

PROJETO ICD

World Press Freedom Score

Participantes:

- CLEYDSON DE SOUZA FERREIRA JUNIOR
- FELIPE GONTIJO SIQUEIRA
- MARIANA DA SILVA MARTINS

Perguntas:

1. A média de liberdade aumenta ou diminui conforme a segurança do país?
2. Como o fator econômico influencia o contexto sociocultural e o de segurança?
3. O quão relacionados estão os índices de contexto econômico e contexto político? São diretamente ou inversamente proporcionais?

Testes

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
```

```
df = pd.read_csv("World Press Freedom Score.csv")
df = df.set_index('Ranking')
```

df

	Country	Political Context 2022	Economic Context
2022 \ Ranking			
1	Norway	94.89	90.38
2	Denmark	94.34	83.67
3	Sweden	91.96	87.66
4	Estonia	91.11	81.97
5	Finland	90.40	82.03
...
176	Myanmar	40.40	29.25
177	Turkmenistan	25.00	11.73
178	Iran	34.14	21.32

179	Eritrea	22.22	19.73
180	North Korea	22.42	0.00
Legal Framework 2022 Safety Score 2022 Sociocultural Context 2022 \ Ranking			
1	92.23	92.03	
93.71			
2	89.18	92.97	
91.17			
3	90.27	84.14	
90.18			
4	87.87	93.20	
90.00			
5	86.64	92.26	
90.77			
...	
...			
176	20.18	4.63	
30.67			
177	29.61	35.69	
23.00			
178	26.71	13.61	
23.78			
179	25.15	11.36	
19.67			
180	22.81	12.38	
12.00			
Abuse Score 2021 Underlying Situation Score 2021 Global Score 2022 \ Ranking			
1	0.00	6.72	
92.65			
2	0.00	8.57	
90.27			
3	0.00	7.24	
88.84			
4	0.00	15.25	
88.83			
5	0.00	6.99	
88.42			
...	
...			
176	0.00	31.12	

25.03			
177	55.21		41.86
25.01			
178	0.00		29.82
23.22			
179	35.55		43.13
19.62			
180	13.86		21.65
13.92			
	Global Score 2021	Global Score 2020	Global Score 2019
Ranking			
1	6.72	7.84	7.82
2	8.57	8.13	9.87
3	7.24	9.25	8.31
4	15.25	12.61	12.27
5	6.99	7.93	7.90
...
176	31.12	30.20	29.00
177	43.94	42.88	43.98
178	29.82	29.79	29.67
179	43.13	42.69	43.63
180	21.59	22.93	22.23

[180 rows x 12 columns]

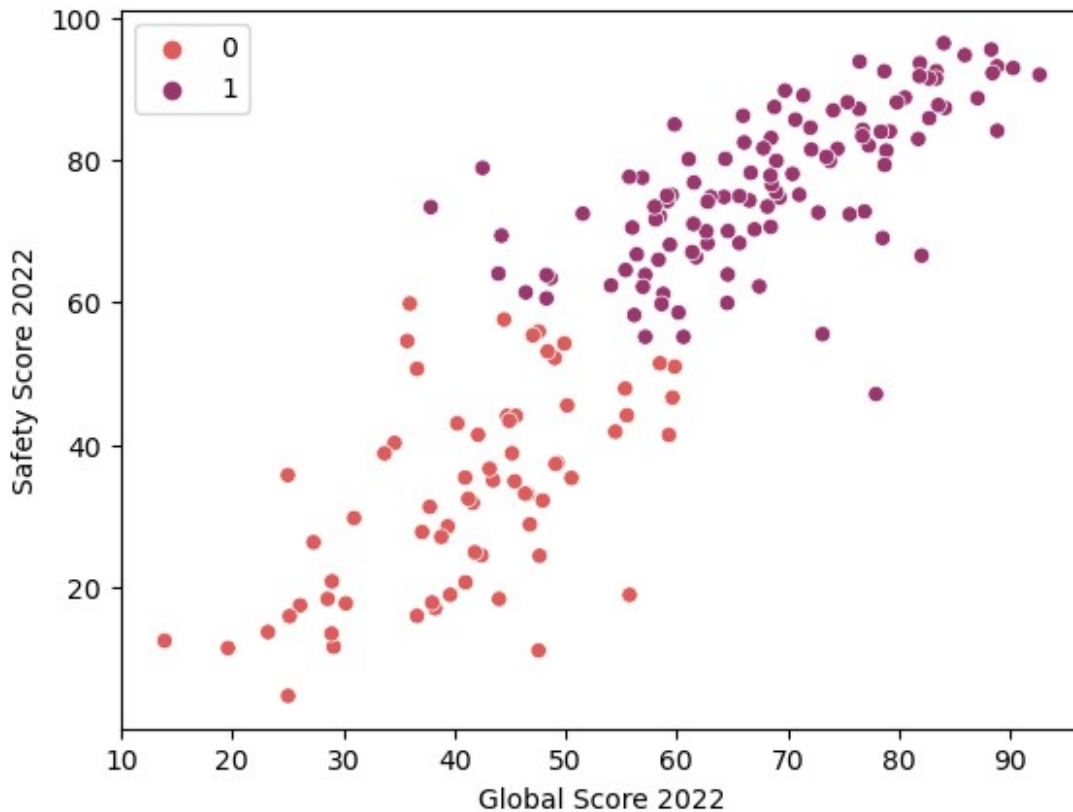
Testes:

- Já de início, no primeiro teste em relação ao Global Score e Safety Score, se percebe uma anomalia, uma vez que, segundo a criadora do dataset, quando maior o Global Score, pior é a condição relacionada ao tema. Dessa forma, analisando o dataset, percebe-se que a coluna do Global score 2022 está invertida. Então, para continuarmos com as nossas análises, vamos invertê-la.

```
df_group = df[['Safety Score 2022', 'Global Score 2022']].dropna()

kmeans = KMeans(n_clusters=2, n_init='auto')
kmeans.fit(df_group)

sns.scatterplot(data = df, x='Global Score 2022', y='Safety Score 2022', hue = kmeans.labels_, palette = 'flare');
```



Inversão da coluna 'Global Score 2022'

- A partir de agora, trabalharemos com esse dataset:

