

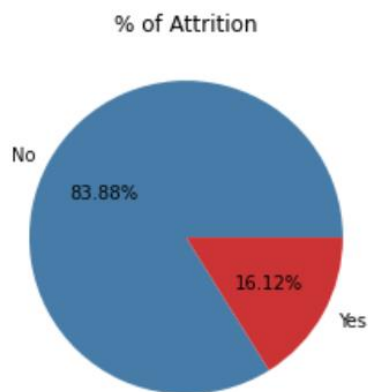
Predicting Employee Attrition

DSC680 Applied Data Science

Question and answers

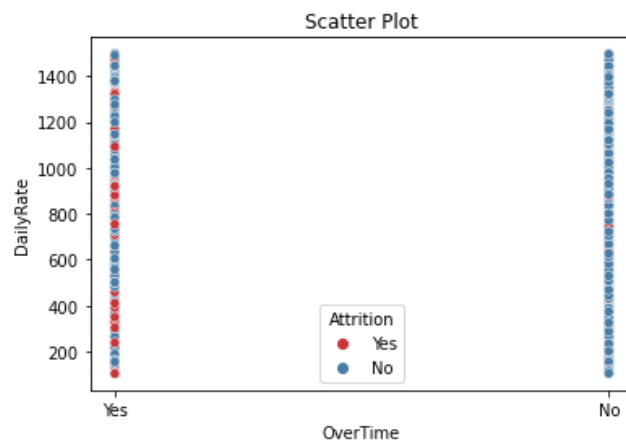
1. What is the percentage of attrition in the provided dataset?

a. There are 16.12% of attrition outlined in the dataset.



b.
2. Does overtime impact the attrition?

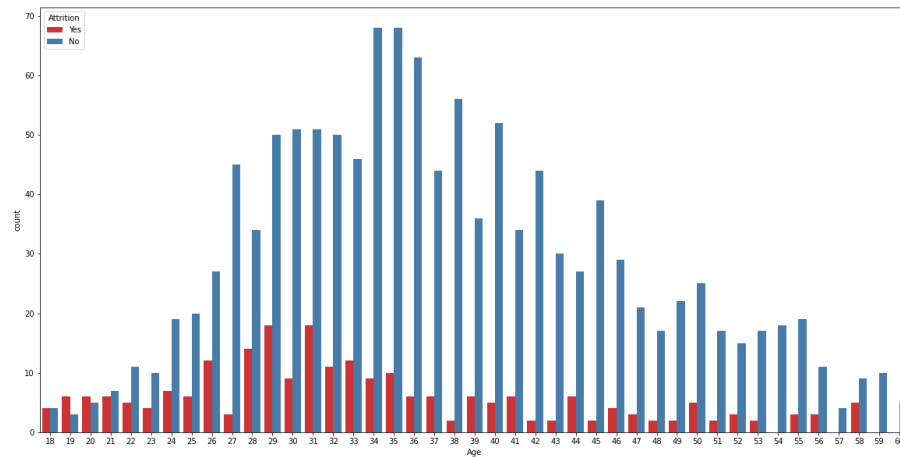
a. Irrespective of daily rates, the overtime plays an important factor in attrition.



b.

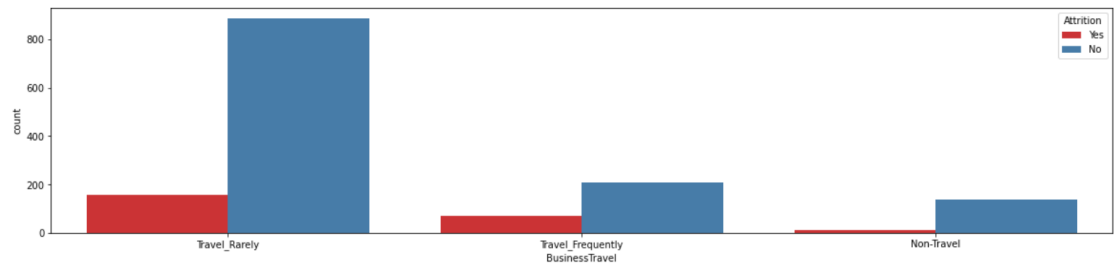
3. Does age impacts attrition?

