

```
In [1]: import matplotlib.pyplot as plt
from sklearn.datasets import make_blobs
import pandas as pd
import numpy as np
%matplotlib inline
```

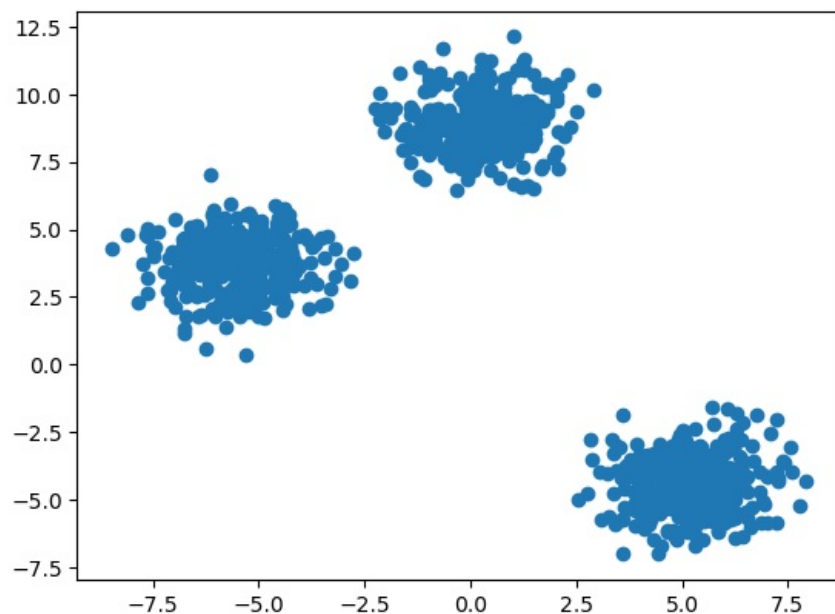
```
In [2]: X,y = make_blobs(n_samples=1000, centers=3, n_features=2, random_state=23)
```

```
In [3]: X.shape
```

```
Out[3]: (1000, 2)
```

```
In [4]: plt.scatter(X[:,0],X[:,1])
```

```
Out[4]: <matplotlib.collections.PathCollection at 0x1eeb253dd50>
```



```
In [5]: from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X,y,test_size=0.33, random_state=42)
```

```
In [6]: from sklearn.cluster import KMeans
```

```
In [7]: # Manual Process
```

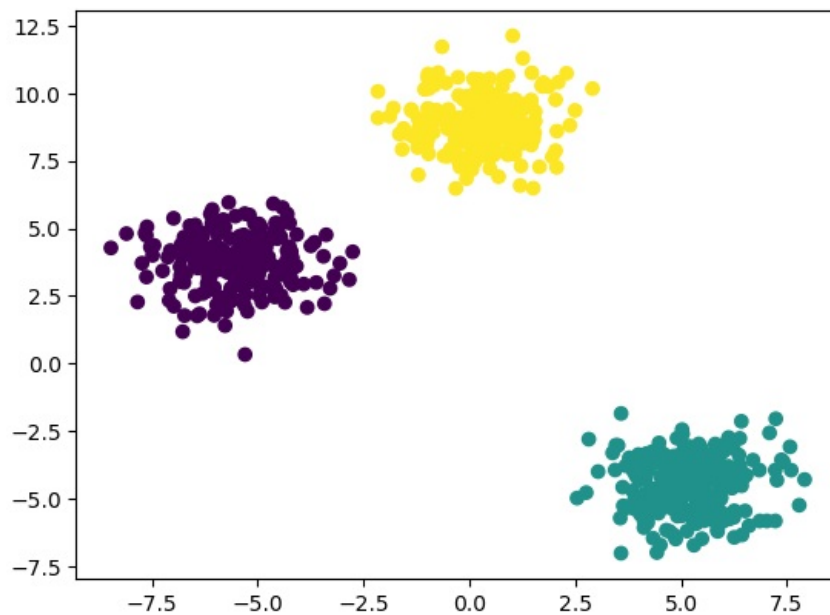
```
In [8]: #1. Elbow Method - to select the k value
```

```
wcss = []
for k in range(1,11):
    kmeans = KMeans(n_clusters=k, init='k-means++')
    kmeans.fit(X_train)
    wcss.append(kmeans.inertia_)
```

```
C:\Users\Garima\anaconda3\lib\site-packages\sklearn\cluster\_kmeans.py:1412: 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)
```

```
In [14]: plt.scatter(X_train[:,0],X_train[:,1],c=y_labels)
```

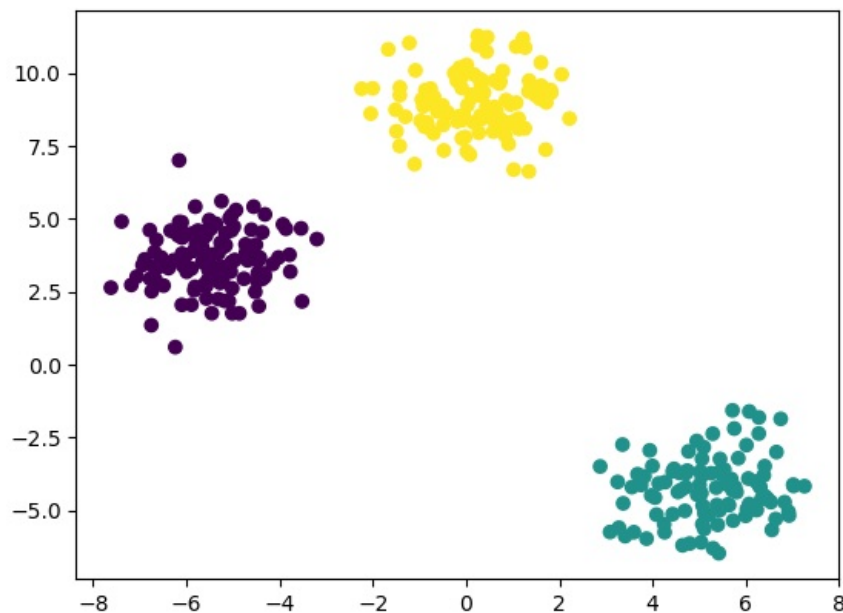
```
Out[14]: <matplotlib.collections.PathCollection at 0x1eeae0dfdf0>
```



```
In [15]: y_test_labels = kmeans.predict(X_test)
```

```
In [16]: plt.scatter(X_test[:,0],X_test[:,1],c=y_test_labels)
```

```
Out[16]: <matplotlib.collections.PathCollection at 0x1eeb3cb7850>
```



```
In [18]: ## 2. Knee locator  
!pip install kneed
```

```
Collecting kneed  
  Downloading kneed-0.8.5-py3-none-any.whl (10 kB)  
Requirement already satisfied: scipy>=1.0.0 in c:\users\garima\anaconda3\lib\site-packages (from kneed) (1.10.0)  
)  
Requirement already satisfied: numpy>=1.14.2 in c:\users\garima\anaconda3\lib\site-packages (from kneed) (1.23.5)  
Installing collected packages: kneed  
Successfully installed kneed-0.8.5
```

```
In [19]: from kneed import KneeLocator
```

```
In [21]: kl = KneeLocator(range(1,11),wcss,curve='convex', direction='decreasing')  
kl.elbow
```

```
Out[21]: 3
```

```
In [23]: ## Performance metrics  
## 3. Silhouette score  
from sklearn.metrics import silhouette_score  
silhouette_coefficients=[]
```

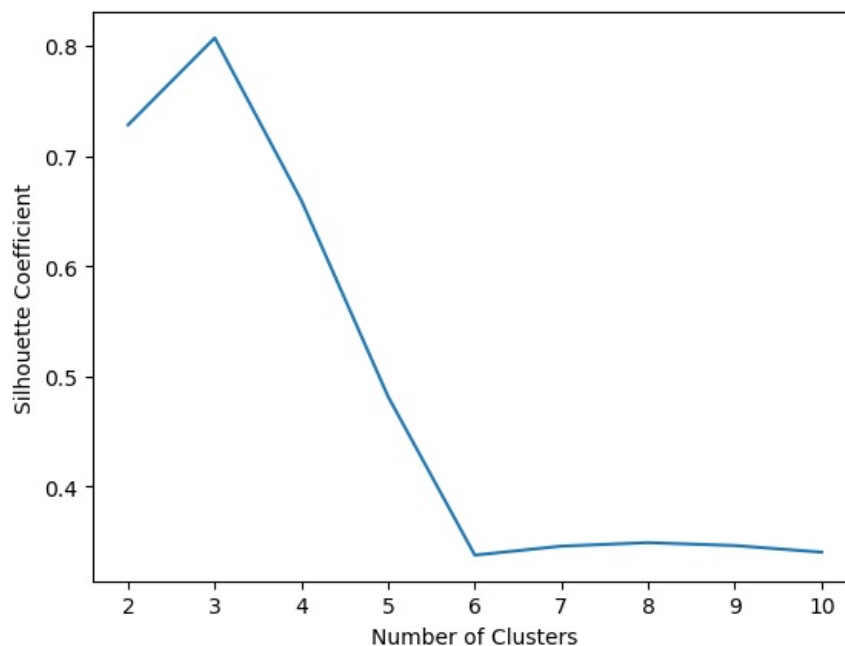
```
for k in range(2,11):
    kmeans = KMeans(n_clusters=k, init='k-means++')
    kmeans.fit(X_train)
    score=silhouette_score(X_train,kmeans.labels_)
    silhouette_coefficients.append(score)
```

C:\Users\Garima\anaconda3\lib\site-packages\sklearn\cluster_kmeans.py:1412: 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)
 C:\Users\Garima\anaconda3\lib\site-packages\sklearn\cluster_kmeans.py:1412: 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)
 C:\Users\Garima\anaconda3\lib\site-packages\sklearn\cluster_kmeans.py:1412: 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)
 C:\Users\Garima\anaconda3\lib\site-packages\sklearn\cluster_kmeans.py:1412: 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)
 C:\Users\Garima\anaconda3\lib\site-packages\sklearn\cluster_kmeans.py:1412: 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)
 C:\Users\Garima\anaconda3\lib\site-packages\sklearn\cluster_kmeans.py:1412: 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)
 C:\Users\Garima\anaconda3\lib\site-packages\sklearn\cluster_kmeans.py:1412: 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)
 C:\Users\Garima\anaconda3\lib\site-packages\sklearn\cluster_kmeans.py:1412: 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)

In [24]: silhouette_coefficients

Out[24]: [0.7281443868598331,
 0.8071181203797672,
 0.6593671074925262,
 0.4814001336222496,
 0.3375447395994776,
 0.3456934811504727,
 0.3488826796246165,
 0.34621003599297945,
 0.34020693560708065]

In [25]: #plotting silhouette_score
 plt.plot(range(2,11),silhouette_coefficients)
 plt.xticks(range(2,11))
 plt.xlabel("Number of Clusters")
 plt.ylabel("Silhouette Coefficient")
 plt.show()



In [26]: #Highest in k=3, so 3 cluster

In []: