import pandas as pd

import matplotlib.pyplot as plt

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

from scipy.stats import pearsonr

dataset.head()

Out[5]:

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.drop\_duplicates()

Out[6]:

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

... ... ... ... ...

4405 42 No ... 0 2

4406 29 No ... 0 2

4407 25 No ... 1 2

4408 42 No ... 7 8

4409 40 No ... 3 9

[4410 rows x 24 columns]

dataset.dropna()

Out[7]:

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

... ... ... ... ...

4404 29 No ... 1 5

4405 42 No ... 0 2

4406 29 No ... 0 2

4407 25 No ... 1 2

4408 42 No ... 7 8

[4382 rows x 24 columns]

dataset.describe()

Out[8]:

Age ... YearsWithCurrManager

count 4410.000000 ... 4410.000000

mean 36.923810 ... 4.123129

std 9.133301 ... 3.567327

min 18.000000 ... 0.000000

25% 30.000000 ... 2.000000

50% 36.000000 ... 3.000000

75% 43.000000 ... 7.000000

max 60.000000 ... 17.000000

[8 rows x 16 columns]

dataset["Attrition"].replace(to\_replace=("No","Yes"),value=(0,1),inplace=True)

dataset["Gender"].replace(to\_replace=("Male","Female"),value=(0,1),inplace=True)

1. **Attrition and Age**

**stats, p=pearsonr(dataset.Attrition, dataset.Age)**

**print (stats, p)**

**-0.15920500686577965 1.996801615886744e-26**

Null : There is no relation between employee age and attrition

Alternate: There is relation between employee age and attrition

**P Less than 0.05 Null Hypothesis rejected**

1. **Attrition and Gender**

**stats, p=pearsonr(dataset.Attrition, dataset.Gender)**

**print (stats, p)**

**-0.01812507887701024 0.22881970951795952**

Null : There is no relation between employee Gender and attrition

Alternate: There is relation between employee ag Gender and attrition

**P greater than 0.05 Null Hypothesis accepted**

1. **Attrition and Job Level**

**stats, p=pearsonr(dataset.Attrition, dataset.JobLevel)**

**print (stats, p)**

**-0.010289713287495035 0.49451717271828405**

Null : There is no relation between employee JobLevel and attrition

Alternate: There is relation between employee Job Level and attrition

**P greater than 0.05 Null Hypothesis accepted**

1. **Attrition and YearsWithCurrManager**

**stats, p=pearsonr(dataset.Attrition, dataset.YearsWithCurrManager)**

**print (stats, p)**

**-0.15619931590162842 1.7339322652900218e-25**

Null : There is no relation between employee YearsWithCurrManager and attrition

Alternate: There is relation between employee Ge YearsWithCurrManager nder and attrition

**P greater than 0.05 Null Hypothesis accepted**

1. **Attrition and DistanceFromHome**

**stats, p=pearsonr(dataset.Attrition, dataset.DistanceFromHome)**

**print (stats, p)**

**-0.00973014101017966 0.5182860428050771**

Null : There is no relation between employee DistanceFromHome and attrition

Alternate: There is relation between employee DistanceFromHome and attrition

**P greater than 0.05 Null Hypothesis accepted**

1. **Attrition and YearsSinceLastPromotion**

**stats, p=pearsonr(dataset.Attrition, dataset.YearsSinceLastPromotion)**

**print (stats, p)**

**-0.03301877514258434 0.028330336189396753**

Null : There is no relation between employee YearsSinceLastPromotion and attrition

Alternate: There is relation between employee YearsSinceLastPromotion and attrition

**P Less than 0.05 Null Hypothesis rejected**

1. **Attrition and MonthlyIncome**

**stats, p=pearsonr(dataset.Attrition, dataset.MonthlyIncome)**

**print (stats, p)**

**-0.031176281698115007 0.03842748490600132**

Null : There is no relation between employee MonthlyIncome and attrition

Alternate: There is relation between employee MonthlyIncome and attrition

**P Less than 0.05 Null Hypothesis rejected**

1. **Attrition and PercentSalaryHike**

**stats, p=pearsonr(dataset.Attrition, dataset.PercentSalaryHike)**

**print (stats, p)**

**0.03253259489105351 0.030743386433355353**

Null : There is no relation between employee PercentSalaryHike and attrition

Alternate: There is relation between employee PercentSalaryHike and attrition

**P Less than 0.05 Null Hypothesis rejected**

1. **Attrition and YearsAtCompany**

**stats, p=pearsonr(dataset.Attrition, dataset.YearsAtCompany)**

**print (stats, p)**

**-0.1343922139899772 3.1638831224877484e-19**

Null : There is no relation between employee YearsAtCompany and attrition

Alternate: There is relation between employee YearsAtCompany and attrition

**P greater than 0.05 Null Hypothesis accepted**

1. **Attrition and TrainingTimesLastYear**

**stats, p=pearsonr(dataset.Attrition, dataset.TrainingTimesLastYear)**

**print (stats, p)**

**-0.04943057624425501 0.0010247061915362814**

Null : There is no relation between employee TotalWorkingYears and attrition

Alternate: There is relation between employee TotalWorkingYears and attrition

**P Less than 0.05 Null Hypothesis rejected**