```
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
        (1000, 2)
Out[3]:
In [4]: plt.scatter(X[:,0],X[:,1])
        <matplotlib.collections.PathCollection at 0x1eeb253dd50>
Out[4]:
         12.5
         10.0
           7.5
           5.0
           2.5
           0.0
         -2.5
         -5.0
         -7.5
                    -7.5
                                      -2.5
                                                                            7.5
                             -5.0
                                                0.0
                                                         2.5
                                                                   5.0
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_)
```

```
f `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 o
           `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 o
         f `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 o
           `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 o
         f `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 o
         f `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 o
         f `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 o
           `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 o
         f `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 o
           `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 [9]: WCSS
 Out[9]: [34827.57682552021,
          7935.437286145418,
          1319.2730531585612,
          1140.4677884655125,
          992.0624178531957,
          853.7610520652662,
          755.285459229354.
          670.8573334280152.
          580.4637798569495,
          534.27295128754061
         #Plotting elbow curve
In [10]:
         plt.plot(range(1,11),wcss)
         plt.xticks(range(1,11))
         plt.xlabel("Number of Clusters")
         plt.ylabel("WCSS")
         plt.show()
            35000
            30000
            25000
            20000
            15000
            10000
             5000
                 0
                            2
                                               5
                                                      6
                                  3
                                                                               10
                      1
                                           Number of Clusters
In [11]: # we got k=3 from elbow curve
In [12]: kmeans = KMeans(n clusters=3,init="k-means++")
```

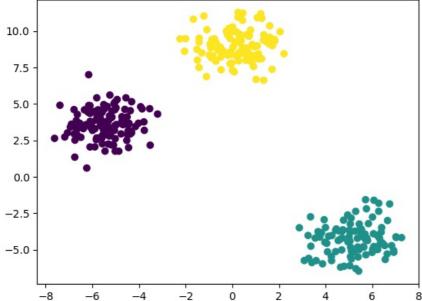
C:\Users\Garima\anaconda3\lib\site-packages\sklearn\cluster_kmeans.py:1412: FutureWarning: The default value of `n init` will change from 10 to 'auto' in 1.4. Set the value of `n init` explicitly to suppress the warning

In [13]: y labels = kmeans.fit predict(X train)

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 o

```
In [14]: plt.scatter(X train[:,0],X train[:,1],c=y labels)
          <matplotlib.collections.PathCollection at 0x1eeae0dfdf0>
Out[14]:
           12.5
           10.0
            7.5
            5.0
            2.5
            0.0
           -2.5
           -5.0
           -7.5
                    -7.5
                              -5.0
                                       -2.5
                                                 0.0
                                                          2.5
                                                                   5.0
                                                                            7.5
In [15]: y_test_labels = kmeans.predict(X_test)
In [16]: plt.scatter(X_test[:,0],X_test[:,1],c=y_test_labels)
          <matplotlib.collections.PathCollection at 0x1eeb3cb7850>
Out[16]:
           10.0
            7.5
            5.0
            2.5
```



```
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:\ \ satisfied: \ scipy>=1.0.0 \ in \ c:\ \ satisfied: \ scipy>=1.0.0 \ in \ c:\ \ satisfied: \ satisfied: \ scipy>=1.0.0 \ in \ c:\ \ satisfied: \ satisfied: \ scipy>=1.0.0 \ in \ c:\ \ satisfied: \ satisfi
                                                                                         Requirement already satisfied: numpy >= 1.14.2 in c: \users \garima\anaconda \lib \site-packages (from kneed) (1.23.) in c: \users \garima\anaconda \lib \site-packages (from kneed) (1.23.) in c: \users \garima\anaconda \site-packages (from kneed) (1.23.) in c: \
                                                                                         5)
                                                                                         Installing collected packages: kneed
                                                                                         Successfully installed kneed-0.8.5
In [19]: from kneed import KneeLocator
```

```
kl = KneeLocator(range(1,11),wcss,curve='convex', direction='decreasing')
In [21]:
Out[21]: 3
In [23]:
         ## Performance metrics
         ## 3. Silhoutte score
         from sklearn.metrics import silhouette_score
         silhouette_coefficients=[]
```

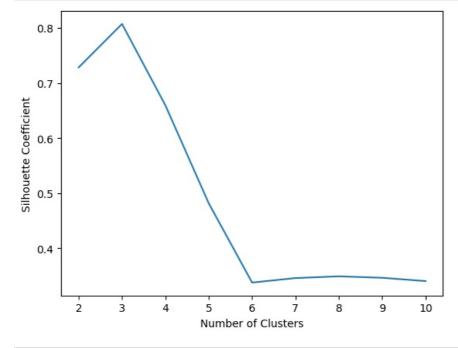
```
kmeans = KMeans(n_clusters=k, init='k-means++')
    kmeans.fit(X train)
    score=silhouette score(X train,kmeans.labels )
    silhouette coefficients.append(score)
C:\Users\Garima\anaconda3\lib\site-packages\sklearn\cluster\ kmeans.py:1412: FutureWarning: The default value o
f `n_init` will change 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 o
  '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 o
 `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 o
  `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 o
  `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 o
  `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 o
f `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 o
  `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
```

for k in range(2,11):

```
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 []: