

Case Study

Name: Raizen Sangalang

Scenario: There is a company that hired male and female developers and they have IT/CS graduates and also non IT/CS graduates some of them are from Mathematics, some of them are career shifters

Case study about Gender in Software engineering field who have background in any IT or CS related Program

Gender Column

In this case 1 represents the Male in Gender

In this case 0 represents the Female in Gender

College Column

1 represents it attends IT related College program

0 it doesn't attend in any IT related College program

```
In [3]: import pandas as pd
import numpy as np
df=pd.read_csv("dataset.csv")
```

```
In [4]: df
```

```
Out[4]:
```

	Gender	Age	Work	College
0	1	18	Quality Engineer	1
1	0	21	Java Developer	0
2	0	22	Software developer	0
3	1	21	Python Developer	0
4	1	20	Frontend Developer	1
...
95	0	21	Java Developer	1
96	1	19	Quality Engineer	0
97	0	20	Database Manager	1
98	1	18	Database Manager	1
99	0	20	Quality Assurance	1

100 rows × 4 columns

```
In [5]: df["Gender"]=df.Gender.replace({
        1:"Male", 0:"Female"
        })
dx = df[["College", "Gender"]].dropna()
pd.crosstab(dx.College,dx.Gender)
```

```
Out[5]:  Gender  Female  Male
College
0         27     20
1         22     31
```

```
In [6]: collegeFemale = 22
collegeMale = 31
nonCollegeFemale= 27
nonCollegeMale = 20
female = collegeFemale + nonCollegeFemale
male = collegeMale + nonCollegeMale
```

```
In [7]: from statsmodels import api as sm
```

```
In [8]: p_female = collegeFemale/(female)
```

```
In [9]: sm.stats.proportion_confint(female * p_female, female)
```

```
Out[9]: (0.3097129211477546, 0.5882462625257148)
```

interpretation (female)

With **95%** confidence, the classification accuracy/error is likely between **0.30971** and **0.5882**, based on **100** samples

```
In [10]: p_male = collegeMale/(male)
```

```
In [11]: sm.stats.proportion_confint(male * p_male, male)
```

```
Out[11]: (0.4738481130108845, 0.7418381614989193)
```

interpretation (male)

With **95%** confidence, the classification accuracy/error is likely between **0.4738** and **0.7418**, based on 100 samples

In []: