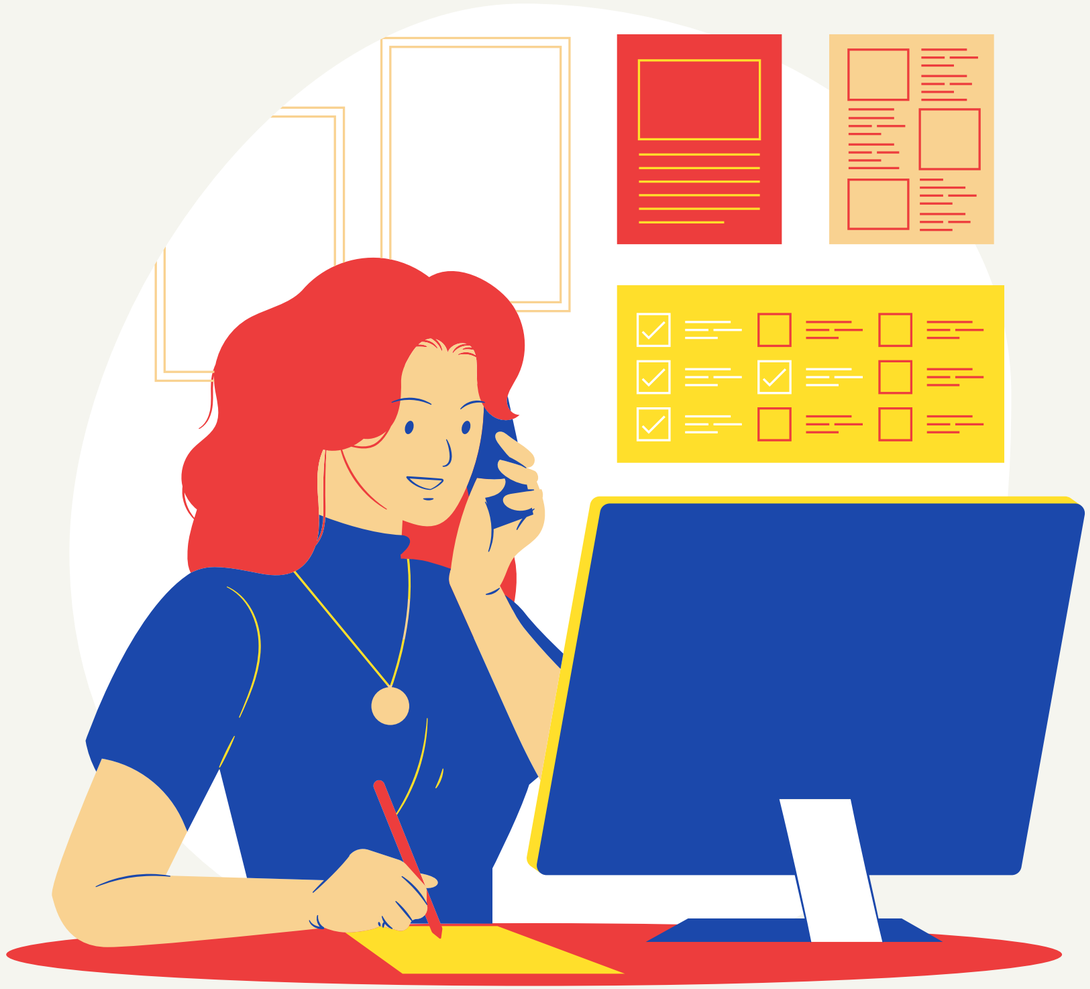
Employee Attrition Analysis and Prediction

A project presented by **Mandy Hong**



Introduction

Part 01

Background

Overview of the project

One of the keys to success in an organization is the ability to hire and retain great talent. It helps minimize the costs in recruiting and contributes to the overall growth of the business. In this project, a dataset on employee attrition will be used to help find insights in the reasoning behind employee attrition.

This project aims to discover insights that will help a company make conscious decisions about hiring talents and improve employee satisfaction and retention.

The Problem

What we want to solve

We want employees to stay longer in the company and reduce the overall attrition rate among employees

Hypothesis

Income, Work Life Balance, Working hours are factors that contribute to employee attrition

Objectives

What we want to achieve

Utilize dataset to uncover insights in the reasoning behind employee attrition and find the most important factors that contribute to attrition

Develop a classification model that helps predict employee attrition

Create action items for the company to prevent future loss of talent

Data Sources and Analysis

Part 02

Data Sources

The dataset is found on kaggle.com, it is a fictional dataset created by data scientists from IBM. It is fictional because it is unlikely that a company would release real data on their employees to protect their personal information.

However, the project will include processes that would be adaptable and helpful for an analysis using a real dataset.

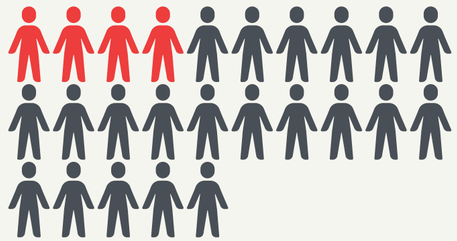
General Dataset Statistics

Shape of Dataset

1470 rows
35 columns

Attrition

4 out of 25 Attrited



Top Correlated Factors

01

Education Field

The probability of attrition depends on the education field of the employee

02

Job Role

Some job roles have a higher rate of attrition

03

OverTime

Whether an employee works overtime contributes highly to whether they will attrit

04

Stock Option Level

The number of stock options provided by the company contributes to attrition rate

05

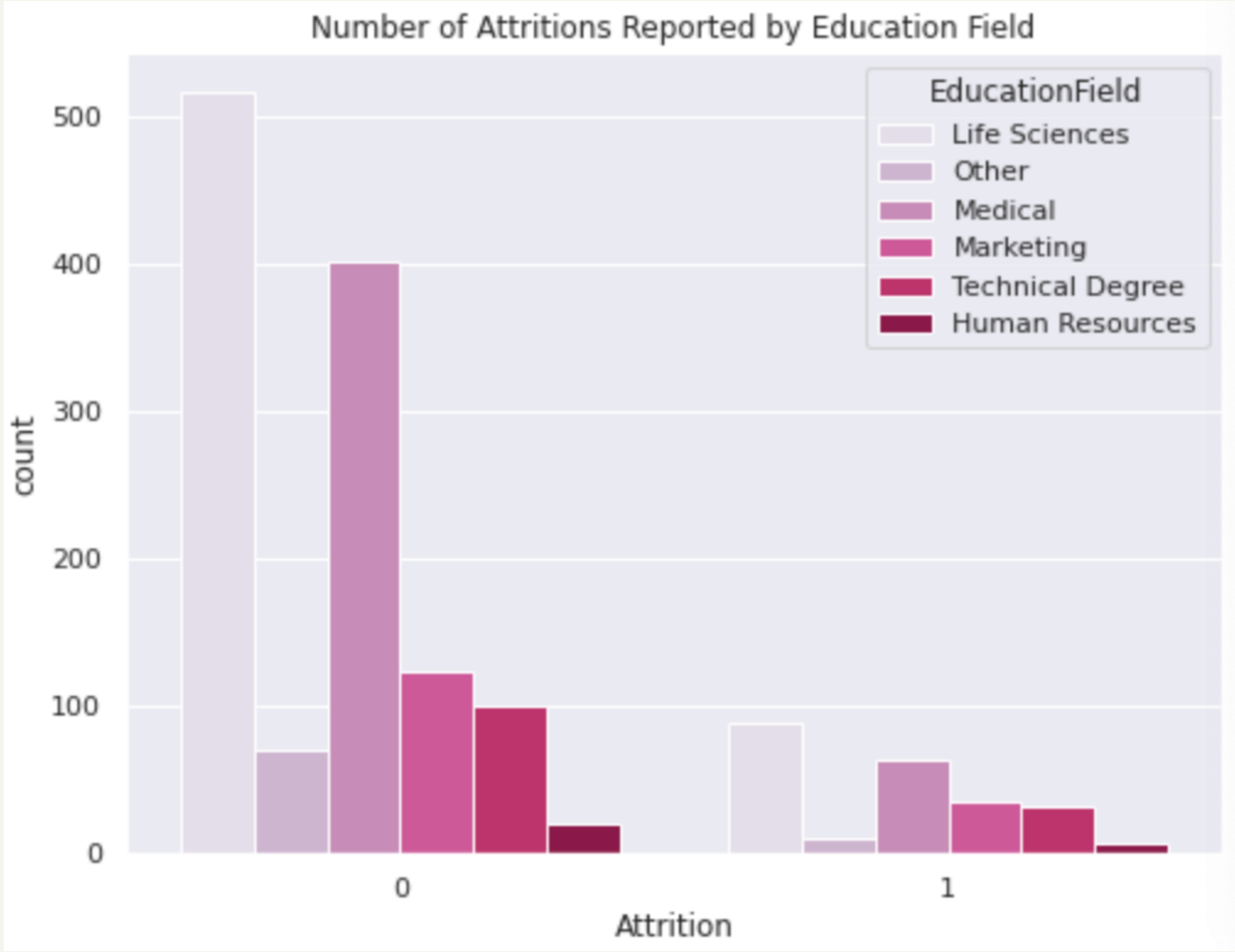
Environment Satisfaction

Environment Satisfaction rating is correlated to attrition

Education Fields

This plot shows the number of attritions reported by Education Field.

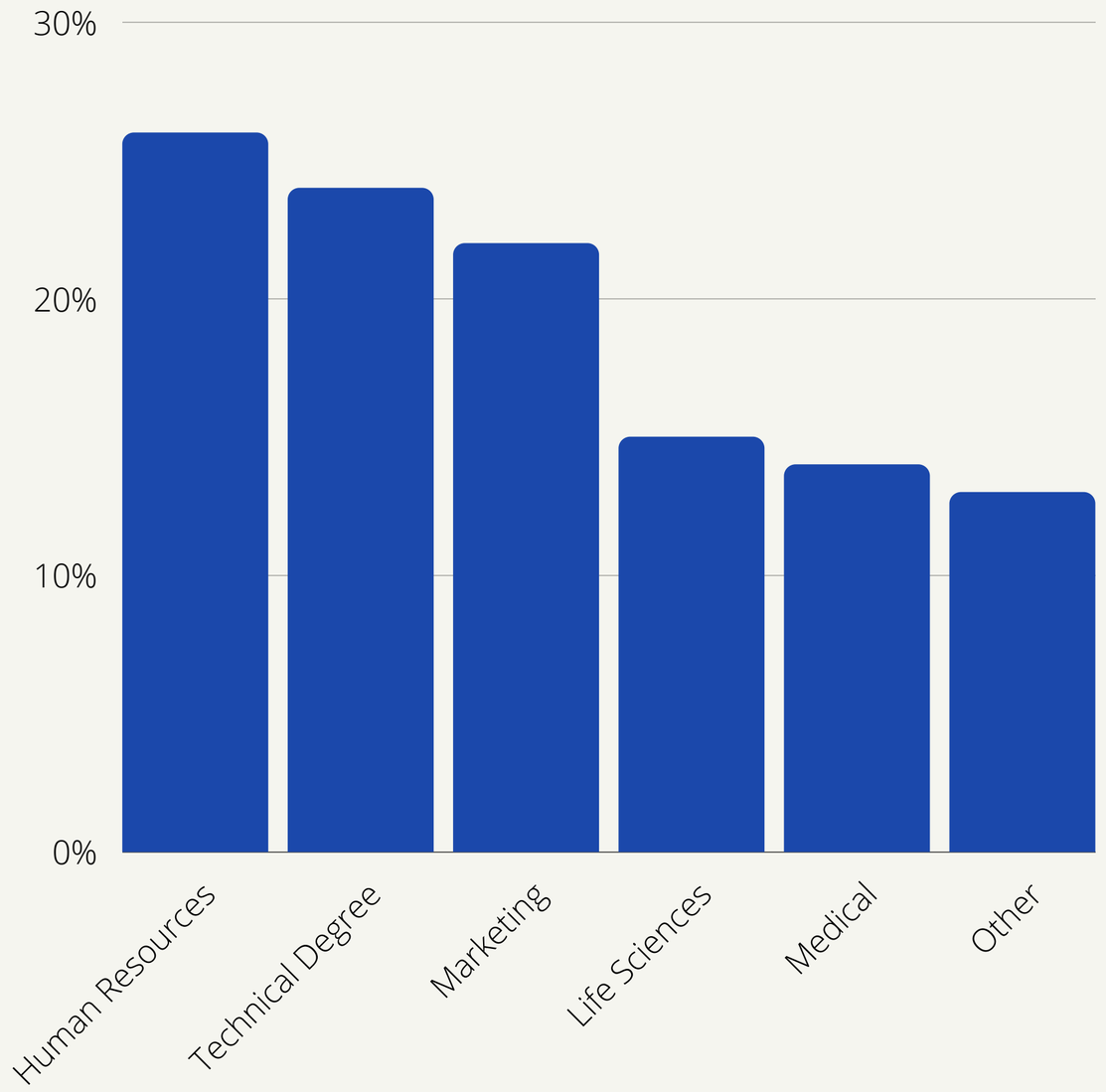
It seems that although technical degree, marketing, and human resources fields make up a small proportion of the dataset, the proportions of the attrited employees are higher.



Education Fields

This plot shows the percentage of attrition by education fields.

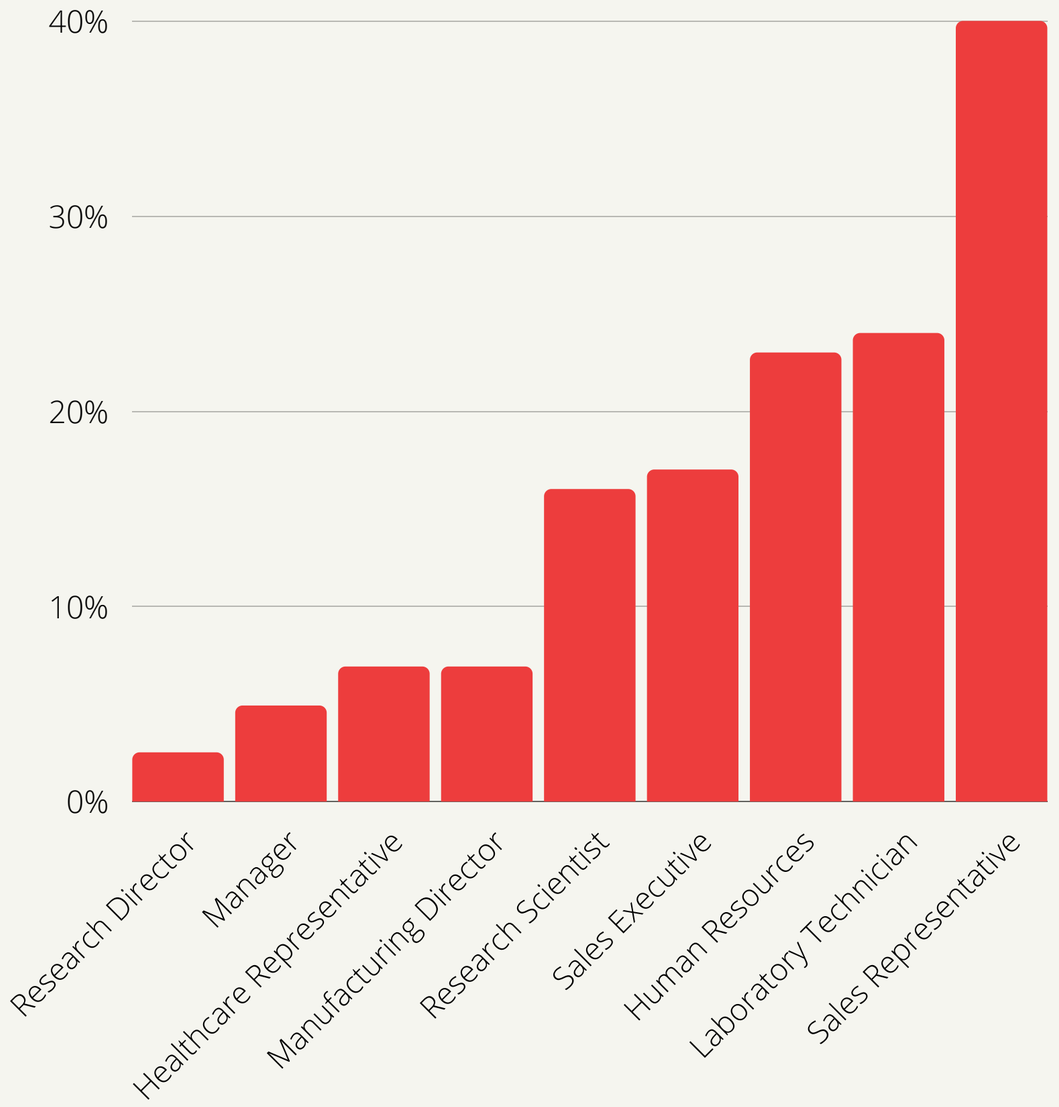
It seems that human resources, technical degree, and marketing have a higher percentage of attrition.



Job Roles

This plot shows the percentage of attritions reported by Job Role.

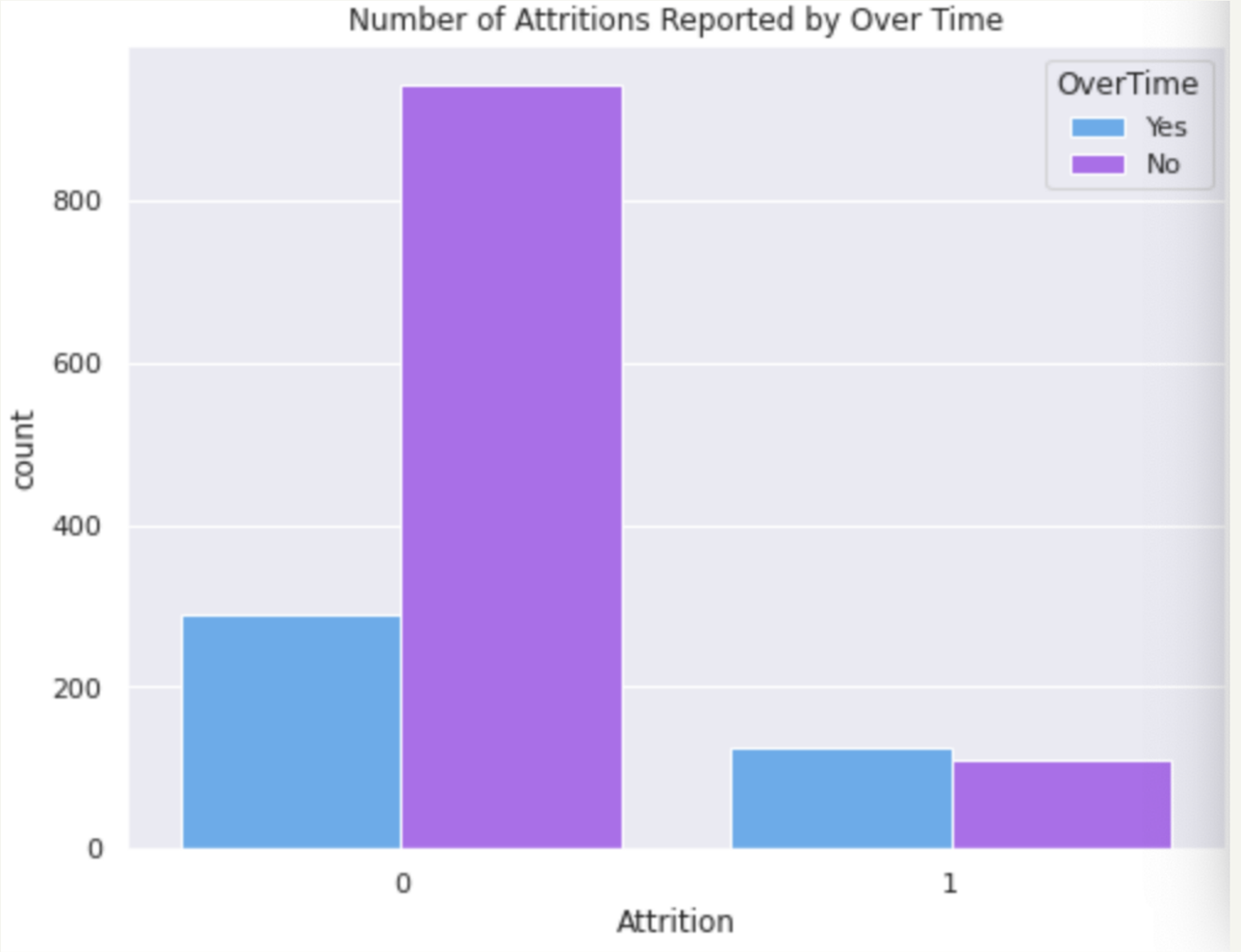
This statistics show that almost 40% of attrited employees are sales representatives, followed by laboratory technicians and human resources positions.



Over Time

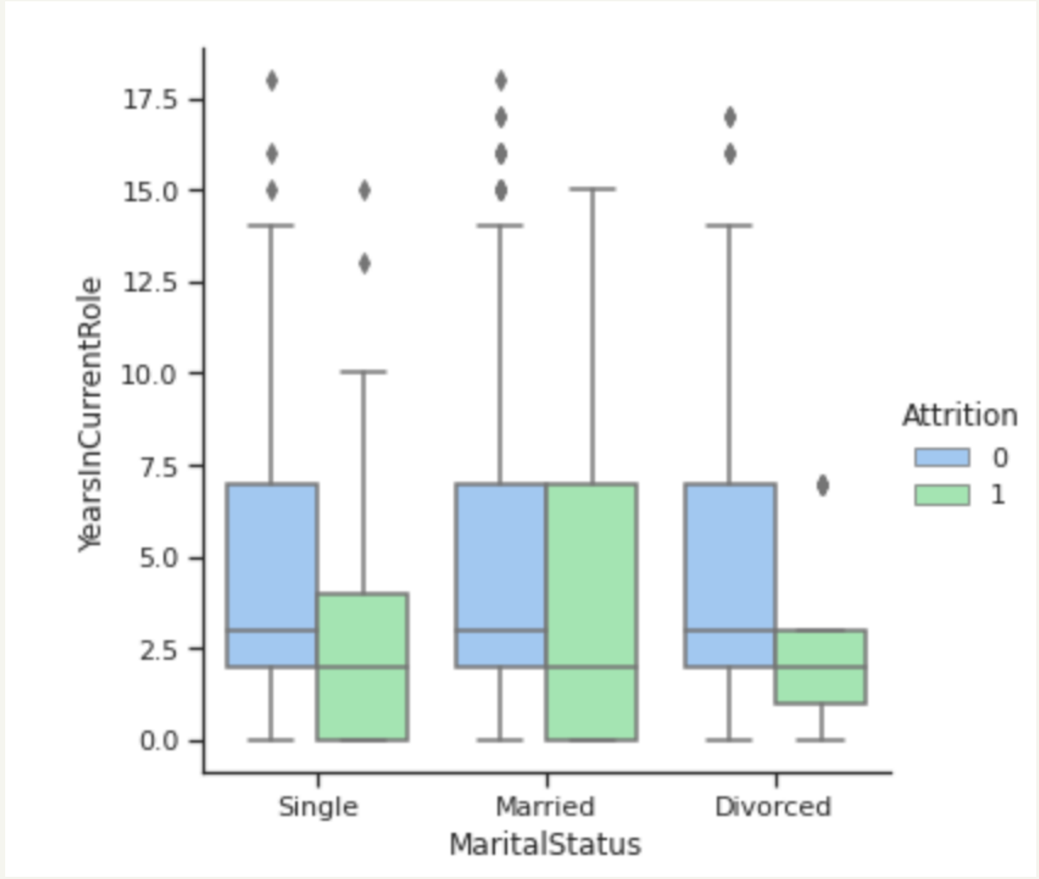
This plot shows the number of attritions reported by Over Time status.

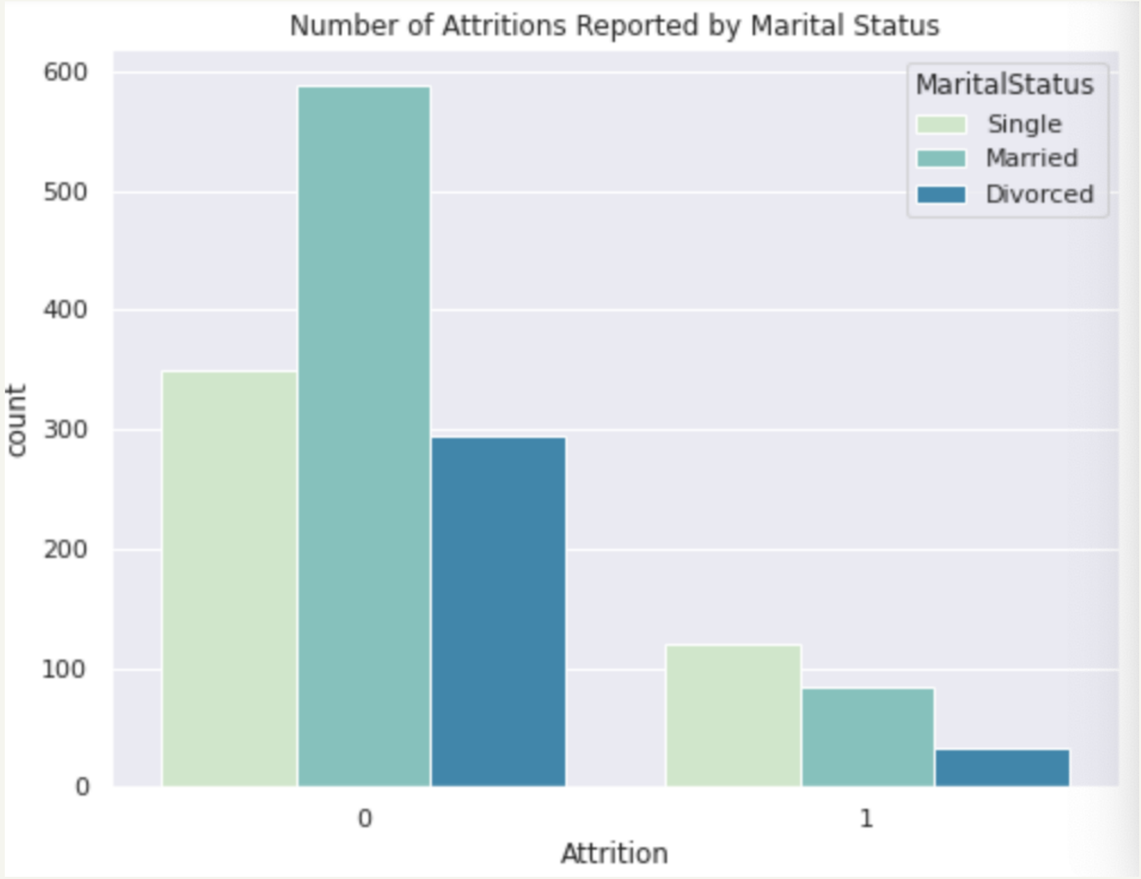
Although the number of overtime workers make up a small proportion of the dataset, about 30.5% of overtime employees have attrited, compared to 10% for non-overtime employees. This means that it is a factor that needs to be accounted for employee attrition.



Marital Status vs Years in Current Role

The boxplots show the distribution of Years in Current Role by Marital Status, Attrition vs No Attrition. It seems that the for non-attrition, an average employee tends to have worked longer at the company. However, for married and attrited employees, the distribution of Years in Current Role span from 0 to 7.5. Overall, about half of the employees are married and only a small percentage of them chooses to leave the company.





Findings and Highlights

Highlight 1

employees with a background in technical degrees, human resources, and marketing fields have a higher tendency in attrition compared to medical, life sciences and other fields.

Highlight 2

30% of overtime employees have attrited compared to employees that don't work overtime. Frequent Business Travels also contribute to attrition.

Highlight 3

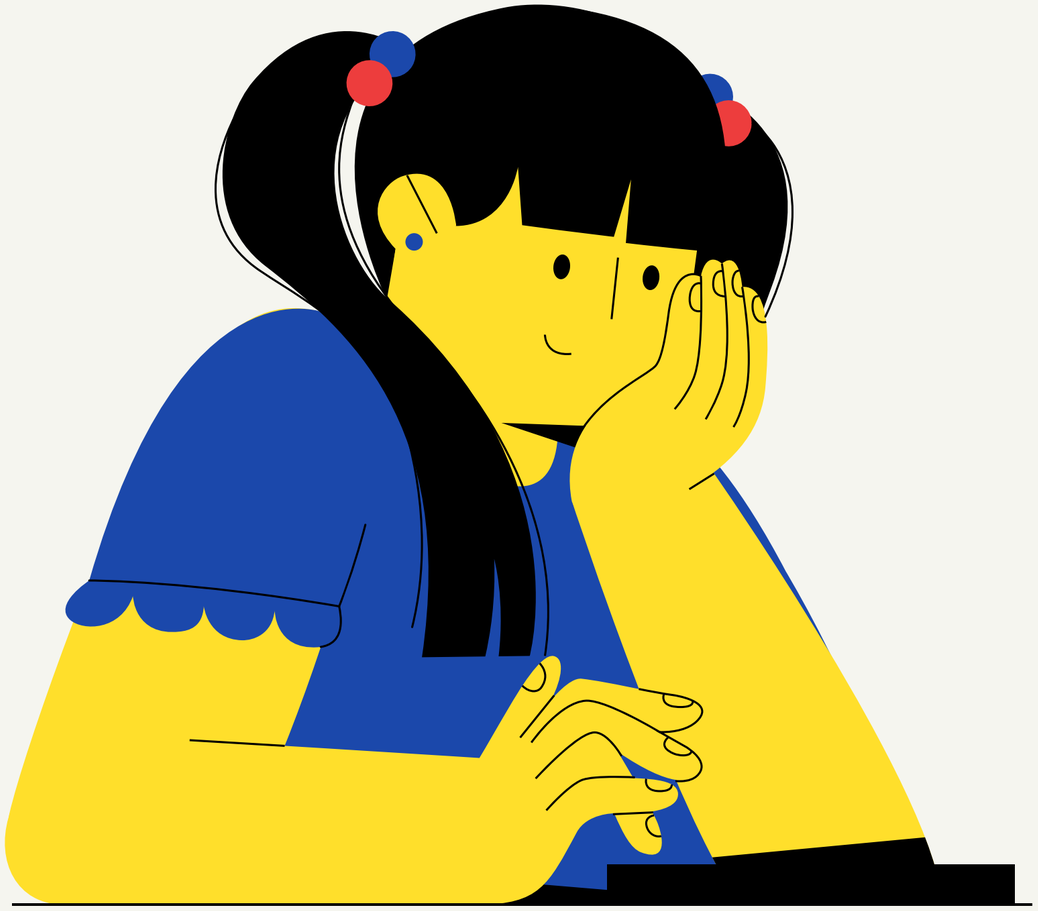
In terms of Job Role, almost 40% of attrited employees are sales representatives, followed by laboratory technicians and human resources positions.

Highlight 4

Stock Option level is also a big factor that contributes to attrition. i.e. if the company offers more stock options for the employees, they are less likely to attrit.

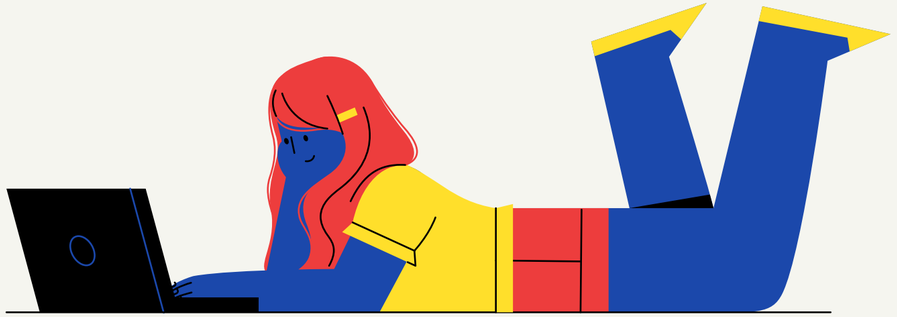
Highlight 5

The higher the employee is involved in their jobs, the less likely they are to leave the company.



Modeling

Part 03

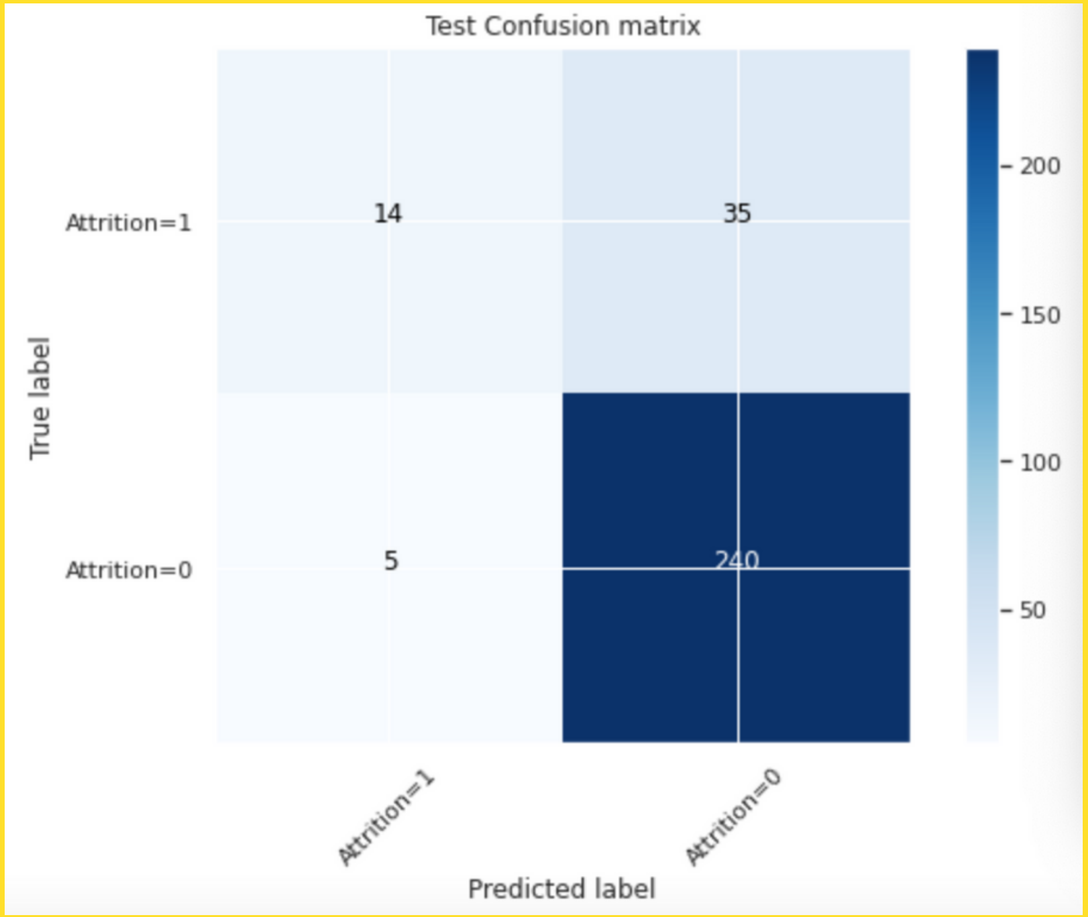


A logistic regression model was created to predict attrition. A heatmap was used to select variables with an absolute correlation value greater than 0.1. And dataset was split into 80% training and 20% testing data.

Testing Confusion Matrix

This confusion matrix shows the number of predictions that were predicted 1/0 by their true labels.

Our f1 score on the testing set was not less than the f1 score on the training set, so we likely did not overfit. The recall score for the testing data is larger than for the training data, this means that our model identifies attrition better in an unfamiliar set of data. The model has an accuracy of 86%.

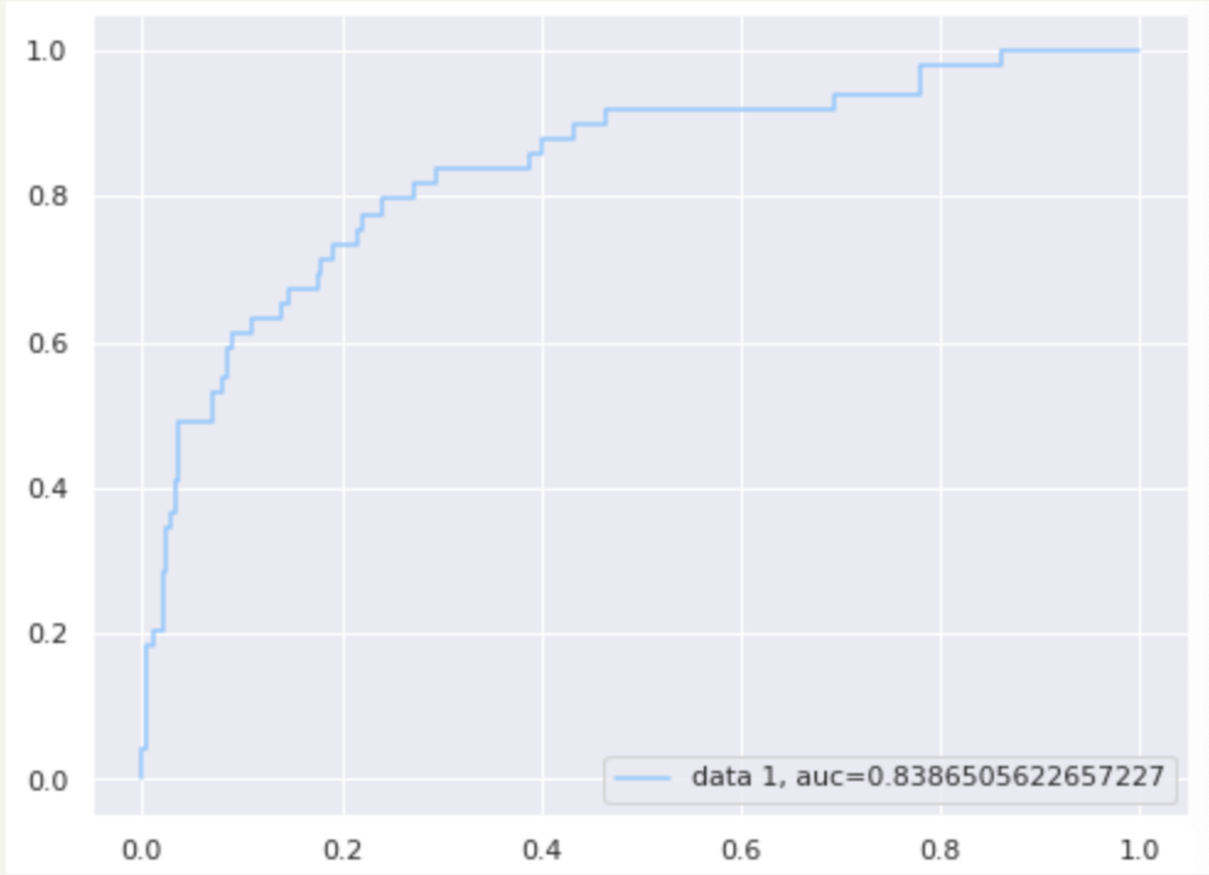


	precision	recall	f1-score	support
0	0.87	0.98	0.92	245
1	0.74	0.29	0.41	49
accuracy			0.86	294
macro avg	0.80	0.63	0.67	294
weighted avg	0.85	0.86	0.84	294

ROC Curve

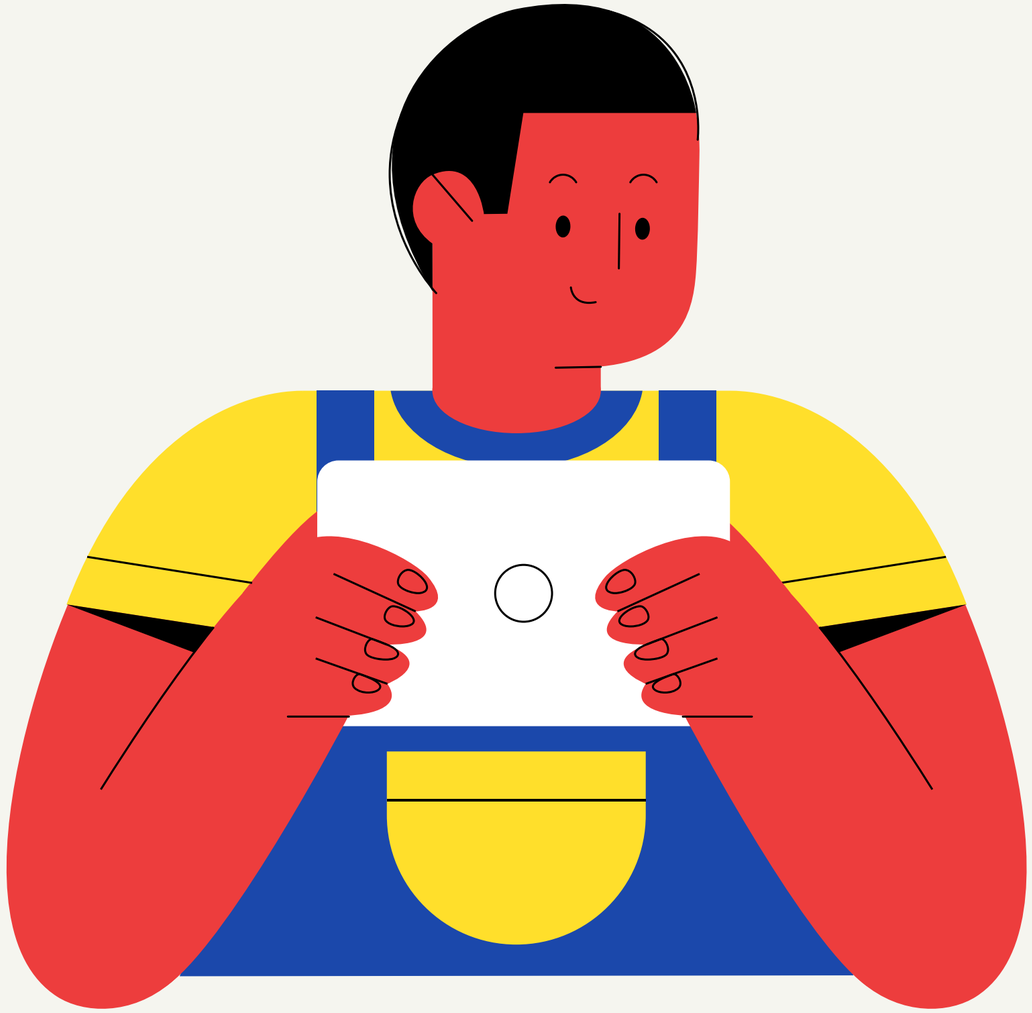
Receiver Operating Characteristic(ROC) curve is a plot of the true positive rate against the false positive rate. It shows the tradeoff between sensitivity and specificity.

Compared to a 0.5, which means the model is not useful, the AUC curve for this model is 0.84, which means it is moderately a good classifier.



Implications

Part 04



Implications

Implication 1

Employees with background in technical degrees, human resources, and marketing fields need to be looked after based on their job roles.

Implication 2

We should reduce employee's overtime work to increase job satisfaction

Implication 3

We should offer more incentives for sales representatives, laboratory technicians and human resources positions.

Implication 4

We should consider offering more stock options for employees.

Implication 5

For all positions, we should help increase employees' involvement in the job to show that they can make a true contribution to the company.

Action Items to Reduce Attrition

01

Overtime Compensation

Provide compensation for overtime workers to increase job satisfaction

02

Involvement

Train managers to be more inclusive of subordinates to make them feel contributed in the job

03

Stock Options

Provide more stock options for employees

04

Benefits

Provide more benefits and incentives for employees with human resources, technical, and marketing background especially who work as a sales representative e.g. reduce work hours / provide flexibility

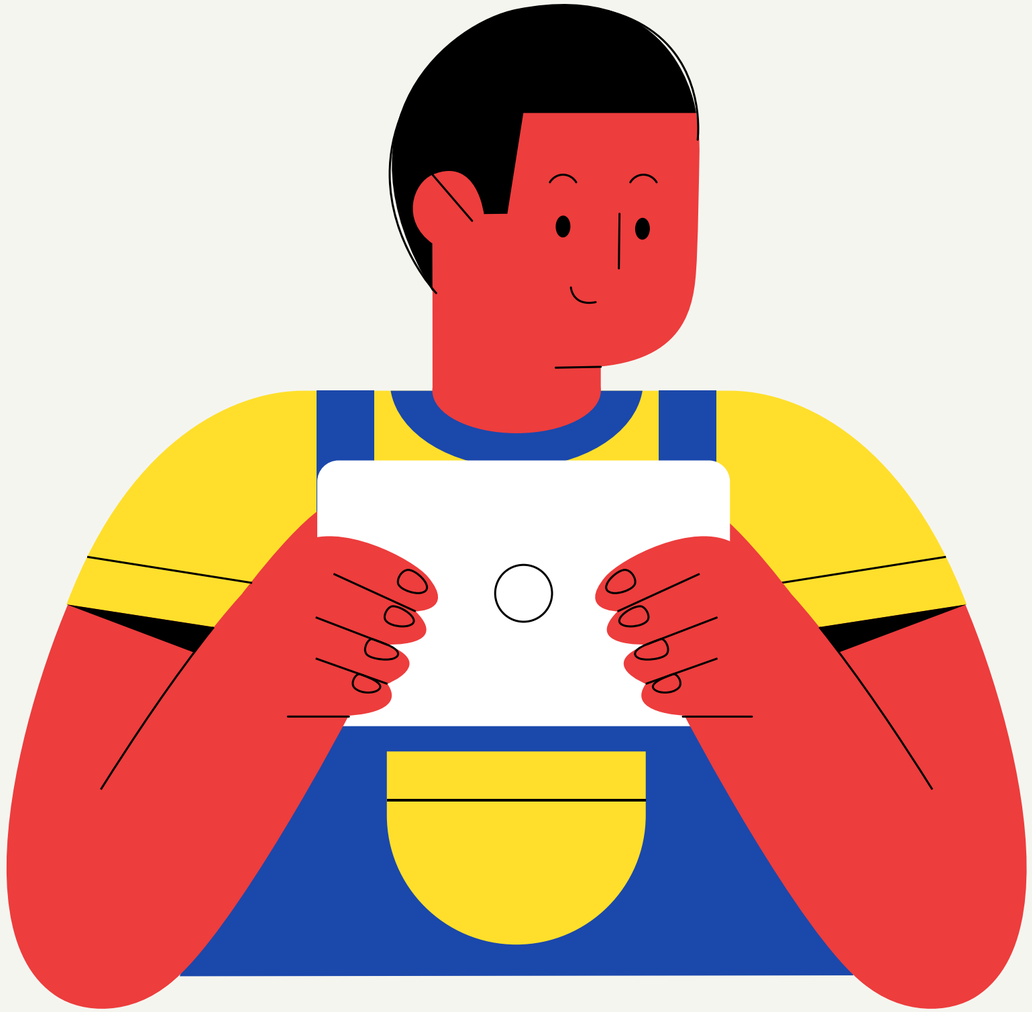
05

Environment

Invest in new items and provide a collaborative and creative workspace for employees

Conclusion

Part 04



Logistic regression analyses were run using the dataset and correlated factors such as environment satisfaction. overtime, job role, education fields were discovered. A model was created to predict attrition with an accuracy of 86%.

UNCOVERED KEY FACTORS FOR ATTRITION

Project highlights

Job involvement is a surprising factor that contributes to attrition. Job roles that offer less job involvement should be provided with more incentives and tasks that help them feel more involved in the job.

NEW UNDERSTANDING OF THE PROBLEM

Significant discovery

Income and work life balance ratings are in fact not important factors that contribute to attrition. We should look more at the human factors such as satisfaction with the environment and the functions behind job roles.

DISAGREEMENTS WITH HYPOTHESIS

Interpretation of findings

Sources

Building a Logistic Regression in Python Step-by-Step

Understanding Logistic Regression in Python

Scikit-Learn Logistic Regression Documentation

Python Sklearn Logistic Regression Tutorial with Example

Precision and Recall