39. Scenario: You work as a data scientist for a marketing agency, and one of your clients is a

large e-commerce company. The company wants to understand the purchasing behavior of its

customers and segment them into different groups based on their buying patterns. The e-commerce

company has provided you with transaction data, including customer IDs, the total amount spent in

each transaction, and the number of items purchased.

Question: Build a clustering model using the K-Means algorithm to group customers based on their

spending and purchase behavior and visualize the clusters using scatter plots or other appropriate

visualizations to gain insights into customer distribution and distinguish different segments.

**import** pandas **as** pd

**import** numpy **as** np

**import** matplotlib.pyplot **as** plt

**from** sklearn.cluster **import** KMeans

**from** sklearn.preprocessing **import** StandardScaler

*# Generate synthetic data*

np.random.seed(42)

num\_customers **=** 50

data **=** {

"CustomerID": [f"CUST{i:**03d**}" **for** i **in** range(1, num\_customers **+** 1)],

"TotalAmountSpent": np.random.normal(500, 150, num\_customers).round(2),

"ItemsPurchased": np.random.poisson(10, num\_customers)

}

df **=** pd.DataFrame(data)

df.to\_csv("ecommerce\_transactions.csv", index**=False**) *# Save to CSV*

*# Load the data*

df **=** pd.read\_csv(r"C:\Users\vara prasad\Downloads\ecommerce\_transactions (1).csv")

*# Prepare features*

X **=** df[['TotalAmountSpent', 'ItemsPurchased']]

scaler **=** StandardScaler()

X\_scaled **=** scaler.fit\_transform(X)

*# KMeans clustering*

kmeans **=** KMeans(n\_clusters**=**3, random\_state**=**42)

df['Cluster'] **=** kmeans.fit\_predict(X\_scaled)

*# Visualization*

plt.figure(figsize**=**(8, 6))

scatter **=** plt.scatter(df['TotalAmountSpent'], df['ItemsPurchased'], c**=**df['Cluster'], cmap**=**'viridis')

plt.xlabel("Total Amount Spent")

plt.ylabel("Items Purchased")

plt.title("Customer Segments Based on Purchasing Behavior")

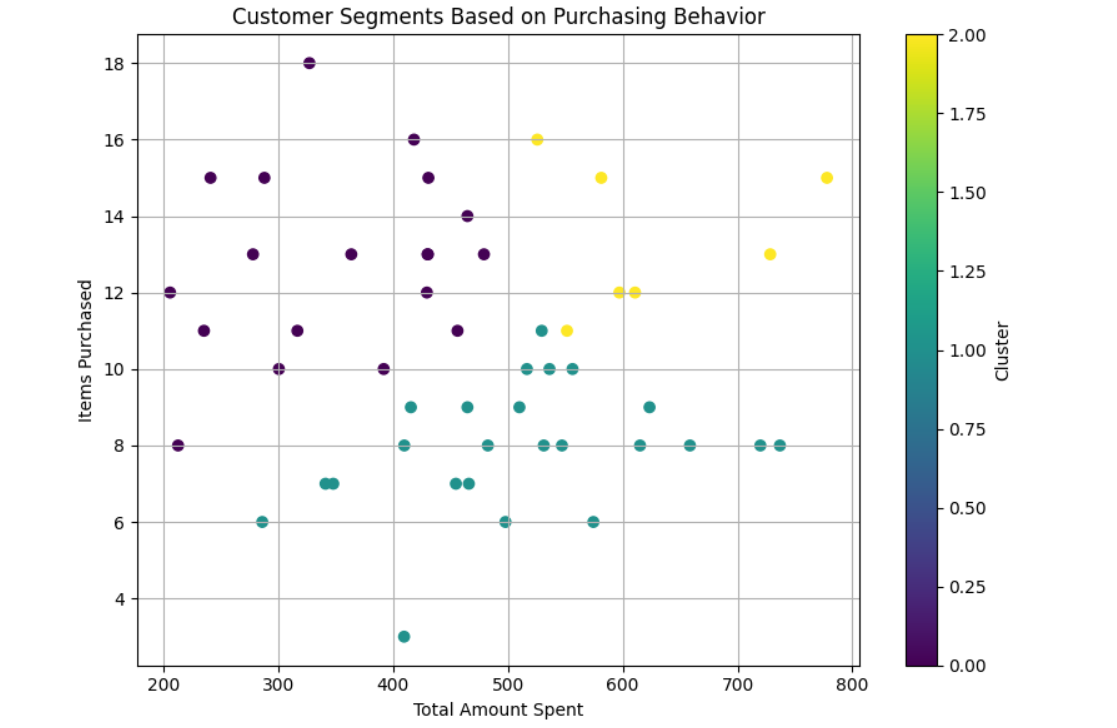
plt.colorbar(scatter, label**=**'Cluster')

plt.grid(**True**)

plt.tight\_layout()

plt.show()

output:



Dataset:

|  |  |  |
| --- | --- | --- |
| CustomerID | TotalAmountSpent | ItemsPurchased |
| CUST001 | 574.51 | 6 |
| CUST002 | 479.26 | 13 |
| CUST003 | 597.15 | 12 |
| CUST004 | 728.45 | 13 |
| CUST005 | 464.88 | 9 |
| CUST006 | 464.88 | 14 |
| CUST007 | 736.88 | 8 |
| CUST008 | 615.12 | 8 |
| CUST009 | 429.58 | 12 |
| CUST010 | 581.38 | 15 |
| CUST011 | 430.49 | 13 |
| CUST012 | 430.14 | 13 |
| CUST013 | 536.29 | 10 |
| CUST014 | 213.01 | 8 |
| CUST015 | 241.26 | 15 |
| CUST016 | 415.66 | 9 |
| CUST017 | 348.08 | 7 |
| CUST018 | 547.14 | 8 |
| CUST019 | 363.8 | 13 |
| CUST020 | 288.15 | 15 |
| CUST021 | 719.85 | 8 |
| CUST022 | 466.13 | 7 |
| CUST023 | 510.13 | 9 |
| CUST024 | 286.29 | 6 |
| CUST025 | 418.34 | 16 |
| CUST026 | 516.64 | 10 |
| CUST027 | 327.35 | 18 |
| CUST028 | 556.35 | 10 |
| CUST029 | 409.9 | 8 |
| CUST030 | 456.25 | 11 |
| CUST031 | 409.74 | 3 |
| CUST032 | 777.84 | 15 |
| CUST033 | 497.98 | 6 |
| CUST034 | 341.34 | 7 |
| CUST035 | 623.38 | 9 |
| CUST036 | 316.87 | 11 |
| CUST037 | 531.33 | 8 |
| CUST038 | 206.05 | 12 |
| CUST039 | 300.77 | 10 |
| CUST040 | 529.53 | 11 |
| CUST041 | 610.77 | 12 |
| CUST042 | 525.71 | 16 |
| CUST043 | 482.65 | 8 |
| CUST044 | 454.83 | 7 |
| CUST045 | 278.22 | 13 |
| CUST046 | 392.02 | 10 |
| CUST047 | 430.9 | 15 |
| CUST048 | 658.57 | 8 |
| CUST049 | 551.54 | 11 |
| CUST050 | 235.54 | 11 |