- a. People with age of 26 to 35 tend to leave the organization.



4. Does Business travel affect attrition?

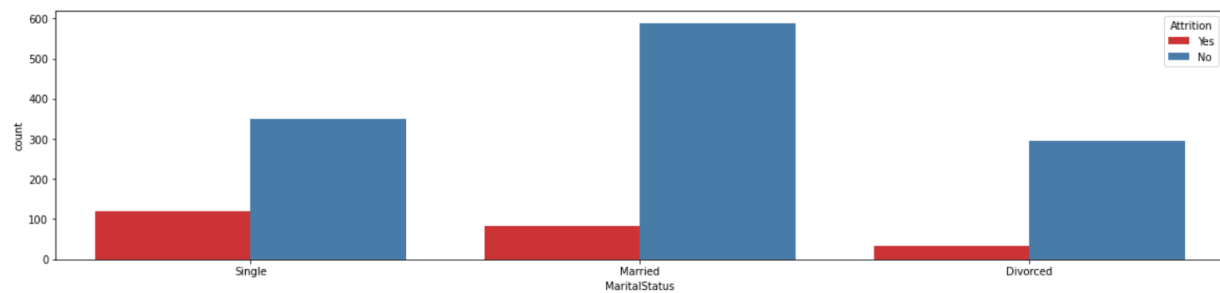
- a. Yes, When there are some business travel. People tend to leave the organization.



b.

5. Which Martial status are more likely to leave the organization.

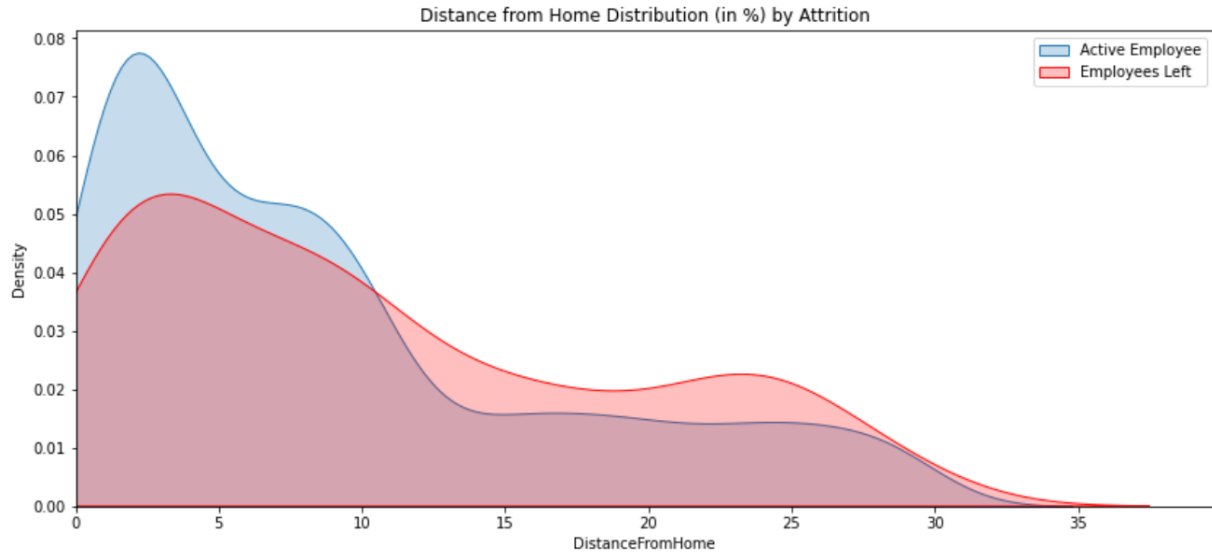
- a. The Martial status single are more likely to leave.



b.

6. Does distance impact in attrition?

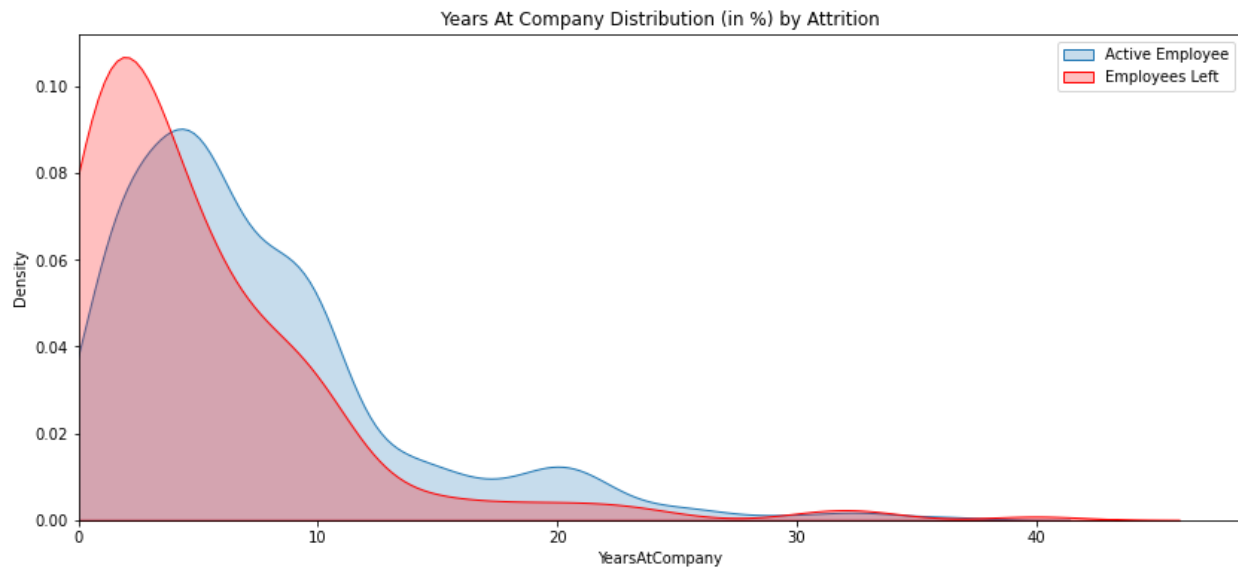
- a. Employees traveling more than 10 miles are more likely to leave.



b.

7. If employee stays in company for less than 2 years. What is the probability of he/she leaving.

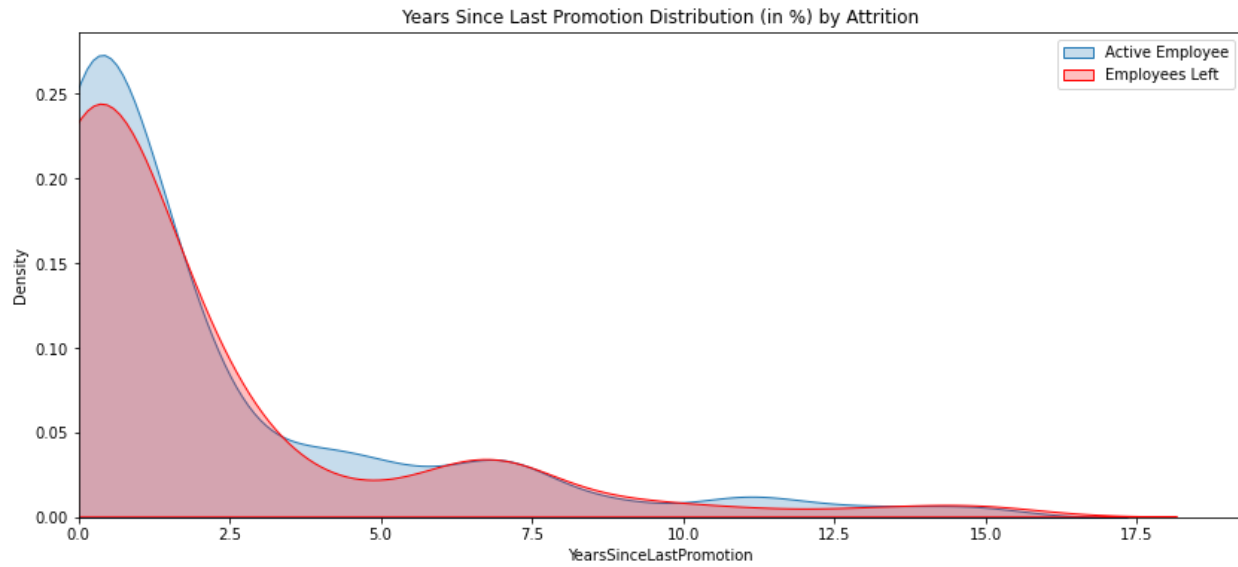
a. Probability of he/she leaving is high. As the graph below show Employees are more likely to leave in the initial couple of years at the company.



b.

8. Does Promotion shows any impact on attrition?

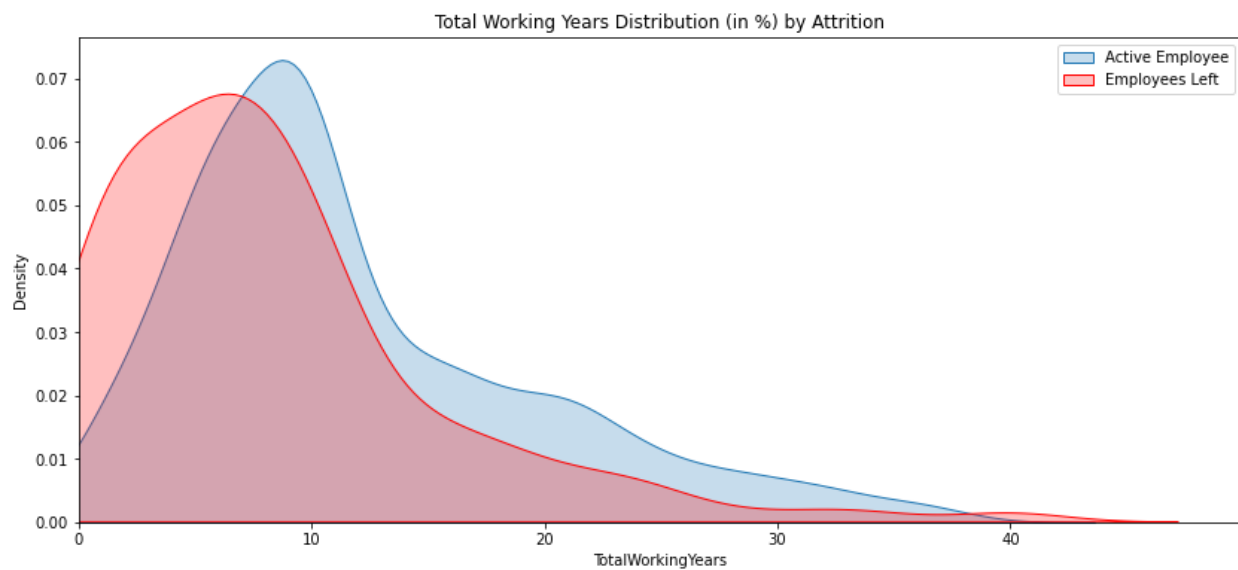
a. Promotion does not show a large effect on Attrition. The distribution is very similar



b.

9. Does total working years impact attrition?

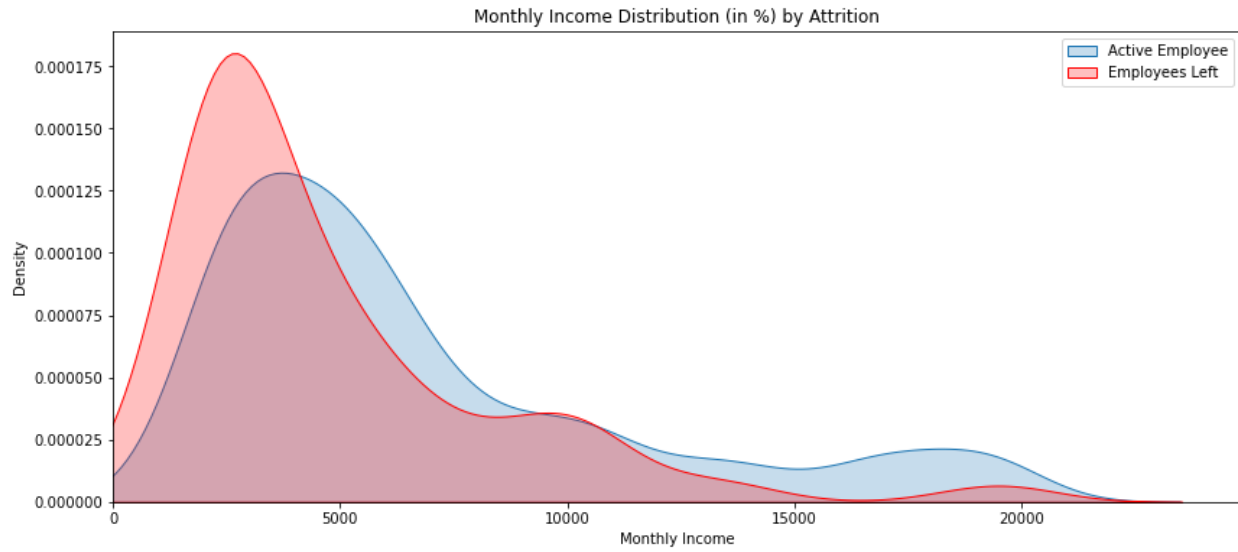
a. If the employee stays in organization for more than 7 years. He is more likely to stay.



b.

10. Does monthly income impact attrition?

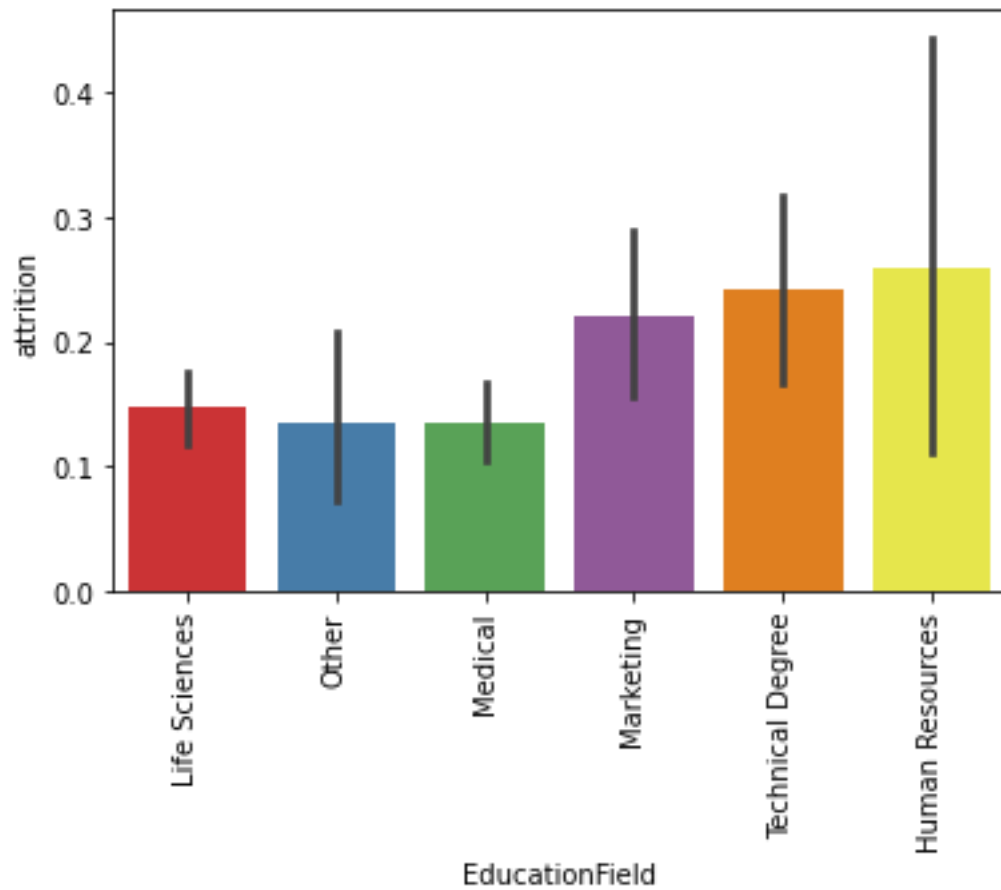
a. When Monthly is less the person will leave.



b.

