

Disclaimer

This is a **practice notebook** that I did as a way to refresh myself with `numpy`, `pandas`, and `scipy.stats`. This is in **NO WAY A VALID STUDY**. The dataset was downloaded from Kaggle and as created by [ruslankl](#). Methodology of the said survey is found [here](#). Again, this entire notebook is **NO WAY A VALID STUDY**. This was msostly done for personal practice and upskilling.

Introduction



The LGBTQIA+ community continues to fight for their rights across parts of the world as it affects aspects of their lives. From Healthcare to Marriage, these sectors do not cater to the community at large. An unspoken aspect of queer and trans lives that is not spoken about is Employment. Many queer and trans people face discrimination in their lives. This is not new to countries inside the European Union (EU). Hence, this data report examined the following questions:

1. Is there any legal provision providing safety against discrimination in Employment for the LGBTQIA+ community within the EU?
2. Are the responses of the respondents correlated to their subset within the LGBTQIA+ community within the EU?
3. Are the responses of the respondents correlated to their country of origin within the EU?

Dataset and Libraries used

The main language used to analyze the data is Python. Using the libraries `numpy`, `pandas`, and `matplotlib.pyplot`, the writer can get a descriptive analysis of the dataset `LGBT_Survey_RightsAwareness.csv`. Likewise, the `researchpy` library and the `scipy.stats` library were imported to gather an inferential analysis of the dataset.

```
In [29]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import researchpy as rp
import scipy.stats as stats

df = pd.read_csv("LGBT_Survey_RightsAwareness.csv")

df.head()
```

```
Out[29]:
```

	CountryCode	subset	question_code	question_label	answer	percentage	notes
0	Austria	Lesbian	d1	In the country where you live, is there a law ...	Yes	53	NaN
1	Austria	Lesbian	d1	In the country where you live, is there a law ...	No	10	NaN
2	Austria	Lesbian	d1	In the country where you live, is there a law ...	Don't know	37	NaN
3	Austria	Gay	d1	In the country where you live, is there a law ...	Yes	45	NaN
4	Austria	Gay	d1	In the country where you live, is there a law ...	No	16	NaN

The dataset was cleaned of any null values using the `.dropna` function to assure the accuracy of the dataset

```
In [49]: df_clean = df.dropna()
df_clean.head()
```

Out[49]:

	CountryCode	subset	question_code	question_label	answer	percentage	notes
16	Belgium	Lesbian	d1	In the country where you live, is there a law ...	No	4	[0]
22	Belgium	Bisexual women	d1	In the country where you live, is there a law ...	No	6	[0]
25	Belgium	Bisexual men	d1	In the country where you live, is there a law ...	No	9	[0]
28	Belgium	Transgender	d1	In the country where you live, is there a law ...	No	9	[0]
40	Bulgaria	Bisexual men	d1	In the country where you live, is there a law ...	No	20	[0]

Survey Questionnaire

The following is the complete questionnaire sent to the respondents.

```
In [7]: questions = df_clean.loc[:, 'question_label']
q_uni = questions.unique()

print(q_uni)
```

['In the country where you live, is there a law that forbids discrimination against persons because of their sexual orientation when applying for a job?']

'In the country where you live, is there a law that forbids discrimination against persons because of their gender identity when applying for a job?'

'Do you know of any organisation in the country where you live that can offer support or advice to people who have been discriminated against because they are Lesbian?'

'Do you know of any organisation in the country where you live that can offer support or advice to people who have been discriminated against because they are Gay?'

'Do you know of any organisation in the country where you live that can offer support or advice to people who have been discriminated against because they are Bisexual?'

'Do you know of any organisation in the country where you live that can offer support or advice to people who have been discriminated against because they are Transgender?'

'In the country where you live, have you ever seen any programme or awareness campaign by either the government or a non-governmental organisation addressing - Discrimination against gay, lesbian and bisexual people?'

'In the country where you live, have you ever seen any programme or awareness campaign by either the government or a non-governmental organisation addressing - Discrimination against transgender people?'

'In the country where you live, have you ever seen any programme or awareness campaign by either the government or a non-governmental organisation addressing - Discrimination on the basis of gender?'

'As far as you know, can same-sex couples legally marry and/or enter registered partnerships in the country where you live?']

For brevity of this report, the writer has only focused on the first two questions (tagged as **d1** and **d2** in the dataset).

Demographic of the Sample

The following are the total respondents based on their subsets within the LGBTQIA+ community. Bisexual Men

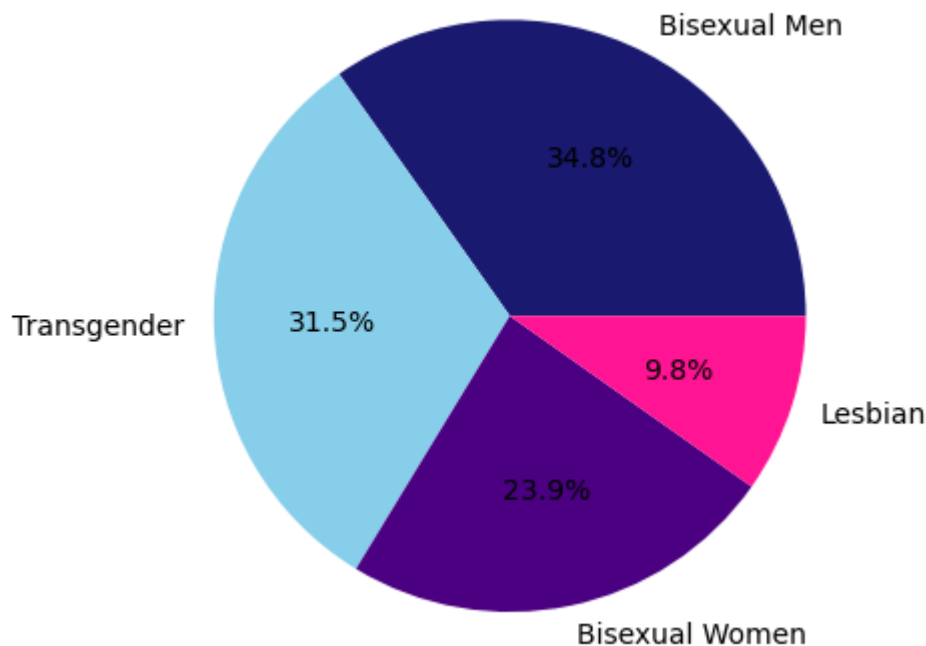
```
In [19]: df_clean = df.dropna()
comm_df = df_clean.loc[df['question_code'] == 'd1']

comm_x = ["Bisexual Men", "Transgender", "Bisexual Women", "Lesbian"]
comm_y = comm_df.loc[:, 'subset'].value_counts()
comm_colors = ['midnightblue', 'skyblue', 'indigo', 'deeppink',]

def comm_pi():
    plt.pie(
        x = comm_y,
        labels = comm_x,
        colors=comm_colors,
        autopct='%1.1f%%'
    )

print(comm_y)
comm_pi()
```

```
Bisexual men      32
Transgender       29
Bisexual women    22
Lesbian           9
Name: subset, dtype: int64
```



```
In [20]: import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv("LGBT_Survey_RightsAwareness.csv")
df_clean = df.dropna()

country_df = df_clean.loc[df['question_code'] == 'd1']
```

```

country_x = ["Cyprus", "Luxembourg", "Malta",
             "Slovenia", "Estonia", "Latvia",
             "Lithuania", "Sweden", "Belgium",
             "Ireland", "Denmark", "Finland",
             "Netherlands", "Czech Republic", "Croatia",
             "Slovakia", "Bulgaria", "Portugal", "Romania",
             "United Kingdom"]
country_y = country_df.loc[:, 'CountryCode'].value_counts()

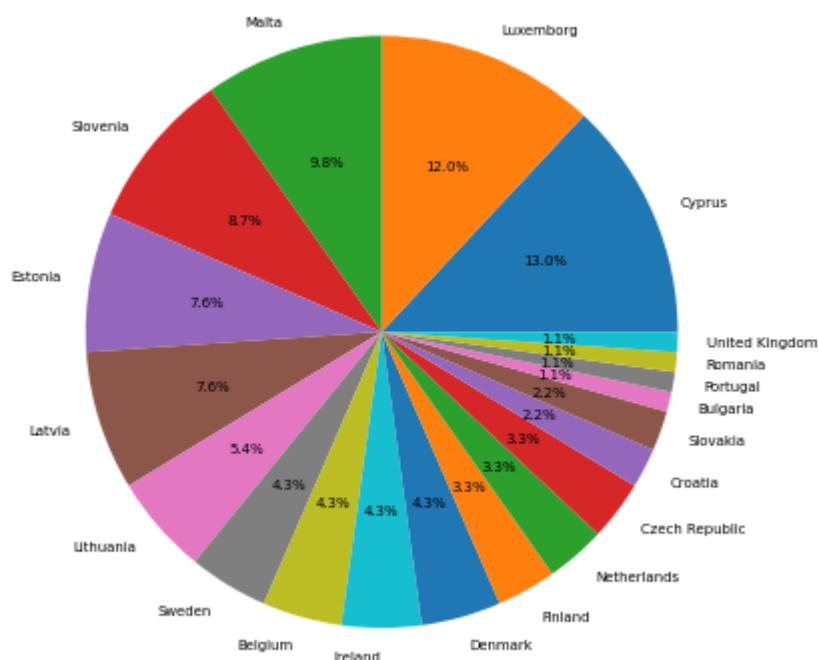
def country_pie() :
    plt.pie(country_y, labels = country_x, autopct='%1.1f%%', textprops={'fontsize': 5})

print(country_y)
country_pie()

```

Cyprus	12
Luxembourg	11
Malta	9
Slovenia	8
Estonia	7
Latvia	7
Lithuania	5
Sweden	4
Belgium	4
Ireland	4
Denmark	4
Finland	3
Netherlands	3
Czech Republic	3
Croatia	2
Slovakia	2
Bulgaria	1
Portugal	1
Romania	1
United Kingdom	1

Name: CountryCode, dtype: int64



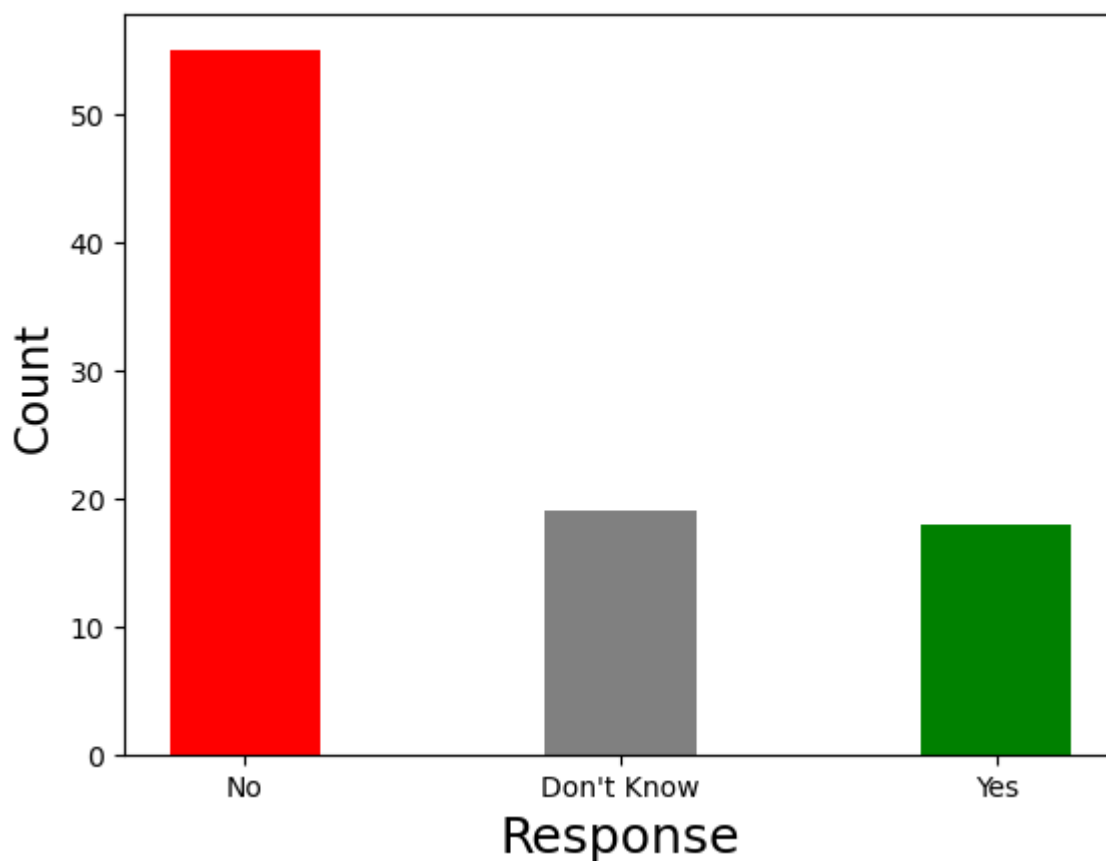
On Workplace Discrimination

```
In [26]: job_disc_2 = df_clean.loc[df['question_code'] == 'd1']
job_disc_a2 = job_disc_2.loc[:, 'answer'].value_counts()
job_disc_l2 = ["No", "Don't Know", "Yes"]
colors = ["Red", "Grey", "Green"]

def job_disc1() :
    plt.bar(job_disc_l2, job_disc_a2, color=colors, width=0.4)
    plt.xlabel("Response", fontsize=18)
    plt.ylabel("Count", fontsize=16)
    plt.show()

print(job_disc_a2)
job_disc1()
```

```
No          55
Don't know  19
Yes         18
Name: answer, dtype: int64
```

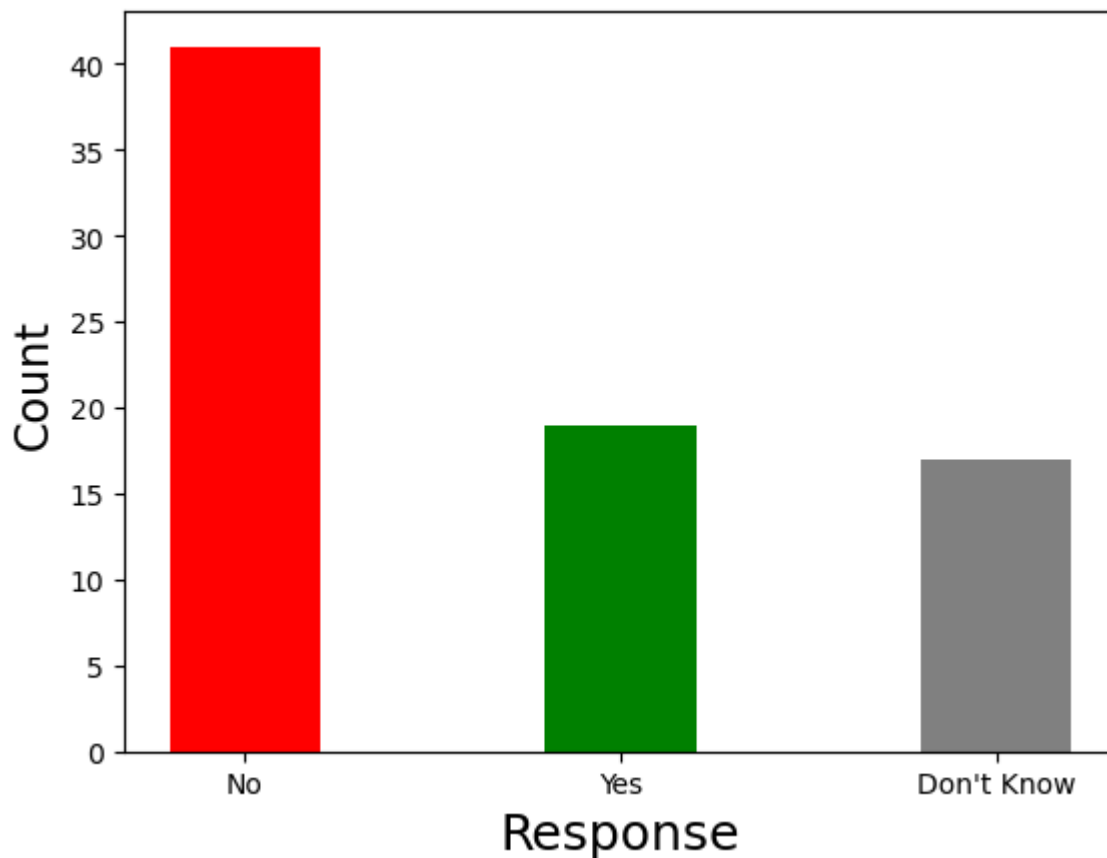


```
In [35]: job_disc_3 = df_clean.loc[df['question_code'] == 'd2']
job_disc_a3 = job_disc_3.loc[:, 'answer'].value_counts()
job_disc_l3 = ["No", "Yes", "Don't Know"]
colors = ["Red", "Green", "Grey"]

def job_disc1() :
    plt.bar(job_disc_l3, job_disc_a3, color=colors, width=0.4)
    plt.xlabel("Response", fontsize=18)
    plt.ylabel("Count", fontsize=16)
    plt.show()

print(job_disc_a3)
job_disc1()
```

```
No          41
Yes          19
Don't know  17
Name: answer, dtype: int64
```



Correlation with Subset within LGBTQIA+ Community

After performing a cross-tabulation of the `subset` with the `answer` column of the dataframe (filtered only to answers under the question code `d1`), the writer has performed a Chi-Square test of Independence.

Hypotheses

H_0 = There is no correlation between the perception of the respondents regarding laws protecting discrimination against their sexual orientation and their subset within the LGBTQIA+ community.

H_a = There is a correlation between the perception of the respondents regarding laws protecting discrimination against their sexual orientation and their subset within the LGBTQIA+ community.

$p - value = 0.05$

```
In [6]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import researchpy as rp
import scipy.stats as stats

df = pd.read_csv("LGBT_Survey_RightsAwareness.csv")
df_clean = df.dropna()
```



```
d1_df = df_clean.loc[df['question_code'] == 'd1']
ctab1 = pd.crosstab(d1_df.subset, d1_df.answer)
ctab1_cs = stats.chi2_contingency(ctab1)

print(ctab1)
print(" ")
print(ctab1_cs)
```

answer	Don't know	No	Yes
subset			
Bisexual men	8	18	6
Bisexual women	3	16	3
Lesbian	1	6	2
Transgender	7	15	7

```
Chi2ContingencyResult(statistic=3.2028929878817136, pvalue=0.7829846084478523, dof=6,
expected_freq=array([[ 6.60869565, 19.13043478,  6.26086957],
[ 4.54347826, 13.15217391,  4.30434783],
[ 1.85869565,  5.38043478,  1.76086957],
[ 5.98913043, 17.33695652,  5.67391304]]))
```

As findings show that the responses in `d1` have a p – *value* of 0.78, the null hypothesis is thus rejected. Therefore, there is a correlation between the perception of the respondents regarding laws protecting discrimination against their sexual orientation and their subset withing the LGBTQIA+ community.

We then performed a cross-tabulation of the `subset` with the `answer` column of the dataframe filtered only to answers under the question code `d2` and performed the same procedures.

Hypotheses

H_0 = There is no correlation between the perception of the respondents regarding laws protecting discrimination against their gender identity and their subset within the LGBTQIA+ community.

H_a = There is a correlation between the perception of the respondents regarding laws protecting discrimination against their gender identity and their subset within the LGBTQIA+ community.

p – *value* = 0.05

```
In [7]: df = pd.read_csv("LGBT_Survey_RightsAwareness.csv")
df_clean = df.dropna()
d2_df = df_clean.loc[df['question_code'] == 'd2']
ctab2 = pd.crosstab(d2_df.subset, d2_df.answer)
ctab2_cs = stats.chi2_contingency(ctab2)

print(ctab2)
print(" ")
print(ctab2_cs)
```

answer subset	Don't know	No	Yes
Bisexual men	6	17	6
Bisexual women	3	14	4
Lesbian	1	3	2
Transgender	7	7	7

```
Chi2ContingencyResult(statistic=5.607260743259722, pvalue=0.46858874115444327, dof=6,
expected_freq=array([[ 6.4025974 , 15.44155844,  7.15584416],
[ 4.63636364, 11.18181818,  5.18181818],
[ 1.32467532,  3.19480519,  1.48051948],
[ 4.63636364, 11.18181818,  5.18181818]]))
```

As findings show that the responses in `d2` have a p – *value* of 0.46, the null hypothesis is thus rejected. Therefore, there is a correlation between the perception of the respondents regarding laws protecting discrimination against their gender identity and their subset withing the LGBTQIA+ community.

Correlation with Country of Origin

After performing a cross-tabulation of the `CountryCode` with the `answer` column of the dataframe (filtered only to answers under the question code `d1`), the writer has performed a Chi-Square test of Independence.

Hypotheses

H_0 = There is no correlation between the perception of the respondents regarding laws protecting discrimination against their sexual orientation and their country of origin.

H_a = There is a correlation between the perception of the respondents regarding laws protecting discrimination against their sexual orientation and their country of origin.

p – *value* = 0.05

```
In [10]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import researchpy as rp
import scipy.stats as stats

df = pd.read_csv("LGBT_Survey_RightsAwareness.csv")
df_clean = df.dropna()
d1_df = df_clean.loc[df['question_code'] == 'd1']
ctab1a = pd.crosstab(d1_df.CountryCode, d1_df.answer)
ctab1a_cs = stats.chi2_contingency(ctab1a)

print(ctab1a)
print(" ")
print(ctab1a_cs)
```

answer	Don't know	No	Yes
CountryCode			
Belgium	0	4	0
Bulgaria	0	1	0
Croatia	0	2	0
Cyprus	4	4	4
Czech Republic	0	3	0
Denmark	1	3	0
Estonia	2	3	2
Finland	0	3	0
Ireland	1	3	0
Latvia	2	3	2
Lithuania	1	3	1
Luxembourg	3	4	4
Malta	3	3	3
Netherlands	0	3	0
Portugal	0	1	0
Romania	0	1	0
Slovakia	0	2	0
Slovenia	2	4	2
Sweden	0	4	0
United Kingdom	0	1	0

```
Chi2ContingencyResult(statistic=29.72285102103284, pvalue=0.8291200485430188, dof=38,
expected_freq=array([[0.82608696, 2.39130435, 0.7826087 ],
[0.20652174, 0.59782609, 0.19565217],
[0.41304348, 1.19565217, 0.39130435],
[2.47826087, 7.17391304, 2.34782609],
[0.61956522, 1.79347826, 0.58695652],
[0.82608696, 2.39130435, 0.7826087 ],
[1.44565217, 4.18478261, 1.36956522],
[0.61956522, 1.79347826, 0.58695652],
[0.82608696, 2.39130435, 0.7826087 ],
[1.44565217, 4.18478261, 1.36956522],
[1.0326087 , 2.98913043, 0.97826087],
[2.27173913, 6.57608696, 2.15217391],
[1.85869565, 5.38043478, 1.76086957],
[0.61956522, 1.79347826, 0.58695652],
[0.20652174, 0.59782609, 0.19565217],
[0.20652174, 0.59782609, 0.19565217],
[0.41304348, 1.19565217, 0.39130435],
[1.65217391, 4.7826087 , 1.56521739],
[0.82608696, 2.39130435, 0.7826087 ],
[0.20652174, 0.59782609, 0.19565217]]))
```

As findings show that the responses in `d1` have a p – *value* of 0.82, the null hypothesis is thus rejected. Therefore, there is a correlation between the perception of the respondents regarding laws protecting discrimination against their sexual orientation and their country of origin.

We then performed a cross-tabulation of the `CountryCode` with the `answer` column of the dataframe filtered only to answers under the question code `d2` and performed the same procedures.

Hypotheses

H_0 = There is no correlation between the perception of the respondents regarding laws

protecting discrimination against their gender identity and their country of origin.

H_a = There is a correlation between the perception of the respondents regarding laws protecting discrimination against their gender identity and their country of origin.

$p - value = 0.05$

```
In [11]: df = pd.read_csv("LGBT_Survey_RightsAwareness.csv")
df_clean = df.dropna()
d2_df = df_clean.loc[df['question_code'] == 'd2']
ctab2a = pd.crosstab(d2_df.CountryCode, d2_df.answer)
ctab2a_cs = stats.chi2_contingency(ctab2a)

print(ctab2a)
print(" ")
print(ctab2a_cs)
```

answer	Don't know	No	Yes
CountryCode			
Austria	0	1	0
Belgium	0	3	0
Bulgaria	0	1	0
Croatia	0	2	0
Cyprus	4	4	4
Denmark	0	2	0
Estonia	2	3	2
Finland	0	1	0
Ireland	0	1	0
Latvia	2	3	2
Lithuania	1	3	1
Luxembourg	3	4	4
Malta	3	3	3
Netherlands	0	1	0
Romania	0	1	1
Slovakia	0	2	0
Slovenia	2	3	2
Sweden	0	2	0
United Kingdom	0	1	0

```
Chi2ContingencyResult(statistic=21.68425079916434, pvalue=0.9713975834877007, dof=36,
expected_freq=array([[0.22077922, 0.53246753, 0.24675325],
[0.66233766, 1.5974026 , 0.74025974],
[0.22077922, 0.53246753, 0.24675325],
[0.44155844, 1.06493506, 0.49350649],
[2.64935065, 6.38961039, 2.96103896],
[0.44155844, 1.06493506, 0.49350649],
[1.54545455, 3.72727273, 1.72727273],
[0.22077922, 0.53246753, 0.24675325],
[0.22077922, 0.53246753, 0.24675325],
[1.54545455, 3.72727273, 1.72727273],
[1.1038961 , 2.66233766, 1.23376623],
[2.42857143, 5.85714286, 2.71428571],
[1.98701299, 4.79220779, 2.22077922],
[0.22077922, 0.53246753, 0.24675325],
[0.44155844, 1.06493506, 0.49350649],
[0.44155844, 1.06493506, 0.49350649],
[1.54545455, 3.72727273, 1.72727273],
[0.44155844, 1.06493506, 0.49350649],
[0.22077922, 0.53246753, 0.24675325]]))
```

As findings show that the responses in `d2` have a $p - value$ of 0.97, the null hypothesis is thus rejected. Therefore, there is a correlation between the perception of the respondents regarding laws protecting discrimination against their gender identity and their country of origin.

Conclusion

Selected LGBTQIA+ residents within the EU have expressed that there is a lack of legal protection for them when it comes to the workplace discrimination. These perceptions are influenced depending on their subset and their country of origin.