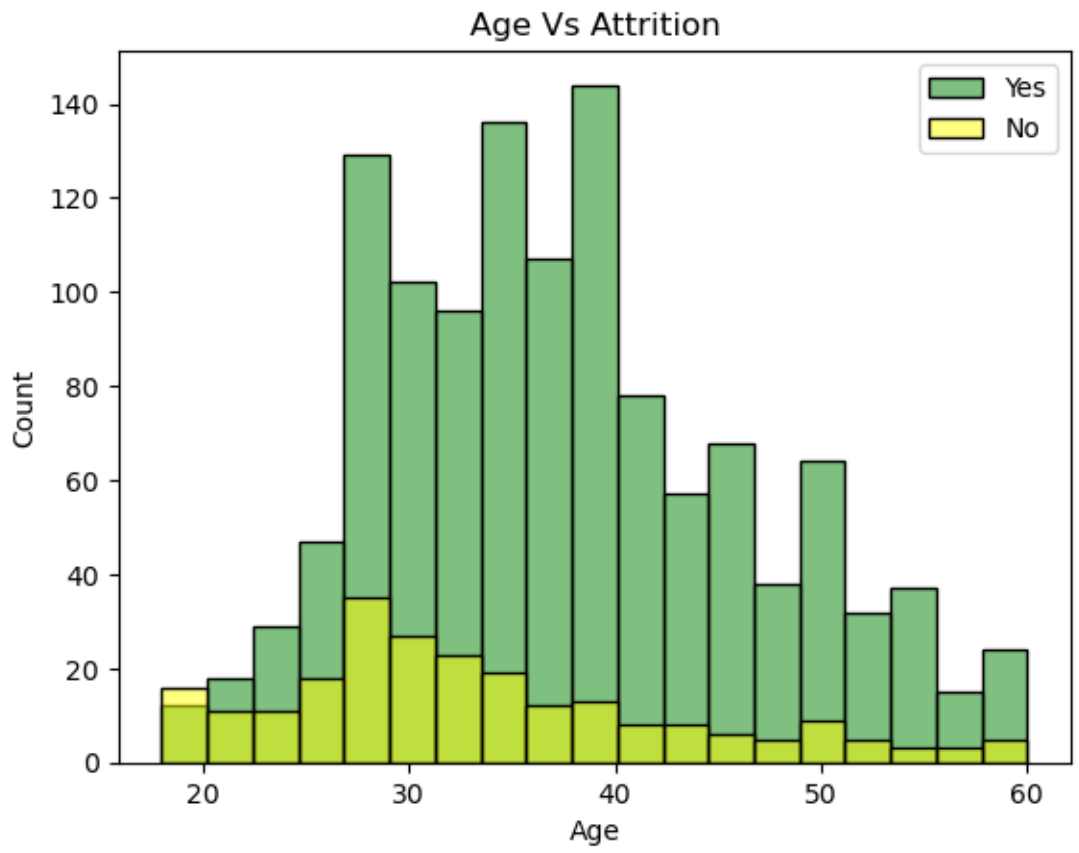
WorkLifeBalance

1 'Bad' 2 'Good' 3 'Better' 4 'Best'

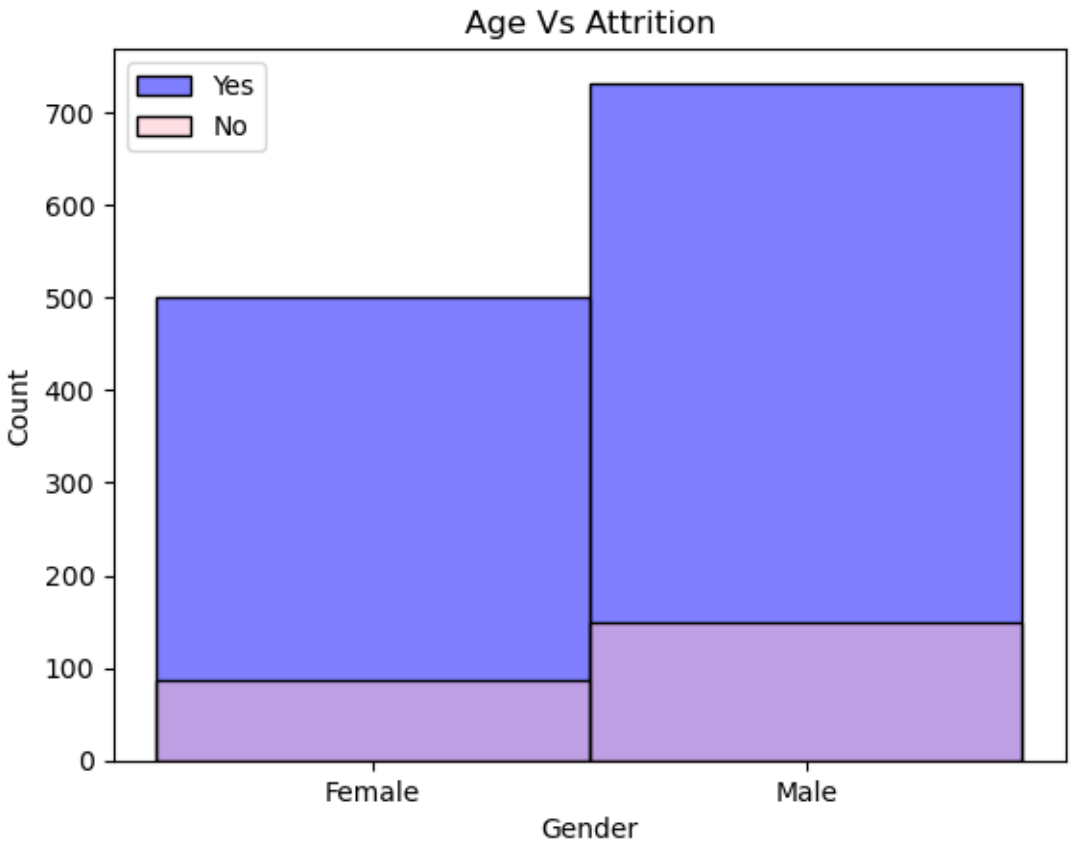
Attrition Description Link: https://www.linkedin.com/pulse/analyzing-employee-attrition-mike-west

Data Link: https://www.kaggle.com/datasets/patelprashant/employee-attrition

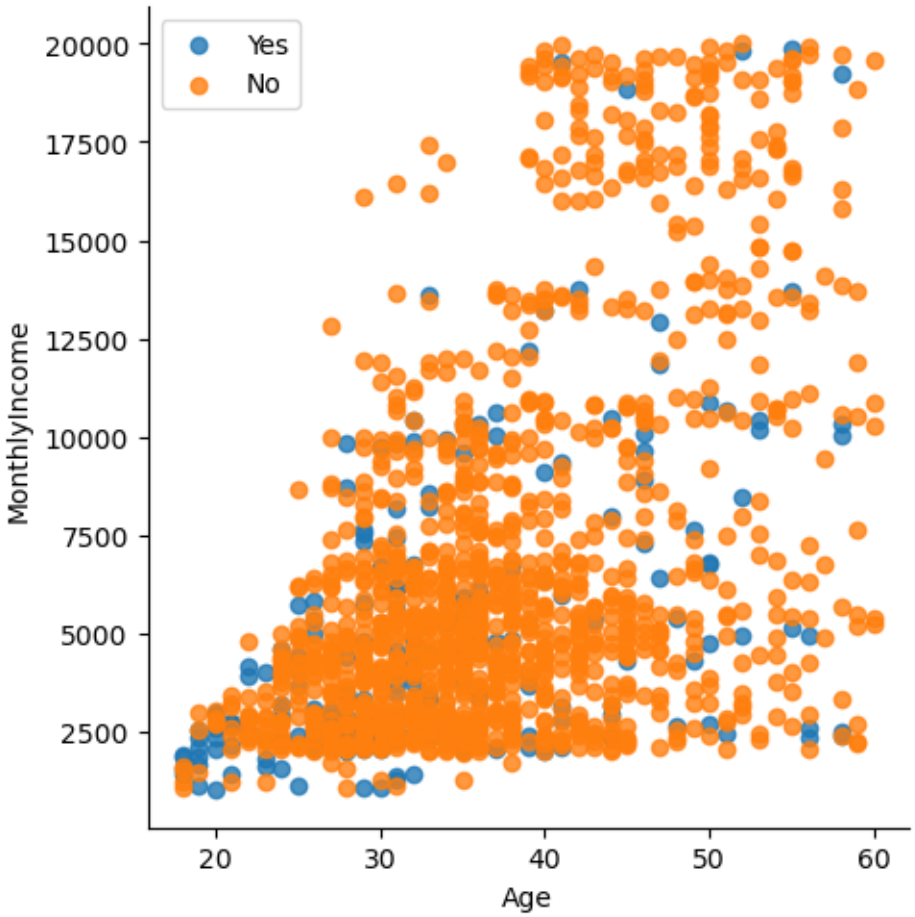
Visual Analysis



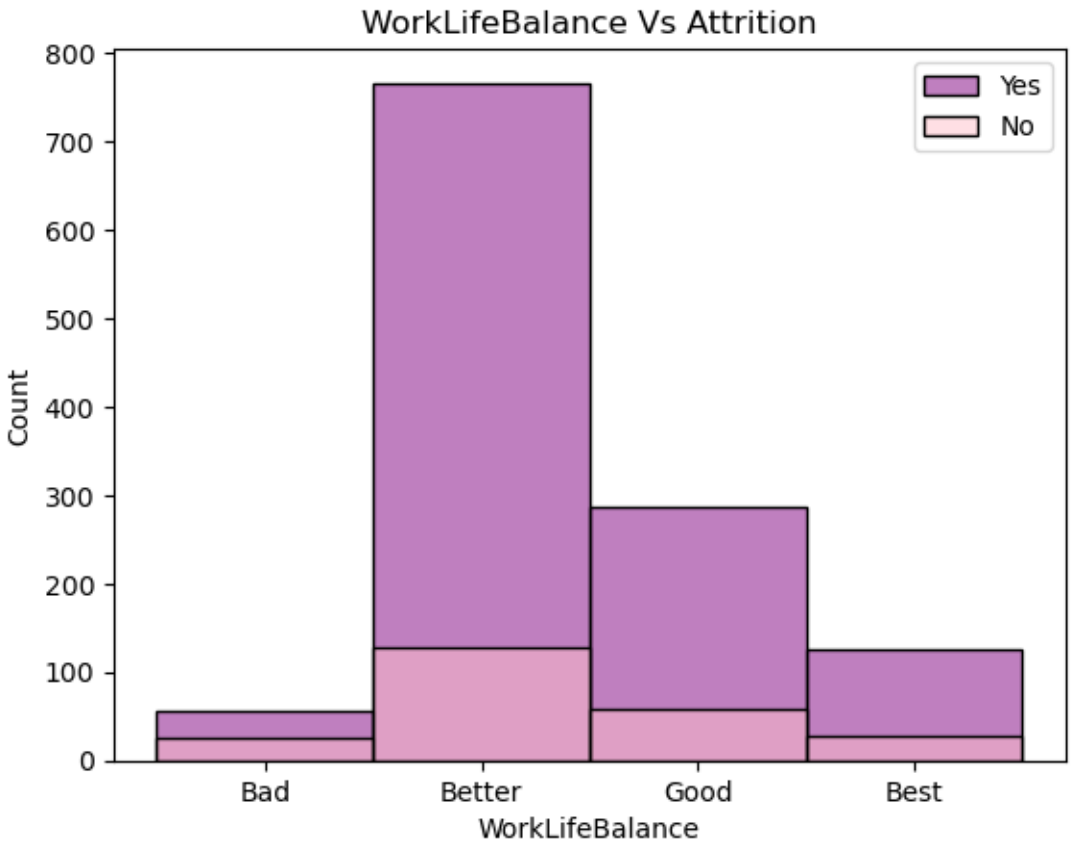
Here is a plot for attrition by age, this plot shows that, the attrition is quite high for the age group of 25 to 40, and attrition became low when the age increase.



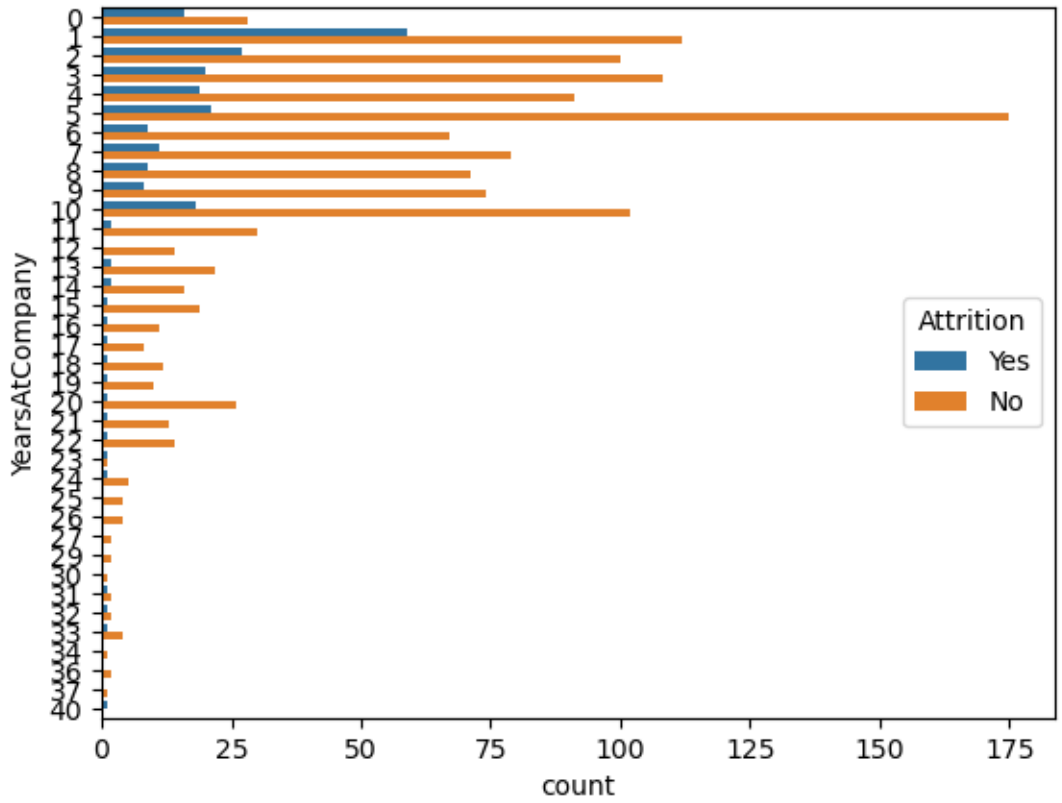
When we do the plot for Attrition by Gender, then it shows that male have higher attrition rate compared to female.



This scatter plot is quite interesting, here I have plotted Age vs Income which clearly shows a pattern that as the age become oler the monthly income increases, but the attrition does not show the same pattern, it shows when montly income is low then the attrition is high because people look for other company to increas the salary, as the salary become higher the attrition rate gradually slow down.



Here I tried to find the attrition rate for the employees, based on their work life balance, I was expecting this attrition rate will be high for bad WorkLifeBalace, but its actully not true, as per the plot the employees who have better worklife balance has really high attrition rate.



Here is the count plot for attrition by years at company, looking at the plot it clearly says when the year at company is low employees are tend to switch jobs and its high around one year working at company, when people start saying longer time, the attrition rate reduces.

Model Fitting for Prediction

I have fit XGBoost model for the prediction purpose to identify the attrition, which give 82% Accuracy for the model fit on the test data.

```
XGBClassifier(base_score=None, booster=None, callbacks=None,
              colsample_bylevel=None, colsample_bynode=None,
              colsample_bytree=None, early_stopping_rounds=None,
              enable_categorical=False, eval_metric=None, feature_types=None,
              gamma=None, gpu_id=None, grow_policy=None,
importance_type=None,
              interaction_constraints=None, learning_rate=0.1, max_bin=None,
              max_cat_threshold=None, max_cat_to_onehot=None,
              max_delta_step=None, max_depth=3, max_leaves=None,
              min_child_weight=10, missing=nan, monotone_constraints=None,
              n_estimators=100, n_jobs=None, num_parallel_tree=None,
              predictor=None, random_state=None, ...)
```

Search, I have fit the model to calculate the R-Square again, it clearly shows that the R-Square has increased a

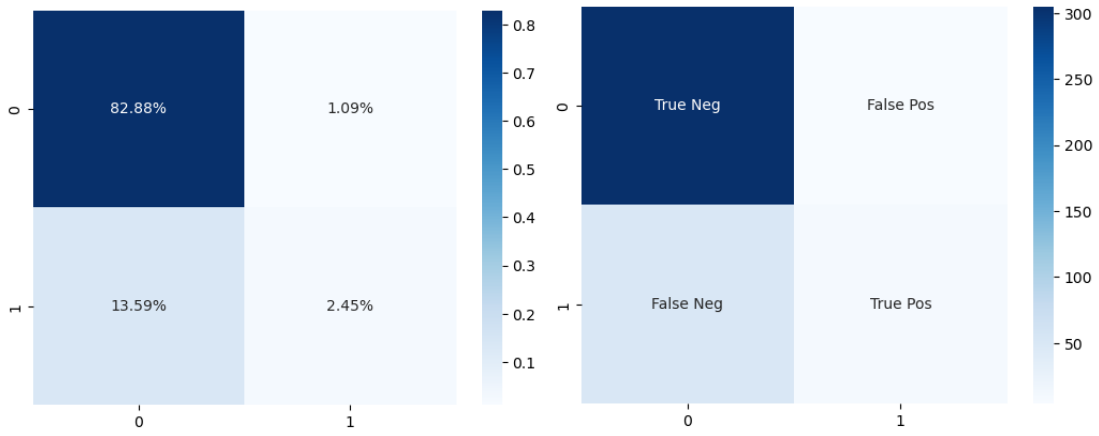
```
In [41]: from sklearn.metrics import precision_score, accuracy_score

         accuracy_score(y_test, y_pred), precision_score(y_test, y_pred)
```

```
Out[41]: (0.8532608695652174, 0.6923076923076923)
```

Conclusion

I have checked the confusion matrix to know the accuracy of the model, what percentage is true positive and what percentage is true negative. The below diagram present the confusion matrix from the model fit.



This prove the quality model and as the problem is related to the human behavior, it can be applied to any employee depending on the organization. All the organization have similar properties depending on the employees like distance from home, hourly rate as these values are different by organization, the model will work accordingly and can be applied to any organization.