```python
import kagglehub
path = kagglehub.dataset_download("kandij/mall-customers")
```

Start coding or generate with AI.

```python
import pandas as pd
import os
csv_file_path = os.path.join(path, "Mall_Customers.csv")
dt = pd.read_csv(csv_file_path)
display(dt.head())
```

	CustomerID	Genre	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

```python
dt.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 200 entries, 0 to 199
Data columns (total 5 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   CustomerID              200 non-null    int64
 1   Genre                   200 non-null    object
 2   Age                     200 non-null    int64
 3   Annual Income (k$)      200 non-null    int64
 4   Spending Score (1-100)  200 non-null    int64
dtypes: int64(4), object(1)
memory usage: 7.9+ KB
```

```python
x = dt.iloc[:,[3,4]].values
x
```

```
array([[ 15,  39],
       [ 15,  81],
       [ 16,   6],
       [ 16,  77],
       [ 17,  40],
       [ 17,  76],
       [ 18,   6],
       [ 18,  94],
       [ 19,   3],
```

```
       [ 19,   72],
       [ 19,   14],
       [ 19,   99],
       [ 20,   15],
       [ 20,   77],
       [ 20,   13],
       [ 20,   79],
       [ 21,   35],
       [ 21,   66],
       [ 23,   29],
       [ 23,   98],
       [ 24,   35],
       [ 24,   73],
       [ 25,    5],
       [ 25,   73],
       [ 28,   14],
       [ 28,   82],
       [ 28,   32],
       [ 28,   61],
       [ 29,   31],
       [ 29,   87],
       [ 30,    4],
       [ 30,   73],
       [ 33,    4],
       [ 33,   92],
       [ 33,   14],
       [ 33,   81],
       [ 34,   17],
       [ 34,   73],
       [ 37,   26],
       [ 37,   75],
       [ 38,   35],
       [ 38,   92],
       [ 39,   36],
       [ 39,   61],
       [ 39,   28],
       [ 39,   65],
       [ 40,   55],
       [ 40,   47],
       [ 40,   42],
       [ 40,   42],
       [ 42,   52],
       [ 42,   60],
       [ 43,   54],
       [ 43,   60],
       [ 43,   45],
       [ 43,   41],
       [ 44,   50],
       [ 44,   46],
```
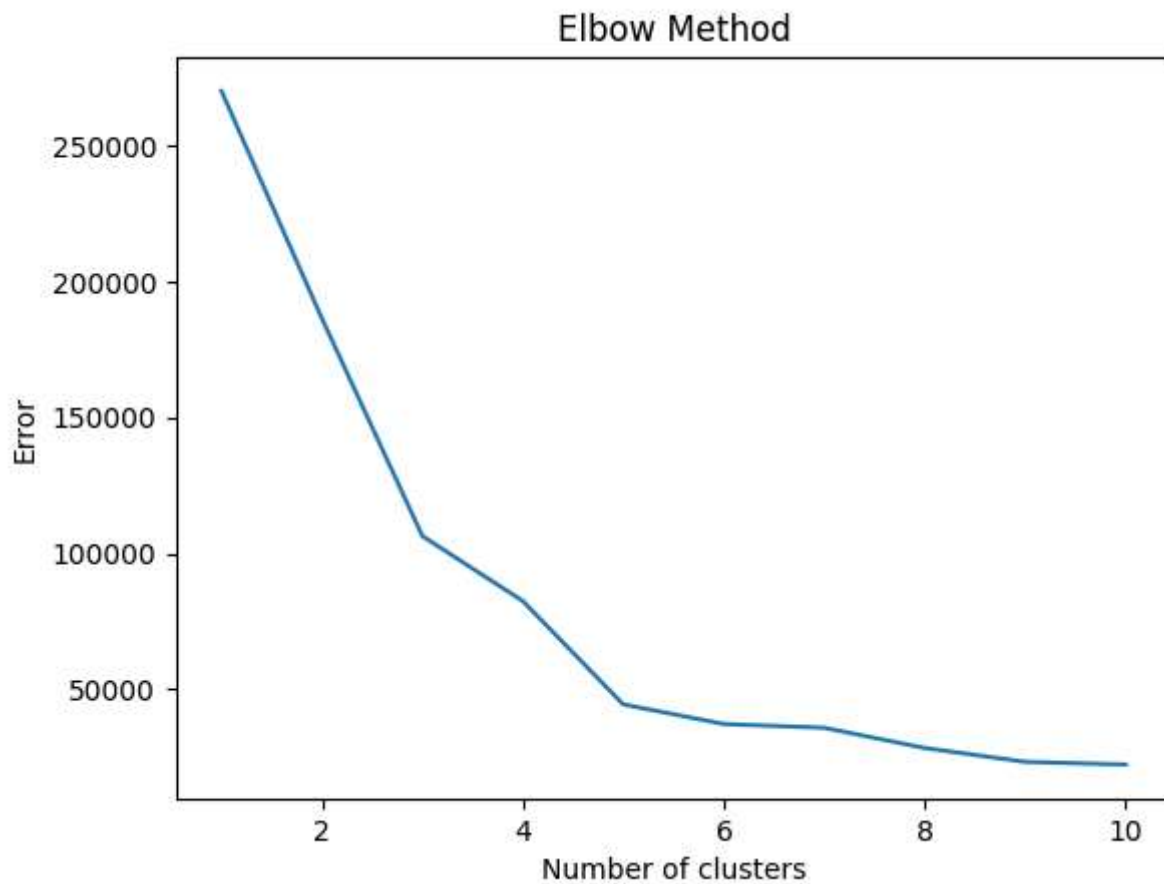
```python
from sklearn.cluster import KMeans
error = []
for i in range(1,11):
    k = KMeans(n_clusters=i, init='k-means++')
    k.fit(x)
    error.append(k.inertia_)
```

```python
import matplotlib.pyplot as plt
plt.plot(range(1,11), error)
plt.title('Elbow Method')
plt.xlabel('Number of clusters')
plt.ylabel('Error')
plt.show()
```



```python
k = KMeans(n_clusters=i, init='k-means++')
```

```python
k.fit(x)
```

```
▼        KMeans          ⓘ ⓘ
KMeans(n_clusters=10)
```

```python
y = k.predict(x)
```
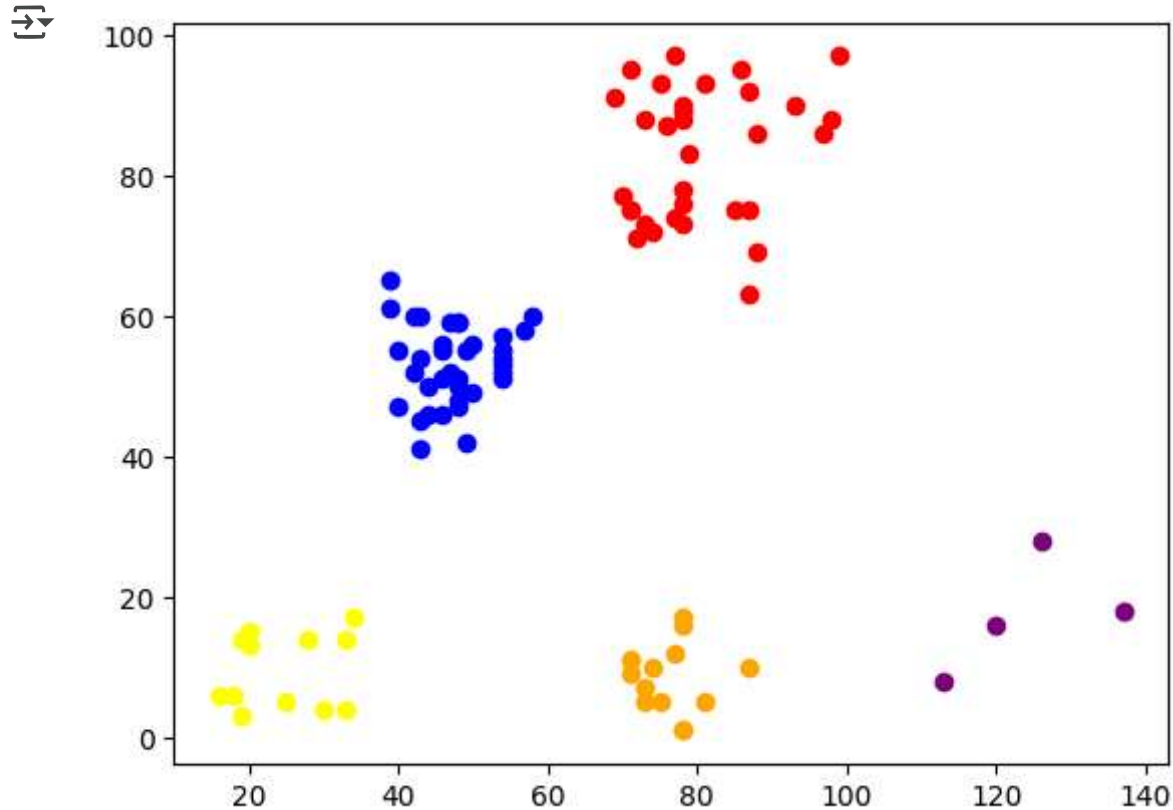
y

```
array([7, 5, 4, 5, 7, 5, 4, 5, 4, 5, 4, 5, 4, 5, 4, 5, 7, 5, 7, 5, 7, 5,
       4, 5, 4, 5, 7, 5, 7, 5, 4, 5, 4, 5, 4, 5, 4, 5, 7, 5, 7, 5, 7, 1,
       7, 1, 1, 1, 7, 7, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 8, 1, 1, 8, 1, 8, 1, 1, 8, 8, 1, 8, 1, 8,
       1, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8,
       8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 2, 6, 2, 8, 2, 3, 2, 3, 2,
       8, 2, 3, 2, 3, 2, 3, 2, 3, 2, 8, 2, 3, 2, 6, 2, 6, 2, 3, 2, 6, 2,
       3, 2, 3, 2, 3, 2, 6, 2, 3, 2, 6, 2, 6, 2, 6, 2, 6, 2, 3, 2, 6, 2,
       6, 2, 6, 2, 6, 2, 6, 2, 6, 2, 6, 9, 6, 9, 6, 9, 0, 9, 0, 9, 0, 9,
       0, 9], dtype=int32)
```

```python
plt.scatter(x[y==0, 0],x[y==0,1], c="purple")
plt.scatter(x[y==1, 0],x[y==1,1], c="blue")
plt.scatter(x[y==2, 0],x[y==2,1], c="red")
plt.scatter(x[y==3, 0],x[y==3,1], c="orange")
plt.scatter(x[y==4, 0],x[y==4,1], c="yellow")
plt.show()
```


