

## ▼ K-Means Clustering

### ▼ Importing the libraries

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
```

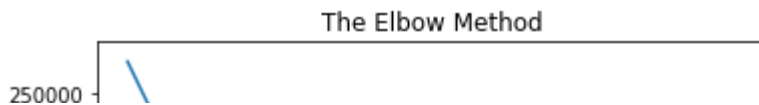
### ▼ Importing the dataset

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

### ▼ Using the elbow method to find the optimal number of clusters

```
from sklearn.cluster import KMeans
wcss=[]
for i in range(1,11,1):
    kmeans=KMeans(n_clusters= i, init='k-means++', random_state=42)
    kmeans.fit(X)
    wcss.append(kmeans.inertia_)
plt.plot(range(1,11,1), wcss)
plt.title("The Elbow Method")
plt.xlabel("No. of Clusters")
plt.ylabel("WCSS")
plt.show()
```





- ▼ Training the K-Means model on the dataset

```
kmeans=KMeans(n_clusters= 5, init='k-means++', random_state=42)
y_kmeans=kmeans.fit_predict(X)
print(y_kmeans)
```

[illegible]

- ▼ 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='black')
plt.title("Cluster of Centroids")
plt.xlabel("Annual Income(k$)")
plt.ylabel("Spending Score(1-100)")
plt.legend()
plt.show()
```

