

Task 3: Customer Segmentation / Clustering

Step 1: Preprocessing for Clustering 1.

****Feature Engineering**:**

- Use customer profile features (e.g., total spent, frequency of transactions) and product-related features (e.g., category, average price).

Step 2: Apply Clustering Algorithm

You can use K-Means clustering or DBSCAN for customer segmentation.

1. ****K-Means Clustering**:**

```
from sklearn.cluster import KMeans
from sklearn.metrics import davies_bouldin_score
```

Apply KMeans

```
kmeans = KMeans(n_clusters=5, random_state=42)
customer_data['Cluster'] = kmeans.fit_predict(customer_features)
```

Evaluate with DB Index

```
db_index = davies_bouldin_score(customer_features,
customer_data['Cluster'])
print(f'Davies-Bouldin Index: {db_index}')
```

Step 3: Visualization of Clusters

1. ****Visualization**:** Use PCA or t-SNE for dimensionality reduction and visualize the clusters.

```
from sklearn.decomposition import PCA
```

```
pca = PCA(n_components=2)
pca_result = pca.fit_transform(customer_features)
```

```
plt.figure(figsize=(10, 6))
sns.scatterplot(x=pca_result[:, 0], y=pca_result[:, 1], hue=customer_data['Cluster'])
plt.title('Customer Segments')
plt.show()
```

