## Module 3: Clustering

Attempt 1	<u> </u>	Review Feedback
		9/22/2025

Attempt 1 Score: 40/40



Anonymous Grading: no

#### **Unlimited Attempts Allowed**

#### ∨ Details

#### Goal of the assignment:

In this assignment, you will analyze the clustering results of the housing dataset we looked at in class. Your goal is to **generate and plot distributions of key features in the different clusters and interpret the characteristics of each cluster,** including the <u>ocean\_proximity</u> feature that can be found in the original housing dataset.

#### Steps:

- 1. Create a new Colab notebook. Name it **your-name-clustering** (replacing your-name with *your actual name*, of course!!) Using the code in the <code>@3-Clustering.ipynb</code> notebook, perform k-Means clustering (with k=4) on the <code>housing\_minus\_coords</code> dataset, which is the z-score standardized housing dataset without the geographical coordinates. You can simply reuse the code in the notebook for this.
- 2. Now that you have the cluster labels (stored in the kmeans.labels\_ object), load the original housing\_data as done in the first cell of the @2-Toolbox-Part3.ipynb notebook.
- 3. Next, paste the clustering labels into the original housing\_data dataframe (hint: I'm doing something similar when adding the cluster labels to the long\_lat\_cl object in 03-clustering.ipynb ...)
- 4. For each of the four clusters, create boxplots or histograms or kernel density plots (whatever you prefer) for the following features: housing\_median\_age, total\_rooms, population, households, median\_income, median\_house\_value, and a barplot (since it's a discrete and not a numeric feature) for the ocean\_proximity feature.
- 5. By comparing the distribution of each feature in each cluster, see if you can find some way to interpret the clusters. For example, how do the features differ between clusters? What can you infer about each cluster based on the ocean\_proximity and other features? (e.g., are certain clusters closer to the coast? Are they associated with higher or lower incomes? Do the clusters contain vacation properties?)
- 6. Write your interpretation of the clusters directly in the notebook as a text cell.

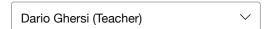
#### **Deliverables:**

A Colab notebook containing all the code and plots, and a cell with your interpretation of the clusters. Please upload that to Canvas. That's all you need to upload.

#### Use of LLMs:

You are allowed (and even encouraged) to ask Gemini (available directly in Colab) or other LLMs for help if you don't know how to do something. The only thing that is not directly provided in the notebooks I shared is how to plot distributions for each cluster, so that might be something you need to do a little bit of research on.

# ∨ View Rubric Select Grader





### Rubric for clustering assignment

Criteria	Ratings					Pts
Data Loading, Preprocessing, and K- Means Clustering		10 pts Good Minor issues with loading the data or clustering, but the approach is correct overall.	applies K-Means but with	4 pts Needs improvement Dataset is not fully loaded or preprocessed. K- Means is applied incorrectly or not as required.	•	12 / 15 pts
Feature Distributions for Clusters	are generated and correctly visualize feature	7 pts Good Plots are mostly correct, but there are minor labeling or clarity issues.	missing or unclear for some	Incomplete or	0 pts Incomplete No plots provided.	10 / 10 pts
Interpretation of Clusters	o .	7 pts Good Good interpretation with mostly clear connections between features and clusters.	insights, but	3 pts Needs improvement Limited or superficial interpretation, with unclear or incorrect conclusions about the clusters.	0 pts Incomplete No interpretation provided.	10 / 10 pts
Code Quality and Presentation	5 pts Excellent Code is clean, well-commented, and easy to follow. Notebook is well-organized.		2 pts Satisfactory Code is functional but messy or lacks adequate comments. Notebook organization is lacking.	1 pts Needs improvement Code has significant issues lacks comments, and the notebook is difficult to follow.	notebook is hard	5 / 5 pts

Total Points: 37

