

## TASK 3

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

from sklearn.preprocessing import StandardScaler
from sklearn.cluster import KMeans
from sklearn.metrics import davies_bouldin_score, silhouette_score

# Load the datasets
customers = pd.read_csv('Customers.csv')
transactions = pd.read_csv('Transactions.csv')

# Merge datasets based on CustomerID
merged_data = transactions.merge(customers, on="CustomerID")

# Aggregating transaction information
customer_data = merged_data.groupby("CustomerID").agg({
    "TotalValue": "sum", # Total transaction value for each customer
    "ProductID": "count" # Number of transactions for each customer
}).rename(columns={"ProductID": "TransactionCount"}).reset_index()

# Merge with customer profile data (if applicable)
customer_data = customer_data.merge(customers, on="CustomerID")
```

```
# Prepare features for clustering
features = customer_data[["TotalValue", "TransactionCount"]]

# Standardize the data
scaler = StandardScaler()
X_scaled = scaler.fit_transform(features)

# Clustering using KMeans
n_clusters = 4 # You can try values between 2 and 10
kmeans = KMeans(n_clusters=n_clusters, random_state=42)
labels = kmeans.fit_predict(X_scaled)
customer_data["Cluster"] = labels

# Evaluate the clustering
db_index = davies_bouldin_score(X_scaled, labels)
silhouette_avg = silhouette_score(X_scaled, labels)

# Print clustering results
print(f"Number of Clusters: {n_clusters}")
print(f"Davies-Bouldin Index: {db_index:.2f}")
print(f"Silhouette Score: {silhouette_avg:.2f}")

# Visualize clusters
plt.figure(figsize=(10, 7))
sns.scatterplot(
    x=features["TotalValue"],
    y=features["TransactionCount"],
    hue=customer_data["Cluster"],
    palette="viridis"
```

```
)  
plt.title(f"Customer Segmentation (n_clusters={n_clusters})")  
plt.xlabel("Total Transaction Value")  
plt.ylabel("Transaction Count")  
plt.legend(title="Cluster", loc="best")  
plt.tight_layout()  
plt.show()  
  
# Save cluster results  
customer_data.to_csv("Customer_Segmentation_Results.csv", index=False)
```

### **Example Output (Simulated):**

- **Davies-Bouldin Index:** 1.09
- **Silhouette Score:** 0.62
- **Cluster Sizes:**
  - Cluster 0: 150 customers
  - Cluster 1: 80 customers
  - Cluster 2: 120 customers
  - Cluster 3: 100 customers

The generated .csv file (Customer\_Segmentation\_Results.csv) contains the following columns:

- CustomerID
- TotalValue
- TransactionCount
- Cluster

