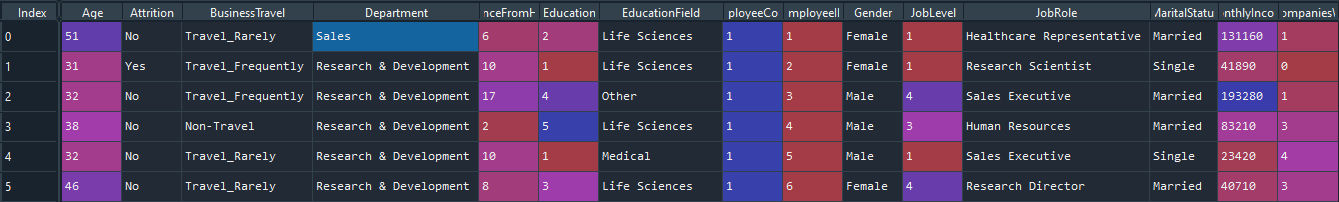
import pandas as pd

import numpy as np

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

dataset.head()

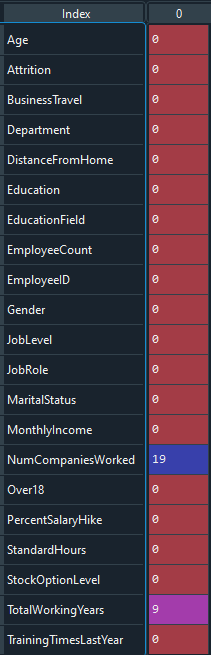


**##### import the package and load the dataset**

# check null values

dataset\_null=dataset.isnull().sum()

dataset = dataset.fillna(dataset[['NumCompaniesWorked','TotalWorkingYears']].mean())



**##### Check the null values and replace by mean**

**## drop unwanted other features**

df = dataset.drop(columns=['EmployeeCount', 'EmployeeID','Over18','StandardHours'])

**Label Encoding technique used to change values**

from sklearn import preprocessing

le = preprocessing.LabelEncoder()

df['Attrition'] = le.fit\_transform(df['Attrition'])

df['BusinessTravel'] = le.fit\_transform(df['BusinessTravel'])

df['Department'] = le.fit\_transform(df['Department'])

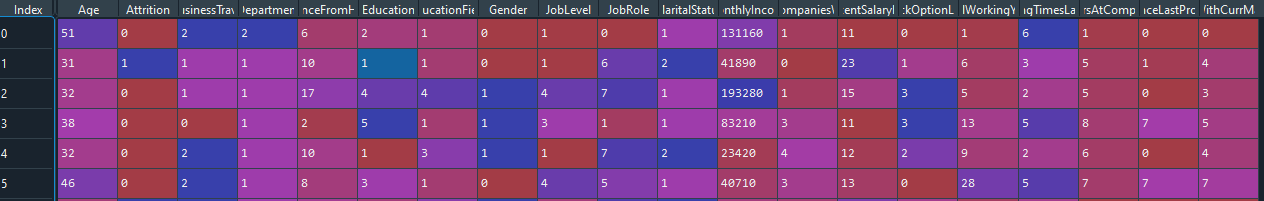
df['EducationField'] = le.fit\_transform(df['EducationField'])

df['Gender'] = le.fit\_transform(df['Gender'])

df['JobRole'] = le.fit\_transform(df['JobRole'])

df['MaritalStatus'] = le.fit\_transform(df['MaritalStatus'])

df.head()



**##### To change label encoding value to number numbers**

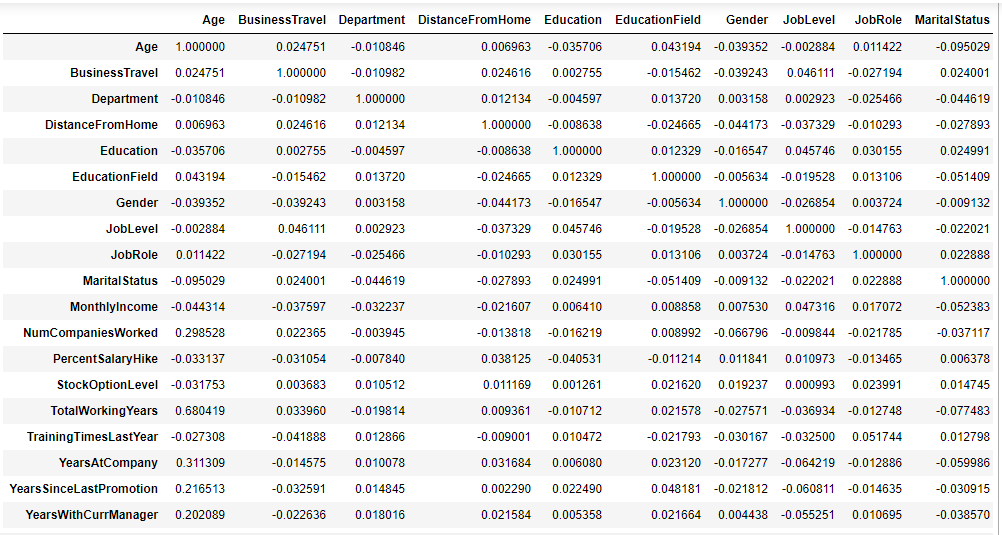
**Variable Selection**

y = df['Attrition']

X = df.drop(columns=['Attrition'])

**##### X is independent variable and y is dependent variable**

**MULTICOLLINEARITY**

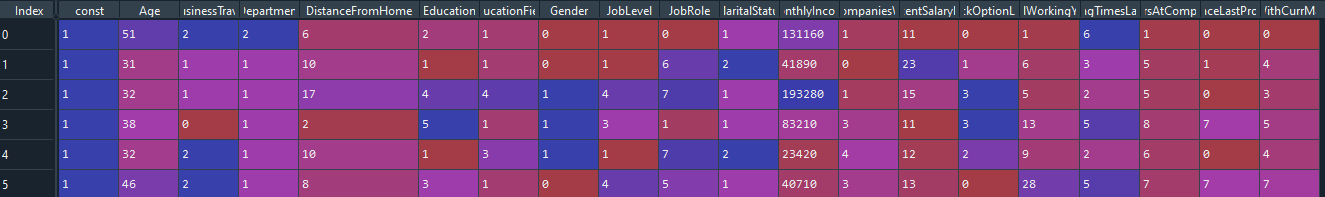


**##### To satisfy multicollinearity test is pass. There is no zero correlation in independent variables.**

**Logistic Regression**

import statsmodels.api as sm

X1 = sm.add\_constant(X)



**##### add the independent variable in Constants**

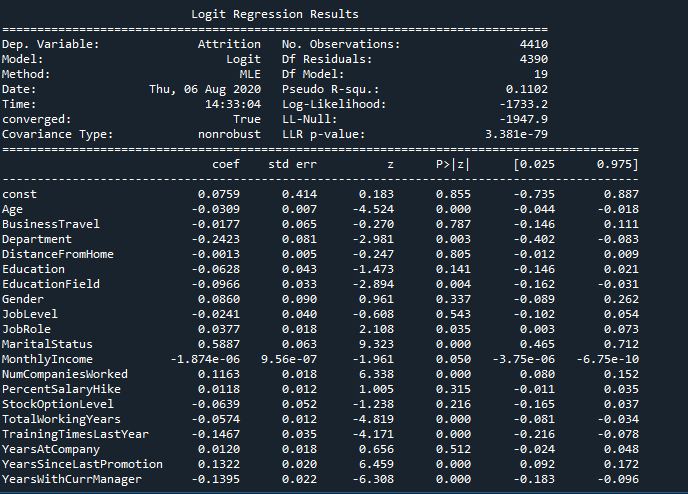
Logistic = sm.Logit(y,X1)

result = Logistic.fit()

**##### Logistic Equation p = 1/(1+e\*\* -(B0+B1X1+….+BnXn))**

**##### Logit is function name and y = dependent variable, X1 – independent variable alone with Constant**

result.summary()



**##### Inference statement**

**Here P values based (p>0.05) Except ” BusinessTravel”,”DistanceFromHome”,”Education “,”Gender”, ”Job Level”, “PercentSalaryHike”,”StockOptionlevel”,”YearAtCompany” are greater than 0.05**

**So that other remaining “Department”,”EducationField”,”Job Role”,”Montly Income” are significantly impact on the Attritions.**

**Most significant impact factor Attritions are “Age”, ”Marital Status”, ”NumCompaniesWorked”, ”TotalWorkingyears”, “TrainingTimesLastYear”, “YearSinceLastProgram”, “YearswithCurrManager”.**