🏠 AI Lab Assignment 1– Housing Data (`housing.csv`)

📓 About the Dataset

The **California Housing dataset** contains information about housing in different block groups across California. Each row represents a block, and the columns describe its geographical, demographic, and economic characteristics.

Column Descriptions

1. **longitude** – How far west a house is located (higher = farther west).
2. **latitude** – How far north a house is located (higher = farther north).
3. **housingMedianAge** – Median age of houses in the block (lower = newer).
4. **totalRooms** – Total number of rooms within the block.
5. **totalBedrooms** – Total number of bedrooms within the block.
6. **population** – Total number of people living in the block.
7. **households** – Total number of households (groups of people living together).
8. **medianIncome** – Median income for households (in tens of thousands of USD).
9. **medianHouseValue** – Median house value (in USD).
10. **oceanProximity** – Distance from the ocean (e.g., “INLAND”, “NEAR OCEAN”, etc.).

🎯 Objective

Use **pandas** to explore and analyze the housing dataset.
Perform data loading, inspection, filtering, and basic analysis operations.

🧩 Instructions

You may complete this assignment using **Google Colab**, **Jupyter Notebook**, or any Python IDE.
Create a notebook named **`Housing_fnamelname_id.ipynb`** and follow the tasks below.

☑ Tasks

1️⃣ *Load the Dataset*

- Upload the file **`housing.csv`**.
- Import the necessary libraries (`pandas`, `io`, `files` if using Colab).
- Read the dataset into a DataFrame called `df`.
- Display the first **5 rows**.

2️⃣ *Basic Information*

- Show the **number of rows and columns**.
- Print all **column names**.
- Display data types and non-null counts (`df.info()`).
- Show the **number of missing values** in each column.

3️⃣ *Display Rows*

- Display **row number 5**.
- Display **rows 10 to 14**.

4️⃣ *Data Analysis*

- Find the **maximum** and **minimum** values in **`medianHouseValue`**.
- Calculate the **mean**, **median**, and **standard deviation** of **`medianHouseValue`**.

5️⃣ *Filter the Data*

- Show rows where **`medianHouseValue` > 300000**.
- Show rows where **`oceanProximity` == 'NEAR OCEAN'**.

6️⃣ *Modify Columns*

- Rename **`medianHouseValue`** → **`HouseValueUSD`**.
- Add a new column **`RowID`** (from 1 to the number of rows).
- Display the first **5 rows** again to confirm changes.

7️⃣ *Sort the Data*

- Sort by **`housingMedianAge`** in **descending order**.
- Display the first **10 rows** after sorting.

8️⃣ *Summary Statistics*

- Display dataset statistics with **`df.describe()`**.
- Calculate the **sum** and **average** of **`HouseValueUSD`**.

9️⃣ *Save and Submit*

- Save your notebook as **PDF** or **.ipynb**.