कौशलम् सर्वसाधनम्
KIET GROUP OF INSTITUTIONS
www.kiet.edu
Delhi-NCR, Ghaziabad

KIET
GROUP OF INSTITUTIONS
Connecting Life with Learning

Assessment Report

on

“Customer Segmentation in E-commerce”

submitted as partial fulfillment for the award of

BACHELOR OF TECHNOLOGY

DEGREE

SESSION 2024-25

In

Computer Science & Engineering

(Artificial Intelligence and Machine Learning)

By

Shreya (202401100400179)

Under the supervision of

“Mr. Abhishek Shukla”

KIET

Group of Institutions

May, 2025

Introduction

In the digital commerce landscape, understanding customer behaviour is key to building effective marketing strategies and improving user experience. One way to do this is through customer segmentation, which involves grouping customers into clusters based on common characteristics.

This project applies KMeans clustering, an unsupervised machine learning technique, to segment customers of an e-commerce platform. The segmentation is based on features like purchase history, browsing patterns, and spending behaviour.

Methodology

The project follows these steps:

a. Data Collection:
The dataset 9. Customer Segmentation in E-commerce.csv contains numeric customer features related to their shopping habits.

b. Data Preprocessing:

- Removed missing values
- Selected only numeric columns for clustering
- Standardized data using StandardScaler

c. Clustering Technique:

- Used KMeans Clustering to group customers
- Determined the optimal number of clusters using the Elbow Method

d. Visualization:

- Applied PCA (Principal Component Analysis) to reduce high-dimensional data into 2D
- Plotted customer clusters for better understanding

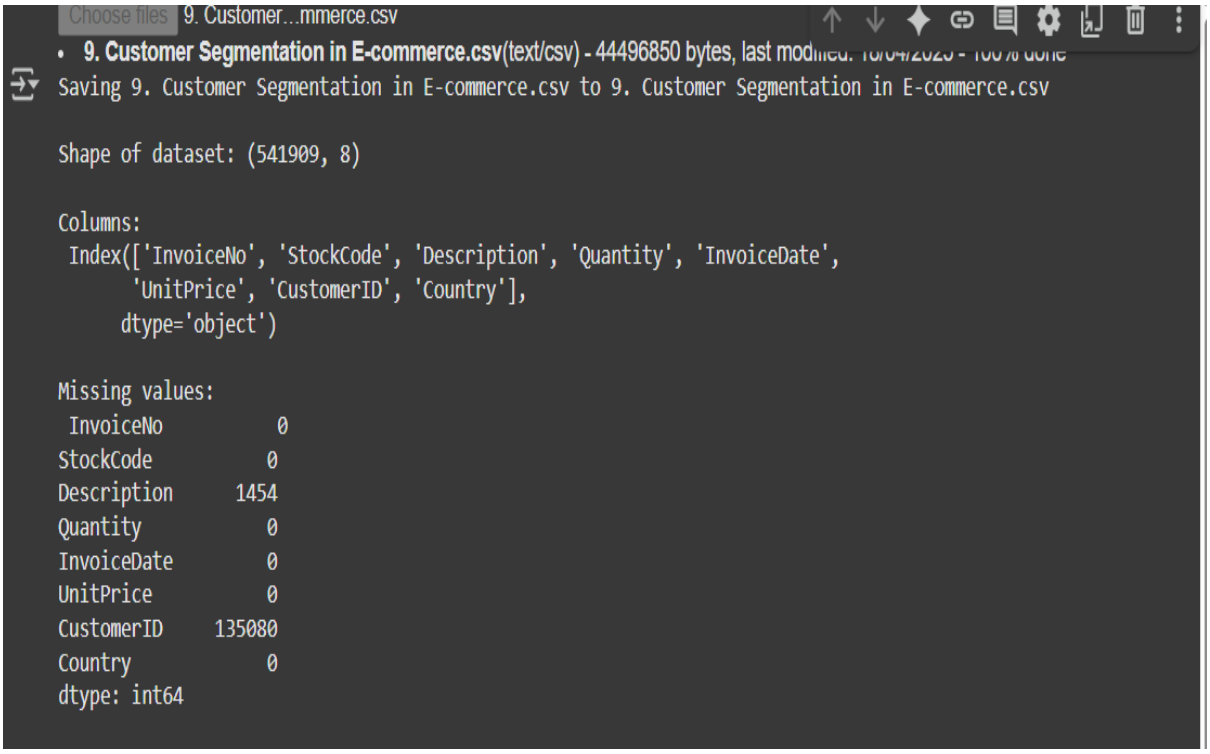
e. Output Export:
Final data with clusters was saved as Customer_Segmentation_Output.csv

