21BDS0340 - Abhinav Dinesh Srivatsa

Using dataset from https://www.kaggle.com/datasets/warcoder/earthquake-dataset

In [ ]:
```
import pandas as pd
from sklearn.preprocessing import StandardScaler
from sklearn.cluster import KMeans
import matplotlib.pyplot as plt
import seaborn as sns
from mpl_toolkits.basemap import Basemap
```

In [ ]:
```
earthquakes = pd.read_csv("./earthquake_data.csv")
earthquakes
```

Out[ ]:

	title	magnitude	date_time	cdi	mmi	alert	tsunami	sig	net	nst
0	M 6.5 - 42 km W of Sola, Vanuatu	6.5	16-08-2023 12:47	7	4	green	0	657	us	114
1	M 6.5 - 43 km S of Intipucá, El Salvador	6.5	19-07-2023 00:22	8	6	yellow	0	775	us	92
2	M 6.6 - 25 km ESE of Loncopué, Argentina	6.6	17-07-2023 03:05	7	5	green	0	899	us	70
3	M 7.2 - 98 km S of Sand Point, Alaska	7.2	16-07-2023 06:48	6	6	green	1	860	us	173
4	M 7.3 - Alaska Peninsula	7.3	16-07-2023 06:48	0	5	NaN	1	820	at	79
...	...	...	...	...	...	...	...	...	...	...
995	M 7.1 - 85 km S of Tungor, Russia	7.1	27-05-1995 13:03	0	9	NaN	0	776	us	0
996	M 7.7 - 249 km E of Vao, New Caledonia	7.7	16-05-1995 20:12	0	4	NaN	0	912	us	0
997	M 6.9 - 27 km NNW of Maubara, Timor Leste	6.9	14-05-1995 11:33	0	6	NaN	0	732	us	0
998	M 6.6 - 10 km W of Aianí, Greece	6.6	13-05-1995 08:47	0	9	NaN	0	670	us	0
999	M 7.1 - 14 km NE of Cabatuan, Philippines	7.1	05-05-1995 03:53	0	7	NaN	0	776	us	0

1000 rows × 19 columns

In [ ]:
```
feature_names = ['magnitude', 'depth', 'gap']
features = earthquakes[feature_names]
features
```

Out[ ]:

	magnitude	depth	gap
0	6.5	192.955	25.0
1	6.5	69.727	40.0
2	6.6	171.371	28.0
3	7.2	32.571	36.0
4	7.3	21.000	172.8
...	...	...	...
995	7.1	11.000	0.0
996	7.7	20.200	0.0
997	6.9	11.200	0.0
998	6.6	14.000	0.0
999	7.1	16.000	0.0

1000 rows × 3 columns

In [ ]:
```
scaler = StandardScaler()
scaled_features = scaler.fit_transform(features)
scaled_features = pd.DataFrame(scaled_features, columns=feature_names)
scaled_features
```

Out[ ]:

	magnitude	depth	gap
0	-1.005072	0.905124	0.166930
1	-1.005072	-0.037366	0.781591
2	-0.776724	0.740042	0.289862
3	0.593361	-0.321548	0.617682
4	0.821709	-0.410047	6.223393
...	...	...	...
995	0.365014	-0.486531	-0.857505
996	1.735099	-0.416166	-0.857505
997	-0.091682	-0.485001	-0.857505
998	-0.776724	-0.463586	-0.857505
999	0.365014	-0.448289	-0.857505

1000 rows × 3 columns

In [ ]:
```
kmeans = KMeans(n_clusters=4, random_state=42)
clusters = kmeans.fit_predict(scaled_features)

earthquakes['cluster'] = clusters
earthquakes
```

```
/Users/abhi/College Work/Year 4 Semester 1 (Sem 7)/env/lib/python3.8/site-packages/sklearn/cluster/_kmeans.py:1416: FutureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
  super()._check_params_vs_input(X, default_n_init=10)
```

Out[ ]:

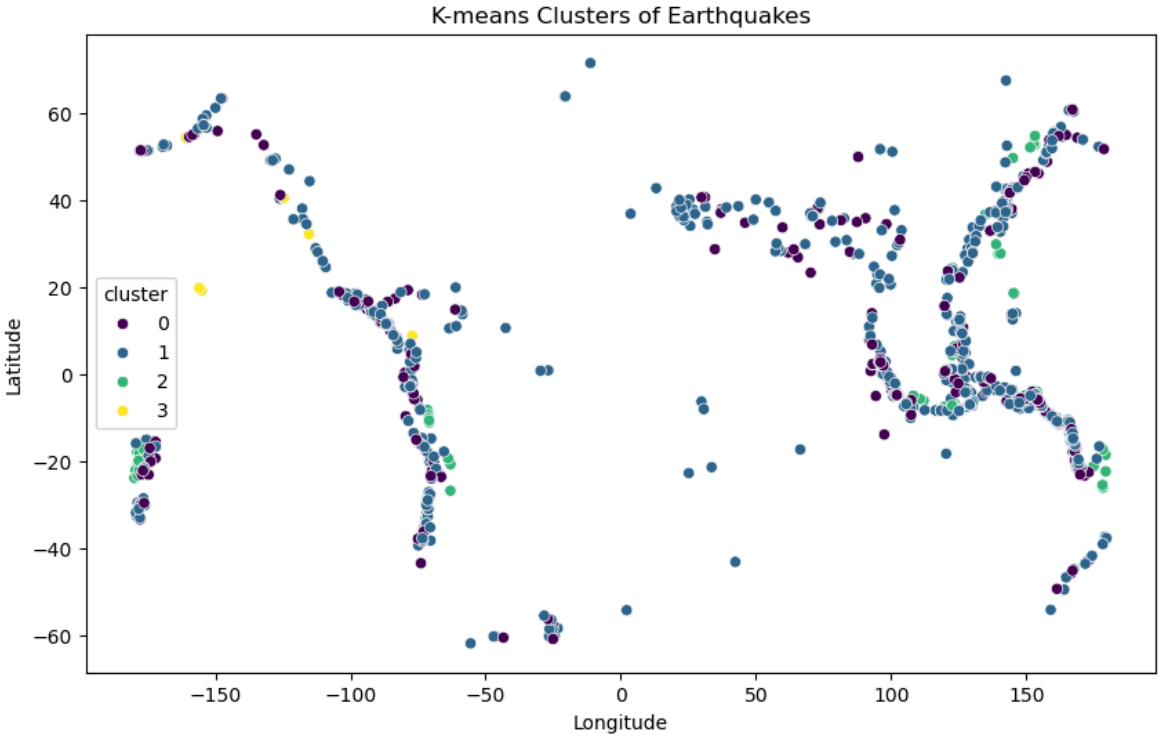
	title	magnitude	date_time	cdi	mmi	alert	tsunami	sig	net	nst
0	M 6.5 - 42 km W of Sola, Vanuatu	6.5	16-08-2023 12:47	7	4	green	0	657	us	114
1	M 6.5 - 43 km S of Intipucá, El Salvador	6.5	19-07-2023 00:22	8	6	yellow	0	775	us	92
2	M 6.6 - 25 km ESE of Loncopué, Argentina	6.6	17-07-2023 03:05	7	5	green	0	899	us	70
3	M 7.2 - 98 km S of Sand Point, Alaska	7.2	16-07-2023 06:48	6	6	green	1	860	us	173
4	M 7.3 - Alaska Peninsula	7.3	16-07-2023 06:48	0	5	NaN	1	820	at	79
...	...	...	...	...	...	...	...	...	...	...
995	M 7.1 - 85 km S of Tungor, Russia	7.1	27-05-1995 13:03	0	9	NaN	0	776	us	0
996	M 7.7 - 249 km E of Vao, New Caledonia	7.7	16-05-1995 20:12	0	4	NaN	0	912	us	0
997	M 6.9 - 27 km NNW of Maubara, Timor Leste	6.9	14-05-1995 11:33	0	6	NaN	0	732	us	0
998	M 6.6 - 10 km W of Aianí, Greece	6.6	13-05-1995 08:47	0	9	NaN	0	670	us	0
999	M 7.1 - 14 km NE of Cabatuan, Philippines	7.1	05-05-1995 03:53	0	7	NaN	0	776	us	0

1000 rows × 20 columns

In [ ]:
```
plt.figure(figsize=(10, 6))
sns.scatterplot(x=earthquakes['longitude'], y=earthquakes['latitude'], hu
plt.title('K-means Clusters of Earthquakes')
```

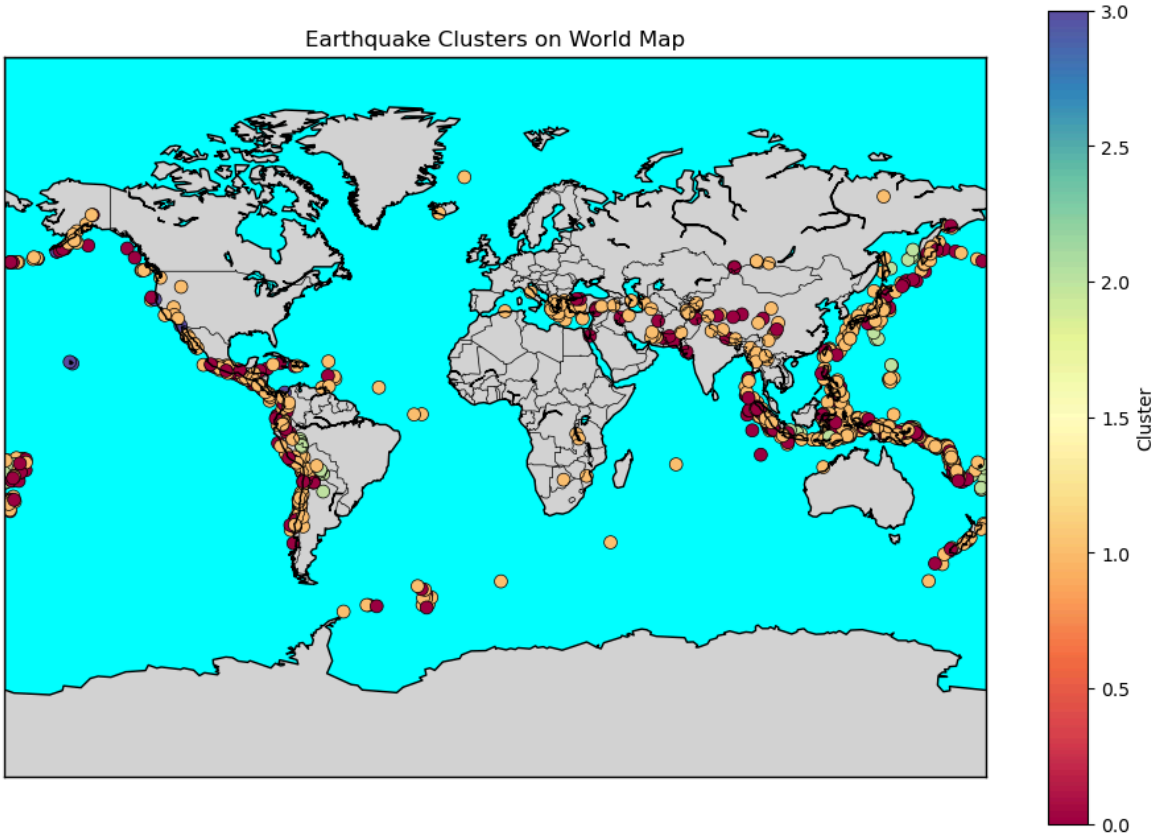
```
plt.xlabel('Longitude')
plt.ylabel('Latitude')
```

Out[ ]: Text(0, 0.5, 'Latitude')



In [ ]:
```
plt.figure(figsize=(12, 8))
m = Basemap(projection='mill',
            llcrnrlat=-90, urcrnrlat=90,
            llcrnrlon=-180, urcrnrlon=180,
            resolution='c')
m.drawcoastlines()
m.drawcountries()
m.drawmapboundary(fill_color='aqua')
m.fillcontinents(color='lightgray', lake_color='aqua')
x, y = m(earthquakes['longitude'].values, earthquakes['latitude'].values)
scatter = m.scatter(x, y, c=earthquakes['cluster'], cmap='Spectral', s=50
plt.colorbar(scatter, label='Cluster')
plt.title('Earthquake Clusters on World Map')
```

Out[ ]: Text(0.5, 1.0, 'Earthquake Clusters on World Map')


Earthquake Clusters on World Map
Cluster
3.0
2.5
2.0
1.5
1.0
0.5
0.0