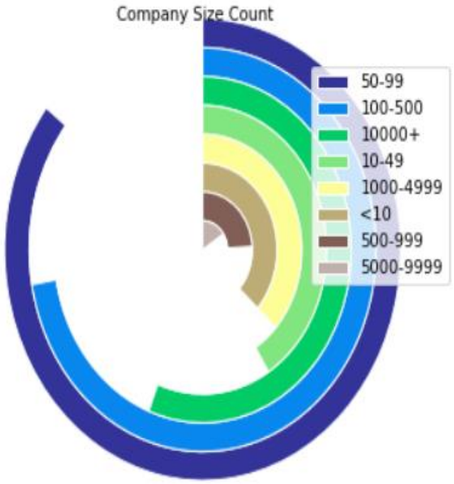
ANALYSIS OF JOB CHANGE OF DATA SCIENTIST

OBJECTIVES

- **To find out the employees who are leaving the company.**
- **To find out the attributes that influence the Employee Decision.**
- **Through this project, we are going to find out the reasons and factors contributing for a job change in employees and find out the ways to retain them.**

DATA ANALYSIS FINDINGS

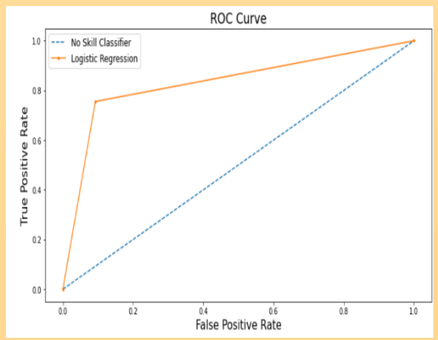
- Around 75% number of employees which are not looking for job change.
- Maximum males are working in the company and maximum of them are not looking for job change even.
- Maximum employees having relevant experience wants to work in Pvt Ltd Companies.
- More of the Graduates are working in this field and more of them are not looking for the job change as well.
- Maximum people with more than 20 years of experience are not looking for job change as maybe they are satisfied and comfortable with their current job.
- Maximum employees are not enrolled in any university and maximum are not looking for job change even.
- Most Employees have undergone about 50 hours of training and some of them have undergone Extensive training of more than 100 hours.



RELATIONSHIP OF VARIABLES

- There is a significant relationship between the Target variable and the Gender having p-value-0.010.
- The Company Type and the Target variable also has a significant relationship at p-value less than 0.05.
- There is some relation between the Education level and the Target at the significance level of 0.05.
- Other Categorical Variables also showed some relationship with the Target variable while doing Ch-square test.
- The Training Hours and the City Development Index column are negatively correlated with the Target variable.

BEST MODELS

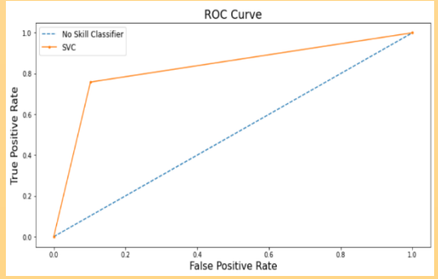


LOGISTIC REGRESSION Model Performance Metrics:

	precision	recall	f1-score	support
0.0	0.79	0.91	0.84	2562
1.0	0.89	0.76	0.82	2597
accuracy			0.83	5159
macro avg	0.84	0.83	0.83	5159
weighted avg	0.84	0.83	0.83	5159

CONFUSION MATRIX :
[[2325 237]
[635 1962]]

NO SKILL CLASSIFIER: ROC AUC=0.500
LOGISTIC REGRESSION: ROC AUC=0.831

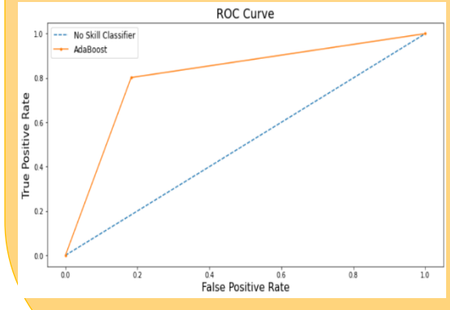


SVC Model Performance Metrics:

	precision	recall	f1-score	support
0	0.79	0.90	0.84	2902
1	0.88	0.76	0.81	2851
accuracy			0.83	5753
macro avg	0.84	0.83	0.83	5753
weighted avg	0.83	0.83	0.83	5753

CONFUSION MATRIX :
[[2605 297]
[689 2162]]

NO SKILL CLASSIFIER: ROC AUC=0.500
SVC: ROC AUC=0.828

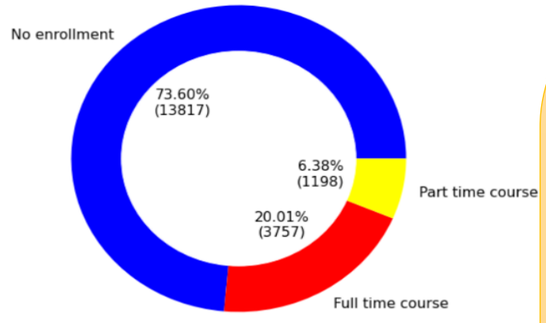


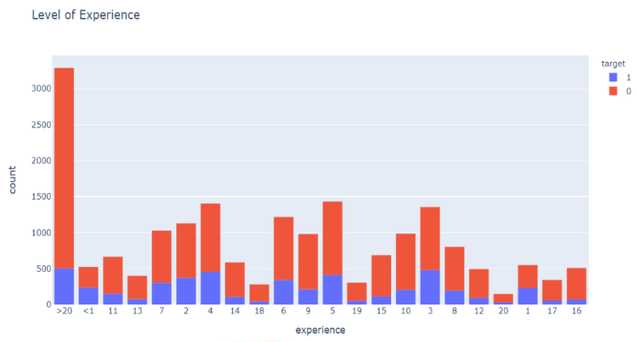
AdaBoost Model Performance Metrics:

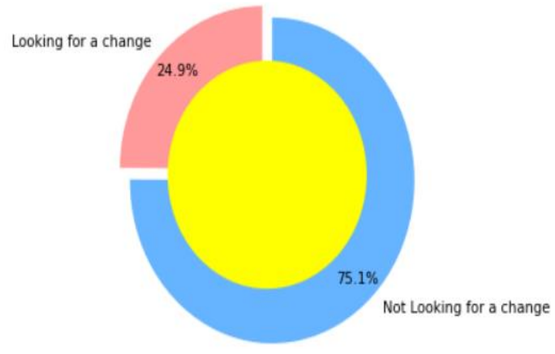
	precision	recall	f1-score	support
0.0	0.81	0.82	0.81	2902
1.0	0.81	0.80	0.81	2851
accuracy			0.81	5753
macro avg	0.81	0.81	0.81	5753
weighted avg	0.81	0.81	0.81	5753

CONFUSION MATRIX :
[[2372 530]
[565 2286]]

NO SKILL CLASSIFIER: ROC AUC=0.500
AdaBoost: ROC AUC=0.810







MODEL FINDINGS

- Out of all the models, AdaBoost Classifier, SVC and Logistic Regression showed the best scores.
- K-Fold Cross Validation scores are a bit higher in AdaBoost Classifier Model.
- It seems to be that SVC and Logistic Regression with 83% are the best performing models followed by AdaBoost.
- As per the Classification Report, maximum predictions were caught correctly in SVC and the AdaBoost Classifier.
- As per the test and the training scores, the Logistic Regression and the Support Vector Classifier came out to be the best fit models.

So, as per the other scores Logistic Regression, Support Vector Classifier and AdaBoost Classifier model are the best performing models.

RECOMMENDATIONS

- Company should focus on retaining more female employees.
- Company should manage the long and extensive training hours.
- Also, company should try to add more employees of other discipline having different education level.
- This Analysis further indicates there is a need to add more features to the model in order to predict the job change status.