11. Which education field has more attrition?

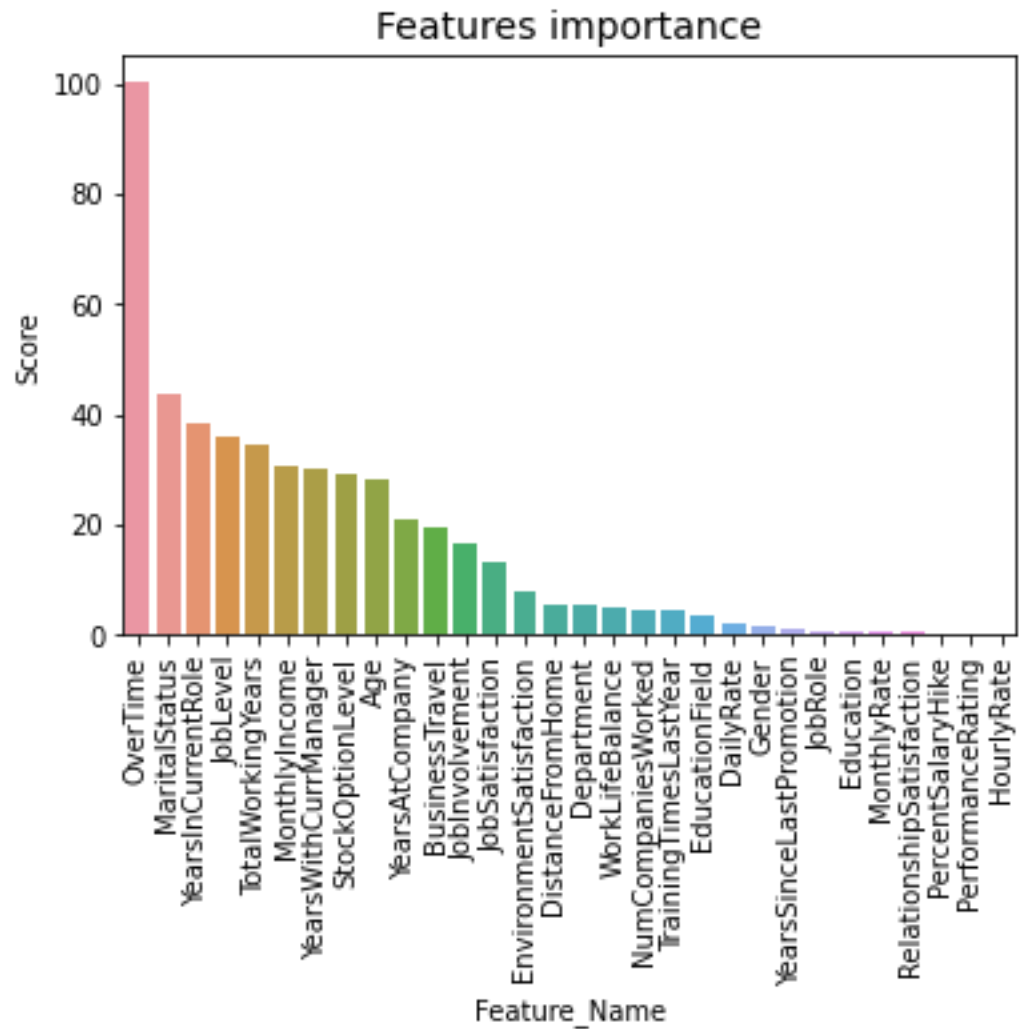
a. Employees with education field Human Resources are more likely to change job.



b.

12. Which is the important feature for attrition?

a. Overtime is an important feature in the dataset.



b.

13. Which is the best algorithm for the analysis?

	roc_auc
RandomForestClassifier	0.795754
DecisionTreeClassifier	0.622729
SGDClassifier	0.768108
LogisticRegression	0.811739

a.

b. The best algorithms are Random Forest Classifier and Logistic Regression for this data set.