Code

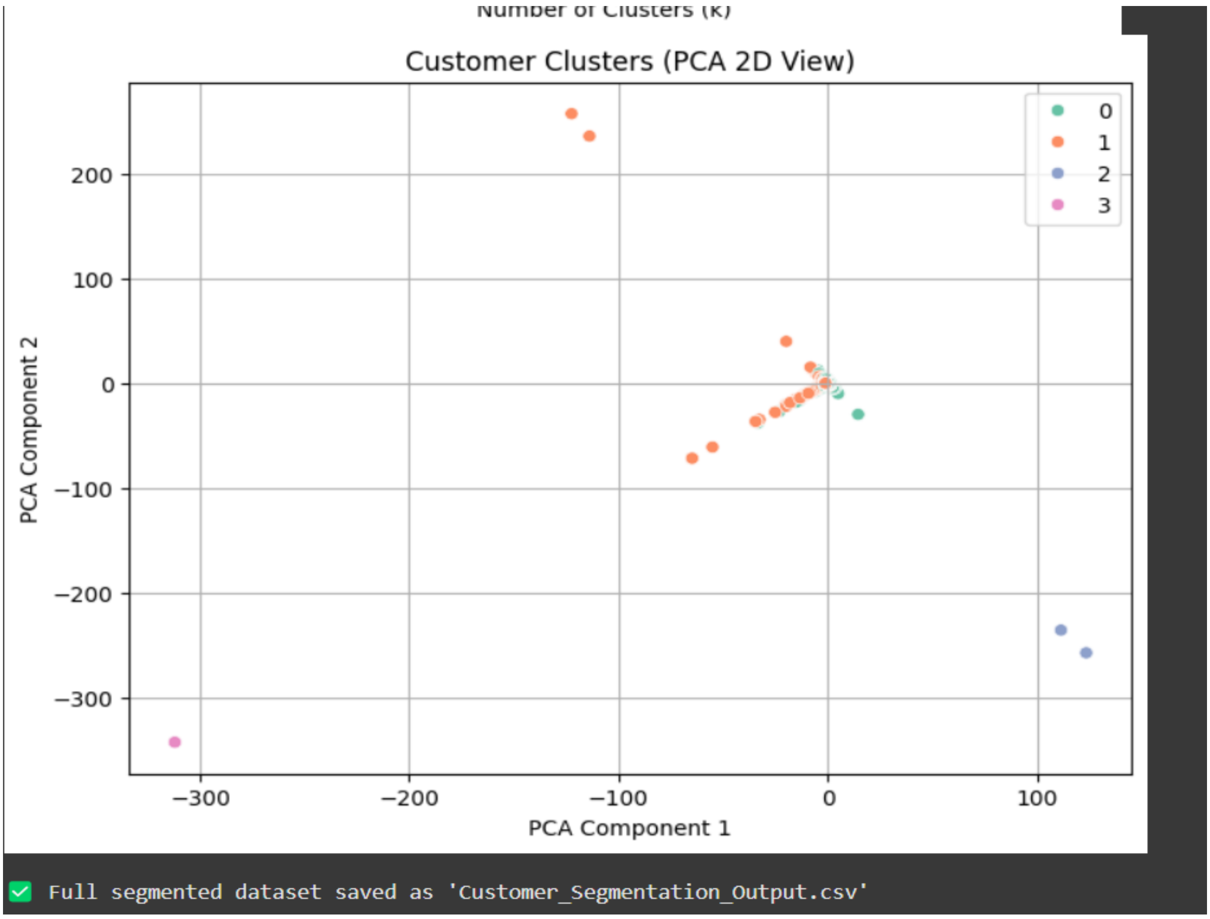
```
import pandas as pd

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.decomposition import PCA
from google.colab import files

uploaded = files.upload()
df = pd.read_csv("9. Customer Segmentation in E-commerce.csv")
df.dropna(inplace=True)
numeric_df = df.select_dtypes(include=[np.number])
scaler = StandardScaler()
scaled_data = scaler.fit_transform(numeric_df)

wcss = []
for i in range(1, 11):
   kmeans = KMeans(n_clusters=i, init='k-means++', random_state=42)
   kmeans.fit(scaled_data)
   wcss.append(kmeans.inertia_)

plt.plot(range(1, 11), wcss, marker='o')
plt.title('Elbow Method')
plt.xlabel('Number of Clusters')
plt.ylabel('WCSS')
plt.grid(True)
plt.show()

kmeans = KMeans(n_clusters=4, init='k-means++', random_state=42)
df['Cluster'] = kmeans.fit_predict(scaled_data)
```

```
pca = PCA(n_components=2)

pca_data = pca.fit_transform(scaled_data)

df['PCA1'] = pca_data[:, 0]

df['PCA2'] = pca_data[:, 1]


sns.scatterplot(x='PCA1', y='PCA2', hue='Cluster', data=df, palette='Set2')

plt.title('Customer Clusters')

plt.show()


df.to_csv("Customer_Segmentation_Output.csv", index=False)
```

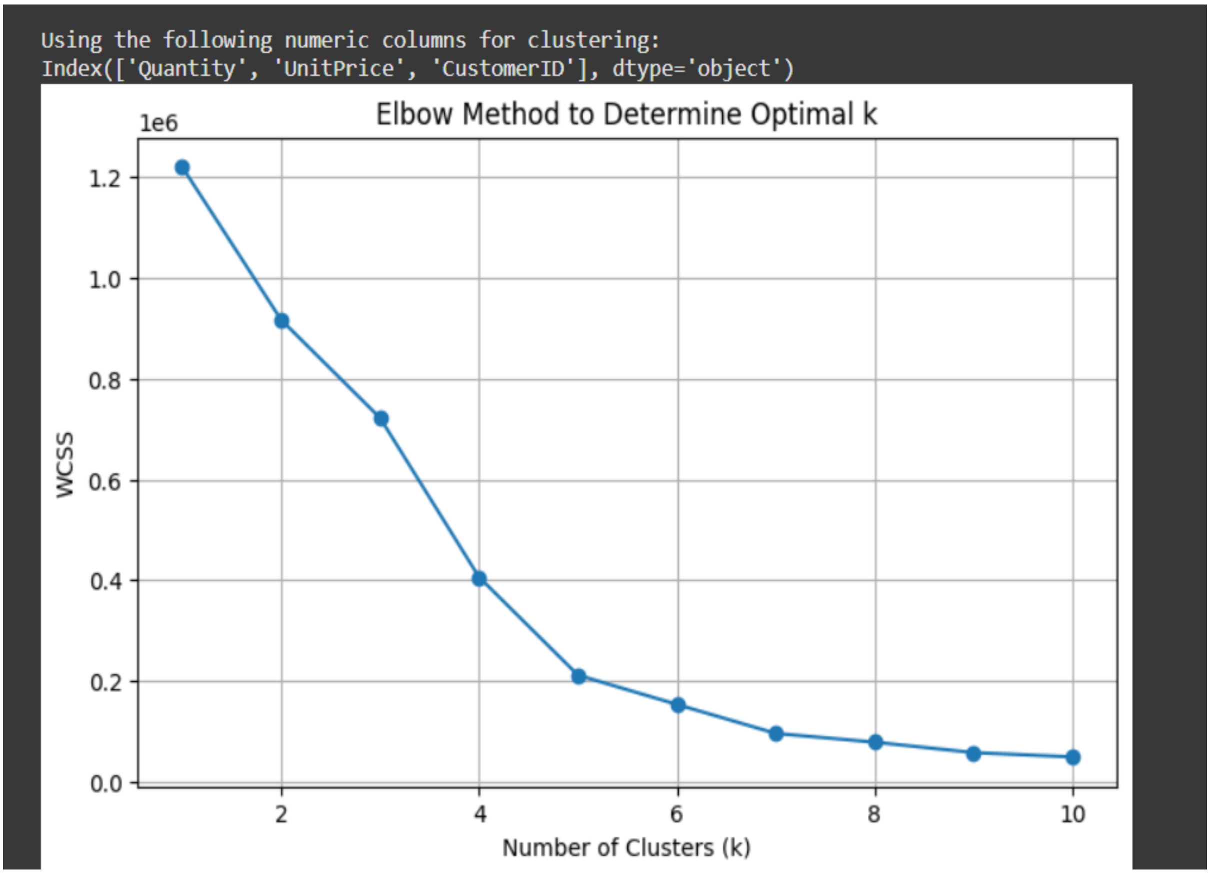
Output



```
• 9. Customer Segmentation in E-commerce.csv(text/csv) - 44496850 bytes, last modified: 10/04/2025 - 100% done
Saving 9. Customer Segmentation in E-commerce.csv to 9. Customer Segmentation in E-commerce.csv

Shape of dataset: (541909, 8)

Columns:
 Index(['InvoiceNo', 'StockCode', 'Description', 'Quantity', 'InvoiceDate',
       'UnitPrice', 'CustomerID', 'Country'],
      dtype='object')

Missing values:
 InvoiceNo         0
StockCode         0
Description    1454
Quantity          0
InvoiceDate       0
UnitPrice         0
CustomerID   135080
Country           0
dtype: int64
```



```
✅ Full segmented dataset saved as 'Customer_Segmentation_Output.csv'
```



```
Using the following numeric columns for clustering:
Index(['Quantity', 'UnitPrice', 'CustomerID'], dtype='object')
```

References

- Dataset provided for academic use: Customer Segmentation in E-commerce.csv
- Python libraries: Scikit-learn, Pandas, Matplotlib, Seaborn
- Google Colab: Used for writing and executing the code
- PCA and KMeans theory: scikit-learn documentation