Clustering Model

Clustering Model

```python
import numpy as np

import matplotlib.pyplot as plt

import pandas as pd


# Importing the dataset
dataset = pd.read_csv(r'D:\Samson - All Data\Naresh IT Institute\New folder\Mall_Customers.csv')

x = dataset.iloc[:, [3, 4]].values


# Using the elbow method to find the optional number of clusters
from sklearn.cluster import KMeans


# we are going to findout the optimal number of cluster & we have to use the elbow
wcss=[]


for i in range(1, 11):
    kmeans = KMeans(n_clusters = i,init="k-means++",random_state=0)
    kmeans.fit(x)
    wcss.append(kmeans.inertia_)
plt.plot(range(1, 11), wcss)
plt.title('The Elbow Method')
plt.xlabel('Number of clusters')
plt.ylabel('WCSS')
plt.show()
```

```python
kmeans = KMeans(n_clusters = 5, init = 'k-means++', random_state = 0)

y_kmeans = kmeans.fit_predict(x)


# Visualising the clusters

plt.scatter(x[y_kmeans == 0, 0], x[y_kmeans == 0, 1], s = 100, c = 'red', label = 'Cluster 1')

plt.scatter(x[y_kmeans == 1, 0], x[y_kmeans == 1, 1], s = 100, c = 'blue', label = 'Cluster 2')

plt.scatter(x[y_kmeans == 2, 0], x[y_kmeans == 2, 1], s = 100, c = 'green', label = 'Cluster 3')

plt.scatter(x[y_kmeans == 3, 0], x[y_kmeans == 3, 1], s = 100, c = 'cyan', label = 'Cluster 4')

plt.scatter(x[y_kmeans == 4, 0], x[y_kmeans == 4, 1], s = 100, c = 'magenta', label = 'Cluster 5')

plt.scatter(kmeans.cluster_centers_[:, 0], kmeans.cluster_centers_[:, 1], s = 300, c = 'yellow', label = 'Centroids')

plt.title('Clusters of customers')

plt.xlabel('Annual Income (k$)')

plt.ylabel('Spending Score (1-100)')

plt.legend()

plt.show()


dataset['cluster'] = y_kmeans
```