

**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 1**

**Course: Applied Data Science with AI**

**Semester:** BSSE 7th  
**Week #:** 1  
**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:**

* *Data Science from Scratch* – Joel Grus (Chapter 1: Introduction)
* *Artificial Intelligence: A Modern Approach* – Introduction
* Official documentation for **Anaconda** and **Jupyter Notebook** (setup and basic usage)

**Key Learnings:**

1. **Understanding the Data Science Process:**  
   I learned that data science follows a workflow — from data collection and cleaning to visualization, modeling, and deployment. Each stage is connected; errors in early stages can affect model accuracy later.
2. **Difference Between Data Science and AI:**  
   Data Science focuses on analyzing data and finding insights, while AI focuses on creating systems that can learn and make decisions. Data science prepares and analyzes the data that AI systems often rely on.
3. **Setting Up the Right Environment:**  
   Using Anaconda helps manage libraries easily and ensures compatibility. Jupyter Notebook provides an interactive environment to write, run, and explain code with text and visualizations — ideal for projects and assignments.

**Reflection:**  
These readings helped me understand **why the environment setup is the foundation** of data science projects. For my **House Price Prediction** project, I’ll collect housing data, clean it, and later build predictive models. Setting up tools like Anaconda and GitHub in Week 1 ensures that my workflow remains organized and reproducible. The AI readings also made it clear that my project is a *supervised learning regression problem* — predicting continuous values (house prices) using past data.

**2. 🏫 Classroom Task Documentation**

**Task Performed in Class:**

* Installed **Anaconda Distribution** (which includes Jupyter Notebook, Python, and important data libraries).
* Launched **Jupyter Notebook** and explored how to create a new notebook file.
* Created a **GitHub account** and **repository** named *Data-Science-AI-Course* to store weekly assignments.

**Code Snippets:**

# Checking Python version

!python --version

# Basic test code in Jupyter Notebook

import pandas as pd

import numpy as np

print("Setup successful! Pandas and NumPy imported.")

**Output:**

Setup successful! Pandas and NumPy imported.

This confirmed that the environment was working correctly.

**3. 📂 Weekly Assignment Submission**

**Assignment Title:** *Introduction to Python and Data Loading*

**Steps Taken:**

1. Downloaded a **sample dataset** (Titanic dataset) from **Kaggle**.
2. Opened **Jupyter Notebook** in Anaconda and imported **Pandas** to load the dataset.
3. Displayed the **first 10 rows** using head() function to ensure data loaded correctly.
4. Saved the notebook as Week1\_DS\_AI\_Week1\_Notebook.ipynb.
5. Uploaded the notebook to my GitHub repository.

**Sample Code:**

import pandas as pd

# Load dataset

df = pd.read\_csv("titanic.csv")

# Show first 10 rows

df.head(10)

**Output:**  
A table displaying the first 10 rows of the Titanic dataset, showing columns like *PassengerId*, *Name*, *Sex*, *Age*, *Survived*, etc.

**Challenges Faced:**

* Initially, I was unsure how to open .ipynb files in Jupyter instead of VS Code.
* The “Select Kernel” option appeared, and I learned that selecting **Python 3 (ipykernel)** is necessary to run code.
* After setup, everything worked smoothly.

**GitHub Link:**  
<https://github.com/Iram-Ahmad/Data-Science-AI-Course>

**4. 🚀 Project Progress Milestone**

**This Week’s Milestone:**  
✅ Installed Anaconda and launched Jupyter Notebook.  
✅ Created GitHub repository for the course.  
✅ Downloaded sample dataset and loaded it in Jupyter.  
✅ Selected my final project topic — *House Price Prediction*.

**Next Week’s Goal:**  
➡ Clean my project dataset (handle missing values, duplicates, and outliers).  
➡ Upload “before vs after cleaning” dataset to GitHub.

**5. ✅ Self-Evaluation**

☑ I completed all tasks on time.  
I successfully installed all tools, ran my first notebook, and uploaded my work to GitHub. Now I understand how to use Jupyter for further assignments.

**6. 💬 Questions for Instructor**

* How can we link our Jupyter Notebook directly with GitHub (for auto-synchronization)?
* Should we use the same dataset for practice (Titanic) and for our final project (House Prices), or switch after Week 1?