In [7]:
```
# import required packages
import pandas as pd
import numpy as np
import scipy.stats as ss
import warnings
warnings.filterwarnings("ignore")
from sklearn.preprocessing import LabelEncoder
```

In [3]:
```
data= pd.read_csv("Interaction+demography+payor.csv")
```

In [4]:
```
data.head()
```

Out[4]:

	patient_id	service	date of interaction	enrolled_channel	enrolled date	gender	age	payor
0	2038844	Support_person	2018-05-22 00:00:00	skype	2018-05-21 05:00:00	M	26-40	Commercial
1	2038844	Support_person	2018-05-22 00:00:00	skype	2018-05-24 05:00:00	M	26-40	Commercial
2	2038844	Support_person	2018-05-22 00:00:00	skype	2018-06-06 05:00:00	M	26-40	Commercial
3	2038844	Support_person	2018-05-22 00:00:00	skype	2019-04-22 05:00:00	M	26-40	Commercial
4	2038844	Collateral	2018-05-23 00:00:00	skype	2018-05-21 05:00:00	M	26-40	Commercial

In [8]:
```
label_encoder = LabelEncoder()
```

In [10]:
```
data[['service', 'gender', 'age']] = data[['service', 'gender', 'age']].apply
```

In [11]:
```
print(data)
```

```
         patient_id  service  date of interaction enrolled_channel  \
0           2038844        8  2018-05-22 00:00:00            skype
1           2038844        8  2018-05-22 00:00:00            skype
2           2038844        8  2018-05-22 00:00:00            skype
3           2038844        8  2018-05-22 00:00:00            skype
4           2038844        1  2018-05-23 00:00:00            skype
...             ...      ...                  ...              ...
2540917  2374501870        2  2018-08-22 00:00:00            PHONE
2540918  2374501870        2  2018-08-22 00:00:00            PHONE
2540919  2374501870        2  2018-08-22 00:00:00            PHONE
2540920  2374501870        2  2018-08-22 00:00:00            PHONE
2540921  2374501870        2  2018-08-22 00:00:00            PHONE

              enrolled date  gender  age       payor
0        2018-05-21 05:00:00       1    1  Commercial
1        2018-05-24 05:00:00       1    1  Commercial
2        2018-06-06 05:00:00       1    1  Commercial
3        2019-04-22 05:00:00       1    1  Commercial
4        2018-05-21 05:00:00       1    1  Commercial
...                      ...     ...  ...         ...
2540917  2016-09-22 05:00:00       0    1  Commercial
2540918  2016-09-26 05:00:00       0    1  Commercial
2540919  2016-09-28 05:00:00       0    1  Commercial
2540920  2016-10-12 05:00:00       0    1  Commercial
2540921  2018-06-26 05:00:00       0    1  Commercial

[2540922 rows x 8 columns]
```

In [12]:
```
# sample data
sample1= data.sample(n=100,random_state=1)
print(sample1)
```

```
          patient_id  service  date of interaction enrolled_channel  \
2534515      2037487        5  2021-04-27 00:00:00              WEB
2360387      1172182        2  2018-09-06 00:00:00              WEB
1002182   3304450181        8  2015-04-06 00:00:00            PHONE
838369      41010201        1  2019-02-08 00:00:00              WEB
2356190    110110141        4  2022-01-25 00:00:00            skype
...              ...      ...                  ...              ...
1429931  45010187328        2  2016-09-02 00:00:00            PHONE
551107    7725014501        0  2020-08-26 00:00:00              WEB
2171381         3728        0  2019-12-30 00:00:00            skype
2084310      7305225        0  2019-09-05 00:00:00              APP
1810004     48543101        0  2022-03-16 00:00:00              WEB

              enrolled date  gender  age       payor
2534515  2019-06-24 05:00:00       0    1  Commercial
2360387  2016-07-20 05:00:00       0    2  Commercial
1002182  2020-05-05 05:00:00       1    1  Commercial
838369   2019-04-03 05:00:00       1    2  Commercial
2356190  2016-08-04 05:00:00       0    1  Commercial
...                      ...     ...  ...         ...
1429931  2015-10-29 05:00:00       1    2  Commercial
551107   2020-08-12 05:00:00       0    2  Commercial
2171381  2017-07-12 05:00:00       0    3  Commercial
2084310  2018-05-08 05:00:00       0    0  Commercial
1810004  2016-04-19 05:00:00       0    2  Commercial

[100 rows x 8 columns]
```

In [18]:
```
# contingency table
contingency_table=pd.crosstab(sample1["age"],sample1["service"])
contingency_table
```

Out[18]:

service	0	1	2	3	4	5	7	8
age								
0	2	0	0	0	0	0	0	0
1	10	0	6	0	2	5	0	5
2	23	2	12	1	4	7	3	10
3	3	0	3	0	0	2	0	0

In [ ]:
```

```

In [19]:
```
# Chisquare test
p=ss.chi2_contingency(contingency_table)
p
```

Out[19]:
```
(11.273647829250546,
 0.9571891401359497,
 21,
 array([[7.600e-01, 4.000e-02, 4.200e-01, 2.000e-02, 1.200e-01, 2.800e-01,
         6.000e-02, 3.000e-01],
        [1.064e+01, 5.600e-01, 5.880e+00, 2.800e-01, 1.680e+00, 3.920e+00,
         8.400e-01, 4.200e+00],
        [2.356e+01, 1.240e+00, 1.302e+01, 6.200e-01, 3.720e+00, 8.680e+00,
         1.860e+00, 9.300e+00],
        [3.040e+00, 1.600e-01, 1.680e+00, 8.000e-02, 4.800e-01, 1.120e+00,
         2.400e-01, 1.200e+00]]))
```

p-value : 0.9571891401359497 There is an association between age and service (or) Age column is dependent on Service column

In [ ]:
```
```