import pandas as pd

dataset = pd.read\_csv('general\_data.csv')

dataset.head()

Out[3]:

Age Attrition ... YearsSinceLastPromotion YearsWithCurrManager

0 51 No ... 0 0

1 31 Yes ... 1 4

2 32 No ... 0 3

3 38 No ... 7 5

4 32 No ... 0 4

[5 rows x 24 columns]

dataset.columns

Out[4]:

Index(['Age', 'Attrition', 'BusinessTravel', 'Department', 'DistanceFromHome',

'Education', 'EducationField', 'EmployeeCount', 'EmployeeID', 'Gender',

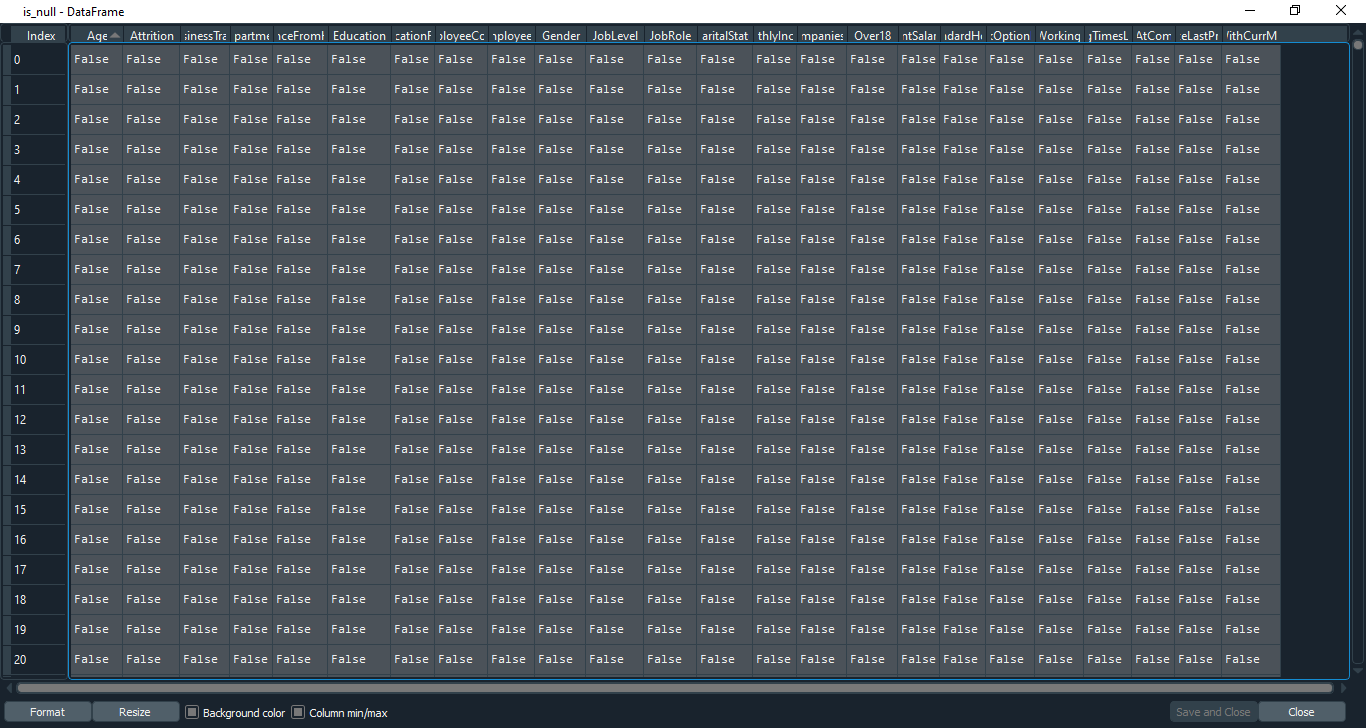
'JobLevel', 'JobRole', 'MaritalStatus', 'MonthlyIncome',

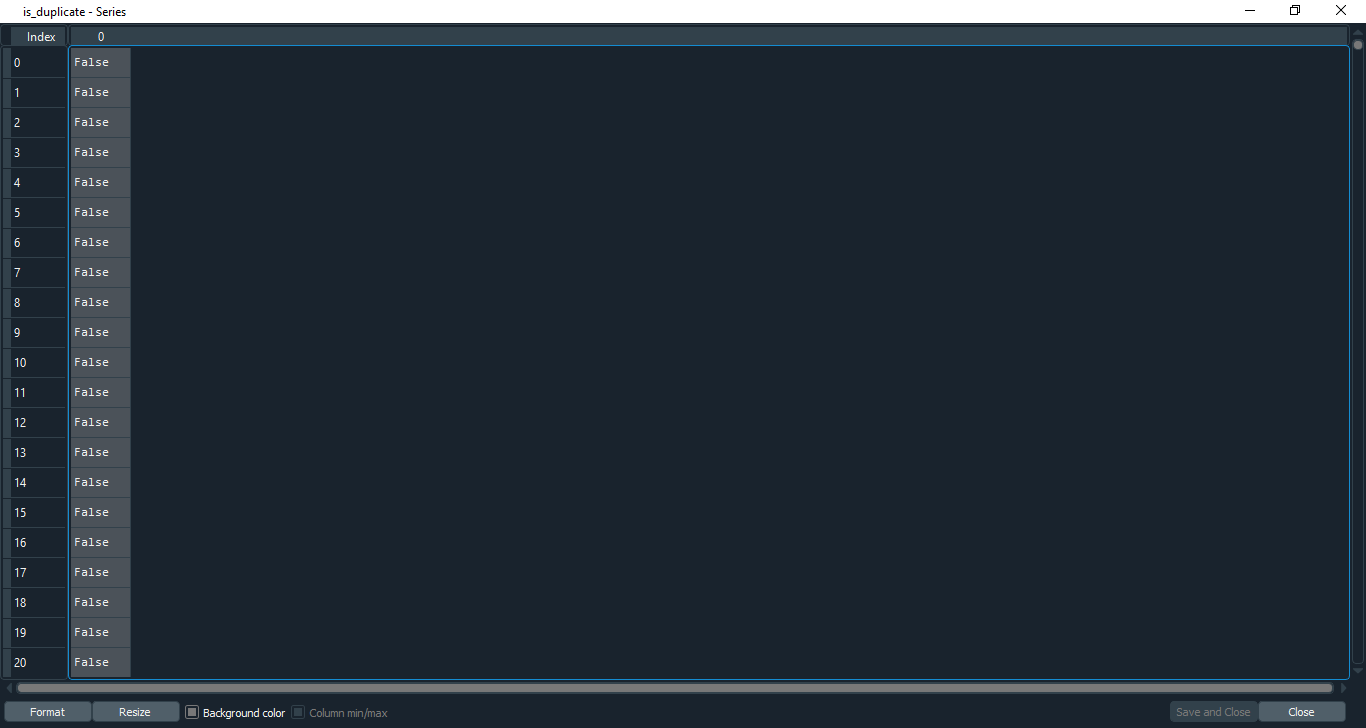
'NumCompaniesWorked', 'Over18', 'PercentSalaryHike', 'StandardHours',

'StockOptionLevel', 'TotalWorkingYears', 'TrainingTimesLastYear',

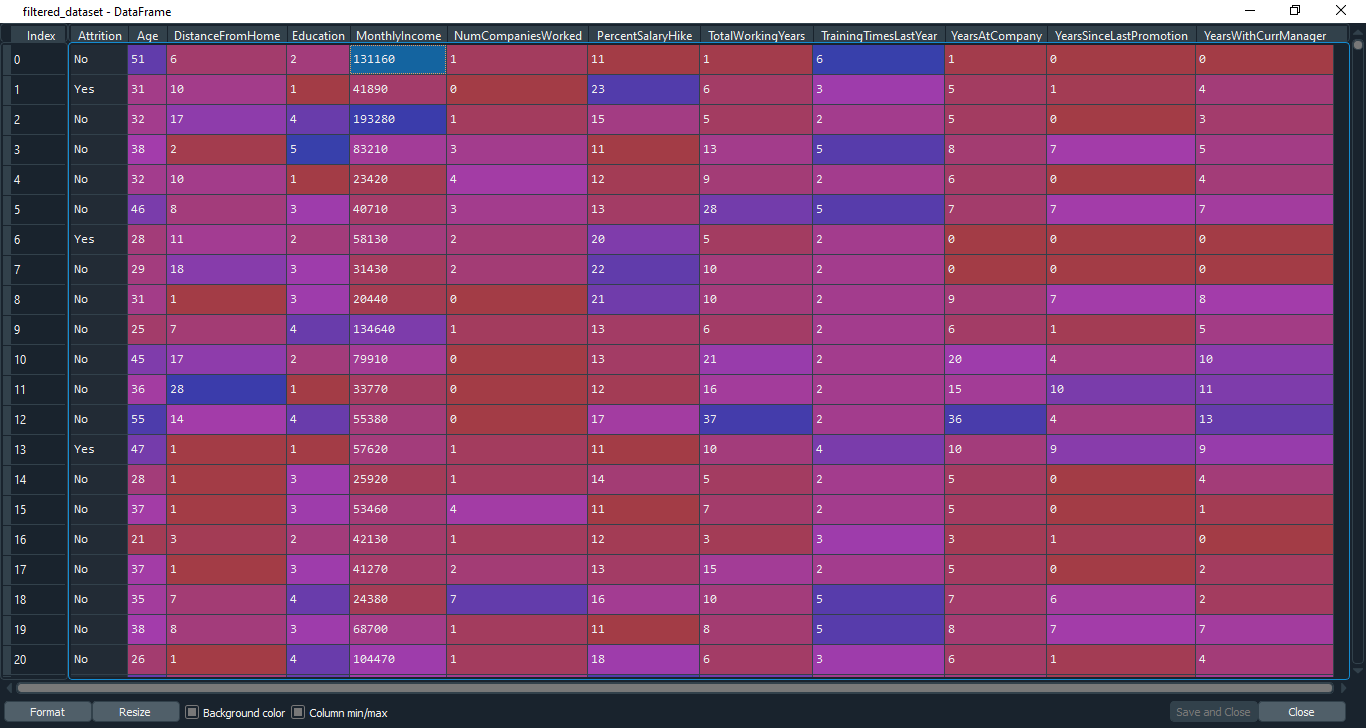
'YearsAtCompany', 'YearsSinceLastPromotion', 'YearsWithCurrManager'],

dtype='object')

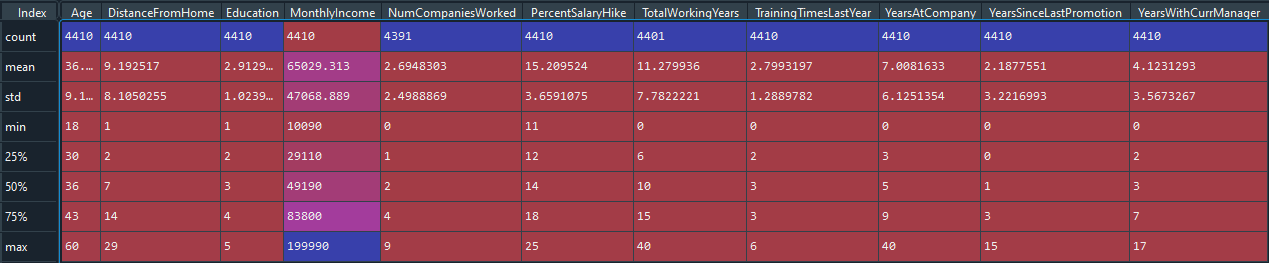
is\_null = dataset.isnull()

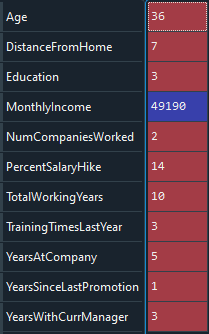
is\_duplicate = dataset.duplicated()

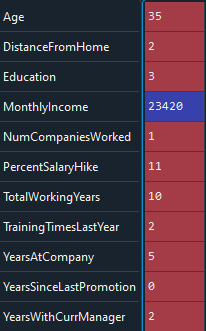
filtered\_dataset = dataset[['Attrition','Age','DistanceFromHome','Education','MonthlyIncome', 'NumCompaniesWorked', 'PercentSalaryHike','TotalWorkingYears', 'TrainingTimesLastYear', 'YearsAtCompany','YearsSinceLastPromotion', 'YearsWithCurrManager']]

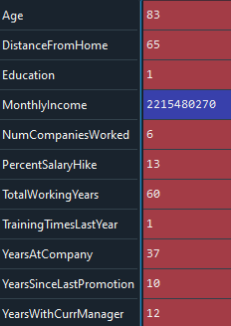


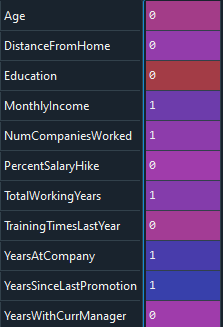
With these filtered data, we will analyse the data

description = filtered\_dataset.describe()

median = filtered\_dataset.median()

mode = filtered\_dataset.mode()

variance = filtered\_dataset.var()

skew = filtered\_dataset.skew()

kurt = filtered\_dataset.kurt()

Data Summary:

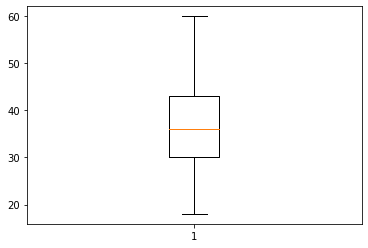
|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Mean | Median | Mode | Variance | Standard  Deviation | IQR | Skewness  (Rounded) | Kurtosis  (Rounded) |
| Age(yrs) | 36.9 | 36 | 35 | 83 | 9 | 13 | 0.413 | -0.406 |
| Distance  From  Home(km) | 9.19 | 7 | 2 | 65 | 8 | 12 | 0.957 | -0.227 |
| Monthly  Income(Rs.) | 65029.313 | 49190 | 23420 | 2215480270 | 47068 | 54690 | 1.369 | 1.000 |
| TotalWorking  Years | 11.28 | 10 | 10 | 60 | 7 | 9 | 1.117 | 0.913 |
| YearsAt  Company | 7.01 | 3 | 5 | 37 | 6 | 6 | 1.763 | 3.923 |
| YearsSince  Last  Promotion | 2.19 | 1 | 0 | 10 | 3 | 3 | 1.983 | 3.602 |
| YearsWith  CurrManager | 4.12 | 3 | 2 | 12 | 3 | 5 | 0.833 | 0.168 |

Inferences from Above data:

1. All above variable show close to normal skewness, but all are slightly positive
2. YearsAtCompany and YearsSinceLastPromotion are leptokurtic, and others are mesokurtic to platykurtic
3. The MonthlyIncome IQR suggests income-bands wide Attrition, and thus that’s not the reason of Attrition
4. Mean age forms a near normal distribution with 13 years of IQR(close to 0 in perfect normality)

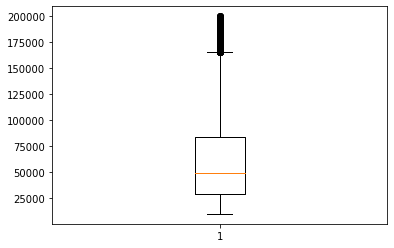
Outliers:

Age:

plt.boxplot(dataset.Age)

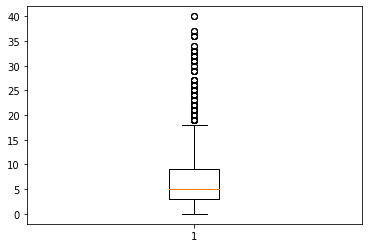
Age is normally distributed without any Outliers.

Monthly Income:

plt.boxplot(dataset.MonthlyIncome)

Monthly income is Positively(Right) skewed with several outliers

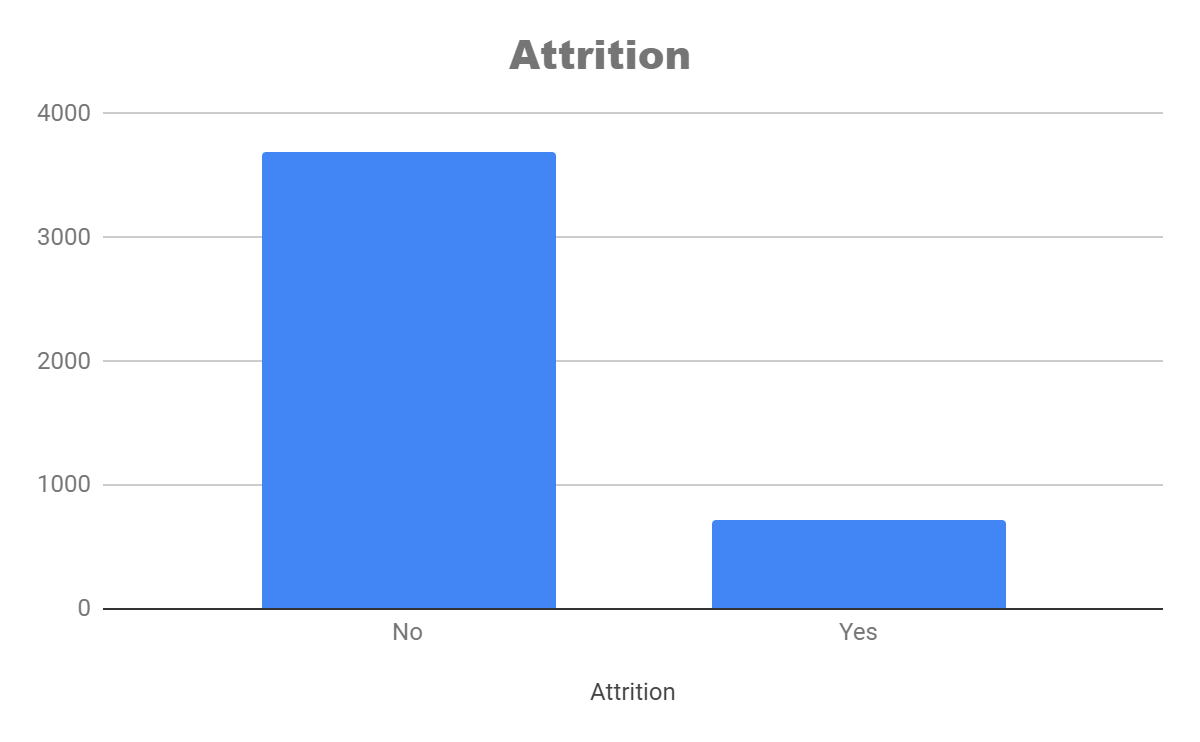
Years at company:

plt.boxplot(dataset.YearsAtCompany)

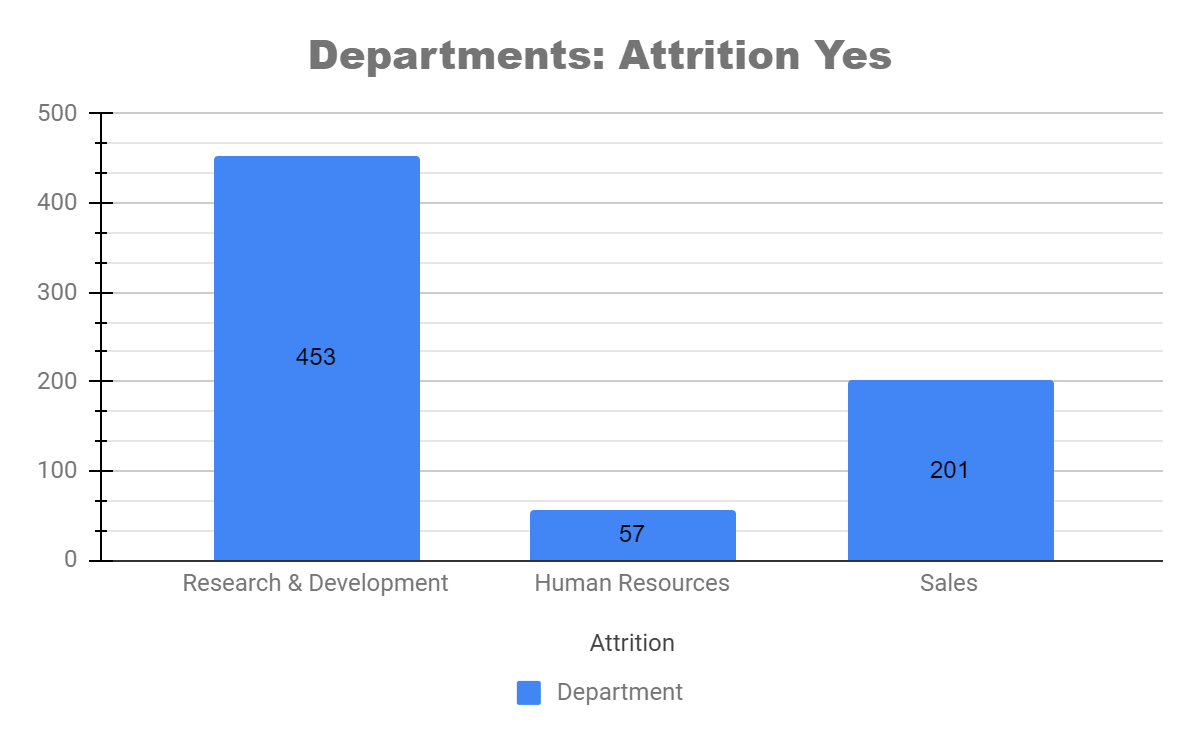
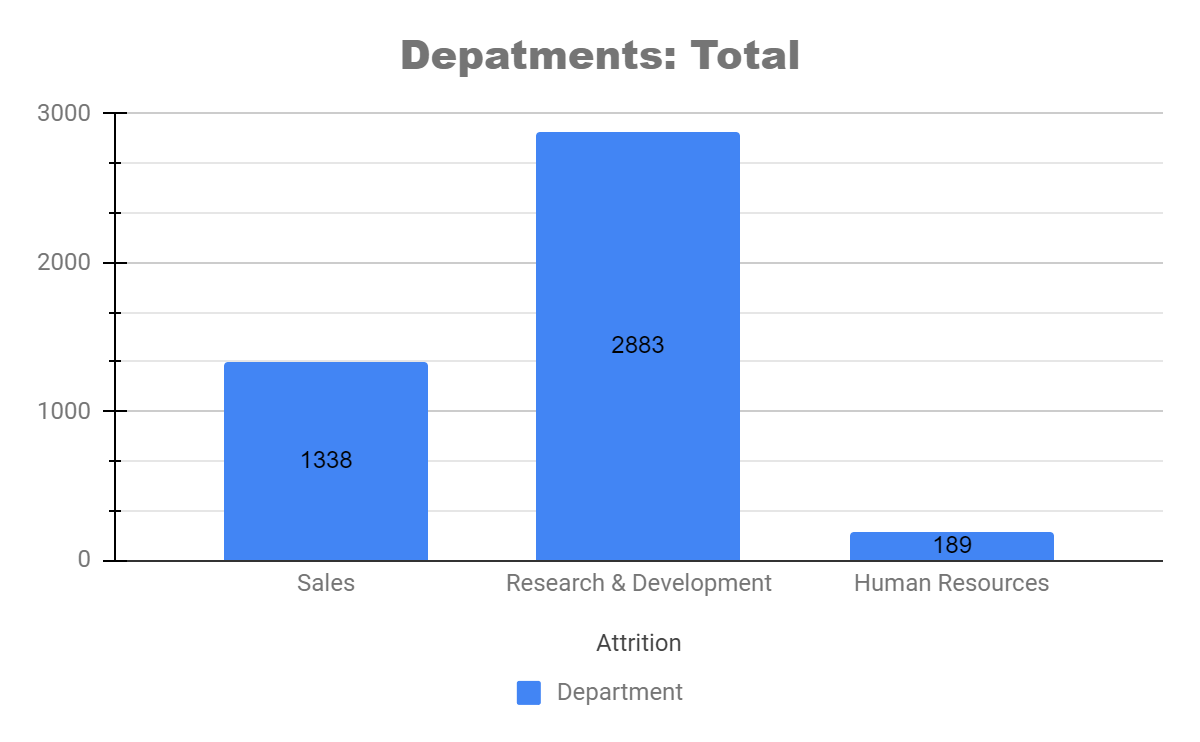
Years at company is also Positively(Right) Skewed with several outliers observed.

**Visual Representation:**

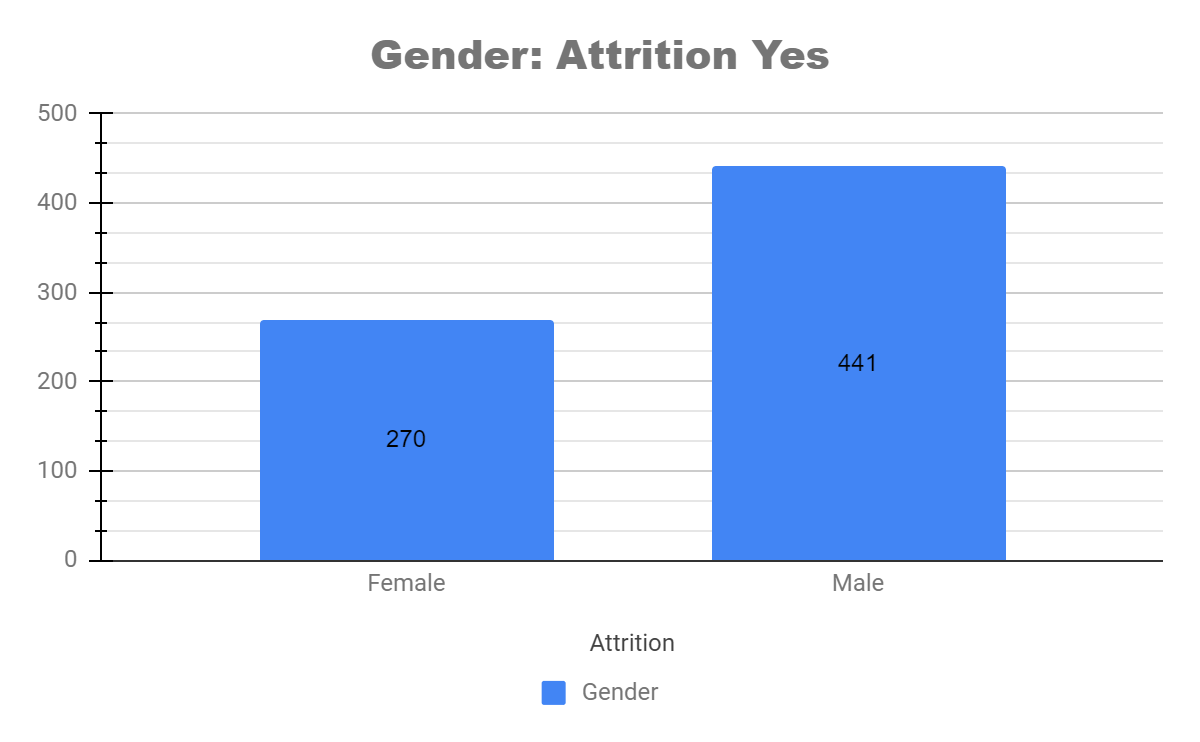
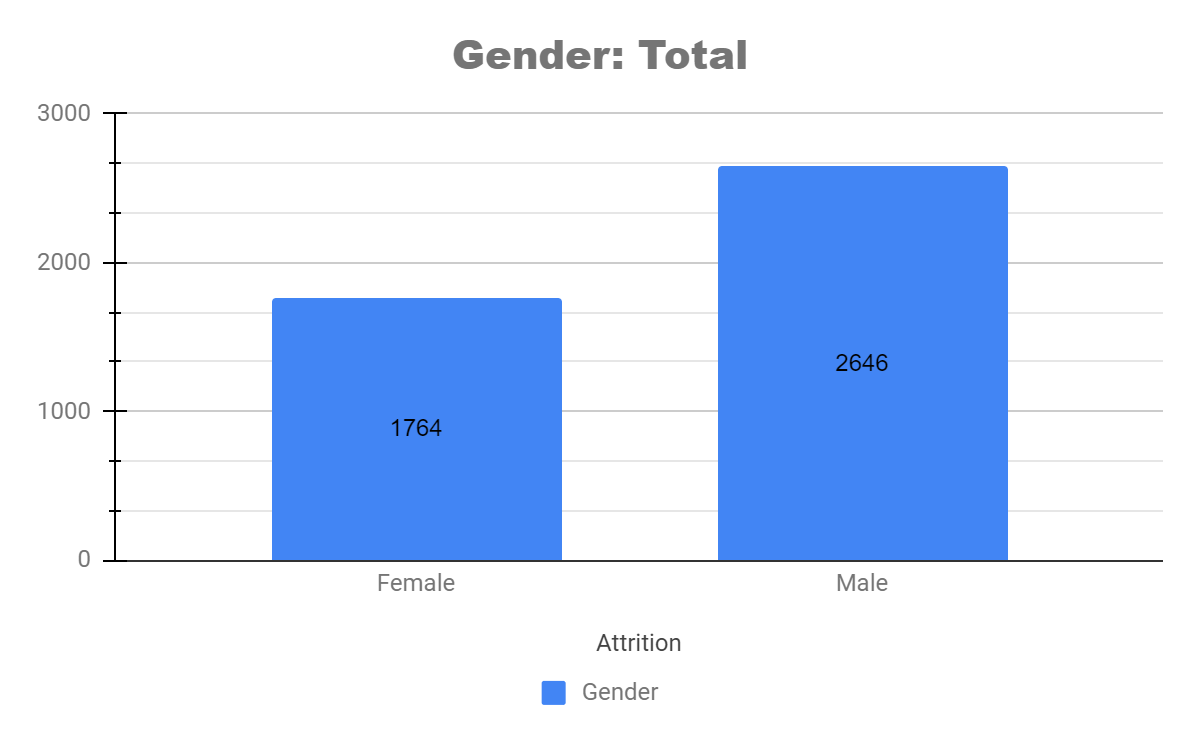
Histogram of Attrition:



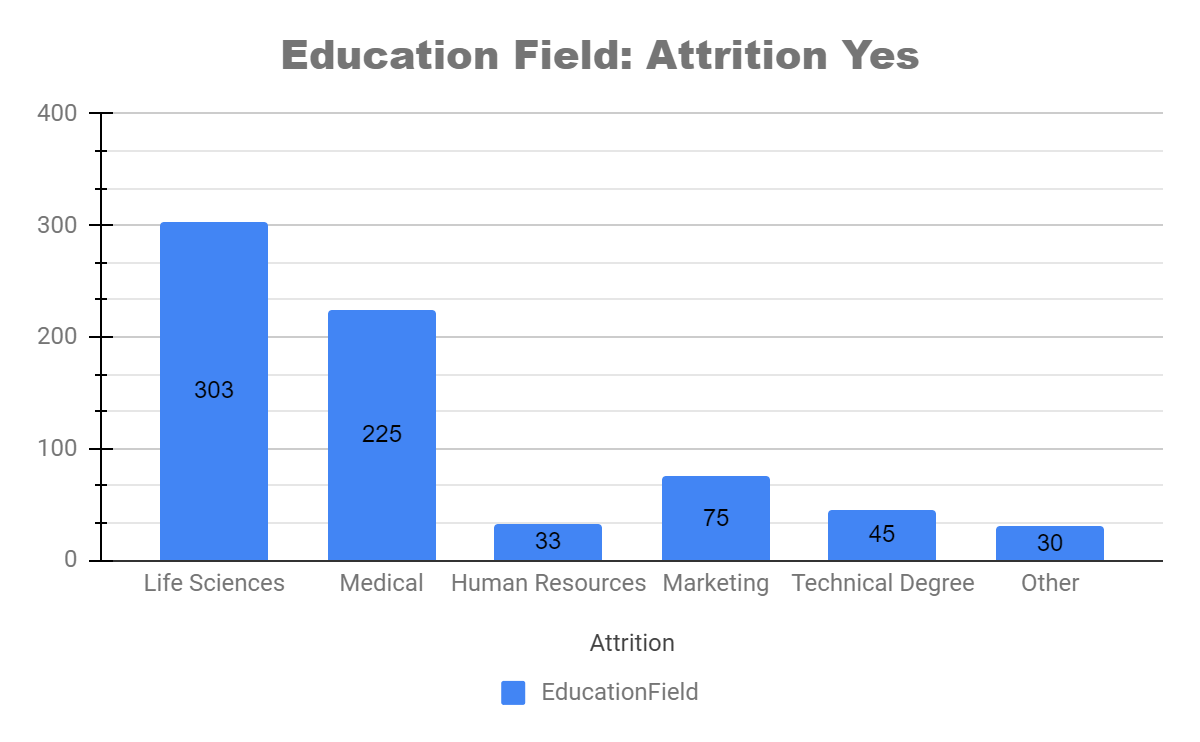
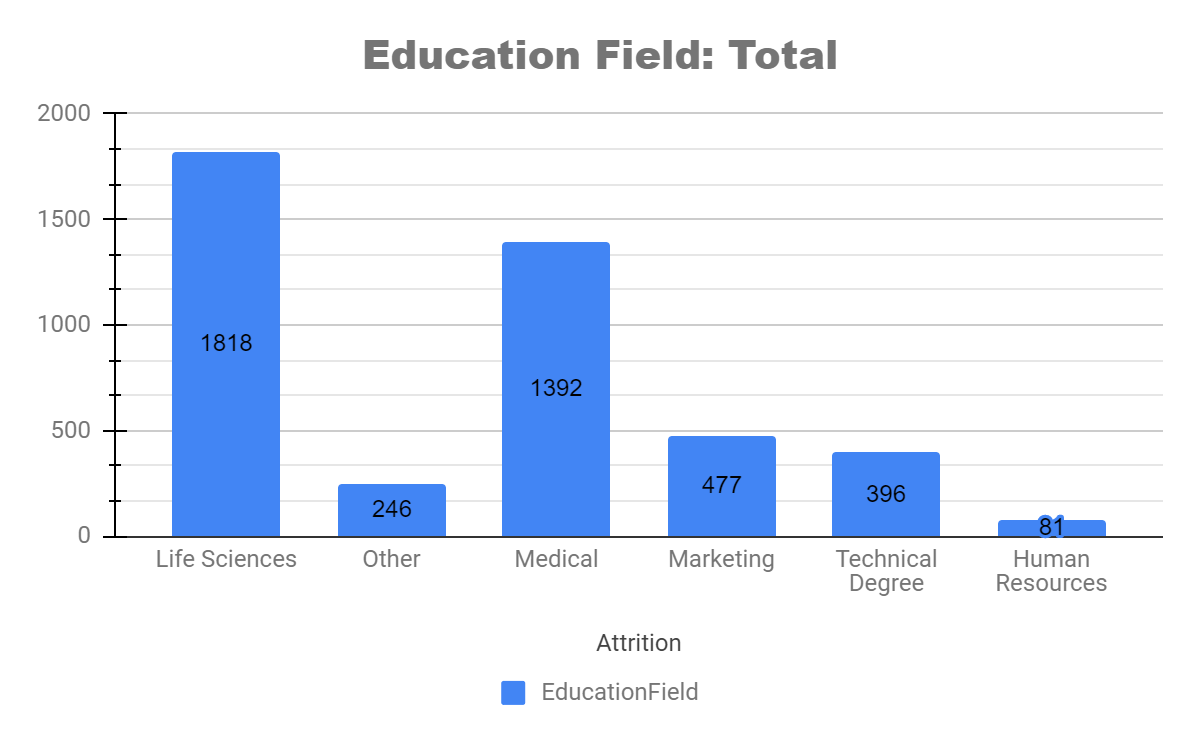
Attrition/Department:



Attrition/Gender:



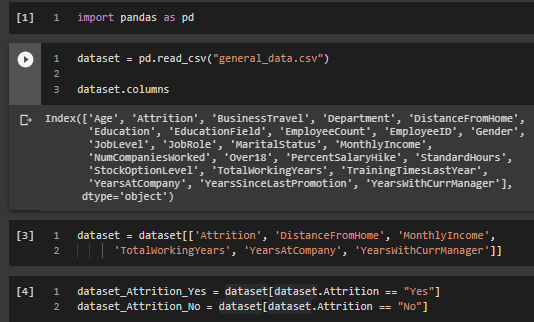
Attrition/Education Field:



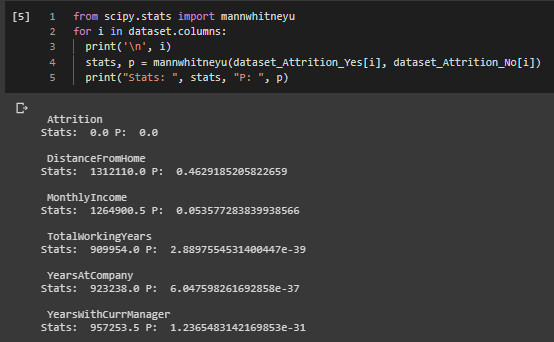
**Statistical tests:**

[Mann-Whitney]

Prerequisites:



Running the tests:



**Distance From Home:**

As p-value(0.4629185205822659) > 0.05:

H0 is accepted, H1 is rejected:

H0: There are no significant differences in the Distance From Home between attrition (Y) and attrition (N)

H1: There are significant differences in the Distance From Home between attrition (Y) and attrition (N)

**Monthly income:**

As p-value(0.053577283839938566) >0.05:

H0 is accepted, H1 is rejected:

H0: There are no significant differences in the Monthly income between attrition (Y) and attrition (N)

H1: There are significant differences in the Monthly income between attrition (Y) and attrition (N)

**Total Working Years:**

As p-value(2.8897554531400447e-39) < 0.05:

H0 is rejected, H1 is accepted:

H0: There are no significant differences in the Total Working Years between attrition (Y) and attrition (N)

H1: There are significant differences in the Total Working Years between attrition (Y) and attrition (N)

**Years At Company:**

As p-value(6.047598261692858e-37) < 0.05:

H0 is rejected, H1 is accepted:

H0: There are no significant differences in the Years At Company between attrition (Y) and attrition (N)

H1: There are significant differences in the Years At Company between attrition (Y) and attrition (N)

**Years With Current Manager:**

As p-value(1.2365483142169853e-31) < 0.05:

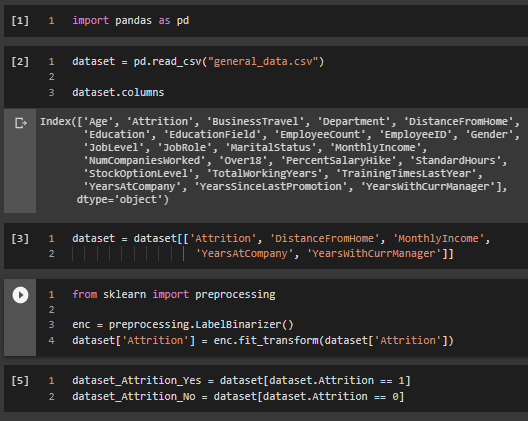
H0 is rejected, H1 is accepted:

H0: There are no significant differences in the Years With Current Manager between attrition (Y) and attrition (N)

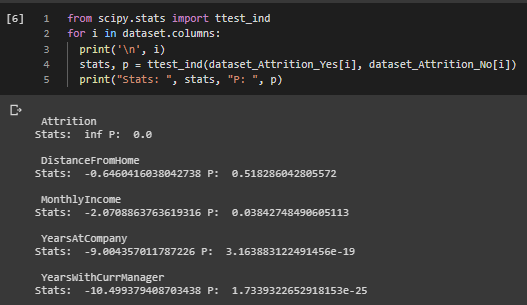
H1: There are significant differences in the Years With Current Manager between attrition (Y) and attrition (N)

[Seprate T-Test]

Prerequisite :



Running The tests:



**Distance From Home:**

As p-value(0.518286042805572) > 0.05:

H0 is accepted, H1 is rejected:

H0: There are no significant differences in the Distance From Home between attrition (Y) and attrition (N)

H1: There are significant differences in the Distance From Home between attrition (Y) and attrition (N)

**Monthly income:**

As p-value(0.03842748490605113) <0.05:

H0 is rejected, H1 is accepted:

H0: There are no significant differences in the Monthly income between attrition (Y) and attrition (N)

H1: There are significant differences in the Monthly income between attrition (Y) and attrition (N)

**Years At Company:**

As p-value(3.163883122491456e-19) <0.05:

H0 is rejected, H1 is accepted:

H0: There are no significant differences in the Years At Company between attrition (Y) and attrition (N)

H1: There are significant differences in the Years At Company between attrition (Y) and attrition (N)

**Years With Current Manager:**

As p-value(1.7339322652918153e-25) < 0.05:

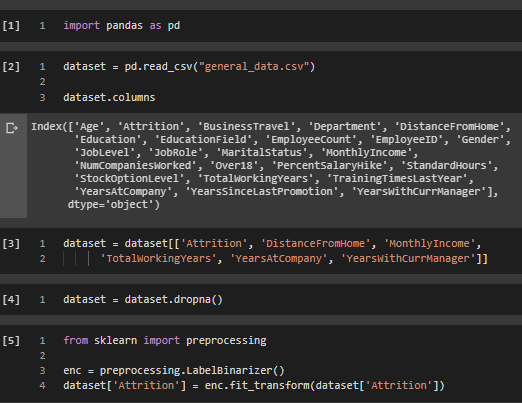
H0 is rejected, H1 is accepted:

H0: There are no significant differences in the Years With Current Manager between attrition (Y) and attrition (N)

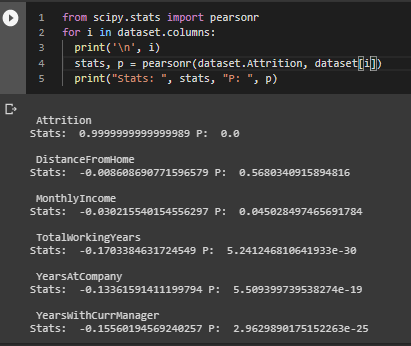
H1: There are significant differences in the Years With Current Manager between attrition (Y) and attrition (N)

[Correlation Analysis]

Prerequisites:



Running the Tests:



**Distance From Home:**

As R = (**-0.0086**), there’s a low negative correlation between Attrition and Distance From Home

As p-value(**0.5680340915894816**) > 0.05:

H0 Accepted, H1 Rejected

H0: There is no significant correlation between Attrition & Distance From Home

H1: There is a significant correlation between Attrition & Distance From Home

**Monthly Income:**

As R =(**-0.03**), there’s a low negative correlation between Attrition and Monthly Income

As p-value(**0.045028497465691784**) < 0.05:

H0 Rejected, H1 Accepted

H0: There is no significant correlation between Attrition & Monthly Income

H1: There is a significant correlation between Attrition & Monthly Income

**Total Working Years:**

As R = (**-0.17**), there’s a medium negative correlation between Attrition and Total Working Years

As p-value(**5.241246810641933e-30**) < 0.05:

H0 Rejected, H1 Accepted

H0: There is no significant correlation between Attrition & Total Working Years

H1: There is a significant correlation between Attrition & Total Working Years

**Years at Company:**

As R =(**-0.1336**), there’s a medium negative correlation between Attrition and Years at Company

As p-value(**5.509399739538274e-19**) < 0.05:

H0 Rejected, H1 Accepted

H0: There is no significant correlation between Attrition & Years at Company

H1: There is a significant correlation between Attrition & Years at Company

**Years with Current Manager:**

As R = (**-0.1556**), there’s a medium negative correlation between Attrition and Years with Current Manager

As p-value(**2.9629890175152263e-25**) < 0.05:

H0 Rejected, H1 Accepted

H0: There is no significant correlation between Attrition & Years with Current Manager

H1: There is a significant correlation between Attrition & Years with Current Manager