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.

The dealers and potential buyers could all be interested, even banks to look for a loan estimate. For potential buyers, they want a price as low as possible for the house they have in mind. For dealers, they have a preference of net profit = selling price - buying price as high as possible. Whereas the bank wants to have as accurate price as possible to assess the risk of funding a loan or morgage.

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.

B,C are unfair. For B, it is mostly considered unfair because people selling the house would be expecting to sell the house higher than the actual value of the house, with this assumption it is hard to estimate the value of a house as lower than its actual price. For C, it is an unfair systems that depreciates any housing transactions and leads to the situations that not one house would be sold in mutual content. It is highly unlikely..

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 regressive taxation gives burden to the black homeowner and disobeys the intentions of giving tax. the corruption and bias in the tax system design, it distributed tax burdens unevenly, placing a disproportionate burden on lower income homeowners, this is caused by the appeal process and how richer residents gets more exposure to it than the less rich residents, which creates a discrepency when assessing the properties, and result in an unfair taxation. one of the cause could be the lack of transparency and corruption since Berrios is close to the tax lawyer industry, and how the system develops in favor of the richer residents.

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?

many non-white property owners had less access to the appeal process to correctly assess their home value, which leads to the lag in the update of their property value and the taxation reflects on this to apply more burden to them. it is also because of the particular arrangement of the property assessment system that enabled the wealthy and privileged to more easily and successfully challenge their assessments.

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

>=, by adding another feature, we are introducing a new feature that can capture more variabilities of the sale price, similar idea applies to neural network in general, by compilating more features/functions, we can better capture the previously hard to capture outliers, since the complexity increases, it is normal the bias decreases, especially when we only have training errors here.

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 [23]: # 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 Log Building Square Feet has more predictive effect than Bedrooms, so the effect of Bedrooms is weakened here.

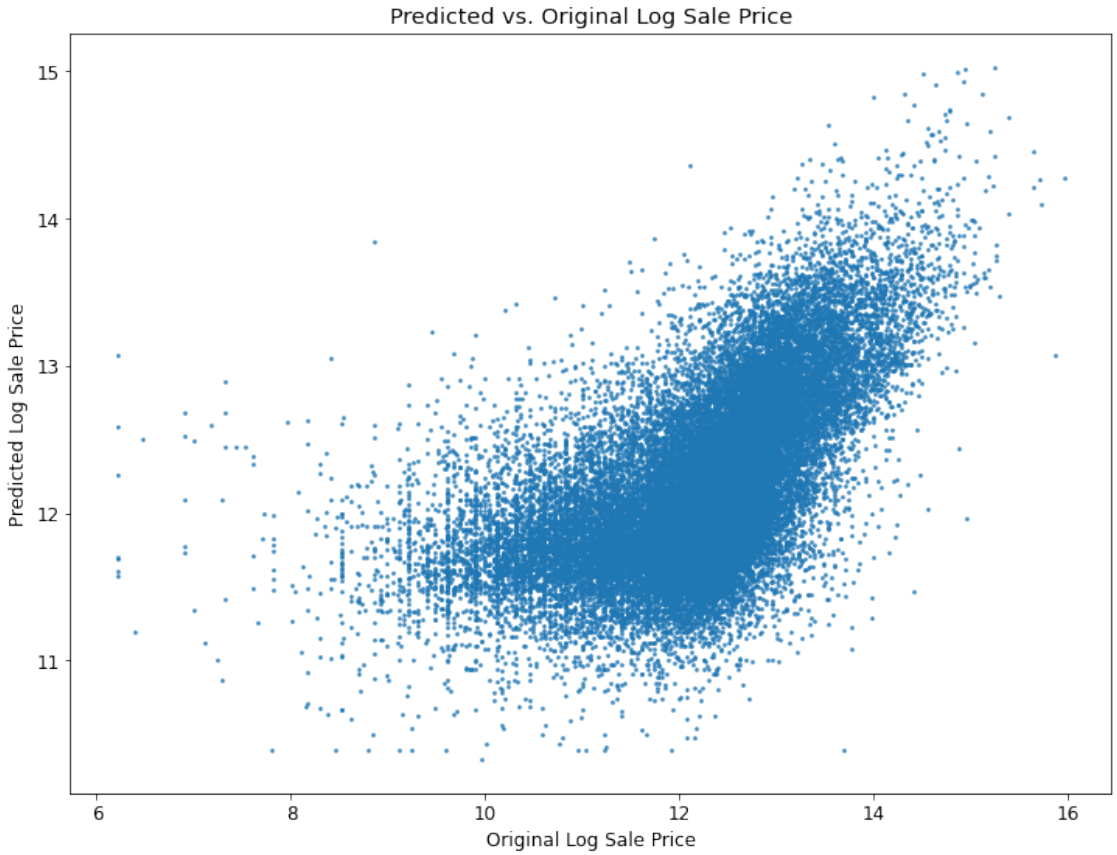
0.4 Question 7

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 [24]: plt.scatter(x=y_test_m2,y=y_predicted_m2,s=3,alpha=.75)
         plt.xlabel('Original Log Sale Price')
         plt.ylabel('Predicted Log Sale Price')
         plt.title("Predicted vs. Original Log Sale Price");
```



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?

when running regression, error is calculated based on the sale price compared to the government's assessment, which means that for a error > 0, property is being not fairly valued and is most likely priced lower than actual, this then by taxation leads to a lower tax, conversely, if it is negative then we have a higher tax.

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.

I would define fairness in favor of the people, I believe the government should encourage or at least maintain the housing value for people to actively engage in the trading of houses so that people can still afford a house and try their best to live a better life, under this assumption a recession taxation is unfair and I believe an equal exposure to all property owners for regular appeal is important. It is only fair if it favors most people, not the certain group of residents. As a student I can see the data analysis has its use but politics is where the true impact happens, so if there is no response data is a powerful tool at least to expose the unfairness to the public, which is something that people did not have before.

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 think it does a better job now with their transparency, at least in the readme doc I can see what features are being calculated into the algorithm, but not looking at the code I am still not aware of what the actual emphasis is and if it is fair and impartial for features like location or income level information. It is already hard enough for nontechnical residents to .