```
a = df.sort_values('Global Score 2022')
a = a['Global Score 2022'].reset_index()
a = a.drop(columns = 'Ranking')
a.index +=1
df['Global Score 2022'] = a
df
```

	Country	Political Context 2022	Economic Context
2022 \ Ranking			
1	Norway	94.89	90.38
2	Denmark	94.34	83.67
3	Sweden	91.96	87.66
4	Estonia	91.11	81.97

5	Finland	90.40	82.03
...
176	Myanmar	40.40	29.25
177	Turkmenistan	25.00	11.73
178	Iran	34.14	21.32
179	Eritrea	22.22	19.73
180	North Korea	22.42	0.00
Legal Framework 2022 Safety Score 2022 Sociocultural Context 2022 \ Ranking			
1	92.23	92.03	
93.71			
2	89.18	92.97	
91.17			
3	90.27	84.14	
90.18			
4	87.87	93.20	
90.00			
5	86.64	92.26	
90.77			
...	
...			
176	20.18	4.63	
30.67			
177	29.61	35.69	
23.00			
178	26.71	13.61	
23.78			
179	25.15	11.36	
19.67			
180	22.81	12.38	
12.00			
Abuse Score 2021 Underlying Situation Score 2021 Global Score 2022 \ Ranking			
1	0.00	6.72	
13.92			
2	0.00	8.57	

19.62		
3	0.00	7.24
23.22		
4	0.00	15.25
25.01		
5	0.00	6.99
25.03		
...
...		
176	0.00	31.12
88.42		
177	55.21	41.86
88.83		
178	0.00	29.82
88.84		
179	35.55	43.13
90.27		
180	13.86	21.65
92.65		

	Global Score 2021	Global Score 2020	Global Score 2019
Ranking			
1	6.72	7.84	7.82
2	8.57	8.13	9.87
3	7.24	9.25	8.31
4	15.25	12.61	12.27
5	6.99	7.93	7.90
...
176	31.12	30.20	29.00
177	43.94	42.88	43.98
178	29.82	29.79	29.67
179	43.13	42.69	43.63
180	21.59	22.93	22.23

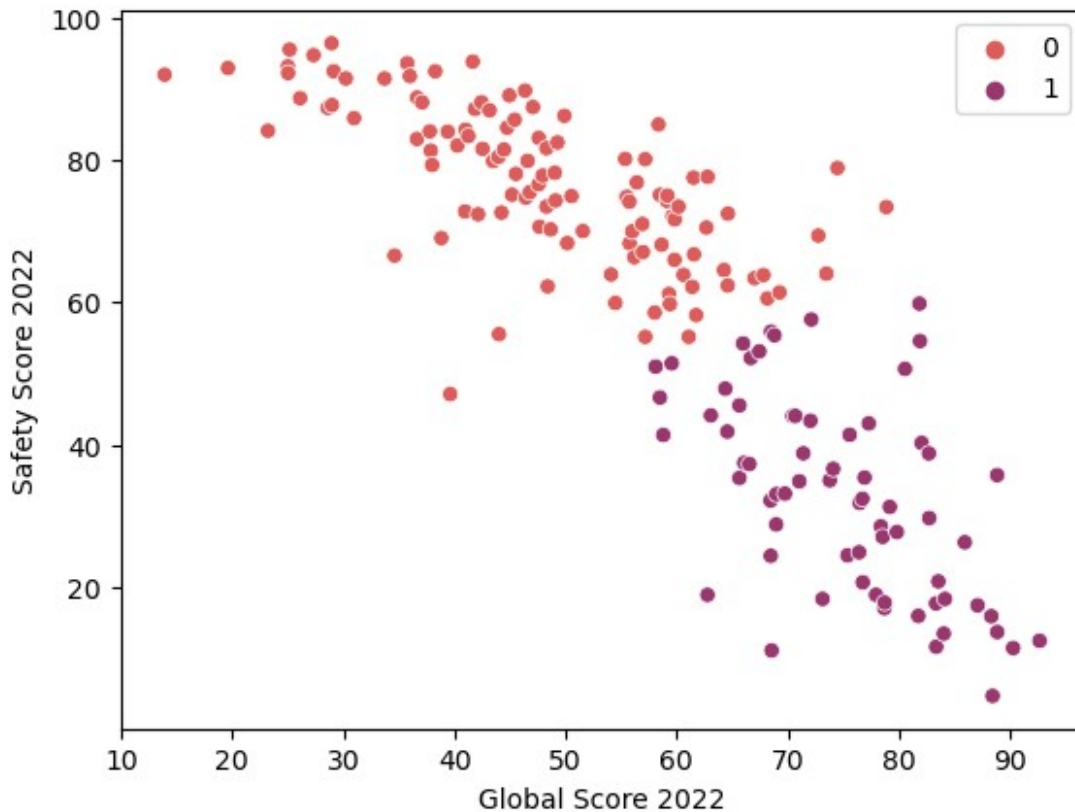
[180 rows x 12 columns]

Testes com os valores corrigidos:

- Após a correção dos valores, podemos observar uma inversão no gráfico, que agora é mais claro e preciso.

```
df_group = df[['Safety Score 2022', 'Global Score 2022']].dropna()
kmeans = KMeans(n_clusters=2, n_init='auto')
kmeans.fit(df_group)

sns.scatterplot(data = df, x='Global Score 2022', y='Safety Score 2022', hue = kmeans.labels_, palette = 'flare');
```



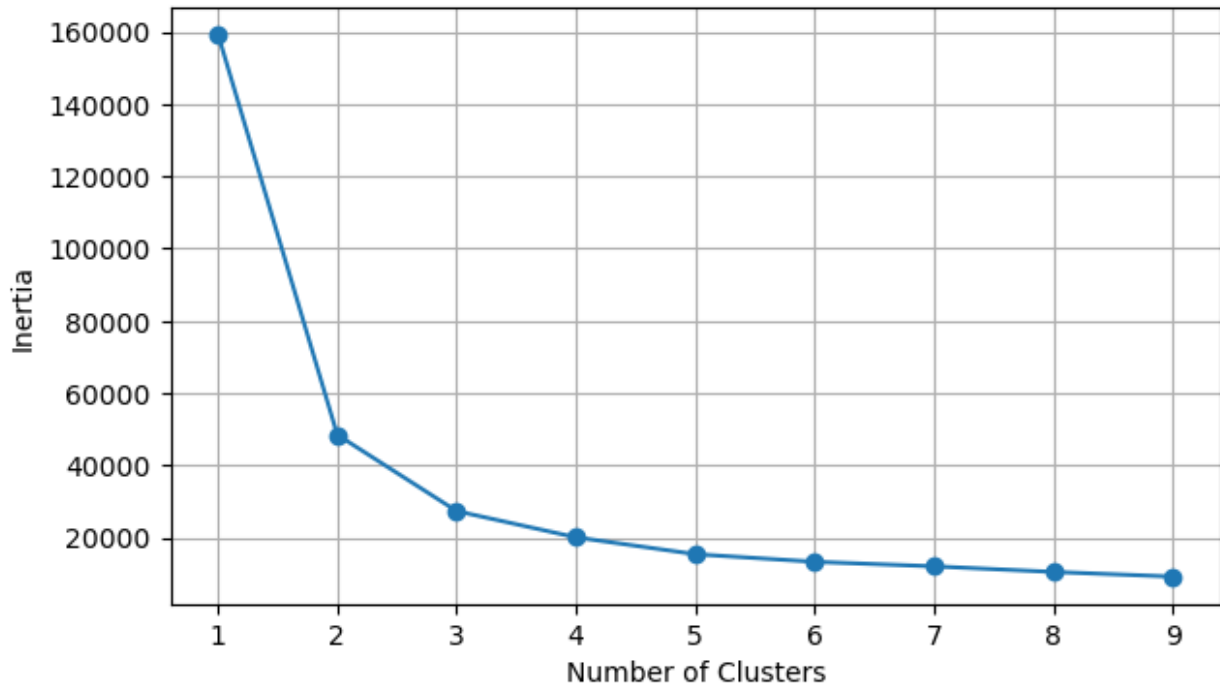
Função 'optimise_kmeans'

Criando função 'optimise_kmeans' para otimizar o número de clusters ao longo do notebook.

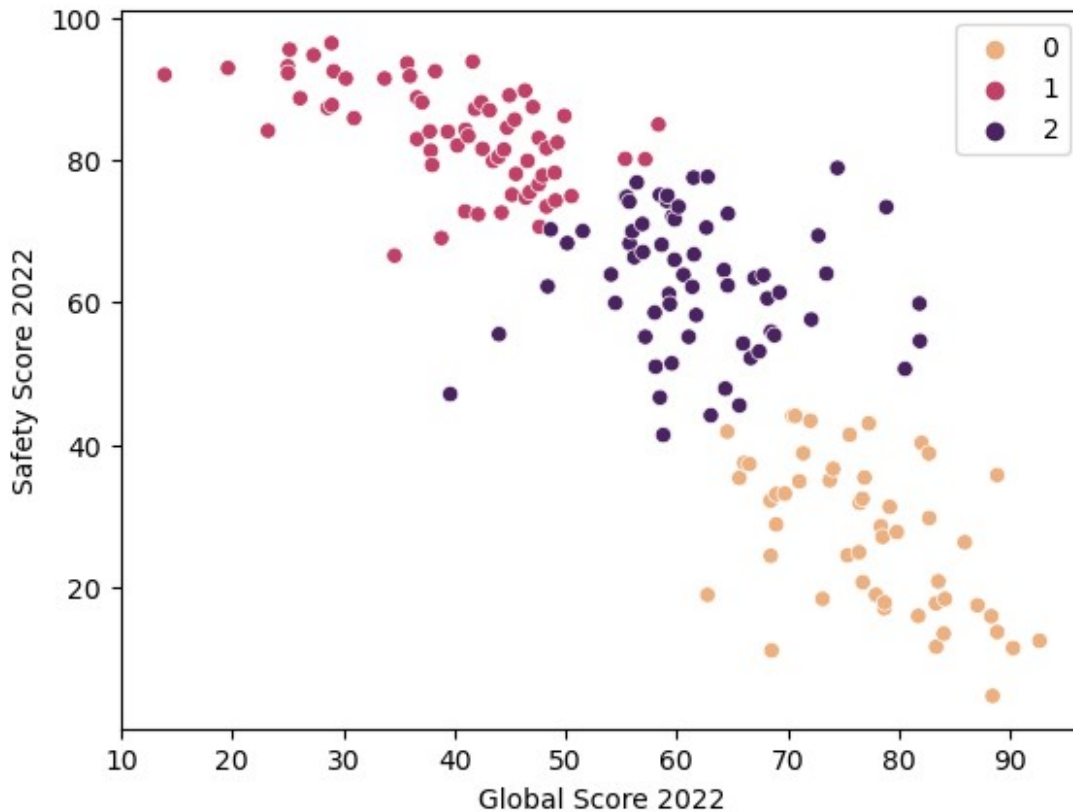
```
def optimise_kmeans(data, max_k):  
  
    means = []  
    inertias = []  
  
    for k in range(1, max_k):  
  
        kmeans = KMeans(n_clusters = k, n_init='auto')  
        kmeans.fit(data)  
  
        means.append(k)  
        inertias.append(kmeans.inertia_)  
  
    fig = plt.subplots(figsize = (7, 4))  
    plt.plot(means, inertias, 'o-')  
    plt.xlabel('Number of Clusters')  
    plt.ylabel('Inertia')  
    plt.grid(True)  
    plt.show()
```

Safety Score 2022 x Global Score 2022

```
optimise_kmeans(df[['Safety Score 2022', 'Global Score 2022']], 10) #  
Scatterplot com Clusters 3
```



```
df_group = df[['Safety Score 2022', 'Global Score 2022']].dropna()  
kmeans = KMeans(n_clusters=3, n_init='auto')  
kmeans.fit(df_group)  
  
sns.scatterplot(data = df, x='Global Score 2022', y='Safety Score  
2022', hue = kmeans.labels_, palette = 'flare');
```

Usando 'SCIPY' para análise relacionada a Regressão Linear

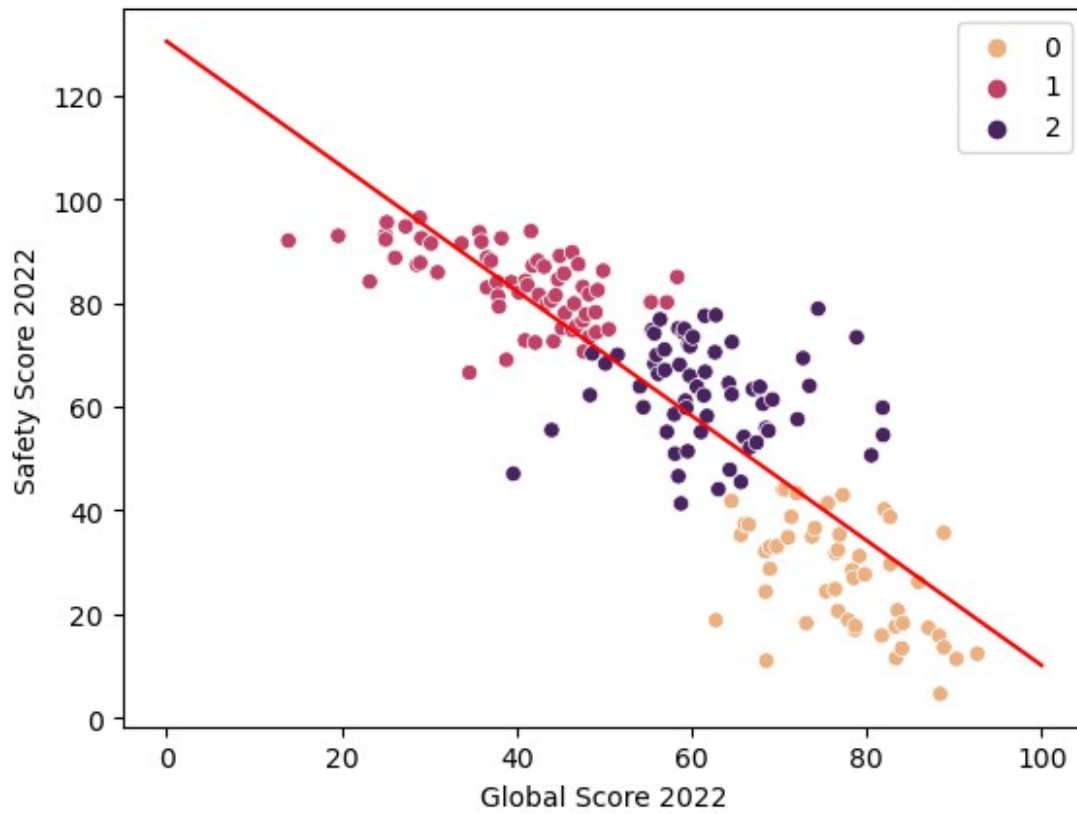
```
from scipy.stats import linregress

def f(x):
    return x*result.slope + result.intercept

result = linregress(df['Global Score 2022'], df['Safety Score 2022'])
result.slope # valor de a
-1.2033007092058736

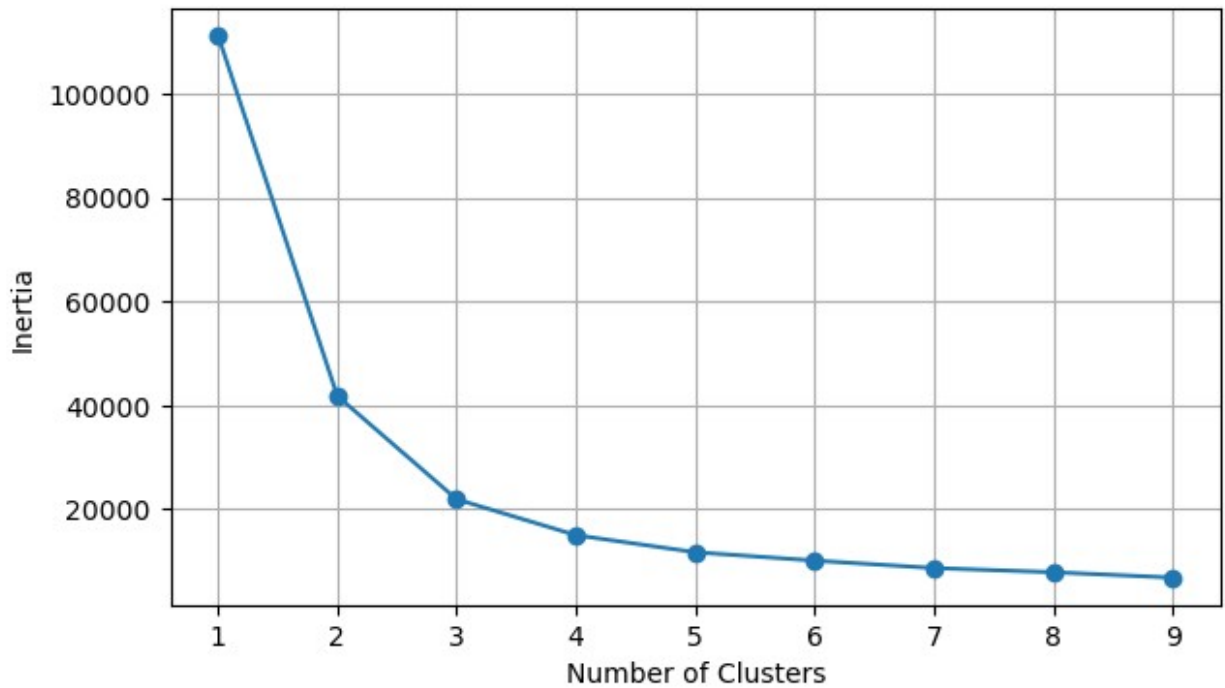
result.intercept # valor de b
130.4727604369319

sns.scatterplot(x= df['Global Score 2022'], y= df['Safety Score
2022'], hue = kmeans.labels_ , palette = 'flare');
plt.plot((0,100), (f(0), f(100)), c = 'red');
```



Political Context 2022 x Legal Framework 2022

```
optimise_kmeans(df[['Political Context 2022', 'Legal Framework 2022']],  
10)
```

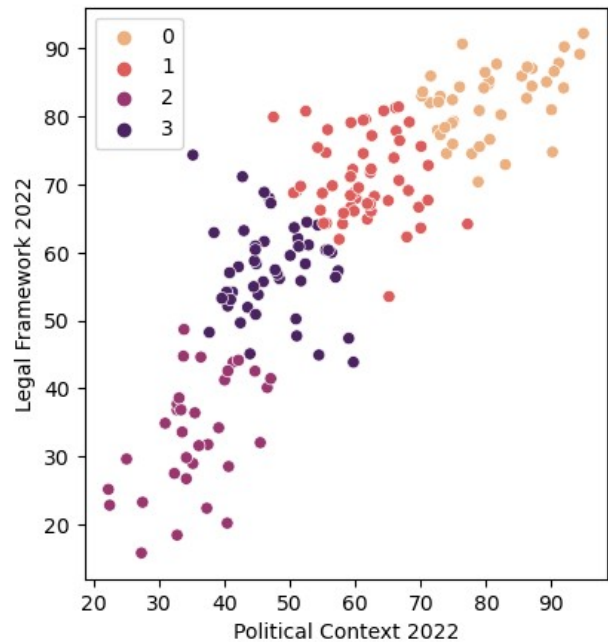
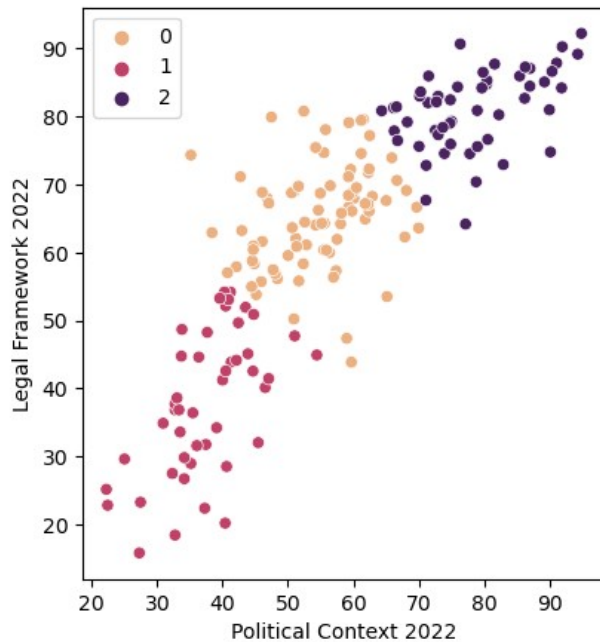


```
# Scatterplot com Clusters 3
kmeans = KMeans(n_clusters = 3, n_init='auto')
kmeans.fit(df[['Political Context 2022', 'Legal Framework 2022']])
klab1 = kmeans.labels_

# Scatterplot com Clusters 4
kmeans = KMeans(n_clusters = 4, n_init='auto')
kmeans.fit(df[['Political Context 2022', 'Legal Framework 2022']])
klab2 = kmeans.labels_

fig, ax = plt.subplots(1, 2, figsize= (10,5))

sns.scatterplot(x= df['Political Context 2022'], y= df['Legal Framework 2022'], hue = klab1, palette = 'flare', ax=ax[0]);
sns.scatterplot(x= df['Political Context 2022'], y= df['Legal Framework 2022'], hue = klab2 , palette = 'flare', ax=ax[1]);
```



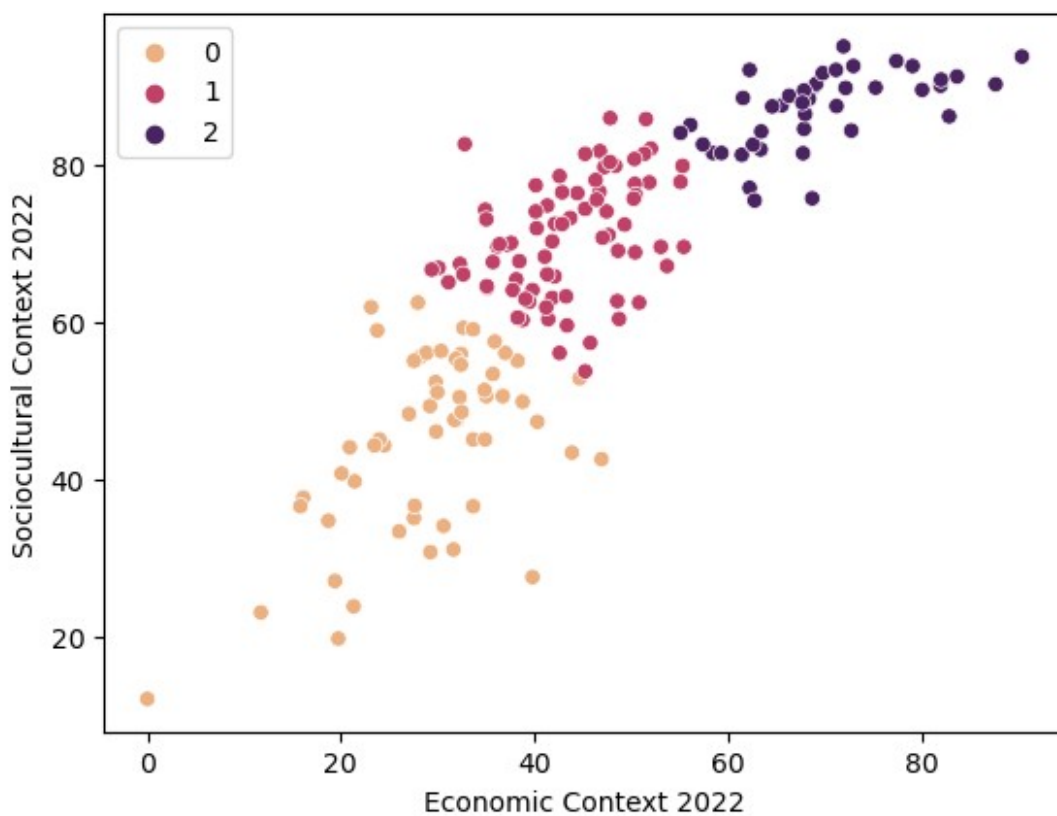
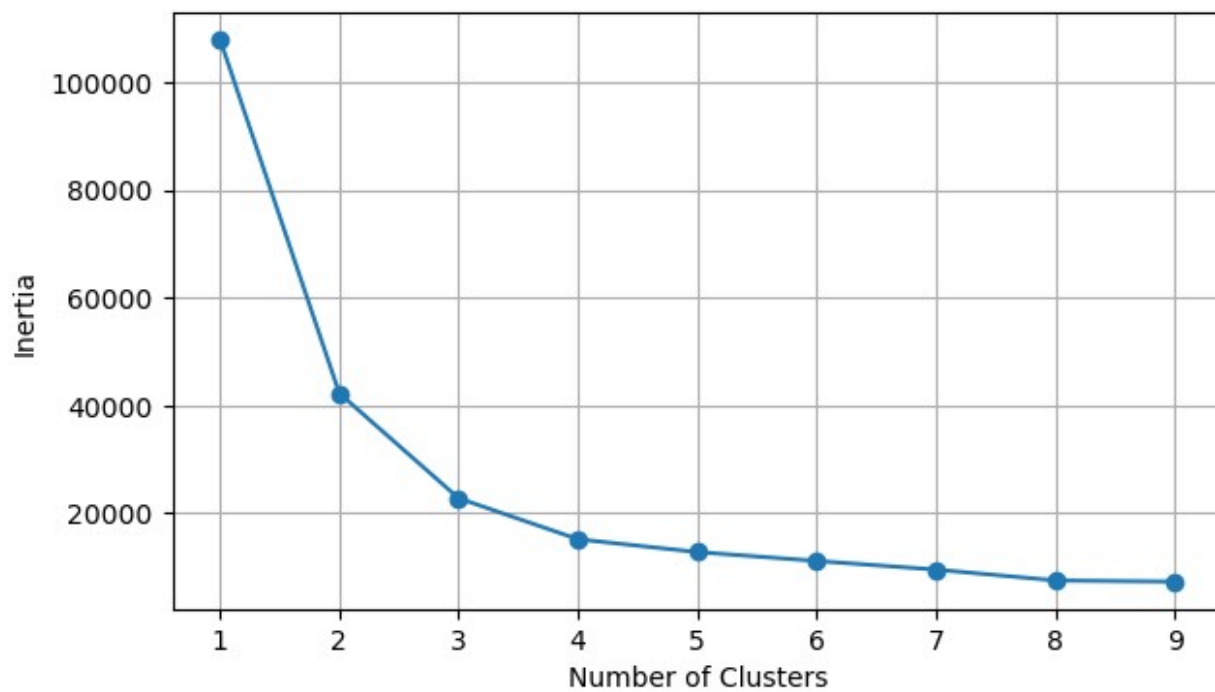
Economic Context 2022 x Sociocultural Context 2022

