In [2]:
```
from sklearn.preprocessing import OneHotEncoder, StandardScaler
from sklearn.metrics.pairwise import cosine_similarity
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

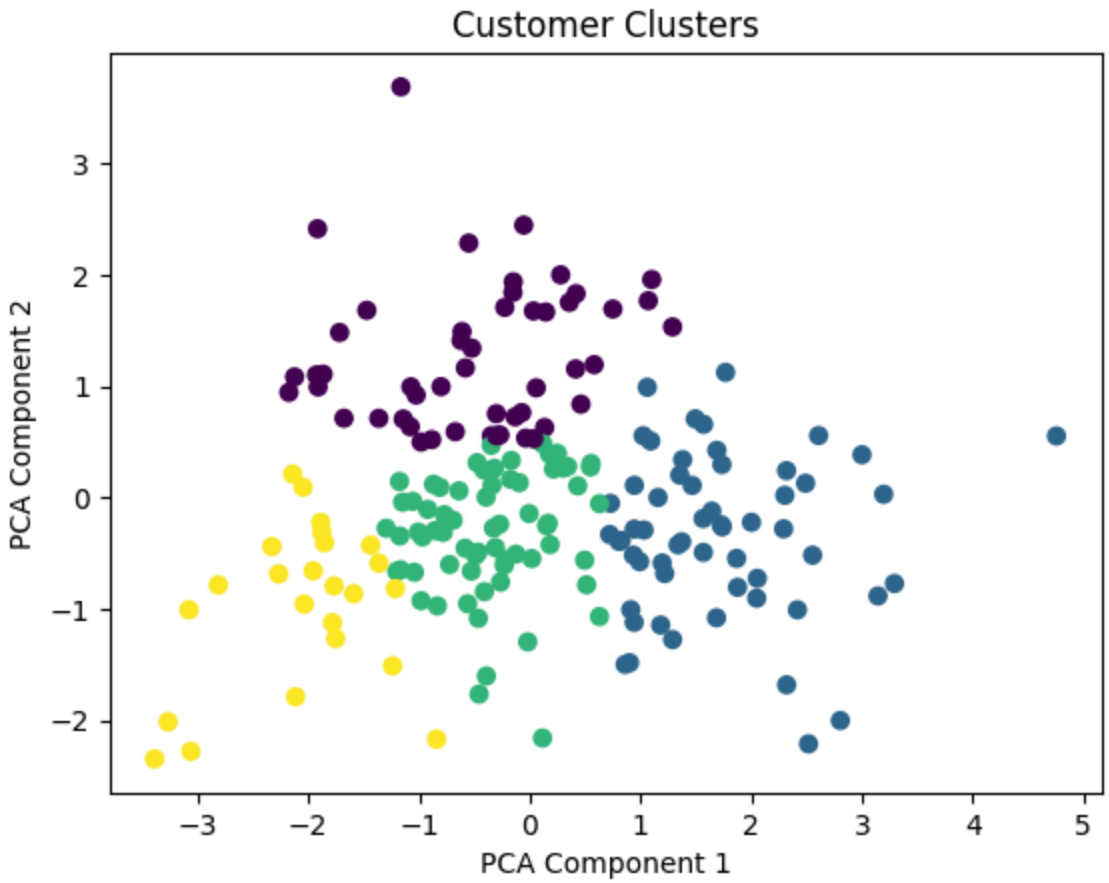
In [3]:
```
customers = pd.read_csv("./Customers.csv")
products = pd.read_csv("./Products.csv")
transactions = pd.read_csv("./Transactions.csv")
customers['SignupDate'] = pd.to_datetime(customers['SignupDate'])
transactions['TransactionDate'] = pd.to_datetime(transactions['TransactionDate
merged_data = transactions.merge(customers, on="CustomerID").merge(products, c
merged_data.head(5)
```

Out[3]:

	TransactionID	CustomerID	ProductID	TransactionDate	Quantity	TotalValue	Price_x	Custom
0	T00001	C0199	P067	2024-08-25 12:38:23	1	300.68	300.68	Andrea
1	T00112	C0146	P067	2024-05-27 22:23:54	1	300.68	300.68	Brittan
2	T00166	C0127	P067	2024-04-25 07:38:55	1	300.68	300.68	Kathryn
3	T00272	C0087	P067	2024-03-26 22:55:37	2	601.36	300.68	Travis (
4	T00363	C0070	P067	2024-03-21 15:10:10	3	902.04	300.68	Timot

In [7]:
```
from sklearn.cluster import KMeans
from sklearn.decomposition import PCA
from sklearn.metrics import davies_bouldin_score, silhouette_score
from sklearn.preprocessing import StandardScaler
import matplotlib.pyplot as plt

customer_profiles = merged_data.groupby('CustomerID').agg({
    'TotalValue': 'sum',
    'Quantity': 'sum',
    'Price_x': 'mean'
}).reset_index()


scaler = StandardScaler()
clustering_features = customer_profiles[['TotalValue', 'Quantity', 'Price_x']]
scaled_features = scaler.fit_transform(clustering_features)


kmeans = KMeans(n_clusters=4, random_state=42)
clusters = kmeans.fit_predict(scaled_features)
customer_profiles['Cluster'] = clusters


db_index = davies_bouldin_score(scaled_features, clusters)
sil_score = silhouette_score(scaled_features, clusters)
print("Davies-Bouldin Index:", db_index)
print("Silhouette Score:", sil_score)


pca = PCA(n_components=2)
reduced_features = pca.fit_transform(scaled_features)
plt.scatter(reduced_features[:, 0], reduced_features[:, 1], c=clusters, cmap='
plt.title("Customer Clusters")
plt.xlabel("PCA Component 1")
plt.ylabel("PCA Component 2")
plt.show()
```

```
Davies-Bouldin Index: 0.9448437495439578
Silhouette Score: 0.33071125179757666

C:\Users\peddi\AppData\Local\Programs\Python\Python311\Lib\site-packages\skle
arn\cluster\_kmeans.py:1412: FutureWarning: The default value of `n_init` wil
l change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to su
ppress the warning
  super()._check_params_vs_input(X, default_n_init=10)
```


Customer Clusters
PCA Component 2
PCA Component 1
3
2
1
0
−1
−2
−3
−2
−1
0
1
2
3
4
5

In [ ]: