Logistic Regression on Bank Loan Data Set

Loading Data and Data Treatment:

```
import pandas as pd

loan_data = pd.read_excel("Bank_Personal_Loan_Modelling.xlsx", sheet_name= "Data")

loan_data.head(2)
```

Out[83]:

```
   ID  Age  Experience  ...  CD Account  Online  CreditCard
0   1   25           1  ...           0       0           0
1   2   45          19  ...           0       0           0

[2 rows x 14 columns]

loan_data.isnull().sum()
```

Out[85]:

```
ID                   0
Age                  0
Experience           0
Income               0
ZIP Code             0
Family               0
CCAvg                0
Education            0
Mortgage             0
Personal Loan        0
Securities Account   0
CD Account           0
Online               0
CreditCard           0
```

```
dtype: int64


loan_data.columns
Out[87]:
Index(['ID', 'Age', 'Experience', 'Income', 'ZIP Code', 'Family', 'CCAvg',
       'Education', 'Mortgage', 'Personal Loan', 'Securities Account',
       'CD Account', 'Online', 'CreditCard'],
      dtype='object')
```

Model Generation:

```
Y = loan_data[['Personal Loan']]


X = loan_data[['Age', 'Experience', 'Income', 'Family', 'CCAvg',
       'Education', 'Mortgage', 'Securities Account',
       'CD Account', 'Online', 'CreditCard']]


Corr_matrix = X.corr()


import statsmodels.api as sm


X1 = sm.add_constant(X)


Logistic = sm.Logit(Y, X1)


result = Logistic.fit()
Optimization terminated successfully.
         Current function value: 0.128435
         Iterations 9
```

result.summary()

Out[95]:

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

```
                           Logit Regression Results
==============================================================================
Dep. Variable:          Personal Loan   No. Observations:                 5000
Model:                          Logit   Df Residuals:                     4988
Method:                           MLE   Df Model:                           11
Date:                Sun, 09 Aug 2020   Pseudo R-squ.:                 0.5938
Time:                        14:13:24   Log-Likelihood:                -642.18
converged:                       True   LL-Null:                       -1581.0
Covariance Type:            nonrobust   LLR p-value:                     0.000
==============================================================================
                     coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const                -12.1928      1.645     -7.411      0.000     -15.417      -8.968
Age                   -0.0536      0.061     -0.874      0.382      -0.174       0.067
Experience             0.0638      0.061      1.046      0.295      -0.056       0.183
Income                 0.0546      0.003     20.831      0.000       0.049       0.060
Family                 0.6958      0.074      9.364      0.000       0.550       0.841
CCAvg                  0.1240      0.040      3.127      0.002       0.046       0.202
Education              1.7362      0.115     15.088      0.000       1.511       1.962
Mortgage               0.0005      0.001      0.856      0.392      -0.001       0.002
Securities Account    -0.9368      0.286     -3.277      0.001      -1.497      -0.377
CD Account             3.8225      0.324     11.800      0.000       3.188       4.457
Online                -0.6752      0.157     -4.298      0.000      -0.983      -0.367
CreditCard            -1.1197      0.205     -5.462      0.000      -1.522      -0.718
==============================================================================
```

Inference:

Features Income, Family, CCAvg, Education, Securities Account, CD Account, Online, Credit are significantly important in getting the loan.

The probability for getting loan P(Y/N) is calculated as follows

P(Y/N) = 1/(1+e^(-k))

Where k = -12.1928 + (-0.0536)(Age) + (0.0638)('Experience') + (0.0546)('Income') + (0.6958)('Family') + (0.1240)('CCAvg') + (1.7362)('Education') + (0.0005)('Mortgage') + (-0.9368)('Securities Account') + (3.8225)('CD Account') + (-0.6752)('Online') + (-1.1197)('CreditCard')