```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import random
```

```
dataset=pd.read_csv("/content/E-commerce Customer Behavior - Sheet1.csv")
dataset.head()
```

	Customer ID	Gender	Age	City	Membership Type	Total Spend	Items Purchased	Average Rating	Discount Applied	Days Since Last Purchase	Satisfaction Level
0	101	Female	29	New York	Gold	1120.20	14	4.6	True	25	Satisfied
1	102	Male	34	Los Angeles	Silver	780.50	11	4.1	False	18	Neutral
2	103	Female	43	Chicago	Bronze	510.75	9	3.4	True	42	Unsatisfied
3	104	Male	30	San Francisco	Gold	1480.30	19	4.7	False	12	Satisfied
4	105	Male	27	Miami	Silver	720.40	13	4.0	True	55	Unsatisfied

Next steps: Generate code with dataset | New interactive sheet

```
features = ['Total Spend', 'Items Purchased', 'Days Since Last Purchase', 'Average Rating']
X = np.array(dataset[features])
```

```
def calculate_distance(a, b):
    return np.sqrt(np.sum((a - b) ** 2))
```

```
def assign_clusters(centroids, X):
    assigned_cluster = []
    for i in X:
        distances = [calculate_distance(i, c) for c in centroids]
        assigned_cluster.append(np.argmin(distances))
    return assigned_cluster
```

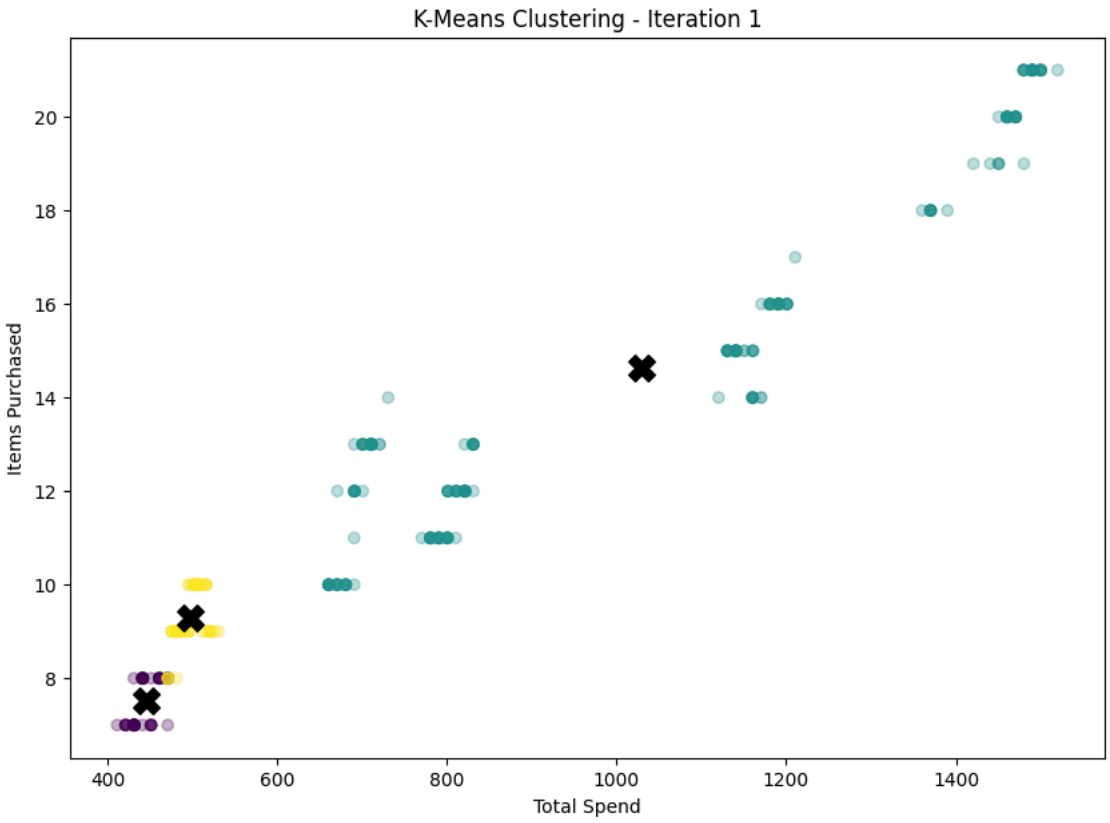
```
def update_centroids(clusters, X):
    new_centroids = []
    df_temp = pd.concat([pd.DataFrame(X), pd.Series(clusters, name='cluster')], axis=1)
    for c in set(df_temp['cluster']):
        current_cluster = df_temp[df_temp['cluster'] == c][df_temp.columns[:-1]]
        cluster_mean = current_cluster.mean(axis=0)
        new_centroids.append(cluster_mean)
    return np.array(new_centroids)
```

