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. House owner - high
2. Potential buyer - low
3. Real estate agent - high so that there is more commission and more sellers, low so that there are more buyers. Overall high

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.

A. Homeowners would have to pay more tax than they should
B. Homeowners would sell their house at a lower price or believe they have less asset than they really do
C. Homeowners of inxpensive propeties will experience A and those of expensive will experience B
D. Homeowners of inxpensive propeties will experience B and those of expensive will experience A

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 problem is that properties were not valued properly and causing properties owners to be paying incorrect property taxes. One of the causes for this issue was because the wealthy could afford lawyers and therefore had more success in appealing their assessments, which led to undervaluations of the wealthier areas and consequently lower taxes for the wealthy. On the other hand, the less wealthy did not appeal as much and consequently had overvaluations and higher taxes. The ones that paid less tax tend to be white as they tend to be more wealthy and those that paid more tend to be minoirities. Moreover, this demographic geographic segregation (i.e. disproportionally white/black neighborhoods) can be traced back to structural racism. Due to racism in the early credit system, it was more difficult for black people to get a mortgage and buy a house as the government did not insure banks to issue mortgages in areas that were deemed risky (i.e. redlining). Moreover, the real estate industry used race as one of the factors for property valuation, which further increased segregation

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?

This is because the wealthy could afford tax lawyers and therefore had more success in appealing their valuations. This led to undervaluations in the wealthier areas and consequently lower taxes for the wealthy. On the other hand, the less wealthy could not appeal as much due to not being able to afford a lawyer or not being aware of this practice. Consequently, their properties were overvalued and they paid higher taxes. The burden fell disproportionally on the non-whites as they tend to be less wealthy

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

$$\geq$$

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 [29]: # 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
```

Adding a new feature simply creates a new regression, which in turn has new coefficients that model the response variable better. It just so happens that by adding Log Building Square Feet as a new feature, the coefficient of Bedrooms (theta1) went from positive to negative.

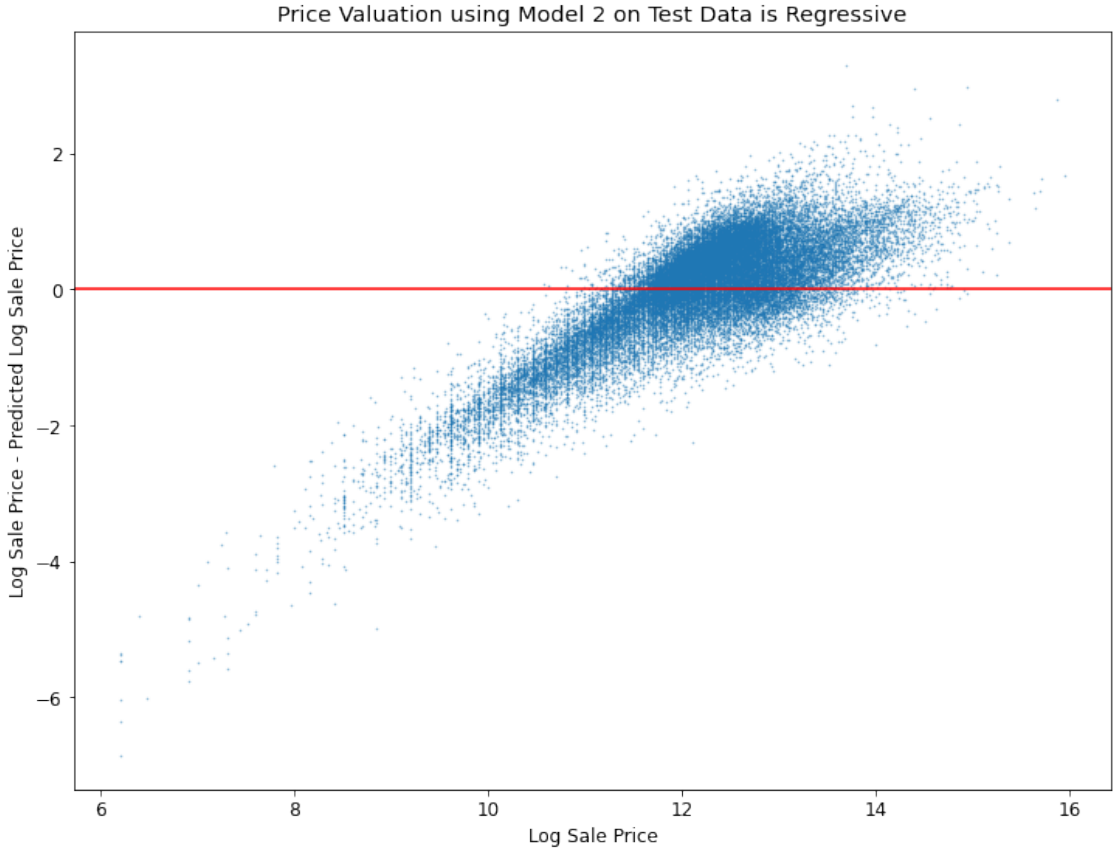
0.4 Question 7

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 [30]: plt.scatter(y_test_m2, y_test_m2 - y_predicted_m2, s = 0.3, alpha = 0.5)
         plt.axhline(0, color = "r")
         plt.title("Price Valuation using Model 2 on Test Data is Regressive")
         plt.ylabel("Log Sale Price - Predicted Log Sale Price")
         plt.xlabel("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?

For an individual owner, the RMSE is the typical difference between the actual value of his/her house and the value the county valuates it at. In terms of property taxes, it would correspond to the amount of tax one pays more or less than he/she should. Namely, if the error is positive for a specific house, the owner would pay less tax since the house is undervalued, and if the error is negative, he/she would pay more as it is overvalued.

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 believe that taxes should be equal for all factors other than the amount one spends or earns. As far as I know, these are the only two means of which taxes are generated. Therefore, I don't think factors like race, gender, or anything else should play a part in how much tax one has to pay. However, I believe that a progressive tax is the most fair, since it moreorless guarantees everyone at least enough to survive. That is, people who earn a lot can pay more and still live comfortably, while those that can barely make ends meet should pay as little as they need to survive. This ensures some sort of general/social living standard. In terms of property assessments, I believe that a house should be priced based on supply and demand. That is, instead of having someone (i.e. a government official) to valuate it, it should be determined by how much someone (i.e. anyone from the public) will be willing to pay for it. The forces of supply and demand makes sure that there is no bias in terms of the system, although the nature of it may be suscpetible to other bias that exists in society per se. Nonetheless, in my opinion, that is the most fair. If using economic forces is impractical, something like a model we've just built would be a good alternative as long as the raw data are inputed without bias and the residuals don't exhibit a pattern.

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?

The documentation does not seem to talk about much about how the model determines a property's value. Instead, it mostly talks about how to use the model as an API and how the model is structured. In other words, it tells us how the steps the model takes to make its decisions, but not the rationale behind those decisions. It is quite difficult for anyone who does not have coding experience to understand really anything, espcially the tutorial for implementation and the lack of sample code. In general, the docs are very non-user-friendly and as someone who has not used R Shiny, I have practically no idea how to use this in my own projects. The main issue with this is, although the idea of transparency is stressed, if the general public cannot understand how the model works or even worse, if the model even exists (i.e. how to get to this page), there is really no transparency.