Day 14 Training Report

10 July 2025

Hands-on: Customer Behavior Clustering

On **Day 14**, students applied **clustering techniques to real-world customer datasets** to understand **how unsupervised learning can uncover hidden patterns**. This day emphasized **practical application**, integrating all prior concepts in clustering, evaluation, and visualization.

---

1. Objectives of the Session

- Understand customer segmentation for **targeted marketing and business strategy**.
- Apply **k-Means clustering** to real datasets to identify groups with similar behavior.
- Learn to interpret **cluster characteristics** for actionable insights.

---

2. Dataset Exploration

Students worked on a dataset containing features like:

- **Customer ID**
- **Annual Income**
- **Spending Score**
- **Age**

Steps taken:

1. **Load dataset using Pandas**: Ensured all required columns were present.
2. **Data cleaning**: Handled missing values, duplicates, and inconsistencies.
3. **Feature selection**: Chose relevant numeric features (Annual Income, Spending Score) for clustering.

---

3. Applying k-Means Clustering

- Determined the optimal number of clusters using the **Elbow Method** and **Silhouette Score**.
- Initialized k-Means with the chosen k and trained the model.
- Assigned **cluster labels** to each customer.

Example Implementation:

```
from sklearn.cluster import KMeans
import matplotlib.pyplot as plt
```

```
X = data[['Annual Income', 'Spending Score']]

# Optimal clusters (from Elbow method)
kmeans = KMeans(n_clusters=5, random_state=42)
data['Cluster'] = kmeans.fit_predict(X)

# Visualizing clusters
plt.figure(figsize=(8,5))
plt.scatter(data['Annual Income'], data['Spending Score'], c=data['Cluster'], cmap='rainbow')
plt.xlabel('Annual Income')
plt.ylabel('Spending Score')
plt.title('Customer Segmentation')
plt.show()
```

---

4. Analyzing the Clusters

Students analyzed **cluster characteristics** to understand customer behavior:

- **High-income, high-spending** → Premium segment, potential for luxury marketing.
- **Low-income, high-spending** → Young or aspirational segment, opportunity for promotions.
- **Low-income, low-spending** → Budget-conscious customers, may respond to discounts.
- **High-income, low-spending** → Conservative spenders, opportunity for loyalty programs.

They discussed how these insights can guide **marketing strategies, product recommendations, and business decisions**.

---

5. Hands-on Challenges

- Visualize clusters using **scatter plots with labels**.
- Experiment with different values of k to see how clusters change.
- Compare clusters using **silhouette scores** for quality evaluation.
- Identify **outliers** or customers that don’t fit any cluster well.