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

```
data=pd.read_csv('/content/product_reviews.csv')
```

```
data.head()
```

```
{"summary":"{\n  \"name\": \"data\",\n  \"rows\": 10,\n  \"fields\":
[\n    {\n      \"column\": \"ProductID\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 3,\n        \"min\": 101,\n
\"max\": 110,\n        \"num_unique_values\": 10,\n
\"samples\": [\n          109,\n          102,\n          106\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"ReviewText\",\n
\"properties\": {\n        \"dtype\": \"string\",\n
\"num_unique_values\": 10,\n        \"samples\": [\n          \"Didn't
meet my expectations. Disappointed.\",\n          \"Worst product I
ever bought. Totally waste.\",\n          \"Good value for money.
Satisfied with purchase.\"\n        ],\n        \"semantic_type\":
\"\",\n        \"description\": \"\"\n      }\n    },\n    {\n
\"column\": \"Rating\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 1,\n        \"min\": 1,\n
\"max\": 5,\n        \"num_unique_values\": 5,\n        \"samples\":
[\n          1,\n          2,\n          3\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    }\n  ]\n}","type":"dataframe","variable_name":"data"}
```

```
data.isnull().sum()
```

```
ProductID     0
ReviewText    0
Rating        0
dtype: int64
```

```
data.dropna(inplace=True)
```

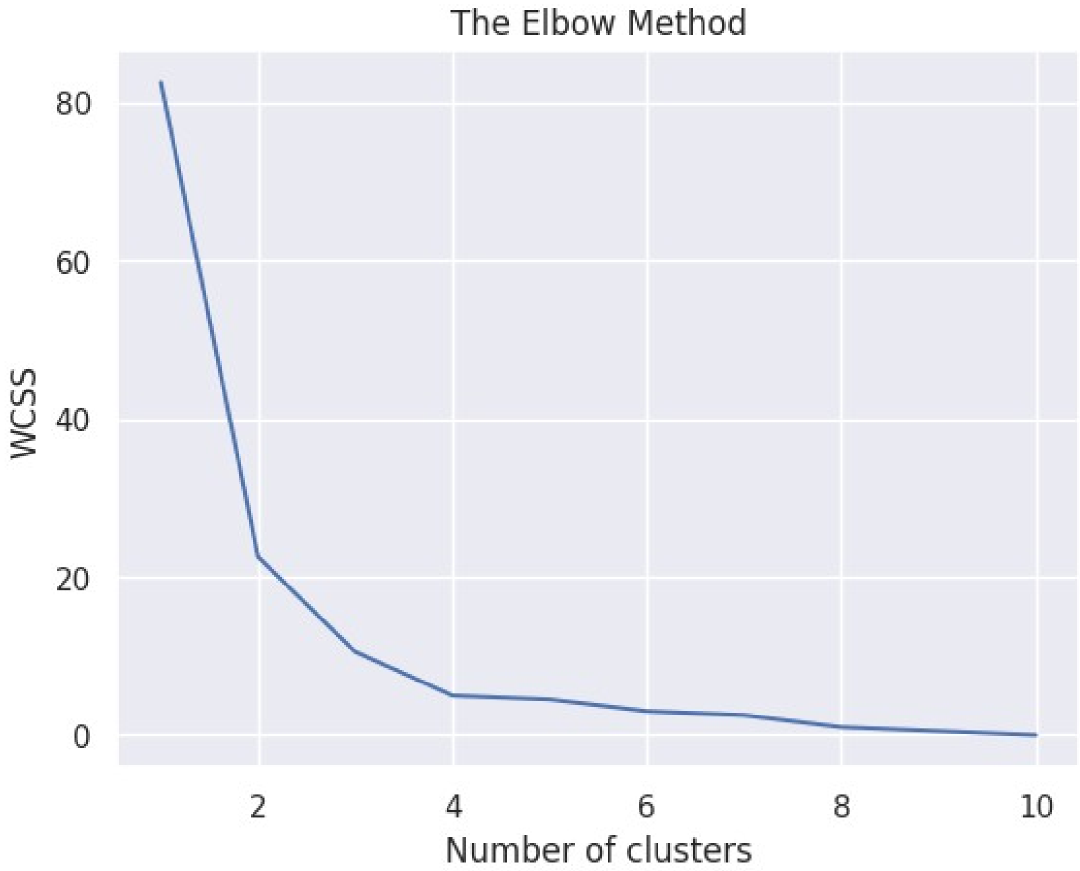
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10 entries, 0 to 9
Data columns (total 3 columns):
 #   Column      Non-Null Count  Dtype 
---  ------      --------------  ----- 
 0   ProductID   10 non-null     int64 
 1   ReviewText  10 non-null     object
 2   Rating      10 non-null     int64 
dtypes: int64(2), object(1)
memory usage: 372.0+ bytes
```

```
x=data.iloc[:,0].values
x=x.reshape(-1,1)

wcss=[]
for i in range(1,11):
  kmeans=KMeans(n_clusters=i,init='k-means++',random_state=42)
  kmeans.fit(x)
  wcss.append(kmeans.inertia_)

plt.plot(range(1,11),wcss)
plt.title('The Elbow Method')
plt.xlabel('Number of clusters')
plt.ylabel('WCSS')
plt.show()
```

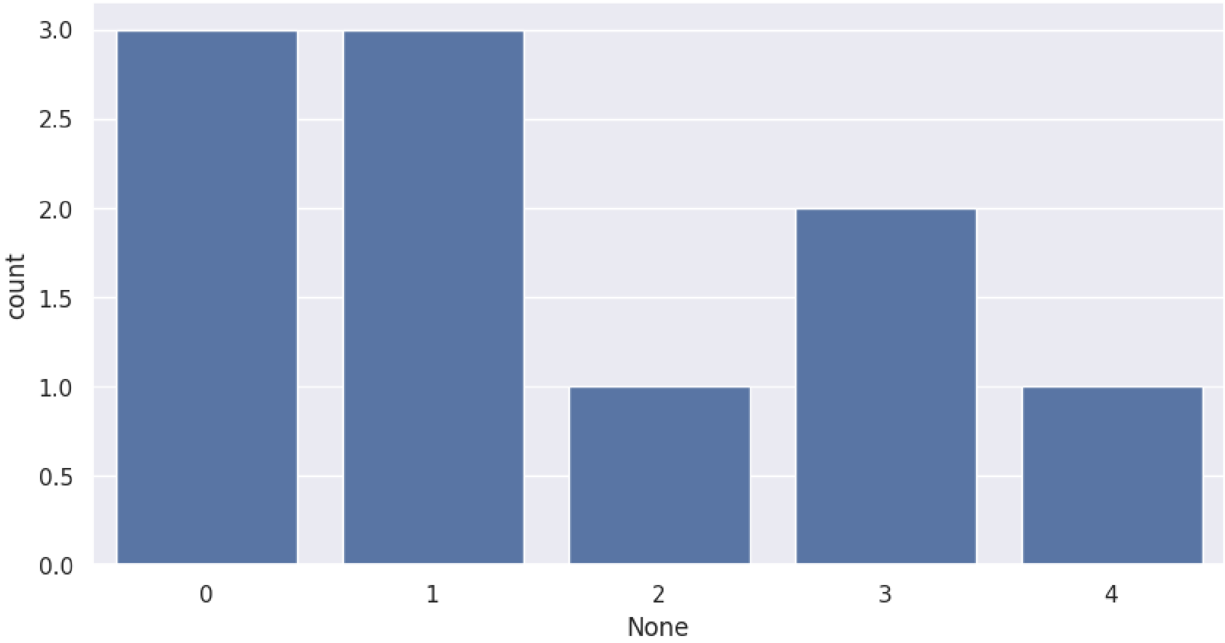


```
kmeans=KMeans(n_clusters=5,init='k-means++',random_state=42)
y_kmeans=kmeans.fit_predict(x)
print(y_kmeans)

[2 4 0 0 0 3 3 1 1 1]
```

```
c=pd.Series(y_kmeans).value_counts()
plt.figure(figsize=(10,5))
sns.barplot(x=c.index,y=c)
```

```
<Axes: xlabel='None', ylabel='count'>
```


count
None