Approach

1. **Data Preparation**:
    - Merge customer and transaction datasets.
    - Generate features such as total transactions, total spending, product categories purchased, and customer region.
2. **Clustering**:
    - Choose an appropriate clustering algorithm (e.g., KMeans).
    - Optimize the number of clusters using metrics like the Davies-Bouldin (DB) Index.
3. **Visualization**:
    - Visualize the clusters using a scatter plot or other relevant techniques.

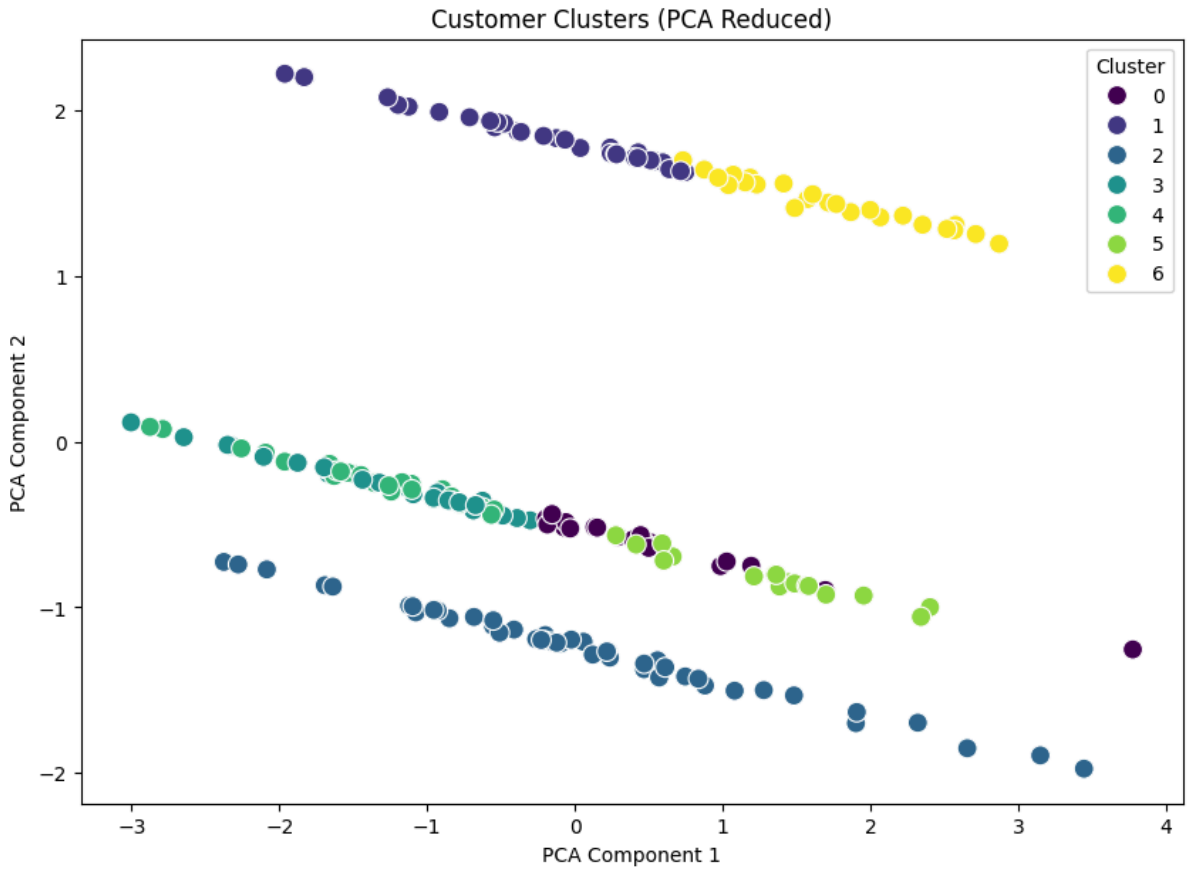
Steps Explained

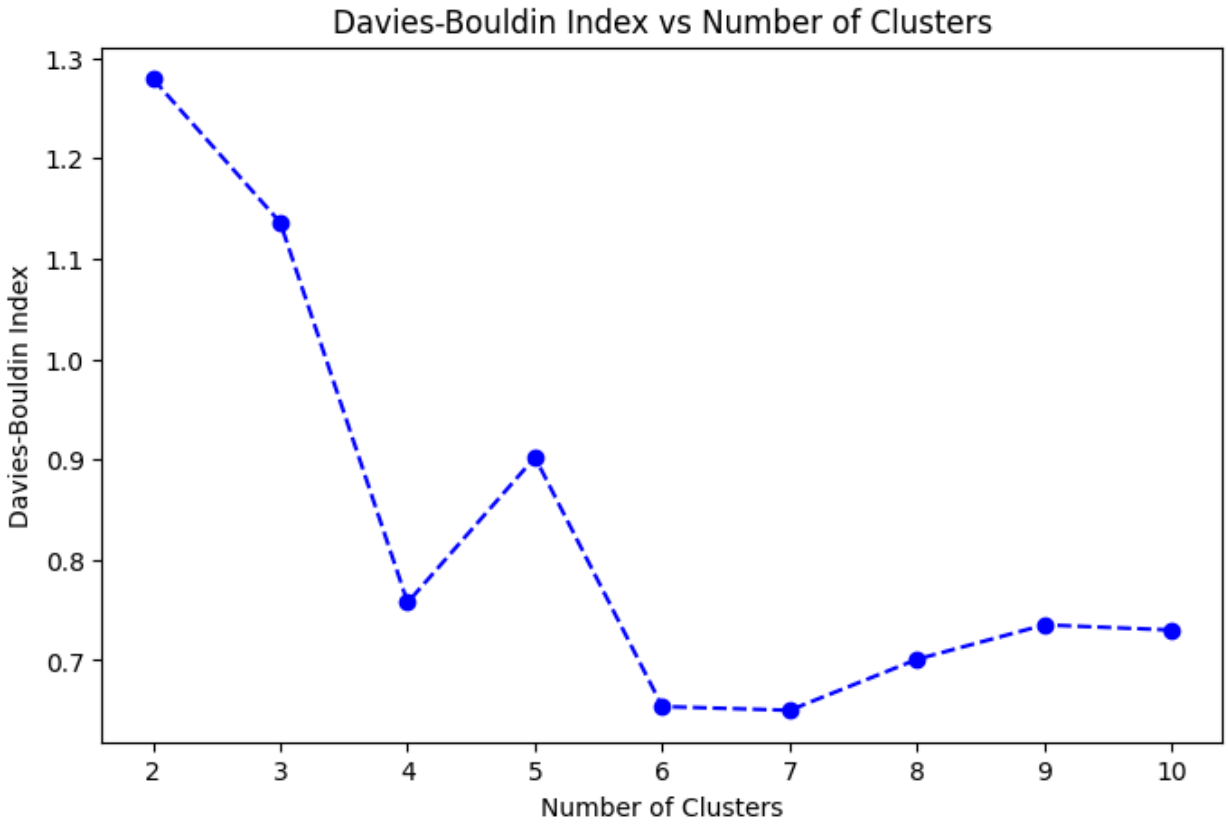
1. **Data Preparation**:
    - Aggregates transactional data to calculate features like `TotalTransactions` and `TotalSpending`.
    - Adds region data from the `Customers.csv` file and performs one-hot encoding for categorical variables.
2. **Davies-Bouldin Index**:
    - Evaluates clustering performance for `k` values (2 to 10).
    - Selects the number of clusters (`optimal_k`) with the lowest DB Index.
3. **Clustering**:
    - Uses the KMeans algorithm with the optimal number of clusters.
    - Assigns cluster labels to customers.
4. **Visualization**:
    - Reduces dimensionality using PCA and visualizes clusters in a 2D scatter plot.
5. **Output**:
    - Saves customer data with cluster assignments as a CSV file.

Output CSV File

The `FirstName_LastName_Clustering.csv` will have the following structure:

CustomerID	Customer Name	Region	TotalTransactions	TotalSpending	Region_Asia	Region_Europe	...	Cluster
C0001	John Doe	Asia	10	2000.50	1	0	...	1
C0002	Jane Smith	Europe	5	1200.75	0	1	...	2




Davies-Bouldin Index vs Number of Clusters
Davies-Bouldin Index
1.3
1.2
1.1
1.0
0.9
0.8
0.7
2
3
4
5
6
7
8
9
10
Number of Clusters