

```
In [1]: #Import all Libraries
import pandas as pd
import numpy as np
from sklearn.feature_selection import chi2
from scipy.stats import chi2_contingency
from sklearn.preprocessing import OrdinalEncoder
```

```
In [2]: data = pd.read_csv("Customer+OrderForm.csv")
data.head()
```

```
Out[2]:
```

	Phillippines	Indonesia	Malta	India
0	Error Free	Error Free	Defective	Error Free
1	Error Free	Error Free	Error Free	Defective
2	Error Free	Defective	Defective	Error Free
3	Error Free	Error Free	Error Free	Error Free
4	Error Free	Error Free	Defective	Error Free

```
In [3]: data.dtypes
```

#datatype of each column is categoricall

```
Out[3]:
```

Phillippines	object
Indonesia	object
Malta	object
India	object

dtype: object

```
In [4]: data.columns
```

```
Out[4]: Index(['Phillippines', 'Indonesia', 'Malta', 'India'], dtype='object')
```

```
In [5]: print(data.Phillippines.value_counts())
print(data.Indonesia.value_counts())
print(data.Malta.value_counts())
print(data.India.value_counts())
```

```
Error Free    271
Defective      29
Name: Phillippines, dtype: int64
Error Free    267
Defective      33
Name: Indonesia, dtype: int64
Error Free    269
Defective      31
Name: Malta, dtype: int64
Error Free    280
Defective      20
Name: India, dtype: int64
```

Chi-Squared Test

```
In [6]: #contingency table
obs = np.array([[271, 267, 269, 280], [29, 33, 31, 20]])
```

```
obs
```

```
Out[6]: array([[271, 267, 269, 280],  
        [ 29,  33,  31,  20]])
```

```
In [7]: val = chi2_contingency(obs)  
val
```

```
Out[7]: (3.858960685820355,  
        0.2771020991233135,  
        3,  
        array([[271.75, 271.75, 271.75, 271.75],  
               [ 28.25,  28.25,  28.25,  28.25]]))
```

```
In [8]: expected_values = val[3]  
expected_values
```

```
Out[8]: array([[271.75, 271.75, 271.75, 271.75],  
               [ 28.25,  28.25,  28.25,  28.25]])
```

```
In [9]: p_value = val[0]  
p_value
```

```
Out[9]: 3.858960685820355
```

```
In [10]: if p_value < 0.05:                                # alpha value is 0.05 or 5%  
        print(" we are rejecting null hypothesis")  
    else:  
        print("we are accepting null hypothesis")
```

we are accepting null hypothesis

Since the p- value is(3.85) > 0.05, there is a significant difference and the defective % varies by centre