

## 0.1 Question 1: Human Context and Ethics

In this part of the project, we will explore the human context of our housing dataset. **You should watch Lecture 15 before attempting this part.**

---

### 0.1.1 Question 1a

“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. Household owners might be interested in the housing price because they need to know how much does their households worth as compared to others. They tend to hope to see the housing price to be high.
2. Renters/buyers of households might be interested in the housing price because they need to know how much they need to pay and compare the prices. They tend to hope to see the housing price to be low.
3. Government might also be interested in the housing price because they want to maintain a stable housing market that benefits both renters/buyers and household owners. They sometimes want to see the housing price to be higher to promote economic growth.



---

### 0.1.2 Question 1b

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.

All four options exhibit unfairness from various perspectives in terms of taxation.

Option A can be seen as unfair to homeowners who may be over-taxed, feeling unjustly burdened by the government with higher taxes than others.

Option B raises issues of unfairness as well, as homeowners may be under-taxed, which not only doesn't trouble the homeowners but also places an unfair burden on others who have to pay extra taxes to compensate for this shortfall.

Option C introduces unfairness for those who own less expensive properties, as they are subjected to a regressive tax system, which creates further imbalances when compared to homeowners with more expensive properties who end up paying less in taxes.

Option D is also unfair as the process primarily affects homeowners with expensive properties who must pay a large amount of tax due to a progressive tax system. While this approach is often used to bridge the wealth gap in many countries, it may still be viewed as unfair by those who must bear the larger tax burden.



---

### 0.1.3 Question 1d

What were the central problems with the earlier property tax system in Cook County as reported by the Chicago Tribune? 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 CCAO has established a regressive tax system that disproportionately affects both rich and poor individuals, ultimately exacerbating income disparities. This system levies lower taxes on affluent property owners, particularly those with expensive properties, while burdening individuals with less expensive properties, many of whom are non-white, with higher tax obligations. This disparity primarily arises from the fact that wealthy homeowners have the means to hire tax lawyers who are well-versed in navigating the system to challenge their property assessments, ultimately reducing the assessed value of their homes. Conversely, lower-income individuals are often unable to access these resources, leaving them unable to pursue such opportunities for tax reduction.



---

#### 0.1.4 Question 1e

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

The property tax system employed redlining practices, rendering it difficult or impossible to secure federally-backed mortgages for home purchases in neighborhoods designated as “risky.” These practices centered around the development of seemingly objective rating systems, often involving information technologies, to assess property values. Notably, these systems incorporated race as a determinant in the valuation process.





---

## 0.2 Question 4a

One way of understanding a model's performance (and appropriateness) 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 such a large dataset, 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 [24]: plt.scatter(x = Y_valid_m2, y = Y_valid_m2 - Y_predicted_m2, s = 3, alpha = 0.30)
          plt.xlabel('Log Sale Price: $y$')
          plt.ylabel('Residuals : $y-\hat{y}$')
          plt.title('Residuals To Log Sale Price of 2nd Model for validation data')
```

```
Out[24]: Text(0.5, 1.0, 'Residuals To Log Sale Price of 2nd Model for validation data')
```



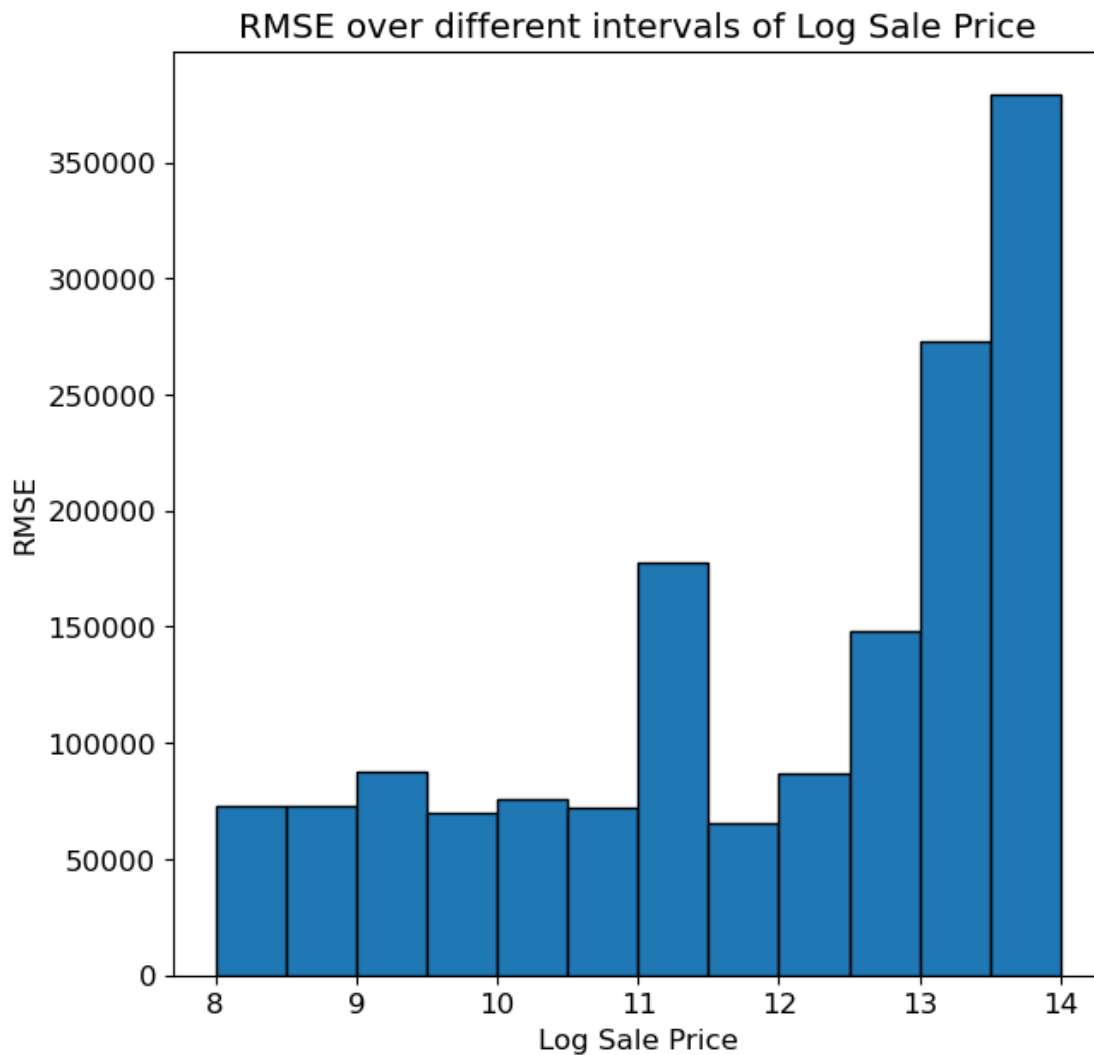


---

### 0.2.1 Question 6c

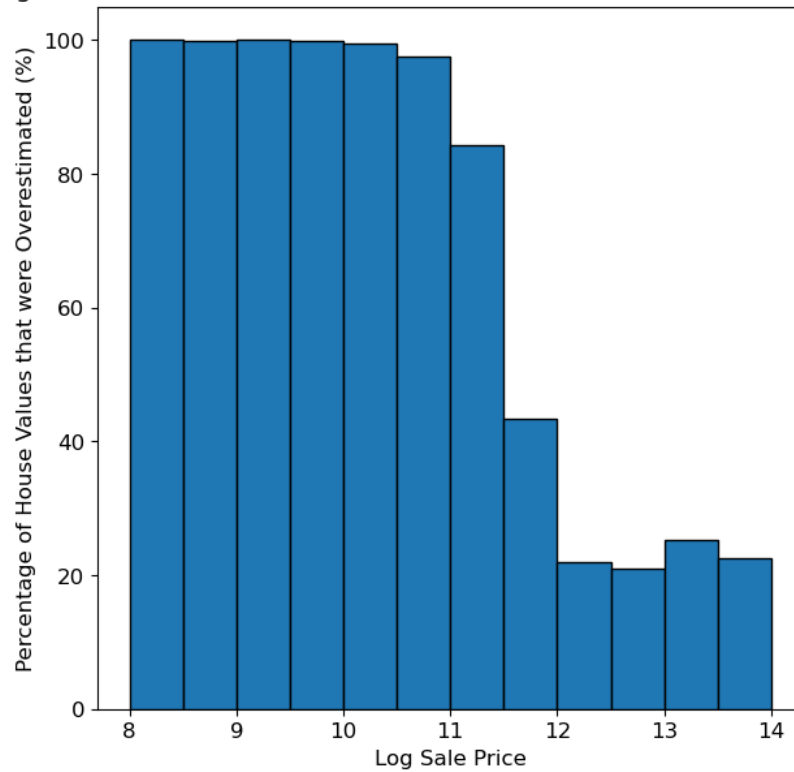
Now that you've defined these functions, let's put them to use and generate some interesting visualizations of how the RMSE and proportion of overestimated houses vary for different intervals.

```
In [43]: # Run the cell below to generate the plot; no further action is needed
rmsees = []
for i in np.arange(8, 14, 0.5):
    rmsees.append(rmse_interval(X, Y, i, i + 0.5))
plt.figure(figsize = (7, 7))
plt.bar(x = np.arange(8.25, 14.25, 0.5), height = rmsees, edgecolor = 'black', width = 0.5)
plt.title('RMSE over different intervals of Log Sale Price')
plt.xlabel('Log Sale Price')
plt.ylabel('RMSE');
```



```
In [44]: # Run the cell below to generate the plot; no further action is needed
props = []
for i in np.arange(8, 14, 0.5):
    props.append(prop_overest_interval(X, Y, i, i + 0.5) * 100)
plt.figure(figsize = (7, 7))
plt.bar(x = np.arange(8.25, 14.25, 0.5), height = props, edgecolor = 'black', width = 0.5)
plt.title('Percentage of House Values Overestimated over different intervals of Log Sale Price')
plt.xlabel('Log Sale Price')
plt.ylabel('Percentage of House Values that were Overestimated (%)');
```

Percentage of House Values Overestimated over different intervals of Log Sale Price



Explicitly referencing **any ONE** of the plots above (using `props` and `rmse`), explain whether the assessments your model predicts more closely align with scenario C or scenario D that we discussed back in 1b. Which of the two plots would be more useful in ascertaining whether the assessments tended to result in progressive or regressive taxation? Provide a brief explanation to support your choice of plot. For your reference, the scenarios are also shown below:

- C. An assessment process that systematically overvalues inexpensive properties and undervalues expensive
- D. An assessment process that systematically undervalues inexpensive properties and overvalues expensive

The assessment that my model predicts more closely align with senario C. According to the second plot that shows the percentage of house values overestimated over different intervals of log sale prices, we observe a decrease in the percentage of overestimated values as log sale price increases. This means that, for less expensive properties, the model tends to overestimate their values, which aligns with senario C. The plot of percentage of house values overestimated will be more useful in determining the type of taxation because it directly shows how the model's assessments compare to the true values for different intervals of Log Sale Price. Since the percentage of overestimated values decreases as Log Sale Price rises, we can infer that the assessments tend to result in regressive taxation.



### 0.3 Question 7: Evaluating the Model in Context

---

#### 0.4 Question 7a

When evaluating your model, we used RMSE. In the context of estimating the value of houses, what does the residual mean for an individual homeowner? How does it affect them in terms of property taxes? Discuss the cases where residual is positive and negative separately.

We utilize our model to assess a property's value. The residual, in this context, represents the difference between the property's market value and the assessment provided by CAAO (Cook County Assessor's Office). When the residual is positive, it implies that the CAAO's assessment undervalues the property. In other words, the property is assessed at a lower value than its actual market price, potentially resulting in lower property tax bills.





---

## 0.5 Question 7b

Reflecting back on your exploration in Questions 6 and 7a, in your own words, what makes a model's predictions of property values for tax assessment purposes "fair"?

This question is open-ended and part of your answer may depend upon your specific model; we are looking for thoughtfulness and engagement with the material, not correctness.

**Hint:** Some guiding questions to reflect on as you answer the question above: What is the relationship between RMSE, accuracy, and fairness as you have defined it? Is a model with a low RMSE necessarily accurate? Is a model with a low RMSE necessarily "fair"? Is there any difference between your answers to the previous two questions? And if so, why?

Fairness, in many respects, remains an ideal that we continuously strive for. In practice, our primary objective is to assess properties accurately, ensuring that they are taxed in proportion to their actual values. This includes refraining from undervaluing or overvaluing any property. The fundamental principle is to follow equitable taxation, treating every homeowner equally in terms of their property.

