

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.metrics.pairwise import cosine_similarity
from sklearn.preprocessing import StandardScaler, LabelEncoder
from sklearn.cluster import KMeans
from sklearn.metrics import davies_bouldin_score
import json

customers = pd.read_csv("Customers.csv")
products = pd.read_csv("Products.csv")
transactions = pd.read_csv("/content/Transactions - Transactions.csv")

# Task 3: Customer Segmentation / Clustering
# Clustering Preparation
kmeans_data = normalized_data # Using normalized customer profile data

# K-Means Clustering
kmeans = KMeans(n_clusters=4, random_state=42)
kmeans_labels = kmeans.fit_predict(kmeans_data)

# Add cluster labels to customer profile
customer_profile['Cluster'] = kmeans_labels

# Evaluate Clustering
db_index = davies_bouldin_score(kmeans_data, kmeans_labels)
print("Davies-Bouldin Index:", db_index)

# Visualize Clusters
plt.scatter(kmeans_data[:, 0], kmeans_data[:, 1], c=kmeans_labels, cmap='viridis', alpha=0.5)
plt.title('Customer Clusters (K-Means)')
plt.xlabel('Feature 1')
plt.ylabel('Feature 2')
plt.show()

# Save Clustering Results
customer_profile.to_csv("Customer_Clusters.csv", index=False)
```

↔ Davies-Bouldin Index: 1.1977479161848428



Start coding or [generate](#) with AI.