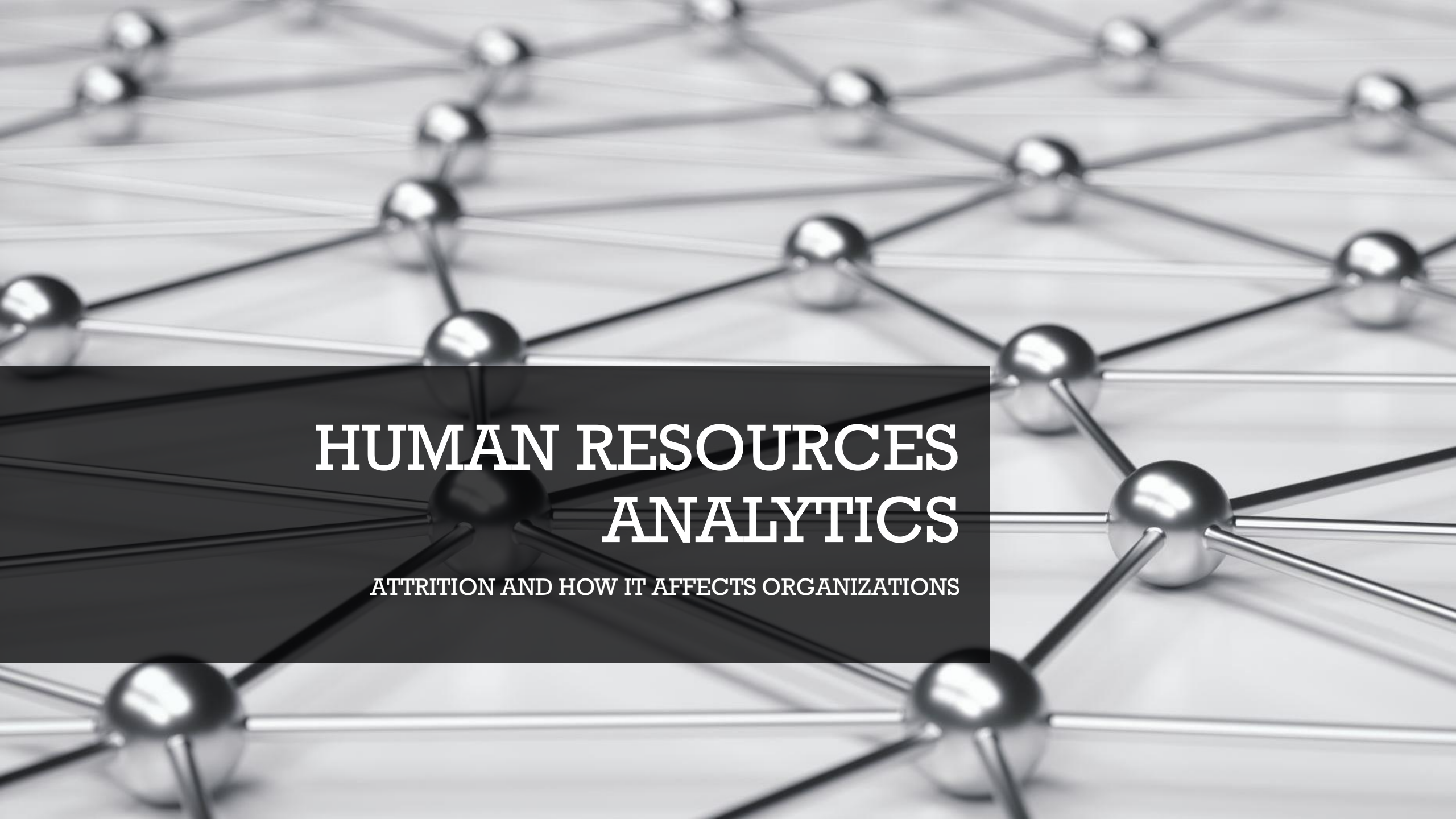


HUMAN RESOURCES ANALYTICS

ATTRITION AND HOW IT AFFECTS ORGANIZATIONS

PROBLEM STATEMENT & GOAL

Employee Attrition is the loss of an employee through a naturally occurring event/process.

The disadvantages of Employee Attrition:

Reduced size of workforce: Exiting an employee and not filling the vacancy obviously means a smaller workforce.

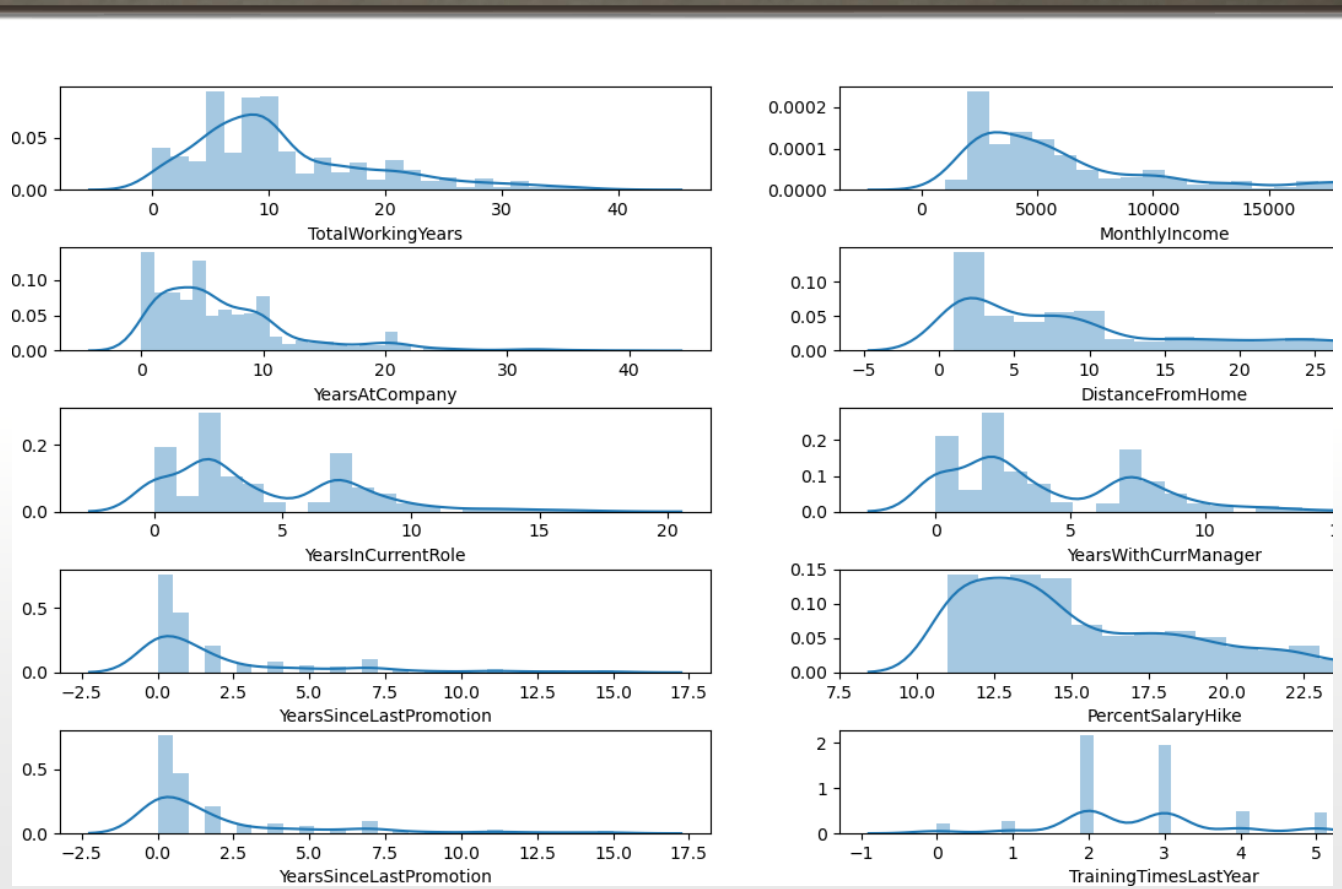
Potential to be costly: If attrition occurs for unplanned reasons, it can potentially prove to be costly

In this project, we will predict attrition rate based on various features in employee data set

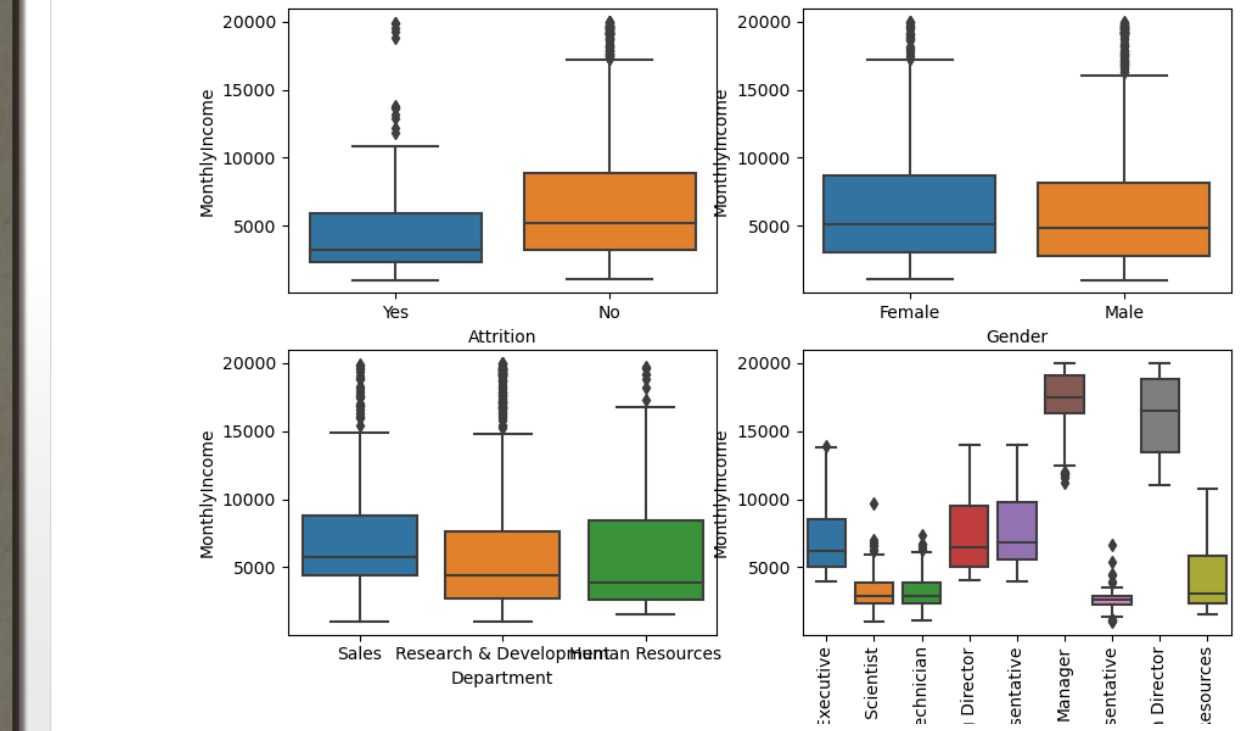
EXPLORATORY DATA ANALYSIS

- Exploratory Data Analysis(EDA) refers to performing critical investigations on the data set
- This can help determine correlations, spot data anaomalies, check our assumptions.
- This can be done in raw statistical methods as well as data visualizations

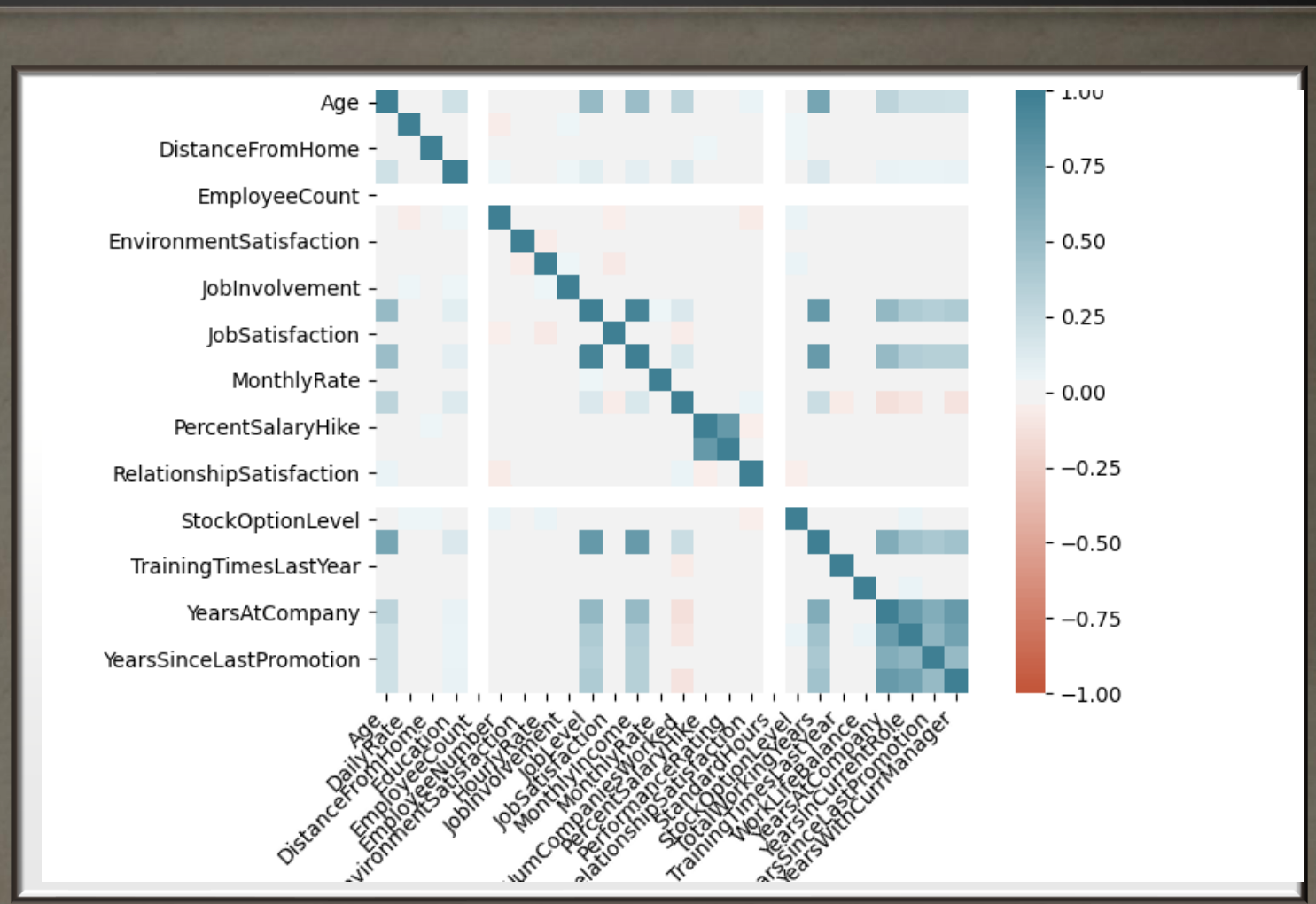
FEATURE VISUALIZATIONS



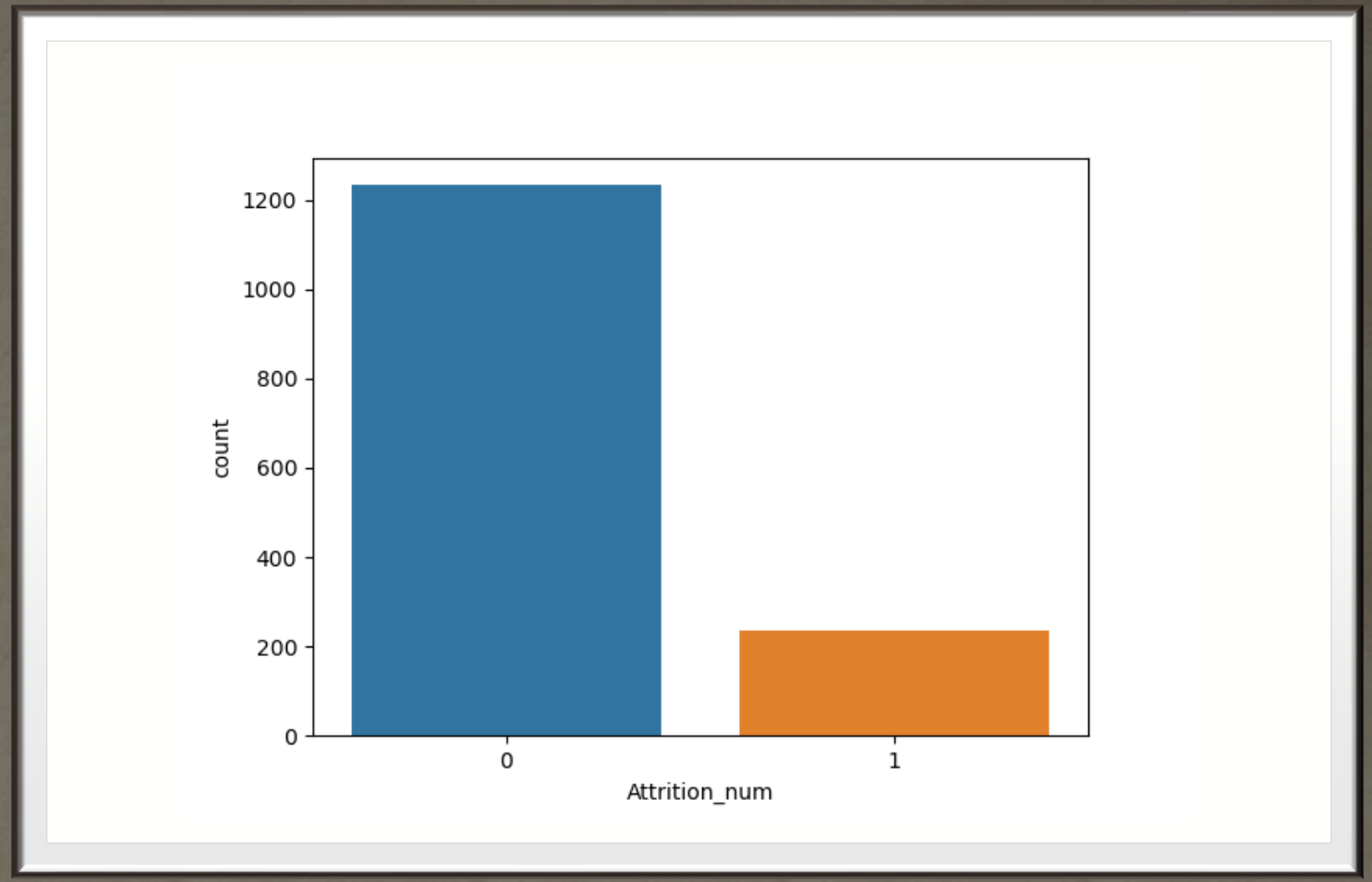
ATTRITION IN VARIOUS FEATURES



CORRELATION MATRIX



ATTRITION COUNT



Our class 0 is in high majority than our class 1. This is an imbalance dataset and sampling techniques to be used

SMOTE

- Synthetic Minority Oversampling TEchnique (SMOTE) is a very popular oversampling method that was proposed to improve random oversampling
- We will use this method in treating our imbalance data set

CROSS VALIDATION

- **Cross-validation** is a resampling procedure used to evaluate machine learning models on a limited data sample.
- In our model, we will use 5-fold cross validation method

ROC CURVES

- ROC Curves summarize the trade-off between the true positive rate and false positive rate for a predictive model using different probability thresholds.
- AUC function takes both the true outcomes (0,1) from the test set and the predicted probabilities for the 1 class. It returns the AUC score between 0.0 and 1.0 for no skill and perfect skill respectively.

MODELS IN PLAY

- Logistic Regression
- Random Forest Classifier
- Gradient Boosting

LOGISTIC REGRESSION

- Train our model using Logistic regression
- Logistic Regression AUC = 0.69
- Accuracy Score for Logistic Regression Model is 69%

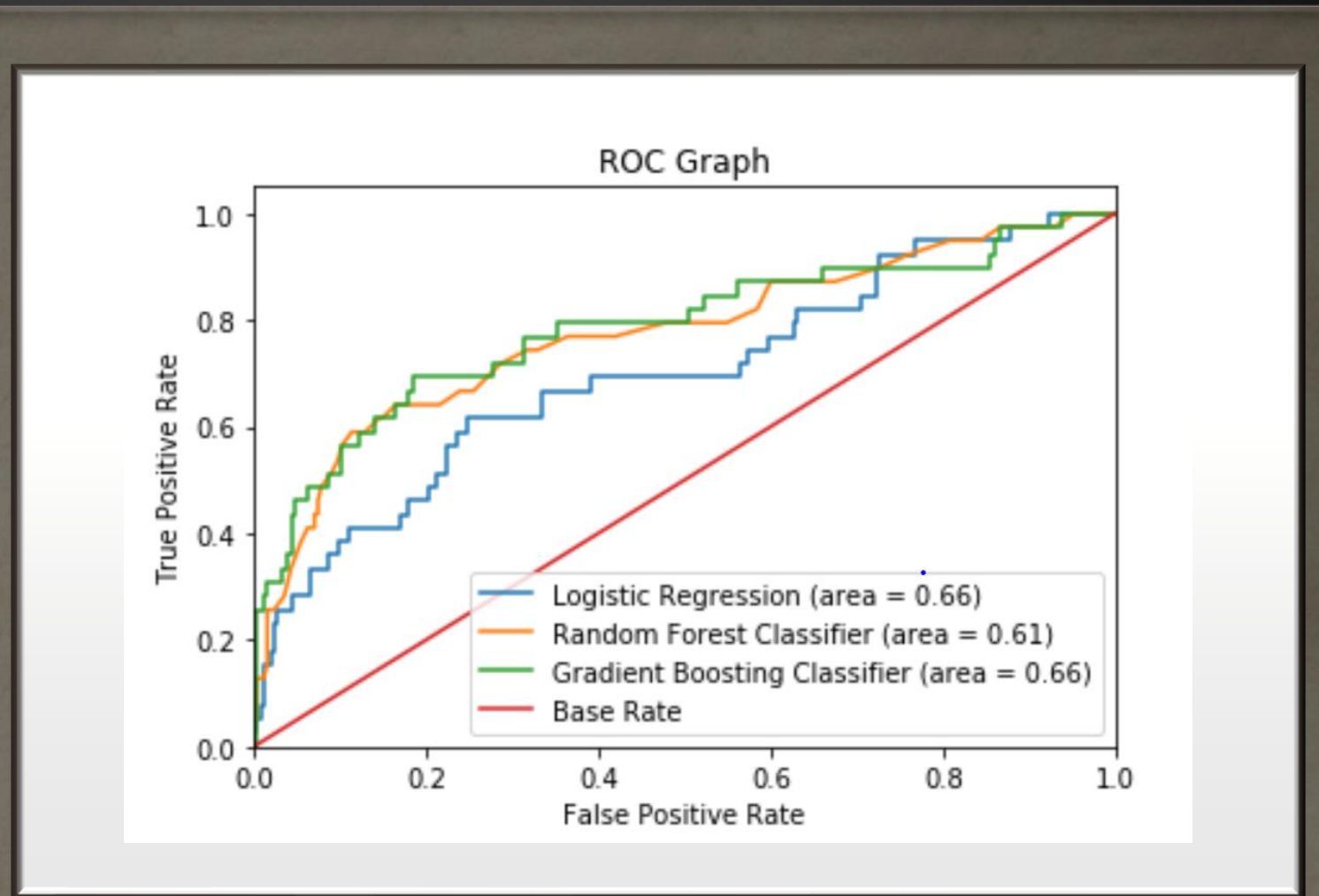
RANDOM FOREST CLASSIFIER

- Train our model using RF Classifier
- Random Forest AUC = 0.88
- Accuracy Score for Random Forest Classifier Model is 88%

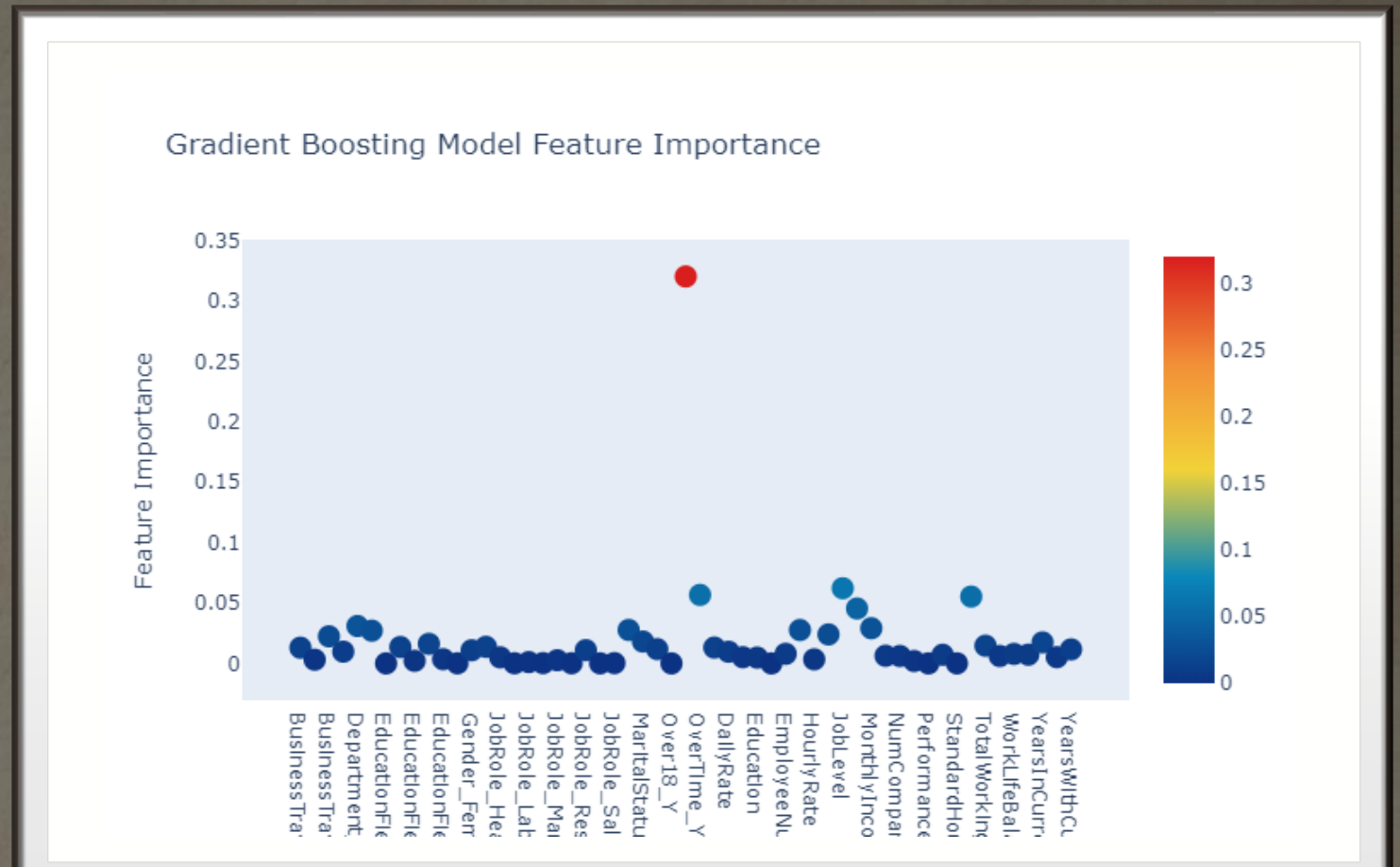
GRADIENT BOOSTING

- Train our model using Gradient Boosting
- Gradient Boosting AUC = 0.88
- Accuracy Score for Gradient Boosting Model is 88%

ROC GRAPH


ROC Graph
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
Logistic Regression (area = 0.66)
Random Forest Classifier (area = 0.61)
Gradient Boosting Classifier (area = 0.66)
Base Rate

FEATURE IMPORTANCES



CONCLUSION

- Based on how we want to optimize recall or precision, management level decisions will be made. Management decisions can be made to optimize features like employee work life balance, overtime etc to decrease the attrition rate.
- Implementing Yoga or Stress Management courses can also help achieve job satisfaction.