Objective:

The objective of this project is to design a **Recommendation Engine** that uses the **ClusterID** (from Milestone 2) to generate **domain-specific recommendations** for different groups of entities.
In this example, we use **student performance data** to provide personalized academic strategies, study routines, and improvement tips based on behavioral clustering.

Dataset Source:

Description:
A synthetic dataset representing student performance metrics such as:

- Average study hours per day
- Attendance percentage
- Number of assignments completed
- ClusterID (generated from previous KMeans clustering in Milestone 2)

Sample Structure:

StudentID	Avg_Study_Hours	Attendance_%	Assignments_Completed	ClusterID
101	2	60	3	2
102	5	85	9	1
103	1	45	1	2
104	6	90	10	0

Steps Followed:

1️⃣Cluster Review & Analysis:

- Reviewed the ClusterID column to understand group distributions.
- Computed average metrics (Avg_Study_Hours, Attendance_%, and Assignments_Completed) per cluster.
- Interpreted behavioral patterns:
    - **Cluster 0:** High performers
    - **Cluster 1:** Moderate performers
    - **Cluster 2:** Low performers needing improvement

2️⃣Cluster Mapping → Recommendation Generation:
Mapped each cluster to a set of actionable, personalized recommendations and tools:

ClusterID	Recommendation	Tools / Techniques
0	Maintain consistency and explore advanced topics.	Try competitive quizzes, coding platforms, mentorship programs.
1	Improve focus and time management for steady growth.	Use Pomodoro timers, set study goals, and track weekly progress.
2	Increase study time and participation	Attend extra tutorials, form study groups,

ClusterID	Recommendation	Tools / Techniques
	in class.	and seek teacher guidance.

Added two new columns:

- Recommendation
- Tools/Techniques

5. Visualization:

Created two visualizations using **Seaborn** and **Matplotlib**:

a. Cluster Distribution:
Shows the number of students per cluster.
sns.countplot(x='ClusterID', data=df, palette='Set2')

b. Recommendation Distribution:
Shows how many students fall under each recommendation type.
sns.countplot(y='Recommendation', data=df, palette='Set3')
Both plots help in understanding which recommendations are most common and which clusters dominate the dataset.

Tools Used:

Tool	Purpose
Python	Core programming language
Pandas	Data manipulation and dataset management
Matplotlib & Seaborn	Data visualization (count plots)
Google Colaboratory	Cloud-based notebook environment
CSV file	Input/output data format for Milestone 2 & 3

Key Insights:

- **Cluster 0 (High Performers):** Students exhibit excellent study habits and attendance. Recommendation: Maintain performance and seek advanced learning opportunities.
- **Cluster 1 (Moderate Performers):** Students are consistent but need better focus and time management. Recommendation: Use productivity techniques like Pomodoro.
- **Cluster 2 (Low Performers):** Students show poor attendance and low assignment completion. Recommendation: Encourage active participation and guided learning sessions.

Visualization:

1. Cluster Distribution

2. Recommendation Distribution

Conclusion

The Recommendation Engine successfully:

- Analyzed clusters from prior milestone results.
- Mapped behavioral patterns to actionable improvement plans.
- Generated structured and domain-relevant recommendations.
- Visualized the distribution of recommendations to identify group sizes and focus areas.

This approach can be extended to other domains such as **Retail (customer segmentation)** or **Healthcare (patient engagement)** by modifying the cluster mappings and recommendation rules.