

**Lahore College for Women University, Lahore**

**Department: Software Engineering**

**Submitted By: Iram Ahmad**

**Roll no: 2225165111**

**Submitted to: Sir Mohsin Nazir**

**Semester: VII**

**Session: 2022-2026**

**Assignment no 8**

**Course: Applied Data Science with AI**

**Semester:** BSSE 7th  
**Week #:** 8  
**Student Name:** Iram Ahmad  
**Roll Number:** 2225165111  
**Project Title:** House Price Prediction  
**GitHub Link:** https://github.com/Iram-Ahmad/Data-Science-AI-Course

**1. Reading Summary (½–1 page)**

**Reading Material for this Week:**

* *Intro to Machine Learning with Python – GitHub*
* *Scikit-Learn Clustering Documentation*

**Key Learnings:**

* **K-Means Clustering:** Learned how to group similar data points based on feature similarity using K-Means, a popular unsupervised learning algorithm.
* **PCA (Principal Component Analysis):** Understood how PCA helps reduce the number of features while retaining most of the information, making visualization easier.
* **Clustering Evaluation:** Learned that clustering doesn’t use labeled data; instead, performance is evaluated using metrics like inertia or silhouette score.

**Reflection:**  
These readings helped me understand how to explore patterns and similarities in data even without target labels. In my **House Price Prediction** project, clustering can help segment houses into groups (e.g., low, medium, and high-priced) and identify patterns in features like area, quality, and location. This insight supports better feature engineering for future predictive models.

**2. Classroom Task Documentation**

**Task Performed in Class:**

* Implemented **K-Means Clustering** and **PCA** on a dataset.
* Visualized clusters in 2D using PCA for better understanding of group separations.

**Screenshots / Code Snippets:**  
*(Plots and PCA visualizations are included in the Jupyter Notebook — K-Means cluster scatter plot using PCA.)*

**3. Weekly Assignment Submission**

**Assignment Title:** *Unsupervised Learning – K-Means and PCA Visualization*

**Steps Taken:**

1. Loaded the cleaned house prices dataset.
2. Selected numerical features and applied **StandardScaler** for normalization.
3. Performed **K-Means Clustering** with different k values to find optimal clusters.
4. Reduced dimensions using **PCA (2D)** for visualization.
5. Visualized results with a scatter plot showing clusters in 2D PCA space.

**Output:**

* Created 3 main clusters representing different house feature groups.
* Visualized PCA-based 2D scatter plot where clusters were clearly separated.

**Challenges Faced:**

* A warning appeared regarding missing font glyphs (“HOUSE BUILDING” symbol), which did not affect the result.
* Deciding on the optimal number of clusters (k) required testing multiple values.

**GitHub Link:**  
[Data-Science-AI-Course/notebooks at main · Iram-Ahmad/Data-Science-AI-Course](https://github.com/Iram-Ahmad/Data-Science-AI-Course/tree/main/notebooks)

**4. Project Progress Milestone**

**This Week’s Milestone:**

* Added **Unsupervised Learning (K-Means Clustering + PCA)** analysis to the House Price Prediction project.

**Next Week’s Goal:**

* Begin working on **Feature Engineering and Model Improvement** using clustering insights.

**5. Self-Evaluation (Check one)**

☑ **I completed all tasks on time.**  
☐ I partially completed the tasks.  
☐ I struggled with this week’s tasks and need help.