Predict คนลาออกจากงาน

Absurdism ภูมิภัทร พินธิระ 6410301032 อนันต์โชติ ธนโชคธัญสิริ 6410301033

Problem

- พนักงานเป็นฟันเฟืองหนึ่งขององค์กร
- การลาออกของพนักงานส่งผลต่อการทำงาน, ประสิทธิภาพขององค์กร
- ฉะนั้นแต่ละองค์กรควรที่จะรับรู้ว่าปัจจัย (factor) ใดที่จะ ส่งผลต่อการลาออกของพนักงาน เพื่อที่จะได้เตรียมการ ในการแก้ไขปัญหา

Goal

• ใช้เพื่อวิเคราะห์ว่าปัจจัยใดที่ส่งผลต่อการลาออกของพนักงาน

Data Set

Employee Attrition Classification Dataset 47 **New Notebook** Discussion (1) Suggestions (1) Data Card Code (30) # Years at Company = # Age △ Gender △ Job Role # Mol gender Years at Company Job Role Month 54% Male 26% Technology 46% 23% Female Healthcare Other (7653) 51% 18 74.5k 52685 36 Male 13 Healthcare 8029 7 30585 35 Male Education 4563 50 54656 Male Education 5583 33442 58 Male 44 Media 5525 15667 39 Male 24 Education 4604 3496 45 Female 30 Healthcare 8104 46775 22 5 Healthcare Female 8700

https://www.kaggle.com/datasets/stealthtechnologies/employee-attrition-dataset/data

Data Understanding

```
RangeIndex: 74498 entries, 0 to 74497
Data columns (total 23 columns):
    Column
                              Non-Null Count Dtype
                              74498 non-null int64
    Age
    Gender
                              74498 non-null object
    Years at Company
                              74498 non-null int64
    Job Role
                              74498 non-null object
    Monthly Income
                              74498 non-null int64
    Work-Life Balance
                              74498 non-null object
    Job Satisfaction
                              74498 non-null object
    Performance Rating
                              74498 non-null object
    Number of Promotions
                             74498 non-null int64
    Overtime
                              74498 non-null object
    Distance from Home
                              74498 non-null int64
 11 Education Level
                              74498 non-null object
 12 Marital Status
                             74498 non-null object
    Number of Dependents
                             74498 non-null int64
    Job Level
                              74498 non-null object
    Company Size
                              74498 non-null object
    Company Tenure
                              74498 non-null int64
    Remote Work
                              74498 non-null object
    Leadership Opportunities 74498 non-null object
    Innovation Opportunities 74498 non-null object
    Company Reputation
                              74498 non-null object
 21 Employee Recognition
                              74498 non-null object
 22 Attrition
                              74498 non-null object
dtypes: int64(7), object(16)
```

data['Attrition'].value_counts()

✓ 0.0s

... Attrition
Stayed 39128
Left 35370
Name: count, dtype: int64

data['Age'].describe()

✓ 0.0s

```
74498.000000
count
           38.529746
mean
std
           12.083456
min
           18.000000
25%
           28.000000
50%
           39.000000
75%
           49.000000
         59.000000
max
Name: Age, dtype: float64
```

Name: count, dtype: int64

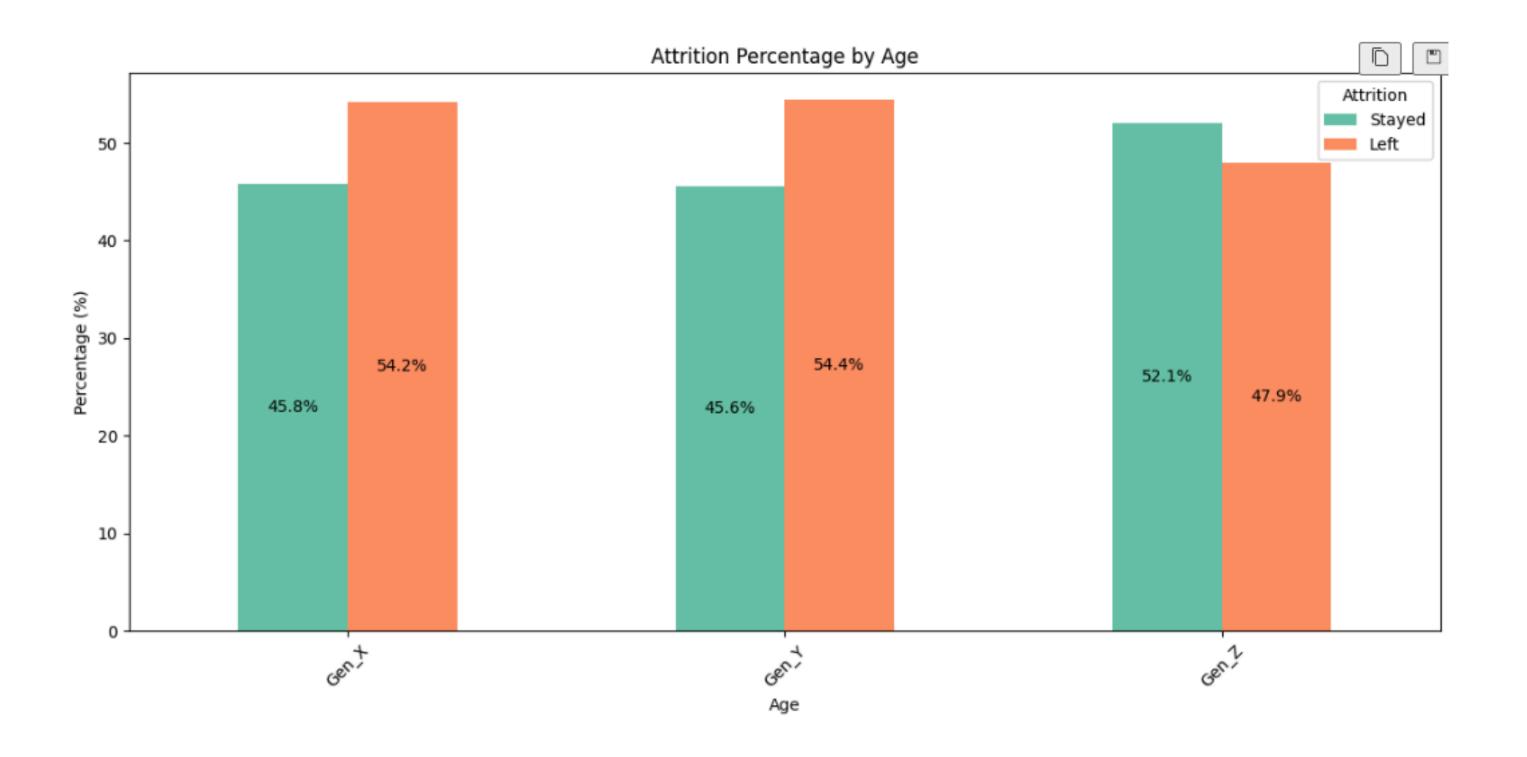
```
D ~
        data['Years at Company'].describe()
      ✓ 0.0s
[10]
              74498.000000
     count
---
                 15.721603
     mean
     std
                 11.223744
     min
                  1.000000
     25%
                  7.000000
     50%
                 13.000000
     75%
                 23.000000
                 51.000000
     max
     Name: Years at Company, dtype: float64
```

```
data['Years at Company'].value_counts()
    ✓ 0.0s
[11]
    Years at Company
          3084
          3056
          3039
          3015
    8
    10
          2987
    9
          2965
          2961
    3
          2952
    6
          2933
    4
          2903
    11
          2893
    12
          2758
    13
          2547
    14
          2349
    15
          2281
    16
          2145
          1994
    17
    18
          1889
    19
          1754
    20
          1729
          1606
    21
    23
          1557
    22
          1537
    24
          1385
```

```
data['Job Role'].value_counts()
 ✓ 0.0s
Job Role
Technology
             19322
Healthcare
             17074
Education
             15658
Media
             11996
Finance
             10448
Name: count, dtype: int64
```

Feature Engineering

```
def classify_birth_year_group(gen):
    birth_year = 2024 - gen
    if birth_year >= 2013:
        return 'Gen_Alpha'
    elif 1995 <= birth_year <= 2012:
        return 'Gen_Z'
    elif 1980 <= birth_year <= 1994:
        return 'Gen_Y'
    elif 1965 <= birth_year <= 1979:
        return 'Gen_X'
    else:
        return 'Baby_Boomer'
  data['Generation'] = data['Age'].apply(classify_birth_year_group)
  data
  # X11 = Prepro_Data[['Generation']]
✓ 0.1s
                                                     Work-
```



Data Preparation

```
from sklearn.preprocessing import StandardScaler, OneHotEncoder, LabelEncoder
continuous_cols = ['Age', 'Years at Company', 'Monthly Income', 'Number of Promotions', 'Distance from Home', 'Company Tenure']
scaler = StandardScaler()
data[continuous_cols] = scaler.fit_transform(data[continuous_cols])
Python
Python
```

```
(count
         7.449800e+04
         2.114517e-16
mean
         1.000007e+00
std
        -1.699008e+00
min
25%
        -8.714242e-01
         3.891746e-02
50%
         8.665008e-01
75%
         1.694084e+00
max
Name: Age, dtype: float64,
Age
             1875
-0.374874
             1861
 0.121676
             1843
 1.197534
 0.535467
             1842
-0.043841
             1834
             1824
 1.363051
             1822
 0.369951
-0.705908
             1818
             1813
 0.038917
             1806
 0.700984
-0.292116
             1806
-0.209358
             1795
 0.618226
             1793
-0.457633
             1791
             1789
-1.616249
-1.285216
             1711
             1708
 1.114776
 0.783742
             1702
-1.699008
             1702
```

```
label_encoder = LabelEncoder()
   data['Attrition'] = label_encoder.fit_transform(data['Attrition'])
   data['Attrition'].value_counts()
Attrition
    39128
    35370
Name: count, dtype: int64
```

1 คือ อยู่ต่อ 0 คือ ลาออก

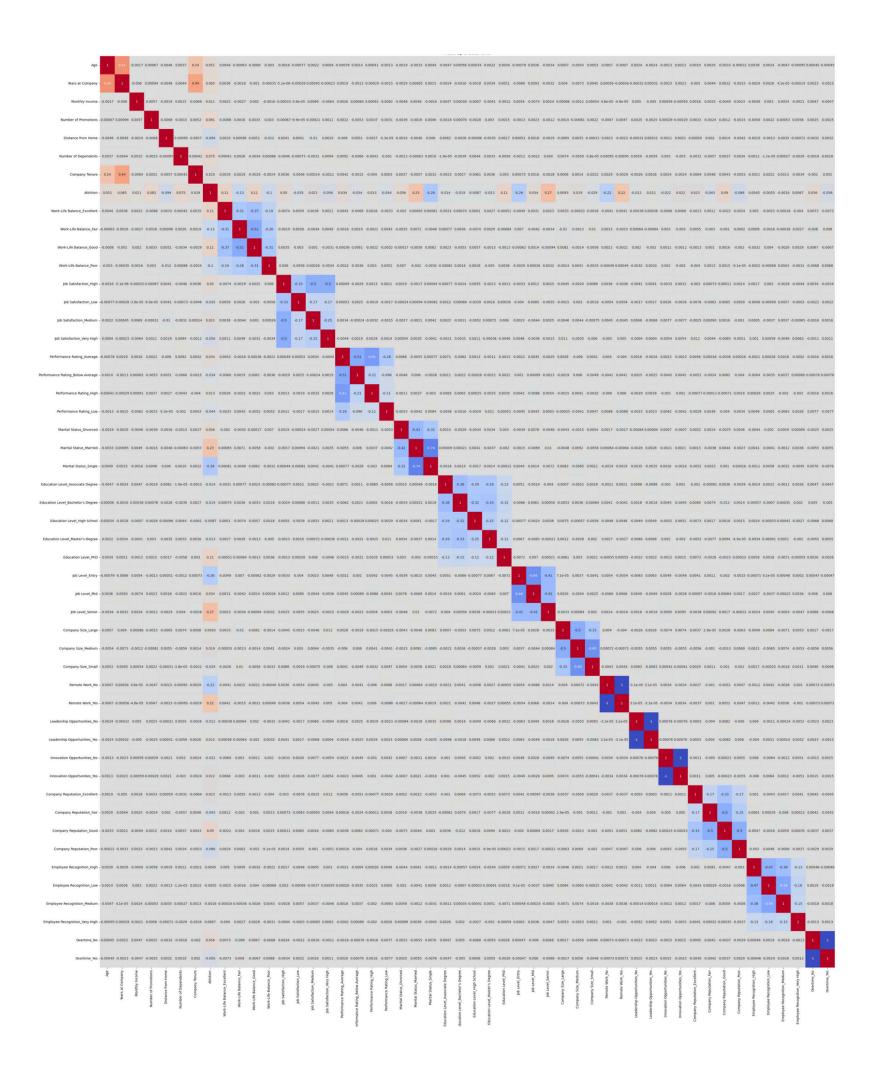
```
# คอลัมน์ที่เป็นตัวแปรประเภท categorical ที่ต้องการทำ One-Hot Encoding categorical_cols = ['Gender', 'Job Role', 'Work-Life Balance', 'Job Satisfaction', 'Performance Rating', 'Marital Status', 'Education Level', 'Job Level', 'Company Size', 'Remote Work', 'Leadership Opportunities', 'Innovation Opportunities', 'Company Reputation', 'Employee Recognition', 'Overtime']

# ทำ One-Hot Encoding โดยใช้ pd.get_dummies()
data = pd.get_dummies(data, columns=categorical_cols)

# แสดงผลลัพธ์
print(data.head())
```

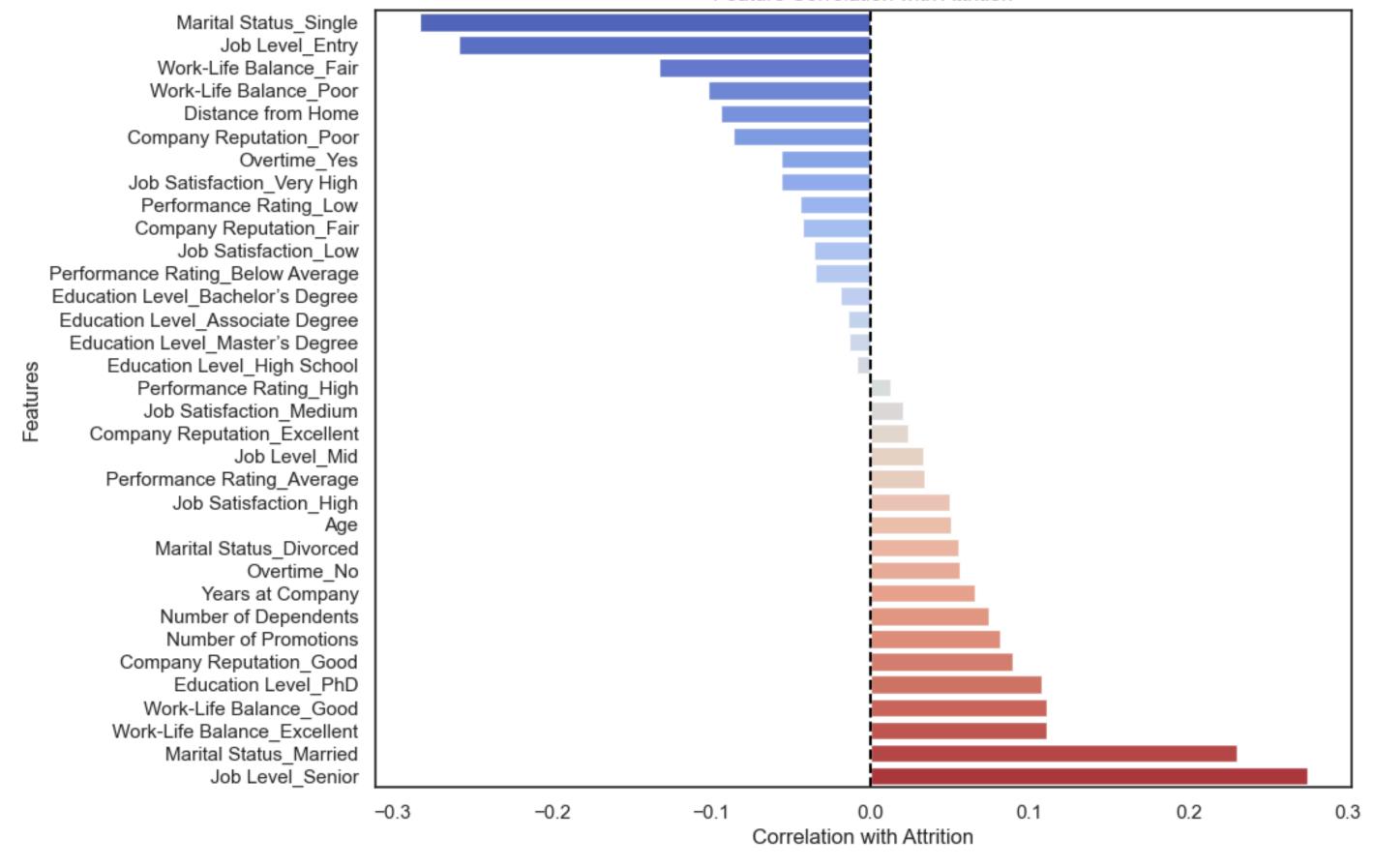
```
57
Index(['Age', 'Years at Company', 'Monthly Income', 'Number of Promotions',
       'Distance from Home', 'Number of Dependents', 'Company Tenure',
       'Attrition', 'Gender_Female', 'Gender_Male', 'Job Role_Education',
       'Job Role Finance', 'Job Role Healthcare', 'Job Role Media',
       'Job Role Technology', 'Work-Life Balance Excellent',
       'Work-Life Balance Fair', 'Work-Life Balance Good',
       'Work-Life Balance Poor', 'Job Satisfaction High',
       'Job Satisfaction_Low', 'Job Satisfaction_Medium',
       'Job Satisfaction Very High', 'Performance Rating Average',
       'Performance Rating Below Average', 'Performance Rating High',
       'Performance Rating Low', 'Marital Status Divorced',
       'Marital Status Married', 'Marital Status Single',
       'Education Level Associate Degree', 'Education Level Bachelor's Degree',
       'Education Level High School', 'Education Level Master's Degree',
       'Education Level PhD', 'Job Level Entry', 'Job Level Mid',
       'Job Level_Senior', 'Company Size_Large', 'Company Size_Medium',
       'Company Size Small', 'Remote Work No', 'Remote Work Yes',
       'Leadership Opportunities No', 'Leadership Opportunities Yes',
       'Innovation Opportunities No', 'Innovation Opportunities Yes',
       'Company Reputation_Excellent', 'Company Reputation_Fair',
       'Company Reputation_Good', 'Company Reputation_Poor',
       'Employee Recognition_High', 'Employee Recognition_Low',
       'Employee Recognition Medium', 'Employee Recognition Very High',
       'Overtime No', 'Overtime Yes'],
      dtype='object')
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 74498 entries, 0 to 74497
Data columns (total 57 columns):
    Column
                                       Non-Null Count Dtype
                                       74498 non-null float64
    Age
 0
    Years at Company
                                       74498 non-null float64
    Monthly Income
                                       74498 non-null float64
    Number of Promotions
                                       74498 non-null float64
    Distance from Home
                                       74498 non-null float64
    Number of Dependents
                                       74498 non-null int64
    Company Tenure
                                       74498 non-null float64
    Attrition
                                       74498 non-null int64
    Gender Female
                                       74498 non-null bool
                                       74498 non-null bool
    Gender Male
                                       74498 non-null bool
    Job Role_Education
 11 Job Role_Finance
                                       74498 non-null bool
   Job Role Healthcare
                                       74498 non-null bool
    Job Role_Media
                                       74498 non-null bool
    Job Role_Technology
                                       74498 non-null bool
 15 Work-Life Balance_Excellent
                                       74498 non-null bool
 16 Work-Life Balance_Fair
                                       74498 non-null bool
 17 Work-Life Balance_Good
                                       74498 non-null bool
    Work-Life Balance Poor
                                       74498 non-null bool
    Job Satisfaction_High
                                       74498 non-null bool
. . .
56 Overtime_Yes
                                       74498 non-null bool
dtypes: bool(49), float64(6), int64(2)
```



ถ้า Correalation Matrix มีค่า 0.03 ขึ้นไป แสดงค่านั้นว่ามีผลต่อการลาออกมาก

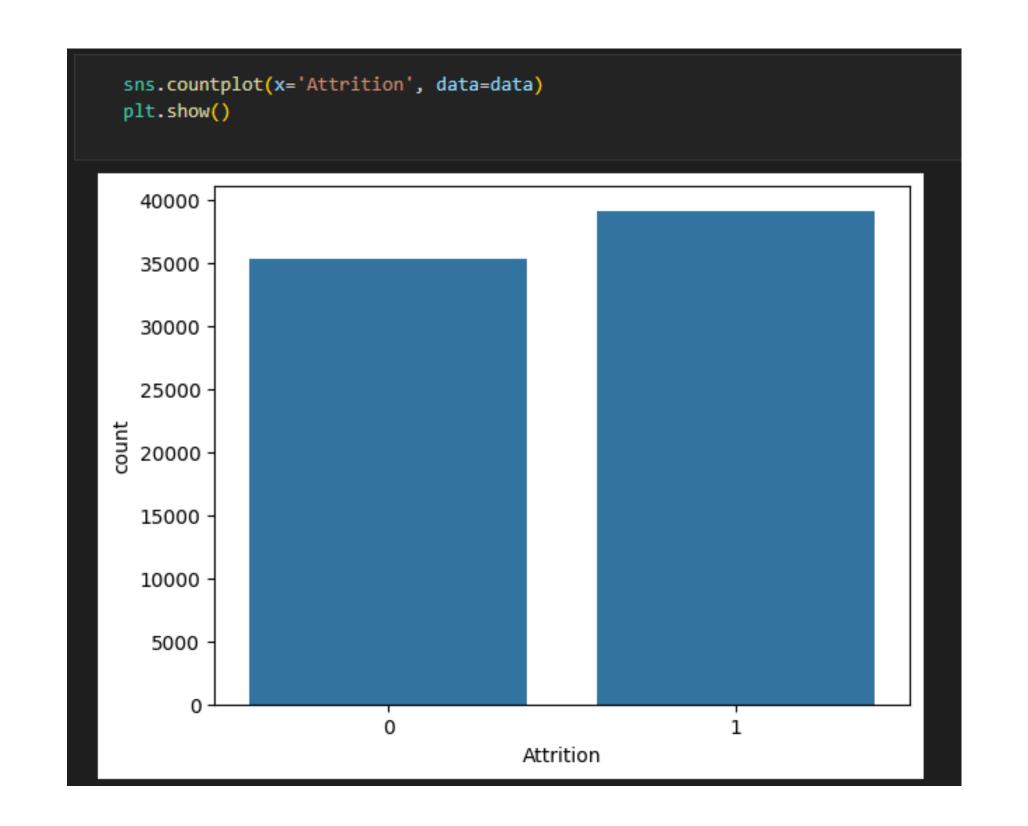
Feature Correlation with Attrition



Python

```
print(len(data.columns))
   print(data.columns)
35
Index(['Age', 'Years at Company', 'Number of Promotions', 'Distance from Home',
       'Number of Dependents', 'Attrition', 'Work-Life Balance Excellent',
       'Work-Life Balance Fair', 'Work-Life Balance Good',
       'Work-Life Balance_Poor', 'Job Satisfaction_High',
       'Job Satisfaction Low', 'Job Satisfaction Medium',
       'Job Satisfaction Very High', 'Performance Rating Average',
       'Performance Rating Below Average', 'Performance Rating High',
       'Performance Rating_Low', 'Marital Status_Divorced',
       'Marital Status Married', 'Marital Status Single',
       'Education Level_Associate Degree', 'Education Level_Bachelor's Degree',
       'Education Level High School', 'Education Level Master's Degree',
       'Education Level_PhD', 'Job Level_Entry', 'Job Level_Mid',
       'Job Level Senior', 'Company Reputation Excellent',
       'Company Reputation_Fair', 'Company Reputation_Good',
       'Company Reputation_Poor', 'Overtime_No', 'Overtime_Yes'],
      dtype='object')
```

																	Heatma	p of Data	aFrame																
Age	1	0.54	0.00067	-0.0046	0.0037	0.051	0.0044	-0.00063	-0.0008	-0.003	-0.0016	-0.00077	0.0022	0.0004	-0.00078	0.0014	0.00041	-0.0013	-0.0019	-0.0033	0.0049	-0.0047	0.00056	0.00034	0.0022	0.0034	-0.00079	0.0036	-0.0034	0.0019	0.0029	-0.0033	-0.00022	0.00045	0.0004
Years at Company	0.54	1	0.00094	-0.0048	0.0044	0.065	0.0036	-0.0016	-0.001	-0.00035	-5.1e-06	-0.00028	0.00045	0.00023	0.0019	-0.0013	-0.00029	-0.0015	-0.0029	0.00065	0.0015	-0.0024	-0.0016	-0.0018	0.0034	0.0051	-0.0066	0.0093	-0.0032	-0.005	0.0044	0.0022	-0.0033	0.0023	-0.002
Number of Promotions	- 0.00067	0.00094	1	-0.0068	-0.0015	0.081	-0.0088	0.0016	0.0033	0.003	-0.00067	-9.9e-05	-0.00021	0.0011	0.0022	-0.0053	0.0037	-0.0031	-0.0039	-0.0016	0.0046	-0.0018	0.00076	-0.0028	0.003	0.0015	-0.0013	0.0023	-0.0012	0.0033	-0.0024	0.0012	-0.0015	0.0025	-0.002
Distance from Home	0.0046	-0.0048	-0.0068	1	-0.00095	-0.094	0.0033	0.00098	0.0051	-0.012	0.0041	0.0041	0.01	0.0019	0.006	0.0051	0.0027	-5.3e-05	-0.0016	0.0046	0.006	0.0062	-0.0028	-0.00096	-0.0035	0.0017	0.00051	0.0018	-0.0029	0.00059	0.002	0.0014	0.0042	-0.0032	0.0032
Number of Dependents	- 0.0037	0.0044	-0.0015	-0.00095	1	0.075	0.00042	0.0026	-0.0034	0.00088	0.0046	-0.00073	-0.0031	0.0094	0.0092	-0.0066	-0.0043	-0.002	-0.0013	-0.00063	0.0016	-1.9e-05	-0.0039	0.0044	0.0033	-0.0058	-0.0012	-0.0022	0.004	-0.0031	-0.0057	0.0037	0.0034	-0.0018	0.0018
Attrition	- 0.051	0.065	0.081	-0.094	0.075	1	0.11	-0.13	0.11	-0.1	0.05	-0.035	0.021	-0.056	0.034	-0.034	0.013	-0.044	0.056	0.23	-0.28	-0.014	-0.019	-0.0087	-0.013	0.11	-0.26	0.034	0.27	0.023	-0.043	0.09	-0.086	0.056	-0.056
Jork-Life Balance_Excellent	- 0.0044	0.0036	-0.0088	0.0033	0.00042	0.11	1	-0.31	-0.37	-0.19	-0.0074	0.0059	0.0038	0.0011	0.0043	-0.0068	0.0026	-0.0033	-0.002	0.00065	0.00081	-0.0031	0.00074	0.0001	0.0027	-0.00051	-0.0049	0.0031	0.0023	-0.0013	0.0012	-0.0022	0.0024	0.0073	-0.007
Work-Life Balance_Fair	0.00063	-0.0016	0.0016	0.00098	0.0026	-0.13	-0.31	1	-0.51	-0.26	-0.0019	0.0026	-0.0044	0.0049	-0.0016	0.0019	-0.0021	0.0043	-0.0035	0.0071	-0.0048	0.00077	0.0036	-0.0074	0.0029	-0.00084	0.007	-0.0042	-0.0034	0.0055	-0.003	-0.001	0.0002	-0.008	0.008
Work-Life Balance_Good	0.0008	-0.001	0.0033	0.0051	-0.0034	0.11	0.37	-0.51	1	-0.31	0.0035	-0.003	0.001	-0.0031	-0.00036	0.0061	-0.0022	-0.0052	-0.00017	0.0058	0.0062	0.0023	-0.0053	0.0057	-0.0013	-0.0013	-0.00062	0.0014	-0.00094	-0.0013	0.001	0.0016	-0.002	0.0067	-0.006
Work-Life Balance_Poor	0.003	-0.00035	0.003	-0.012	0.00088	-0.1	-0.19	-0.26	-0.31	1	0.006	-0.0058	0.00026	-0.0034	-0.0022	-0.0036	0.003	0.0052	0.007	-0.002	-0.0032	-0.00082	0.0018	0.0018	-0.005	0.0036	-0.0029	0.00026	0.0032	-0.004	0.0013	0.0015	-9.1e-05	-0.0068	0.0068
Job Satisfaction_High	0.0016	-5.1e-06	-0.00067	0.0041	-0.0046	0.05	-0.0074	-0.0019	0.0035	0.006	1	-0.33	-0.5	-0.5	0.00049	-0.0029	0.0013	0.0011	0.0019	-0.0017	0.00044	-0.00077	-0.0024	0.0055	-0.0013	-0.0013	-0.0033	0.0012	0.0025	-0.003	0.00073	0.00011	0.0014	0.0034	-0.003
Job Satisfaction_Low	0.00077	-0.00028	-9.9e-05	0.0041	-0.00073	-0.035	0.0059	0.0026	-0.003	-0.0058	-0.33	1	-0.17	-0.17	0.00053	0.0025	-0.0019	-0.0017	-0.00024	0.00094	-0.00081	0.0012	0.00066	-0.0039	0.0016	0.00039	-0.004	0.0085	-0.0055	-0.0076	-0.0083	0.0065	0.0059	-0.0022	0.0022
Job Satisfaction_Medium	- 0.0022	0.00045	-0.00021	-0.01	-0.0031	0.021	0.0038	-0.0044	0.001	0.00026	-0.5	-0.17	1	-0.25	0.0034	-0.00024	-0.0032	-0.0015	-0.0027	-0.0021	0.0042	0.0022	-0.0011	-0.0052	0.00072	0.006	0.0023	-0.0044	0.0025	-0.0025	0.00093	0.0016	-0.001	0.0016	0.001
Job Satisfaction_Very High	- 0.0004	-0.00023	0.0011	0.0019	0.0094	-0.056	0.0011	0.0049	-0.0031	-0.0034	-0.5	-0.17	-0.25	1	-0.0044	0.0019	0.0029	0.0014	0.00054	0.0035	-0.0041	-0.0021	0.0035	0.0011	-0.00036	-0.0046	0.0048	-0.0036	-0.0015	0.012	0.0044	-0.0065	-0.0051	-0.0011	0.001
rformance Rating_Average	0.00078	0.0019	0.0022	-0.006	0.0092	0.034	0.0043	-0.0016	-0.00036	-0.0022	0.00049	0.00053	0.0034	-0.0044	1	-0.51	-0.61	-0.28	0.0066	-0.0055	0.00077	0.0071	-0.0062	0.0013	-0.0011	-0.0015	-0.0021	0.0045	-0.0029	0.0056	0.00016	-0.0036	0.00016	-0.0016	0.0016
ince Rating_Below Average	- 0.0014	-0.0013	-0.0053	0.0051	-0.0066	-0.034	-0.0068	0.0019	0.0061	-0.0036	-0.0029	0.0025	-0.00024	0.0019	-0.51	1	-0.21	-0.096	0.0046	0.006	-0.0028	0.0011	0.0022	-0.00028	-0.0022	-0.0022	0.001	0.00089	-0.0023	-0.0051	-0.0024	0.0082	-0.004	0.00076	0.0007
Performance Rating_High	- 0.00041	-0.00029	0.0037	0.0027	-0.0043	0.013	0.0026	-0.0021	-0.0022	0.003	0.0013	-0.0019	-0.0032	0.0029	-0.61	-0.21	1	-0.11	-0.0011	0.0037	-0.003	-0.0065	0.0065	0.00025	-0.0025	0.0035	0.0042	-0.0086	0.0054	-0.00077	-0.00011	-0.00071	0.0016	-0.0016	0.0016
Performance Rating_Low	0.0013	-0.0015	-0.0031	-5.3e-05	-0.002	-0.044	-0.0033	0.0043	-0.0052	0.0052	0.0011	-0.0017	-0.0015	0.0014	-0.28	-0.096	-0.11	1	-0.0053	-0.0042	0.0084	-0.0058	-0.0016	-0.0029	0.011	0.00053	-0.0045	0.0043	0.0003	-0.0029	0.0038	-0.004	0.0034	0.0077	-0.007
Marital Status_Divorced	0.0019	-0.0029	-0.0039	-0.0016	-0.0013	0.056	-0.002	-0.0035	-0.00017	0.007	0.0019	-0.00024	-0.0027	0.00054	0.0066	-0.0046	-0.0011	-0.0053	1	-0.42	-0.31	0.0015	-0.0029	-0.0034	0.0034	0.003	-0.0039	0.0078	-0.0048	0.0052	0.0019	-0.0075	0.0036	-0.0025	0.0021
Marital Status_Married	0.0033	0.00065	-0.0016	-0.0046	-0.00063	0.23	0.00065	0.0071	-0.0058	-0.002	-0.0017	0.00094	-0.0021	0.0035	-0.0055	0.006	0.0037	-0.0042	-0.42	1	-0.74	0.00049	0.00021	0.0041	-0.0037	-0.002	-0.0015	-0.0069	0.01	0.0013	-0.0038	0.0044	-0.0027	-0.0055	0.0055
