0.1 Question 0: Human Context and Ethics

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. People who want to buy the house or the house owner. They just need or want to know that.
2. The porperty company may interested in this question because it helps thier bussniss. They could do bussiniss strategy analysis or something.
3. The goverment may interested in this question because it helps them to esstimate how much tax is resonable.

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 are unfair.

A: House owner will pay more tax. And when they sell their house they may sell for the same price as others, who have same level properties or at same county, but pay less tax than our poor house owner.

B:Similar to A situation, but oppsite. House owner will pay less tax than others, who have same level properties or at same county. It is unfair for other citizens.

C:Regressive tax will show up. Peopol who have inexpensive properties will pay more tax. For rich people they pay less tax.

D:There is a progressive tax. Peopol who have expensive properties will pay more tax than they should pay. It is unfair too.

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)

Cook County Assessor's Office systematically overvalues inexpensive properties and undervalues expensive properties. That resulting in a regressive property tax system. People have less expensive properties pay more disproportionate amount of taxes and peopel have more expensive properties pay less tax.

The reason for this mainly because richer people can afford tax lawyers and lawyers can help them to lower the assessed value of their home. On the other hand people have less properties normally don't spend money on thoes lawyers.

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?

Even they find there is disproportionate amount of taxes that they have to pay, unlike white wealther, they not always appealed their assessments. This may because they don't have time and money to deal with it. They may not affort expensive lawyer fee. So the situation is hard to get better.

0.2 Question 2a

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

It should be greater or equal sign( ). We are only considering the training loss. When adding more feature, it increase the complexity of model and fit more on training data. So the training loss will be less and less if not same.

0.3 Question 3b

You should oberseve that θ_1 change from positive to negative when we introduce an additional feature in our 2nd model. Provide a reasoning why this may occur. **Hint:** which feature is more useful is predicting `Log Sale Price`?

When we add add one more feature(Log Building Square Feet) on 2nd model, it is more useful is predicting Log Sale Price. Or say the feature 'Log Building Square Feet' is more important(weight more) than 'Bedrooms' feature. Then Bedrooms feature is decreased it's importance(weight).

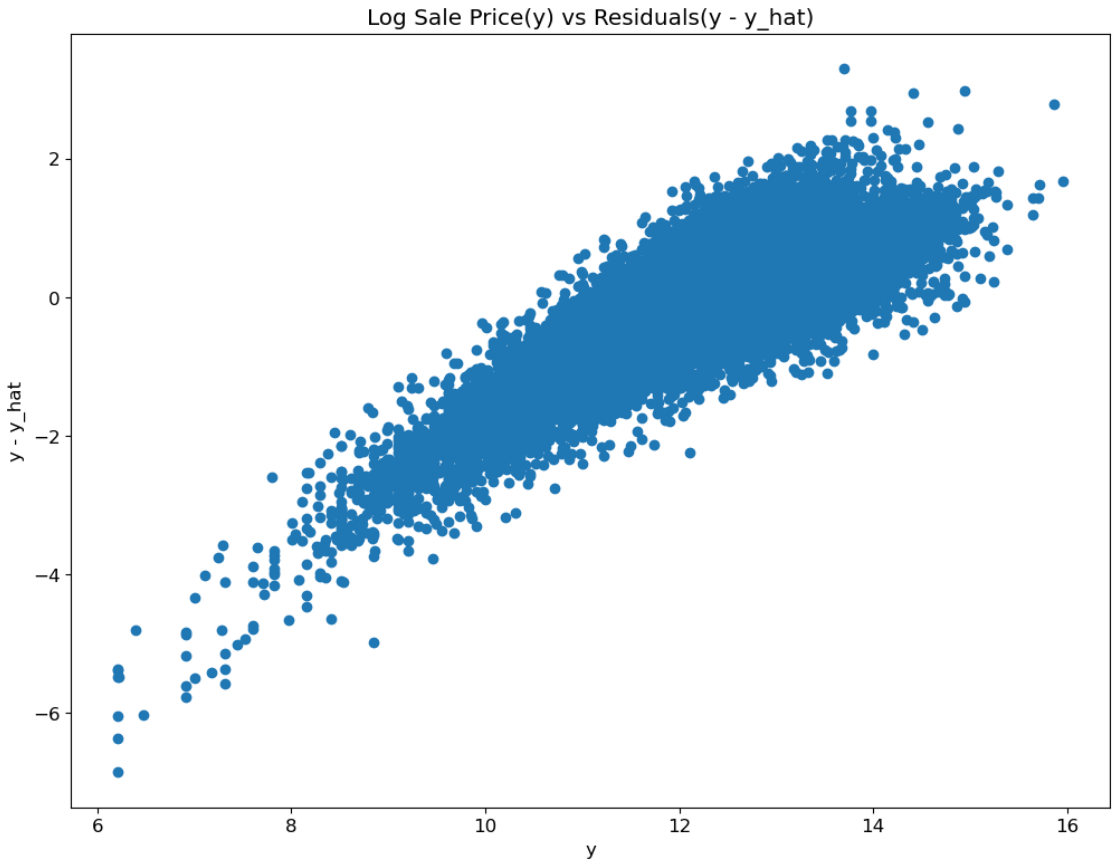
0.4 Question 3c

Another 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 diffult 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 [298]: plt.scatter(x=y_valid_m2, y=y_valid_m2 - y_predicted_m2)
          plt.title("Log Sale Price(y) vs Residuals(y - y_hat)");
          plt.xlabel("y")
          plt.ylabel("y - y_hat")
```

```
Out[298]: Text(0, 0.5, 'y - y_hat')
```



0.5 Question 5

In building your model in question 4, what different models have you tried? What worked and what did not? Brief discuss your modeling process.

Note: We are not looking for a single correct answer. Explain what you did in question 4 and you will get point.

I removed the outliers at the begining. I used linear regression, but RMSE is too big. So after that I tried adding features. I used OneHotEncoder function to manage some columns. I used logtransform and powered some columns. By adding all thoese features It ends up a good EMSE range.

0.6 Question 6 Evaluating Model in Context

0.7 Question 6a

When evaluating your model, we used root mean squared error. In the context of estimating the value of houses, what does 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.

The residual is the difference between the predicted value of a home and its actual value. If the residual is positive, it means that we overvalues actual value this may cause regressive taxation. On the other hand, if it is negative, we undervalues its actual value, progressive taxation may happen.

0.8 Question 6b

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

Fairness in property assessments and taxes is the amount of tax owed by property owners accurately reflects the value of their property, and the assessment process is unbiased. The assessment process should be regular update.Transparency and accountability would be important components of a fair system.

0.9 Question 6c

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 think it is sufficiently explains how the residential valuation model works. Because it contains many parts and each parts have data and explaination. The terminolgies such as "AWS" and "LEGDAT" might be difficult for nontechnical audiences to understand.