Analysis HR Employee Attrition (Group 6) (Link)

Ayushi Mishra
mishra37@wisc.edu

Neil Bhutada
nbhutada@wisc.edu

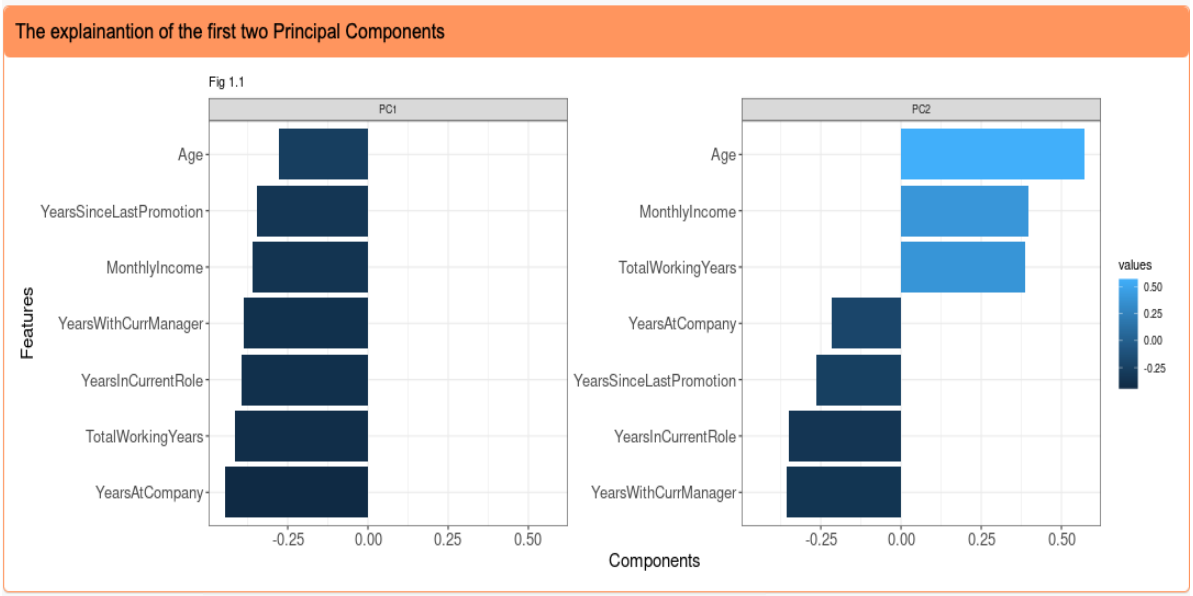
Harshita Narayanan
hbnarayanan@wisc.edu

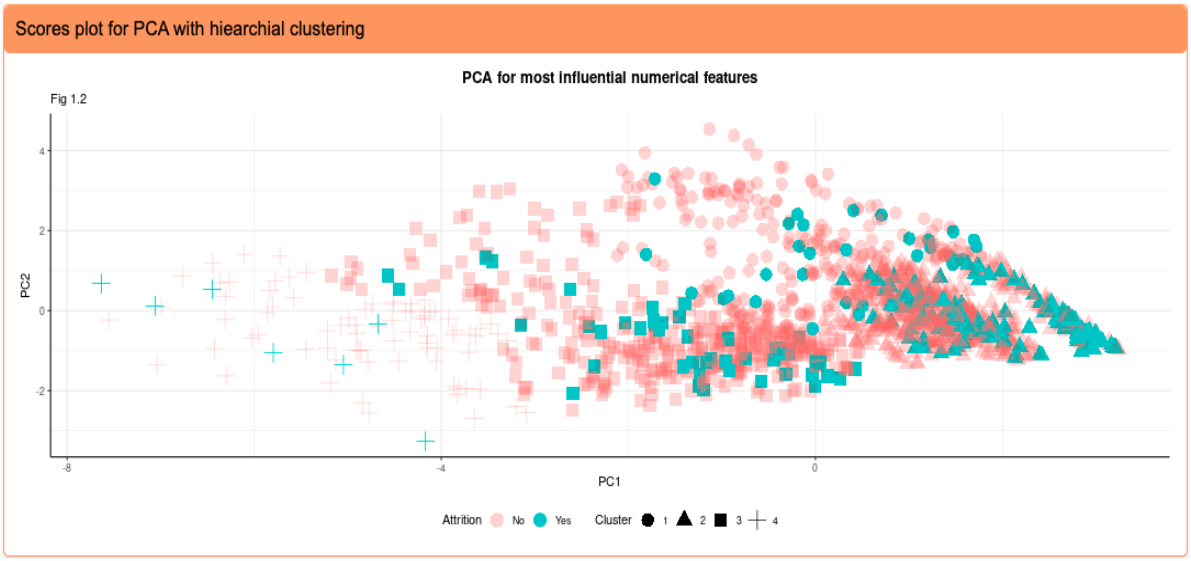
Purpose:

One big problem that companies often face is *Employee Attrition* - the resignation of an employee due to specific reasons. Employee Attrition heavily impacts the capital cost of the company, therefore knowing some of the key reasons of why it is occurring can help companies prevent employee attrition in the future. The main goal of this analysis is to identify any specific interventions that could reduce employee attrition. For this project, we used the mock data set that IBM has publicly available about attrition.

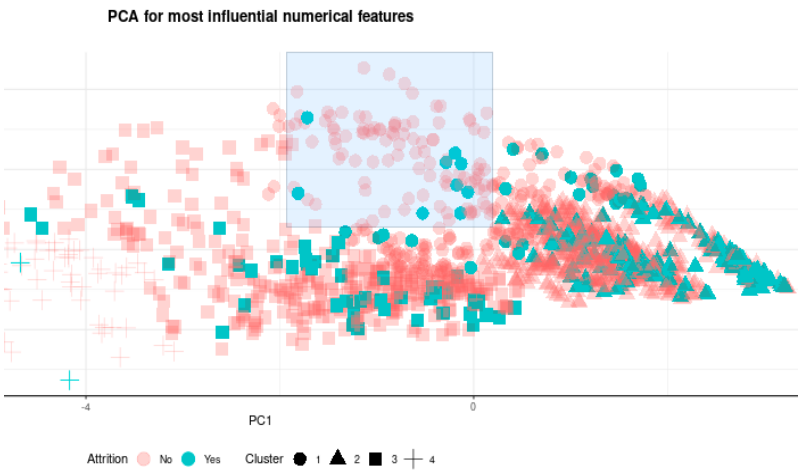
Discussion:

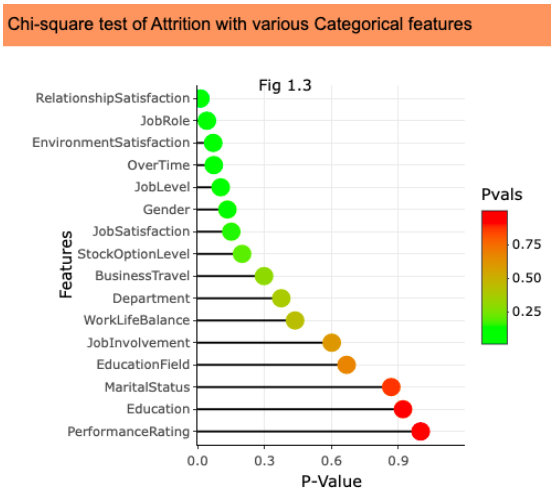
We visualized an interactive PCA plot with hierarchical clustering, incorporated with the brushing feature. PCA allows us to analyze multiple numerical features, while hierarchical clustering helped us understand similar employees. We performed PCA on numerical features *Age*, *Years Since Last Promotion*, *Monthly Income*, *Years With Current Manager*, *Years In Current Role*, *Total Working Years*, and *Years At Company*. The user can analyze categorical variables and their relation to attrition by selecting points in the scores plot Fig 1.2 with *chi-square tests* and *interactive bar plots*. Observing the *component explanation plot* in Fig. 1.1, we can see that higher value of PC1 would mean lower values of all the numeric variables. On the other hand, higher value of PC2 would mean higher values of *Age, Monthly Income,* and *Total Working Years*. We can conclude through the components plot that *Age* and *Years At Company* are the most influential features in PC1 and PC2.



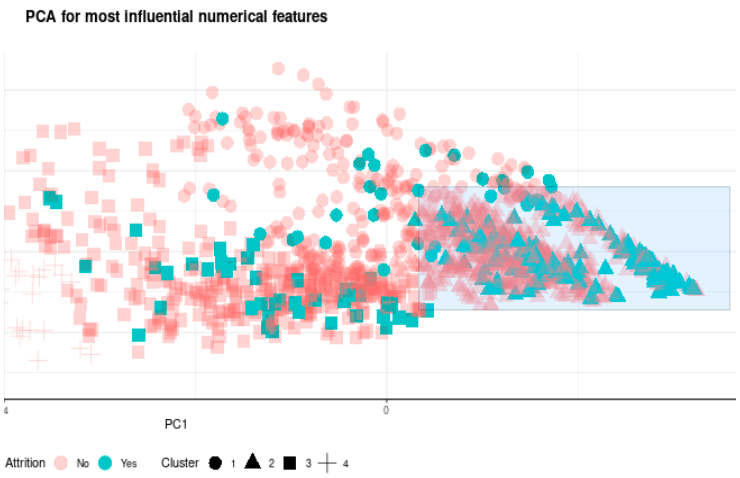


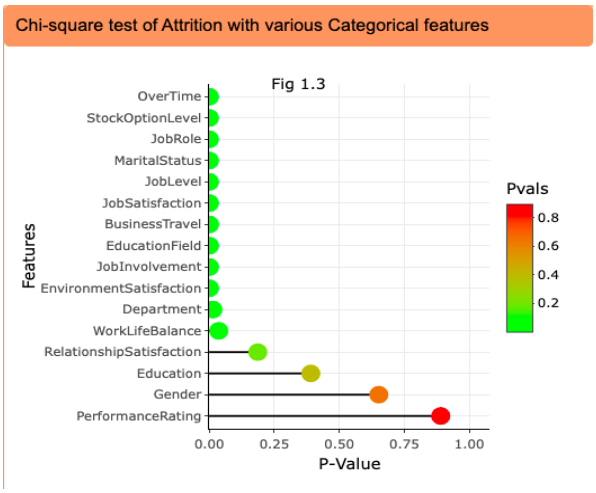
From the brushed points in Cluster 1, we can observe that *Relationship Satisfaction* and *Job Role* seem to be the most influential towards Attrition. For this cluster, since the PC2 values are highest and PC1 are mid-level, it shows that the selected points generally have higher age, income, and experience but mid-level working years at the current company. A solution to reduce attrition in Cluster 1 can be to have employees with higher job levels and include surveys to improve employee relations.



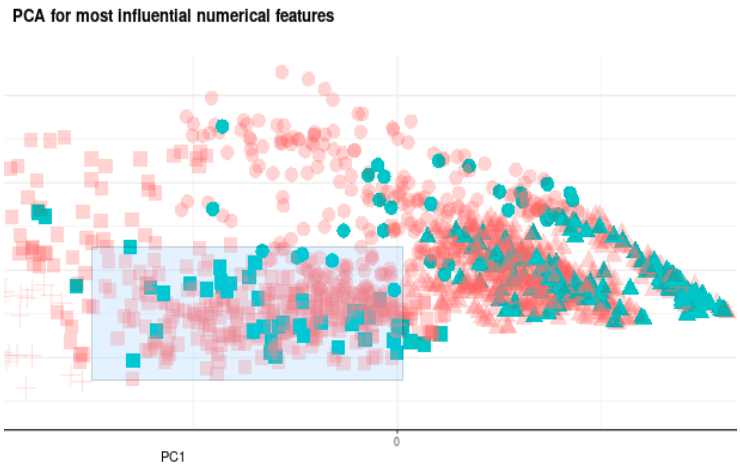


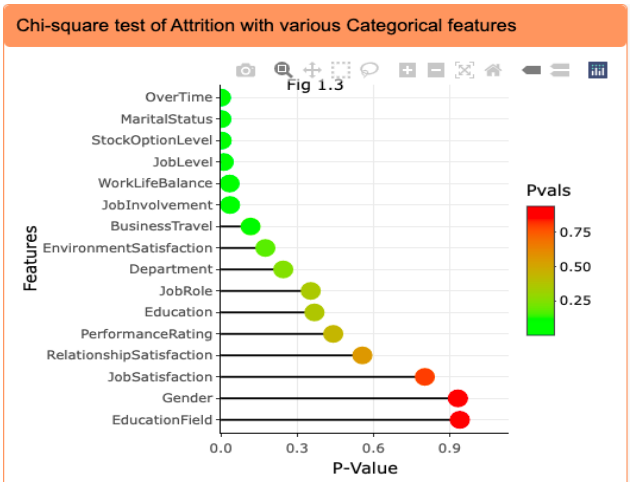
Similarly, the points in cluster 2, being young and less experienced leads to higher rates of attrition. From the points in cluster 2, features such as *Overtime* and *Stock Options* are the most significant from the chi-square test. To reduce attrition in cluster 2, companies can restrict employees from working overtime. Also, increasing stock options could reduce attrition.



