

# Logistic Regression on Attrition Data set.

```
import pandas as pd
```

```
import statsmodels.api as sm
```

```
dataset=pd.read_csv("Attrition Rate Analysis.csv")
```

```
dataset.columns
```

```
Out[1]:
```

```
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')
```

```
from sklearn import preprocessing
```

```
le=preprocessing.LabelEncoder()
```

```
dataset["Attrition"]=le.fit_transform(dataset["Attrition"])
```

```
dataset["BusinessTravel"]=le.fit_transform(dataset["BusinessTravel"])
```

```
dataset["Department"]=le.fit_transform(dataset["Department"])
```

```
dataset["EducationField"]=le.fit_transform(dataset["EducationField"])
```

```
dataset["Gender"]=le.fit_transform(dataset["Gender"])
```

```
dataset["MaritalStatus"]=le.fit_transform(dataset["MaritalStatus"])
```

```
dataset["JobRole"]=le.fit_transform(dataset["JobRole"])
```

```
dataset1=dataset.drop(['EmployeeCount', 'EmployeeID', 'Over18', 'StandardHours'],axis=1)
```

```
dataset1.columns
```

```
Out[3]:
```

```
Index(['Age', 'Attrition', 'BusinessTravel', 'Department', 'DistanceFromHome',  
      'Education', 'EducationField', 'Gender', 'JobLevel', 'JobRole',
```

```
'MaritalStatus', 'MonthlyIncome', 'NumCompaniesWorked',  
'PercentSalaryHike', 'StockOptionLevel', 'TotalWorkingYears',  
'TrainingTimesLastYear', 'YearsAtCompany', 'YearsSinceLastPromotion',  
'YearsWithCurrManager'],  
dtype='object')
```

```
dataset2=dataset1.dropna()
```

```
dataset3=dataset2.drop_duplicates()
```

```
Y=dataset3.Attrition
```

```
dataset3.columns
```

```
Out[7]:
```

```
Index(['Age', 'Attrition', 'BusinessTravel', 'Department', 'DistanceFromHome',  
      'Education', 'EducationField', 'Gender', 'JobLevel', 'JobRole',  
      'MaritalStatus', 'MonthlyIncome', 'NumCompaniesWorked',  
      'PercentSalaryHike', 'StockOptionLevel', 'TotalWorkingYears',  
      'TrainingTimesLastYear', 'YearsAtCompany', 'YearsSinceLastPromotion',  
      'YearsWithCurrManager'],  
      dtype='object')
```

```
X=dataset3[['Age', 'BusinessTravel', 'Department', 'DistanceFromHome',  
            'Education', 'EducationField', 'Gender', 'JobLevel', 'JobRole',  
            'MaritalStatus', 'MonthlyIncome', 'NumCompaniesWorked',  
            'PercentSalaryHike', 'StockOptionLevel', 'TotalWorkingYears',  
            'TrainingTimesLastYear', 'YearsAtCompany', 'YearsSinceLastPromotion',  
            'YearsWithCurrManager']]
```

```
X1=sm.add_constant(X)
```

```
Logistic_Attrition=sm.Logit(Y,X1)
```

```
result=Logistic_Attrition.fit()
```

Optimization terminated successfully.

Current function value: 0.392756

Iterations 7

```
result.summary()
```

```
Out[13]:
```

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

```
''''
```

#### Logit Regression Results

```
=====
```

Dep. Variable: Attrition No. Observations: 1470

Model: Logit Df Residuals: 1450

Method: MLE Df Model: 19

Date: Wed, 12 Aug 2020 Pseudo R-squ.: 0.1108

Time: 11:53:39 Log-Likelihood: -577.35

converged: True LL-Null: -649.29

Covariance Type: nonrobust LLR p-value: 3.295e-21

```
=====
```

```
=====
```

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

```
-----
```

const	0.0650	0.717	0.091	0.928	-1.340	1.470
-------	--------	-------	-------	-------	--------	-------

Age	-0.0306	0.012	-2.583	0.010	-0.054	-0.007
-----	---------	-------	--------	-------	--------	--------

BusinessTravel	-0.0166	0.113	-0.146	0.884	-0.239	0.206
----------------	---------	-------	--------	-------	--------	-------

Department	-0.2421	0.141	-1.720	0.085	-0.518	0.034
------------	---------	-------	--------	-------	--------	-------

DistanceFromHome	-0.0014	0.009	-0.145	0.884	-0.020	0.017
------------------	---------	-------	--------	-------	--------	-------

Education	-0.0625	0.074	-0.847	0.397	-0.207	0.082
-----------	---------	-------	--------	-------	--------	-------

EducationField	-0.0965	0.058	-1.669	0.095	-0.210	0.017
----------------	---------	-------	--------	-------	--------	-------

Gender	0.0869	0.155	0.560	0.576	-0.217	0.391
--------	--------	-------	-------	-------	--------	-------

JobLevel	-0.0249	0.069	-0.363	0.717	-0.159	0.110
JobRole	0.0378	0.031	1.219	0.223	-0.023	0.099
MaritalStatus	0.5885	0.109	5.379	0.000	0.374	0.803
MonthlyIncome	-1.868e-06	1.66e-06	-1.128	0.259	-5.11e-06	1.38e-06
NumCompaniesWorked	0.1184	0.032	3.729	0.000	0.056	0.181
PercentSalaryHike	0.0117	0.020	0.576	0.565	-0.028	0.052
StockOptionLevel	-0.0645	0.089	-0.721	0.471	-0.240	0.111
TotalWorkingYears	-0.0593	0.021	-2.856	0.004	-0.100	-0.019
TrainingTimesLastYear	-0.1465	0.061	-2.406	0.016	-0.266	-0.027
YearsAtCompany	0.0136	0.032	0.428	0.669	-0.049	0.076
YearsSinceLastPromotion	0.1323	0.035	3.732	0.000	0.063	0.202
YearsWithCurrManager	-0.1396	0.038	-3.642	0.000	-0.215	-0.064

=====

=====

!!!!

**The Variables Age, Marital Status, NumcompaniesWorked, Totalworkingyears, TrainingTimesLastYear , YearsSinceLastPromotion , YearsWithCurrManager are significantly important for the Attrition rate in the company**