

```
In [ ]: import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
%matplotlib inline
```

```
In [ ]: dataset = pd.read_csv('../Social_Network_Ads.csv')
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
dataset['Gender']=le.fit_transform(dataset['Gender'])
```

```
In [ ]: print(dataset.head())
```

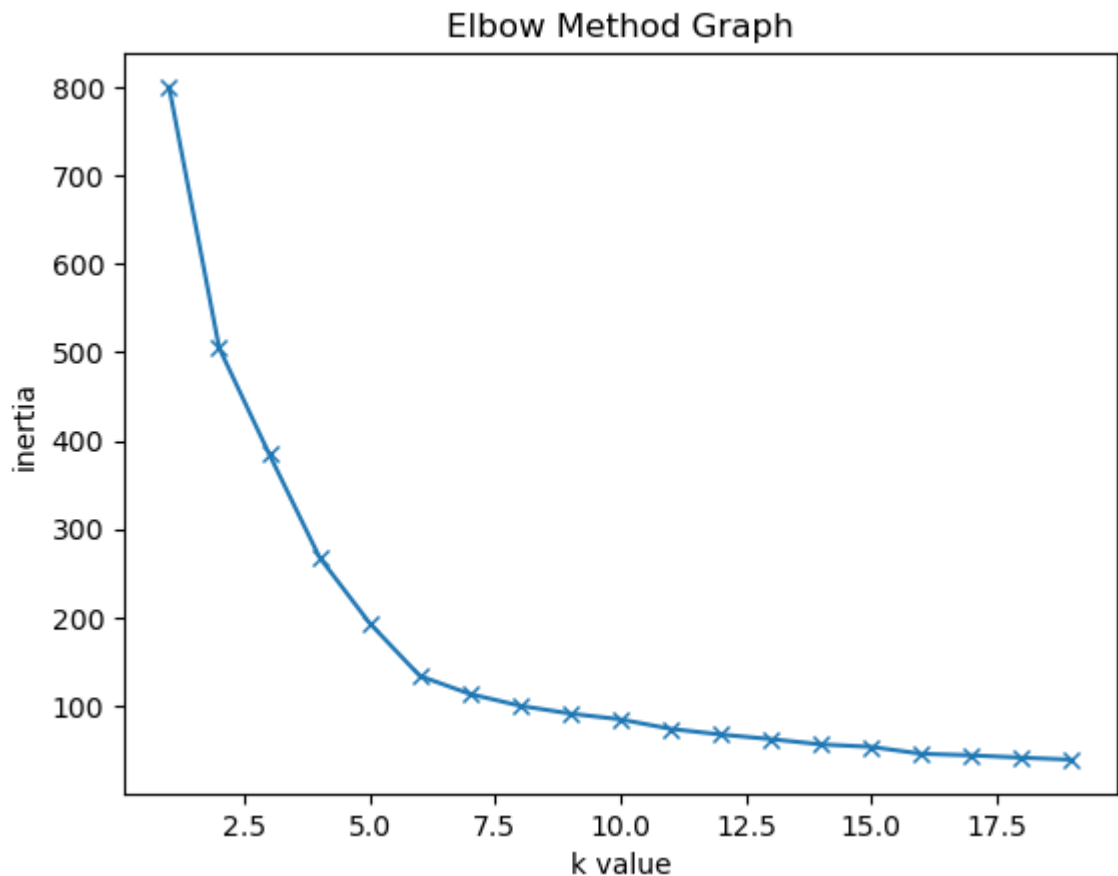
	User ID	Gender	Age	EstimatedSalary	Purchased
0	15624510	1	19	19000	0
1	15810944	1	35	20000	0
2	15668575	0	26	43000	0
3	15603246	0	27	57000	0
4	15804002	1	19	76000	0

```
In [ ]: X = dataset.iloc[:, 2:4].values
y = dataset.iloc[:, -1].values
```

```
In [ ]: from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
X_scaled_data = scaler.fit_transform(X)
```

```
In [ ]: from sklearn.cluster import KMeans
wcss = []
for i in range(1, 20):
    kmeans = KMeans(n_clusters = i, random_state = 0,
                    init='k-means++', n_init='auto')
    kmeans.fit(X_scaled_data)
    wcss.append(kmeans.inertia_)
```

```
In [ ]: plt.plot(range(1,20),wcss,marker = 'x')
plt.title('Elbow Method Graph')
plt.xlabel('k value')
plt.ylabel('inertia')
plt.show()
```



```
In [ ]: kmeans = KMeans(n_clusters=4, random_state=0,  
                        init='k-means++', n_init='auto')  
kmeans.fit(X_scaled_data)
```

```
Out[ ]: ▼ KMeans  
KMeans(n_clusters=4, n_init='auto', random_state=0)
```

```
In [ ]: import seaborn as sns  
labels = kmeans.predict(X_scaled_data)  
centroids = kmeans.cluster_centers_
```

```
In [ ]: import matplotlib.pyplot as plt  
  
plt.figure(figsize=(8, 6))  
scatter = plt.scatter(X_scaled_data[:, 0], X_scaled_data[:, 1], c=labels,  
                      cmap='tab20b', s=50, label='Data Points')  
  
plt.scatter(centroids[:, 0], centroids[:, 1], c='red',  
            marker='X', s=200, label='Centroids')  
  
plt.xlabel('Age')  
plt.ylabel('Estimated Salary')  
  
plt.legend()  
  
plt.title('K-means Clustering Results')  
plt.show()
```

