TASK-2

Sure! Here's how you can implement a **K-means clustering algorithm** using the **Mall Customer Dataset**, typically used to segment customers based on purchase behavior.

🛍️ **Goal:**

Group customers into clusters based on their features like:

* Age
* Annual Income (k\$)
* Spending Score (1–100)

✅ Step-by-Step Implementation Using Python (Scikit-learn):

---

1. **Import Required Libraries**

```python
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler
```

---

2. **Load the Dataset**

```python
df = pd.read_csv('Mall_Customers.csv')  # Replace with your actual file

# Preview
print(df.head())
```

---

3. **Select Features for Clustering**

TASK-2

For example, using **Annual Income** and **Spending Score**:

```python
X = df[['Annual Income (k$)', 'Spending Score (1-100)']]
```

---

4. **Feature Scaling (Optional but Recommended)**

```python
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)
```

---

5. **Find Optimal Number of Clusters Using Elbow Method**

```python
inertia = []
K = range(1, 11)

for k in K:
    kmeans = KMeans(n_clusters=k, random_state=42)
    kmeans.fit(X_scaled)
    inertia.append(kmeans.inertia_)

# Plot the elbow graph
plt.plot(K, inertia, 'bo-')
plt.xlabel('Number of Clusters (k)')
plt.ylabel('Inertia')
plt.title('Elbow Method to Determine Optimal k')
plt.show()
```

---

6. **Apply KMeans Clustering**

Let's say optimal `k = 5` (from elbow curve):

```python
kmeans = KMeans(n_clusters=5, random_state=42)
```

TASK-2

df['Cluster'] = kmeans.fit_predict(X_scaled)
```

---

### 7. **Visualize the Clusters**

```python
plt.figure(figsize=(8,6))
sns.scatterplot(data=df, x='Annual Income (k$)', y='Spending Score (1-100)', hue='Cluster', palette='Set2')
plt.title('Customer Segments')
plt.show()
```

🧠 Optional Enhancements:

* Add `Age` to clustering features.
* Perform 3D clustering using Plotly if needed.

📁 Notes:

* Dataset often named `Mall_Customers.csv`
* Typical structure: `CustomerID`, `Gender`, `Age`, `Annual Income (k$)`, `Spending Score (1-100)`

```