

Dados: happiness

- Contém características que podem representar o nível de felicidade da população de um país
- Além do nome do país, há atributos de índices que representam o produto interno bruto, suporte social, expectativa de vida, liberdade, generosidade e corrupção

Agrupe os países de acordo com o nível de felicidade da população

- Utilize o método K-means para criar 10 clusters com os dados de treinamento
- Avalie em qual grupo pertence cada país

Avalie o resultado dos clusters criados com dados de teste

In [68]:

```
from sklearn.cluster import KMeans
from sklearn.datasets import make_blobs
import matplotlib.pyplot as plt
import pandas as pd

df = pd.read_csv('../datas/happiness_treino.csv')

df.head(10)
```

Out[68]:

	Country or region	GDP per capita	Social support	Healthy life expectancy	Freedom to make life choices	Generosity	Perceptions of corruption
0	Afghanistan	0.350	0.517	0.361	0.000	0.158	0.025
1	Afghanistan	0.332	0.537	0.255	0.085	0.191	0.036
2	Albania	0.947	0.848	0.874	0.383	0.178	0.027
3	Albania	0.916	0.817	0.790	0.419	0.149	0.032
4	Angola	0.730	1.125	0.269	0.000	0.079	0.061
5	Argentina	1.092	1.432	0.881	0.471	0.066	0.050
6	Argentina	1.073	1.468	0.744	0.570	0.062	0.054
7	Armenia	0.850	1.055	0.815	0.283	0.095	0.064
8	Armenia	0.816	0.990	0.666	0.260	0.077	0.028
9	Australia	1.372	1.548	1.036	0.557	0.332	0.290

In [69]:

```
df = df.drop_duplicates(subset=['Country or region'])
df.head(10)
```

Out[69]:

	Country or region	GDP per capita	Social support	Healthy life expectancy	Freedom to make life choices	Generosity	Perceptions of corruption
0	Afghanistan	0.350	0.517	0.361	0.000	0.158	0.025
2	Albania	0.947	0.848	0.874	0.383	0.178	0.027
4	Angola	0.730	1.125	0.269	0.000	0.079	0.061
5	Argentina	1.092	1.432	0.881	0.471	0.066	0.050
7	Armenia	0.850	1.055	0.815	0.283	0.095	0.064
9	Australia	1.372	1.548	1.036	0.557	0.332	0.290
11	Austria	1.376	1.475	1.016	0.532	0.244	0.226

	Country or region	GDP per capita	Social support	Healthy life expectancy	Freedom to make life choices	Generosity	Perceptions of corruption
13	Azerbaijan	1.043	1.147	0.769	0.351	0.073	0.182
15	Bahrain	1.362	1.368	0.871	0.536	0.255	0.110
17	Bangladesh	0.562	0.928	0.723	0.527	0.166	0.143

In [70]:

```
seed = 18
kmeans = KMeans(n_clusters=10, random_state=seed)

kmeans.fit(df[["GDP per capita", "Social support", "Healthy life expectancy", "Freedom to make life choices", "Generosity", "Perceptions of corruption"]])

df.loc[:, 'labels'] = kmeans.labels_
```

In [71]:

```
predicao = kmeans.predict(df[["GDP per capita", "Social support", "Healthy life expectancy", "Freedom to make life choices", "Generosity", "Perceptions of corruption"]])
```

In [72]:

```
dfsort = df.sort_values(by=['labels'])
```

In [73]:

```
# print(dfsort.to_string())
```

In [74]:

```
dft = pd.read_csv('../datas/happiness_teste.csv')

dft.head(10)
```

Out[74]:

	Country or region	GDP per capita	Social support	Healthy life expectancy	Freedom to make life choices	Generosity	Perceptions of corruption
0	Algeria	1.002	1.160	0.785	0.086	0.073	0.114
1	Bolivia	0.776	1.209	0.706	0.511	0.137	0.064
2	Brazil	1.004	1.439	0.802	0.390	0.099	0.086
3	Cambodia	0.574	1.122	0.637	0.609	0.232	0.062
4	Germany	1.373	1.454	0.987	0.495	0.261	0.265
5	Italy	1.294	1.488	1.039	0.231	0.158	0.030

In [75]:

```
predicao = kmeans.predict(dft[["GDP per capita", "Social support", "Healthy life expectancy", "Freedom to make life choices", "Generosity", "Perceptions of corruption"]])
predicao
```

Out[75]:

```
array([8, 0, 0, 6, 7, 4])
```

Análise

- Algeria - grupo 8 se destaca por ter uma baixa livre escolha dos cidadãos e uma baixa generosidade, países que também fazem parte do grupo: Iran, Morocco, Iraq ... Tunisia.
- Bolivia, Brazil - grupo 0 se destaca por ser um dos grupos com mais membros (tendo características mais na média), países que também fazem parte do grupo: South Africa, Colombia ... Ecuador.
- Cambodia - grupo 6 se destaca pela baixa renda percapita e baixa generosidade, países que também fazem parte do grupo: Ghana, Nepal ... India.

- **Germany** - grupo 7 se destaca pela alta renda percapita, alta expectativa de vida (países desenvolvidos), países que também fazem parte do grupo: **Canada, Malta, ... Denmark.**
- **Italy** - grupo 4 se destaca pela alta renda percapita, expectativa de vida mediana, países que também fazem parte do grupo: **Mexico, Turkey, ... Portugal.**