

# Introduction:

The dataset for this project is sourced from a real estate database, comprising information on 100 randomly selected residential properties in a specific County registry. Each property is identified by a unique real estate ID number, and 10 variables are recorded for each property, covering diverse aspects such as year built, square footage, adjusted land value, zip code, and more.

The project aims to provide a comprehensive understanding of the dataset through a series of analytical tasks. The goal is to uncover patterns, relationships, and insights within the real estate data.

**Data set:** [data.csv](#)

# Report Guidelines:

- Each student is required to submit their paper individually.
- Submit a well-documented report that can be easily understood.
- Provide clear explanations for each step.
- Include relevant visualizations to support your analysis.
- Include all R code used in the analysis in the appendix and organize the code by sections corresponding to the report.

## Format:

- **File type:** pdf, doc
- **Font:** Please use a readable font such as Times New Roman or Arial.
- **Font Size:** Maintain a consistent font size of 12 points for the main text.
- **Line Spacing:** Keep the line spacing consistent at 1~1.5 for the main text.
- **Page Numbers:** Limit the main report (not including R codes) to 4-10 pages.

# Project Overview and

# Expectations:

The project aims to provide a comprehensive understanding of the dataset through a series of analytical tasks. The goal is to uncover patterns, relationships, and insights within the real estate data.

## Variable Descriptions:

- **ID:** The county-given identification number for the property.
- **YearBuilt:** The listed year in which the structure was built (by year).
- **SqFt.:** The area of the floor plan in square feet (in square feet).
- **Story:** How many stories the structure has.
- **Acres:** How many acres are included in the property.
- **N\_Baths:** The number of bathrooms at the residence.
- **Fireplace:** Whether the residence has a fireplace.
- **TotalPrice:** The total assessed value of the property (in dollars).
- **LandPrice:** The assessed value of the land (in dollars).
- **BuildingPrice:** The assessed value of the building (in dollars).
- **Zipcode:** The zip code of the property.

## Tasks:

1. **Detect and Handle Missing Data:**
  - Detect any missing values within the dataset and provide the IDs of observations with missing data.
  - Remove observations with missing data before proceeding with the analysis.
  - Ensure the dataset is clean and ready for analysis.
2. **Variable Summarization:**
  - **Select at least two variables of interest.**
  - Summarize each variable with relevant statistics and graphical tools.
  - Identify any unusual data.
  - Provide insightful descriptions of variable distributions.
3. **Price Comparison: Does the presence of a fireplace make a difference?**
  - Investigate whether the presence of a fireplace is related to property price.
  - **You may use any of the three prices (TotalPrice, LandPrice, or BuildingPrice).**
  - Utilize numerical and graphical tools to compare property prices.
  - Provide insights regarding any observed differences in property prices.
4. **Numerical Relationship Exploration:**

- Identify continuous variables.
  - Explore potential relationships among different continuous variables.
  - Identify at least one pair of continuous variables that you speculate may exhibit a discernible relationship. Briefly explain the rationale behind your speculation.
  - Utilize appropriate graphical tools to visualize and understand the relationship between the variables of your choice.
  - Detect and interpret any unusual patterns.
5. **Linear Regression Analysis:**
- Conduct a linear regression analysis on two selected variables utilizing insights from the previous step.
  - Present the regression model and interpret estimated coefficients.
  - Evaluate the goodness of fit using relevant metrics.

## Additional Tips:

- **Clarity:** Clearly label each section and subsection.
- **Visuals:** Use clear and appropriately labeled visuals (tables, charts, graphs).
- **Consistency:** Maintain consistency in formatting throughout the report.
- **Proofreading:** Proofread the report for grammar and spelling.