



HR Case Study

Group Names:

1. Sayan Chowdhury
2. Rohit Rajagopal



Problem Statement

- XYZ company has an employee strength of 4000.
- Attrition rate is 15% annually.
- This results in difficulty in delivering projects within given timelines.
- Also new employees takes time to get onboarded and get going.
- Providing training to them also consumes time and money.



Objective of Analysis

1. Understand the reason for attrition.
2. Check for various factors from company data to reduce attrition rate.
3. Identify the factors which are causing this and find ways to rectify them.

Source of Data- Company Database



Problem Solving Methodology

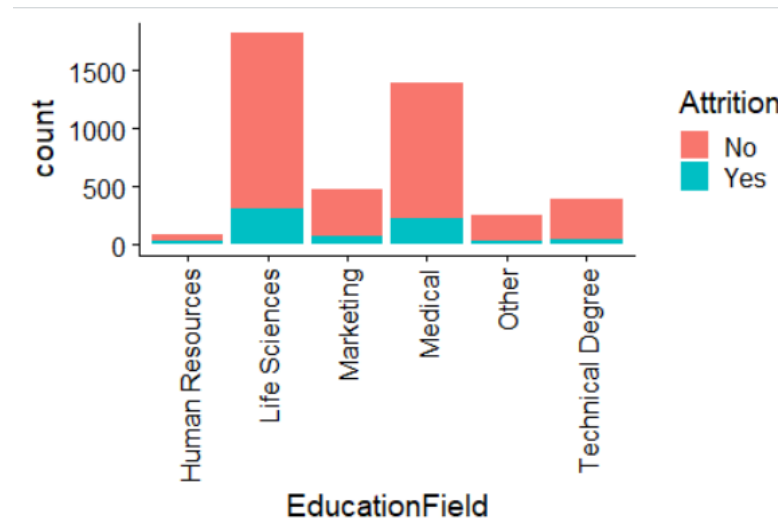
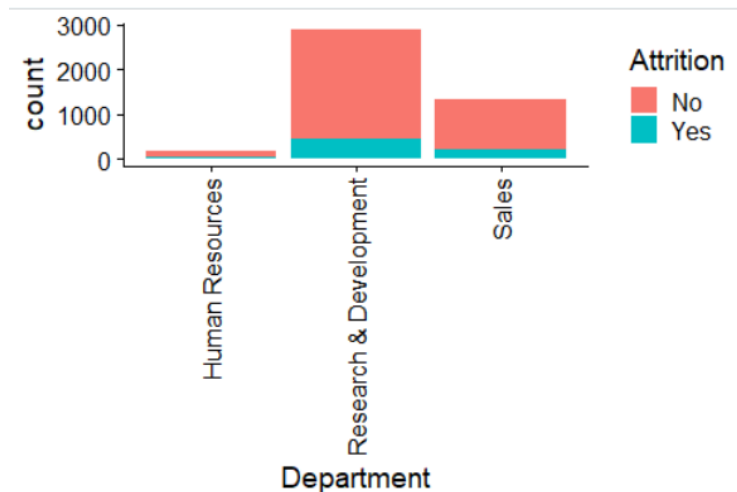
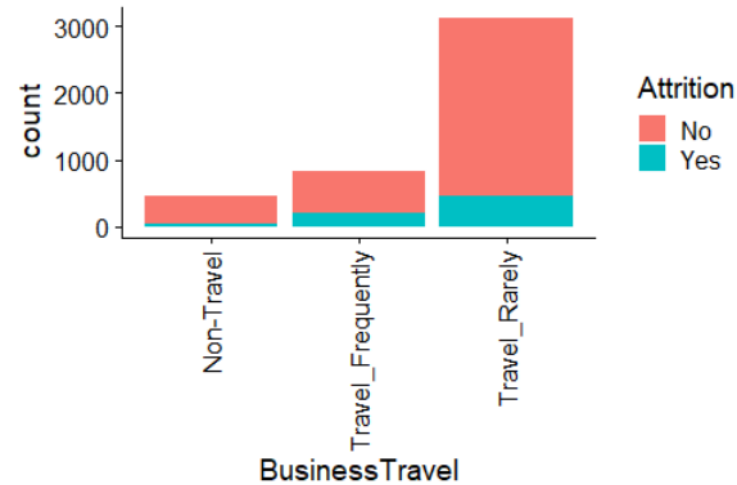
- Sourcing Data from various sources in Company Database
- Cleaning data and merging them in a single master dataframe to do analysis.
- Exploratory data analysis to find patterns and draw insights.
- Using logistic regression to create models and identify parameters which influences employee morale and propose ways to rectify them.
- Model evaluation to check model performance.



Data Cleaning



Analysis



Analysis of various factors affecting attrition shows us some insightful trends.

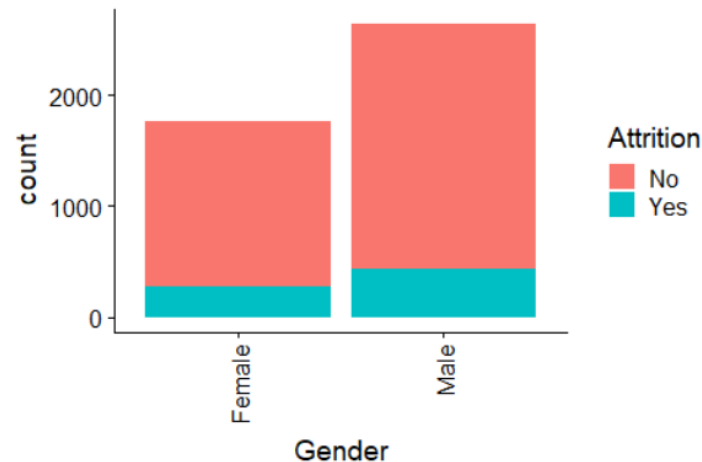
Proportionately people who travel frequently are more prone to churn than people who rarely or don't travel.

There is high attrition in people from Education field and also from people in R&D department.

Analysis



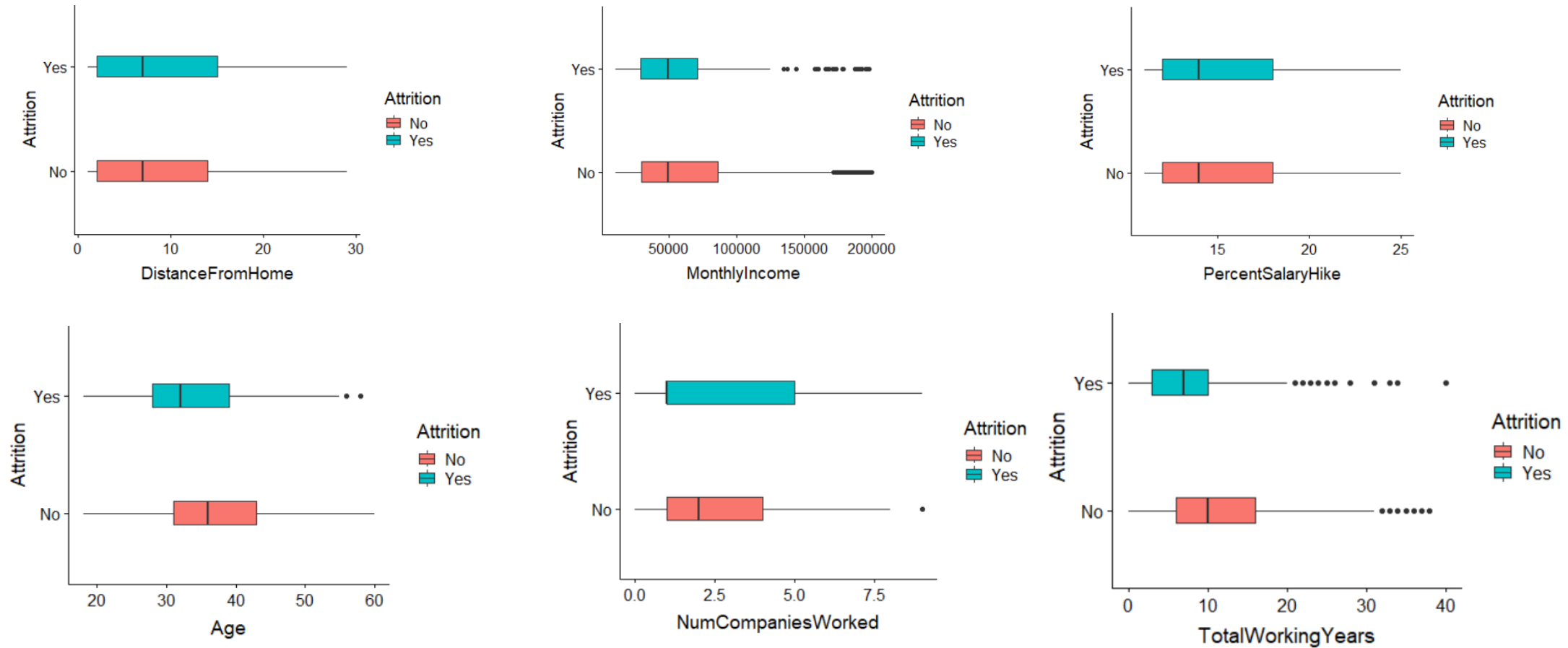
The following graphs shows the relative attrition across various job roles and genders.



We will be implementing logistic regression to identify and classify factors on which the attrition rate is influenced the most.

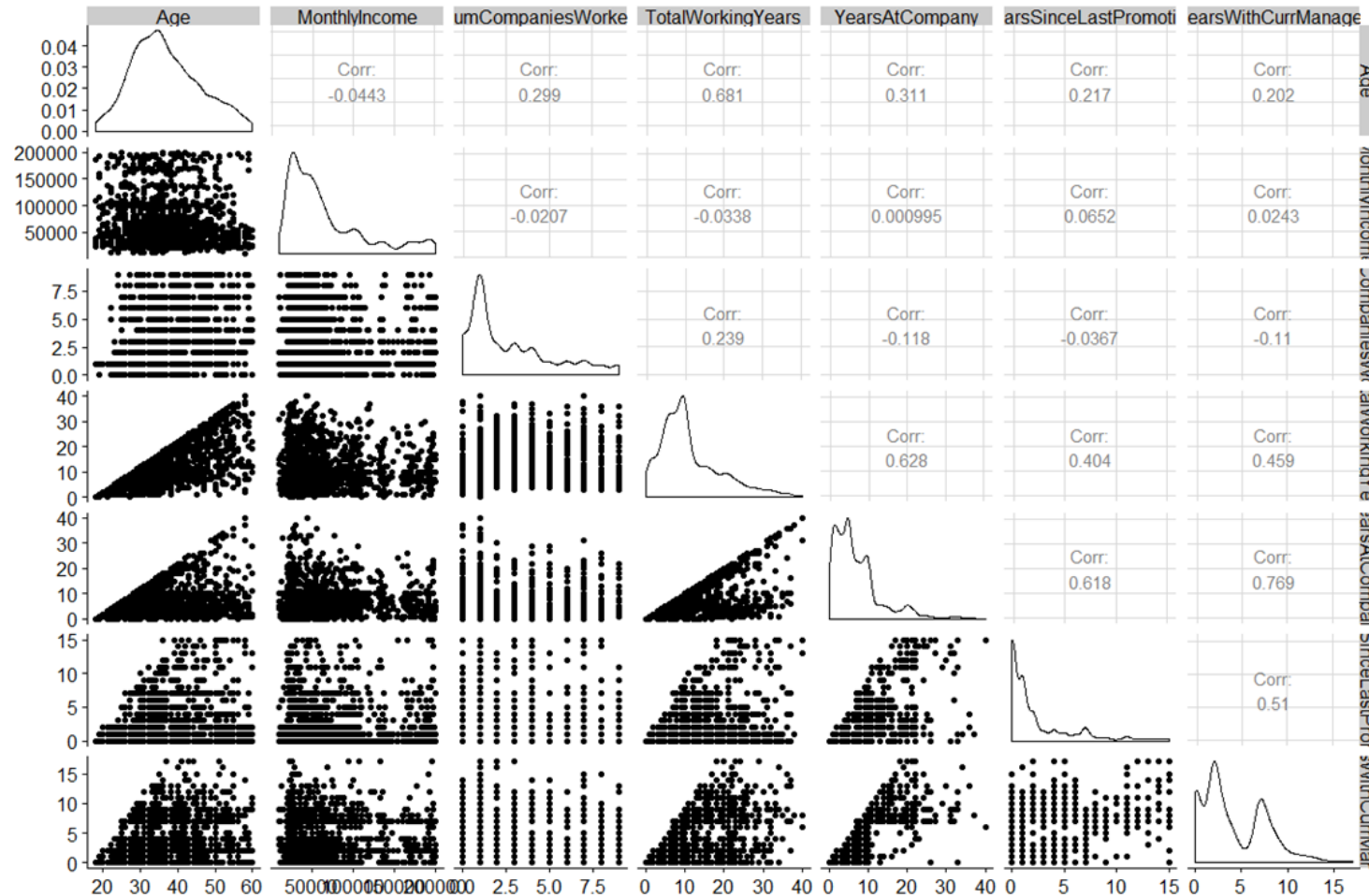
Also model evaluation to check sensitivity and specificity and accuracy is done to find the best cutoff probability for the model.

