Clicked Data Analytics

Bain and Company - Attrition

REPORT

By

Jennifer Opal

Ibukunoluwa Adeyemi

April, 2024

Introduction

Attrition is a critical issue that affects organizations across various industries, and it is essential to identify the factors contributing to it to develop effective retention strategies. As a data analyst at Bain and Company, we have been given the opportunity to work closely with Kim Wexler, Head of HR, to analyze the HR data set and gain insights into the company's unprecedented employee attrition rates.

The data set provided by Kim Wexler is rich in information, including various variables that may have an impact on employee attrition. These variables include demographic information, job-related factors, and employee satisfaction measures

In this report, we will be using empirical evidence and analytical techniques to construct a productive and develop a robust and effective strategy to reduce employee turnover and promote a positive, productive work environment.
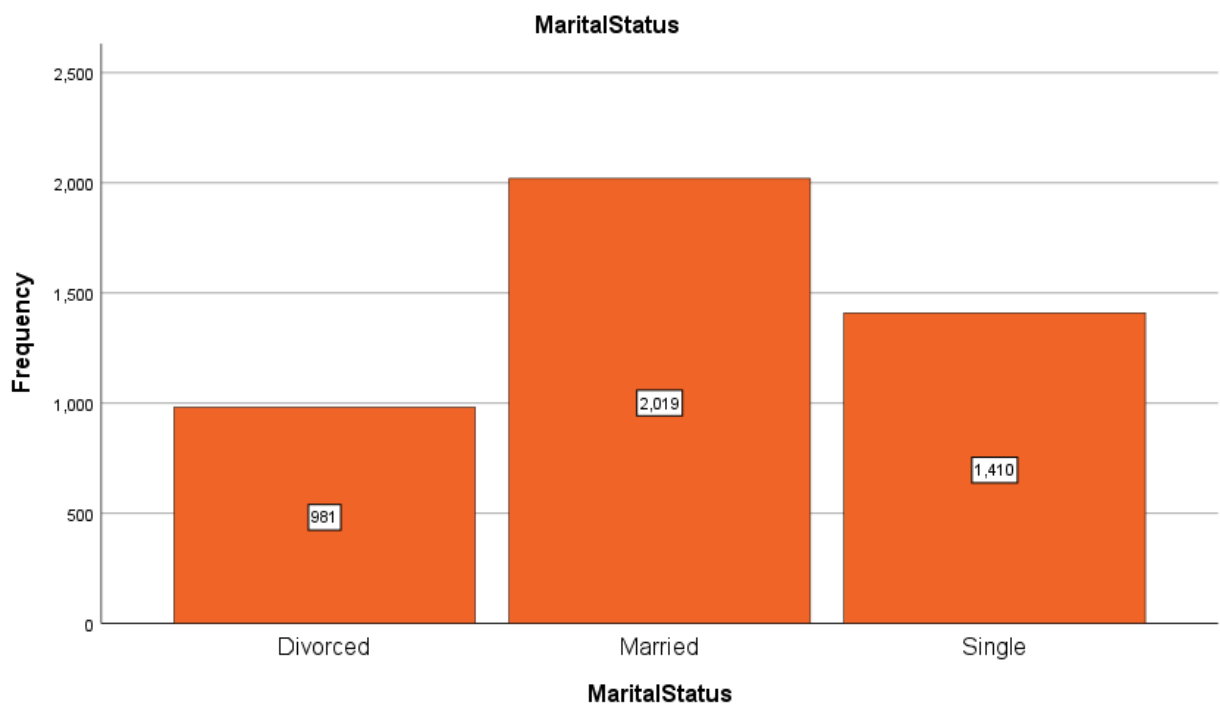
Descriptive Statistics of Workers in Bain and Company
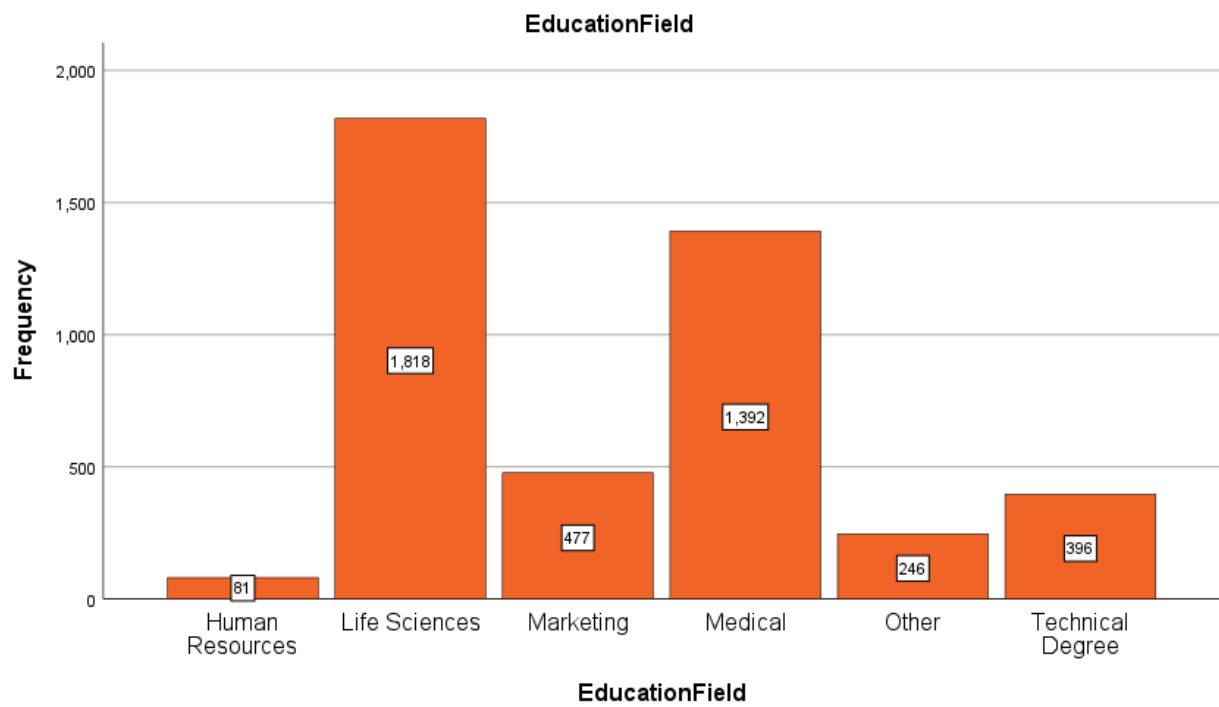

Fig 1: Marital Status of Respondents
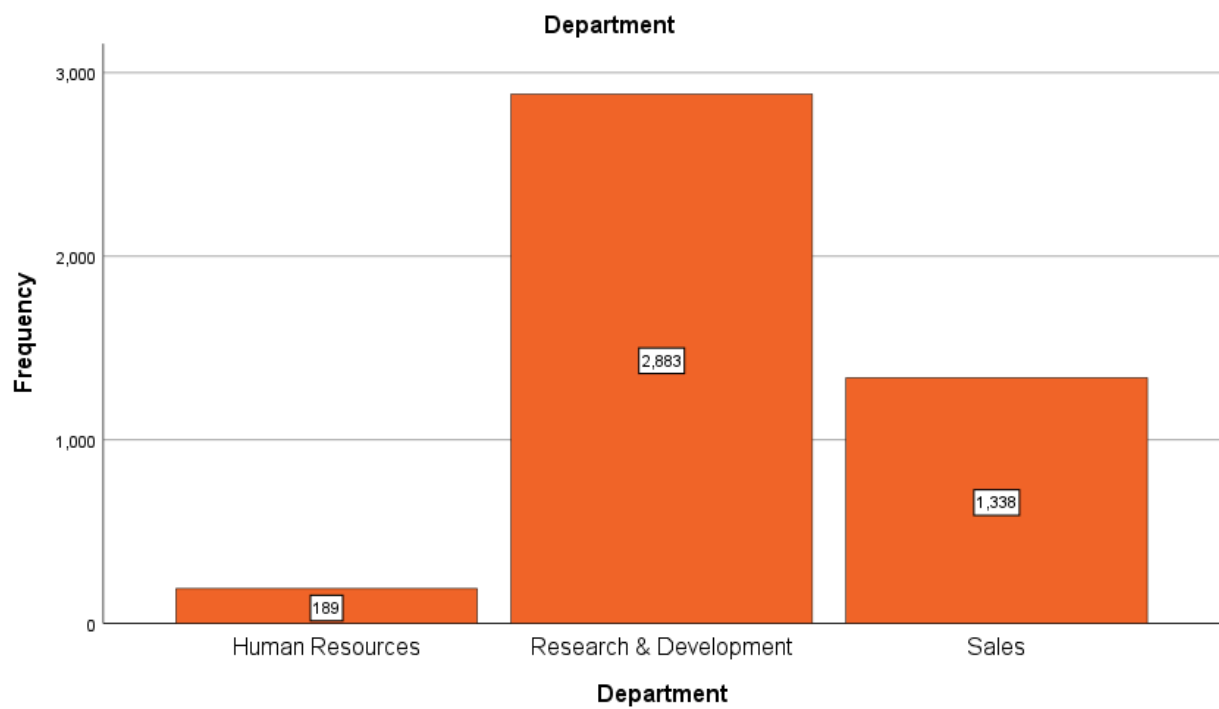
EducationField