```
optimise_kmeans(df[['Economic Context 2022', 'Sociocultural Context
2022']], 10) # número de clusters = 3

data = df[['Economic Context 2022', 'Sociocultural Context
2022']].dropna()

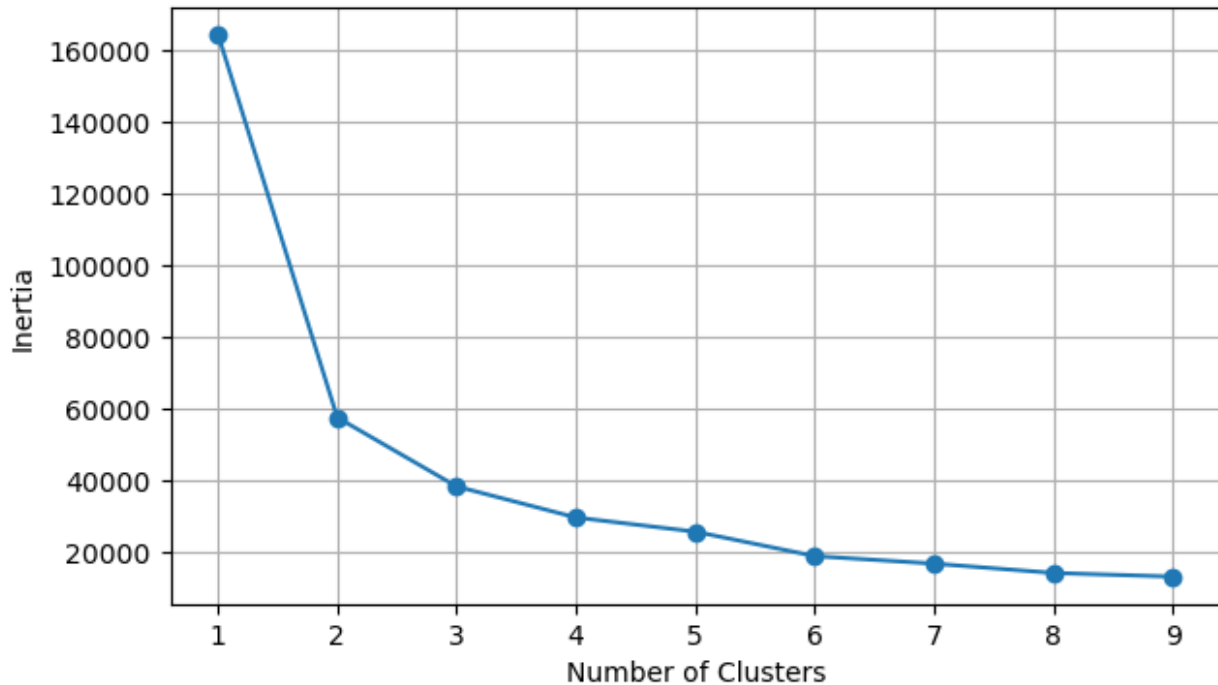
kmeans = KMeans(n_clusters = 3, n_init='auto')
kmeans.fit(data)
klab = kmeans.labels_

sns.scatterplot(x= data['Economic Context 2022'], y=
data['Sociocultural Context 2022'], hue = kmeans.labels_ , palette =
'flare');
```



Safety Score 2022 x Sociocultural Context 2022

```
optimise_kmeans(df[['Safety Score 2022', 'Sociocultural Context 2022']], 10)
```



```
# Scatterplot com Clusters 3
```

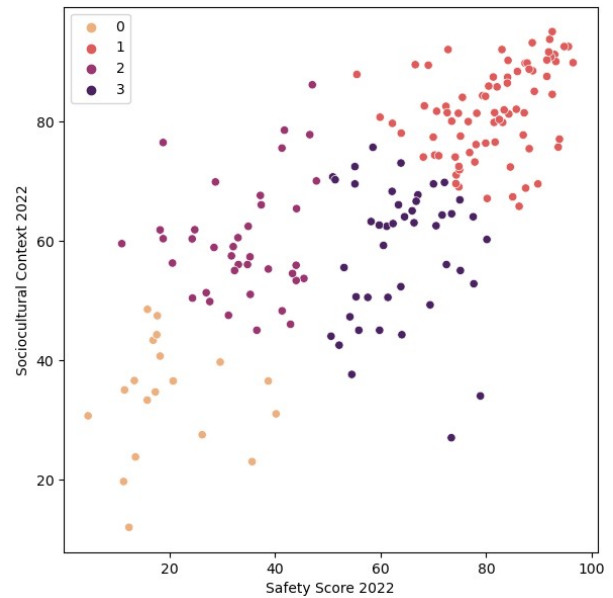
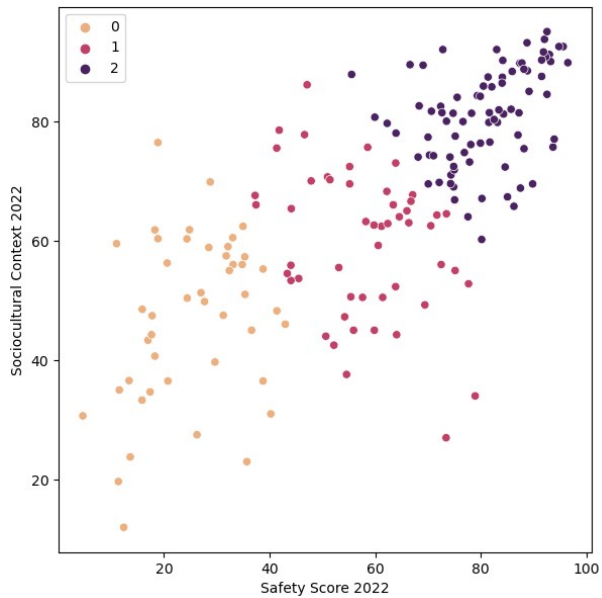
```
kmeans = KMeans(n_clusters = 3, n_init='auto')  
kmeans.fit(df[['Safety Score 2022', 'Sociocultural Context 2022']])  
kmeans1 = kmeans.labels_
```

```
# Scatterplot com Clusters 4
```

```
kmeans = KMeans(n_clusters = 4, n_init='auto')  
kmeans.fit(df[['Safety Score 2022', 'Sociocultural Context 2022']])  
kmeans2 = kmeans.labels_
```

```
fig, ax = plt.subplots(1, 2, figsize= (15, 7))
```

```
sns.scatterplot(x= df['Safety Score 2022'], y= df['Sociocultural Context 2022'], hue = kmeans1, palette = 'flare', ax=ax[0]);  
sns.scatterplot(x= df['Safety Score 2022'], y= df['Sociocultural Context 2022'], hue = kmeans2, palette = 'flare', ax=ax[1]);
```



Usando 'SCIPY' para análise relacionada a Regressão Linear

```
result = linregress(df['Safety Score 2022'], df['Sociocultural Context 2022'])
```

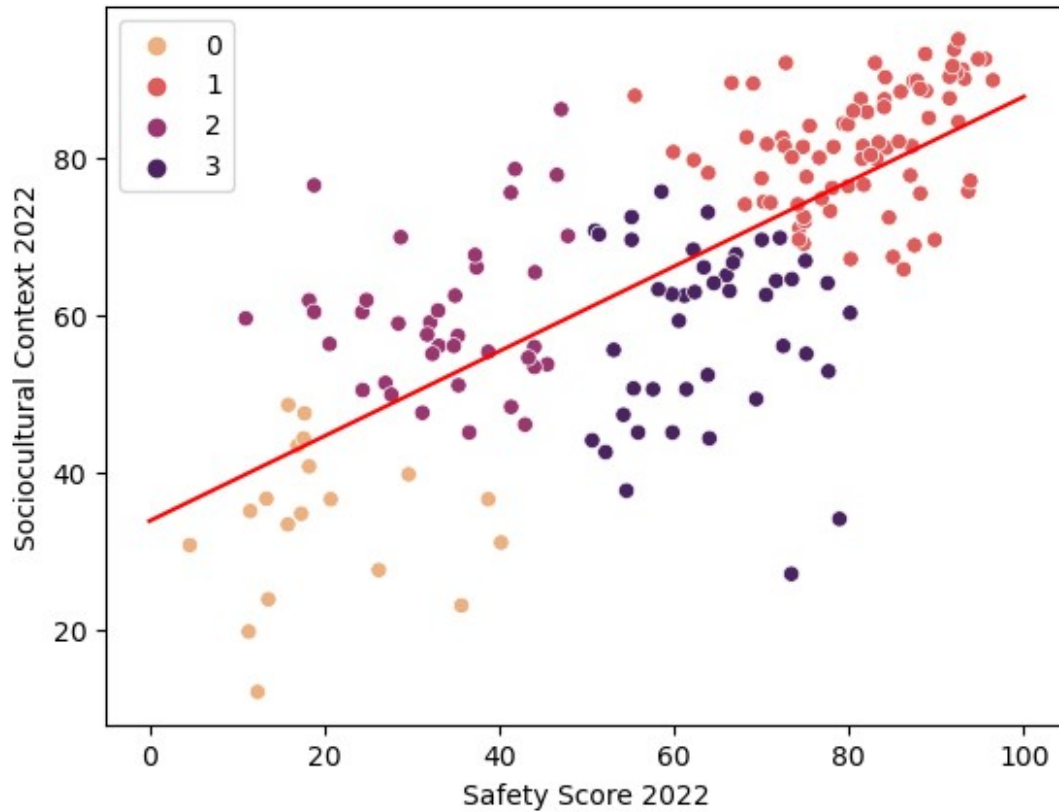
```
result.slope # valor de a
```