Analysis



Few important parameters influencing attrition

Analysis



This graph shows the correlation matrix between various factors.

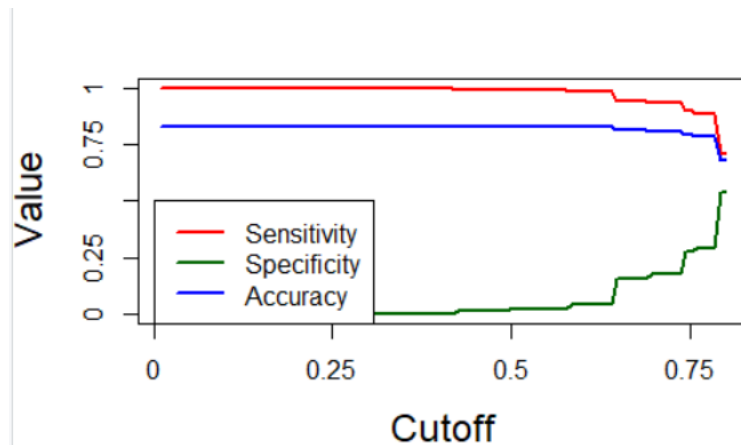
We can see that YearsAtCompany and YearsWithCurrentManager are correlated (corr 0.76).

Also YearsAtCompany and YearsSinceLastPromotion are somewhat correlated (corr 0.618).

Conclusion

	test pred at	
test actual at	No	Yes
No	4	218
Yes	7	1086

Confusion matrix



Accuracy
0.7855513
Sensitivity
0.8856359
Specificity
0.2927928

Chose the cutoff value 0.78 at the intersection of Accuracy Sensitivity and Specificity. We predicted the categorical variables

BusinessTravelNon.Travel,
BusinessTravelTravel_FrequentlyE
ducation2 ie college,
EducationFieldHuman.Resources,
MaritalStatusDivorced,
MaritalStatusMarried

to cause attrition for 78% employee
with of *Accuracy 78.55%, Sensitivity*
88.56% and Specificity 29.27%