

Target Mart Use Case: Clustering (Data Mining/ customer segmentation)

Objective: *Use K-Means clustering to find customer insights.*

Constraints :

1. Must use of *K-Means* Clustering .
 2. Dataset (Min. 5 features, no trivial 2D datasets).
 3. Preprocessing (Must handle missing values, normalization, and feature engineering).
 4. No Pre-Labeled Clusters (Each team must define cluster labels post-analysis).
-

Assigned Problem Statement

"Analyze customer behavior in the dataset to identify distinct clusters (customer segmentation)."

Provided Dataset :

- [Online Retail Dataset \(UCI\)](#):
-

Project Phases

Phase 1: Data PreProcessing

Phase 2: K-Means Clustering

Phase 3: Cluster Interpretation

- Deliverables:
 - i. Assign meaningful labels to clusters (e.g., "High-Value Repeat Buyers").
 - ii. Provide actionable business insights (e.g., "Cluster 3 responds to discounts").

Phase 4: Deliver a report (document)

- Report Must Include:
 - i. Why K-Means was a poor/good choice
 - ii. How dataset flaws impacted results.
 - iii. What is the choice of k based on?
 - iv. Give a label to each cluster? And why?
 - v. Charts (**Bonus**)

Bonus: use another algorithm for clustering, add backend and frontend to offer Machine Learning(AI As a Service)

Grading Rubric (100 pts)

Category	Points	Criteria
Data Preprocessing	25	Handling missing data, feature engineering, normalization.
K-Means Implementation	30	Correct k selection, distance metric justification, clustering validity.
Cluster Analysis	25	Meaningful labels, business insights, visualization quality.
Critical Thinking	20	Limitations, algorithm tradeoffs, reproducibility.

Deliverables

1. Code (.ipynb or .py):
 - Clean, commented, with explanations for key steps.
 2. Report (PDF):
 - pages: methods, results, business recommendations, limitations.
-