Mall Customers

```
In [1]: import matplotlib.pyplot as plt
        import pandas as pd
        from sklearn.datasets import make_blobs
        from sklearn.cluster import KMeans
        from sklearn.metrics import silhouette_score
        from sklearn.preprocessing import StandardScaler
        import seaborn as sns
```

```
In [2]: df=pd.read_csv('D:\\Downloads\\Mall_Customers.csv')
        df
```

Out[2]:

	CustomerID	Gender	Age	Annual Income (k$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40
...	...	...	...	...	...
195	196	Female	35	120	79
196	197	Female	45	126	28
197	198	Male	32	126	74
198	199	Male	32	137	18
199	200	Male	30	137	83

200 rows × 5 columns

In [3]: `df.isnull()`

Out[3]:

	CustomerID	Gender	Age	Annual Income (k$)	Spending Score (1-100)
0	False	False	False	False	False
1	False	False	False	False	False
2	False	False	False	False	False
3	False	False	False	False	False
4	False	False	False	False	False
...	...	...	...	...	...
195	False	False	False	False	False
196	False	False	False	False	False
197	False	False	False	False	False
198	False	False	False	False	False
199	False	False	False	False	False

200 rows × 5 columns

In [4]: `df.head()`

Out[4]:

	CustomerID	Gender	Age	Annual Income (k$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40

In [5]:
```python
sns.heatmap(df.isna())
```

Out[5]: `<AxesSubplot:>`
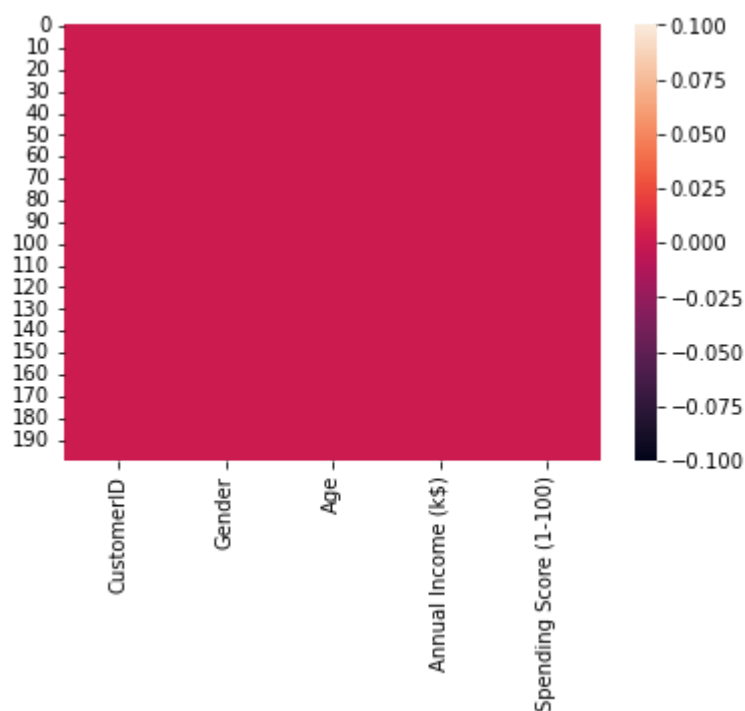
In [6]:
```python
df['Gender'] = df['Gender'].map({'Male':1,'Female':0})
```
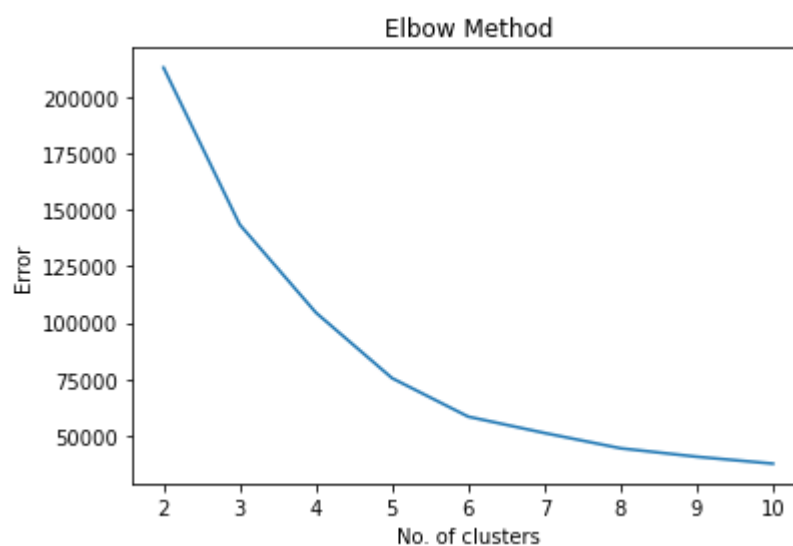
In [7]:
```python
df
```

Out[7]:

	CustomerID	Gender	Age	Annual Income (k$)	Spending Score (1-100)
0	1	1	19	15	39
1	2	1	21	15	81
2	3	0	20	16	6
3	4	0	23	16	77
4	5	0	31	17	40
...	...	...	...	...	...
195	196	0	35	120	79
196	197	0	45	126	28
197	198	1	32	126	74
198	199	1	32	137	18
199	200	1	30	137	83

200 rows × 5 columns

In [10]:
```python
X=df.drop('CustomerID',axis=1)
```

In [11]:
```python
Error=[]                                    #for selecting no. of clusters
for i in range(2,11):
    kmeans=KMeans(n_clusters=i).fit(X)
    Error.append(kmeans.inertia_)
import matplotlib.pyplot as plt
plt.plot(range(2,11),Error)
plt.title('Elbow Method')
plt.xlabel('No. of clusters')
plt.ylabel('Error')
plt.show()
```


In [ ]:

In [ ]:

In [ ]: