



## IIT KHARAGPUR AI4ICPS I HUB FOUNDATION

Hands-on Approach to AI, Cohort-4, 2025

### Programming Assignment 4: Unsupervised Learning in Python

**Due date:** Sunday 14<sup>th</sup> December 2025, EOD – IST

#### Important Instructions about Programming Assignments

1. Programming assignments will be evaluated automatically. **Do not** change the skeleton code provided to you.
2. Write your code **only in the designated places** in the skeleton code and process the input data provided to you in the designated variables. **Do not alter** the input-output structure in the skeleton code.
3. **Do not import** any additional libraries. **Do not use any additional files** for the processing (*other than those mentioned in the skeleton code*).
4. Failure to comply with these instructions may lead to you getting **zero marks** for the assignment, even if the solution is largely correct.
5. The **evaluation is binary**, i.e., either you get full marks if your output exactly matches the evaluation test cases (*different from Sample Test Cases provided in the question*) otherwise you get a zero. There is **no provision for step marking**.

**Objective:** This assignment has two objectives:

1. Understanding how to implement the K-Means clustering algorithm for a real dataset.
2. Develop a pipeline to solve the problem and provide accurate output using basic programming logic.

**Problem:** Write a Python program using the provided template to find the best K in the KMeans algorithm using the Silhouette Coefficient metric for the 'housing.csv' dataset. The program should accept two integers that define the range of K values and output only the best K and nothing else.

**Instructions:** Download the template program and write your code in the designated location mentioned in the comments. Initialize the K-Means model: You must use `n_init='auto'` and `random_state=0` as KMeans model parameters. Use the given dataset (housing.csv) and calculate the Silhouette Coefficient for the range of K provided. Output the best K with respect to the Silhouette Coefficient. Any additional text other than the actual output will be treated invalid. Full marks will be awarded only if the output is an exact match of the evaluation test cases.

You should read the skeleton code to understand the exact mechanism of input from the command line and provide the required output.

**Execution Syntax:** `python assignment.py <number> <number>`

**Sample Test Cases:**

Input	2 4	3 7	13 17	21 25	7 10
Output	2	3	14	24	8