Understanding Employee Attrition with Machine Learning

Presented by: TAWANDA NIGEL CHITAPI.

Why

LOSS OF EMPLOYEE PRODUCTIVITY

BAD COMPANY REPUTATION

LOW MORALE

TRAINING COSTS
(TIME & MONEY)

MISSED PROFIT OPPORTUNITY

Problem Statement

- Employee attrition adversely impacts business productivity and profitabily with the loss of experienced employees

Value Add

- Identify factors that lead to employee attrition and recommend measures that the business should take to retain their employees
- Build a model that predicts attrition before it occurs

Data Collection

- Sythentic data from the IBM HR Analytics, Watson Dataset created by IBM Data Scientists

Data Description

TARGET VARIABLE - 'ATTRITION'

- 0 = Employee stayed on the Job.
- 1 = Employee left their Job.

INDEPENDENT VARIABLES - 34 COLUMNS

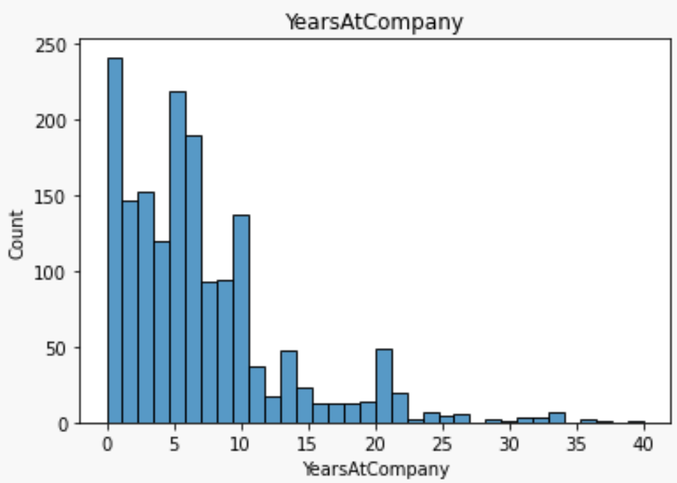
- Represent various employee attributes such as JobRole, MonthlyIncome, YearsInCurrentRole amongst many others.

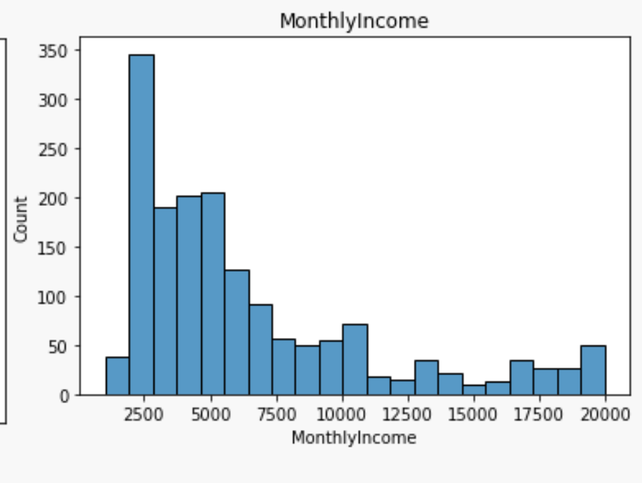
DATA OBSERVATIONS

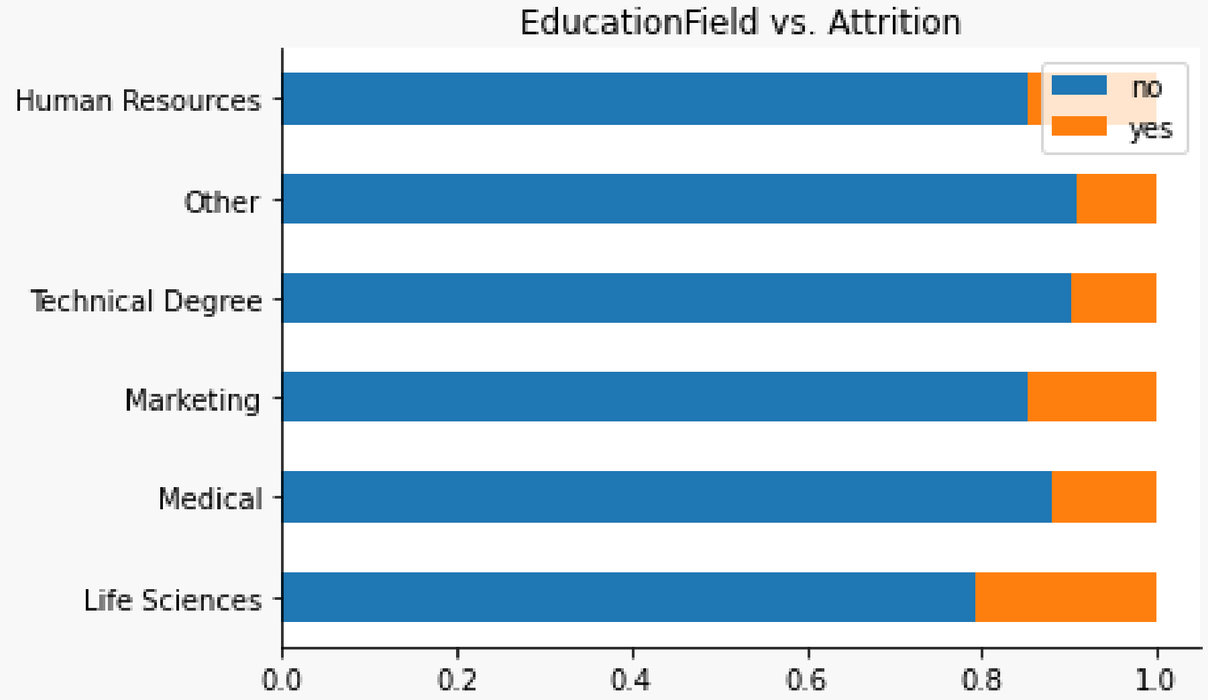
- Represent individual employees.
- There a total of 1676 employees recorded.

Completed Clean-Up and EDA

- Univariate Analysis
- Bivariate Analysis against the target variable



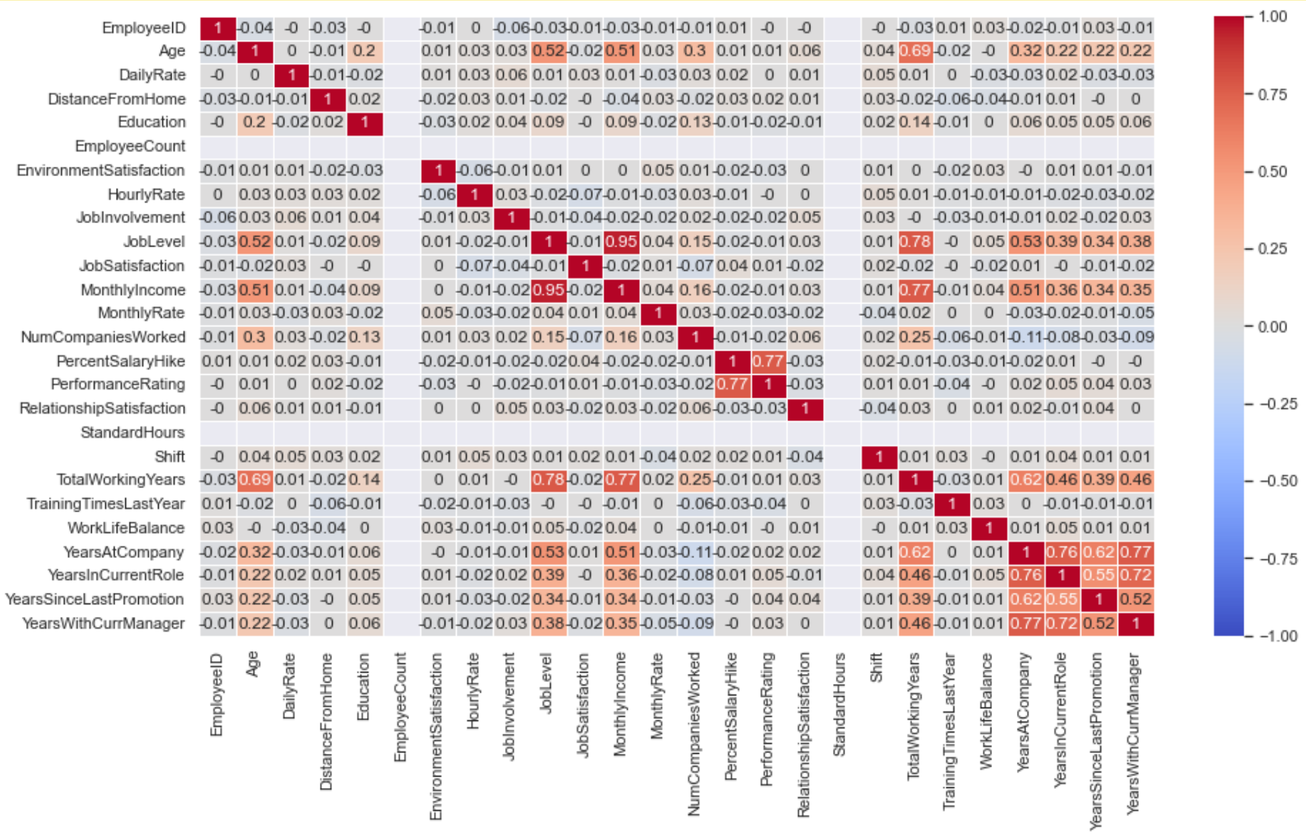




Completed Clean-Up and EDA

- Checked for multicollinearity
- Feature engineering and selection performed
  - One Hot Encoding
  - Fisher Score - Chi Squared Test
  - Cut X-features from 50 - 19
- Upsampled Data
  - SMOTE Technique

Correlation Heat Map



Completed Modelling

- Data fit onto Logistic Regression Model
  - Tested accuracies
- Optimized Hyperparameters
  - Graphically
  - Pipeline GridSearch
- Model evaluation performed
  - Confusion Matrix
  - Classification Report

1

Logistic Regression Classification Report

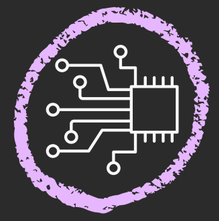
	precision	recall	f1-score	support
0	0.94	0.98	0.96	443
1	0.81	0.57	0.67	60
accuracy			0.93	503
macro avg	0.88	0.77	0.81	503
weighted avg	0.93	0.93	0.93	503

2

Decision Tree Classification Report

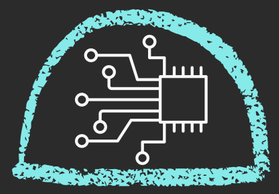
	precision	recall	f1-score	support
0	0.94	0.88	0.91	443
1	0.40	0.58	0.47	60
accuracy			0.84	503
macro avg	0.67	0.73	0.69	503
weighted avg	0.88	0.84	0.86	503

Plan of Action



Fit Random Forest

Ensamble multiple Decision Tree models
Evaluate Model



Put Into Production Trial

Questions