

PROJETO ICD - World Press Freedom Score

Participantes:

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

Motivação e Objetivos da análise

Pós ano de eleições, com a vinculação de notícias a todo vapor, foi possível vivenciar e observar a maneira qual a sociedade trata os profissionais da área jornalística. Além disso, diante os constantes ataques contra os canais midiáticos e a crescente guerra cultural e retórica do ódio no Brasil, surgiu a necessidade de comparação com o restante do globo para analisar e compreender se esses eventos se dão apenas no nosso país ou se essa onda de ódio tem tomado conta de todo o mundo.

Perguntas

Dados Usados

Para o projeto, utilizamos o dataset World Press Freedom Score ('Liberdade de imprensa') do kaggle. Este é um conjunto de dados composto por **180 países**, no quais foram ranqueado e avaliados em temas como *Abuse Score, Global Score 2019-22, Political Context, Economic Context, Legal Framework, Safety Score, Sociocultural Context*.

Analizamos os dados e concluiremos o projeto, com o conceito que a liberdade de imprensa é definida como a capacidade dos jornalistas como indivíduos e coletivos de selecionar, produzir e divulgar notícias de interesse público, independentemente de interferência POLÍTICA, ECONÔMICA, LEGAL e SOCIAL e na ausência de ameaças à sua SEGURANÇA física e mental.

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

```
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]			

Pré-processamento

Análise geral:

- Durante o estudo das informações concedidas e a análise dos dados disponíveis, foi possível notar uma anomalia em relação a distribuição dos valores do Global Score em 2022, em que os países de níveis mais altos do ranking se encontravam com scores muito ruins e divergentes dos outros anos. Nesse sentido, para que nossa análise fosse correta e prática, houve a necessidade de inversão da coluna Global Score 2022, afim de que seus valores ficassem na ordem correta na relação.

- Para uma melhor análise, se fez necessário "dropar" os valores nulos de algumas colunas para conseguir tratar os dados com valores faltantes.

```
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]

Análise específica:

- Para obter os dados relacionados ao Brasil, foi necessário fazer uma busca por comparação na coluna "Country" a fim de achar os valores correspondentes ao país, quando obtidos percebeu-se a faltante de "Abuse Score" do ano de 2021 então dropamos tal coluna para que a análise fosse mais limpa.

```
brazil = df[df['Country'].str.match('Brazil', na = False)]
a = brazil.drop(columns = ['Abuse Score 2021'])
dfmm = pd.DataFrame(a)
dfmm.set_index("Country").T
```

Country	Brazil
Political Context 2022	51.62
Economic Context 2022	37.59
Legal Framework 2022	69.74
Safety Score 2022	47.86
Sociocultural Context 2022	70.00
Underlying Situation Score 2021	37.40
Global Score 2022	64.37
Global Score 2021	37.40
Global Score 2020	36.82
Global Score 2019	36.77

- Para otimizar a pesquisa e trazer mais praticidade para a análise definimos uma função para visualização das cluster, relacionado ao elbow method que é uma heurística usada para determinar o número de clusters em um conjunto de dados. Com esse método também conseguimos definir e simplificar a análise de determinadas colunas da base de dados.

```
from sklearn.cluster import KMeans
def optimise_kmeans(data, max_k):

    means = []
    inertias = []

    for k in range(1, max_k):
```

```

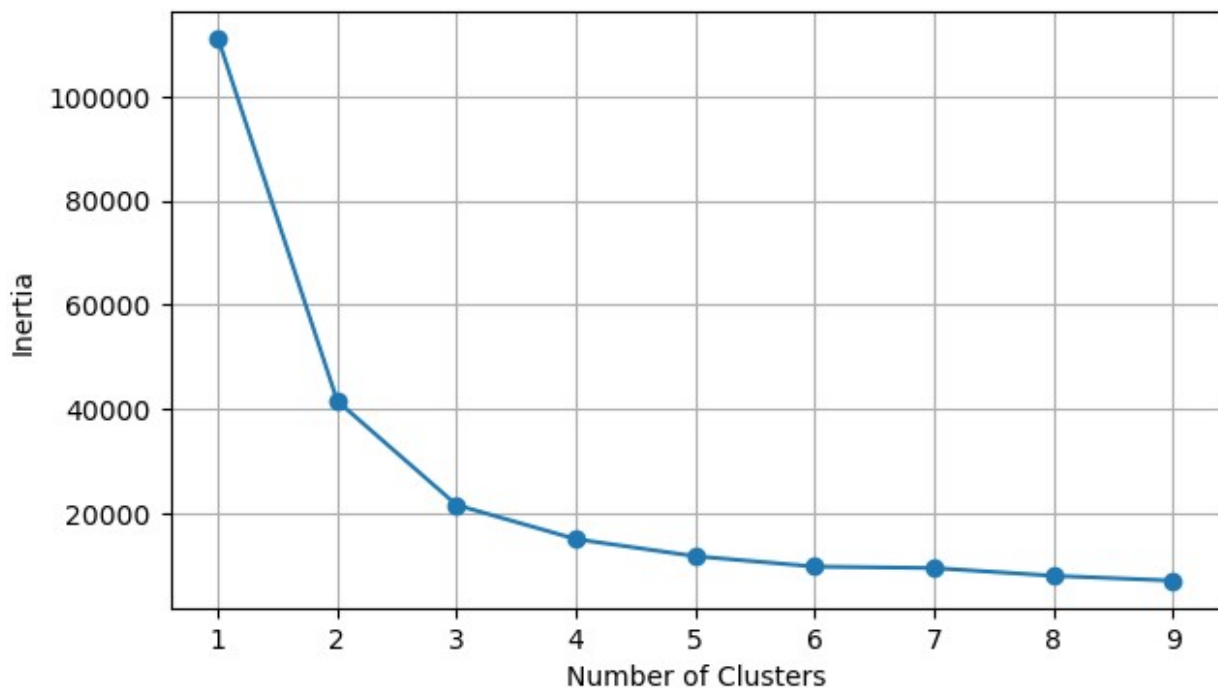
kmeans = KMeans(n_clusters = k, n_init='auto')
kmeans.fit(data)

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()

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

```



Conclusão

Após nossas análises partidas de algumas hipóteses, observamos certos padrões e, com isso, concluímos alguns pontos:

1. **O investimento na segurança de um país é intrínseco para que os cidadãos possam, de fato, exercer da liberdade.**

O score de segurança de um país se baseia nos índices de acesso da população às armas, criminalidade, número de homicídios e o total de pessoas ativas nas Forças Armadas. Tendo em vista os dados mostrados acima dos scores brasileiros, percebemos um aumento no número do

Global Score (ou seja, um indicativo negativo) e, segundo o Anuário de Segurança Pública, divulgado em 28 de junho de 2022, o número de registros de armas de fogo, no Brasil, subiu de 117.467, em 2018, para 673.818, em 2022, um aumento de aproximadamente 473%. Portanto, entende-se claramente que, quanto mais amplo o acesso às armas, piores são os índices de segurança e liberdade em um país.

1. A segurança de um país é diretamente ligada ao contexto econômico, porém, pouco ligada ao contexto sociocultural.

O tamanho da economia de uma nação é normalmente expresso como seu produto interno bruto, ou PIB, que mede o valor de produção de todos os bens de consumo e serviços produzidos dentro de um país em um ano. O Contexto Sócio-Cultural designa um grupo de variáveis contextuais com influência no desempenho e na atividade da organização, e reflete os valores, costumes e tradições da sociedade e influencia as trocas e os sistemas de trabalho.

Os gráficos "Scatterplot" que configuramos, interligando os números de Security Score com os de Economic Context, mostraram uma relação muito forte entre esses dados, uma vez que são diretamente proporcionais: ou seja, quanto melhor o contexto econômico de um país, mais seguro ele é. Todavia, não acontece o mesmo com os dados de Security Score e Sociocultural Context, uma vez que os mesmos mal apresentam uma ligação entre seus números.

1. O contexto econômico está fortemente ligado ao contexto político.

O contexto político pode impactar em fatores como inflação e câmbio, dois elementos que afetam aplicações e empresas. Se o país estiver em dificuldade econômica, também poderá atrasar pagamentos. Ele envolve a situação de governo pela qual o país passa, em relação a como suas forças político-administrativas atuam e são percebidas pela sociedade. No campo financeiro, o contexto político gera maior segurança ou insegurança para investimentos. Em um contexto positivo, há harmonia e cooperação entre: agentes que compõem a força política, como partidos de diferentes ideologias; instituições distintas e complementares dos três poderes (executivo, legislativo, judiciário); mercado financeiro; sociedade civil e forças armadas.

Em nossas análises, podemos observar claramente essa interligação entre esses contextos. Um dos pontos que chamam mais atenção são os poucos "outliers" (valores discrepantes) existentes, o que indica que quase nenhum país possui uma fraca relação entre esses dados.