```
import numpy as np
import pandas as pd
from sklearn.neighbors import KNeighborsClassifier
from matplotlib import pyplot as plt

import kagglehub

# Download latest version
path = kagglehub.dataset_download("abdallahwagih/mall-customers-
segmentation")

print("Path to dataset files:", path)
```

```
Downloading from
https://www.kaggle.com/api/v1/datasets/download/abdallahwagih/mall-
customers-segmentation?dataset_version_number=1...

100%|██████████| 1.56k/1.56k [00:00<00:00, 2.12MB/s]

Extracting files...
Path to dataset files:
/root/.cache/kagglehub/datasets/abdallahwagih/mall-customers-
segmentation/versions/1
```

```
data=pd.read_csv("/root/.cache/kagglehub/datasets/abdallahwagih/mall-
customers-segmentation/versions/1/Mall_Customers.csv")
df=pd.DataFrame(data)
print(df)
```

```
     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
..          ...     ...  ...                 ...                     .
..
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 x 5 columns]
```

```
def make_groups(score):
    if score < 40:
        return 0  # Low Spender
    elif score < 70:
        return 1  # Average Spender
    else:
        return 2  # High Spender

df['Spending_Group'] = df['Spending Score (1-100)'].apply(make_groups)

X = df[["Age", "Annual Income (k$)"]]

y = df["Spending_Group"]

print("Data is ready for training.")
print(X.head())
print(y.head())
```

```
Data is ready for training.
   Age  Annual Income (k$)
0   19                  15
1   21                  15
2   20                  16
3   23                  16
4   31                  17
0    0
1    2
2    0
3    2
4    1
Name: Spending_Group, dtype: int64
```

```
k = 5  # Look at the 5 nearest neighbors
knn = KNeighborsClassifier(n_neighbors=k)

# 4. Train the model
knn.fit(X,y)
```

```
KNeighborsClassifier()
```

```
new_data = [[32, 80]]

prediction = knn.predict(new_data)
print(f"Prediction result: {prediction[0]}")
```

```python
if prediction[0] == 0:
    print("This customer is a LOW spender.")
elif prediction[0] == 1:
    print("This customer is an AVERAGE spender.")
else:
    print("This customer is a HIGH spender!")
```

```
Prediction result: 2
This customer is a HIGH spender!
```

```
/usr/local/lib/python3.12/dist-packages/sklearn/utils/
validation.py:2739: UserWarning: X does not have valid feature names,
but KNeighborsClassifier was fitted with feature names
  warnings.warn(
```