```
0.5399413628862646
```

```
result.intercept # valor de b
```

```
33.72303740662127
```

```
sns.scatterplot(x= df['Safety Score 2022'], y= df['Sociocultural Context 2022'], hue = kmeans.labels_, palette = 'flare');  
plt.plot((0,100), (f(0), f(100)), c = 'red');
```

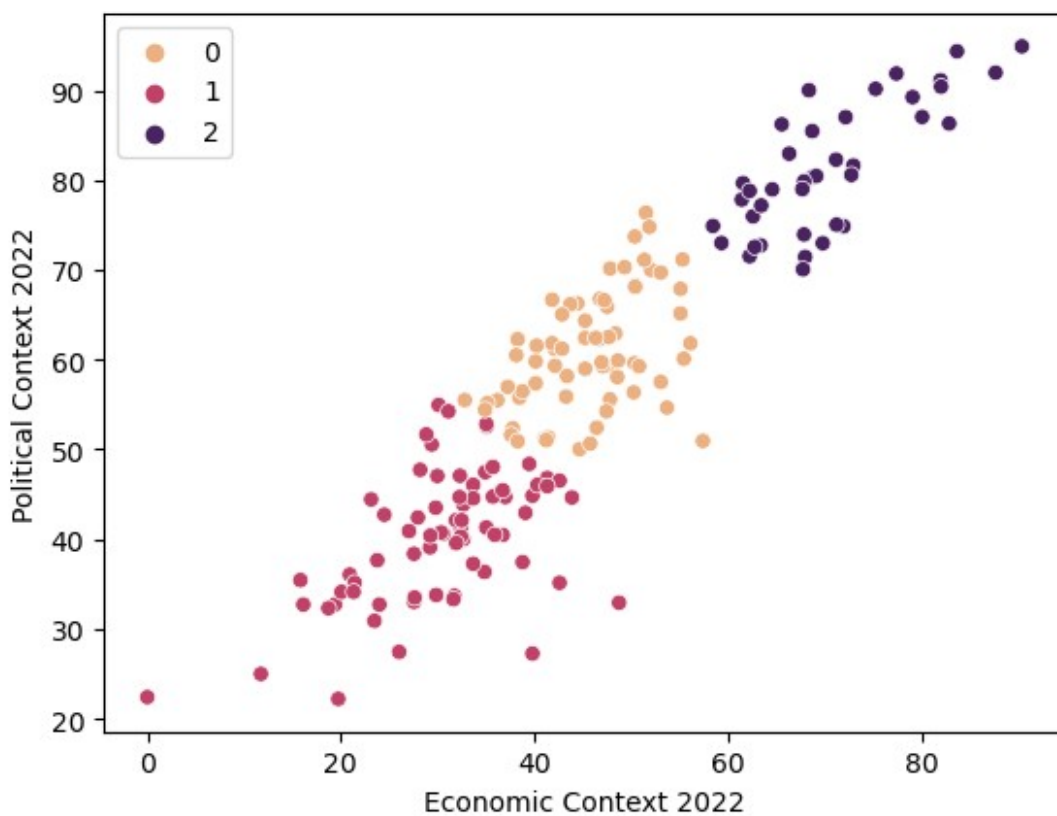
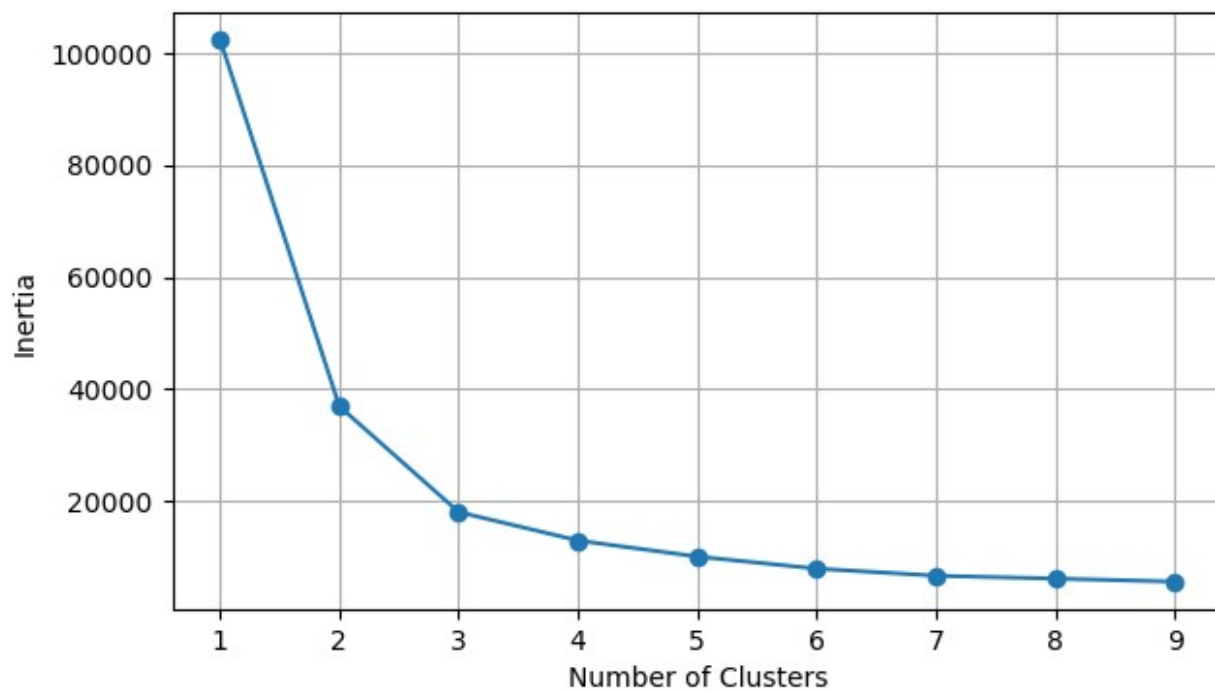


Economic Context 2022 x Political Context 2022

```
optimise_kmeans(df[['Economic Context 2022', 'Political Context
2022']], 10) # número de clusters = 3

kmeans = KMeans(n_clusters = 3, n_init='auto')
kmeans.fit(df[['Economic Context 2022', 'Political Context 2022']])

sns.scatterplot(x= df['Economic Context 2022'], y= df['Political
Context 2022'], hue = kmeans.labels_ , palette = 'flare');
```

Usando 'SCIPY' para análise relacionada a Regressão Linear

```
result = linregress(df['Economic Context 2022'], df['Political Context 2022'])
result.slope