Fig 2: Education Field of Respondents

Department

Fig 3: Department of Respondents

Gender

Legend:
- Female
- Male

40.00%

60.00%

Fig 4: Gender of Respondents

Yes — 16%

No — 84%
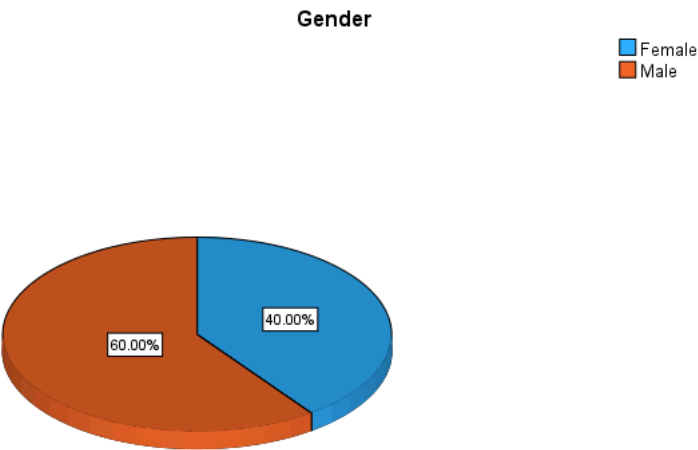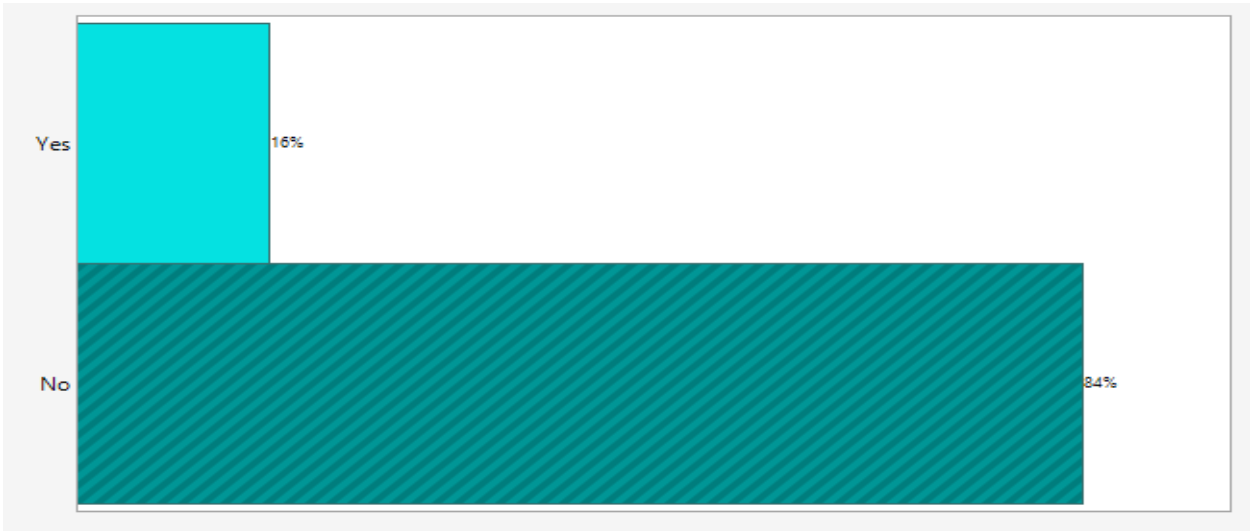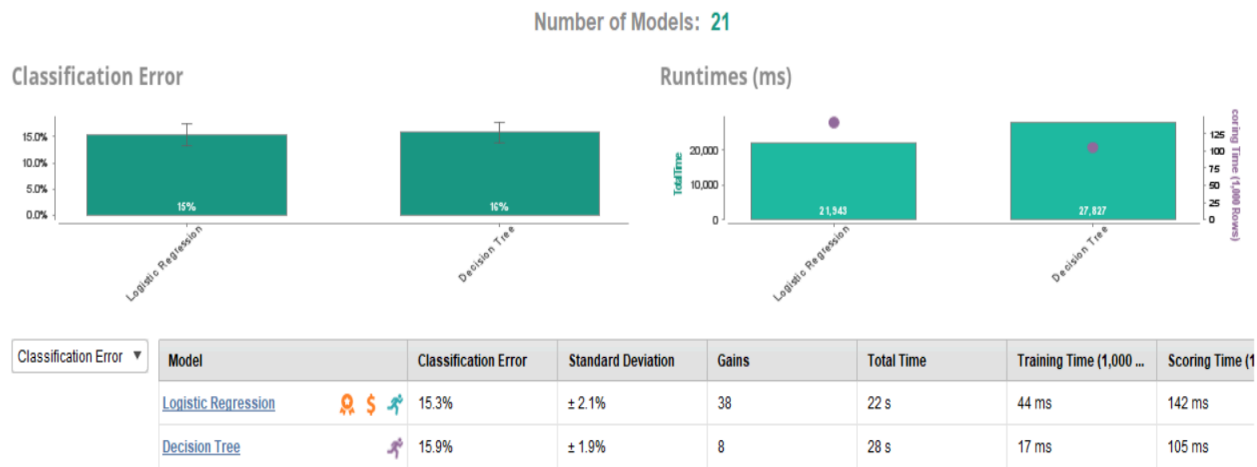
Fig 5: Attrition Rate

Prediction Model

For our Model, we made use of Total working years, age, years with current manager, years at company, environment satisfaction and job satisfaction to predict our attrition rate. There were 6 variables in total. We used this based of their weight by correlation and believe it will have great impact in predicting our attrition rate.

Weights by Correlation

Attribute	Weight
TotalWorkingYears	0.270
Age	0.159
YearsWithCurrManager	0.156
YearsAtCompany	0.134
EnvironmentSatisfaction	0.122
JobSatisfaction	0.109

Overview of Prediction Model used

Overview

Number of Models: 21

Classification Error — Logistic Regression 15%, Decision Tree 16%

Runtimes (ms) — Logistic Regression 21,343, Decision Tree 27,827

Classification Error ▼	Model		Classification Error	Standard Deviation	Gains	Total Time	Training Time (1,000 ...	Scoring Time (1
	Logistic Regression	👤 $ 🏃	15.3%	± 2.1%	38	22 s	44 ms	142 ms
	Decision Tree	🏃	15.9%	± 1.9%	8	28 s	17 ms	105 ms

We made use of Logistic Regression Model and Decision Tree since the goal is to accurately predict the attrition rate in production based on relevant independent variables.

Logistic Regression Performance

Logistic Regression - Performance

Profits from Model: 874   Profits for Best Option (No): 836   Gain: 38   [Show Costs / Benefits...]

Performances

Criterion	Value	Standard Deviation
Accuracy	84.7%	± 2.1%
Classification Error	15.3%	± 2.1%
AUC	70.9%	± 7.4%
Precision	85.7%	± 1.7%
Recall	97.9%	± 0.9%
F Measure	91.4%	± 1.3%
Sensitivity	97.9%	± 0.9%
Specificity	18.7%	± 9.9%

0.021110215997254562

Logistic Regression Confusion Matrix

Confusion Matrix

	true Yes	true No	class precision
pred. Yes	41	22	65.08%
pred. No	171	1026	85.71%
class recall	19.34%	97.90%	