

## Importing the required modules

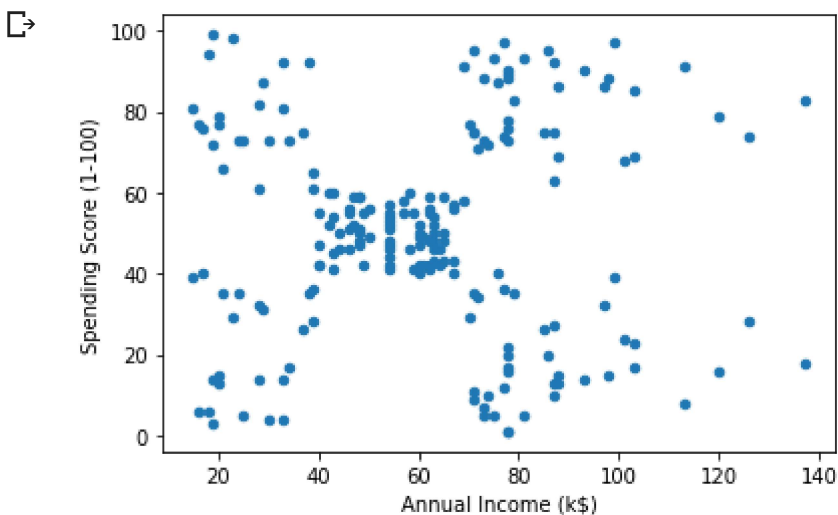
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
```

Initiaising the dataset to variable named "data" and storing the values of last two attributes of data:

```
dataset = pd.read_csv('Mall_Customers.csv')
X = dataset.iloc[:, [3,4]].values
```

Plotting the initial dataset, with no clustering.

```
dataset.plot.scatter(x='Annual Income (k$)', y = 'Spending Score (1-100)')
plt.show()
```



Aliasing the KMeans module as to kmeans. Moreover, we are fixing k value as 5.

```
kmeans = KMeans(n_clusters=5)
```

Predicting to which cluster each datapoint would belong to and storing it into y\_mean.

```
y_mean = kmeans.fit_predict(X)
```

Plotting the clusters.

```
plt.figure(1, figsize = (17, 8))
plt.scatter(X[y_mean == 0, 0], X[y_mean == 0, 1], s = 10, c = 'red', label = 'Cluster 1')
plt.scatter(X[y_mean == 1, 0], X[y_mean == 1, 1], s = 10, c = 'blue', label = 'Cluster 2')
plt.scatter(X[y_mean == 2, 0], X[y_mean == 2, 1], s = 10, c = 'green', label = 'Cluster 3')
plt.scatter(X[y_mean == 3, 0], X[y_mean == 3, 1], s = 10, c = 'orange', label = 'Cluster 4')
plt.scatter(X[y_mean == 4, 0], X[y_mean == 4, 1], s = 10, c = 'purple', label = 'Cluster 5')
```

```

plt.scatter(X[y_mean == 1, 0], X[y_mean == 1, 1], s = 10, c = 'yellow', label = 'Cluster 2')
plt.scatter(X[y_mean == 2, 0], X[y_mean == 2, 1], s = 10, c = 'aqua', label = 'Cluster 3')
plt.scatter(X[y_mean == 3, 0], X[y_mean == 3, 1], s = 10, c = 'violet', label = 'Cluster 4')
plt.scatter(X[y_mean == 4, 0], X[y_mean == 4, 1], s = 10, c = 'lightgreen', label = 'Cluster 1')
plt.title('Clustering based on Annual income v/s Spending Score')
plt.scatter(kmeans.cluster_centers_[0], kmeans.cluster_centers_[1], s = 30, c = 'navy')
plt.legend()
plt.show()

```



Clustering based on Annual income v/s Spending Score

