Redesigning Common Mind & Business Towards Excellence

3P urpose rocess eople Culture

sns INSTITUTIONS

www.snsgroups.com

1st GenAI Powered Design Thinking FrameWork

Build an Entrepreneurial Mindset Through Our Design Thinking FrameWork

SNS COLLEGE OF TECHNOLOGY

Coimbatore-35
An Autonomous Institution

Accredited by NBA – AICTE and Accredited by NAAC – UGC with 'A++' Grade
ASPproved by AICTE, New Delhi & Affiliated to Anna University, Chennai

DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

23AMB201 - MACHINE LEARNING

II YEAR IV SEM

UNIT III – UNSUPERVISED LEARNING ALGORITHMS

TOPIC 11 – Implement Market Segmentation using K-Means

K-means

```
[ ]  import pandas as pd
     import numpy as np
     import matplotlib.pyplot as plt
     import seaborn as sns
     from sklearn.cluster import KMeans
     from sklearn.preprocessing import StandardScaler
```

```
# Simulated dataset (could be customer data)
data = {
    'CustomerID': [1, 2, 3, 4, 5, 6, 7, 8, 9, 10],
    'Age': [25, 34, 22, 45, 52, 23, 40, 60, 30, 27],
    'AnnualIncome': [40000, 60000, 35000, 80000, 90000, 30000, 70000, 100000, 50000, 45000],
    'SpendingScore': [60, 70, 40, 80, 90, 30, 70, 95, 50, 65]
}
```

```
df = pd.DataFrame(data)
# Drop CustomerID for clustering
X = df.drop('CustomerID', axis=1)
```

Customer Segmentation

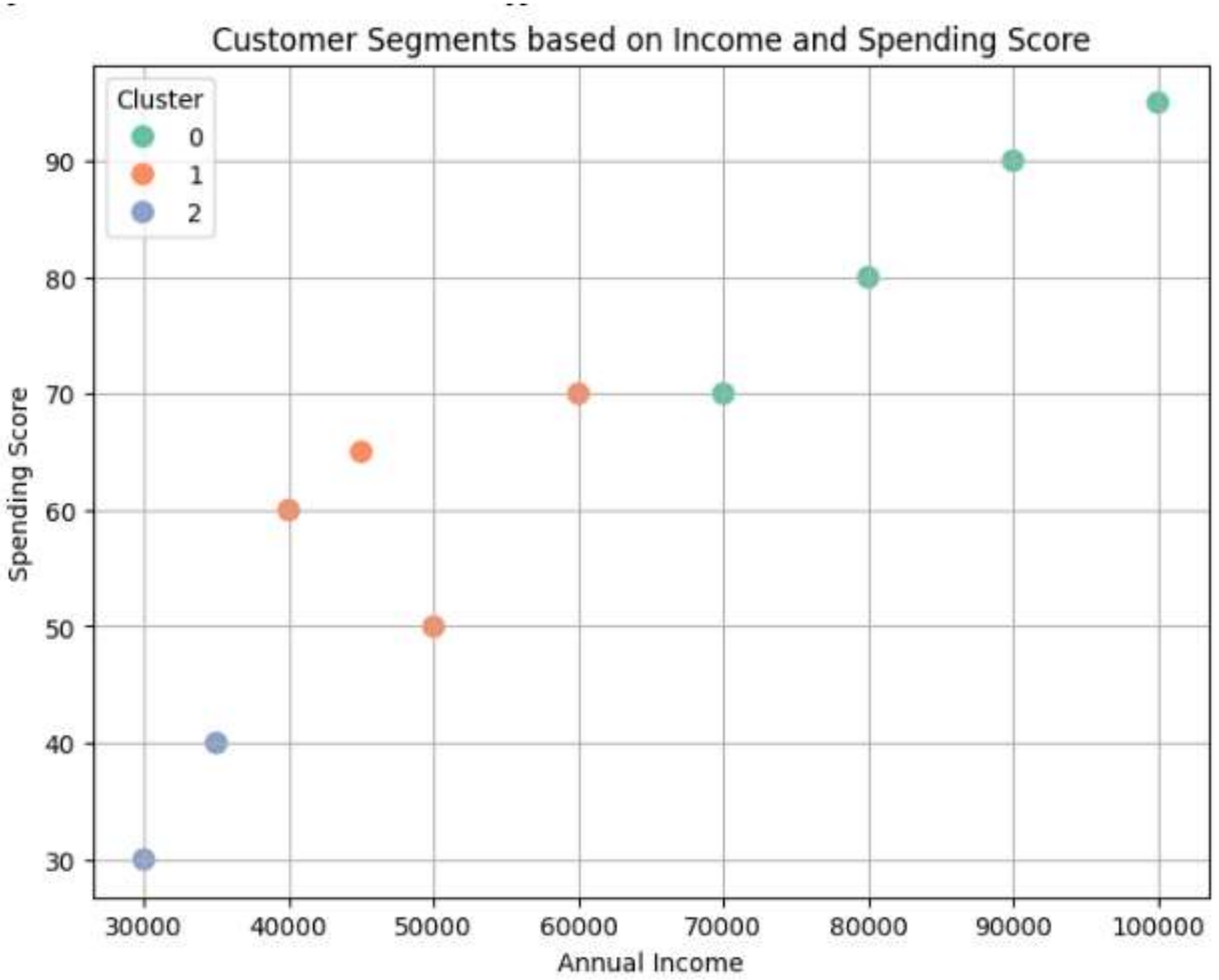
```
# Scale features
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)
# K-Means clustering
kmeans = KMeans(n_clusters=3, random_state=42)
df['Cluster'] = kmeans.fit_predict(X_scaled)
```

```
Cluster Centers (scaled):
[[ 1.08538354  1.09108945  0.94944316]
 [-0.5487441  -0.49099025 -0.18988863]
 [-1.07327889 -1.2001984  -1.51910905]]
```

```
# Show cluster centers (scaled)
print("\nCluster Centers (scaled):")
print(kmeans.cluster_centers_)
```

```
# Visualize clusters
plt.figure(figsize=(8, 6))
sns.scatterplot(data=df, x='AnnualIncome', y='SpendingScore', hue='Cluster', palette='Set2', s=100)
plt.title("Customer Segments based on Income and Spending Score")
plt.xlabel("Annual Income")
plt.ylabel("Spending Score")
plt.grid(True)
plt.show()
```

Visualization




```
Cluster Centers (scaled):
[[ 1.08538354  1.09108945  0.94944316]
 [-0.5487441  -0.49099025 -0.18988863]
 [-1.07327889 -1.2001984  -1.51910905]]
```