Marital Status_Single	- 0.0049	0.0015	0.0046	0.006	0.0016	-0.28	0.00081	-0.0048	0.0062	-0.0032	0.00044	-0.00081	0.0042	-0.0041	0.00077	-0.0028	-0.003	0.0084	-0.31	-0.74	1	-0.0016	0.0019	-0.0017	0.0014	-0.00015	0.0045	0.0014	-0.0072	-0.0052	0.0025	0.001	0.00016	0.0076	-0.007
on Level_Associate Degree	0.0047	-0.0024	-0.0018	0.0062	-1.9e-05	-0.014	-0.0031	0.00077	0.0023	-0.00082	-0.00077	0.0012	0.0022	-0.0021	0.0071	0.0011	-0.0065	-0.0058	0.0015	0.00049	-0.0016	1	-0.38	-0.29	-0.29	-0.13	0.0051	-0.0019	-0.004	-0.001	-0.00081	0.0036	-0.0029	0.0047	-0.004
on Level_Bachelor's Degree	- 0.00056	-0.0016	0.00076	-0.0028	-0.0039	-0.019	0.00074	0.0036	-0.0053	0.0018	-0.0024	0.00066	-0.0011	0.0035	-0.0062	0.0022	0.0065	-0.0016	-0.0029	0.00021	0.0019	-0.38	1	-0.32	-0.33	-0.15	-0.0066	0.0061	0.00056	0.0084	0.0074	-0.012	0.0014	0.005	-0.005
ducation Level_High School	- 0.00034	-0.0018	-0.0028	-0.00096	0.0044	-0.0087	0.0001	-0.0074	0.0057	0.0018	0.0055	-0.0039	-0.0052	0.0011	0.0013	-0.00028	0.00025	-0.0029	-0.0034	0.0041	-0.0017	-0.29	0.32	ı.	-0.25	-0.12	-0.00077	-0.0024	0.0038	-0.0073	0.0017	0.0018	0.0015	-0.0068	0.0068
tion Level_Master's Degree	- 0.0022	0.0034	0.003	-0.0035	0.0033	-0.013	0.0027	0.0029	-0.0013	-0.005	0.0013	0.0016	0.00072	0.00036	-0.0011	-0.0022	0.0025	0.011	0.0034	-0.0037	0.0014	-0.29	-0.33	-0.25	1	-0.12	0.0067	-0.0065	-0.00021	-0.0052	-0.0077	0.0094	-6.9e-05	-0.0055	0.0055
Education Level_PhD	- 0.0034	0.0051	0.0015	0.0017	-0.0058	0.11	-0.00051	-0.00084	-0.0013	0.0036	-0.0013	0.00039	0.006	-0.0046	-0.0015	-0.0022	0.0035	0.00053	0.003	-0.002	-0.00015	-0.13	-0.15	-0.12	-0.12	1	-0.0072	0.007	0.00021	0.0073	-0.0028	-0.0023	0.00023	0.0026	-0.002
job Level_Entry	0.00079	-0.0066	-0.0013	0.00051	-0.0012	-0.26	-0.0049	0.007	-0.00062	-0.0029	-0.0033	-0.004	0.0023	0.0048	-0.0021	0.001	0.0042	-0.0045	-0.0039	-0.0015	0.0045	0.0051	-0.0066	-0.00077	0.0067	-0.0072	1	-0.66	-0.41	0.0041	0.0011	-0.002	-0.0015	0.00047	0.0004
Job Level_Mid	- 0.0036	0.0093	0.0023	0.0018	-0.0022	0.034	0.0031	-0.0042	0.0014	0.00026	0.0012	0.0085	-0.0044	-0.0036	0.0045	0.00089	-0.0086	0.0043	0.0078	-0.0069	0.0014	-0.0019	0.0061	-0.0024	-0.0065	0.007	-0.66	1	-0.41	-0.00097	-0.0018	0.00064	0.0017	-0.006	0.006
job Level_Senior	0.0034	-0.0032	-0.0012	-0.0029	0.004	0.27	0.0023	-0.0034	-0.00094	0.0032	0.0025	-0.0055	0.0025	-0.0015	-0.0029	-0.0023	0.0054	0.0003	-0.0048	0.01	-0.0072	-0.004	0.00056	0.0038	-0.00021	0.00021	0.41	-0.41	1	-0.0038	0.00092	0.0017	-0.00022	0.0068	-0.006
pany Reputation_Excellent	- 0.0019	-0.005	0.0033	0.00059	-0.0031	0.023	-0.0013	0.0055	-0.0013	-0.004	-0.003	-0.0076	-0.0025	0.012	0.0056	-0.0051	-0.00077	-0.0029	0.0052	0.0013	-0.0052	-0.001	0.0084	-0.0073	-0.0052	0.0073	0.0041	-0.00097	-0.0038	1	-0.17	-0.33	-0.17	0.0045	-0.004
Company Reputation_Fair	- 0.0029	0.0044	-0.0024	0.002	-0.0057	0.043	0.0012	-0.003	0.001	0.0013	0.00073	-0.0083	0.00093	0.0044	0.00016	-0.0024	-0.00011	0.0038	0.0019	0.0038	0.0025	-0.00081	0.0074	0.0017	-0.0077	-0.0028	0.0011	-0.0018	0.00092	-0.17	1	-0.5	-0.25	0.0042	-0.004
Company Reputation_Good	0.0033	0.0022	0.0012	0.0014	0.0037	0.09	-0.0022	-0.001	0.0016	0.0015	0.00011	0.0065	0.0016	-0.0065	-0.0036	0.0082	-0.00071	-0.004	-0.0075	0.0044	0.001	0.0036	-0.012	0.0018	0.0094	-0.0023	-0.002	0.00064	0.0017	-0.33	-0.5	1	-0.5	-0.0037	0.0037
Company Reputation_Poor	0.00022	-0.0033	-0.0015	-0.0042	0.0034	-0.086	0.0024	0.0002	-0.002	-9.1e-05	0.0014	0.0059	-0.001	-0.0051	0.00016	-0.004	0.0016	0.0034	0.0036	-0.0027	0.00016	-0.0029	0.0014	0.0015	-6.9e-05	0.00023	-0.0015	0.0017	-0.00022	-0.17	-0.25	-0.5	1	-0.0029	0.0021
Overtime_No	- 0.00045	0.0023	0.0025	-0.0032	-0.0018	0.056	0.0073	-0.008	0.0067	-0.0068	0.0034	-0.0022	-0.0016	-0.0011	-0.0016	-0.00076	-0.0016	0.0077	-0.0025	-0.0055	0.0076	0.0047	0.005	-0.0068	-0.0055	0.0026	0.00047	-0.006	0.0068	0.0045	0.0042	-0.0037	-0.0029		
Overtime_Yes	- 0.00045	-0.0023	-0.0025	0.0032	0.0018	0.056	-0.0073	0.008	-0.0067	0.0068	0.0034	0.0022	0.0016	0.0011	0.0016	0.00076	0.0016	-0.0077	0.0025	0.0055	-0.0076	-0.0047	-0.005	0.0068	0.0055	0.0026	0.00047	0.006	-0.0068	-0.0045	0.0042	0.0037	0.0029	4	1
	Age	Years at Company	Number of Promotions -	Distance from Home -	Number of Dependents .	Attriben -	Work-Life Balance_Excellent -	Work-Life Balance_Fair	Work-Life Balance_Good -	Work-Life Balance_Poor	job Satisfaction_High.	job Satisfaction_Low -	job Satisfaction_Medium -	job Satisfaction_Very High .	Performance Rating_Average -	Performance Rating_Below Avetage	Performance Rating_High -	Performance Rating_Low -	Marital Status_Divorced -	Markal Status, Married -	Marral Status_Single	Education Level_Associate Degree -	Education Level_Bachelor's Degree -	Education Level_High School -	Education Level_Master's Degree -	Education Level_PhD -	Job Level_Entry -	bb Level_Mid	Job Level_Senior	Company Reputation_Excellent -	Company Reputation_Fair -	Company Reputation_Good -	Company Peputation_Poor -	Overtime_No -	Overtime_Yes -



```
# แบ่งข้อมูล

X = data.drop(columns=['Attrition']) # Features (ทั้งหมดยกเว้น Attrition)

y = data['Attrition'] # Target column

# แบ่งข้อมูล

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

print(f"Train set size: {X_train.shape}, Test set size: {X_test.shape}")

Train set size: (59598, 34), Test set size: (14900, 34)
```

```
model = LogisticRegression(random_state=42)

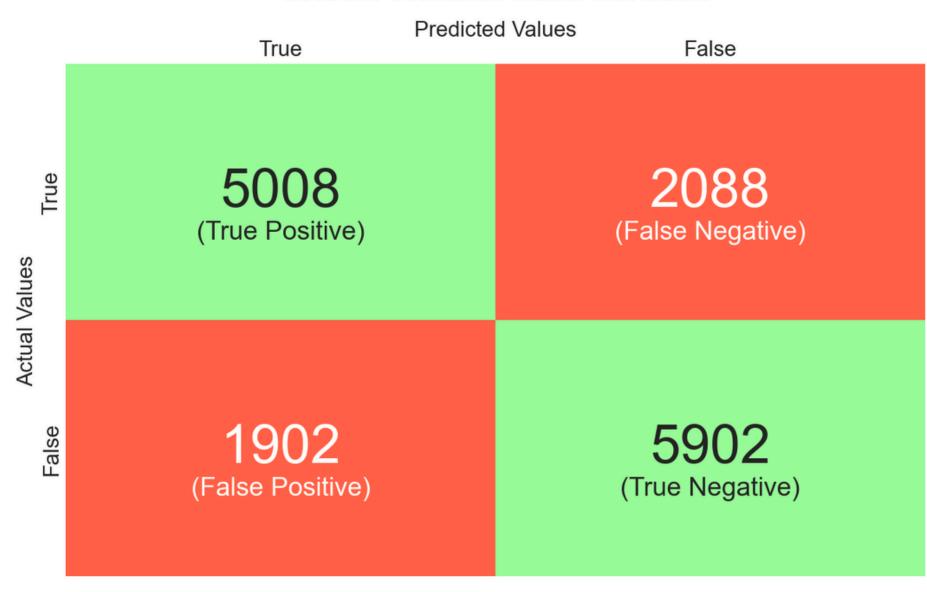
# ฝึกโมเดล
model.fit(X_train, y_train)

# หดสอบโมเดล
y_pred = model.predict(X_test)

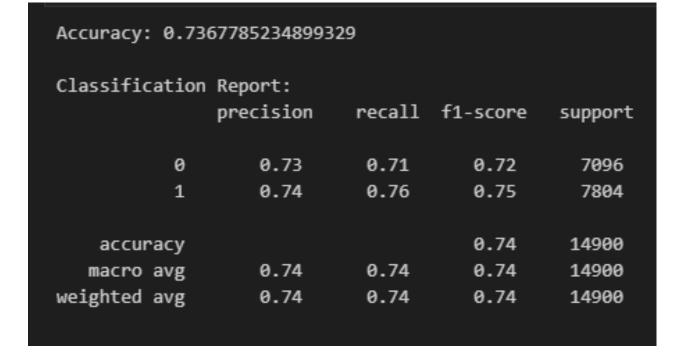
# ประเมินโมเดล
print("Accuracy:", accuracy_score(y_test, y_pred))
print(classification_report(y_test, y_pred))

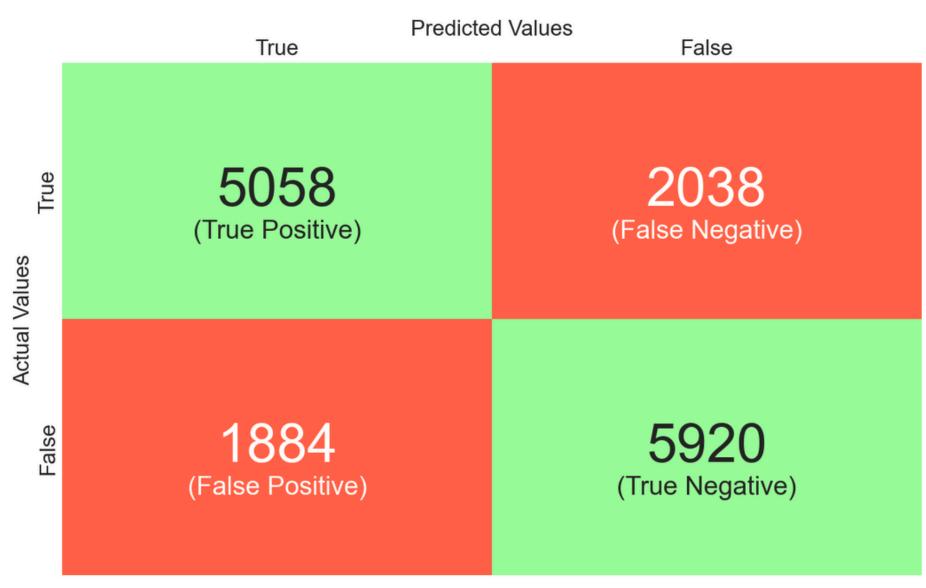
cftmt_lo = confusion_matrix(y_test, y_pred)
print(cftmt_lo)
custom_plt_confution_matrix(cftmt_lo)
```

Accuracy: 0.7	322147651006	5712			
	precision	recall	f1-score	support	
	0.70	0.74	0.70	7006	
0	0.72	0.71	0.72	7096	
1	0.74	0.76	0.75	7804	
accuracy			0.73	14900	
macro avg	0.73	0.73	0.73	14900	
weighted avg	0.73	0.73	0.73	14900	



```
gb_model = GradientBoostingClassifier(n_estimators=100, learning_rate=0.1,
max_depth=3, random_state=42)
# เทรนโมเดลด้วยข้อมูล train
gb_model.fit(X_train, y_train)
# ทำนายผลด้วยข้อมูล test
y_pred_gb = gb_model.predict(X_test)
# ประเมินผลโมเดล
print("Accuracy:", accuracy_score(y_test, y_pred_gb))
print("\nClassification Report:\n", classification_report(y_test, y_pred_gb))
print("\nConfusion Matrix:\n", confusion_matrix(y_test, y_pred_gb))
gBoostCunMatrix = confusion_matrix(y_test, y_pred_gb)
```





```
k = 5

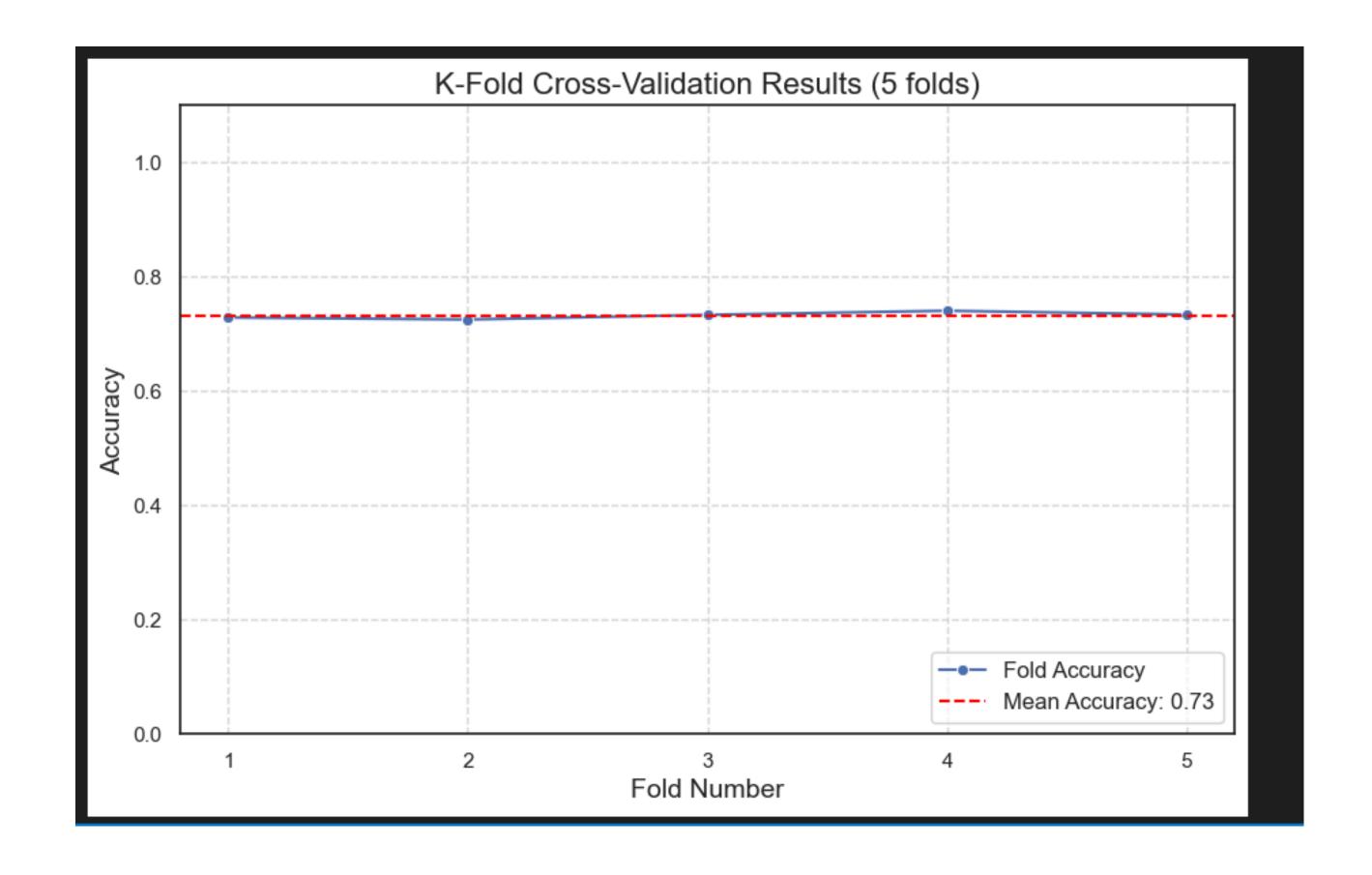
# สร้างโมเดล Gradient Boosting
gb_model = GradientBoostingClassifier(n_estimators=100, learning_rate=0.1, max_depth=3, random_state=42)

# สร้าง K-Fold Cross-Validator (Stratified เพื่อให้คลาสสมดุลในแต่ละ Fold)
kfold = StratifiedKFold(n_splits=k, shuffle=True, random_state=42)

# ใช้ cross_val_score ประเมินโมเดล
scores = cross_val_score(gb_model, X, y, cv=kfold, scoring='accuracy')

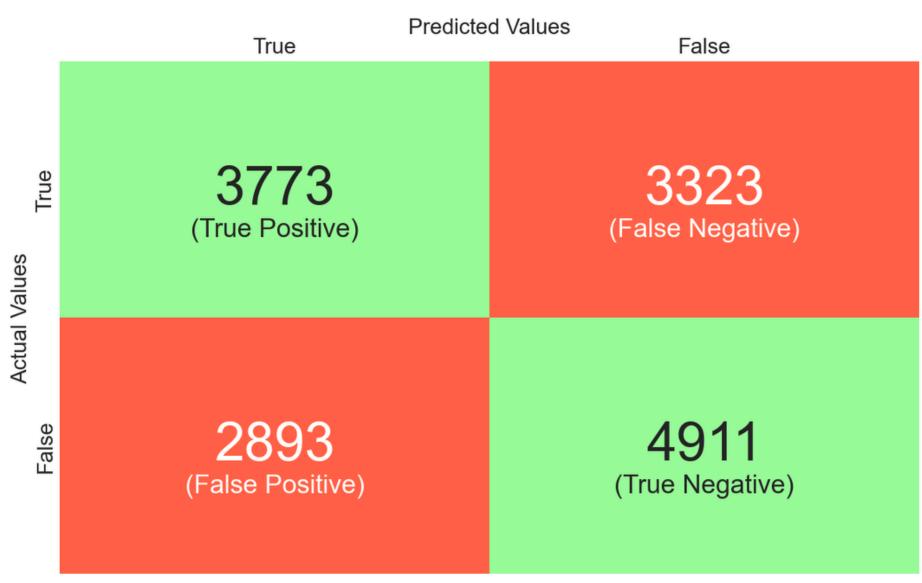
# แสดงผลตะแนน
print(f"K-Fold Cross-Validation Results ({k} folds):")
print(f"Scores: {scores}")
print(f"Mean Accuracy: {scores.mean():.2f}")
print(f"Standard Deviation: {scores.std():.2f}")
```

```
K-Fold Cross-Validation Results (5 folds):
Scores: [0.72865772 0.72516779 0.73328859 0.74045238 0.73380764]
Mean Accuracy: 0.73
Standard Deviation: 0.01
```



```
model = LogisticRegression(random_state=42)
# ฝึกโมเดล
model.fit(X_train, y_train)
# ทดสอบโมเดล
y_pred = model.predict(X_test)
# ประเมินโมเดล
print("Accuracy:", accuracy_score(y_test, y_pred))
print(classification_report(y_test, y_pred))
cftmt_lo = confusion_matrix(y_test, y_pred)
print(cftmt_lo)
custom_plt_confution_matrix(cftmt_lo)
```

Accuracy: 0.	5828187919463	3087		
	precision	recall	f1-score	support
9	0.57	0.53	0.55	7096
1	0.60	0.63	0.61	7804
accuracy			0.58	14900
macro avg	0.58	0.58	0.58	14900
weighted avg	0.58	0.58	0.58	14900

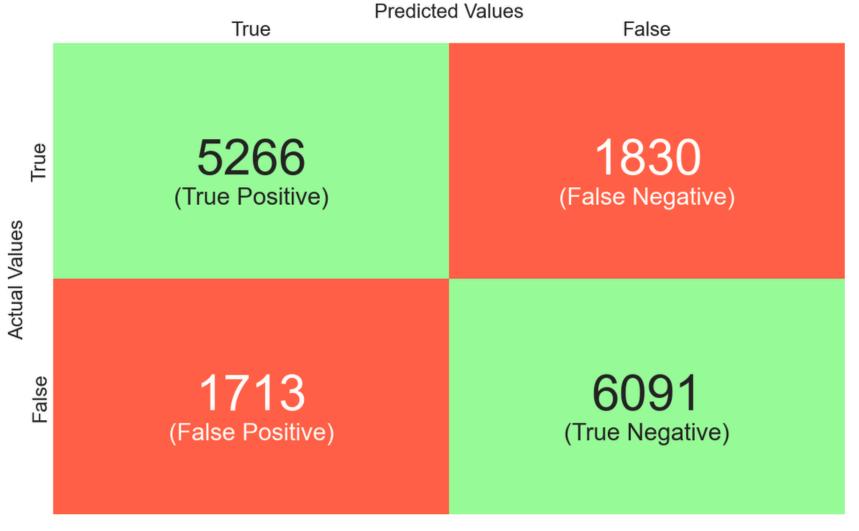


```
gb_model = GradientBoostingClassifier(n_estimators=100, learning_rate=0.1, max_depth=3, random_state=42)
# เทรนโมเดลด้วยข้อมูล train
gb_model.fit(X_train, y_train)

# ทำนายผลด้วยข้อมูล test
y_pred_gb = gb_model.predict(X_test)

# ประเมินผลโมเดล
print("Accuracy:", accuracy_score(y_test, y_pred_gb))
print("\nClassification Report:\n", classification_report(y_test, y_pred_gb))
print("\nConfusion Matrix:\n", confusion_matrix(y_test, y_pred_gb))
gBoostConMatrix = confusion_matrix(y_test, y_pred_gb)
```

Accuracy: 0.762	221476510067	11			
Classification	Report: precision	recall	f1-score	support	
0	0.75	0.74	0.75	7096	
1	0.77	0.78	0.77	7804	
accuracy			0.76	14900	
macro avg	0.76	0.76	0.76	14900	
weighted avg	0.76	0.76	0.76	14900	
Confusion Matri	ix:				
[[5266 1830] [1713 6091]]					



Logistic Regression เปรียบเทียบความแตกต่าง

เปรียบเทียบ	แบบไม่ทำอะไรเลย	แบบทำ Standardization & Drop features			
Accuracy	58.2%	73.2%			
Precision (0,1)	(0.57, 0.60)	(0.72, 0.74)			
Recall (0,1)	(0.53, 0.63)	(0.71, 0.76)			
Confusion Matrix (TP, FN, FP, TN)	(3773, 3323, 2893, 4911)	(5008, 2088, 1902, 5902)			

Gradient Boosting เปรียบเทียบความแตกต่าง

Metric	Gradient Boosting (ไม่ทำอะไร เลย)	Gradient Boosting (ทำ Feature Engineering)
Accuracy	76.2%	73.6%
Precision (0,1)	(0.75, 0.77)	(0.73, 0.74)
Recall (0,1)	(0.75, 0.78)	(0.71, 0.76)
Confusion Matrix (TP, FN, FP, TN)	(5266, 1830, 1713, 6091)	(5058, 2038, 1884, 5920)



Accuracy: 0.69	892617449664	43		Gr	adient Boosting
Classification	Report:				G
	precision	recall	f1-score	support	
0	0.68	0.69	0.69	7096	
1	0.72	0.71	0.71	7804	
accuracy			0.70	14900	
macro avg	0.70	0.70	0.70	14900	
weighted avg	0.70	0.70	0.70	14900	
Confusion Matr [[4907 2189] [2297 5507]]	ix:				

หลังจากดรอป Feature ที่มีผลต่อ y มากที่สุดออก มีค่า Accuracy: 0.6 น้อยลง data = data.drop((['Gender_Female', 'Gender_Male', 'Job Role_Education',
 'Job Role_Finance', 'Job Role_Healthcare', 'Job Role_Media',
 'Job Role_Technology']), axis=1)
print(data.columns)

Accuracy: 0.7331543624161074									
	precision	recall	f1-score	support					
0	0.73	0.71	0.72	7096					
1	0.74	0.76	0.75	7804					
accuracy			0.73	14900					
macro avg	0.73	0.73	0.73	14900					
weighted avg	0.73	0.73	0.73	14900					
[[5009 2087] [1889 5915]]	Lo	gistic	e Regr	ession					

Accuracy: 0.73	778523489932	89		
Classification	Report: precision	recall	f1-score	support
ø	0.73	0.71	0.72	7096
1	0.74	0.76	0.75	7804
accuracy macro avg weighted avg	0.74 0.74	0.74 0.74	0.74 0.74 0.74	14900 14900 14900
weighted avg	0.74	0.74	0.74	14900
Confusion Matr [[5040 2056] [1851 5953]]	rix:	Gra	dien	t Boosting

หลังจากดรอป Feature ที่มีผลต่อ y น้อย มีค่า Accuracy: 0.73 ไม่ต่างกันมาก

Note

- 1. ถ้า Correalation Matrix มีค่า 0.03 ขึ้นไปให้ใช้ค่านั้น (ว่ามีผลต่อการลาออก)
- 2. [1 คืออยู่, 0 คือลาออก]
- 3. (ตอนนี้เรา predict ว่าปัจจัยใดที่มีผลต่อการลาออกได้แล้ว) [27 Jan 2025]