**Attrition Project**

**import pandas as pd**

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

**from sklearn.preprocessing import LabelEncoder**

**le=LabelEncoder()**

**dataset["NumCompaniesWorked"]=le.fit\_transform(dataset["NumCompaniesWorked"])**

**dataset["TotalWorkingYears"]=le.fit\_transform(dataset["TotalWorkingYears"])**

**dataset["Department"]=le.fit\_transform(dataset["Department"])**

**dataset["BusinessTravel"]=le.fit\_transform(dataset["BusinessTravel"])**

**dataset["Attrition"]=le.fit\_transform(dataset["Attrition"])**

**dataset["EducationField"]=le.fit\_transform(dataset["EducationField"])**

**dataset["Gender"]=le.fit\_transform(dataset["Gender"])**

**dataset["Over18"]=le.fit\_transform(dataset["Over18"])**

**dataset["JobRole"]=le.fit\_transform(dataset["JobRole"])**

**dataset["MaritalStatus"]=le.fit\_transform(dataset["MaritalStatus"])**

**dataset.head()**

Out[201]:

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]

**import statsmodels.api as sm**

**Y=dataset.Attrition**

**X=dataset[['Age','BusinessTravel','Department','DistanceFromHome','Education','EducationField','Gender','NumCompaniesWorked','PercentSalaryHike','StockOptionLevel','TotalWorkingYears','TrainingTimesLastYear','YearsAtCompany','YearsSinceLastPromotion','YearsWithCurrManager','JobLevel','JobRole','MaritalStatus','MonthlyIncome']]**

**X1= sm.add\_constant(X)**

**X1= sm.add\_constant(X)**

**logistic=sm.Logit(Y,X1)**

**result=logistic.fit()**

Optimization terminated successfully.

Current function value: 0.394467

Iterations 7

**result.summary()**

Out[218]:

<class 'statsmodels.iolib.summary.Summary'>

"""

Logit Regression Results

==============================================================================

Dep. Variable: Attrition No. Observations: 4410

Model: Logit Df Residuals: 4390

Method: MLE Df Model: 19

Date: Sat, 08 Aug 2020 Pseudo R-squ.: 0.1069

Time: 21:36:40 Log-Likelihood: -1739.6

converged: True LL-Null: -1947.9

Covariance Type: nonrobust LLR p-value: 1.576e-76

===========================================================================================

coef std err z P>|z| [0.025 0.975]

-------------------------------------------------------------------------------------------

const 0.1219 0.413 0.295 0.768 -0.687 0.931

Age -0.0312 0.007 -4.593 0.000 -0.045 -0.018

BusinessTravel -0.0217 0.065 -0.332 0.740 -0.150 0.106

Department -0.2385 0.081 -2.943 0.003 -0.397 -0.080

DistanceFromHome -0.0011 0.005 -0.206 0.837 -0.012 0.009

Education -0.0635 0.043 -1.495 0.135 -0.147 0.020

EducationField -0.0996 0.033 -2.991 0.003 -0.165 -0.034

Gender 0.0759 0.089 0.850 0.395 -0.099 0.251

NumCompaniesWorked 0.0866 0.016 5.422 0.000 0.055 0.118

PercentSalaryHike 0.0130 0.012 1.108 0.268 -0.010 0.036

StockOptionLevel -0.0630 0.052 -1.221 0.222 -0.164 0.038

TotalWorkingYears -0.0458 0.011 -4.082 0.000 -0.068 -0.024

TrainingTimesLastYear -0.1453 0.035 -4.132 0.000 -0.214 -0.076

YearsAtCompany 0.0012 0.018 0.069 0.945 -0.034 0.036

YearsSinceLastPromotion 0.1330 0.020 6.514 0.000 0.093 0.173

YearsWithCurrManager -0.1391 0.022 -6.306 0.000 -0.182 -0.096

JobLevel -0.0279 0.040 -0.704 0.481 -0.105 0.050

JobRole 0.0365 0.018 2.044 0.041 0.002 0.072

MaritalStatus 0.5886 0.063 9.332 0.000 0.465 0.712

MonthlyIncome -1.855e-06 9.56e-07 -1.939 0.052 -3.73e-06 1.96e-08

===========================================================================================

**"""**

**Inference:**

**1.The more significant Variables are:**

**Age, Marital Status, NumCompaniesWorked, TotalWorkingYears, TrainingTimesLastYear ,YearsSinceLastPromotion ,YearsWithCurrManager**

**2.The less significant Variables are:**

**Gender, Department, EducationField, JobRole, Monthly, income, Stock option level, Job level ,Percentage salary hike**

**3.Not significant Variables are:**

**Business Travel ,Distance From Home ,years at company**