

## 0.1 Question 0: Human Context and Ethics

In this part of the project, we will explore the human context of our housing dataset. **You should watch Lecture 14 before attempting this part.**

---

### 0.1.1 Question 0a

“How much is a house worth?” Who might be interested in an answer to this question? **Please list at least three different parties (people or organizations) and state whether each one has an interest in seeing the housing price to be high or low.**

*1. Homebuyers: They want the housing price to be low, so they can purchase with a lower price. 2. Homeowners: They want the housing price to be high, so they can have a higher selling price. 3. Real-estate agent: They also want the housing price to be high, so they can earn more commission.*



---

### 0.1.2 Question 0b

Which of the following scenarios strike you as unfair and why? You can choose more than one. There is no single right answer, but you must explain your reasoning. Would you consider some of these scenarios more (or less) fair than others? Why?

- A. A homeowner whose home is assessed at a higher price than it would sell for.
- B. A homeowner whose home is assessed at a lower price than it would sell for.
- C. An assessment process that systematically overvalues inexpensive properties and undervalues expensive properties.
- D. An assessment process that systematically undervalues inexpensive properties and overvalues expensive properties.

*Answer: A, this is unfair for the homeowner since they may pay higher property taxes. B, this is unfair for the community and local government since the homeowner may pay less property taxes. C, this is unfair for inexpensive properties and benefits the expensive properties. D, this is unfair for the expensive properties and benefits the inexpensive properties. I think C and D are more unfair than the other scenarios because they represent systematic unfairness of a group of homeowners rather than individual cases above.*



---

### 0.1.3 Question 0d

What were the central problems with the earlier property tax system in Cook County as reported by the Chicago Tribune ? And what were the primary causes of these problems? (Note: in addition to reading the paragraph above you will need to watch the lecture to answer this question)

*The central problems were the CCAO's model was producing a **regressive** tax system which disproportionately burdened Black and Latinx homeowners in Cook County. The primary causes of these problems are: 1, The outdated methods for property valuation, lack of updates to reflect the latest property market. 2, The systemic racial biases. 3, The system lacked of transparency and overly complex, people who can afford to hire tax lawyers to appeal for the valuation of their houses will eventually not be overvalued.*



---

#### 0.1.4 Question 0e

In addition to being regressive, how did the property tax system in Cook County place a disproportionate tax burden on non-white property owners?

*Because of the Jim Crow laws last century, the housing segregation caused people with color live in far away predominately white homeowners neighborhood. The neighborhoods with predominately non-white homeowners often had their properties to be overvalued, which means they need to pay higher taxes compared to predominately white neighborhoods.*





---

## 0.2 Question 3a

One way of understanding the performance (and appropriateness) of a model is through a plot of the residuals versus the observations.

In the cell below, use `plt.scatter` to plot the residuals from predicting Log Sale Price using **only the 2nd model** against the original Log Sale Price for the **validation data**. With a data size this large, it is difficult to avoid overplotting entirely. You should also ensure that the dot size and opacity in the scatter plot are set appropriately to reduce the impact of overplotting as much as possible.

```
In [22]: plt.scatter(y_valid_m2, y_valid_m2-y_predicted_m2, marker = '.', alpha = 0.3)
plt.title('Residual of prediction for 2nd model Log Sale Price')
plt.xlabel('Original validation Log Sale Price')
plt.ylabel('Residual  $(y - \hat{y})$ ')
plt.axhline(y=0, color='r');
```