dt[y==1]

	CustomerID	Genre	Age	Annual Income (k$)	Spending Score (1-100)
43	44	Female	31	39	61
45	46	Female	24	39	65
46	47	Female	50	40	55
47	48	Female	27	40	47
50	51	Female	49	42	52
51	52	Male	33	42	60
52	53	Female	31	43	54
53	54	Male	59	43	60
54	55	Female	50	43	45
55	56	Male	47	43	41
56	57	Female	51	44	50
57	58	Male	69	44	46
58	59	Female	27	46	51
59	60	Male	53	46	46
60	61	Male	70	46	56
61	62	Male	19	46	55
62	63	Female	67	47	52
63	64	Female	54	47	59
64	65	Male	63	48	51
65	66	Male	18	48	59
66	67	Female	43	48	50
67	68	Female	68	48	48
68	69	Male	19	48	59
69	70	Female	32	48	47
70	71	Male	70	49	55
71	72	Female	47	49	42
72	73	Female	60	50	49
73	74	Female	60	50	56
75	76	Male	26	54	54
76	77	Female	45	54	53

78	79	Female	23	54	52

```
dt[y==1]
```

81	82	Male	38	54	55
84	85	Female	21	54	57
86	87	Female	55	57	58
88	89	Female	34	58	60

	CustomerID	Genre	Age	Annual Income (k$)	Spending Score (1-100)
43	44	Female	31	39	61
45	46	Female	24	39	65
46	47	Female	50	40	55
47	48	Female	27	40	47
50	51	Female	49	42	52
51	52	Male	33	42	60
52	53	Female	31	43	54
53	54	Male	59	43	60
54	55	Female	50	43	45
55	56	Male	47	43	41
56	57	Female	51	44	50
57	58	Male	69	44	46
58	59	Female	27	46	51
59	60	Male	53	46	46
60	61	Male	70	46	56
61	62	Male	19	46	55
62	63	Female	67	47	52
63	64	Female	54	47	59
64	65	Male	63	48	51
65	66	Male	18	48	59
66	67	Female	43	48	50
67	68	Female	68	48	48
68	69	Male	19	48	59
69	70	Female	32	48	47
70	71	Male	70	49	55
71	72	Female	47	49	42
72	73	Female	60	50	49
73	74	Female	60	50	56
75	76	Male	26	54	54
76	77	Female	45	54	53