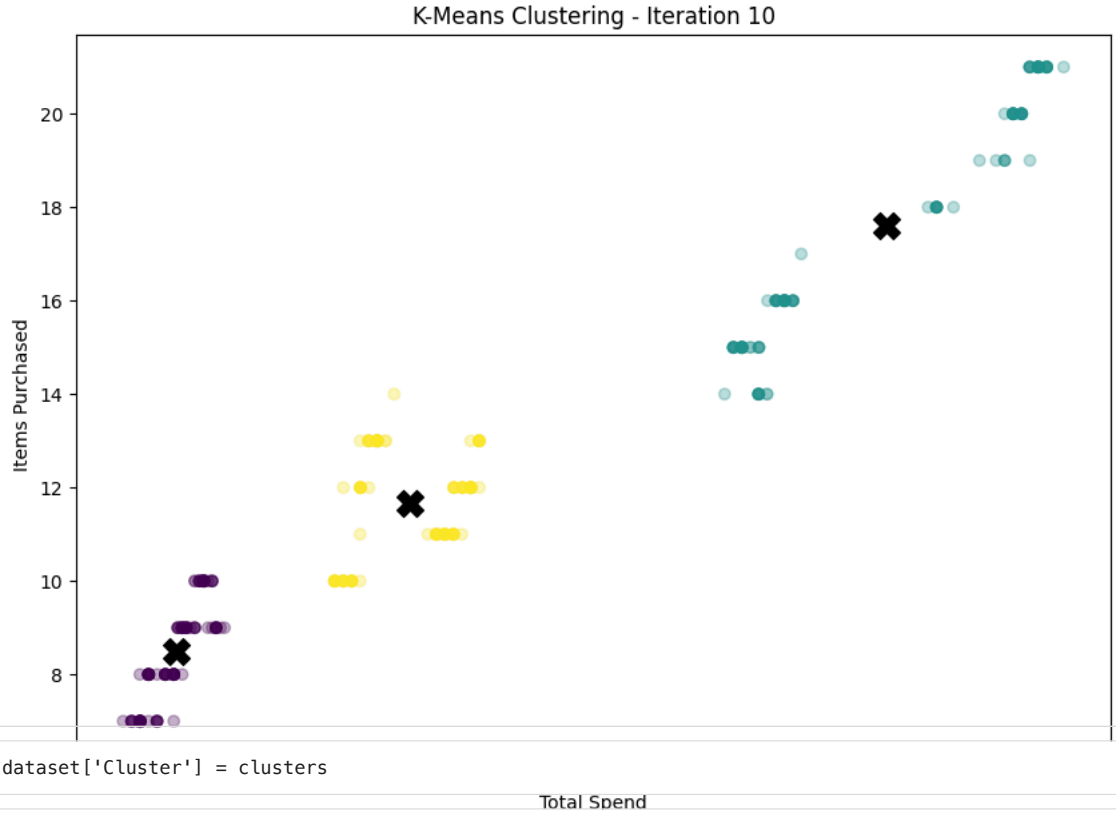
```
k = 3
init_indices = random.sample(range(len(X)), k)
centroids = np.array([X[i] for i in init_indices])
print("Initial Centroids:\n", centroids)
```

```
Initial Centroids:
 [[430.5    8.    21.     3.1 ]
 [800.9   12.    17.     4.1 ]
 [505.75  10.    39.     3.3 ]]
```

```
epochs = 10
for i in range(epochs):
    clusters = assign_clusters(centroids, X)
    centroids = update_centroids(clusters, X)

    # Visualize first and last iteration
    if i == 0 or i == epochs - 1:
        plt.figure(figsize=(10,7))
        plt.scatter(X[:,0], X[:,1], c=clusters, alpha=0.3)
        plt.scatter(centroids[:,0], centroids[:,1], color='black', marker='X', s=200)
        plt.title(f'K-Means Clustering - Iteration {i+1}')
        plt.xlabel(features[0])
        plt.ylabel(features[1])
        plt.show()
```






```
dataset['Cluster'] = clusters
```

```
for i in range(k):
    print(f"\nCluster {i} Summary:")
    print(dataset[dataset['Cluster']==i][features].describe())
```

```
Cluster 0 Summary:
       Total Spend  Items Purchased  Days Since Last Purchase  Average Rating
count   116.000000       116.000000                116.000000      116.000000
mean    473.388793         8.491379                 31.612069        3.325000
std      31.299435         1.050844                  9.504938        0.197759
min     410.800000         7.000000                 18.000000        3.000000
25%     440.900000         8.000000                 23.000000        3.175000
50%     475.250000         8.500000                 33.500000        3.300000
75%     500.750000         9.000000                 39.000000        3.500000
max     530.400000        10.000000                 49.000000        3.600000

Cluster 1 Summary:
```

```
       Total Spend  Items Purchased  Days Since Last Purchase  Average Rating
count   117.000000       117.000000                117.000000      117.000000
mean   1311.144444        17.615385                 17.940171        4.675214
std     151.929971         2.548990                  7.535308        0.167077
min    1120.200000        14.000000                  9.000000        4.300000
25%    1160.600000        15.000000                 11.000000        4.500000
50%    1210.600000        17.000000                 18.000000        4.700000
75%    1470.500000        20.000000                 25.000000        4.800000
max    1520.100000        21.000000                 36.000000        4.900000

Cluster 2 Summary:
       Total Spend  Items Purchased  Days Since Last Purchase  Average Rating
count   117.000000       117.000000                117.000000      117.000000
mean    748.432479        11.658120                 30.256410        4.051282
std      60.594250         1.107672                 16.820633        0.174007
min     660.300000        10.000000                 12.000000        3.700000
25%     690.600000        11.000000                 15.000000        3.900000
50%     770.200000        12.000000                 21.000000        4.100000
75%     800.900000        13.000000                 47.000000        4.200000
max     830.900000        14.000000                 63.000000        4.400000
```