In [ ]:
```
# Name : Shubham Sapkal
# Roll No. : 2012118
# subject: ML DL
# practical no. : 7
```

Clustering

You are given much more country data. Using the same methodology as the one in the lecture, group all the countries in 2 clusters. Try with other numbers of clusters and see if they match your expectations. Maybe 7 is going to be a cool one! Plot the data using the c parameter to separate the data by the clusters we defined. Note: c stands for color

Import Libraries

In [ ]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
sns.set()
from sklearn.cluster import KMeans
```

Load The Data

In [ ]:
```
df = pd.read_csv('Countries_exercise.csv')
```

Check The Data

In [ ]:
```
df.head()
```

Out[ ]:

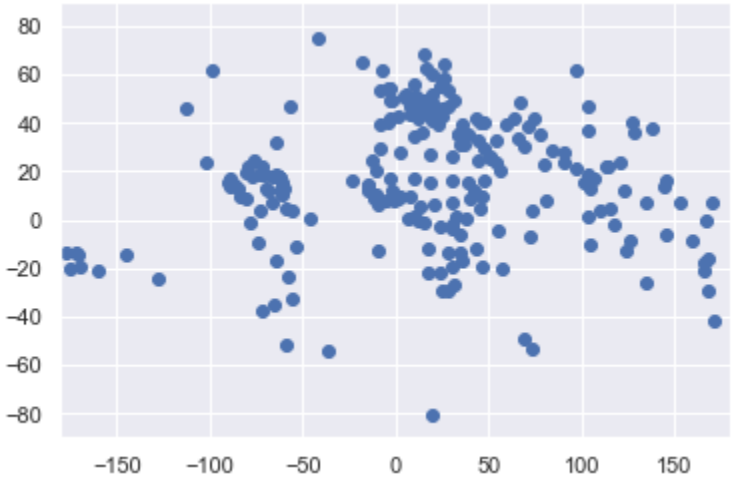
	name	Longitude	Latitude
0	Aruba	-69.982677	12.520880
1	Afghanistan	66.004734	33.835231
2	Angola	17.537368	-12.293361
3	Anguilla	-63.064989	18.223959
4	Albania	20.049834	41.142450

Remove the duplicate index column from the dataset.

In [ ]:
```
data = df.copy()
```

Plot The Data

In [ ]:
```
plt.scatter(data['Longitude'], data['Latitude'])
plt.xlim(-180,180)
plt.ylim(-90, 90)
plt.show()
```



Select The Features

Create a copy of that data and remove all parameters apart from Longitude and Latitude.

In [ ]:
```
x = data.iloc[:,1:3]
```

In [ ]:
```
x
```

Out[ ]:

	Longitude	Latitude
0	-69.982677	12.520880
1	66.004734	33.835231
2	17.537368	-12.293361
3	-63.064989	18.223959
4	20.049834	41.142450
...	...	...
236	-172.164851	-13.753243
237	47.586762	15.909280
238	25.083901	-29.000341
239	27.774759	-13.458242
240	29.851441	-19.004204

241 rows × 2 columns

Clustering

Here's the actual solution: Simply change kmeans = KMeans(2) to kmeans = KMeans(3)

In [ ]:
```
kmeans = KMeans(3)
```

In [ ]:
```
kmeans.fit(x)
```

Out[ ]:
```
KMeans(n_clusters=3)
```

Clustering Results

In [ ]:
```
identified_clusters = kmeans.fit_predict(x)
identified_clusters
data_with_clusters = data.copy()
data_with_clusters['Cluster'] = identified_clusters
data_with_clusters
```

Out[ ]:

	name	Longitude	Latitude	Cluster
0	Aruba	-69.982677	12.520880	0
1	Afghanistan	66.004734	33.835231	1
2	Angola	17.537368	-12.293361	1
3	Anguilla	-63.064989	18.223959	0
4	Albania	20.049834	41.142450	1
...	...	...	...	...
236	Samoa	-172.164851	-13.753243	0
237	Yemen	47.586762	15.909280	1
238	South Africa	25.083901	-29.000341	1
239	Zambia	27.774759	-13.458242	1
240	Zimbabwe	29.851441	-19.004204	1

241 rows × 4 columns

In [ ]:
```
plt.scatter(data['Longitude'], data['Latitude'],c=data_with_clusters['Cluster'], cmap =
plt.xlim(-180,180)
plt.ylim(-90, 90)
plt.show()
```