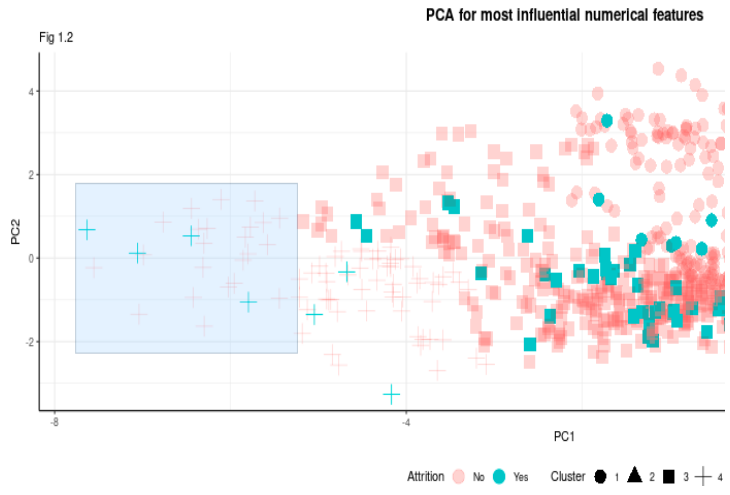


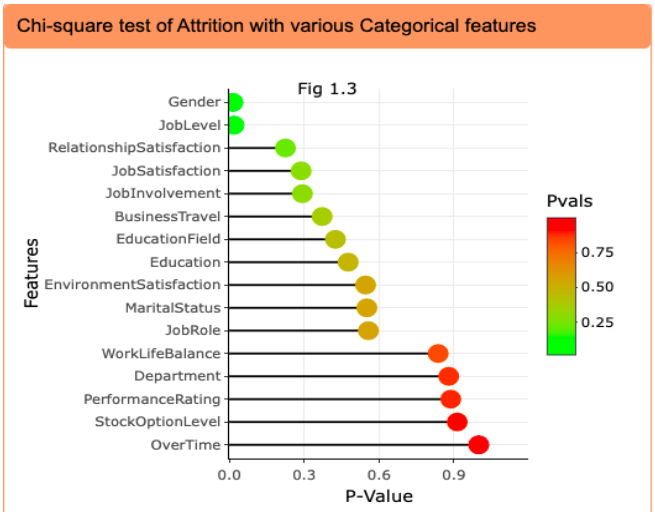
From the points in Cluster 3, the principal components have lower ages and monthly income but medium amounts of experience. We can also observe that *Over Time* and Stock options are highly influential towards Attrition. Similar incentives to cluster 2 can be implemented.





For the points in cluster 4, the components show old ages and high experience leading to less attrition. *Gender* seems to be one of the most influential features towards Attrition. Females have a higher chance of attrition compared to men. The company can improve gender equality to reduce attrition in Cluster 4.





We used a decision tree to further understand different routes towards attrition. The decision tree model has high accuracy and precision, but lower recall rate; this implies we can trust the paths from the decision tree to predict attrition. However, the decision tree was not able to reveal all the causes to attrition. *Overtime* is the starting point to predict attrition. The decision tree states that employees working overtime and are paid less than $761 are likely to resign. Surprisingly, employees who are single and have not been promoted in 3 years will are likely to resign. Hence, HR could keep a policy to promote employees after every 3 years at least until a certain seniority.

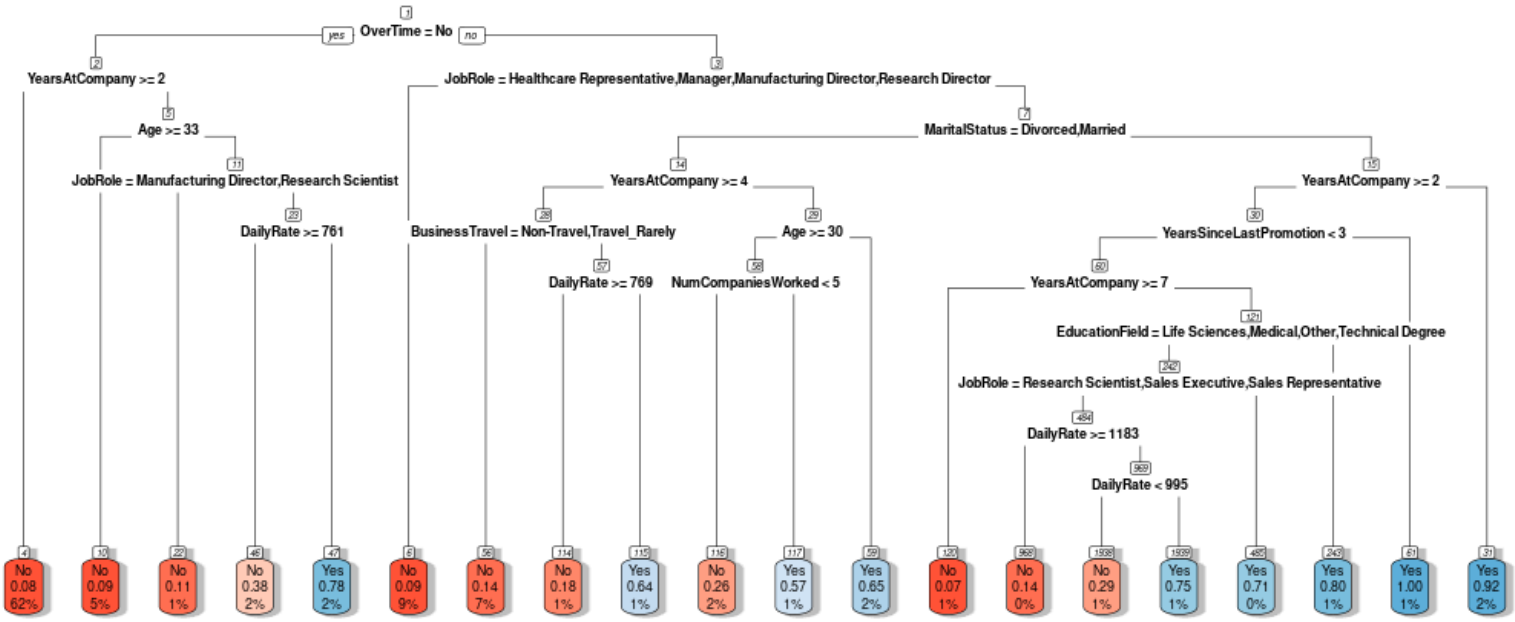


Fig 2.1

The profiles for *Age* and *Years At Company* had a slope. According to the CBs the chances of attrition reduced when the age was beyond 30 and the *Years At Company* were above. This implies the company should focus on applying schemes for reducing attrition for employees lesser than 30 years-old and have less than 10 years at the company. *Age* and *Years At Company*, employees with a Marketing degree, and Employees that are Laboratory Technicians and Sales Representative are most likely to resign.

To understand combinations of categorical and numerical variables that strongly interact with attrition, we have faceted bar plots and ridge density plots controlled by a slider with p-values ranging from 0 to 0.1. The p-values were generated from by performing Likelihood-Ratio Tests (LRT) on combinations of numerical-categorical and categorical-categorical variables. For both plots, there are many interactions that are useful to understand attrition rates because of their significant p-values. From categorical-numerical interactions, the top three most significant interactions are Job Role-Monthly Income, Job Role-Years at the Company, Job Level-Years at Company. From categorical-numerical interactions, the top three interactions were between Education-Job Role, Job Role-Relationship Satisfaction, Marital Status-Performance Rating.

Categorical-Numerical Interactions LRT

Show 10 entries Search:

	Categorical	Numerical	PValue
1	JobRole	MonthlyIncome	0.00005
2	JobRole	YearsAtCompany	0.000061
3	JobLevel	YearsAtCompany	0.000182
4	JobLevel	MonthlyIncome	0.000204
5	Education	YearsAtCompany	0.001749
6	MaritalStatus	YearsAtCompany	0.002482
7	JobRole	YearsSinceLastPromotion	0.003071
8	JobRole	YearsInCurrentRole	0.003479
9	Gender	NumCompaniesWorked	0.004365
10	MaritalStatus	YearsWithCurrManager	0.00478

Showing 1 to 10 of 29 entries Previous 1 2 3 Next

Categorical-Categorical Interactions LRT

Show 10 entries Search:

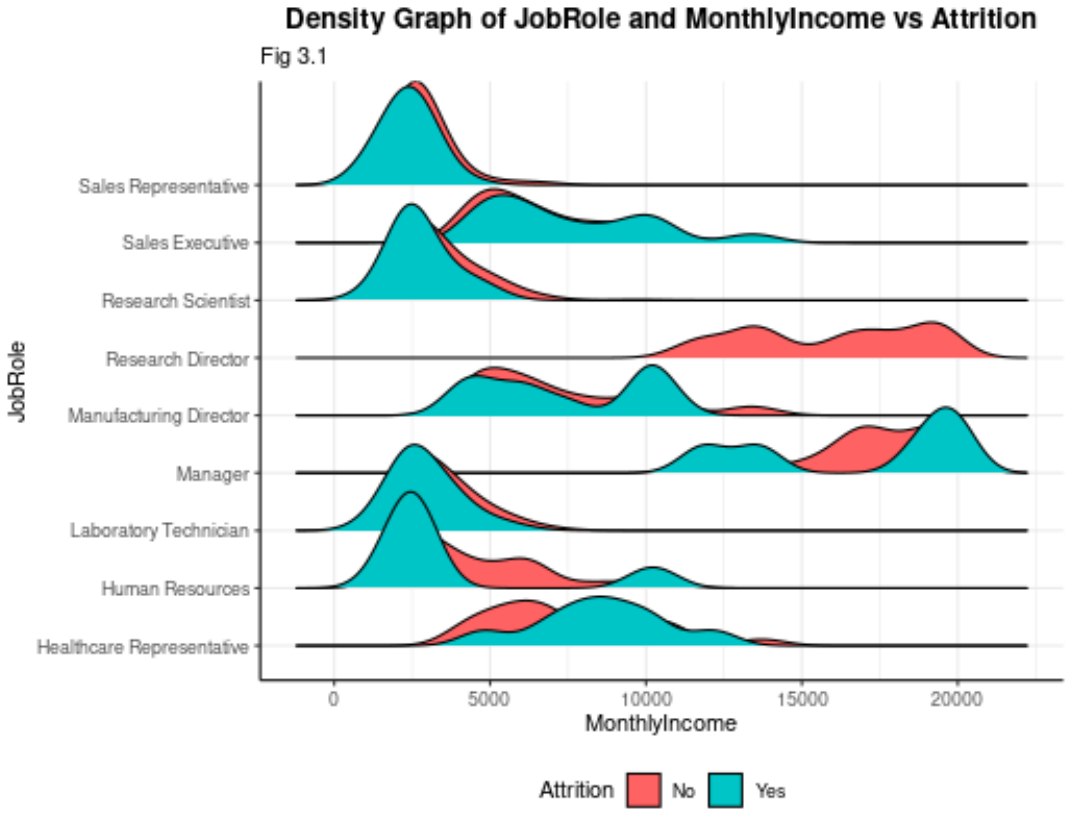
	Categorical 1	Categorical 2	PValue
1	EducationField	JobRole	0
2	JobRole	RelationshipSatisfaction	0.003987
3	MaritalStatus	PerformanceRating	0.017059
4	Education	JobInvolvement	0.019109
5	JobInvolvement	StockOptionLevel	0.029501
6	BusinessTravel	JobInvolvement	0.04296
7	JobRole	WorkLifeBalance	0.043765
8	Department	MaritalStatus	0.044506
9	EnvironmentSatisfaction	JobInvolvement	0.04579

Showing 1 to 9 of 9 entries Previous 1 Next

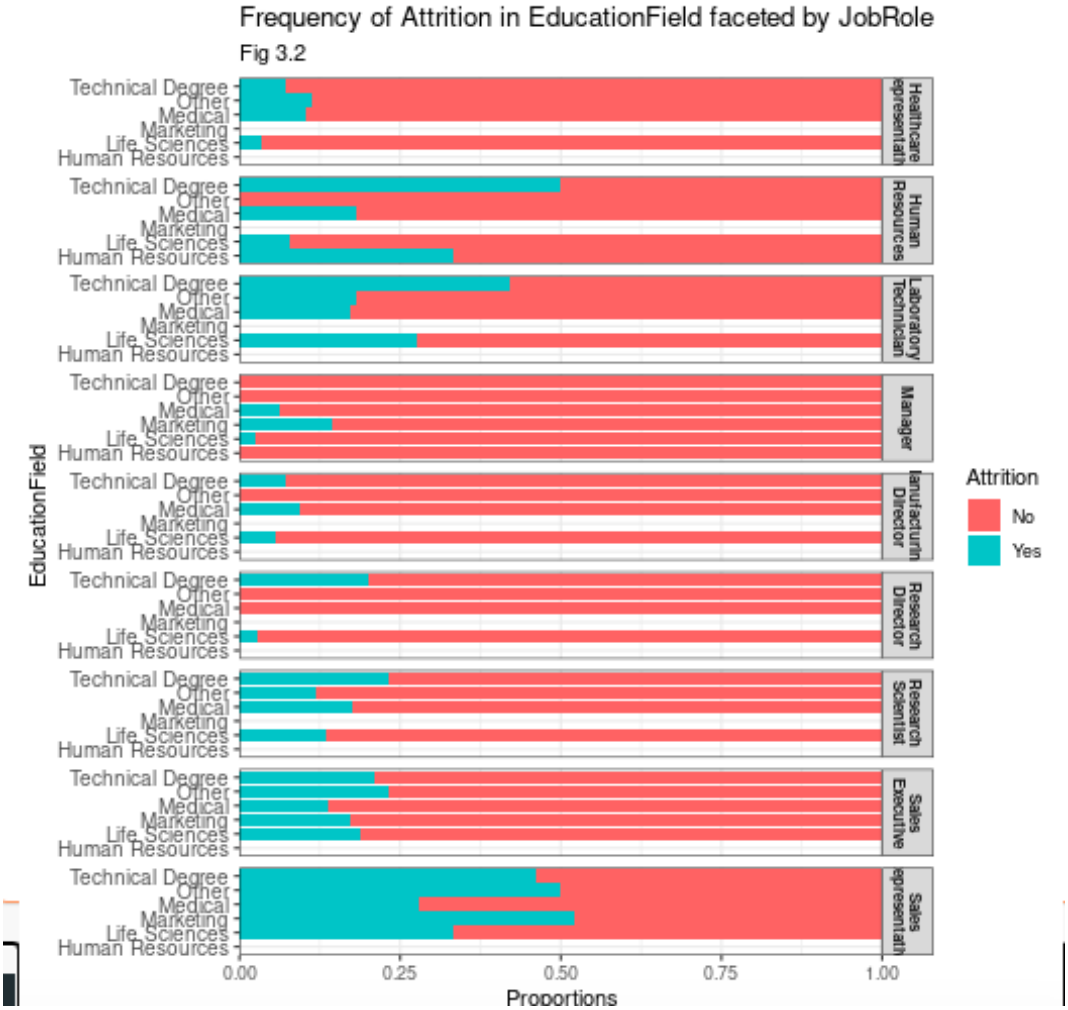
Analyzing the most significant interactions, i.e., *Job Role-Education Field* and *Job Role-Monthly Income*:

- Sales representatives have high attrition rates for all educational backgrounds, but sale representatives with marketing degrees seem to leave the most. However, sales representatives with monthly incomes between $3000 to $5000 have lesser attrition rates. A solution to avoid attrition of Sale representatives could be to increase pay or stock options, especially to marketing degree sale representatives.
- Employees that work in HR and have technical degrees have higher chances of attrition. The company should cautiously hire technical degree employees in HR. Higher attrition can be seen for HR employees with monthly income lesser than $5,000 and greater than $10,000. HR employees with monthly incomes between $5,000 to $10,000 have lower attrition.

Ridge-Density plots for Categorical-Numerical Interaction with Attrition



Bar plots for Categorical-Categorical Interaction with Attrition



Common Analysis and Conclusion:

Age and work-related experience seem to be common features in all models that lead to attrition. Common influential factors amongst models were:

- Overtime was common in PCA and Decision Tree
- Marketing Degrees Sale representatives in Feature Interactions and Decision Tree

Few generalized solutions overall are:

- Providing Stock Options to lower age and job level employees can reduce attrition.
- Having strict measures and incentives in-place to avoid Over Time work
- Providing more incentives Sales representatives with marketing degrees