0.1 Question 0

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 value be high or low.

1. Real estate company who builds and sells the houses, who wants the value to be high as profits.
2. House buyer who wants to find an optimal house with their budget, who wants the value to be low as cost.
3. Real estate agents who buys the house from owners and sells it to others, who wants the value to be low when buying and wants it to be high when selling.

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.

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 of four scenarios seem to be unfair to me because the home shall be assessed and sold for the same price in order to be fair. There is no reason for any excessive profits to exist or go to any party.

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)

There fairness of assessing property tax is not achieved. Poorer people pay more property tax.

Primary causes: 1. Real estate and racial/class inequality in the United States. 2. Human impact cause the assessment was made by human.

0.1.4 Question 0e

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

Because due to possible racial discrimination, the County tended to overassess non-white property owners' houses so they will have higher tax to pay.

0.2 Question 2

Without running any calculation or code, complete the following statement by filling in the blank with one of the comparators below:

$$\geq$$
$$\leq$$
$$=$$

Suppose we quantify the loss on our linear models using MSE (Mean Squared Error). Consider the training loss of the 1st model and the training loss of the 2nd model. We are guaranteed that:

Training Loss of the 1st Model_____Training Loss of the 2nd Model

$\geq$ since the first model has less feature than second model so its complexity is smaller than the second one leading to that the pattern of data is less caught by first model. So there will be more or equal training loss of model one than model 2.

0.3 Question 6

Let's compare the actual parameters (θ_0 and θ_1) from both of our models. As a quick reminder,

for the 1st model,

$$\text{Log Sale Price} = \theta_0 + \theta_1 \cdot (\text{Bedrooms})$$

for the 2nd model,

$$\text{Log Sale Price} = \theta_0 + \theta_1 \cdot (\text{Bedrooms}) + \theta_2 \cdot (\text{Log Building Square Feet})$$

Run the following cell and compare the values of θ_1 from both models. Why does θ_1 change from positive to negative when we introduce an additional feature in our 2nd model?

```
In [22]: # Parameters from 1st model
         theta0_m1 = linear_model_m1.intercept_
         theta1_m1 = linear_model_m1.coef_[0]

         # Parameters from 2nd model
         theta0_m2 = linear_model_m2.intercept_
         theta1_m2, theta2_m2 = linear_model_m2.coef_

         print("1st Model\n 0: {}\n 1: {}".format(theta0_m1, theta1_m1))
         print("2nd Model\n 0: {}\n 1: {}\n 2: {}".format(theta0_m2, theta1_m2, theta2_m2))
```

```
1st Model
 0: 10.571725401040084
 1: 0.4969197463141442
2nd Model
 0: 1.9339633173823696
 1: -0.030647249803554506
 2: 1.4170991378689644
```

Because second model has an additional feature `Log Building Square Feet` whose value is relatively larger and has more prediction power (building sqft is more directly related to the sales price). In this way, more positive coefficient was assigned to `Log Building Square Feet` thus `Bedrooms` has to be less in order to make accurate predictions.

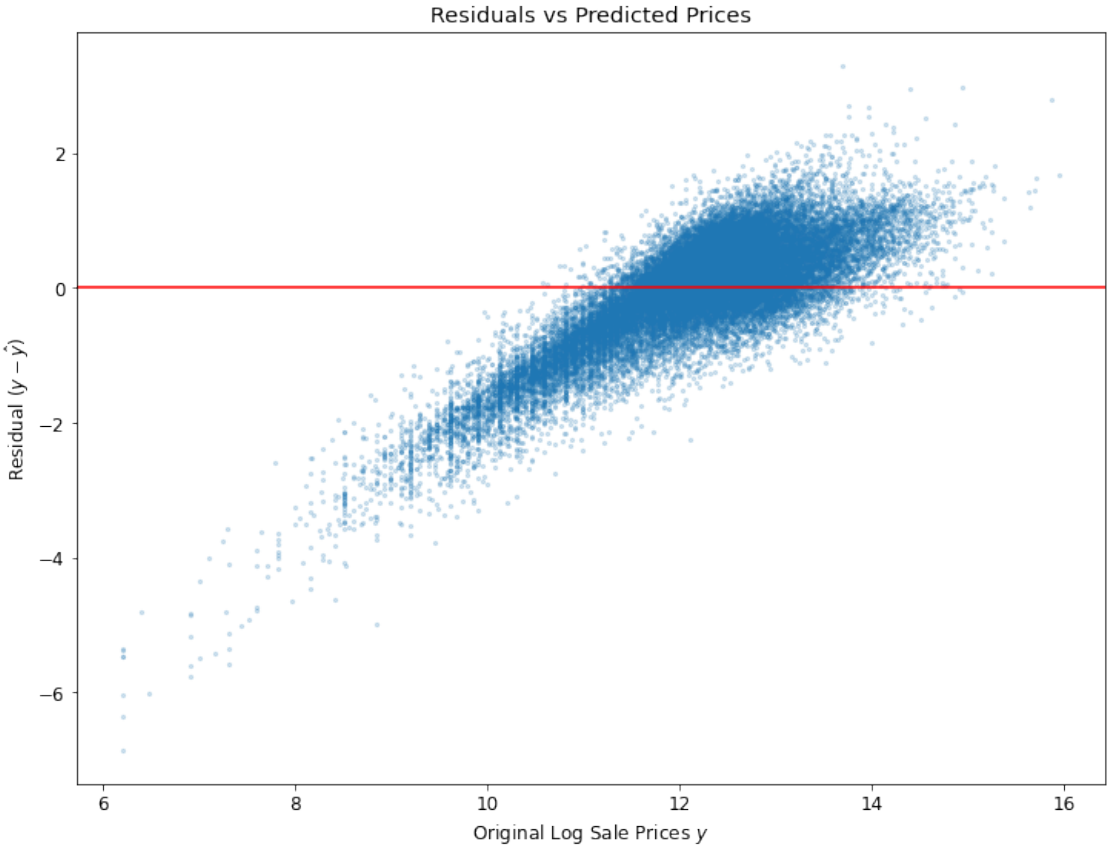
0.4 Question 7

0.4.1 Question 7a

Another way of understanding the performance (and appropriateness) of a model is through a plot of the model 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 **test data**. You should also ensure that the dot size and opacity in the scatter plot are set appropriately to reduce the impact of overplotting.

```
In [23]: plt.scatter(y_test_m2, y_test_m2 - y_predicted_m2, s = 5, alpha=0.2)
         plt.ylabel("Residual $(y - \hat{y})$")
         plt.xlabel("Original Log Sale Prices $y$")
         plt.title("Residuals vs Predicted Prices")
         plt.axhline(y = 0, color='r');
```



0.5 Question 9

When evaluating your model, we used root mean squared error. In the context of estimating the value of houses, what does error mean for an individual homeowner? How does it affect them in terms of property taxes?

The error means the difference between the CCAO's assessment and actual market value, which also generates more/less property tax than a fair tax they should pay. If the assessment is made higher than the market value, they will pay more unfair tax accordingly. If the assessment is made smaller than the market value, they will pay less unfair tax accordingly.

In the case of the Cook County Assessor's Office, Chief Data Officer Rob Ross states that fair property tax rates are contingent on whether property values are assessed accurately - that they're valued at what they're worth, relative to properties with similar characteristics. This implies that having a more accurate model results in fairer assessments. The goal of the property assessment process for the CCAO, then, is to be as accurate as possible.

When the use of algorithms and statistical modeling has real-world consequences, we often refer to the idea of fairness as a measurement of how socially responsible our work is. But fairness is incredibly multifaceted: Is a fair model one that minimizes loss - one that generates accurate results? Is it one that utilizes "unbiased" data? Or is fairness a broader goal that takes historical contexts into account?

These approaches to fairness are not mutually exclusive. If we look beyond error functions and technical measures of accuracy, we'd not only consider *individual* cases of fairness, but also what fairness - and justice - means to marginalized communities on a broader scale. We'd ask: What does it mean when homes in predominantly Black and Hispanic communities in Cook County are consistently overvalued, resulting in proportionally higher property taxes? When the white neighborhoods in Cook County are consistently undervalued, resulting in proportionally lower property taxes?

Having "accurate" predictions doesn't necessarily address larger historical trends and inequities, and fairness in property assessments in taxes works beyond the CCAO's valuation model. Disassociating accurate predictions from a fair system is vital to approaching justice at multiple levels. Take Evanston, IL - a suburb in Cook County - as an example of housing equity beyond just improving a property valuation model: Their City Council members recently approved reparations for African American residents.

0.6 Question 10

In your own words, describe how you would define fairness in property assessments and taxes.

Fairness here should be the assessments equal to the real market price and the tax should be collected proportionally: Higher value of property comes with higher taxes, and lower value property comes with lower taxes in a reasonable proportion.

0.7 Question 11

Take a look at the Residential Automated Valuation Model files under the Models subgroup in the CCAO's GitLab. Without directly looking at any code, do you feel that the documentation sufficiently explains how the residential valuation model works? Which part(s) of the documentation might be difficult for nontechnical audiences to understand?

I don't feel like the documentation sufficiently explains because the structure of this documentation is sort of messy. I don't get each folders' meaning in a glance. The glossary in the hand book is especially hard to read due to lots of equations.