LOGISTIC REGRESSION RESUT FOR ATTRITION DATASET:

STEPS:

1. Import pandas package.

2. Loading the dataset.

3. Cleaning the dataset.

```
In [38]: import os
         os.chdir(r"C:\Users\User\Downloads")
         import pandas as pd
         dataset=pd.read_excel("general_data.xlsx",sheet_name=0)
         dataset.head()
         dataset=dataset.dropna()
         dataset.isna().sum()

Out[38]: Age                        0
         Attrition                  0
         BusinessTravel             0
         Department                 0
         DistanceFromHome           0
         Education                  0
         EducationField             0
         EmployeeCount              0
         EmployeeID                 0
         Gender                     0
         JobLevel                   0
         JobRole                    0
         MaritalStatus              0
         MonthlyIncome              0
         NumCompaniesWorked         0
         Over18                     0
         PercentSalaryHike          0
         StandardHours              0
         StockOptionLevel           0
         TotalWorkingYears          0
         TrainingTimesLastYear      0
         YearsAtCompany             0
         YearsSinceLastPromotion    0
         YearsWithCurrManager       0
         dtype: int64
```

4. Assigning dependent variable as “attrition” and all other significant columns as independent variables.

5. Import statistical model and adding constants to dependent variable.

   import statsmodels.api as sm
   x1=sm.add_constant(x)

6. Applying logistic regression formula using “Logit” method and optimizing the result.

   log=sm.Logit(y,x1)
   result=log.fit()

7. Result:

```
In [47]: result.summary()
```

Out[47]:

Logit Regression Results

Dep. Variable:	Attrition	**No. Observations:**	4382
Model:	Logit	**Df Residuals:**	4371
Method:	MLE	**Df Model:**	10
Date:	Fri, 07 Aug 2020	**Pseudo R-squ.:**	0.06926
Time:	22:40:39	**Log-Likelihood:**	-1799.2
converged:	True	**LL-Null:**	-1933.1
Covariance Type:	nonrobust	**LLR p-value:**	9.866e-52

	coef	std err	z	P>|z|	[0.025	0.975]
Age	-0.0297	0.007	-4.457	0.000	-0.043	-0.017
Education	-0.0562	0.041	-1.357	0.175	-0.137	0.025
EmployeeCount	0.5747	0.332	1.730	0.084	-0.076	1.226
JobLevel	-0.0404	0.039	-1.038	0.299	-0.117	0.036
MonthlyIncome	-2.34e-06	9.41e-07	-2.487	0.013	-4.18e-06	-4.96e-07
PercentSalaryHike	0.0165	0.011	1.441	0.150	-0.006	0.039
TotalWorkingYears	-0.0365	0.011	-3.392	0.001	-0.058	-0.015
TrainingTimesLastYear	-0.1265	0.034	-3.667	0.000	-0.194	-0.059
YearsAtCompany	-0.0213	0.017	-1.232	0.218	-0.055	0.013
YearsSinceLastPromotion	0.1258	0.020	6.262	0.000	0.086	0.165
YearsWithCurrManager	-0.1284	0.021	-5.988	0.000	-0.170	-0.086

As we can observe the p value for the following is less than 0.05 which means they are significant attributes.

1. Age
2. Monthly income
3. Total working years
4. Training times last year
5. Years since last promotion
6. Years with current manager