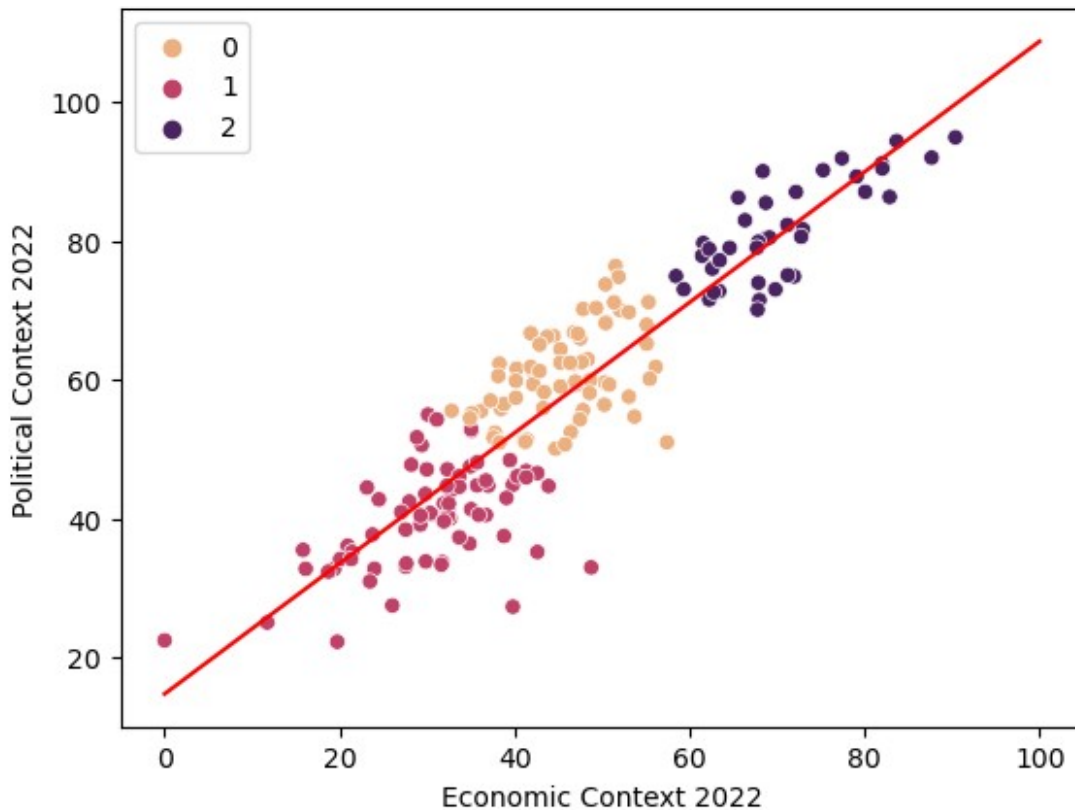
0.9399703167306787

result.intercept

14.710262078100577

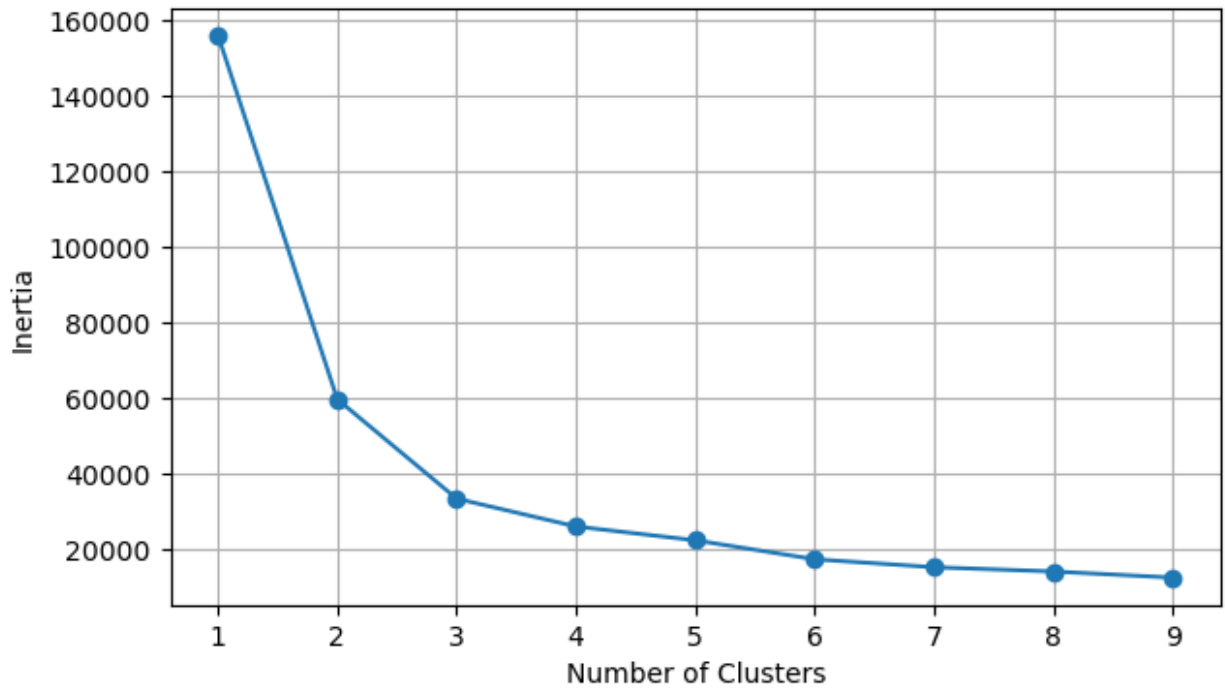
def f(x):
    return x*result.slope + result.intercept

sns.scatterplot(x= df['Economic Context 2022'], y= df['Political Context 2022'], hue = kmeans.labels_ , palette = 'flare');
plt.plot((0,100), (f(0), f(100)), c = 'red');
```



Economic Context 2022 x Safety Score 2022

```
data2 = df[['Economic Context 2022', 'Safety Score 2022']].dropna()
optimise_kmeans(data2[['Economic Context 2022', 'Safety Score 2022']],
10)
```

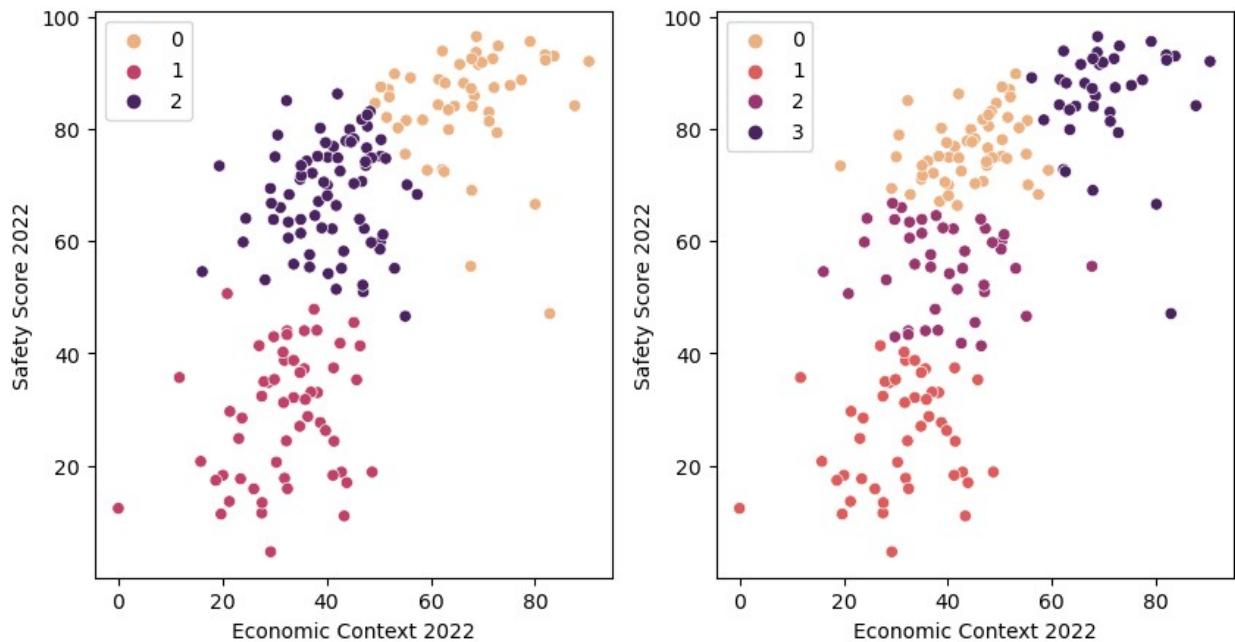


```
# Scatterplot com Clusters 3
kmeans = KMeans(n_clusters = 3, n_init='auto')
kmeans.fit(data2)
klab3 = kmeans.labels_

# Scatterplot com Clusters 4
kmeans = KMeans(n_clusters = 4, n_init='auto')
kmeans.fit(data2)
klab4 = kmeans.labels_

fig, ax = plt.subplots(1, 2, figsize= (10,5))

sns.scatterplot(x= data2['Economic Context 2022'], y= data2['Safety
Score 2022'], hue = klab3, palette = 'flare', ax=ax[0]);
sns.scatterplot(x= data2['Economic Context 2022'], y= data2['Safety
Score 2022'], hue = klab4 , palette = 'flare', ax=ax[1]);
```



Algum padrão foi detectado?

Ao observar as análises feitas acima, percebem-se informações úteis para responder as questões previamente levantadas. Vale ressaltar que todas as análises foram baseadas nos índices do ano de 2022.

1. A média de liberdade aumenta ou diminui conforme a segurança do país?

Os índices de liberdade (Global Score) e de segurança (Security Score) são inversos, sendo a menor pontuação do "Global Score" um indicativo de alto grau de liberdade no país, e a menor pontuação do "Security Score" um indicativo de péssima segurança. Nota-se que o padrão é inversamente proporcional (quanto maior o Security Score, menor o Global Score), e isso conclui que há uma ligação direta entre esses dois valores, pois quando a segurança de um país é melhor, maior é sua liberdade.

1. Como o fator econômico influencia o contexto sociocultural e o de segurança?

Como no caso acima, é perceptível que a economia do país (Economic Context) possui uma relação direta com o contexto sociocultural (Sociocultural Context), de forma a existir uma alta relação entre esses índices.

Já quando analisamos o fator econômico em relação à segurança, observamos um comportamento diferente, pois não há uma relação concreta entre esses dados. Podemos visualizar isso diretamente no modelo de gráfico "Scatterplot", pois as "bolinhas" estão muito dispersas no gráfico.

1. O quão relacionados estão os índices de contexto econômico e contexto político? São diretamente ou inversamente proporcionais?

Esses dois índices possuem uma relação intensa, existindo pouquíssimos valores discrepantes (outliers). Quando analisamos o gráfico Scatterplot com esses dados, observamos esse padrão

de proximidade e, além disso, percebemos que são diretamente proporcionais: ou seja, quanto melhor o contexto político (Political Context) de um país, melhor seu contexto econômico.

Alguma nova pergunta foi descoberta ao analisar os agrupamentos?

A princípio, a pergunta de item 3 seria "Qual país possui o pior contexto econômico juntamente com menor histórico de abuso?", porém, percebemos que sua resposta não seria expressada de forma clara. Por isso, surgiu uma curiosidade ao longo da análise, e observamos um padrão com dois valores fortemente interligados entre si, sendo eles os índices de contexto econômico juntamente com o contexto político. Deste modo, podemos respondemos a seguinte pergunta: "O quão relacionados estão os índices de contexto econômico e contexto político? São diretamente ou inversamente proporcionais?"