In [1]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import scipy
import scipy.stats as stats
import pylab
```

In [2]:

```
BR = pd.read_csv('/Users/acer/Sandesh Pal/Data Science Assgn/Hypothesis/BuyerRatio.csv')
```

In [3]:

```
BR
```

Out[3]:

	Observed Values	East	West	North	South
0	Males	50	142	131	70
1	Females	435	1523	1356	750

In [4]:

```
br = BR.drop(['Observed Values'], axis=1)
```

In [5]:

```
#Inputs are 4 discrete variables(east,west,north,south).
#Output is also discrete.
#We are trying to find out if proportions of male and female are similar or not across the regions
#Hence, we'll proceed with chi-square test
```

In [6]:

```
#Create hypothesis
#Ho= Proportions of Male and Female are same
#Ha= Proportions of Male and Female are not same
```

In [7]:

```
from scipy.stats import chi2_contingency
```

In [8]:

```
br
```

Out[8]:

	East	West	North	South
0	50	142	131	70
1	435	1523	1356	750

In [9]:

```
chi2_stat, p_val, dof, ex =stats.chi2_contingency(br)

print("===Chi2 Stat===")
print(chi2_stat)
print("\n")
print("===Degrees of Freedom===")
print(dof)
print("\n")
print("===P-Value===")
print(p_val)
print("\n")
print("===Contingency Table===")
print(ex)
```

```
===Chi2 Stat===
1.595945538661058


===Degrees of Freedom===
3


===P-Value===
0.6603094907091882


===Contingency Table===
[[  42.76531299  146.81287862  131.11756787   72.30424052]
 [ 442.23468701 1518.18712138 1355.88243213  747.69575948]]
```

In [10]:

```
#Since p-value (0.66)> alpha (0.05), hence we can't reject the null hypothesis
#Conclusion: proportion of male and female across regions is same.
```

In [ ]: