dataset=pd.read\_excel("general\_data.xlsx",sheet\_name=0)  
  
from scipy.stats import pearsonr  
  
from sklearn.preprocessing import LabelEncoder  
  
le = LabelEncoder()  
  
dataset['Attrition'] = le.fit\_transform(dataset['Attrition'])  
  
dataset.head()  
Out[23]:  
   Age  Attrition  ... YearsSinceLastPromotion YearsWithCurrManager  
0   51          0  ...                       0                    0  
1   31          1  ...                       1                    4  
2   32          0  ...                       0                    3  
3   38          0  ...                       7                    5  
4   32          0  ...                       0                    4  
  
[5 rows x 24 columns]  
  
data\_yes = dataset[dataset["Attrition"] == 1]  
  
data\_no = dataset[dataset["Attrition"] == 0]  
  
dataset.head()  
Out[26]:  
   Age  Attrition  ... YearsSinceLastPromotion YearsWithCurrManager  
0   51          0  ...                       0                    0  
1   31          1  ...                       1                    4  
2   32          0  ...                       0                    3  
3   38          0  ...                       7                    5  
4   32          0  ...                       0                    4  
  
[5 rows x 24 columns]  
  
dataset.columns  
Out[27]:  
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')

Attrition & DistanceFromHome  
  
stats,p=pearsonr(dataset.Attrition,dataset.DistanceFromHome)  
  
print(stats,p)  
-0.009730141010179674 0.5182860428050771

As r = -0.009, there’s low negative correlation between Attrition and DistanceFromHome

As the P value of 0.518 is > 0.05, we are accepting H0 and hence there’s no significant correlation between Attrition & DistanceFromHome

Attrition & Monthly Income

stats, p=pearsonr(dataset.Attrition, dataset.MonthlyIncome)  
  
print(stats,p)  
-0.031176281698115007 0.03842748490600132

As r = -0.031, there’s low negative correlation between Attrition and MonthlyIncome

As the P value of 0.038 is < 0.05, we are accepting Ha and hence there’s significant correlation between Attrition & MonthlyIncome

Attrition & YearsAtCompany

stats, p=pearsonr(dataset.Attrition, dataset.YearsAtCompany)  
  
print(stats,p)  
-0.1343922139899772 3.1638831224877484e-19

As r = -0.1343, there’s low negative correlation between Attrition and YearsAtCompany

As the P value is < 0.05, we are accepting Ha and hence there’s significant correlation between Attrition & YearsAtCompany

Attrition & YearsWithCurrManager

stats, p=pearsonr(dataset.Attrition, dataset.YearsWithCurrManager)  
  
print(stats,p)  
-0.15619931590162847 1.7339322652896276e-25

As r = -0.1561, there’s low negative correlation between Attrition and YearsWithCurrManager

As the P value is < 0.05, we are accepting Ha and hence there’s significant correlation between Attrition & YearsWithCurrManager