A 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)

By

Uday Singh Kushwaha (202401100300269)

Under the supervision of

"Mstr. Abhishek Shukla Sir"

KIET Group of Institutions, Ghaziabad

Affiliated to

Dr. A.P.J. Abdul Kalam Technical University, Lucknow

(Formerly UPTU)

April, 2025

Introduction

Customer segmentation is a key strategy in e-commerce to group customers based on their behavior and preferences. The goal is to understand different types of shoppers by analyzing their purchasing habits and browsing behavior. This helps businesses personalize marketing, improve customer service, and optimize product recommendations.

In this study, we aim to identify meaningful customer clusters using unsupervised learning (K-Means clustering) on an e-commerce dataset.

Methodology:

Step 1: Data Upload & Cleaning

A CSV file containing customer data was uploaded.
Null values were removed, and only numerical features were used for clustering.

Step 2: Feature Scaling

StandardScaler was used to normalize features to ensure fair clustering.

Step 3: Finding Optimal Clusters

The Elbow Method was applied by plotting Within-Cluster Sum of Squares (WCSS) for cluster counts from 1 to 10.
The "elbow point" helped decide the optimal number of clusters (typically where the WCSS starts flattening).

Step 4: K-Means Clustering

KMeans algorithm was applied with the selected number of clusters (e.g., 4).
Each customer was assigned to one of the clusters.

Step 5: PCA for Visualization

Principal Component Analysis (PCA) reduced data to 2 dimensions for visualization.

Scatter plot was used to display distinct clusters.

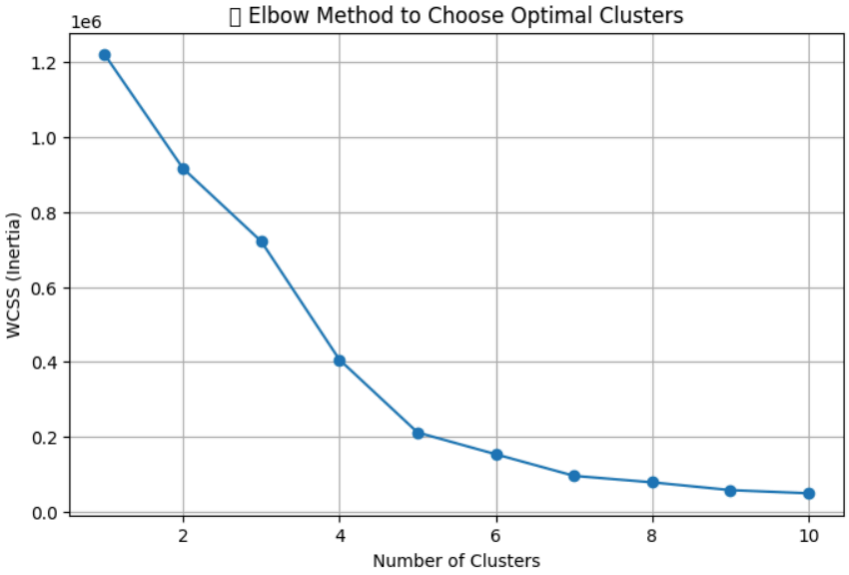
Code

```
import pandas as pd
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
import io

# Upload CSV file
uploaded = files.upload()

# Load the CSV file
for filename in uploaded.keys():
    df = pd.read_csv(io.BytesIO(uploaded[filename]))
    print(f"✅ Loaded file: {filename}")

# Clean and preprocess
df_cleaned = df.dropna()
df_numeric = df_cleaned.select_dtypes(include=['float64', 'int64'])
scaler = StandardScaler()
df_scaled = scaler.fit_transform(df_numeric)

# Elbow method to find optimal clusters
wcss = []
for i in range(1, 11):
    kmeans = KMeans(n_clusters=i, random_state=42)
    kmeans.fit(df_scaled)
    wcss.append(kmeans.inertia_)

# Plot the Elbow curve
plt.figure(figsize=(8,5))
plt.plot(range(1, 11), wcss, marker='o')
```

```
plt.title('Elbow Method - Find Optimal Clusters')
plt.xlabel('Number of Clusters')
plt.ylabel('WCSS')
plt.grid(True)
plt.show()

# Apply KMeans with optimal clusters
optimal_k = 4  # Adjust based on Elbow curve
kmeans = KMeans(n_clusters=optimal_k, random_state=42)
df_cleaned['Cluster'] = kmeans.fit_predict(df_scaled)

# Visualize clusters with PCA
pca = PCA(n_components=2)
pca_components = pca.fit_transform(df_scaled)

plt.figure(figsize=(8,6))
sns.scatterplot(x=pca_components[:,0], y=pca_components[:,1],
hue=df_cleaned['Cluster'], palette='Set2')
plt.title('Customer Segments (PCA)')
plt.xlabel('PCA 1')
plt.ylabel('PCA 2')
plt.show()
```

Output


Customer Segmentation Visualization (PCA)
Cluster
0.0
1.0
2.0
3.0
Principal Component 2
Principal Component 1
200
100
0
−100
−200
−300
−300
−200
−100
0
100


Elbow Method to Choose Optimal Clusters
1e6
WCSS (Inertia)
Number of Clusters
1.2
1.0
0.8
0.6
0.4
0.2
0.0
2
4
6
8
10

References/Credits

- Dataset: Provided CSV file – *"9. Customer Segmentation in E-commerce.csv"*
- Libraries Used: pandas, scikit-learn, matplotlib, seaborn
- Algorithms: K-Means Clustering, PCA
- Elbow Method Reference: Wikipedia - Elbow Method