35. Scenario: You work as a data scientist for a retail company that operates multiple stores. The

company is interested in segmenting its customers based on their purchasing behavior to better

understand their preferences and tailor marketing strategies accordingly. To achieve this, your team

has collected transaction data from different stores, which includes customer IDs, the total amount

spent in each transaction, and the frequency of visits.

Question: Your task is to build a clustering model using the K-Means algorithm to group

customers into distinct segments based on their spending patterns.

Code:

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

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

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

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

*# Step 1: Load the data*

df **=** pd.read\_csv(r"C:\Users\jampa\Downloads\customer\_transactions.csv")

*# Step 2: Preprocess the data*

features **=** df[['TotalAmountSpent', 'Frequency']]

scaler **=** StandardScaler()

scaled\_features **=** scaler.fit\_transform(features)

*# Step 3: Apply K-Means clustering*

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

df['Cluster'] **=** kmeans.fit\_predict(scaled\_features)

*# Step 4: Visualize the clusters*

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

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

plt.xlabel('Total Amount Spent')

plt.ylabel('Frequency')

plt.title('Customer Segments based on Spending Behavior')

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

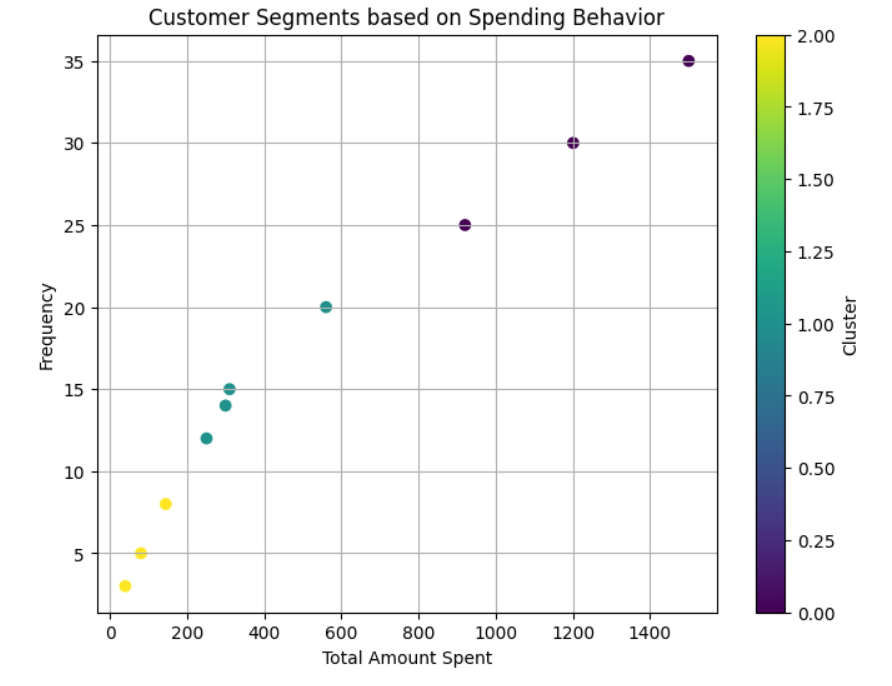
plt.grid(**True**)

plt.show()

*# Optional: Save the clustered data*

df.to\_csv("clustered\_customers.csv", index**=False**)

output:



Dataset:

|  |  |  |
| --- | --- | --- |
| CustomerID | TotalAmountSpent | Frequency |
| 1001 | 250.75 | 12 |
| 1002 | 145 | 8 |
| 1003 | 560.4 | 20 |
| 1004 | 1200.6 | 30 |
| 1005 | 80.9 | 5 |
| 1006 | 300 | 14 |
| 1007 | 920.5 | 25 |
| 1008 | 40.25 | 3 |
| 1009 | 1500 | 35 |
| 1010 | 310.4 | 15 |