# Educational Escape Room Design for Machine Learning: Week 3 - Unsupervised Learning

**Title:** "The Lost Archive: Decoding the Unknown"

**Target Week/Topic:** Week 3 – Unsupervised Learning (k-Means Clustering & Dimensionality Reduction)

**Narrative/Theme:**  
The players are recruited as digital archaeologists attempting to restore a corrupted archive of alien knowledge discovered in a deep-space probe. The archive has no labels and contains unstructured information from different alien species. They must reverse-engineer and group the data using unsupervised learning methods to unlock the meaning and retrieve the message before the system resets in 60 minutes.

**Learning Objectives:**

* Understand how k-means clustering groups similar data points.
* Recognise the impact of dimensionality reduction on data interpretability.
* Apply visualisation to interpret clusters.
* Experience the process of choosing number of clusters (k).

**Session Time:** 75 minutes total

* Escape room gameplay: 60 minutes
* Debrief and reflection: 15 minutes

**Mode:** Collaborative, small groups of 3-4 students

**Structure:** Non-linear (three parallel clusters of puzzles leading to final meta-puzzle)

**Tools/Platform:**

* Google Forms or Unity WebGL (optional)
* Jupyter notebooks or Colab environment pre-set with synthetic data
* Puzzle cards or PDFs if offline

**Puzzle Structure**

**Entry Puzzle (5 mins)**

* Students decode a corrupted alien message using frequency analysis (introduces idea of pattern discovery without labels).
* Unlocks access to the alien archive dataset.

**Three Puzzle Streams (Non-linear):**  
Players can tackle these streams in any order. Each is themed around one alien species' unlabelled data.

**1. The Crystal Codex (Clustering Puzzle)**

* Dataset: Alien mineral samples with unknown categories.
* Task: Use k-means in a notebook to cluster the samples.
* Puzzle: Match the cluster centroids to specific crystal types based on visual cues.
* Output: A numerical code from the cluster labels.

**2. The Forgotten Glyphs (Dimensionality Reduction Puzzle)**

* Dataset: High-dimensional glyph features.
* Task: Apply PCA to visualise in 2D.
* Puzzle: Identify patterns that match a glyph-based alien alphabet.
* Output: A symbol-password for the central console.

**3. The Fragmented Signal (Interpreting Clusters)**

* Dataset: Alien communication logs.
* Task: Students match clustered conversations to hypothetical topics.
* Puzzle: Use silhouette scores to decide optimal k.
* Output: Colour code sequence.

**Final Meta Puzzle (10-15 mins)**

* Combine outputs from the three puzzle paths to reconstruct and decrypt the final message.
* Puzzle: A logic grid combining numeric, symbolic, and colour codes. Successful reconstruction plays an audio clip or animation of the restored message.

**Assessment & Debrief**

* Google Form quiz or reflection sheet:
  + What did each cluster represent?
  + How did PCA help interpret the data?
  + What challenges arose from unlabelled data?
* Group discussion on real-world unsupervised learning challenges.

**Optional Extension**

* Students modify k or test different dimensionality reduction methods (e.g., t-SNE) after the game to see how results change.

**Notes:**

* Include hints and scaffolding for weaker groups.
* Escape room can be adapted to linear mode by gating each puzzle in sequence if desired.

# Solution

Here is the expected solution for the **Week 3: Unsupervised Learning Escape Room**, designed around **k-means clustering** and **dimensionality reduction (PCA)**.

**🔓 Puzzle 1: The Data Vault (Exploratory Data Analysis)**

**Task:**  
Players are given a CSV file with unlabeled data (e.g., synthetic customer or anomaly data). They must:

* Identify the number of features
* Plot pairwise scatter plots
* Observe natural groupings or anomalies

**Expected Solution:**

* Use pandas and matplotlib/seaborn to load and visualise
* Spot roughly 3 natural groupings in 2D scatter plots

import pandas as pd

import seaborn as sns

data = pd.read\_csv("mystery\_data.csv")

sns.pairplot(data)

**Clue revealed:** “Three paths lie hidden, but only the right lens will let you see.”

**🔓 Puzzle 2: The Clustering Chamber**

**Task:**  
Apply k-means clustering with a guessed k=3, and find the centroids.

**Expected Solution:**

from sklearn.cluster import KMeans

kmeans = KMeans(n\_clusters=3)

kmeans.fit(data)

labels = kmeans.labels\_

centroids = kmeans.cluster\_centers\_

**Clue revealed:** Coordinates of the cluster centres are part of the room unlock code.

**🔓 Puzzle 3: The Projection Mirror (Dimensionality Reduction)**

**Task:**  
Use PCA to reduce to 2D, visualise the cluster assignments and identify which cluster is which (e.g., use colour to match original data to PCA clusters).

**Expected Solution:**

from sklearn.decomposition import PCA

import matplotlib.pyplot as plt

pca = PCA(n\_components=2)

reduced = pca.fit\_transform(data)

plt.scatter(reduced[:,0], reduced[:,1], c=labels)

plt.title("PCA Projection")

plt.show()

**Clue revealed:** A geometric symbol appears (e.g., triangle pointing left) representing a code for the final lock.

**🔓 Final Puzzle: Decoding the AI Guardian**

**Task:**  
Solve a riddle combining all previous elements (e.g., number of clusters + centroid coordinates + PCA explained variance). Requires entering a code into a simulated input field in the escape room interface.

**Expected Solution:**

* Combine cluster count 3
* Add centroid x-coordinates (rounded): e.g., 2 + 5 + 4 = 11
* Use PCA explained variance: e.g., 0.67

Final code = 3-11-67

**Technical Skills Involved:**

* Reading and inspecting datasets
* Applying unsupervised learning (k-means)
* Using PCA for dimensionality reduction
* Data visualisation
* Combining quantitative outputs for reasoning